A Time Series Model for the China-to-U.S. Ocean Freight Pricing

by

Yuchen Yvonne Cao

Bachelor of Arts, Business, Brandeis University, 2017

Bachelor of Arts, Politics, Brandeis University, 2017

SUBMITTED TO THE PROGRAM IN SUPPLY CHAIN MANAGEMENT
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF

MASTER OF APPLIED SCIENCE IN SUPPLY CHAIN MANAGEMENT
AT THE
MASSACHUSETTS INSTITUTE OF TECHNOLOGY

May 2020

Signature of Author: ____

Yuchen Yvonne Cao
Department of Supply Chain Management
May 1, 2020

Signature of Author: ____

Department of Supply Chain Management
May 1, 2020

Certified by: ____

Dr. Josué C. Velázquez Martínez
Executive Director, Supply Chain Management Program
Capstone Advisor

Certified by: ____

Dr. Ozden Tozanli
Postdoctoral Associate
Capstone Co-Advisor

Accepted by: ____

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

A Time Series Model for the China-to-U.S.
Ocean Freight Pricing
By
Yuchen Yvonne Cao
Submitted to the Program in Supply Chain Management
on May 1, 2020 in Partial Fulfillment of the
Requirements for the Degree of Master of Applied Science in Supply Chain Management

ABSTRACT

Ocean freight forwarding on the China-to-U.S. lane is a key service that C.H. Robinson, the sponsoring company, offers to the firm's international clients. The rates on that lane have experienced volatility in the past few years which led to uncertainties to the future pricing trend. A statistically predictive model that forecasts the future pricing trend can help to resolve this challenge. This capstone studies two approaches: a time series forecasting model, and a time series forecasting model with exogenous factors. These models are used to build a predictive model to forecast the future ocean freight rate. Economic indicators are selected as the independent variables in this research. After comparing 14 time series models including Autoregressive Integrated Moving Average (ARIMA) and Exponential Smoothing model, the results show that a Multiplicative Seasonality (with no trend) Exponential Smoothing Model provides the best-fit forecasting metric. We also discovered that the best-fit model is less sensitive to error when the analysis assigns more weight to the most recent observation and the error rate increases rapidly right after the model experiences a sharp drop in the historical data rates. Some economic indicators show a correlation with the historical ocean freight rates; however, they do not improve the accuracy of the model with or without lags in the period. Therefore, we concluded that a multiplicative seasonality (with no trend) exponential smoothing model can best predict the future pricing of the China-to-U.S. ocean freight rates.

Capstone Advisor: Dr. Josué C. Velázquez Martínez
Title: Executive Director, Supply Chain Management Program
Capstone Co-Advisor: Dr. Ozden Tozanli
Title: Postdoctoral Associate

ACKNOWLEDGMENTS

I want to thank the sponsor company, C.H. Robinson, for giving me full support from understanding the dynamics of the ocean market to finalizing the methodologies and results. I also want to thank my advisors, Josue and Ozden, for providing me the guidance and support wherever it was needed. I want to thank Dr. Tugba Efendigil for being a key collaborator for the first steps of my capstone. I want to thank Toby and Pamela for the incredible helps in structuring the project report and revising the writings. Lastly, I want to thank MIT and MIT Supply Chain Management program for giving me such a wonderful opportunity to work on this project and providing supports throughout the journey of my master's program. Thank you.

Yvonne Cao

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

1 INTRODUCTION

The project is sponsored by C.H. Robinson (NASDAQ: CHRW), North America's largest 3rd party logistics service provider with annual revenue of $15.3 billion in 2019 and net income from operation of $790.0 million (C.H. Robinson, 2020). In 2018, C.H. Robinson handled over 1.5 million ocean freight forwarding shipments, equivalent to 8.8% of its total business volume (C.H. Robinson, 2019). The ocean business has also ranked as the largest non-vessel operating common carrier (NVOCC) for the route from The China-to-U.S. (C.H. Robinson, 2019). The success of the ocean route from the China-to-U.S. is crucial to C.H. Robinson's global forwarding business and therefore, greatly contributed to the overall success of the company.

According to the CHR executives interviewed for this project, currently there are no predictive models to assist with forecasting next year's budget. The current forecast is mainly based on personal experience and judgment on the historical trend and major events in the market. Since freight rates are mainly impacted by the market demand for international goods volume and the market supply of vessel volume, major changes in these areas cause the rates to fluctuate. Moreover, the ocean freight rate market is volatized with the minimum and maximum varying by over 35% (C.H. Robinson, 2020). If we can guide the industry on the pricing trend, not only will C.H. Robinson benefit from a more accurate rate forecast, but also the shippers can obtain a better judgment on their demand planning based on the transportation expense forecast.

The objective of the project is to create a predictive model (i.e. time series model) to forecast the future pricing of the China-to-U.S. lane by using the data from C.H. Robinson's historical ocean pricing, public economic indicators, and other carrier data resources.

The main methodology of the project is to explore forecasting models, including but not limited to Auto-Regressive Integrated Moving Average (ARIMA) model, Exponential Smoothing model, ARIMA with an exogenous factor model that can predict the future trend of the China-to-U.S. ocean freight rates.

Additionally, we leverage quantitative correlation analysis to identify exogenous economic indicators with positive contributions to improve the accuracy of the forecasting model. This capstone focuses on determining whether rates have seasonality and cyclical patterns or any positive or negative trend that shows the tendency of future rates, or whether the rates are independent from historical rates but are impacted by market movements. In the literature review section, we explore information on the history of ocean freight market, different exogenous indicators leveraged by previous researches to predict future pricing and the methodologies applied for statistical predictive modeling. This capstone further develops methodologies to approach the research question to predict future ocean freight pricing and provide findings and business interpretation on the result.

2 LITERATURE REVIEW

The literature review is structured into three different sections to illustrate the market dynamics of the ocean freight industry, to explore different economic indicators that can potentially influence the pricing, and to demonstrate forecasting models that can consider all these effects and predict the future trend. We leverage similar factors and methodologies, and apply to the China-to-U.S. ocean freight rate.

2.1 Market Dynamics

Containerized ocean freight is a volatilized market within the transportation industry, especially after the 2008 financial crisis. Literatures focus on using independent and unaccompanied market events to explain the variation within the ocean freight forwarding market. The reasonings can be grouped into the followings: overcapacity, carrier consolidation, and alliances, shifting demand, changes in trade policies and variation in the cost of fuel.

After the 2008 financial crisis, major ocean carriers started to build mega vessels to pursue economies of scale. From 2009 to 2018, the annual containership capacity grew at an average rate of 6.1% (Alphaliner, 2019). Even though supply has increased consistently year over year, demand has failed to catch up as previously forecasted. Using gross domestic product (GDP) as the indicator of the overall demand, the ratio between the growth of 20 foot equivalent units to the growth of GDP has decreased from 3.4 in 1990 – 1999 to 2.6 in 2000 – 2008, then further declined to 1.4 in 2010 to 2018,(Alphaliner, 2019). Shipbuilding has long lead time, and when the market realized the market growth rate is much lower than previously anticipated, it was too late to pause the production. By 2017, the market had reached substantial overcapacity and after years of financial suffering, one of the major Asian carriers, Hanjin Shipping Co., Ltd, declared bankruptcy (Nam, 2017).

To improve their challenging financial situations, the carriers started to form alliances, which could help them to be more flexible in terms of the routings and to leverage the economies of scale of the group, and therefore share the risk and investment. In 2017, carrier consolidation and alliances reached their peak, with 20 independent carriers before consolidating into 12 alliances (Laxmana, 2017).

Within the US market, there are also rate fluctuations between the U.S. West Coast and U.S. East Coast. The port worker strike on the United States West Coast in 2014 forced some shipments to shift to the East Coast and some of the demand has permanently stayed with East Coast (Pinsker, 2015). Moreover, the economy of the Southeast region in the United States is growing faster than in other regions, which affected the demand distribution within the country. Since the East Coast is closer to the Southeast region than the West Coast, more demand has transitioned to be cleared on the East Coast. Lastly, the expansion of the Panama Canal in 2016 allowed bigger and heavier ships to reach the East Coast (Link, 2017); as a result, the East Coast capacity has increased, and demand has grown. Although the East Coast only handled around one-third of the total shipping capacity in the United States from 2014 to 2019 (C.H. Robinson, 2019), it is playing an increasingly critical role and bringing changes to the dynamics of the traditionally West Coast dominated shipping market.

2.2 Exogenous Factors: Economic Indicators

Previous literature has demonstrated different exogenous factors that can potentially impact rates in different transportation modes for both international and domestic lanes. In early 1946, Marx researched the impact of war on ocean freight rates within West Europe region. He concluded based on qualitative findings that oversupply in shipping capacities, the shift in demand driven by GDP and employment rates, the cost of fuel to operate vessels, the carrier's economy of scale, and the social policies such as war, affect ocean freight rates (Marx, 1946).

