Forecasting Short Term Trucking Rates

By

Xiwen Bai

B.S., Maritime Studies, Nanyang Technological University, 2014

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

Master of Engineering in Supply Chain Management

at the

Massachusetts Institute of Technology

June 2018

Signature of Author..…………………………………………………………………………………
Master of Engineering in Supply Chain Management Program, Engineering Systems Division
May 10, 2018

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

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

Forecasting Short Term Trucking Rates

By

Xiwen Bai

Submitted to the Engineering Systems Division in Partial Fulfillment of the Requirements for the Degree of Master of Engineering in Supply Chain Management

Abstract

Transportation costs constitute an important part of total logistics costs and have a dramatic impact on all kinds of decisions across the supply chain. Accurate estimation of transportation costs can help shippers make better decisions when planning transportation budgets and can help carriers estimate future cash flows. This study develops a forecasting model that predicts both contract and spot rates for truckload transportation on individual lanes for the next seven days. This study considers several input variables, including lagged values of spot and contract rates, rates on adjacent routes and volumes. The architectural approach to short-term forecasting is a neural network based on Nonlinear Autoregressive Models with eXogenous input (NARX) models. NARX models are powerful when modelling complex, nonlinear and dynamic systems, especially time series. Traditional time series models, including autoregressive integrated moving average (ARIMA), are also used and results from different models are compared. Results show that the NAR model provides better short-term forecasting performance for spot rates than the ARIMA model, while the ARIMA model performs slightly better for contract rates. However, for a longer-term forecast, the NARX model provides better results for contract rates. The results from this study can be applied to industrial players for their own transportation rate forecasting. These results provide guidelines for both shippers and carriers regarding what model to use, when to update the model with new information, and what forecasting error can be normally expected from the model.

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

Acknowledgements

First and foremost, I would like to express my sincere gratitude to my supervisor, Dr. Chris Caplice, for the continuous support of my research, for his patience, enthusiasm, motivation, and immense knowledge. His guidance helped me in all the time of research and writing of this thesis.

I would like to thank my family for all their love and care throughout the journey and my life in general. For my parents who have consistently supported me in all my pursuits. And for my loving, encouraging, patient and supportive husband Yilin, who has given me unconditional and faithful support throughout the years. This thesis is dedicated to you all.

Table of Contents

List of Figures

List of Tables

1 Introduction

Transportation is a service within a supply chain that moves raw materials, product parts, and finished products from point to point. It links many supply chain activities from the extraction of natural resources, the fabrication of industrial, commercial, and consumer products to the distribution of final products to wholesalers, retailers and consumers. The US Department of Commerce recorded that in 2015, $1.5 trillion was spent on logistics and transportation, which constitutes 8% of annual gross domestic product (GDP) (CFRA, 2017). Nowadays, because manufacturers, wholesalers and retailers put more emphasis on lean production and inventory minimization, the roles played by transportation providers have become more significant than before.

From a shipper's perspective, up to 50% of total logistics costs can be attributed to transportation costs. For that reason, transportation cost has a dramatic impact on all kinds of decisions across the supply chain. Accurate estimation of transportation costs could aid not only shippers' decision-making process towards better transportation budget planning, but also decisions across the entire supply chain regarding facility location, vehicle routing, economic order quantity (EOQ) and inventory replenishment policies (Swenseth & Godfrey, 1996). As a result, transportation cost is often incorporated in lot sizing or inventory replenishment decisions (Burwell, Dave, Fitzpatrick, & Roy, 1997; Carter & Ferrin, 1996; Mendoza & Ventura, 2009; Swenseth & Godfrey, 2002).

The objective of this project is to develop a forecasting model that predicts both contract and spot rates for truckload transportation on individual lanes for the next seven days. More accurate rate forecasting could aid the decision-making process for carriers with respect to determining future cash flows. It also provides useful guidance for third-party providers and shippers regarding potential price fluctuations and resulting risks. Furthermore, such price information could be used for budget planning by large companies.

In this study, the architectural approach for short-term forecasting is a neural network based on Nonlinear Autoregressive Models with eXogenous input (NARX) models, also known as NARX recurrent neural networks (Lin, Horne, Tino, & Giles, 1996). NARX models are powerful when modelling complex, nonlinear and dynamic systems, especially time series. Compared to other networks, learning is more effective, the convergence rate is much faster and generation is better for NARX models (Gao & Er, 2005; Lin et al., 1996). Traditional time series models including autoregressive integrated moving average (ARIMA) are also used and results from different models are compared.

This study has made several contributions to both academics and practice. Firstly, this research has for the first time introduced the hybrid model (NAR and NARX) into the transportation forecasting field. Such a model combines the advantages of both neural networks (model complex and non-linear relationships) and time series models (model series dependencies). It is proven to have good forecasting abilities in many fields. Secondly, this study introduces the framework of when and how to update the model with new information. TL rates are dynamically changing and when new rate information comes in, a decision must be made as to whether to update the model with such information. To help address this issue, this paper introduces the criterion of error reduction rates. Then, the best time to update the model will be determined by the threshold set for the error reduction rate. Last but not least, the model can be applied to industrial players for their own forecasting. It is found that NARX models provide satisfactory predicting abilities for contract rates, and decent but not highly stable forecasts for spot rates.

In this study, the scope for rate forecasting has been limited to long-haul dry van full truckload (TL) shipments. Long-haul is defined as any shipments over 250 miles. TL constitutes the majority (97% according to Costello (2013)) of the trucking industry in the US. Thus, forecasting TL rates has more practical significance than Less than Truckload (LTL) rates. In the forecasting part, one high-volume lane is used as a sample lane for forecasts. However, the same methodology can be applied to all different lanes.

This thesis is organized as follows. Section 2 reviews past literature on transportation rate forecasting and identifies major literature gaps. The methodology and dataset used are presented in Section 3. Section 4 describes results of different models and the best fitted models for both spot and contract rates will be selected. The implications are also discussed. Last but not least, Section 5 provides a conclusion and Section 6 points out future research directions.

2 Literature Review

This section provides an overview of the US trucking industry and reviews rate forecasting in the trucking industry. Trucking rate forecasting has shifted gradually from approximated rate function using distance and weight as inputs to actual rate forecasting. However, actual truck rate forecasting studies have been limited due to the high number of lanes available, which means increased model complexity. Furthermore, no study has considered the interactions between spot and contract rates, between rates for a particular lane and its adjacent routes, and between rates and volumes, to forecast TL rates. Methodologies used for ocean transportation forecasting have also been compared. As one mode of transportation, ocean freight forecast has been widely researched in the literature. In a tramp shipping market, a ship has no fixed routing or schedule. A ship can load any cargo from any port to any port. The truckload (TL) and tramp shipping markets share similar characteristics, for example, they both have steady contract markets and volatile spot markets. By studying the methods used in ocean transportation rate forecasting, potential references can be drawn to TL rates.

2.1 The US trucking industry

The trucking industry is regarded as the lifeblood of the US economy (Costello, 2013). In 2015, trucks moved roughly 70% of the United States' freight by weight, which amounts to 10.49 billion tons (ATA, 2017). In terms of revenue, trucking earned \$726.4 billion in 2015, constituting 81.5% of the nation's freight transportation revenue (ATA, 2017). A truck can haul almost everything from raw materials to finished goods, linking manufacturers, distributors, retailers and households.

The trucking industry can be divided into two primary categories: for-hire and private carriers. For-hire carriers predominantly lift freight for other companies, while private carriers are operated by those companies that have a fleet of trucks to support their own business. In 2012, according to the estimation made by the American Trucking Association (ATA), for-hire and private carriers hauled respectively 54.5% and 45.5% of total truck tonnage in the United States (CFRA, 2017).

The for-hire industry is a competitive and fragmented market with around 586,000 companies at the end of 2015, even with some consolidation over the years. Within the truckload industry, participants vary in size from one truck to more than 10,000 trucks, with most being small businesses (Costello, 2013). In 2015, the top 20 for-hire carriers accounted for only 11% of the total TL industry (American Trucking Trends, 2016). Compared to the high operating margins for the rail industry (over 20%), intense competition leads to lower margins (less than 5%) for the trucking industry (American Trucking Trends, 2016).

2.1.1 Trends in the US trucking industry

Before 1945, the rail industry hauled more freight tonnage than the trucking industry. However, since then, a few developments have pushed the trucking sector to today's leading position. First, the construction of the Interstate Highway System between the late 1950s and the mid-1970s enabled trucks to move more efficiently across most parts of the country. Second, the deregulation of the motor carrier sector in 1980 opened up competition in the industry (Özkaya, Keskinocak, Roshan Joseph, & Weight, 2010). Since deregulation, freight rates have been pushed down due to competition and excess capacity (Baker, 1991). Before 1980, the increase in freight rates tracked the changes in inflation closely. However, since the deregulation, TL rates have no longer been tied to specific commodities. From 1980 to 2000, real revenue per mile dropped by 47.2%. Although real revenue per mile recovered by around 30% from 2000 to 2012, it is still significantly lower than the level in 1980, as shown in Figure 2.1.

Figure 2.1 Real average revenue per mile (1980–2012)

Source: Costello (2013).

The third trend is the Just in Time (JIT) practice in the supply chain field, which enables supply chain participants to hold less inventory. The flexibility of trucks enables the JIT practice, and as JIT practices became more popular, the industry quickly expanded its market share in the 1980s. In certain industries such as auto manufacturing, trucks could act as warehouses with limited inventories.

2.1.2 Truckload (TL) and Less than Truckload (LTL) market

The for-hire market can be further classified into the Truckload (TL) and Less than Truckload (LTL) market. The threshold between TL and LTL is 10,000 pounds. Any shipments weighing 10,000 pounds or less are considered as LTL. TL carriers typically move from one point to another point directly hauling a full truckload for one customer. On the other hand, LTL carriers will go through a series of intermediate terminals before heading to the final destination, operating more like a hub and spoke system (Costello, 2013). In 2012, TL hauls 97% of total for-hire tonnage and LTL transported 3% (Costello, 2013). TL and LTL have different pricing systems. In this thesis, we primarily focus on the TL market due to its dominant market size.

2.1.3 Trailer type

Different trailer types carry different products. The main types include box trailer, flatbed, refrigerated/temperature-controlled trailer and tank. The box trailer (also referred to as a dry van) is the most common type of trailer. According to Costello (2013), around 45% of the tractors pull dry vans. The flatbed trailer is the next most common type (15% of tractors pulling), which is used to carry construction materials or large machinery, as the trailer does not have a side wall or ceiling. Food and medicines are normally hauled by temperature-controlled (TC) trailers. 10% of tractors pull this type of trailer. The last type is tanks, which are used to haul refined oil products or chemicals in the liquid form.

2.1.4 Contract and spot markets

Shippers can cover their freight transportation requirements by two types of services: through long-term contracts or on the spot market (Garrido, 2007). Shippers in most cases use contract carriers to haul their truckload products. The contract is typically a one-year commitment, which consists of origin/destination, service requirement, volume and any other factors that affect the price. A contract rate is a lane-rate with a non-binding price expectation of volume commitment, which means the shipper does not provide a minimum volume guarantee for the carrier. The contract typically lasts for a year but can vary between shippers. For each lane (unique origin and destination pair), the shipper uses a route guide or a list of carriers with different characteristics, such as price, on-time delivery, familiarity with existing operations and how easy it is to do business with (Sheffi, 2004). Often, price is the primary consideration for the shippers. The carriers agree to haul the products on a specified route for a specified price.

The shipper will rank carriers based on preferences. When the shipper requires transportation for a load, it will issue a tender to the first carrier according to the route guide. The carrier will then either accept or reject the tender. Although both parties have agreed in advance on the contract rate, the carrier may not have the truck available for load at the place where the shipper requests. If the carrier rejects the tender, the shipper will source from the other carriers based on the list, until the tender is accepted.

On occasions when the contract rate fails, when a tender is not accepted or the lane/rate does not exist in the route guide, the shipper needs to go to the spot market to obtain a rate. If shippers experience a surge in freight volume that could not be covered sufficiently under the long-term contracts, they are also forced to procure additional capacities on the spot market. The rate is decided based on the market condition at the time of the transaction (Sheffi, 2004).

The contract market, while providing stability for the shipper, has less flexibility compared to the spot market, thus being less responsive to changes in fuel prices, new technologies or changing market conditions. Such drawbacks of the contract market could reduce the overall trucking industry efficiency. On the other hand, the spot market could decrease such inefficiencies to a certain extent by being overly responsive to market changes and is not consistent for budgeting. However, it also results in higher levels of volatility and fluctuations compared to contract rates (Garrido, 2007).

The choice between contract and spot market could be influenced by the number of carriers available for a particular lane, as discussed by Hubbard (2001). He concluded that with higher numbers of carriers available, spot rates could be pushed down by increasing competition between the carriers. Thus, the likelihood of procuring on the spot market is higher than the situation where there exists a limited number of carriers and spot rates are typically higher due to less competition and less capacity, which is referred to as market liquidity.

2.2 Transportation rate forecasting

Transportation rate forecasting is typically made for either spot freight rates or contract rates depending on the type of decisions to be made. For example, for a carrier looking for an immediate load in the spot market, spot rates are more of its concern. Contract rates have been more frequently studied than spot rates due to their use for long-term decisions. Earlier research since the 1990s has mostly focused on approximating contract freight rates and finding appropriate rate function forms in terms of distance for TL, or distance and weight for LTL. Different forms have been proposed, for example, linear function or power function (Ballou, 1991; Swenseth & Godfrey, 1996; Tyworth & Ruiz-Torres, 2000). More recently, there has been an increasing trend in the forecasting of freight rates using more route- and market-specific variables (Budak, Ustundag, & Guloglu, 2017; Kay & Warsing, 2009; Özkaya et al., 2010).

Several approximated freight rate functions have been investigated by researchers to emulate the actual freight rates (Swenseth & Godfrey, 1996). Freight rates are often defined as a function of distance and weight. By using such function, a simple market rate for a lane can be calculated for a given origin and destination. The functional form varies across literature. Broadly speaking, freight rate function can be classified into two categories: discrete and continuous functions. TL rates are typically based on per-mile, while LTL rates are reported in cost per hundredweight (CWT). Swenseth & Godfrey (1996) evaluated five different forms of rate functions, including a constant rate per CWT (linear), a constant price per shipment (inverse), a proportional function, an exponential function and an adjusted inverse function. The accuracy of each function is evaluated by comparing the rate from the functional form and the actual rate. Mendoza and Ventura (2009) used the proportional function proposed by Swenseth & Godfrey (1996) and the power function by Tyworth & Ruiz-Torres (2000), when computing the freight rates in evaluating inventory replenishment and supplier selection decisions.

However, such approximated freight rate functions, which only incorporate distance for TL or distance and weight for LTL, have major limitations. They lose the accuracy of the actual rates, and do not contain any information on the current market conditions, such as capacity availability, carrier's availability and differences, and general economic situations. Many times, they do not represent the inherent nature of TL transportation. Furthermore, such functions do not take into account route-specific rates. Such a formula generally assumes a uniform rate for all routes with the same length. However, in reality, due to geographical difference and market dynamics changes, rates differ a lot across different routes. Neglecting route specific information would result in inaccurate rate forecasting.

Recently, increasing research attention has been given to the prediction of actual freight rates both in the TL and LTL markets. Instead of pure distance and weight factors, scholars have incorporated more lane- and market-specific variables in order to depict the actual freight market conditions more accurately. Smith, et al. (2007) collected US LTL carrier data and examined the revenues generated from different customers on different lanes by regressions. They then compared the difference between estimated and actual shipment rates and pointed out that when the two rates are systematically different, there may exist opportunities for re-negotiation. However, the limitation of this study exists, as the authors only used a single carrier dataset. In fact, for each lane, it is comprised of different shippers using various carriers and the rate is a determined by all different carriers and shippers. Caldwell & Fisher (2008) used ordinary least square (OLS) regression to determine the relationship between TL rates and various factors, including distance, origin and destination of the load, tender rejections, economics of scale, carrier size among others. They identified a signifcant positive relationship between transportation costs and tender rejections. Kay & Warsing (2009) adopted non-linear regression to estimate LTL rates utilizing a set of explanatory variables including load density, shipment weight, and distance between O-D pair. Özkaya, et al. (2010) used multiple regressions to model the US Less-than-Truckload (LTL) market rates. In this study, the authors considered both tangible and intangible market (not captured in the dataset) factors. Tangible factors included weight, distance, freight class, carrier type and origin/destination. Intangible factors were captured by an expert survey and included freight desirability, negotiation power of the shipper, economic value estimates and perceived freight class.

The traditional regression methods are less effective in modelling complex short-term fluctuations which are often found in spot market rates. Budak, et al. (2017) compared artificial neural network (ANN) and quantile regression methods in predicting TL spot market price in Turkey. They considered both route-specific forecasting and a more general method where data in all routes were used in one framework. Fourteen independent variables were defined, including place of departure/arrival, return load, distance, vehicle and freight type, tonnage, fuel price, way stop and location difficulty level, to name just a few. They concluded that the route-specific approach achieved higher forecasting accuracy compared to the aggregated approach. For route-specific models, the ANN model performed better than the quantile regression model.

As one mode of transportation, ocean freight forecast has been widely researched in the literature. TL market and tramp shipping markets share similar characteristics, for example, they both have steady contract markets and volatile spot markets. In a shipping market, a ship can be hired for a certain period of time, known as time-charter, or it can be found in the spot market. In shipping literature, to facilitate more accurate forecasting, more research has shifted from building large-scale econometric or simulation models to more direct specifications of freight rate process itself or reduced forms (Glen, 2006). By doing so, the statistical properties could be obtained. Appropriate forecasting techniques could enable players in the shipping business to make better informed decisions. By studying the methods used in ocean transportation rate forecasting, we can draw potential references (such as different methodologies) to TL rates.