Even though the research was done nearly 80 years ago, there are similarities between the condition now and then. In 2008, research shows the impact of supply, demand and social events on ocean freight rate quantitatively for the period of 1850s (Klovland, 2008). A financial model to predict ocean freight rates was built based on the freight rate index, price level, and shipping tonnage. The exchange rate has also been determined as a relevant economic indicator in another study where the researcher examined the freight flows between the U.S. and China (Chi, 2016). According to our interviews with the C.H. Robinson team, the executives suggested that the Consumer Price Index (CPI) and Purchasing Manager Index (PMI) could potentially affect the ocean freight, as the two indexes are greatly associated with the purchasing demand of the shipper (C.H. Robinson, 2020). In this research, the forecasting models explore similar economic indicators, as those discussed in the literature for different regions and times, to understand the relationship between the exogenous indicators and future ocean pricing.

2.3 Forecasting Model

Forecasting models consist of two main categories, statistical forecasting, and judgmental forecasting. Statistical forecasting relies heavily on the historical data of the set, while judgmental forecasting focuses on the current known factors (Silver, Pyke and Peterson, 1998). Additionally, Elarbim (2013) shows that if a predictive model is solely developed based on judgmental forecasting, the result will not be optimal since it is not based on both quantitative and qualitative analysis. Elarbi (2013) further categorized statistical forecasting into four methods: qualitative analysis, time-series analysis, causal analysis and simulation. Time-series is the most commonly adopted. According to Elarbi, the time-series analysis involves five key parameters: level (a), trend (b), seasonal variations (F), cyclical movement (C) and irregular random fluctuation (E) (Elarbi, 2013). In 2018, Adland, Benth, and Koekebakker (2018) proposed a multivariate co-integrated time-series model to predict regional spot freight rates. In their research, ocean freight rates were decomposed into a non-stationary market factor and some stationary factors that are correlated and vary by region.

To incorporate exogenous economic indicators into the prediction, autoregressive integrated moving average with exogenous variables (ARIMAX) models can be another approach to explore. In 2013, Andrews, Dean, Swain and Cole (2013) conducted research on building an ARIMAX models to predict long-term disability benefit application-rates by using external indicators such as the competitor's activities, the economic and governmental regulations. This model is capable of determining the underlying behavior of time-series data and to measure the influence of other environmental factors. The data structure of the disability-benefit application-rates data set is very similar to the ocean freight data set in natures where both include dates for a consecutive time periods and rates associate with the dates. Moreover, the environmental factors selected in the literature have some similarities to the ones that are mentioned in Section 2.2.

A review of the current market circumstances indicates that recent historical rate movements are mainly associated with individual events such as financial crisis, labor strike, carrier consolidation, and policy implementation. By looking at academic research focusing on other regions and timelines, this research can leverage similarities such as using different economic indicators to predict future pricing. Lastly, by exploring different forecasting models, especially time series, the analysis can leverage the pattern of the data's own historical values to foresee future trends. In this research, the analysis compares different time series forecasting models with or without exogenous variables to determine the model with the best fit.

3 DATA AND METHODOLOGY

3.1 Methodology Overview

In the methodology section, each necessary step is illustrated and led to the final conclusion as shown in Figure 1.

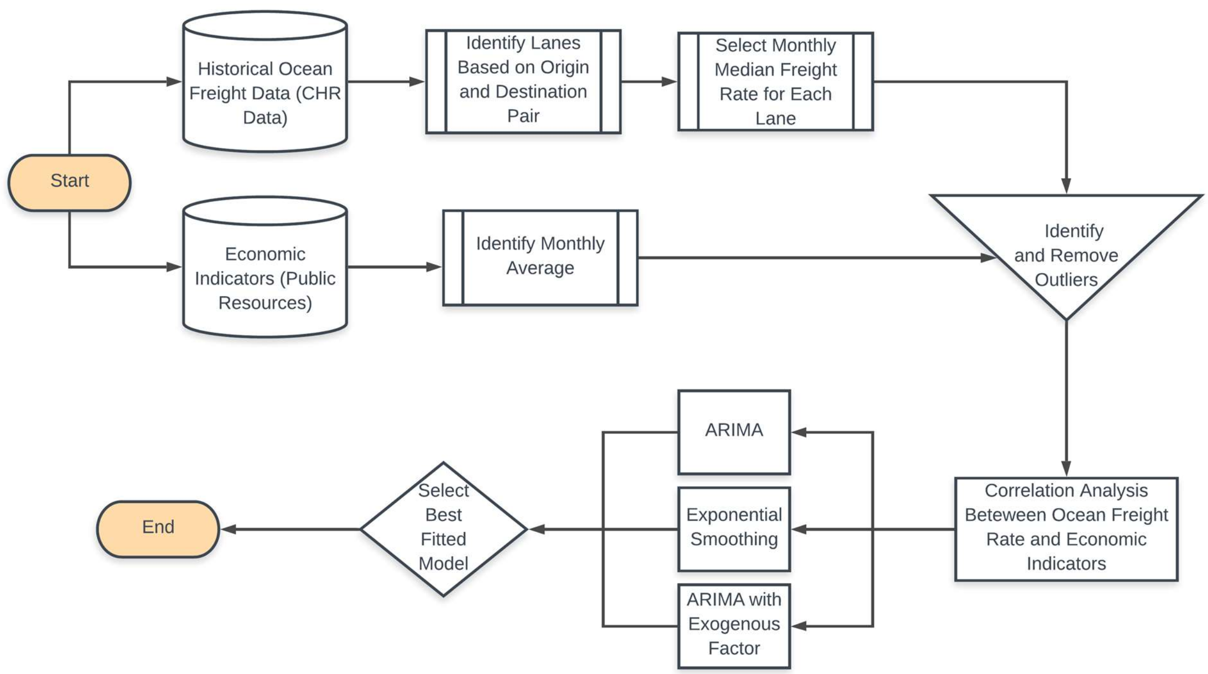

Figure 1. Project Methodology Flowchart

1. **Data Collection:** This step laid the foundation for the quantitative analysis of the project and ensures the accuracy of the research moving forward by retrieving information regarding historical ocean freight data and economic indicators data.

2. **Data Cleaning:** In this step, we identified lanes based on the original and destination pairs, collect the monthly average data and remove any outliers for both the ocean freight and economic indicators.
3. **Correlation Analysis:** We conducted a correlation analysis, a statistical evaluation to measure the strength of a relationship between two continuous variables (Cohen, Cohen, West, and Aiken, 2003). In this project, we ran the correlation analysis between all potential economic indicators defined in Section 3.2 with the historical ocean freight rates to determine the most correlated indicator for different ocean routes.
4. **Time Series Model:** We selected the best-fit time series forecast among 14 different models referenced in Table 9, among three different categories, Auto-Regressive Integrated Moving Average (ARIMA), Exponential Smoothing, and ARIMA model with an exogenous factor.

3.2 Data Collection

We collected the 2014 - 2019 historical ocean freight rates from our sponsoring company, C.H. Robinson with over 15,000 records as the main data set to initialize the analysis. We also collected data on economic indicators and oil prices as potential variables that affect ocean freight rates.

3.2.1 Dependent Variable: Ocean Freight Rate

In correlation analysis, an independent variable is defined as a variable whose variation does not depend on others, and the dependent variable is defined as a variable whose value is depended on the independent variable (Cohen, Cohen, West, and Aiken, 2003). In this project, the dependent variable, ocean freight rate, was given by the sponsor company. As a freight forwarder, C.H. Robinson negotiated these rates with different shippers either as fixed contracts or spot rates. The data is structured as shown in Table 1 below.

Table 1. Summary of Ocean Freight Rates File

	Field	Definition
1	ValidDate	Date that the rate is valid for
2	Origin	Origin of the shipment
3	POL	Port of Loading, the port at which the goods are loaded on to the vessel
4	POD	Port of Destination, the port at which the goods are landed
5	Destination	Destination of the shipment
6	Currency	The currency of the rate
7	Ocean	Rate
8	ContainerType	Type of container on the vessel

The date is structured by individual dates from 2014 to 2019 For each date, there should be multiple entries since different carriers offer different rates. “Origin” and “Destination” are indicated by the three-letter code of the origin of the shipments and the three-letter code of the destination of the shipments. The Origin can be the same or can be different from the Port of Loading (POL) and the Destination can be the same or different from the Port of destination (POD). For this research, we only selected origin and destination pairs based on the POL and POD. The sponsor company selected six Ports of Loading for analysis, including CNSHA, CNNGB, CNQIN, CNYTN, CNSZX, and CNXMN. For POD, the sponsor company selected 11 destinations which are separated into two categories, U.S. West Coast and U.S. East Coast. For East Coast destinations, C.H. Robinson also separated them into four subcategories as shown in Table 2. For us to collect sufficient data, the rate is separated into two groups, the China-to-U.S. West Coast (CNWC) and the China-to-U.S. East Coast (USEC). Currency contains USD as the sole currency for the analysis and container type is set at 40ft.

Table 2. Categorize Destinations

Destination

West Coast		East Coast			
WC1	WC2	EC1	EC2	NewYork	Gulf
TIW	LAX	CHS	BOS	NYC	HOU
		SAV	MIA		
		ORF	JAX		
			BAL		

3.2.2 Economic Indicators

For the independent variables, based on the literature review and executive interviews, the sources were identified as shown in Table 3. Among the indicators, GDP is shown as a monthly data point and all other indicators are shown as daily data points.

Table 3. Summary of Independent Parameters and Sources

Economic Indicator	**Data Source / Comments**
U.S. GDP growth rate (%), U.S. GDP per capita growth rate (%) U.S. GDP ($), GDP per capita ($)	U.S. Bureau of Economic Analysis (2019)
China GDP growth rate (%), China GDP per capita growth rate (%) China GDP ($), GDP per capita ($)	The World Bank (2019)
U.S. Consumer price index (CPI) – Non-seasonally adjusted	U.S. Bureau of Labor Statistics (2019)
China Consumer price index (CPI) – Non-seasonally adjusted	The World Bank (2019)

Currency exchange rate	International Monetary Fund Currency Exchange Database (2019)
Purchasing Manager Index	Institute for Supply Management (ISM)
Crude oil price	Brent crude: NASDAQ database (2019)

3.3 Data Cleaning

Since GDP is an important economic indicator measuring the total value of everything produced in the country, the data point must be efficiently reflected in the analysis. Therefore, we categorize all data sets monthly to accommodate the GDP's monthly data points. For ocean freight rate, there are multiple data points in a day, based on different shippers and origin to destination routes. Therefore, there is a high possibility that using the average value is impacted by outliers or rates that are placed in the system but were never utilized. Therefore, in our analysis, we use the median of every month as the reference point. For other economic indicators, there is only one rate for each day. Using the average will not expose to a higher risk of data inconsistency, therefore we selected the average of every month as the reference point for these data sets. Once all data is converted into monthly data, the proposed model merges all data sets into one master file.

Once the master file was collected, it was necessary to remove any outliers that can impact the correlation analysis conducted later in the research. The analysis is designed to cover 95% of the data which is two standard deviations around the mean to eliminate the outliers, data points that significantly differ from other observations (Cohen, Cohen, West, and Aiken, 2003). Out of 62 data points, we identified 4 outliers including the data points of Feb 2015, Jun 2016, July 2016, and Aug 2016. After removing the outliers, we conducted a normalization process to bring all data sets to the same scale while keeping the result of the correlation analysis unaffected. The analysis was completed by calculating the average and standard deviations of the data set to transform the original date based on Z score using Equation 1, where x is the raw data, μ is the population mean and σ is the population standard deviation of the data sets.

$$z = (x - \mu)/\sigma \tag{1}$$

3.4 Correlation Analysis

Correlation analysis is used to test hypotheses generated by researchers from different area including in the business and industry. They can be generated based on previous experience, literature review or formal theories to test some type of relationship between one or more indicators (Cohen, Cohen, West, and Aiken, 2003). There are usually two variables in correlation analysis, and the approach is targeted to find a statistical relationship between the two variables. Using correlation analysis, we can understand the relationship between ocean freight rates, a dependent variable, and an economic indicator, an independent variable, and show how strongly they are correlated and related to each other, whether in a negative or positive direction.

In the correlation analysis, the Pearson Product Moment correlation coefficient (R) is a statistic that measures the linear correlation between the independent and dependent variables (Benesty, Chen, Huang, Cohen, 2009). It is the sole indicator to quantify the strength of the relationship between two variables. The indicator is ranged between -1 and +1. When R is between -0.7 to -1, the two variables are strongly negatively correlated, where one variable increases as the other decreases. When R is between +0.7 to +1, the two variables are strongly positively correlated, where one variable increases as the other increase (Cohen, Cohen, West, and Aiken, 2003).

3.5 Time Series Model

According to Glass (1964), time series forecasting was invented by John Graunt and is defined as a set of data that consists of data points ordered by time and by successive equally spaced sequence. The model is used to understand meaningful statistical significance based on the data's historical observed value and against time. Compared to correlation analysis, which is discussed in Section 3.4, the time series model does not involve other independent factors in the analysis to determine the future forecasting outcomes (Adland, Benth, and Koekebakker, 2018).

Analyzing the history of the data, the model is designed to provide insights on the level, trend, and seasonality of the series. If the data shows a level, there is an initial value set as the average value in the series. If a trend is shown in the ocean freight data set, we can conclude that there is an upward or downward shift consistently represented in the data set. For example, if we find an upward trend in the series, it can conclude that the China-to-U.S. freight rates have been trending upwards for the past 5 years despite volatilities and is expected to go higher on average in the future years. If the data shows a seasonality cycle, it shows a repetitive pattern based on the season or other time cycles. For example, we might find that the ocean freight rate will go up every year in September right before the volume peak and go down every December after the holiday. It means that there is a seasonality pattern and the same cycle will likely happen again next year.

ARIMA and Exponential Smoothing are used to forecast future expected values. ARIMA represents a category of time series models that take into consideration of its historical value and the lags and the lagged forecast errors of previous entries. An ARIMA model includes three main parts: the autoregressive term (p), the difference (d), and the moving average term (q). The autoregressive term represents the regression of its prior observations. The current observation is a function or a percentage of the previous observations. The difference term is how the model treats the dataset. To replace the original dataset's observation with the differencing values between the current observation and the previous observation, differencing methods allow the model to remain stationary and fit the dataset better. The moving average term represents the errors of previous observations that are used to calculate current observations (Nau, 2014).

For Exponential Smoothing models, instead of applying the same weight equally for each historical observation, the models give more weight to the most recent observations and decrease exponentially for more distant observations. Data degrades over time and weighs less than the newer observations as time goes by. Exponential smoothing is a forecasting analysis that blends the value of new and old information. A smoothing factor is assigned between 0 to 1 to demonstrate how much weight is assigned to the most recent observation. If the model has a high smoothing factor, the model gives more weight to the more recent historical rate and has a less smoothing effect. It becomes more nervous, volatile, and naïve. If the model has a low smoothing factor, the model gives more equal consideration to all data point, therefore, the forecast becomes more calm, staid, and cumulative (Gelper, Fried and Croux, 2009). The analysis can be decomposed into the following three components, level, trend, and seasonality, which are explained below:

1. Level: The weighted average value of the dataset. It is the base for every observation of the dataset where the demand hovers (Goodwin, 2010). The formula is shown in Equation 2, where:

   $\hat{a}_t$: Forecasted level at period t

   α: Exponential Smoothing factor for level (0 < α < 1)

   x_t: Most recent observation

$$\hat{a}_t = \alpha x_t + (1-\alpha)(\hat{a}_{t-1}) \qquad 0 \le \alpha \le 1 \tag{2}$$

2. Trend: There is a change in the underlying level of the dataset, either in decreasing or increasing manner (Goodwin, 2010). The formula is shown in Equation 3, where:

   $\hat{b}_t$: Forecasted level at period t

   β: Exponential Smoothing factor for trend (0 < β < 1)

   $\hat{b}_{t-1}$: Forecasted Trend for time period t-1

$$\hat{b}_t = \beta(\hat{a}_t - \hat{a}_{t-1)} + (1-\beta)\hat{b}_{t-1} \qquad 0 \le \beta \le 1 \tag{3}$$

3. Seasonality: Seasonality is the percent of the average demand for a fixed period (Goodwin, 2010). The formula is shown in Equation 4, where:

   $\hat{F}_{t-p}$: Forecasted Seasonality Index for time period t-p

   γ: Exponential Smoothing factor for seasonality (0 < γ < 1)

   p: Number of time periods within the seasonality

$$\hat{F}_t = \gamma(\frac{x_t}{\hat{a}_t}) + (1-\gamma)\hat{F}_{t-p} \qquad 0 \le \gamma \le 1 \tag{4}$$

The common approach to determine the best-fit model is to use numerical optimization to search for the coefficients that resulted in the lowest error. Rob Hyndman (2019) explained in his book Forecasting: Principles and Practice:

"[…] a more robust and objective way to obtain values for the unknown parameters included in any exponential smoothing method is to estimate them from the observed data. […] the unknown parameters and the initial values for any exponential smoothing method can be estimated by minimizing the errors" (Hyndman, 2019).

We compared different statistical indicators to calculate statistics errors to determine the best-fit models mainly using the following two measurements:

1. Root Mean Square Error (RMSE): It is a statistical measurement to evaluate how big the residual errors are. It shows the distance between the actual data points and the best-fit model. In Equation 5, f equals to the forecasts and o equals to the observed value.

$$\text{RMSE} = \sqrt[2]{(f-o)^2} \tag{5}$$

2. Mean Absolute Percentage Error (MAPE): It is a statistical measurement to assess the performance of the model versus the data used for the analysis, as shown in Equation 6, where n is the period, A_t is the actual value and F_t is the forecast value:

$$\text{MAPE} = \frac{1}{n}\sum_{t=1}^{n} \left| \frac{A_t - F_t}{A_t} \right| \tag{6}$$

4 RESULTS AND BUSINESS INSIGHTS

4.1 Results

To find the best-fit model, we apply different forecasting models to the master data file to determine which model produces the minimum error rates, RMSE, and MAPE. After generating the best-fit time series forecasting model based on historical data only, we add the most correlated economic indicator as an exogenous variable to the analysis to determine if the accuracy of the model improves. The report includes the two best-fit models for the route, China to U. S. West Coast (USWC), and China to U.S. East Coast (USEC).