Recent years have seen the increasing popularity of machine learning techniques, such as artificial neural networks (ANNs) for financial time series analysis. ANNs have shifted from simple pattern recognition to varied application fields. They have been adopted in the ocean transportation rate forecasting (Fan, Ji, Gordon, & Rickard, 2013; Li & Parsons, 1997; Lyridis, Zacharioudakis, Mitrou, & Mylonas, 2004). Such methods have good fit for complex nonlinear function (Han, Yan, Ning, & Yu, 2014). Li & Parsons (1997) were the first to use a neural network model to forecast freight rates of crude oil tankers in the Mediterranean using data from 1980 to 1995. Three variables, including spot freights, the Drewry's tanker demand index and the total capacity of active tankers were considered by the authors. They developed two ANN models for freight rate forecasting, the first one utilizing only information from the self-correlation of freight rates, while the second one made use of all information from the three variables. To compare the results, they also established two parallel auto-regressive moving average (ARMA) models, and the results show that ANN performs better than the ARMA models in all cases. Lyridis, et al. (2004) implemented ANN in the VLCC market to forecast Ras Tanura-Rotterdam spot freight. Eleven variables are identified as input, including demand for oil transportation, active fleet, newbuilding and secondhand prices, etc. Fan, et al. (2013) adopted Wavelet Neural Networks to forecast Baltic Dirty Tanker Index. The main characteristic of this method is that it chains the properties of time frequency localization and neural networks' adaptive learning nature.

However, artificial intelligence is often regarded as a black-box method as the causal relations cannot be detected and no formal testing can be done.

2.3 Literature gaps

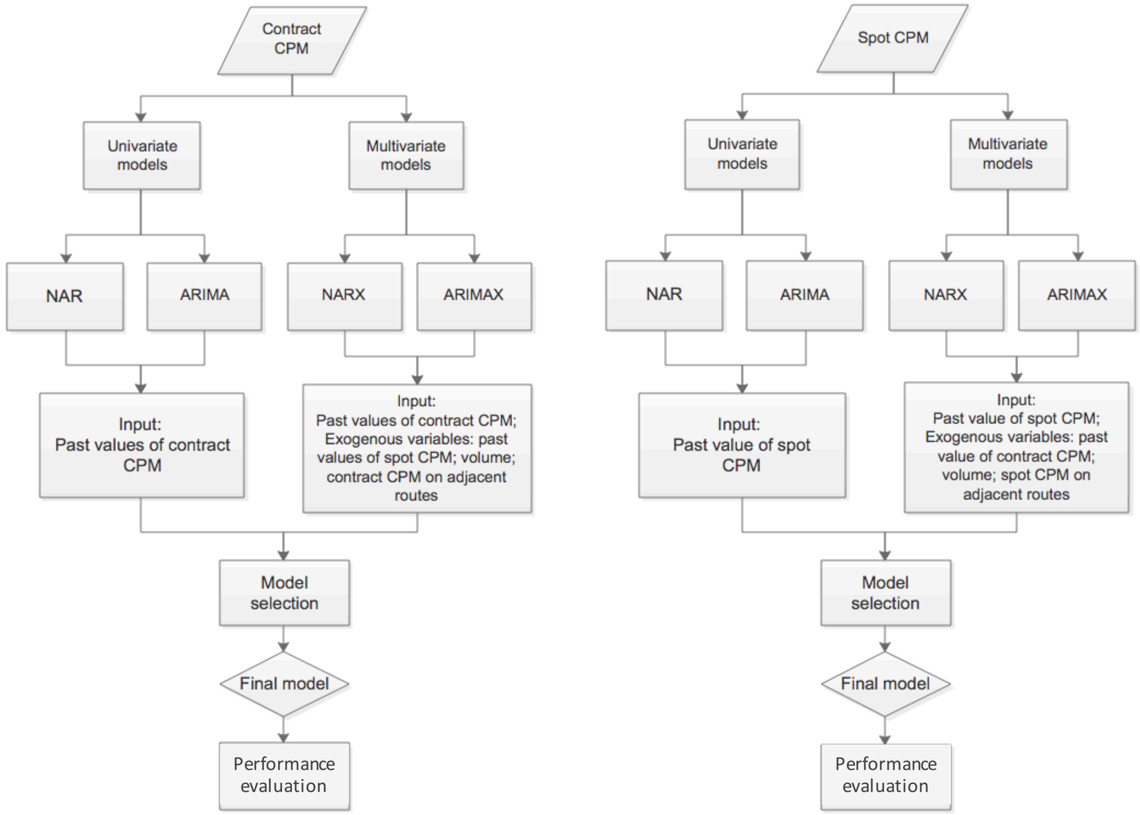
So far, most studies on TL rate forecasting have relied on linear regression methods for truck rate forecasting. Moreover, most literature focuses on contract rate forecasting. Spot market rates have been less studied in the literature, partially due to the complex and volatile nature of the spot market, which makes it hard to achieve a good forecasting accuracy. Furthermore, a uniform rate is often derived based on distance and weight. Such a method would be helpful for an initial understanding of the rate structure. However, it is not accurate enough since it disregards market information, such as time of year, volume, and the relationship between contract and spot rates. No study has been found that models and forecasts contract and spot rates for TL shipment on individual lanes, considering interactions between spot and contract rates, between rates for a particular lane and its adjacent routes, or between rates and volumes. This thesis aims to fill this gap by using neural network models to provide an estimate for the TL contract and spot rates.

3 Methodology and Data

3.1 Methodology

In this study, different models are proposed to forecast both contract and spot rates, specifically neural networks and traditional time series models. Initially, univariate Autoregressive integrated moving average (ARIMA) and Nonlinear autoregressive (NAR) neural network models are used to forecast both TL contract and spot market rates. Then, new variables are added (including volume and rates on adjacent trading routes), and multivariate ARIMA (ARIMAX) and Nonlinear autoregressive with exogenous inputs (NARX) neural network models are used. The forecasting performance of each model will be compared and the best predictive model will be selected. Last but not least, the performance over time will be evaluated for the selected model. The methodology process flow chart is shown in Figure 3.1.

Figure 3.1 Methodology flowchart



Source: Author.

3.1.1 Artificial Neural Network (ANN)

Artificial Neural Networks (ANN) are powerful non-parametric tools used in many applications, including pattern recognition, interpolation and time series forecasting. One of the main advantage of ANN over other econometric models is its ability to find complex and nonlinear associations between the parameters of the model without a priori assumption of the nature of the relationship (Zou, Xia, Yang, & Wang, 2007). Specifically, it is not necessary to assume a functional relationship between the variables. However, appropriate input variables need to be selected to make a sound estimate. Traditional time series models, such as ARIMA, are based on linear assumptions, which may not be suitable to model complex nonlinear relationships. Furthermore, strict hypotheses need to be made regarding the error terms for statistical models, however, ANN parameters are extremely adaptable. Almost no specification is needed for error distributions even for complex ANN models. Moreover, the problem of multicollinearity (the high correlation between independent variables) often found in statistical models is less a concern for ANN models. ANN models do not need the assumption of no correlation between independent variables (Karlaftis & Vlahogianni, 2011). Another advantage of an ANN model is that it has higher tolerance for errors in the data compared to other models. ANN model is quite robust to noise in the training data.

Each neural network consists of several layers. The first layer is the input layer, which is a group of input variables, $\{x_i\}, i = 1, \ldots, k$. The last layer is the output layer, with a group of output variables, $\{y_i\}, i = 1, \ldots, k$. The layers in between are called hidden layers. The number of hidden layers in a network can be zero, one, or more. In a network, neurons are connected between the layers, where the connection is activated when reaching a threshold. Such threshold is decided by the transfer function based on input parameters. Weights and bias are assigned for all the connections. Each neuron has three components: inputs, an activation function and outputs. In a neuron, the following calculation takes place:

$$y = f(b_0 + \sum_i \omega_i x_i)$$

where y is the output, x is the input, ω is the weight vector, b_0 is the bias and $f(b_0, \omega, x)$ is the activation function, which performs a transformation on the results calculated.

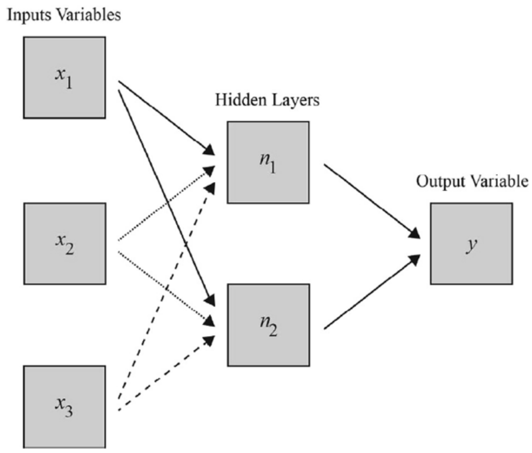
Each layer can have different numbers of neurons. The number of neurons for the input and output layers equal to the number of independent and dependent variables respectively. The input and output variables can be continuous, discrete, or a combination of the two (Kristjanpoller & Minutolo, 2015). A typical network diagram is shown in Figure 3.2.

A neural network modelling consists of two steps. The first step is the training of the network. The second step involves testing the forecasting performance of the network by minimizing the sum of squared difference between the output result and the dependent variable, also called the target. There are several learning algorithms available for testing the network. Among them, back-propagation has been the most popular and most widely implemented algorithm. This is an algorithm for supervised learning and utilizes gradient descent to minimize the quadratic error. Specifically, the algorithm begins with the output layer and propagates backwards using gradient rules to update the deviations and weights through multiple iterations, see Rumelhart, Hinton, & Williams (1986) for more details.

One important task for the neural network is to define the input variables, choose an appropriate number of hidden layers and number of neurons in each layer. Increasing the number of layers strengthen the model's ability to remember, but increases the computational time and the danger of overfitting, which may result in poor out-of-sample forecasting performance. On the other hand, reducing the number of layers decreases the time to learn, but may weaken the network's ability to generalize (Zou et al., 2007). In practice, one or two hidden layers are widely used and have proven to perform well. With small dataset in this thesis, the number of hidden layers is set to one.

There are no consistent formulas to select the optimal number of hidden neurons. Often, the number of hidden neurons are selected based on experiments. However, several rules of thumb have been developed to aid the selection process and they vary across different researchers. Bailey & Thompson (1990) suggests that the number of neurons in the hidden layer for a three-layer neural network should be around 70% of the number of input neurons. Ersoy & Hong (1990) recommended to double the number of hidden neurons until the network's performance deteriorates for the testing data. In this study, with same dataset, the number of hidden nodes is selected based on experiments.

Figure 3.2 An ANN network diagram



Source: Author.

3.1.1.1 Nonlinear autoregressive (NAR) and Nonlinear autoregressive with exogenous inputs (NARX) models

In this thesis, we adopt the Nonlinear autoregressive (NAR) and nonlinear autoregressive with exogenous inputs (NARX) models as our neural network designs. NAR and NARX models are capable of modelling complex, dynamic and nonlinear real-word time series data, thus providing a powerful tool for time series analysis and predictions (Gao & Er, 2005).

In the neural network setup, NAR and NARX models have recurrent neural architectures (Chen, Billings, & Grant, 1990). However, unlike other recurrent neural models, feedback structures for NARX models are limited only from the output neuron, instead of from hidden states. The NAR model is constructed to predict the value of an observation y_i based on the past observations of y, the function can be expressed as:

$$y(t+1) = f\left[y(t), \dots y\left(t - d_y + 1\right)\right]$$

where $y(t)$ represents output of network at discrete time t. $d_y \geq 1$ is the memory delay. d_y defines how far back past values will be included in the model.

The NARX model builds upon NAR model and further incorporates past values of exogenous variable u. According to Lin et al. (1996), the function is defined as

$$y(t+1) = f\left[y(t), \dots y\left(t - d_y + 1\right); u(t), u(t-1), \dots u(t - d_u + 1)\right],$$

where $u(t)$ and $y(t)$ represent input and output of network at discrete time t. $d_u \geq 1$, $d_y \geq 1$ and $d_u \leq d_y$ are memory delays.

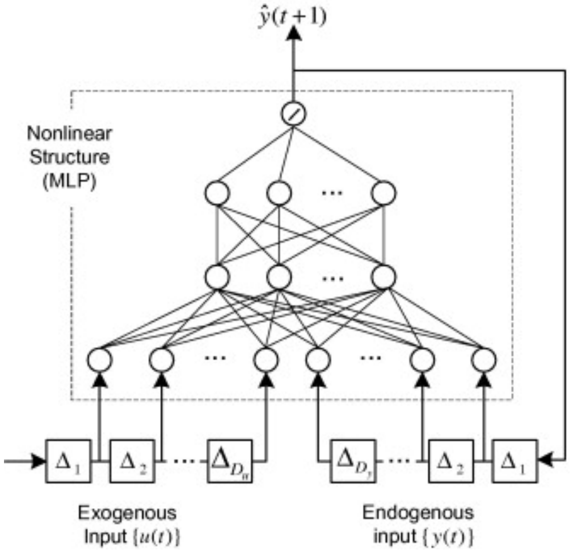
The formula can be further written in vector form as

$$y(t+1) = f[\boldsymbol{y}(t); \boldsymbol{u}(t)]$$

where $\boldsymbol{y}(t)$ and $\boldsymbol{u}(t)$ are vectors representing the output and input regressors. The nonlinear mapping $f(\cdot)$ can be approximated by a standard multilayer perceptron (MLP) network with back-propagation algorithm. A multilayer perceptron (MLP) is a class of artificial neural network described in Section 3.1.1. Then, the resulting connected architecture is called NARX network. It is a powerful class of dynamic models. The model is computationally as strong as fully connected recurrent networks, and also Turing machines (Siegelmann, Horne, & Giles, 1997). A typical structure of a two-hidden-layer NARX network is shown in Figure 3.3. The weights and bias values in the training network are updated following Levenberg-Marquardt optimization. During training, the real values of $y(t)$ are used as the input for the feed forward network, instead of the estimated ones.

In this study, we specify the number of hidden layers to be one. It is noted one hidden layer is usually sufficient for small sample sizes. The number of neurons N_h in the hidden layer are chosen based on the lowest mean squared error (MSE) for the validation set.

Figure 3.3 NARX network with D_u delayed inputs and D_y delayed outputs

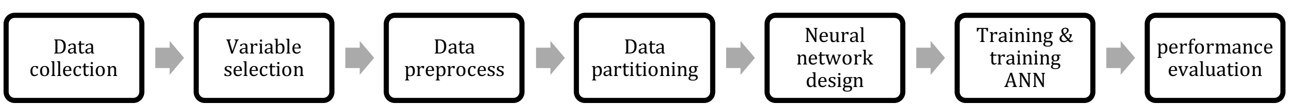


Source: Andalib & Atry (2009).

3.1.1.2 Process to build an ANN

A proper process has to be followed to build an ANN model. This study follows the procedures as listed out in Figure 3.4, which include data collection, variable selection, data pre-processing, data partitioning, neural network design, training and testing ANN, and performance evaluation. The detailed descriptions on each step are provided below.

Figure 3.4 ANN model process flow



Source: Author.

1) Data collection

Our data consists of the daily contract and spot cost per lane (CPL) and volume for each individual lane from origin city to destination city across the US. Since there are enormous transportation lanes based on city pairs, we introduce the regional corridor concept. The regional corridor is defined by the two-digit zip code, plus the cardinal direction (East, West, South, North and Central). The entire US geography is divided into 92 regions. The forecast is thus performed on a region to region basis. The original CPL data is then transformed to cost per mile (CPM) by using CPL divided by distance from the origin city to the destination city. The average daily CPM from one origin region to destination region is calculated by taking the average of all CPM values from cities in the origin region to cities in the destination region. Therefore, the final dataset used for forecasting models are daily contract and spot CPM, as well as total volumes for each individual lane from the origin region to the destination region from 1 Apr 2016 to 31 Mar 2017 (a total of 365 spot and contract observations each).

The detailed description on the dataset is provided in section 3.2.

2) Variable selection

The primary rule for variable selection is that input variables shall be as predictive as possible. To select the appropriate variables and the number of delays (lags), the autocorrelation coefficient of output variables and the cross-correlation coefficient of input and output variables are calculated.

Suppose Y is a stochastic process and Y_t is the value produced by a given run of the process at time t. The process has mean u_t and variance ${\sigma_t}^2$ at time t, for each t. The definition of autocorrelation between times s and t is defined as:

$$\rho(s,t) = \frac{E[(Y_t - u_t)(Y_s - u_s)]}{\sigma_t \sigma_s}$$

where $E[]$ indicates the expected value.

The term cross-correlation is used to define the correlation between two random variables X and Y at different time periods. Let X and Y be two wide-sense stationary process, the cross-correlation is given by:

$$\rho_{XY}(\tau) = \frac{E[(X_t - u_X)(Y_{t+\tau} - u_Y)]}{\sigma_X \sigma_Y}$$

where u_X and u_Y are the means for X and Y, while σ_X and σ_Y are the standard deviations. τ is the time difference.

In this study, the number of delays for the neural network and time series models are selected based on autocorrelation and cross-correlation values.

3) Data pre-processing

Analysis has shown that pre-processing data can influence the performance of prediction models (Zhang, 2003). The input and target values are normalized to keep them within the interval [-1,1]. This help to simplify the problem of potential outliers for the network. The following equation is used to normalize the data:

$$x_{i_{-1\ to\ 1}} = \frac{x_i - (x_{max} + x_{min})/2}{(x_{max} - x_{min})/2}$$

4) Data Partitioning

The data is split into training, validation and testing data. The training set is used to train the model and the validation set is used to select the optimal number of hidden neurons and epoch times to avoid over-fitting. The test set is used to evaluate the forecasting performance on unseen data. The split ratio between training, validation and testing data is 70:15:15 based on time sequence.

5) Neural Network design

The neural network uses NAR and NARX architecture as described in section 3.1.1.1.