Before running statistical analysis, we plot a visual figure to show the level, trend, and seasonality of the historical ocean freight rate based on the two different routes as shown in Figure 3.

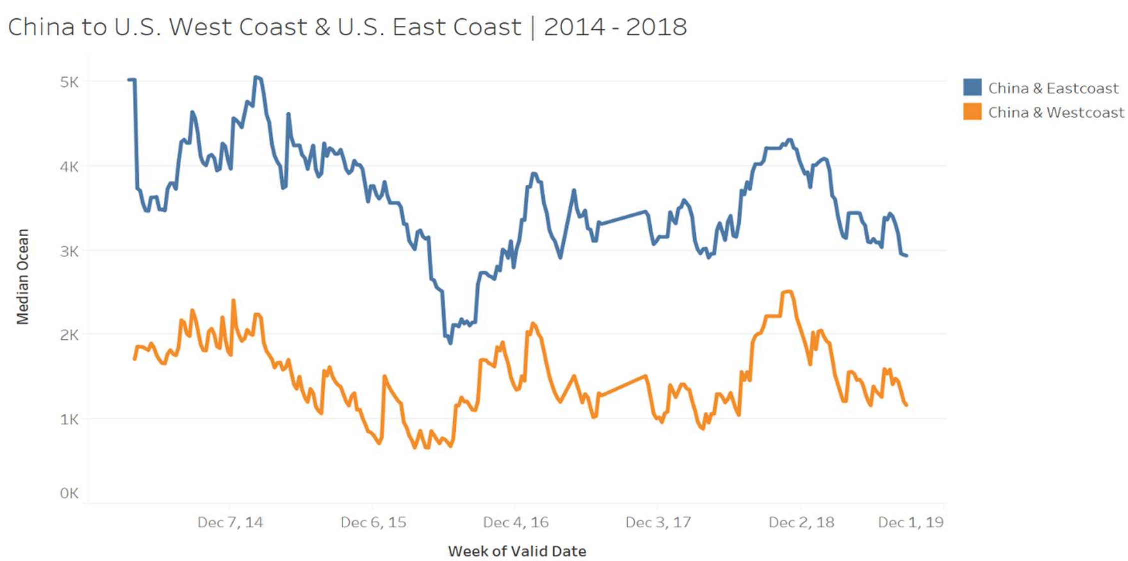

Figure 2~~3~~.The China-to-U.S. West Coast and U.S. East Coast Ocean Freight Rate

SPSS[1] and MS Office Excel[2] are used as the main software to apply statistical analysis in the research, including correlation regression analysis and time series forecasting. After transforming the series' dates into SPSS format, the SPSS software can support either Expert Modeler, which automatically selects the best-fit model or running the analysis manually through each model. Several time series models are tested for the ability to forecast the historical ocean freight rate. The forecast accuracy measurements used are RMSE and MAPE. SPSS is used for the initial assessment, where all exponential smoothing models are tested for their forecast errors. The 6 exponential smoothing models include (1) Multiplicative Seasonality (2) Holt-Winter' Multiplicative (3) Holt's Linear Trend (4) Brown's Linear Trend, (5) Damped Trend (6) ARIMA(0,1,1) . The results are summarized in Table 4.

Table 4. Forecasting Errors for Exponential Smoothing and ARIMA

Forecast Method	RMSE	MAPE
Multiplicative Seasonality	213.721	8.3
Holt-Winter's Multiplicative	291.445	10.0
Damped Trend	289.082	10.8
Holt's Linear Trend	286.472	10.8
Brown's Linear Trend	314.695	11.6
ARIMA(0,1,1)	292.131	10.693

Since SPSS can only provide limited statistical insights, after the initial analysis, we selected two models to run the same analysis manually in excel. We selected Multiplicative Seasonality Exponential Smoothing and Holt's Winter Multiplicative based on the RMSE and MAPE value. The RMSE of the Holt-Winter's Multiplicative model is 0.8% and 1.7% higher than RMSE's Damped Trend and Holt's Linear Trend, respectively. However, since MAPE of Holt-Winter's Multiplicative performs better than Damped Trend and Holt's Linear Trend with an 8% improvement (10.0% vs 10.8%). Therefore, we selected Multiplicative Seasonality and Holt's Winter Multiplicative to conduct further analysis.

1 Source: www.ibm.com/products/spss-statistics
2 Source: www.microsoft.com/en-us/microsoft-365/excel

1. Multiplicative Seasonal Exponential Smoothing (With no trend): Multiplicative Seasonal Exponential Smoothing with no trend is defined as a multiplicative model in that the seasonality for each period is the result of the level and that period's seasonality factor (Glass, 1964). The formula is shown in Equation 7, where:

   $\widehat{X}_{t,t+1}$: Forecasted value for time period t+1

   $\hat{a}_t$: Forecasted level at period t

   α: Exponential Smoothing factor for level (0 < α < 1)

   x_t: Most recent observation

   τ: The forecasting period

   $\widehat{F}_{t-p}$: Forecasted Seasonality Index for time period t-p

   γ: Exponential Smoothing factor for seasonality (0 < γ < 1)

   p: Number of time periods within the seasonality

   $$\hat{x}_{t,t+1} = \hat{a}_t \hat{F}_{t+\tau-p} \tag{7}$$

   $$\hat{a}_t = \alpha\left(\frac{x_t}{\hat{F}_{t-p}}\right) + (1-\alpha)(\hat{a}_{t-1}) \qquad 0 \le \alpha \le 1$$

   $$\hat{F}_t = \gamma\left(\frac{x_t}{\hat{a}_t}\right) + (1-\gamma)\hat{F}_{t-p} \qquad 0 \le \gamma \le 1$$

2. Holt's Winter Multiplicative: The dataset has a linear trend, a level, and a seasonality pattern (Glass, 1964). Holt's Winter Multiplicative Exponential Smoothing (Level, Trend, and Seasonality) is defined as the model that assumes a linear trend with a multiplicative seasonality effect over both level and trend (Glass, 1964)

   The standard formula for Holt's Winter multiplicative is shown in Equation 8, where:

   $\widehat{X}_{t,t+1}$: Forecasted value for time period t+1

   $\hat{a}_t$: Forecasted level at period t

   α: Exponential Smoothing factor for level (0 < α < 1)

x_t: Most recent observation

τ: The forecasting period

$\hat{F}_{t-p}$: Forecasted Seasonality Index for time period t-p

γ: Exponential Smoothing factor for seasonality (0 < γ < 1)

p: Number of time periods within the seasonality

$\hat{b}_t$: Forecasted level at period t

β: Exponential Smoothing factor for trend (0 < β < 1)

$\hat{b}_{t-1}$: Forecasted Trend for time period t-1

$$\hat{x}_{t,t+1} = (\hat{a}_t + \tau\hat{b}_t)\,\hat{F}_{t+\tau-p} \tag{8}$$

$$\hat{a}_t = \alpha\left(\frac{x_t}{\hat{F}_{t-p}}\right) + (1-\alpha)(\hat{a}_{t-1} + \hat{b}_{t-1}) \qquad 0 \le \alpha \le 1$$

$$\hat{b}_t = \beta(\hat{a}_t - \hat{a}_{t-1)} + (1-\beta)\hat{b}_{t-1} \qquad 0 \le \beta \le 1$$

$$\hat{F}_t = \gamma\left(\frac{x_t}{\hat{a}_t}\right) + (1-\gamma)\hat{F}_{t-p} \qquad 0 \le \gamma \le 1$$

To minimize the forecast error of the analysis, the best-fit model is determined as Multiplicative Seasonality with a smoothing coefficient factor for the level, α, and a smoothing seasonality coefficient factor, γ. In this model, 100*α percent of the weight is assigned to the most recent observation, and 100*(1- α) percent of the weight is assigned to the rest of the previous observations. Similarly, 100*γ percent of the weight is assigned to the most recent de-leveled seasonality factor, and 100*(1- γ) percent of the weight is assigned to the de-leveled seasonality factors of prior observations.

To initiate the model, we calculate an initial seasonality index as shown in Figure 4, a measure of how a particular season through some cycle compares with the average season of that cycle. We analyze the percentage of average demand for the month, then compared with the total demand of the year. The result for USWC and USEC seasonality indexes are shown in Table 6. The below seasonality index indicates how the value of the freight rate in a certain month is compared to the average of the full year. The data reveals that the freight rates are below-average during spring for both the USWC and USEC, while above-average during fall and winter.

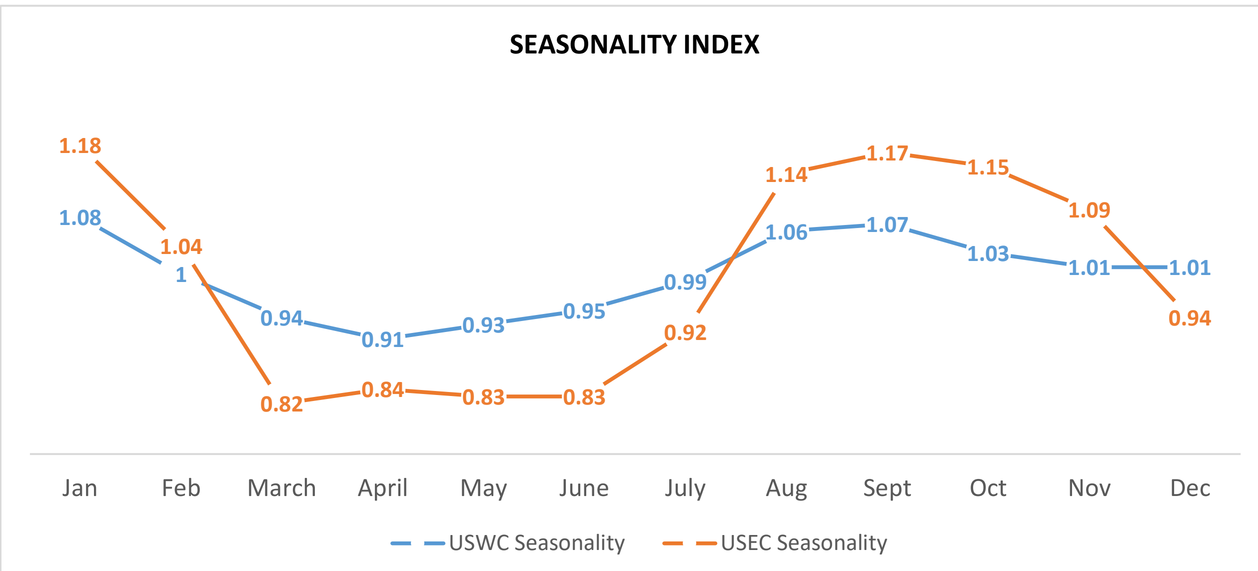

Figure ~~3~~4. Seasonality Index

To initialize the level, this research calculated the latest actual observation and deseasonalized it by dividing it by the seasonal factor of the same month. For example, using the USWC dataset as shown in Appendix 1, the initial data point is August 2014. To calculate the initial level, the model divided the actual observation for the freight rate of August 2014, $2065, by the seasonality factor of the same month, 114% and concluded that the initial starting point for $\hat{a}_t$ is at $1813.8. The first forecast for USWC is calculated as $2127 in September 2014, using the level multiplied by the seasonality index.

Comparing to the actual observation of September 2014, the forecast has an error of $21, a good prediction with minimum error. Using the MAPE formula, finding the absolute value of ($2106-

$2127)/$2106, this resulted in a 1% forecast error. Similarly, RMSE is calculated by applying the square root of ($2106-$2127)^2, therefore the RMSE for September 2014 is $20.59.

Based on the result shown in Table 6, for U.S. West Coast, the best-fit time series model with the lowest RMSE assigned smoothing factor, α at 0.689, and γ at 0, which means 68.9% of the weight is assigned to the latest observation, and the current period's seasonality is simply the previous most recent estimate for that period's seasonality.

For U.S. East Coast, the best-fit time series model with the lowest RMSE assigned smoothing factor, α at 0.727, and γ at 0, which means 72.7% of the weight is assigned to the latest observation, and the current period's seasonality is simply the previous most recent estimate for that period's seasonality.

Table 5~~6~~. Multiplicative Seasonality Statistics

Model Fit (China to U.S. West Coast)

Fit Statistic	Mean	Smoothing Factor	
RMSE	204.441	Alpha	0.689
MAPE	11	Gamma	0

Model Fit (China to U.S. East Coast)

Fit Statistic	Mean	Smoothing Factor	
RMSE	223.001	Alpha	0.727
MAPE	5.6	Gamma	0

Equation 9 and Equation 10 after updating the smoothing factors are shown below:

China to USWC Multiplicative Seasonality Model:

$$\hat{x}_{t,t+1} = [0.689 * \left(\frac{x_t}{\hat{F}_{t-p}}\right) + 0.311(\hat{a}_{t-1})] * \hat{F}_{t+\tau-p} \qquad 0 \le \alpha \le 1 \qquad 0 \le \gamma \le 1 \qquad (9)$$

China to USEC Multiplicative Seasonality Model:

$$\hat{x}_{t,t+1} = [0.727 * \left(\frac{x_t}{\hat{F}_{t-p}}\right) + 0.273(\hat{a}_{t-1})] * \hat{F}_{t+\tau-p} \qquad 0 \le \alpha \le 1 \qquad 0 \le \gamma \le 1 \qquad (10)$$

Additionally, we ran the analysis using Holt's Winter Multiplicative as shown in Table 8. Using the same approach to find level and seasonality, an additional step for Holt-Winter Multiplicative is to determine the trend of the dataset. To initiate the trend, the model requires two observed values. Using USEC rates as an example, the observed value for August 2014 is 4285 and the observed value for

September 2014 is 4508, with the seasonality index of August and September at 114% and 117%, respectively. The initial value of the level is calculated by de-seasonalized September 2014's value using 4508 divided by 117%.

The model fits as shown in Table 7, where USWC shows only a level smoothing factor, alpha, and 0 for both trend and seasonality smoothing factor. It means that Holt's Winter Multiplicative fits USWC the most when the model weighs more to the most recent observation to determine the level and weighs observations equally when looking for level and seasonality. On the other hand, USEC shows a higher trend and seasonality smoothing factor, which means that the best-fit Holt's Winter Multiplicative model for USEC assigns more weights to the most recent observation than the observations in the past. The model fits also is concluded that both RMSE and MAPE for USWC and USEC yield higher RMSE and MAPE rates than Multiplicative Seasonality. The finding shows that by adding a trend component to the forecasting model has a negative effect and make the model less accurate.

Table ~~7~~6. Holt's Winter Statistics

Model Fit (China to U.S. West Coast)

Fit Statistic	Mean	Smoothing Factor	
RMSE	213.27	Alpha	0.765
MAPE	10.40	Betta	0.000
		Gamma	0.000

Model Fit (China to U.S. East Coast)

Fit Statistic	Mean	Smoothing Factor	
RMSE	369.62	Alpha	0.515
MAPE	9.50	Betta	0.110
		Gamma	0.643

The updated Equation 11 and Equation 12 after updating the smoothing factors are shown below:

China to USWC Holt's Winter Model:

$$\hat{x}_{t,t+1} = [0.765(\frac{x_t}{\hat{F}_{t-p}}) + 0.235(\hat{a}_{t-1} + \hat{b}_{t-1}) + \hat{b}_{t-1}]\, \hat{F}_{t+1-p} \quad (11)$$

$0 \leq \alpha \leq 1 \qquad 0 \leq \beta \leq 1 \qquad 0 \leq \gamma \leq 1$

The China to USEC Holt's Winter Model:

$$\hat{x}_{t,t+1} = (\hat{a}_t + \tau\hat{\mathrm{b}}_{\mathrm{t}})\ \hat{F}_{t+\tau-p} \tag{12}$$

$$\hat{a}_t = 0.515(\frac{x_t}{\hat{F}_{t-p}}) + 0.485(\hat{a}_{t-1} + \hat{b}_{t-1}) \qquad 0 \le \alpha \le 1$$

$$\hat{\mathrm{b}}_{\mathrm{t}} = 0.11(\hat{\mathrm{a}}_{\mathrm{t}} - \hat{\mathrm{a}}_{\mathrm{t-1)}} + 0.89\hat{\mathrm{b}}_{\mathrm{t-1}} \qquad 0 \le \beta \le 1$$

$$\hat{\mathrm{F}}_{\mathrm{t}} = 0.643\left(\frac{\mathrm{x_t}}{\hat{\mathrm{a}}_{\mathrm{t}}}\right) + 0.357\hat{\mathrm{F}}_{\mathrm{t-p}} \qquad 0 \le \gamma \le 1$$

Since multiplicative seasonality shows better performance, the model does not predict an upward or downward-facing trend but a cycling pattern and level. For USWC, the forecast projected to remain at \$1,500 level, with an average of \$1,550. Based on the forecast, the lowest point occurs in February 2020, priced at \$1,271 and the highest point occurs in December 2020 at \$1,832 as shown in Figure 5. Historically, based on the average of 5 years data from 2014 – 2019, the lowest month is in March and the highest month is in October. However, if we only look at 2018 and 2019 data, the lowest month aligned with 2020's forecast for both 2018 and 2019 and occurred in February at \$1006 and \$1315, respectively. The highest month for 2018 is also aligned with the forecast and occurred in December at 2415. It proves that the research gives more weight to more recent data and the forecast is aligned with the more recent observations. Similarly, for China to USEC lane, the forecast projected that the rate would remain at \$2,500 to \$3,500 level with an average of \$3,212. Based on the forecast, the lowest point occur in March 2020 priced at \$2,927, and the highest point occurs in December 2020 at \$3,453 as shown in Figure 6. This also aligned with USWC forecast and also the historical data in 2018 and 2019 where the peak and through appeared in the same period.