6) Training and testing ANN

In the training stage, we use the series-parallel (SP) configuration (Menezes & Barreto, 2008), where the output regressor is formed by the actual values of the network's output.

$$y(t+1) = f\left[\boldsymbol{y_{sp}}(t); \boldsymbol{u}(t)\right],$$

$$= f\left[y(t), \ldots y\left(t - d_y + 1\right); u(t), u(t-1), \ldots u(t - d_u + 1)\right],$$

In this model, no estimated values are fed back to the input layer. The forecast $(y(t+1))$ is made using past values of actual outputs $(y(t), \ldots y\left(t - d_y + 1\right))$ and exogenous variables $(u(t), u(t-1), \ldots u(t - d_u + 1))$. d_y and d_u are feedback and input delays, which specify how many periods back of observations shall be included in the model. There are some advantages to use true outputs instead of feeding back the estimated ones. Firstly, it provides more accurate training models. Secondly, it will result in a purely feedforward network structure. This leads to less computational effort, as static backpropagation can be then adopted for training.

In the testing stage, as the series-parallel network can only predict one-step-ahead, the system is changed to parallel (P) mode to allow for multi-step ahead forecasts. The predicted outputs in this network are fed back to the input of the feedforward neural network:

$$y(t+1) = f\left[\boldsymbol{y_p}(t); \boldsymbol{u}(t)\right],$$

$$= f\left[\hat{y}(t), \ldots \hat{y}\left(t - d_y + 1\right); u(t), u(t-1), \ldots u(t - d_u + 1)\right]$$

Where $\hat{y}()$ is the estimated values. The forecast $(y(t+1))$ is calculated by using past values of predicted outputs ($\hat{y}(t), \ldots \hat{y}\left(t - d_y + 1\right)$) and exogenous variables $(u(t), u(t-1), \ldots u(t - d_u + 1))$. The network could continue to predict using internal feedbacks, even when external feedbacks are missing. It can make as many predictions as the time steps of the input series.

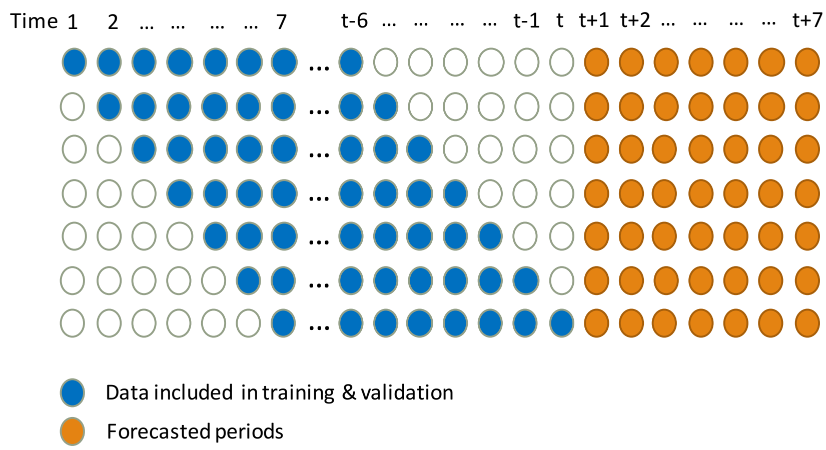
With different initial weights and bias, the model could yield different results even with same inputs and neural network structures. This is because gradient decent always finds local optimal, thus, with different initial starting points, the local optimal point will be different. To stabilize the results, each model is run 10 times and the average of the estimated outputs are taken.

7) Performance evaluation

The model is evaluated in multi-step-ahead prediction tasks. The metric used to evaluate performance is the mean squared error (MSE). After the best model is selected with lowest MSE, rolling-window forecasts are performed to test the network performance over time. Specifically, we perform daily and weekly updates of the training set when new information comes in and re-evaluate the model performance on the testing set. The hypothesis is that with new information, the forecasting accuracy will increase. In other words, closer forecasts should perform better than distant forecasts.

Daily update: 338-344 observations (7 days) are held out for testing. Models are trained with different training and validation sets in each round, namely: 1 to 331 observations for the first round, 2 to 332 observations for the second round, until 7:337 observations for the seventh round. The spilt ratio between training and validation sets is 70:15. The process is shown in Figure 3.5.

Figure 3.5 Daily rolling window model updates



Weekly update: Instead of daily rolling window update, the model is updated every week with weekly new information coming in. Specifically, 338-365 observations (4 weeks) are held out for testing. The training sets included in each round are 1:295 observations for the first round, 8:302 observations for the second round, until 42:337 observations for the seventh round. It is noted that the number of observations used for each training model is the same.

3.1.2 ARIMA and ARIMAX models

Autoregressive integrated moving average (ARIMA) model, proposed by Box & Jenkins (1970), is a common and popular tool to model time series. The model has the capability to capture the series dependence, and is relatively simple in applications. The model uses past values of a univariate time series to analyse the trend and forecast future cycles. An ARIMA model in most cases can be described by an Autoregressive Moving Average (ARMA) model with the integrated term set to zero. An ARMA(p, q) process for a stationary series Y_t can be defined as:

$$Y_t = \phi_0 + \phi_1 Y_{t-1} + \cdots + \phi_p Y_{t-p} + a_t + \theta_1 a_{t-1} + \cdots + \theta_q a_{t-q}$$

where a_t is a white noise. It can also be written as

$$\phi(B) \cdot Y_t = \theta(B) \cdot a_t$$

where B is a lag operator and $Y_{t-1} = BY_t$. The lag operator performs on one point of a time series to produce the previous point. The lag operator can be raised to powers. The polynomials can also be formed. The autoregressive polynomial and moving average polynomial can be represented by $\phi(B)$ and $\theta(B)$ respectively.

$$\phi(B) = \phi_0 + \phi_1 B + \phi_2 B^2 + \cdots + \phi_p B^p$$

$$\theta(B) = \theta_1 B + \theta_2 B^2 + \cdots + \theta_q B^q$$

The above equations are only valid under the condition of a stationary process. A stationary process is a stochastic process whose unconditional joint probability distribution remains the same in different times. If a series is not stationary, operator $(1 - B)$ can be applied d times to make the series stationary. An general ARIMA (p, d, q) model is derived as follows:

$$(1 - B)^d \phi(B) \cdot Y_t = \theta(B) \cdot a_t$$

where d is the difference times.

In many cases, including leading indicators could improve the model performance. This leads to the introduction of ARIMAX model. The ARIMAX model is an extension of ARIMA model, and is often known as a dynamic regression model. The explanatory variables can be inserted into the univariate model to derive the multivariate ARIMAX model.

In the general case of more than one independent variable, an ARIMAX model can be written as:

$$(1 - B)^d \phi(B) \cdot Y_t = \theta(B) \cdot a_t + \sum_{i=1}^{K_1} \beta_{t-i}^{(1)} X_{t-i}^{(1)} + \sum_{i=1}^{K_2} \beta_{t-i}^{(2)} X_{t-i}^{(2)} + \cdots$$

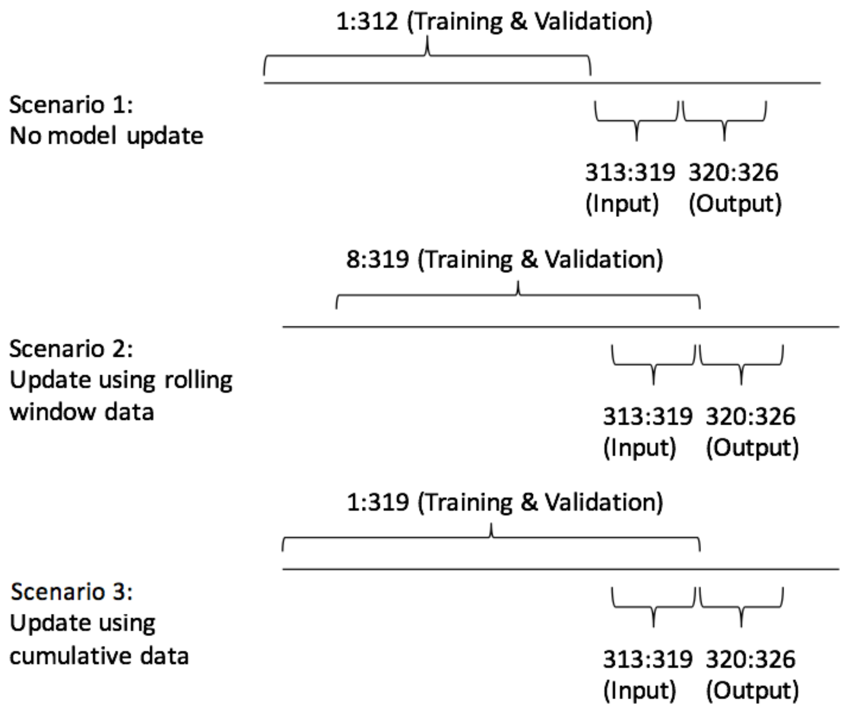
where $(1-B)^d \phi(B) \cdot Y_t$ and $\theta(B) \cdot a_t$ terms are the same as the original ARIMA model. The rest terms are explanatory variables. $X_{t-i}^{(j)}$ is the j th independent variable at time $t-i$ and $\beta_{t-i}^{(j)}$ is the corresponding parameter. K_j is the order for $X^{(j)}$.

3.1.3 Model updates with new information

Once the model has been selected and trained, one of the important decisions to make is when to modify the model when new information comes. In practice, too frequent updates may not be necessary when the original model still has sound forecasting performance. However, too infrequent updates mean that the model is not learning from the new data, and the new data does not make contributions to future forecasting, leading to a waste of information. Therefore, it is critical to decide the best time to update the original model. Furthermore, when updating the model, it is also important to decide how much information is needed to train the new model. In this thesis, different scenarios are tested, which include different update times and different information sizes for model updating. All tests are performed on the test set, which includes unseen information. Figure 3.6 indicates different scenarios for model updating.

In the first scenario, no model update is performed. When the new information comes, it is fed directly into the model to provide next 7 days' forecast. For example, data for days 313:319 is fed into the model as input, and forecast is given for days 320 to 326. In the next period (one week), again data for days 320 to 326 is fed into the model as input, and forecasts for days 327 to 333 are provided. In the second scenario, model is updated every week using a rolling window for data. For example, with new information of days 313 to 319, the model is trained using periods 8 to 319 as training and validation sets (ratio: 70:15). Periods 1 to 7 are considered as oldest information and can be discarded when updating a new model. In the last scenario, unlike scenario 2, all the previous information and new information (periods 1 to 319) are used to train the new model. Different models are tested for the next 7 periods (one week considered as one period) and MSEs for different models at each period are compared. The models for each period ahead are run 10 times to stabilize the results.

Figure 3.6 Scenarios for model updating



Source: Author.

To measure the performance of models with updating and without updating, the MSE reduction rate is used. The MSE reduction rate is defined as (MSE original model- MSE updated model)/MSE original model*100%. The error for the updated model is calculated by taking the difference between the averaged outputs from 10 runs and the targets.

3.2 Data

The initial dataset consists of the US nationwide weekly long-haul Truckload data, including origin region, destination region, year, week, model type (LHDV for long-haul dry van and LHTC for long-haul temperature controlled), rate type (contract or spot), volume, average cost per load (CPL) and average distance during the period of week 40, 2012 to week 9, 2017. The volume is the sum of volume hauled over the week, and average CPL is the averaged CPL for all shipments on a specific lane over a week. This dataset is used to analyse trends in volumes and rates over the years, as well as geographical allocations of volumes over the States.

After performing aggregate data analysis on the initial dataset, we further extract the daily contract and spot cost per mile (CPM), and volume information for each origin region to destination region over a one-year period (1 Apr 2016 to 31 Mar 2017) as stated in Section 3.1.1.2. Such dataset is used to build the forecasting model. Specifically, contract and spot dry van rates on one region to region corridor (Georgia Central (GA_C) to Florida Central (FL_C)) are used as an empirical example for forecasting.

3.2.1 LHDV and LHTC aggregate weekly data analysis

In this section, volume and rate developments for LHDV and LHTC are analysed on a national basis. This helps to provide an overview of the entire US trucking industry and its trends over the years.

3.2.1.1 LHDV and LHTC volume analysis

Dry vans on average haul more freight compared to temperature controlled trucks. Furthermore, more than 90% of the volume are hauled on contract basis for both dry van and reefer. Table 3.1 below summaries the contract volume and spot volume for LHDV and LHTC, as well as the year-on-year growth rate. As can be seen from the table, the average yearly volume for LHDV is around 4 times of LHTC. Within each sector, most volumes are hauled on contract basis, and only a small percentage is sourced on the spot market. The percentage of spot volume over total volume ranges from 2.7% to 5.83% for LHDV, and 1.24% to 4.26% for LHTC over

2013-2016. The surge in spot volume for both LHDV and LHTC in 2014 can be attributed to several factors, ranging from disruptive winter weather, to improved US economic conditions, coupled with seasonal freight influxes. Consequently, shippers tended to rely more on third party logistics providers (3PL) and brokers to meet their growing freight demand during that period.

Table 3.1 Contract and spot volume for LHDV/LHTC over 2013 to 2016

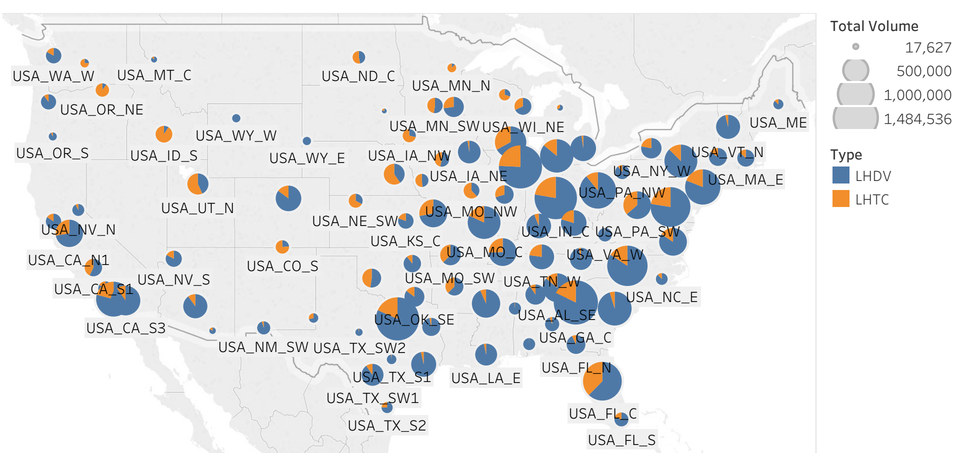
	LHDV					LHTC				
Year	**Contract Volume**	**yoy**	**Spot volume**	**yoy**	**% Spot**	**Contract Volume**	**yoy**	**Spot volume**	**yoy**	**% Spot**
2013	4,469,154		123,878		2.70%	1,418,138		17,849		1.24%
2014	5,091,637	*14%*	315,353	*155%*	5.83%	1,480,049	*4%*	65,873	*269%*	4.26%
2015	6,811,733	*34%*	333,287	*6%*	4.66%	1,588,537	*7%*	53,626	*-19%*	3.27%
2016	7,606,387	*12%*	321,866	*-3%*	4.06%	1,873,386	*18%*	67,293	*25%*	3.47%
Yearly Avg	5,994,728		273,596			1,590,028		51,160		

Note: spot data is self-reported by each company.

3.2.1.2 Volume analysis for origin and destination regions

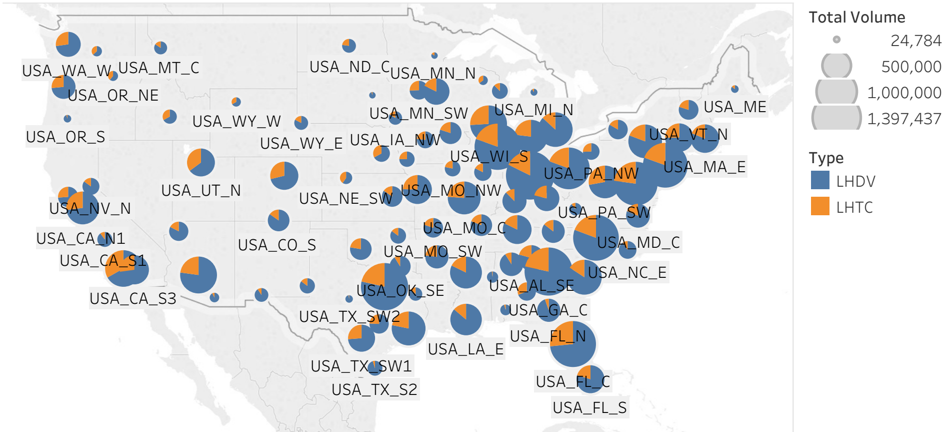
Geographically speaking, Northeast, Southeast and west coasts have high inflows and outflows, which is in line with state population distribution. Figure 3.7 and Figure 3.8 show the volume distribution by origin region and destination region. The size of the pie indicates the total volume in that region.

Figure 3.7 Total TL volume by origin region (week 40, 2012 to week 9, 2017)



Source: Author.

Figure 3.8 Total TL volume by destination region (week 40, 2012 to week 9, 2017)



Source: Author.

The origin regions and destination regions with top ten average weekly volume for LHDV and LHTC over the entire data period are reported in Table 3.2. The volumes in each region are not evenly distributed as shown in Table 3.3. For LHDV, the average weekly volume in or out from a region ranges from 30 to 5,187, with a mean of 1299, while for LHTC, the weekly volume spans from around zero to 1825, with mean being 345. For LHDV, Georgia (GA), Texas (TX), Illinois (IL), Indiana (IN) and North Carolina (NC) are high volume nodes. Florida (FL), Georgia (GA), Pennsylvania (PA), Indiana (IN) and Texas (TX) are nodes with high volumes for LHTC.