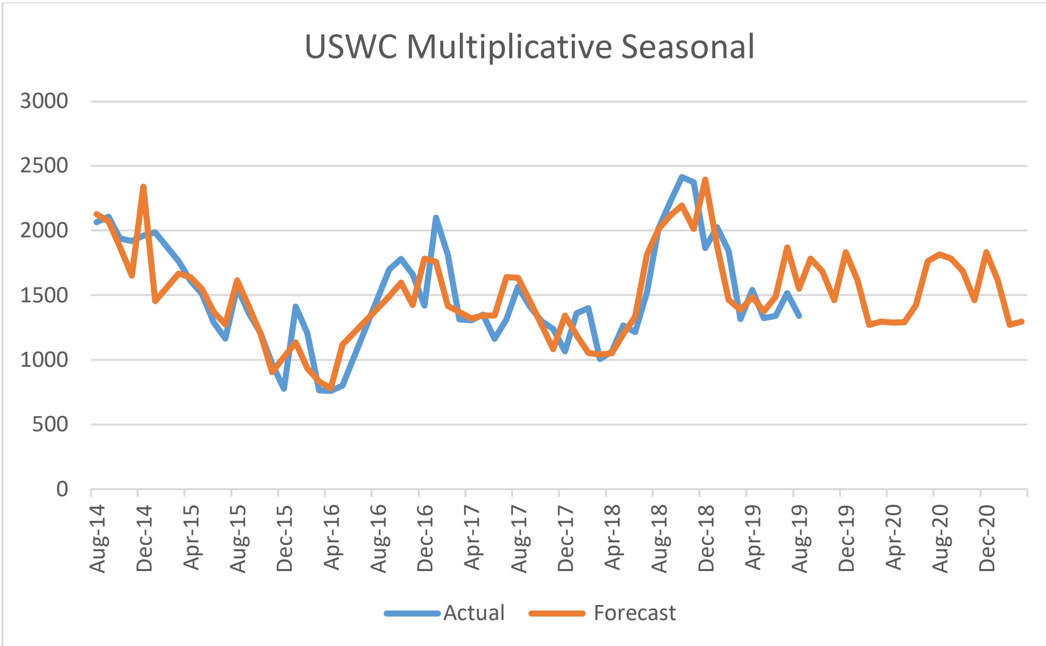

Figure 4~~5~~. U.S. West Coast Forecast

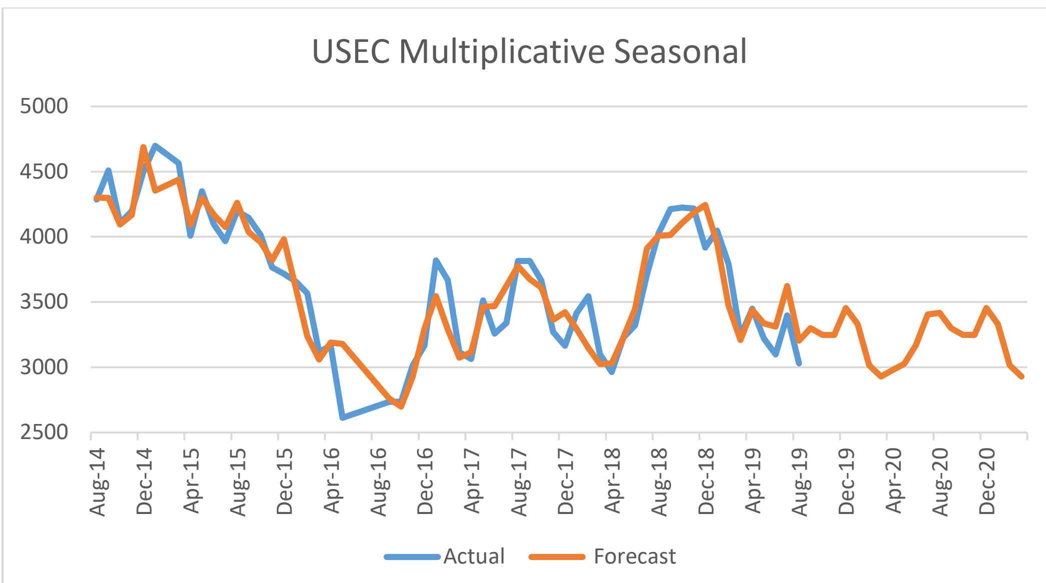

Figure 5~~6~~. U.S. West Coast Forecast

After running the time series models, we also analyzed the correlations between the ocean freight rate and economic indicators. According to the Pearson correlation analysis as defined in Section 3.4, a correlation between 0.1 and 0.3 is identified as weakly correlated, a correlation between 0.3 and 0.5 is identified as moderately correlated, and a correlation between 0.5 and 1 considered as strongly correlated. The result of Pearson correlation analysis is shown in Table 8.

Table 7~~8~~. U.S. Pearson Correlation Analysis

Correlation Coefficient	China to USWC	China to USEC
Oil_Price	0.491**	0.305*
CNY_USD	0.122	-0.441**
China_CPI	0.021	-0.313**
US_CPI	0.098	-0.345**
CN_GDP	0.119	-0.306**
US_GDP	0.072	-0.344**
CN_PMI	-0.033	-0.261**
US_PMI	0.350**	0.113

The majority of the parameters that are correlated are between 0.1 to 0.3, which is considered as weakly correlated. Among these parameters, oil price and US PMI to USWC and exchange rate and oil price to USEC are the two most correlated indicators above 0.3, at 0.491, 0.350, -0.441, 0.305, respectively. These results are aligned with the industry expectations where the price goes up when the operating cost goes up due to oil price increase. As mentioned in 3.1, we will run ARIMA with exogenous factors analysis between the historical ocean freight rates and the most correlated factors. Therefore, we selected US PMI, exchange rate, and oil price as the exogenous factors for the analysis. After implementing the two exogenous variables into the models, by running the regression model between ARIMA time series analysis and the exogenous factor, we find that the model fit became worse, with a significant increase in RMSE and MAPE than previous with the more detailed result shown in Appendix 1.

We further analyzed the leading and lagging factors, where the dependent variable runs periods faster or slower than the independent variable, by adjusting the lag period by one. Since the dataset is arranged by monthly average and median, it means that the lead indicator runs one month ahead of the historical ocean freight rate. For the West Coast, using the same ARIMA model, we applied lag = 1, where the ocean freight rates are one period behind the economic indicator, oil price, and US PMI. For the east coast, using the same ARIMA model, we applied lag =1, where the ocean freight rates are one period behind economic indicator exchange rate and oil price. We applied ARIMA with exogenous indicator analysis using SPSS software, and the results are shown in Appendix 2.

The statistical indicators summary of all analyzed models is shown in Table 9. Determined by statistical indicators, RMSE, and MAPE, the research find the best-fit model to be Multiplicative Seasonality exponential smoothing.

Table 8~~9~~. SPSS Forecasting Errors Summary

Forecast Method	RMSE	MAPE
Multiplicative Seasonality	213.721	8.3
Holt's Winter Multiplicative	291.445	9.95
Damped Trend	289.082	10.75
Holt's Linear Trend	286.472	10.76
Brown's Linear Trend	314.695	11.637
ARIMA (0,1,1)	292.131	10.693
ARIMA and Oil Price, U.S. West Coast	287.705	14.92
ARIMA and PMI, U.S. West Coast	287.707	14.969
ARIMA and Exchange Rate, U.S. East Coast	289.466	6.594
ARIMA and Oil Price, U.S. East Coast	288.02	6.493
ARIMA and Oil Price, U.S. West Coast, Lag =1	360.849	20.36
ARIMA and PMI, U.S. West Coast, Lag =1	376.967	19.326
ARIMA and Exchange Rate, U.S. East Coast, Lag =1	381.57	8.284
ARIMA and Oil Price, U.S. East Coast, Lag =1	373.649	8.188

After building a time series forecast based on the multiplicative seasonality model, we incorporate two economic indicators as the exogenous variables to the model. The exogenous variables did not improve the accuracy of the model. We further analyze the lead and lag factor, and adjust the lag by one period of time which means that the historical ocean freight rate runs one period of time later than the leading indicator suggested. However, the statistics measurements did not improve. The project concluded that Multiplicative Seasonality with no trend is the best forecasting model to predict future ocean freight rates for the China-to-U.S. route.

4.2 Smoothing Factor Sensitivity Analysis

We study the impact of level smoothing factor, α, weighing to the overall accuracy of the model. For USWC, the RMSE plotted graph shows a convex shape with the minimum at 0.689 and with the lowest RMSE at 204. The analysis shows that the forecast error is more sensitive to change when the α value is low, between 0 to 0.68, which means less weight should be assigned to past observations to avoid a higher forecast error. With a 20% increase of α value from 0.4 to 0.6, RMSE decreased over 20. On the other hand, when the alpha value is higher than 0.68, between 0.68 to 1, the RMSE is less sensitive to changes. With the same 20% increase, α value increases from 0.6 to 0.8, and RMSE increased for less than 5. We discover a similar trend for USEC, the plotted graph shows a convex shape with the minimum at 0.727. RMSE decreased for 30 by changing α from 0.4 to 0.6, and only increased for 10 by changing α from 0.727 to 1.