Table 3.2 Origin and destination with top ten average weekly volume for LHDV and LHTC

LHDV				LHTC			
Origin Region	Avg. Weekly Volume	Destination Region	Avg. Weekly Volume	Origin Region	Avg. Weekly Volume	Dest Region	Avg. Weekly Volume
USA_GA_C	5,187	USA_IN_C	4,893	USA_FL_C	1807	USA_FL_C	1416
USA_TX_C	4,920	USA_GA_C	4,606	USA_IL_CHI	1450	USA_GA_C	1257
USA_IL_CHI	4,560	USA_TX_C	4,451	USA_IN_C	1288	USA_TX_C	1207
USA_IN_C	4,521	USA_IL_CHI	4,370	USA_PA_S	1140	USA_CA_LA	1120
USA_NC_C	4,442	USA_NC_C	4,285	USA_GA_C	1123	USA_PA_S	1104
USA_PA_S	3,961	USA_NJ_C	4,103	USA_TX_C	1099	USA_IL_CHI	1051
USA_OH_C	3,818	USA_FL_C	3,946	USA_WI_S	1045	USA_IN_C	1046
USA_SC_C	3,535	USA_PA_S	3,754	USA_PA_SW	897	USA_NC_C	1018
USA_NJ_C	3,254	USA_OH_C	3,565	USA_NE_E	828	USA_NJ_C	987
USA_CA_LA	3,118	USA_AZ_C	2,661	USA_NC_C	818	USA_WI_S	890

Note: Regions are shown as US state abbreviation. Detailed geographical locations of regions are provided in Appendix A.

Table 3.3 Descriptive statistics for origin and destination weekly volume for LHDV and LHTC for different origin and destination regions

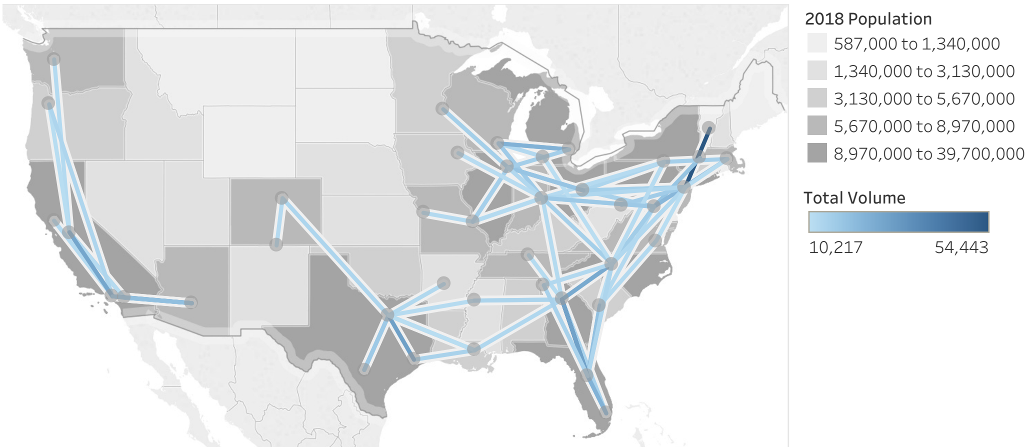
	LHDV		LHTC	
Descriptive statistics	Origin Volume	Destination Volume	Origin Volume	Destination Volume
Mean	1,299.22	1,299.22	345.52	345.52
Median	735.48	939.94	203.52	212.51
Standard Deviation	1,278.66	1,192.22	369.59	344.48
Kurtosis	1.06	1.43	2.73	0.92
Skewness	1.33	1.43	1.60	1.31
Range	5,156.86	4,805.17	1,825.85	1,423.15
Minimum	30.09	87.94	0.05	3.40
Maximum	5,186.94	4,893.10	1,825.89	1,426.55

3.2.1.3 Lane specific volume analysis

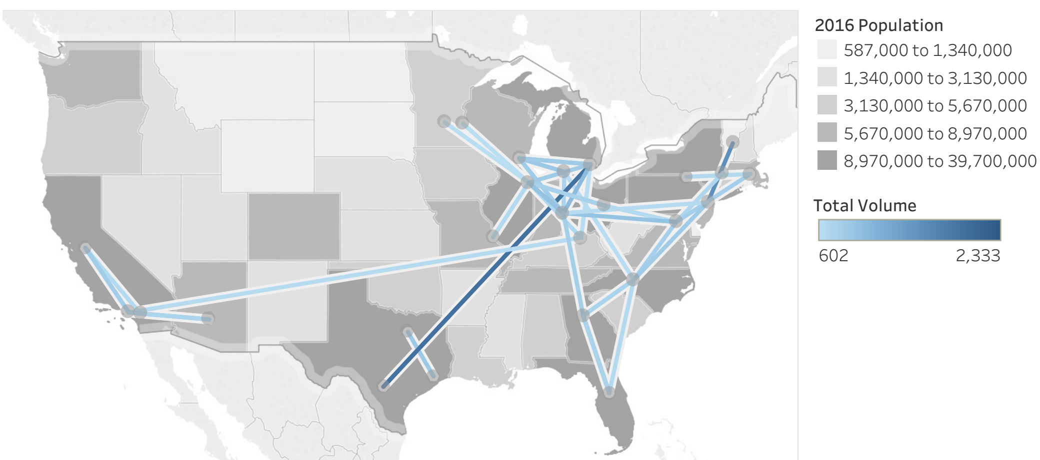
The US trucking system is a complex network, with freight flowing over thousands of lanes across the country. A total of 8023 and 6366 lanes are active during the data period for dry van and reefer respectively. The contract and spot high volume lanes for both LHDV and LHTC are plotted in the maps below (Figure 3.9). The color of the line between regions represents the total volume hauled on a specific lane over the entire data period. The darker the blue line, the higher the volume is. The map coloring indicates the state population. As can be seen from the graphs, volume flows follow state population levels. Higher volume flows between regions with larger populations. Another observation is that for high volume lanes, the average distance for LHDV is shorter than that for LHTC. This implies that dry van tends to move high volume freight between adjacent regions.

Figure 3.9 Maps for high volume lanes

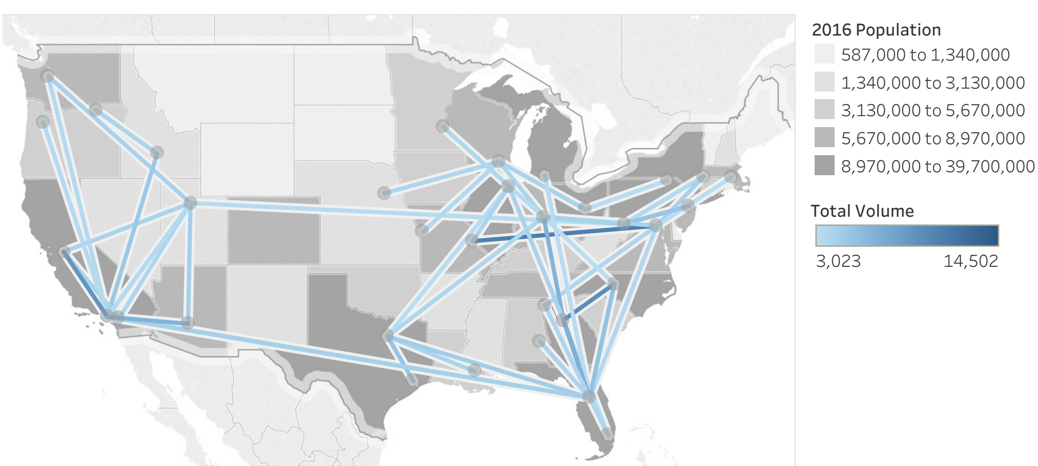
a) LHDV contract high volume links and 2016 US population (Total volume > 15,000)

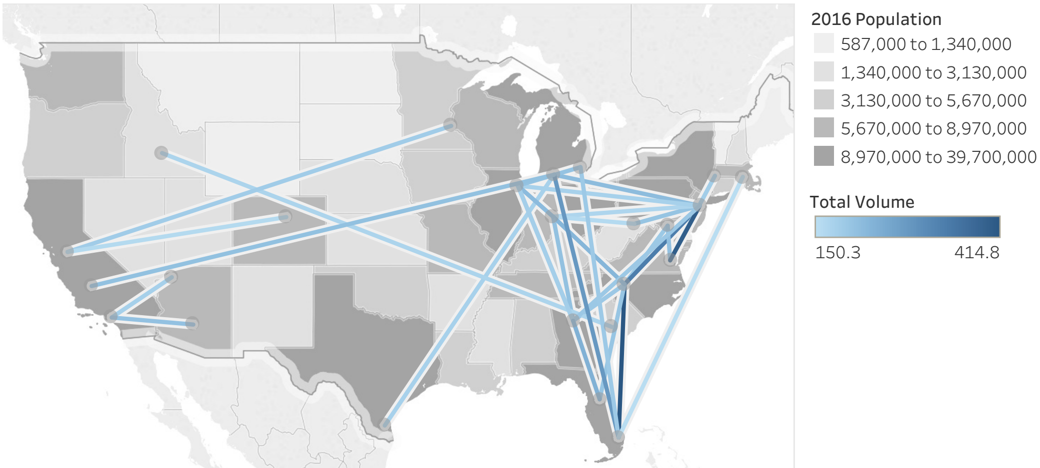


b) LHDV spot high volume links (volume > 600)



c) LHTC contract high volume links (volume > 4,000)





d) LHTC spot high volume links (volume > 150)

Note: map coloring shows 2016 population by state. The color of the line shows the volume on the link. The darker the blue line means more volume.

Source: Author.

Table 3.4 and Table 3.5 show top ten volume links for LHDV and LHDC respectively. The highest links for LHDV are from Vermont North (VT_N) to New Jersey Central (NJ_C), Georgia Central (GA_C) to Florida Central (FL_C) and California Los Angeles (CA_LA) to Arizona Central (AZ_C). For LHTC, the highest volume links are GA_C to FL_C, Pennsylvania South (PA_S) to Missouri Central (MO_C).

The regional corridor - GA_C to FL_C appears in the top 2 lists both for LHDV and LHTC, indicating the important role this link plays in the US trucking network. Thus, in the following sections, rates on this link are used as the forecasting data.

Table 3.4 Top average weekly volume links for LHDV

OD Pair	Average weekly volume	Average weekly contract volume	Average weekly spot volume
USA_VT_N|USA_NJ_C	1071	1043	28
USA_GA_C|USA_FL_C	1042	1012	30
USA_CA_LA|USA_AZ_C	734	709	25
USA_CA_S3|USA_AZ_C	629	620	9
USA_TX_C|USA_TX_SE	588	572	17
USA_FL_C|USA_GA_C	570	556	14
USA_CA_LA|USA_CA_N1	546	530	16
USA_GA_C|USA_NC_C	531	516	15
USA_FL_C|USA_FL_S	469	458	10
USA_OH_C|USA_IL_CHI	457	427	30

Table 3.5 Top average weekly volume links for LHTC

OD Pair	Average weekly volume	Average weekly contract volume	Average weekly spot volume
USA_GA_C|USA_FL_C	279	273	5
USA_PA_S|USA_MO_C	207	206	1
USA_GA_C|USA_NC_C	187	184	4
USA_CA_LA|USA_CA_N1	175	172	3
USA_CA_LA|USA_AZ_C	130	125	5
USA_IN_C|USA_GA_C	129	125	4
USA_FL_C|USA_GA_C	129	127	2
USA_PA_SW|USA_NJ_C	126	119	6
USA_UT_N|USA_CA_LA	111	109	2
USA_TX_C|USA_TX_SE	103	100	3

3.2.1.4 LHDV and LHTC rate analysis

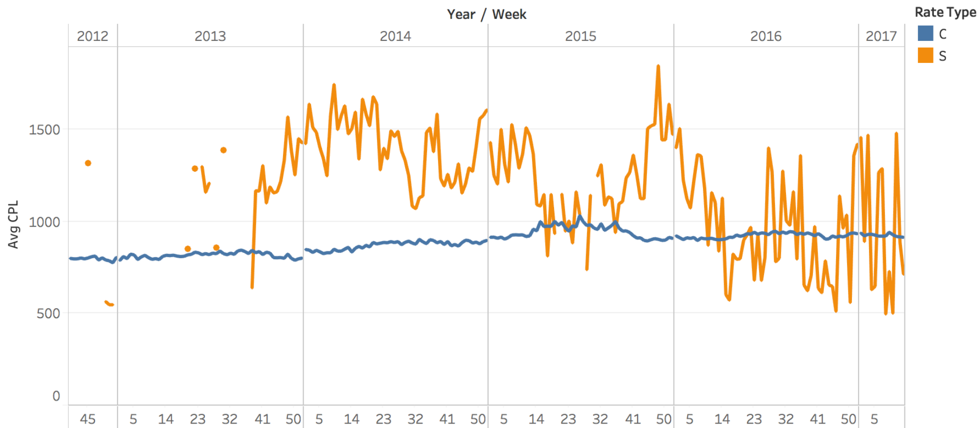
As mentioned in Section 2, a contract rate is a rate quote issued to a shipper that is supposed to hold static for a period of time, normally a year. In reality, contract rates could also be adjusted or renegotiated. However, contract rates are far less volatile compared to spot rates.

The TL spot market is an extremely dynamic place where rates respond quickly to both the demand for transportation services and the truckload capacity availability to move the freight. Spot market rates vary in different seasons, and carriers normally obtain high rates during periods of peak demand. Spot rates normally fluctuate around contract rates. A shipper can adopt a mix of contract or spot rates when designing their freight transportation, based on the market conditions. For example, if the economic condition deteriorates, and truck capacity seems to be abundant, rates are expected to be falling. In this case, shippers can make use of spot quotes to save transportation costs. On the other hand, if all factors indicate a rising rate, shippers would rather lock in contract rates as a hedge against rate increase. This will also help shippers capture the increasingly scarce capacity with raising rates. From the carriers' perspective, they can also take advantages of spot market rate movements by positioning trucks into regions where rates are expected to increase.

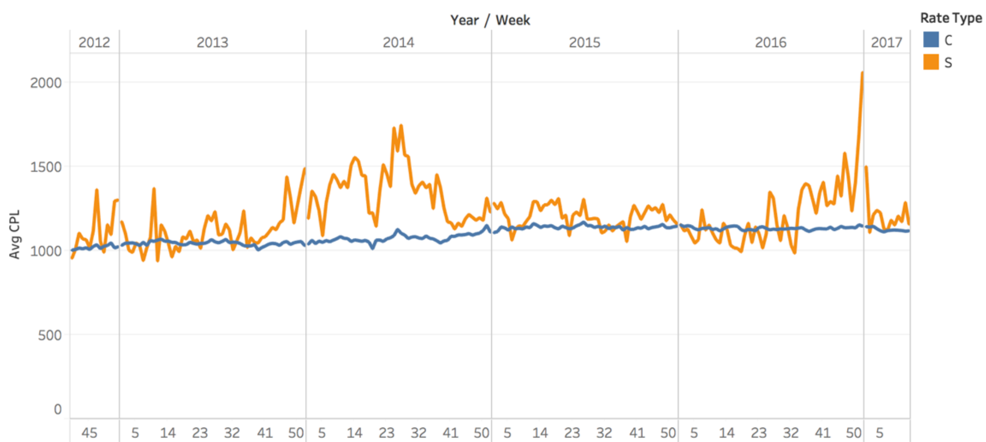
The weekly freight movement information is collected for high volume lanes and plotted in Figure 3.10. The high volatility in spot rates is further evidenced by the graphs. It is also noted that 2014 was a strong year for the spot market, where spot rates stay above contract rates for all lanes shown in the figure. For the rest times, spot rates fluctuate around contract rates. The descriptive statistics for CPL development for the selective lanes are shown in Table 3.6. As can be seen from the data, the standard deviation ratio for spot over contract rates ranges from 2.9x to 5.4x for dry van, and 5.5x to 7.0x for reefer, indicating much higher variance in the spot market. Another observation is that the mean value for spot rates are higher than that for contract rates for all three lanes for both dry van and reefer, indicating generally higher transportation costs in the spot market.