We conclude that the model is more sensitive to change when α is low, therefore when the model assigns less value to the most recent observation, the model makes a poor prediction and performs with higher error. When the model assigns higher α value, the most recent observation is given more weight than the best-fit solution, the model's error will increase but less significantly as shown in Figure 7.

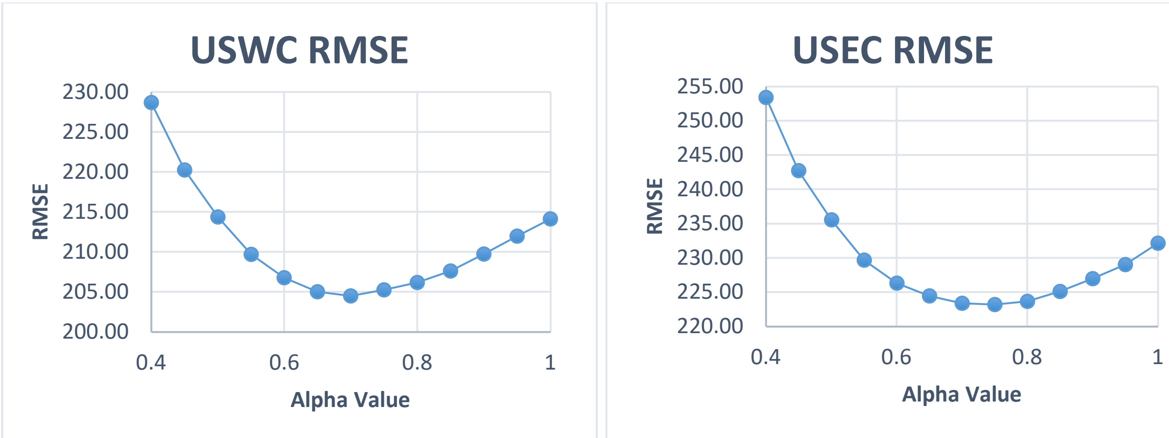

Figure 6~~7~~. USWC & USEC Smoothing Factor Sensitivity Analysis

In the future, the sponsor company can use this insight to determine how to approach freight rate prediction based on the historical data, by paying more attention to the recent observations, rather than the past observations. Prediction performance can be poor if less weight is assigned to the more recent data. However, the forecast has a higher tolerance if the analysis overly assigned more weight to the most recent observation. In another word, it shows that past observation does not value as much as the most recent ones. If the company wants to make a quick judgment on predicting the future rates, it is advised to use the most recent observation rather than allocating the same weight to all data points. For example, if the company wants to predict June 2020's freight rate, the analysis should rely more heavily on May 2020's freight rate rather than the one of May 2014.

4.3 Business Insight

The research shows that the China-to-U.S. historical ocean freight rates are best predicted by a Multiplicative Seasonality (with no trend) exponential smoothing model with an α at 0.689 for USWC and 0.727 for USEC. It means that 68.9% and 72.7% of the weight, respectively, are given to the most recent observation. Comparing to Multiplicative Seasonality (with no trend), Holt's Winter Multiplicative Exponential Smoothing Model includes one more component which is a trend. However, the result of Holt's Winter Multiplicative forecasting is less accurate than the multiplicative seasonality with 20% more MAPE and 36% more RMSE. The difference between the two models is that Multiplicative Seasonality does not consider a trend in the data set, while Holt's Winter Multiplicative does. Since Multiplicative Seasonality has lower error rates, RMSE and MAPE, therefore we conclude that the dataset does not show a linear trend based on its historical data.

Based on the sensitivity analysis in Section 4.2, there is a higher chance that yielding more weight to the most current events will provide a more accurate forecast than giving more weight to the more distant historical events. When the smoothing factor is lower than the optimal solution, it accelerates the increase of the error rates. Therefore, if there is not enough weight assigned to the recent observations, there is a higher risk of receiving an inaccurate forecast,

Moreover, we discover that economic indicators are weakly correlated with ocean freight rates, that is the movement of the ocean freight rates is weakly related to the movement of the exogenous factors. U.S. East Coast freight rates are correlated with more economic indicators than U.S. West Coast freight rates. For U.S. East Coast freight rates, the historical ocean freight rates are positively correlated the oil price, that is, when oil price increases (or decreases), ocean freight increases (decreases), as well. And the ocean freight rates are negatively correlatedwith the appreciation of USD versus CNY (i.e. when one increases, the other deacreases, and vice versa). It is also negatively correlated with China CPI, US CPI, CN GDP, and US GDP. For U.S. West Coast ocean freight rates, the historical ocean freight rates are positively correlated with the oil price and US PMI and are not negatively correlated with any other economic indicators examined in Table 8. This finding aligns with the expectation that when fuel cost increased, the cost to operate increased, and therefore the rate increased. Even though there are correlations between economic indicators and the ocean freight rates, adding additional variables to the time series model resulted in a less accurate model. We also applied lag functions, i.e. setting ocean freight rates a period behind the correlated economic indicators, to the analysis but did not yield a better result. This analysis shows that economic indicators and oil prices are not leading factors to predict ocean freight rates in the future.

The model forecasted the next peak and through for the upcoming years. For U.S. West Coast and U.S. East Coast freight rates, the data indicated that December has the highest rate of the year. This finding matches with what business observes in reality. Since the majority of the U.S. sellers are looking to stock up inventory right before the Christmas shopping season. The demand in the market goes up while capacity stays the same. U.S. West Coast and U.S East Coast show different months for the low season, but the two months are close to each other in February and March, respectively. These observations align with industry practices, where the low season is driven by the decline after Christmas sale, and manufactory factories closure for Chinese New Year.

Since the model is designed to explain regular cycle patterns, significant increases, or drops associated with individual industry event may cause a negative impact to the model for the following months after the event occurred. For example, between 2014 to 2018, the ocean freight industry has experienced two significant decreases, one in March 2016 and one in February 2017. After March 2016's drop, the actual observation recovered gradually from March to August. Even though the forecast adapted to the drop, the model presents a much higher error rate, RMSE, for the next few periods with a delay from May 2016 to November 2016 as shown in Figure 8.

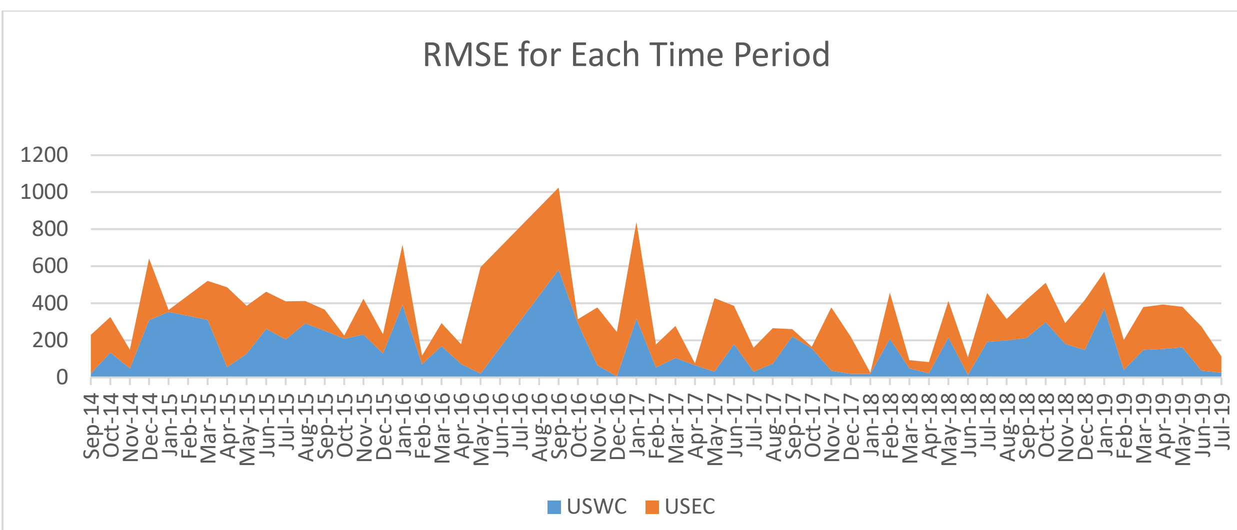

Figure 7~~8~~. RMSE for Each Time Period

4.4 Limitations

We analyzed the historical ocean freight rates to determine future patterns. It also researched the impact of economic indicators on the future China-to-U.S. ocean freight rates. However, as indicated in the literature review, freight rates are dynamic and are also impacted by other internal factors, such as the number of vessels available at a certain time, size of the vessels and port volume. It is difficult to obtain this information from both public websites and private databases.