Figure 3.10 Weekly CPL development for selected dry van lanes

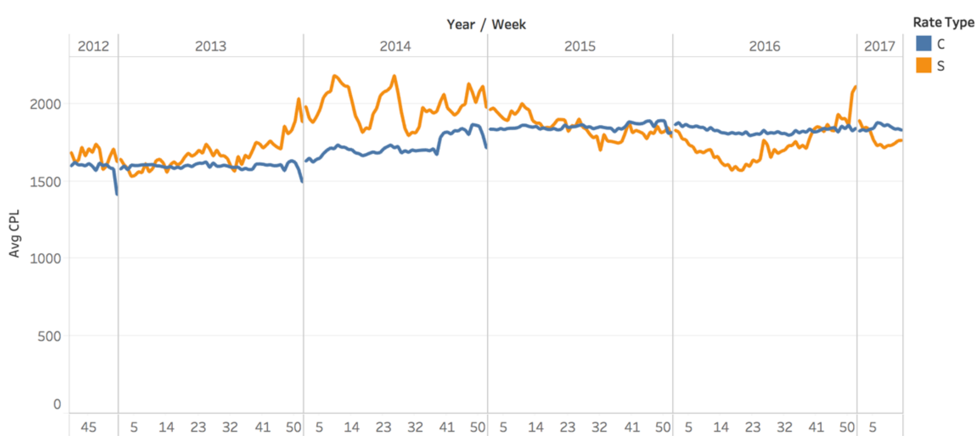
a) Dry van VT_N to NJ_C



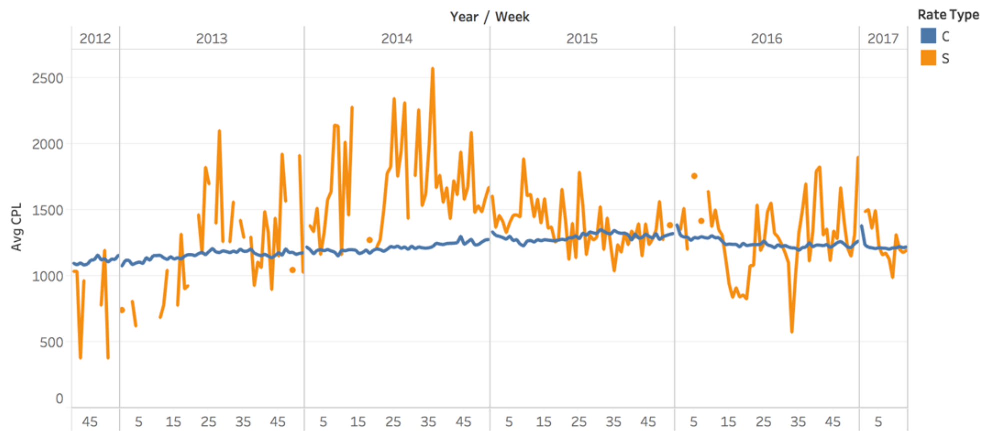
b) Dry van GA_C to FL_C



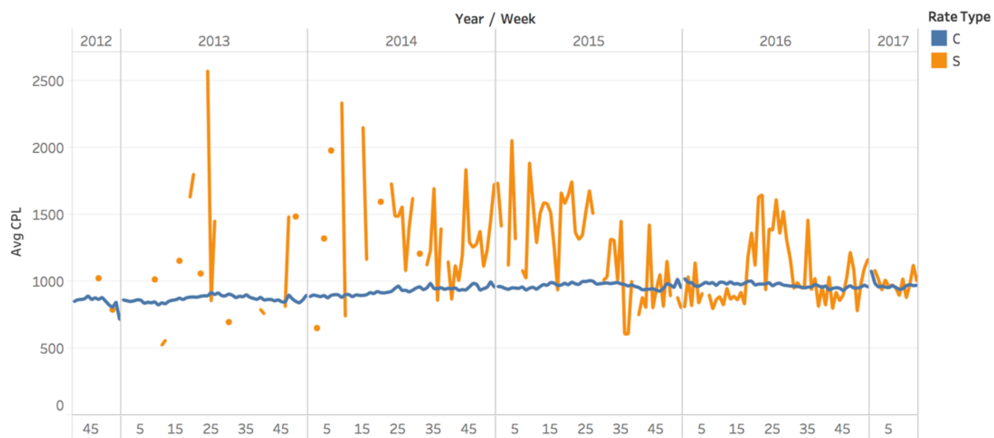
c) Dry van CA_LA to AZ_C



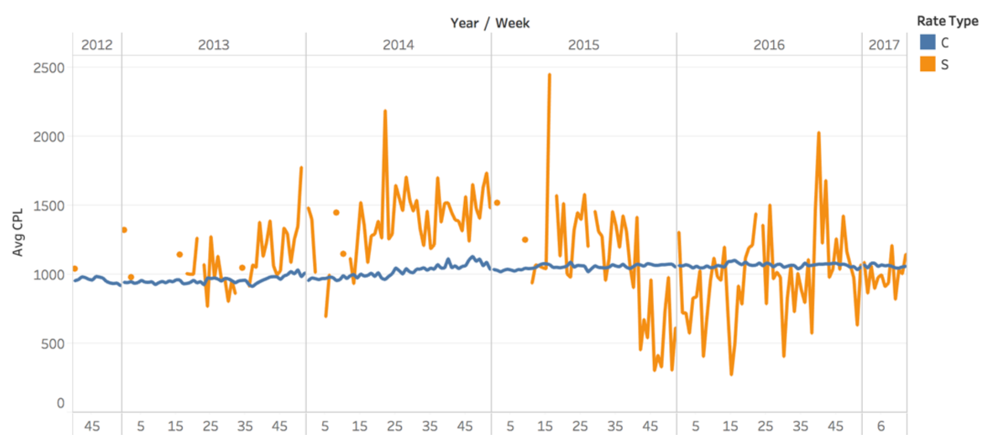
d) TC GA_C to FL_C



e) TC GA_C to NC_C



f) TC CA_LA to AZ_C



Note: C stands for contract rate and S stands for spot rates for the rate type label.

Source: Author.

Table 3.6 Descriptive statistics for CPL rates on selective lanes

			Mean	SD	Kurtosis	Skewness	Minimum	Maximum	Count
LHDV	VT_N|NJ_C	C	885.48	56.66	-0.91	-0.07	775.41	1031.47	234
		S	1178.60	304.13	-0.55	-0.47	496.98	1845.73	192
	GA_C|FL_C	C	1095.52	45.03	-1.39	-0.32	1003.98	1171.49	235
		S	1216.27	159.45	3.37	1.30	939.13	2059.91	234
	CA_LA|AZ_C	C	973.64	52.87	-1.42	-0.38	876.56	1075.04	235
		S	1056.12	152.14	0.73	0.65	751.58	1605.01	235
LHTC	GA_C|FL_C	C	1222.81	63.21	-0.40	-0.06	1074.93	1388.10	235
		S	1378.77	348.20	1.11	0.31	378.02	2572.90	212
	GA_C|NC_C	C	930.85	51.77	0.23	-0.57	715.96	1078.54	235
		S	1182.51	360.58	0.94	0.89	523.56	2571.84	160
	CA_LA|AZ_C	C	1023.44	51.71	-1.16	-0.40	913.40	1132.46	235
		S	1140.43	340.42	1.18	0.11	273.45	2450.25	186

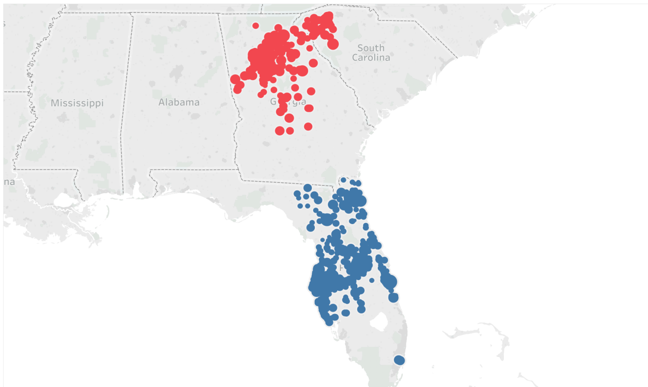
Note: C stands for contract and S stands for spot.

3.2.2 Daily disaggregated dry van rate analysis

In this study, forecasting is performed for the daily GA_C to FL_C dry van contract and spot rates. This section examines closely the daily rate development for GA_C to FL_C, as well as relevant variables that would affect daily rates, such as volumes and rates on adjacent routes.

The dataset contains daily contract and spot rates per route, rate type (contract/spot), the origin/destination cities in GA_C and origin/destination cities in FL_C during 1 Apr 2016 to 31 Mar 2017. In other words, the data includes every link in and out from GA_C and FL_C. Specifically for GA_C to FL_C, 717 links are active with origin city in the region GA_C and destination city in the region FL_C, where 85 cities act as origins and 132 cities act as destination. Figure 3.11 indicates the different origin cities in GA_C and destination cities in FL_C. The numerous number of links makes forecasting each city to city link almost impossible. Therefore, we convert the cost per lane (CPL) to cost per mile (CPM) for all the links to make the cost comparable, controlling the distance effect. Then, city to city daily rates are aggregated to region to region level. Forecasting is performed on regional corridor basis.

Figure 3.11 Origin cities in GA_C and destination cities in FL_C



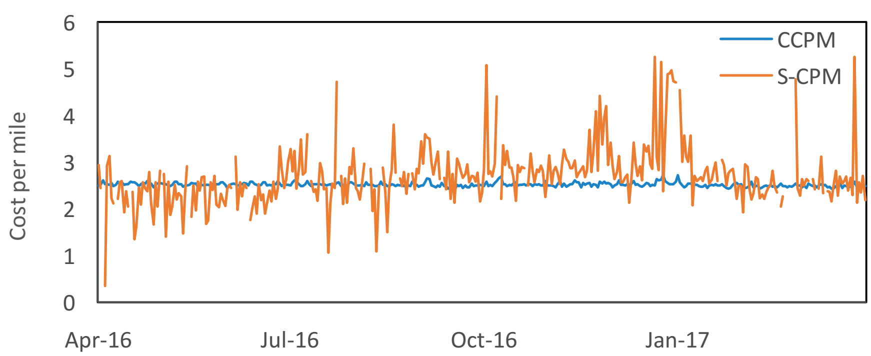
Source: Author.

3.2.2.1 Target and input variables for forecasting model

3.2.2.1.1 GA_C to FL_C daily spot and contract rates

As stated above, GA_C to FL_C daily spot and contract rates are the targets to forecast in this study. Figure 3.12 plots the spot and contract rates for this route. Contract rates are very stable, while spot rates are more volatile and fluctuates around contract rates.

Figure 3.12 GA_C|FL_C daily spot vs contract CPM



Source: Author.

3.2.2.1.2 Input decision variables

Using the past knowledge of the trucking industry, a few variables are identified as the potential candidates to be included in the forecasting model, which includes

1) The past values of contract/spot rates to account for autocorrelation of the time series;
2) The past values of spot rates for contract rates forecasting and vice versa, to account for interactions between spot and contract rates;
3) The past values of contract/spot rates on adjacent routes to account for the interaction of rates between adjacent routes, and regional supply/demand dynamics.

   Specifically, in this case, all routes with origin GA_C are selected and then the routes adjacent to GA_C to FL_C are chosen as candidates. Here, routes from GA_C to FL_N (Florida north), FL_S (Florida south) and SC_C (South Carolina south) are selected as potential candidates.

3.2.2.1.3 Descriptive data analysis

Before model building, the descriptive statistics analysis for the dataset has been performed. On many days, there is no spot volume moved on a lane, so there is missing data for spot rates. The number of missing data for the different routes are calculated. If the number of missing data is less than 10% of the entire series, the imputation method is used to impute the missing values. The imputation method is based on ordinary least square regression. In this case, spot rates are regressed on contract rates. The missing values in spot rates will be then replaced by predictive values obtained from the regression. If there is too much missing data, the data series is omitted from the input variables. The numbers of missing values for spot rates from GA_C to FL_C, FL_N, FL_S and SC_C are 26, 320, 127 and 126 respectively. Therefore, the missing values for GA_C|FL_C spot rates are imputed, while variables GA_C|FL_N-, GA_C|FL_S- and SC_C-spot rates are omitted due to higher number of missing values. Table 3.7 summarizes the descriptive statistics for the rest of the series. As can be seen, the range for spot rates on GA_C to FL_C are from 0.34 to 5.26, while that for contract rates are from 2.38 to 2.70. Furthermore, the coefficient of variation for spot rates is around 15 times of contract rates. This indicates higher variability in spot rates. The mean contract CPM for GA_C|FL_C, GA_C|FL_N, GA_C|FL_S are very similar, indicating that same contract rates are often set for these three lanes. On the other hand, mean contract CPM on GA_C|SC_C lane is significantly lower than the other lanes.

One of the prerequisite for ARIMA and ARIMA-X model is that the series need to be stationary. The Augmented Dickey-Fuller (ADF) test (with constant and no trend) is performed to test the stationarity of the time series. The null hypothesis is that the series has a unit root. Thus, a significant p value leads to the rejection of the null hypothesis. The ADF tests for all series are significant, indicating that all series are stationary. Thus, the series can be directly used for ARIMA and ARIMA-X model building.

Table 3.7 Descriptive statistics of the decision variables (CPM rates and volume)

	Contract GA_C|FL_C	Spot GA_C|FL_C	Contract GA_C|FL_N	Contract GA_C|FL_S	Contract GA_C|SC_C	Volume
Mean	2.50	2.70	2.48	2.48	2.05	193.50
SD	0.04	0.66	0.16	0.06	0.12	61.03
Coefficient of variation	0.017	0.241	0.066	0.025	0.058	0.315
Min.	2.38	0.34	1.70	2.32	1.56	15.00
Max.	2.70	5.26	3.05	2.70	2.66	297.00
Skewness	1.05	1.30	-0.06	0.45	1.64	-0.54
Kurtosis	2.97	3.94	2.59	0.37	6.71	-0.89
ADF test statistics	-5.22**	-4.19**	-7.01**	-4.68**	-6.59**	-10.3**

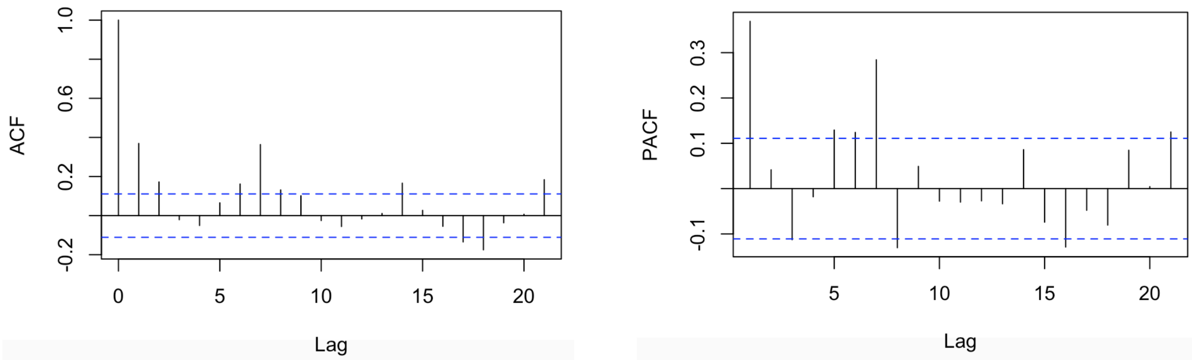
Note: ADF is the unit root test for stationarity. * and ** indicates significance at 5% and 1% levels respectively.

3.2.2.1.4 Variables' autocorrelation and cross-correlation analysis

As stated in section 3.1.1.2, one of the important procedures for building ANN is variable selection. In order for the future forecast to be robust and valid, it is important to cross-correlate the time series for forecasting with the historical independent input time series. Hence, the autocorrelation of contract and spot rates for GA_C to FL_C, and the cross-correlations with the other series, are calculated. Figure 3.13 and Figure 3.14 plot the autocorrelation (ACF) and partial autocorrelation (PACF) of a time series by lags for contract and spot rates. As can be seen from the figures, contract rates seem to have weekly seasonality patterns, where the correlation between current values and values of the seventh lag is significantly higher than the other lags. However, such effect is less observable for spot rates. Table 3.8 and Table 3.9 show the cross-correlations between GA_C to FL_C contract rates and other variables, and between GA_C to FL_C spot rates and other variables. As can be seen from the table, contract rates are moderately correlated with lagged values of volumes, while there seems to be no obvious correlation between spot rates and lagged values of volumes. It is also notable that the correlation between contract rates and past values of volume (when the lag equals to 6 or 7) is negative. This means that if the volume on a day in the previous week is high, the contract rate is likely to be lower on the same day this week. Furthermore, among all adjacent routes, GA_C to FL_C contract rates seem to be more correlated to the lagged values of contract rates for GA_C to FL_S, especially for the seventh lag (correlation = 0.40). The cross-correlation between spot and contract rates are modest at best, indicating that there is no obvious lead-lag relationship or information transmission between spot and contract rates in a short period (1-14 days).

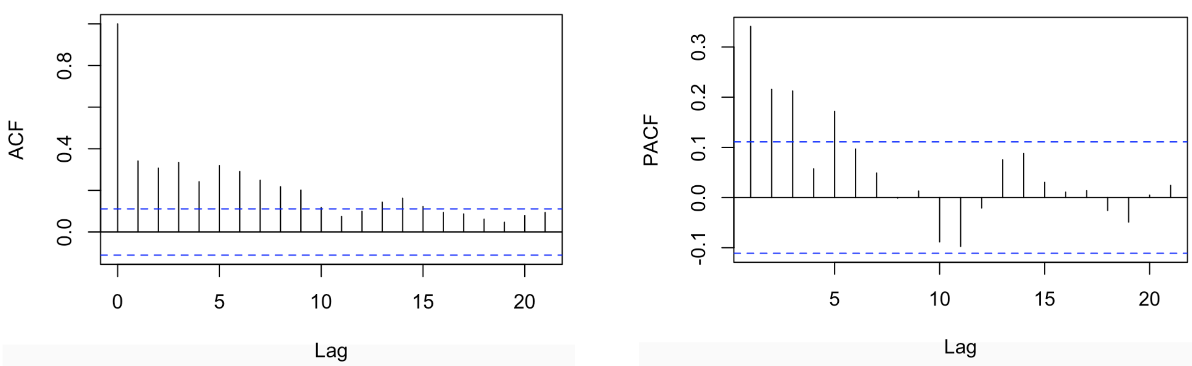
Therefore, based on the autocorrelation and cross-correlation analysis, the potential input variables are further refined. For contract rate forecasting, the potential variables include the lagged values of contract rates, spot rates, contract rates for GA_C to FL_S and volumes. On the other hand, the potential variables for spot rates include lagged values of spot rates and contract rates. The number of lags/delays included (d_y for feedback delays and d_u for input delays) are decided both based on the correlation analysis and forecasting performance.

Figure 3.13 ACF and PACF plots for GA_C to FL_C contract rates



Source: Author.

Figure 3.14 ACF and PACF plots for GA_C to FL_C spot rates



Source: Author.

Table 3.8 Cross-correlations between GA_C to FL_C contract rates (CCPM) and other variables

Number of lags	[CCPM, CCPM]	[CCPM, SCPM]	[CCPM, Contract GA_C|FL_N]	[CCPM, Contract GA_C|FL_S]	[CCPM, Contract GA_C|SC_C]	[CCPM, Volume]
0	1.00	0.25	0.05	0.50	0.08	-0.41
1	0.38	0.11	-0.11	0.23	0.12	-0.03
2	0.22	0.10	-0.05	0.16	0.13	0.23
3	0.04	0.13	-0.07	0.13	0.03	0.33
4	0.01	0.17	0.03	0.10	0.16	0.15
5	0.12	0.19	0.12	0.22	0.08	-0.05
6	0.21	0.16	0.02	0.26	0.07	-0.32
7	0.42	0.14	0.05	0.40	0.06	-0.34
8	0.19	0.10	-0.14	0.16	0.02	0.01
9	0.16	0.02	-0.04	0.05	0.03	0.20
10	0.02	0.06	-0.03	0.04	0.02	0.26
11	0.00	0.03	0.07	0.04	0.09	0.11
12	0.05	0.06	0.07	0.13	-0.02	-0.08
13	0.07	0.01	0.03	0.22	0.03	-0.33
14	0.24	0.02	0.01	0.30	-0.01	-0.32

Table 3.9 Cross-correlations between GA_C to FL_C spot rates (SCPM) and other variables

Number of lags	[SCPM, SCPM]	[SCPM,CCPM]	[SCPM, Volume]
0	1.00	0.25	-0.18
1	0.30	0.21	-0.09
2	0.28	0.23	-0.01
3	0.30	0.11	0.04
4	0.23	0.13	0.02
5	0.27	0.06	0.05
6	0.26	0.12	-0.08
7	0.23	0.14	-0.16
8	0.20	0.11	-0.10
9	0.19	0.08	-0.03
10	0.11	-0.01	0.04
11	0.08	-0.01	0.06
12	0.11	-0.07	0.04
13	0.13	0.00	-0.07
14	0.16	0.02	-0.18

3.3 Summary

This section describes the methodologies for TL rate forecasting and elaborates on the dataset used for forecasting. As a short summary, it is found that a majority of TL volumes are hauled on contract basis, with less than 10% in the spot market, both for dry van and reefer TLs. Furthermore, there are over 8000 active dry-van TL lanes across the states on a regional corridor to corridor basis, with volumes unevenly distributed. Rates on a high-volume lane (GA_C to FL_C) are used as our forecasting data in this study. Spot CPM rates on this route are much more volatile than contract CPM rates. Nevertheless, the autocorrelation of contract and spot rates for GA_C to FL_C, and the cross-correlation analysis with other series have been conducted. The potential input variables are further refined based on this analysis. In the next section, both NAR/NARX and ARIMA/ARIMAX models described above will be applied to GA_C|FL_C dataset and the forecasting performance will be evaluated and compared.

4 Results and discussions

In this section, the results of different neural networks and ARIMA models are presented and compared. Results show that the NAR model provides better short-term forecasting performance for spot rates than the ARIMA model, while the ARIMA model performs slightly better for contract rates. However, for a longer-term forecast, the NARX model provides better results for contract rates.

4.1 Artificial Neural Network (ANN) model results

The following sections present the results for ANN models. Specially, different NAR and NARX models are built for spot and contract rate forecasting. For spot rates, the NAR model that incorporates 7 feedback delays performs the best, while for contract rates, the NARX model that includes contract rates with feedback delay of 7 and spot rates with input delays of 7 is selected with the best forecasting performance. Then, the decision regarding when and how to update the model with new information has been discussed. Different scenarios including no update, rolling window update and cumulative update have been compared. The decision criterion is based on MSE reduction rate. For spot rates, training the model with new information does not necessarily improve the model's performance. For contract rates, updating the model every three to four weeks is recommended, as from that time, MSE reduces significantly compared to no model update scenario.

Last but not least, the model's performance over time has been evaluated. Same data set is used for forecasting, while different periods of data are used for training. The hypothesis is that if training data is closer to the forecasting data, the MSE should be smaller. For example, forecasts made 1 period ahead should be more accurate that forecasts made 7 periods ahead. Such hypothesis is sustained for contract rates, however, not that obvious for spot rates.

4.1.1 ANN model results for spot rates

As stated in Section 3.2.2.1.4, potential input variables for spot rate forecasting are historical values of spot rates and contract rates. Thus, ANN models are built that incorporate past values of spot rates (NAR model), as well as past values of spot and contract rates (NARX model). The results show that, with feedback delay equals to 7 and number of hidden nodes equals to 4, the model has the best predicting performance among all tested models.

4.1.1.1 NAR model for spot rates

Based on the autocorrelation analysis, the potential feedback delays (d_y) are set in the range of 7 to 14. The best model for each d_y is selected based on lowest MSE for the validation dataset. The best number of hidden nodes (N_h) is selected based on the same criteria. A loop is included in the model to test the performance of the model with N_h ranging from 1 to 20. Last but not least, d_y is chosen based on the lowest MSE for validation data, because the validation data assesses the model's performance on unseen data and avoids overfitting issues.

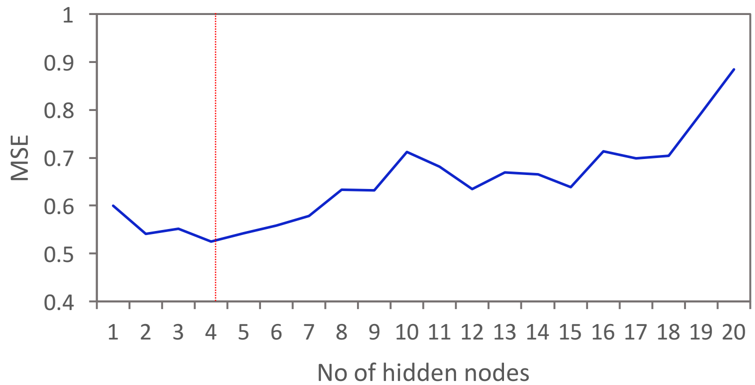
The results of the model with d_y equals to 7 are presented first. Table 4.1 shows MSEs for the validation set for different N_h, using NAR model with d_y(=7), conducted 10 times in order to get a stable result. As can be seen, the MSE is smallest when N_h equals to 4. Furthermore, the MSE decreases first when the number of nodes increases, but then increases with more nodes added that cause more complexities for the model, as shown in Figure 4.1. Therefore, for this specific model, N_h is set to 4. Then, the next important decision to make is when to stop iterations. Here, iterations stop when the error for the validation set stops decreasing. Such method can help prevent potential overfitting issues, because the training error will always decrease with more iterations.

Figure 4.2 illustrates the MSEs for training, validation and testing sets. The MSE for the validation set is lowest at epoch times = 4. This shall be the point where training stops. Figure 4.3 plots the one-day ahead forecasts for all data points.

Table 4.1 MSE for the validation set with different numbers of hidden nodes (N_h) (d_y=7)

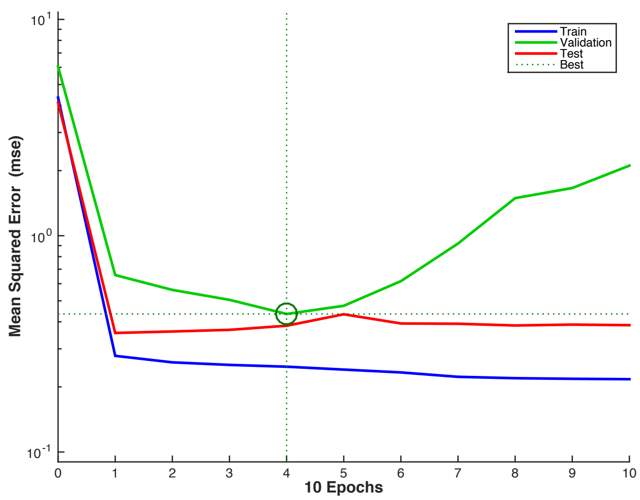
	MSE for the validation set for different numbers of runs										
No of runs / N_h	1	2	3	4	5	6	7	8	9	10	Average
1	0.595	0.579	0.655	0.580	0.568	0.639	0.640	0.600	0.632	0.507	0.599
2	0.679	0.442	0.679	0.522	0.647	0.526	0.419	0.496	0.395	0.605	0.541
3	0.537	0.541	0.514	0.513	0.602	0.520	0.483	0.541	0.589	0.681	0.552
4	0.524	0.599	0.607	0.591	0.488	0.470	0.502	0.458	0.488	0.528	**0.525**
5	0.460	0.552	0.546	0.661	0.537	0.509	0.637	0.463	0.567	0.485	0.542
6	0.518	0.566	0.574	0.436	0.560	0.882	0.408	0.398	0.763	0.478	0.558
7	0.583	0.621	0.691	0.552	0.499	0.707	0.645	0.522	0.433	0.527	0.578
8	0.614	0.536	1.111	0.700	0.480	0.581	0.555	0.533	0.720	0.497	0.633
9	0.644	0.609	0.536	0.695	0.688	1.087	0.542	0.442	0.508	0.570	0.632
10	0.840	0.497	0.521	1.331	0.548	0.792	0.489	0.840	0.530	0.733	0.712
11	0.562	0.672	0.862	0.762	1.105	0.436	0.517	0.441	0.565	0.894	0.682
12	0.898	0.608	0.529	0.482	0.604	0.564	0.520	0.758	0.910	0.482	0.635
13	0.468	0.504	0.700	0.946	0.605	0.363	0.795	1.169	0.521	0.622	0.669
14	0.675	0.575	0.663	0.601	0.907	0.636	0.537	0.517	0.974	0.563	0.665
15	0.397	0.476	0.972	0.646	0.456	0.594	0.828	0.672	0.453	0.889	0.638
16	0.496	0.497	0.618	1.299	0.599	0.521	0.552	0.556	0.881	1.115	0.713
17	0.734	0.796	0.742	0.752	0.522	0.756	0.586	0.451	0.723	0.926	0.699
18	0.557	0.652	0.535	1.521	0.801	0.455	0.631	0.808	0.566	0.519	0.705
19	0.892	1.266	0.854	0.848	0.541	0.542	0.498	0.549	0.721	1.229	0.794
20	0.634	0.836	0.766	1.295	0.624	0.688	0.774	0.893	0.715	1.616	0.884

Figure 4.1 MSE for the validation set with different numbers of hidden nodes (N_h) (d_y=7)



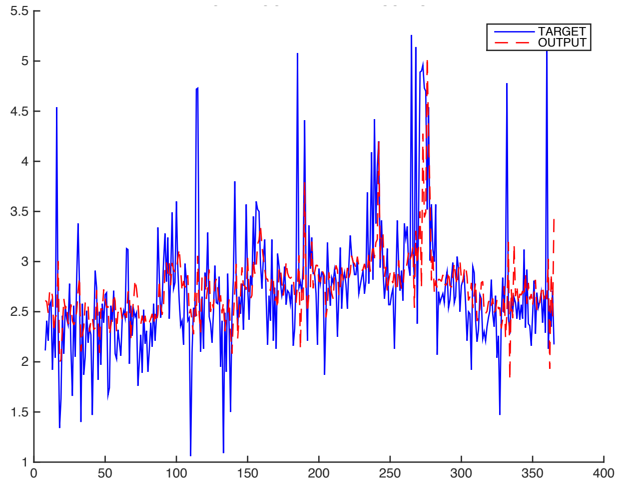
Source: Author.

Figure 4.2 MSE for training, validation and testing sets (d_y=7; N_h=4)



Source: Author.

Figure 4.3 One day ahead forecasts for spot rates (d_y=7; N_h=4)



Source: Author.

Similar procedures are performed for models with d_y ranging from 7 to 14. The best number of hidden nodes for each d_y, the corresponding MSE for the validation set (based on 10 runs) are shown in Table 4.2. The MSEs for the validation set is lowest when d_y equals to 7. The result is understandable, as it models the weekly effect. Therefore, the final NAR model for spot rates is d_y=7 and N_h=4.

Table 4.2 NAR model results for spot rates with different feedback delays (d_y)

Feedback delays (d_y)	Hidden nodes (N_h)	MSE for validation set
7	4	**0.494**
8	2	0.507
9	5	0.543
10	2	0.527
11	3	0.526
12	4	0.539
13	2	0.558
14	2	0.544

4.1.1.2 NARX model for spot rates

In this section, the historical values of contract rates are included in the model. The feedback delay d_y is set to 7, based on the results from NAR model, while the input delay d_u is initially set in the range of 1 to 7 based on cross-correlation results between spot rates and past values of contract rates. The best number of hidden nodes for each d_u is selected by the lowest MSE for the validation set. Then, the best d_u is chosen based on the best forecasting accuracy, represented by the lowest MSE for the validation set. As can be seen from Table 4.3, the lowest MSE for the validation set is 0.501, when d_u equals to 5 and N_h equals to 3. However, comparing this result with NAR model, adding historical values of contract rates as input variables does not improve forecasting accuracies.

In a short summary, the best model selected for forecasting spot rates is NAR model with d_y=7 and N_h=4.

Table 4.3 NARX model results for spot rates with different input delays of contract rates (d_u)

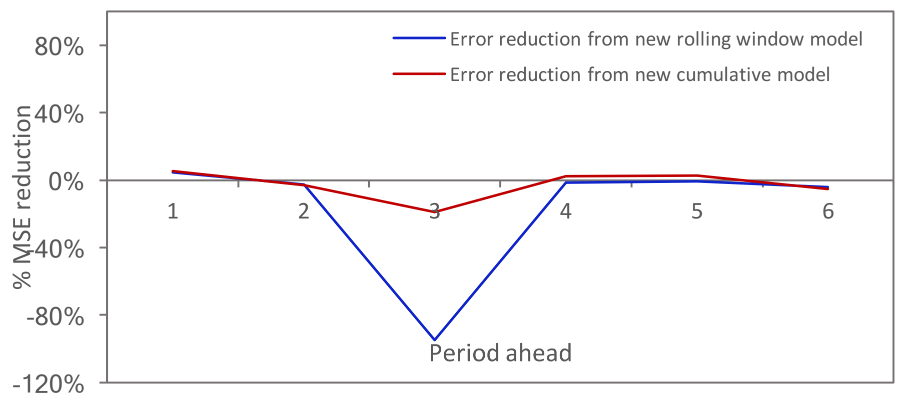
Input delays (d_u)	Hidden nodes (N_h)	MSE for validation set
1	6	0.544
2	5	0.513
3	4	0.511
4	2	0.552
5	3	**0.501**
6	2	0.530
7	2	0.559

Note: d_y=7. The same model for each d_u and N_h is run 10 times to get stable results. MSEs are the average values of 10 runs.

4.1.1.3 Model updates with new information

As previously mentioned in Section 3.1.3, once the model has been selected and trained, one of the important decisions to make is when and how to update the model when new information comes. Figure 4.4 shows MSE reductions by different models for spot rate forecasting. All the models here are using the best model selected for spot rates in the previous section, namely the NAR model with d_y=7 and N_h=4. The x-axis of the figure is the time step for update. For example, period ahead = 1 means using days 313 to 319 to forecast days 320 to 326 (one week), while period ahead =2 represents using days 320 to 326 to forecast days 327 to 333, and so forth. As can be seen, updated models do not perform better than the original model in most cases. One of the main reasons for the underperformance of updated models could be that the parameters trained in the original model may not be stable due to the high level of volatility and noise in the original data. Thus, when updating the model with original numbers of hidden nodes and feedback delays, the new model may not be the best fitted model.

Figure 4.4 MSE reductions by updating models with new information for spot rates

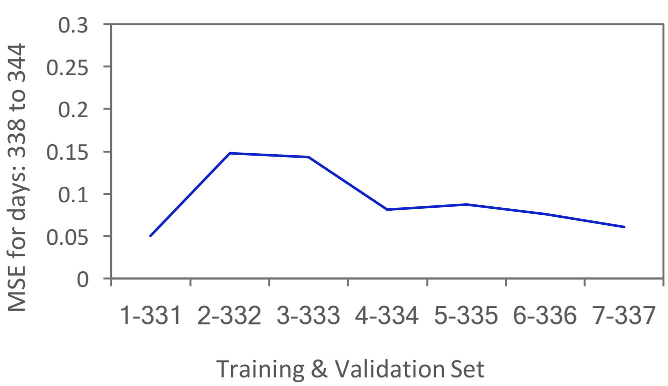


Note: 10 runs for each period ahead, d_y=7 and N_h=4.
Source: Author.

4.1.1.4 Model performance over time

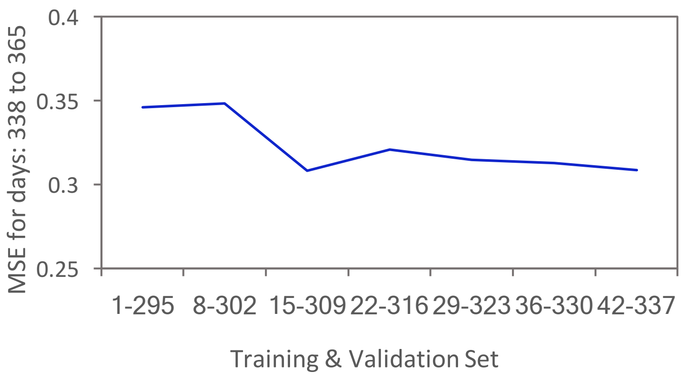
The model's performance over time is also of interest. As discussed in Section 3.1.1.2, the model's performance over time is tested over daily and weekly updates. For this performance evaluation, the rolling window update method is used as described in the methodology part. For daily updates, the MSEs are calculated for days 338 to 344. For weekly updates, the MSEs are calculated for days 338 to 365. The x-axis is the data period used for training and validation. Different periods of data are used to forecast the same data period. Figure 4.5 and Figure 4.6 plot the MSEs with rolling window daily and weekly updates respectively. As can be seen, for daily update there is no clear pattern indicating that as the date comes closer to the forecasting period, the MSE decreases. On the other hand, such pattern is observed for the weekly update. One explanation for such differences is that daily information contains high level of noise, and adding new information of a single day or several days does not help improve the model performance. On the other hand, weekly new data contains more valuable information. In this case, the MSE drops by 11% from 28-day forecasts made 7 weeks before and forecasts made 1 week before.

Figure 4.5 MSEs with rolling window daily updates for spot rates



Note: 10 runs for each period, d_y=7 and N_h=4.
Source: Author.

Figure 4.6 MSEs with rolling window weekly updates for spot rates



Note: 10 runs for each period, d_y=7 and N_h=4.
Source: Author.

4.1.2 ANN model results for contract rates

The potential input variables for predicting contract rates include the lagged values of contract rates, spot rates, contract rates for GA_C to FL_S, and volumes. In the next sections, results are presented for various NAR and NARX models.

4.1.2.1 NAR model for contract rates

The autocorrelation analysis in Section 3.2.2.1.4 indicates that lag 7 has the highest correlation with the present value of contract rates, implying strong weekly seasonality effects. Therefore, the feedback delay (d_y) is set to 7. The number of hidden nodes is set to 3, based on the lowest MSE for the validation set. The MSE for the validation set is 0.0023.

4.1.2.2 NARX models for contract rates

The input delays (d_u) for all potential input variables are all set to 7, due to higher cross-correlations with current value of contract rates. The feedback delay (d_y) for all models is set to 7 based on the results from the previous section. Table 4.4 shows the results of different NARX models for contract rate forecasting. Different combinations of input variables are included. As can be seen, the model with contract rate delays and volume delays performs best on the generalized data (validation set). Adding additional input variables does improve the forecasting performance for contract rates. In fact, the model with both contract rate delays and volume delays outperforms the model with only contract delays by 20% for the validation set. This can be partially due to the fact of tiered rates based on the lane volume. On the other hand, adding spot rate delays does not contribute to higher forecasting accuracies. This suggests relatively weak information transformations between spot and contract rates in a short period of time (weekly). Adding contract rate delays on adjacent routes also helps improve the model performance. However, the improvement is not as high as the model with the input variable-volumes. The reason why the model with both volume- and GA_C|FL_S contract rate-delays does not outperform the model with only volume delays, could be that volumes and GA_C|FL_S contract rates together contain redundant information.

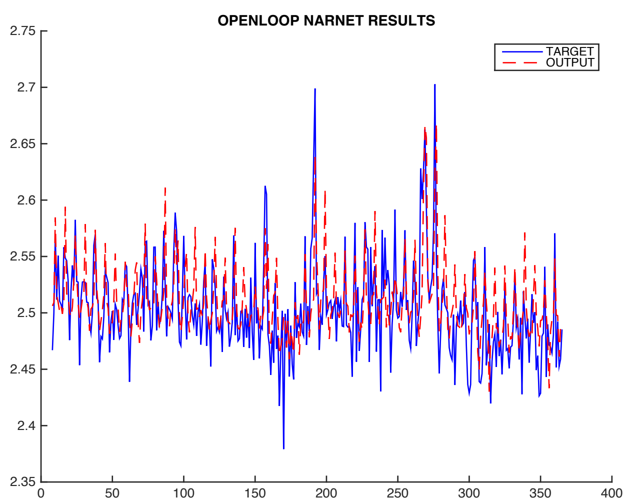
Therefore, the final model for contract rates is the model with volumes and contract rates as input variables, d_u=7, d_y=7 and N_h=2. Figure 4.7 shows the one day ahead forecasts for contract rates using the final model.

Table 4.4 NARX model results for contract rates

Input variables	Input delay (d_u)	Hidden nodes (N_h)	MSE for validation set
CR	7	3	0.00239
CR, SR	7	3	0.00235
CR, GA_C|FL_S	7	1	0.00205
CR, Volume	7	2	**0.00192**
CR, SR, GA_C|FL_S	7	2	0.00226
CR, SR, Volume	7	1	0.00210
CR, GA_C|FL_S, Volume	7	4	0.00211
CR, SR, GA_C|FL_S, Volume	7	1	0.00200

Note: Feedback delay (d_y) is set to 7 for all cases. CR stands for spot rates, SR for spot rates, GA_C|FL_S for contract rates on GA_C|FL_S. The same model for each d_u and N_h is run 10 times to get stable results. MSEs are the averaged values of 10 runs.

Figure 4.7 One day ahead forecasts for contract rates (d_u=7; d_y=7; N_h=2)

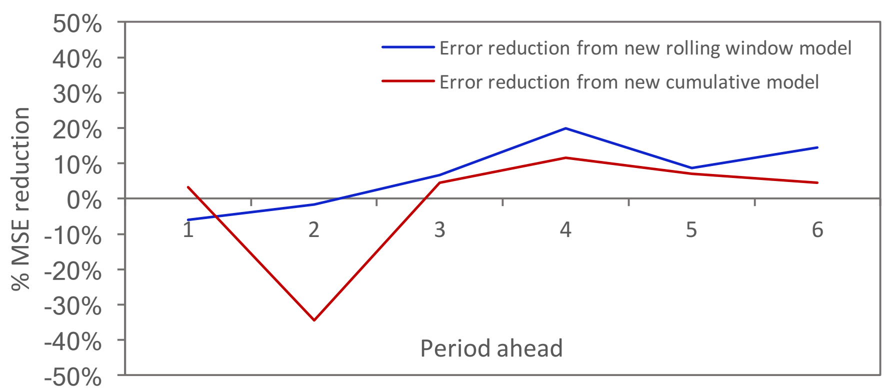


Source: Author.

4.1.2.3 Model update with new information

Similar to spot rates, the MSE reduction from updating the model with new information has been calculated. As can be seen from Figure 4.8, overall speaking, the updated model performs better than the original model from 3-period ahead onwards. Furthermore, the rolling window model performs better than the cumulative model. This might be due to the fact that contract rates are more stable, and adding more information does not help explain the model better. With regard to the question of when to update, it depends on the MSE reduction rate that the decision maker aims to achieve. In this case, the MSE reductions in 3-, 4-, 5-, 6-period ahead for rolling window models are 7%, 20%, 9% and 14% respectively. As a general rule, for contract rates, the model is recommended to update every three weeks or one month based on the results.

Figure 4.8 MSE reductions by updating models with new information for contract rates

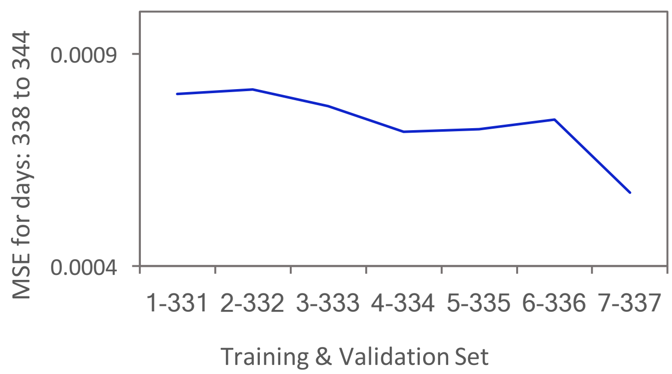


Note: 10 runs for each period ahead, d_u=7, d_y=7 and N_h=2.
Source: Author.

4.1.2.4 Model performance over time

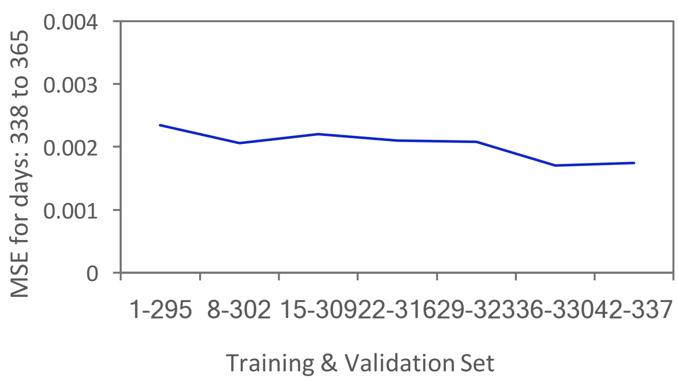
The model's performance over time is also evaluated for contract rate forecasting. The procedures are exactly the same as described in Section 4.1.1.4 for spot rates. However, unlike spot rates, the daily and weekly updates all result in MSE reductions with newer and closer information set. The MSE reduces by around 28% from 7-day forecasts made 7 days before to forecasts made 1 day before, and 30% from 28-day forecasts made 7 weeks before to forecasts made 1 week before, as shown in Figure 4.9 and Figure 4.10. This can be explained by the stable rate structure and less noise in contract rates.

Figure 4.9 MSEs with rolling window daily updates for contract rates



Note: 10 runs for each period, d_u=7, d_y=7 and N_h=2.
Source: Author.

Figure 4.10 MSEs with rolling window weekly updates for contract rates



Note: 10 runs for each period, d_u=7, d_y=7 and N_h=2.
Source: Author.

4.2 ARIMA and ARIMAX model results

4.2.1 ARIMA model results for spot rates

The number of observations for training data used of ARIMA models are set equal to the number of observations for training and validation data used for ANN models, while the testing data has the same number of observations for two types of models. Therefore, the training and testing split ratio is 85:15 for ARIMA model. As discussed in Section 3.2.2.1.3, both spot and contract rates are stationary in level forms, thus, the integrated term (d) for ARIMA (p, d, q) model is zero. Ljung-Box (LB) Q-statistic is conducted first for the autocorrelation test and the test result suggests that the null hypothesis of non-correlation is rejected at 1% level and the series demonstrates significant autocorrelation. As such, an autoregressive process would be appropriate.

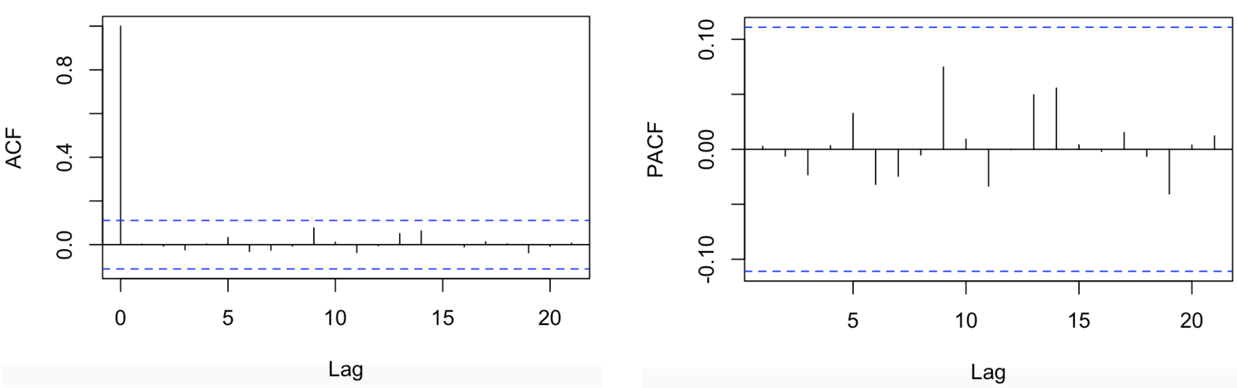
The optimal p and q lags for ARIMA (p, d, q) model are selected based on Akaike information criterion (AIC). A mean value is also included in the model. The results show that ARIMA (6,0,2) has the lowest AIC value. The coefficients of the model are presented in Table 4.5. ϕ_0 is the mean value, $\phi_p, p = 1, \dots 6$ is the AR term, and $\theta_q, p = 1,2$ is the MA term. Most of the coefficients are significant at 5% level. The MSE for the test set is 0.3189. Figure 4.11 plots the ACF and PACF for the residuals from the ARIMA model. As can be seen, there are no further remaining autocorrelation and partial autocorrelation effects, indicating good model fit. This is further confirmed by insignificant Ljung-Box (LB) Q-statistic performed on residuals from the ARIMA model.

Table 4.5 ARIMA (6,0,2) results for spot rates

	Coefficients	Std. Error
ϕ_0	2.7449**	0.0982
ϕ_1	1.632**	0.1033
ϕ_2	-0.9761**	0.1424
ϕ_3	0.1712	0.1224
ϕ_4	-0.1681	0.1292
ϕ_5	0.2732*	0.1202
ϕ_6	-0.0567	0.0683
θ_1	-1.4731**	0.0885
θ_2	0.8525**	0.0891
AIC	564.42	
Loglikelihood	-272.21	
Q (5)	2.327	

Note: ** and * represent significance at 1% and 5% levels respectively. Ljung-Box (LB) Q-statistic is the test for residual autocorrelation, conducted using 5 lags.

Figure 4.11 ACF and PACF plots for ARIMA (6,0,2) residuals



Source: Author.

4.2.2 ARIMAX model results for spot rates

Since it is shown in ANN models that including contract rates as an explanatory variable for spot rate forecasting does not improve the predicting performance, the ARIMAX model that includes contract rates as an exogenous regressor will not be discussed.

4.2.3 ARIMA model results for contract rates

Ljung-Box (LB) Q-statistic is conducted first for the autocorrelation test and the test statistics is 114.5 (significance < 1%), conducted using 5 lags. A significant Q-statistic suggests the existence of autocorrelation effects for contract rates, thus an autoregressive process could be used.

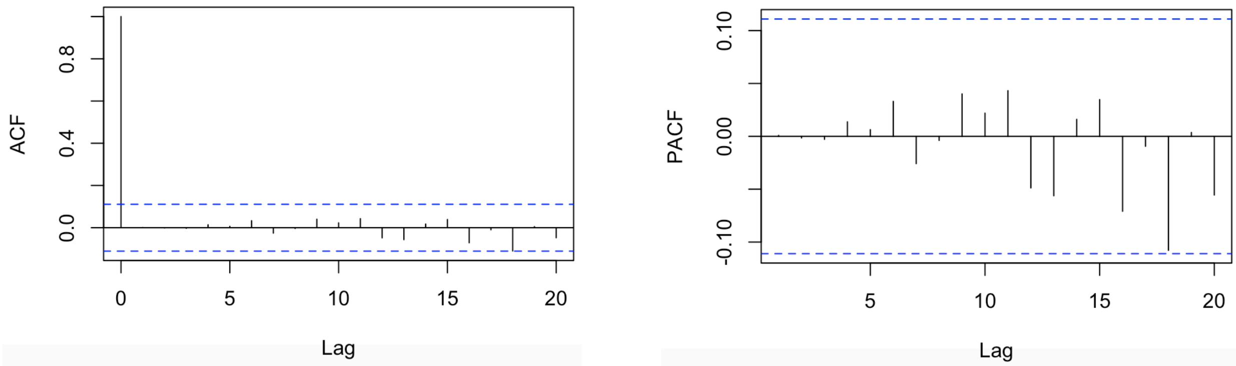
The best ARIMA model for contract rates is ARIMA (7,0,1) based on AIC. The result is provided in Table 4.6. AR (2), AR (7), and MA (1) coefficients are significant at 1% level. The MSE for the testing set is 0.00149. The residual ACF and PACF plots (Figure 4.12), and insignificant Ljung-Box (LB) Q-statistic on residuals all indicate no remaining autocorrelation and partial correlation effects.

Table 4.6 ARIMA (7,0,1) results for contract rates

	Coefficients	Std. Error
ϕ_0	2.5077**	0.0051
ϕ_1	-0.0801	0.1267
ϕ_2	0.2142**	0.068
ϕ_3	-0.0394	0.0563
ϕ_4	-0.0868	0.0569
ϕ_5	0.0276	0.0557
ϕ_6	0.0601	0.0565
ϕ_7	0.3457**	0.054
θ_1	0.4287**	0.1309
AIC	-1174.04	
Loglikelihood	597.02	
Q (5)	1.3081	

Note: ** and * represent significance at 1% and 5% levels respectively. Ljung-Box (LB) Q-statistic is the test for residual autocorrelation, conducted using 5 lags.

Figure 4.12 ACF and PACF plots for ARIMA (7,0,1) residuals



Source: Author.

4.2.4 ARIMAX model results for contract rates

Based on results from the NARX model for contract rates in Section 4.1.2.2, the potential lagged predictors include volumes and contract rates on GA_C|FL_S. The predictors are considered for up to 7 lags, that is, the model may include values of predictors one day before, and up to 7 days before that. The best model is the one with the smallest AIC value. Suppose that one predictor (either volume or contract rate on GA_C|FL_S) is added to the ARIMA model each time. Recall the equation in Section 3.1.2 for a model that allows for lagged effects:

$$y_t = \phi_0 + \beta_1 x_{t-1} + \beta_2 x_{t-2} + \cdots \beta_k x_{t-k} + n_t,$$

where n_t is an ARIMA process, the value of k can be determined using AIC. The value of p and q for ARIMA error is set to 7 and 1 based on ARIMA model in Section 4.2.3. Table 4.7 presents the AIC value for ARIMAX models with different predictors and lag lengths. As can be seen, the ARIMAX model with volume as the predictor and k set to 6 has the lowest AIC value. However, comparing this with AIC value from ARIMA (7,0,1) model, the simpler ARIMA (7,0,1) model has smaller AIC. Thus, the final ARIMA/ARIMAX model for contract rates is ARIMA (7,0,1).

Table 4.7 AIC values for ARIMAX models with different predictors and lag lengths

	AIC Value	
Lag length (k)	Predictor: Volume	Predictor: GA_C|FL_S Contract rates
1	-1150.35	-1144.70
2	-1148.56	-1142.74
3	-1147.64	-1143.98
4	-1153.72	-1145.28
5	-1157.32	-1143.38
6	**-1158.25**	-1145.95
7	-1157.35	-1144.65

4.3 Comparison between NAR and ARIMA models

In this section, the overall comparison is made between NAR models and ARIMA models. The models are compared based on Root Mean Squared Error (RMSE) due to easier interpretability. Table 4.8 shows the RMSE results for the best NAR and ARIMA models for both spot and contract rates. RMSEs for the testing set (which includes 53 day forecasts) and for 7-days rolling forecasting have been calculated. The 7-days rolling forecast is done by first using days 306 to 312 as inputs and days 313 to 319 as forecasting periods, then days 313 to 319 as inputs and days 320 to 326 as forecasting periods, …, until days 348 to 354 as inputs and days 355 to 361 as forecasting periods. Then the mean RMSE is calculated. The % difference between NAR and ARIMA models is also provided to show performance differences between different models.

For spot rates, the ARIMA model performs slightly better than the NAR model for the entire testing set. However, the NAR model has a lower RMSE than the ARIMA model for short-term forecasts. For 7-day rolling forecasts, the spot error is around $0.565 per mile over a $2.7 per mile average using the NAR model, with a coefficient of variation (COV) of 0.209. On the other hand, for contract rates, the NARX model performs much better over longer period forecasts, but not as good as the ARIMA model for short-term forecasts (although the difference is small). This is because contract rates are much stable and structured, so a simple ARIMA model could model well the short-term rate movements. For 7-day rolling forecasts, the contract error is around \$0.031 per mile over a \$2.5 per mile average using the ARIMA model and \$0.033 per mile using the NARX model. The COVs are 0.012 and 0.013 using ARIMA and NARX models respectively, indicating almost equally sound performance of the two models. Furthermore, the COV for contract rate forecasting is much smaller than spot rate forecasting, implying much less forecasting variability and much higher accuracy for contract rates. On the other hand, as it is shown in section 4.1.2.3, for contract rates, the model is recommended to update every three weeks. However, it is noted that the comparison between NARX and ARIMA model is done without model updates. Thus, the NARX model for contract rates could have a better performance if updated with new data. Furthermore, for longer term forecasts, a NARX model provides better results.

Table 4.8 RMSE comparisons between NAR and ARIMA models

Rate type	Model type	Best model	RMSE for testing set (53 days)	RMSE for 7 days (rolling forecast)
Spot	NAR/NARX	NAR with d_y=7	0.58864	0.56511
	ARIMA/ARIMAX	ARIMA (6,0,2)	0.56472	0.60167
	% Difference		4%	-6%
Contract	NAR/NARX	NARX with d_y=7, d_u=7	0.03302	0.03286
	ARIMA/ARIMAX	ARIMA (7,0,1)	0.03860	0.03082
	% Difference		-14%	7%

Note: % Difference is calculated as $(\mathrm{RMSE}_{\mathrm{NAR/NARX}} - \mathrm{RMSE}_{\mathrm{ARIMA/ARIMAX}})/\mathrm{RMSE}_{\mathrm{ARIMA/ARIMAX}}$ as a way to measure relative performance of two types of models.

4.4 Summary

This section compares different NAR/NARX and ARIMA/ARIMAX models. Results show that overall speaking, contract rates have much higher forecasting accuracy and less forecasting variability compared to spot rates. Furthermore, for spot rates, the NAR model has better short-term forecasting results compared to the ARIMA model, while for contract rates, NARX and ARIMA models provide almost equally good forecasting results. For a longer-term forecast, the NARX model is more accurate than the ARIMA model for contract rates.

5 Conclusions

This study has developed a forecasting model that predicts both contract and spot rates for truckload transportation on individual lanes for the next seven days. The model used is a neural network model based on Nonlinear Autoregressive Models with eXogenous input (NARX). It is a hybrid model that incorporates both non-linear features of neural network models and autoregressive features of time series models. This study considers several input variables, including lagged values of spot and contract rates, rates on adjacent routes and volumes. The best NAR/NARX models for spot and contract rates are selected based on highest forecasting accuracy on the validation set. The NAR/NARX model is also compared with traditional time series models (ARIMA and ARIMAX). This section summaries the major findings and contributions, and identifies research limitations.

There are five key findings of this thesis. First, generally speaking, the NAR model provides better short-term forecasting performance for spot rates than the ARIMA model, while the ARIMA model performs slightly better for contract rates. However, for a longer-term forecast, the NARX model provides better results for contract rates.

Second, the best NAR/NARX model for spot rates is NAR model with seven days feedback delay. The best ARIMA/ARIMAX model is ARIMA (6,0,2). Adding additional information, such as past values of contract rates and volumes, does not improve the model's performance. Furthermore, updating the model with new information does not help improve the forecasting accuracy. In addition, the forecasting accuracy increases only slightly as the dates come closer to the forecasting periods. All these results indicate that spot rates contain high levels of noise, and the resulting neural network is unstable.

Third, the best NAR/NARX model for contract rates is NARX model with seven days feedback delay and seven days input delay of volumes. The best ARIMA/ARIMAX model is the ARIMA (7,0,1) model. For contract rates, retraining of the model increases the model's performance, and using newer and closer information sets for the same model also improve the model's performance. Therefore, the general guideline for contract rate forecasts is to update new information using the original model every week and retrain the original model every month.

Furthermore, the model's prediction of contract rates is much more accurate and has much less forecasting variability than spot rates. A high forecasting accuracy can be achieved for contract rates either using NARX or ARIMA models. However, spot rates are difficult to forecast due to high variability.

Last but not least, results show that there exists no short-term information transmission between spot and contract rates. This can be due to the fact that contract rates are often negotiated for a one-year period, which reflects future market expectation at the time of contract negotiations. On the other hand, spot rates often reflect current market supply/demand dynamics.

This study has made several contributions. First, this research has made methodological advancements by introducing the hybrid neural network and time series model (NAR and NARX) into the transportation forecasting field. The model is shown to have better short-term forecasting abilities for spot rates, as well as accurate forecasts for contract rates in both shorter (seven days) and longer terms (two months). Second, the results from this study can be applied to industrial players for their own forecasting. These results provide guidelines for both shippers and carriers regarding how to select input variables, what model to use, when to update the model with new information, and what forecasting error is normally expected from the model.

Limitations of this study exist. The classifications between spot rates and contract rates in the dataset are reported by various companies. However, some rates can be misclassified, thus resulting in higher modelling errors. In addition, this study only considers one year of data, which makes it hard to model in monthly seasonality effects.

6 Future research

This thesis has considered various forecasting techniques and different input variables for TL rate forecasting. However, due to research scope and data limitations, there are potential areas that were not fully explored, but would be worth investigating in the future:

- It is interesting to see whether a classification technique could be used to automatically classify a new rate (either spot or contract), based on rate behaviour rather than self-reported.
- It is also worthwhile to investigate seasonal effects (for example, the month of the year effect) on spot rate forecasting using data that lasts for a longer period.
- This study only considers rate forecasting for one specific high-volume lane. Future research could potentially extend the current model to different lanes and test the model's performance across lanes.
- The impact of neighbouring or adjacent lanes/regions could be further explored. This study only considers three adjacent lanes for GA_C to FL_C data. In the future work, an algorithm to automatically filter out highly correlated lanes among all possible neighbouring lanes could be developed.

- The information transmission between spot and contract rates over a longer period of time is also of interest. For example, how the spot rate movements in the past year affect this year's contract negotiation rates could be examined.

Through applying the models established in this thesis and investigating the future research areas noted above, both shippers and carriers will be able to better estimate transportation costs. This can help shippers make better decisions when planning transportation budgets and help carriers estimate future cash flows.

7 References

American Trucking Trends. (2016). U.S. Freight Transportation Forecast to 2027. Retrieved November 20, 2017, from https://www.epa.gov/sites/production/files/2017-04/documents/sw-feight-trends-2017-03-29.pdf

Andalib, A., & Atry, F. (2009). Multi-step ahead forecasts for electricity prices using NARX: A new approach, a critical analysis of one-step ahead forecasts. *Energy Conversion and Management*, *50*, 739–747. https://doi.org/10.1016/j.enconman.2008.09.040

ATA. (2017). American Trucking Associations, Industry data. Retrieved November 15, 2017, from http://www.trucking.org/News_and_Information_Reports_Industry_Data.aspx

Bailey, D., & Thompson, D. (1990). How to Develop Neural-network Applications. *AI Expert*, *5*(6), 38–47.

Baker, J. A. (1991). Emergent Pricing Structures in LTL Transportation. *Journal of Business Logistics; Hoboken*, *12*(1), 191.

Ballou, R. H. (1991). The accuracy in estimating truck class rates for logistical planning. *Transportation Research Part A: General*, *25*(6), 327–337. https://doi.org/10.1016/0191-2607(91)90011-E

Box, G. E. P., & Jenkins. (1970). *Time Series Analysis: Forecasting and Control*. San Francisco: Holden-Day.

Budak, A., Ustundag, A., & Guloglu, B. (2017). A forecasting approach for truckload spot market pricing. *Transportation Research Part A: Policy and Practice*, *97*(Supplement C), 55–68. https://doi.org/10.1016/j.tra.2017.01.002

Burwell, T. H., Dave, D. S., Fitzpatrick, K. E., & Roy, M. R. (1997). Economic lot size model for price-dependent demand under quantity and freight discounts. *International Journal of Production Economics*, *48*(2), 141–155. https://doi.org/10.1016/S0925-5273(96)00085-0

Caldwell, E. R. (Erik R., & Fisher, B. C. (2008). *The impact of lead time on truckload transportation rates* (Thesis). Massachusetts Institute of Technology. Retrieved from http://dspace.mit.edu/handle/1721.1/45252

Carter, J. R., & Ferrin, B. G. (1996). Transportation costs and inventory management: Why transportation costs matter. *Production and Inventory Management Journal; Alexandria*, *37*(3), 58–62.

CFRA. (2017). CFRA's Industry Surveys Reports Road&Rail.pdf.

Chen, S., Billings, S. A., & Grant, P. M. (1990). Non-linear system identification using neural networks. *International Journal of Control*, *51*(6), 1191–1214. https://doi.org/10.1080/00207179008934126

Costello, B. (2013). The Trucking Industry: The Lynchpin of the U.S. Economy. *Business Economics*, *48*(3), 195–201. https://doi.org/10.1057/be.2013.16

Ersoy, O. K., & Hong, D. (1990). Parallel, self-organizing, hierarchical neural networks. *IEEE Transactions on Neural Networks*, *1*(2), 167–178. https://doi.org/10.1109/72.80229

Fan, S., Ji, T., Gordon, W., & Rickard, B. (2013). Forecasting Baltic Dirty Tanker Index by Applying Wavelet Neural Networks. *Journal of Transportation Technologies*, *03*(01), 68–87. https://doi.org/10.4236/jtts.2013.31008

Gao, Y., & Er, M. J. (2005). NARMAX time series model prediction: feedforward and recurrent fuzzy neural network approaches. *Fuzzy Sets and Systems*, *150*(2), 331–350. https://doi.org/10.1016/j.fss.2004.09.015

Garrido, R. A. (2007). Procurement of transportation services in spot markets under a double-auction scheme with elastic demand. *Transportation Research Part B: Methodological*, *41*(9), 1067–1078. https://doi.org/10.1016/j.trb.2007.04.001

Glen, D. R. (2006). The modelling of dry bulk and tanker markets: a survey. *Maritime Policy & Management*, *33*(5), 431–445. https://doi.org/10.1080/03088830601020562

Han, Q., Yan, B., Ning, G., & Yu, B. (2014). Forecasting Dry Bulk Freight Index with Improved SVM [Research article]. https://doi.org/10.1155/2014/460684

Hubbard, T. N. (2001). Contractual Form and Market Thickness in Trucking. *The RAND Journal of Economics*, *32*(2), 369–386. https://doi.org/10.2307/2696415

Karlaftis, M. G., & Vlahogianni, E. I. (2011). Statistical methods versus neural networks in transportation research: Differences, similarities and some insights. *Transportation Research Part C: Emerging Technologies*, *19*(3), 387–399. https://doi.org/10.1016/j.trc.2010.10.004

Kay, M. G., & Warsing, D. P. (2009). Estimating LTL rates using publicly available empirical data. *International Journal of Logistics Research and Applications*, *12*(3), 165–193. https://doi.org/10.1080/13675560802392415

Kristjanpoller, W., & Minutolo, M. C. (2015). Gold price volatility: A forecasting approach using the Artificial Neural Network–GARCH model. *Expert Systems with Applications*, *42*(20), 7245–7251. https://doi.org/10.1016/j.eswa.2015.04.058

Li, J., & Parsons, M. G. (1997). Forecasting tanker freight rate using neural networks. *Maritime Policy & Management*, *24*(1), 9–30. https://doi.org/10.1080/03088839700000053

Lin, T., Horne, B. G., Tino, P., & Giles, C. L. (1996). Learning long-term dependencies in NARX recurrent neural networks. *IEEE Transactions on Neural Networks*, *7*(6), 1329–1338. https://doi.org/10.1109/72.548162

Lyridis, D. V., Zacharioudakis, P., Mitrou, P., & Mylonas, A. (2004). Forecasting Tanker Market Using Artificial Neural Networks. *Maritime Economics & Logistics*, *6*(2), 93–108. https://doi.org/10.1057/palgrave.mel.9100097

Mendoza, A., & Ventura, J. A. (2009). Estimating freight rates in inventory replenishment and supplier selection decisions. *Logistics Research*, *1*(3–4), 185–196. https://doi.org/10.1007/s12159-009-0018-5

Menezes, J. M. P., & Barreto, G. A. (2008). Long-term time series prediction with the NARX network: An empirical evaluation. *Neurocomputing*, *71*(16), 3335–3343. https://doi.org/10.1016/j.neucom.2008.01.030

Özkaya, E., Keskinocak, P., Roshan Joseph, V., & Weight, R. (2010). Estimating and benchmarking Less-than-Truckload market rates. *Transportation Research Part E: Logistics and Transportation Review*, *46*(5), 667–682. https://doi.org/10.1016/j.tre.2009.09.004

Rumelhart, D. E., Hinton, G. E., & Williams, R. J. (1986). Learning representations by back-propagating errors. *Nature*, *323*(6088), 533–536. https://doi.org/10.1038/323533a0

Sheffi, Y. (2004). Combinatorial Auctions in the Procurement of Transportation Services. *Interfaces*, *34*(4), 245–252. https://doi.org/10.1287/inte.1040.0075

Siegelmann, H. T., Horne, B. G., & Giles, C. L. (1997). Computational capabilities of recurrent NARX neural networks. *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, *27*(2), 208–215. https://doi.org/10.1109/3477.558801

Smith, L. D., Campbell, J. F., & Mundy, R. (2007). Modeling net rates for expedited freight services. *Transportation Research Part E: Logistics and Transportation Review*, *43*(2), 192–207. https://doi.org/10.1016/j.tre.2005.11.001

Swenseth, S. R., & Godfrey, M. R. (1996). Estimating freight rates for logistics decisions. *Journal of Business Logistics; Hoboken*, *17*(1), 213–231.

Swenseth, S. R., & Godfrey, M. R. (2002). Incorporating transportation costs into inventory replenishment decisions. *International Journal of Production Economics*, *77*(2), 113–130. https://doi.org/10.1016/S0925-5273(01)00230-4

Tyworth, J. E., & Ruiz-Torres, A. (2000). Transportation's role in the sole- versus dual-sourcing decision. *International Journal of Physical Distribution & Logistics Management*, *30*(2), 128–144. https://doi.org/10.1108/09600030010318847

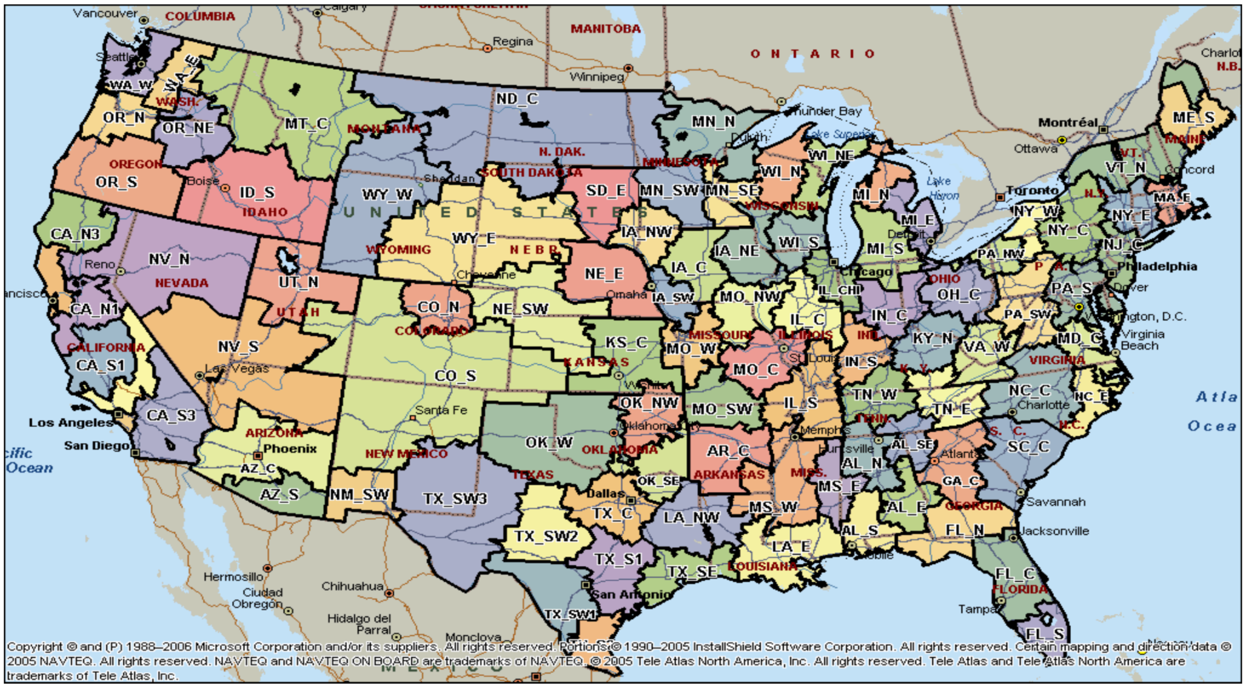
Zhang, G. P. (2003). Time series forecasting using a hybrid ARIMA and neural network model.

Neurocomputing, *50*, 159–175. https://doi.org/10.1016/S0925-2312(01)00702-0

Zou, H. F., Xia, G. P., Yang, F. T., & Wang, H. Y. (2007). An investigation and comparison of artificial neural network and time series models for Chinese food grain price forecasting. *Neurocomputing*, *70*(16), 2913–2923. https://doi.org/10.1016/j.neucom.2007.01.009

8 Appendix

Appendix A: US region breakdown



Source: MIT CTL Freight Lab, 2018.