We only examined a limited number of economic indicators based on the literature review and executive interviews. Other economic indicators may have a stronger correlation with ocean freight rates. We only had a limited number of data points that are available to retrieve. Since U.S. and China GDP can only be obtained as monthly data, the analysis can only apply monthly data points for all other data sets, including the historical ocean freight rates. The analysis may provide different findings if the model applied a different period, for example, weekly ocean freight rates or weekly economic indicator figures.

As a time, series forecasting model, forecasts presented for the near future, especially for the next few months are more accurate than forecast projected for the far future, in the next two or three years. It is suggested to re-run the analysis after new data points are recorded, to receive the most accurate projection.

5 CONCLUSION

The project scope of the analysis is to build a predictive forecasting model to quantify the future ocean freight rates of China to the U.S. lane. By applying time series models with and without exogenous factors, we determined the best-fit model based on statistical measurements, RMSE, and MAPE. Analyzing the history of the data, the model is designed to provide insights on the trend, seasonal cycles, pulses, steps, and outliers of the series. ARIMA, Exponential Smoothing, and time series with exogenous factor models are used for forecasting. ARIMA represents a category of time series models that take into consideration its historical value and the lags and the lagged forecast errors of previous entries. For Exponential Smoothing models, the analysis can identify the level, trend, and seasonality in the structure. The time series model with an exogenous factor runs regression analysis between the ARIMA model and an economic indicator to imbed an exogenous factor into the time series analysis.

The best-fit model, based on the goodness of fit, has been determined as a multiplicative seasonality time series model RMSE at 213.7, and MAPE at 8.3. The model demonstrates that the data does not show a linear increase or decrease for the future years. Moreover, the data shows a cycling pattern, which means that the data will repeat the same pattern after periods. The best-fit multiplicative seasonality model is sensitive to the assigned smoothing value, especially when we give less weight to the most recent observation. The error rates increased more rapidly, comparing to assigning too much weight to the most recent observation. The research is also sensitive to significant events occurred in the period, where error rates peaked during the next few months after the event.

After running the time series model, the project also analyzed the correlations between the ocean freight rate and economic indicators. Among these parameters, oil price and US PMI is the most correlated indicator for U.S. West Coast and exchange rate and oil price is the most correlated indicator for U.S. East Coast After implementing the two exogenous variables into the models, we found that the statistical fit became worse than the previous model.

The research further analyzed the lead and lag factor by adjusting the lag period by one. Since the dataset is arranged by the monthly average and median, the lead indicator will run one month ahead of historical ocean freight rates. For the U.S. West coast, we applied lag = 1 to the economic indicator oil price. For U.S. East Coast, we applied lag = 1 to the economic indicator exchange rate. We also applied ARIMA with an exogenous indicator using SPSS software, however, the statistical measurements did not improve.

Based on the findings, the research has shown that the historical ocean freight rate contains seasonality. The peaks indicated that December will experience the highest rates of the year and identified February as the low season. This finding matches with industry experiences and is associated with the holidays' inventory planning.

There are some limitations to the research. The research only analyzed the effect of oil price and economic indicators’ effect on ocean freight rates but did not include other endogenous factors, such as vessel sizes and vessel supplies. There is a limited number of economic indicators applied with limited data points. As a time series model, the model can best predict the near future pricing but has limitations towards forecasting long term future pricing.

For future researches, the analysis can incorporate regression analysis involving supply and demand indicators or other internal indicators such as vessel size to discover correlations with the historical rates. Moreover, future researches can use different time intervals, for example, weekly average or daily average to conduct the same time series analysis with indicators that incorporate weekly data points. Other time series model may be a better fit for the historical data with different time interval segments. Other lanes of ocean freight rates can also be researched upon to verify the same analysis. Whether seasonality is shown in all ocean freight lanes, or it is specifically characterized for the China-to-U.S. lane. The research can also leverage this classical forecasting approach to create more sophisticated analysis, for example, neural networks, a machine learning method to predict the pricing for China to U.S. ocean freight rates.

REFERENCES

Alphaliner Weekly Newsletter (2019). Alphaliner, 2019 (32). Retrieved from www.alphaliner.com

Alphaliner Weekly Newsletter (2019). Alphaliner, 2019 (38). Retrieved from www.alphaliner.com

Benesty, J., Chen, J., Huang, Y., & Cohen, I. (2009). Pearson Correlation Coefficient.

Binkley J.K. and Harrer, B. (1981). Major Determinants of Ocean Freight Rates for Grains: An Econometric Analysis. American Journal of Agricultural Economics, 63(1), 47–57

Bureau of Economic Analysis (2019). Gross Domestic Product (2009-2018). Retrieved from https://www.bea.gov/data/gdp/gross-domestic-product

Bureau of Labor Statistics (2019). Consumer Price Index (2009-2018). Retrieved from https://download.bls.gov/pub/time.series/ap/ap.data.0.Current

C.H. Robinson (2019). C.H. Robinson Reports 2018 forth quarter and full-year results. Retrieved from https://investor.chrobinson.com/News-and-Events/Press-Releases/press-release-details/2019/CH-Robinson-Reports-2018-Fourth-Quarter-and-Full-Year-Results/default.aspx

C.H. Robinson Annual Report 2018. (n.d.).

CAI Transportation Log (2017) How will IMO 2020 affect ship capacity and freight rates? Retrieved from http://blog.capps.com/how-will-imo-2020-affect-ship-capacity-and-freight-rates

Chi, J. (2016). Exchange rate and transport cost sensitivities of bilateral freight flows between the US and China. Transportation Research Part A, 89, 1-13

Cohen, J., Cohen, P., West, S. G., & Aiken, L. S. (2015). *Applied multiple regression/correlation analysis for the behavioral sciences*. New York: Routledge, Taylor & Francis Group.

Ding, J.F et.al. (2017). Evaluating key factors influencing the development of multi-country consolidation for ocean freight forwarders in Taiwan. Journal of Engineering for the Maritime Environment, 231(1), 342–352

Elarbi, A.E. (2013). Statistical Model to Validate Sales Force Driven Forecast. MIT Scale Research Report. MISI-2013-6

Flexport Market Update: Ocean and Air Freight Market Update: February 2017. (n.d.). Retrieved May 09, 2020, from https://www.flexport.com/market-updates/freight-market-update-february-2017/

Goodwin, P. (2010). The Holt's Winters Approach to Exponential Smoothing: 50 Years Old and Going Strong.

D. V. Glass, D. V. (1964). John Graunt and his Natural and political observations.

Hyndman, R. J. (2013). Forecasting: Principles and practice. Retrieved 2020.

International Monetary Fund Currency Exchange Database (2019). Exchange rate (2009-2019). Retrieved from https://www.imf.org/external/np/fin/ert/GUI/Pages/CountryDataBase.aspx

Klovland, J.T. (2008). New evidence on the fluctuations in ocean freight rates in the 1850s. Explorations in Economic History, 46, 266–284

Laxmana, S. (2017). Global Shipping Trends: What to Expect in 2018. Retrieved from https://www.chrobinson.com/blog/global-forwarding/global-shipping-trends-what-to-expect-in-2018/

Link. J (2017). The Panama Canal Expansion's Massive Ripple Effect on US Ports and Shipping. Retrieved from https://www.autodesk.com/redshift/ports-and-shipping/

Makridakis, S. (1986). The art and science of forecasting: an assessment and future directions. International Journal of Forecasting, 2(1), 15-39.

Marx, D. (1946). The Determination of Postwar Ocean Freight Rates. The American Economic Review, 36(2), 561-574.

Nam, I.-S. (2017, February 17). Hanjin Shipping Is Declared Bankrupt. Retrieved from http://www.wsj.com/articles/hanjin-shipping-is-declared-bankrupt-1487296151

NASDAQ database (2019). Brent crude (2009-2019). Retrieved from https://www.nasdaq.com/market-activity/commodities/bz%3Anmx

Nau, R. (2014). *The mathematical structure of Arima models*.

Pinkser, J (2015). How 14,000 Workers Managed to Slow Down the Entire Economy https://www.theatlantic.com/business/archive/2015/02/how-only-14000-workers-briefly-slowed-down-the-entire-economy/385858/

Poskus, N (2018). IMO 2020: What Shippers Need to Know Now https://www.flexport.com/blog/imo-2020-what-shippers-need-to-know-now

Silver, E.A., Pyke, D.F. and Peterson, R. (1998). Inventory Management and Production Planning and Scheduling. 3rd Edition, John Wiley, New York.

Yunianto, I.T. Lazuardi, S.D., and Hadi. F. (2018). Freight Calculation Model: A Case Study of Coal Distribution. Earth and Environmental Science, 135, 012013.

6 Appendix

Appendix 1: Multiplicative Seasonality Exponential Smoothing

Table 9~~10~~. USWC Multiplicative Seasonality Exponential Smoothing

Date	USWC_freight	x^ (t, t+1)	a^	F^(t)	f^(t-p)	f^(t+T-P)	MAPE	RMSE
Aug-14	2065	2127	1813.8	114%		117%		
Sep-14	2106	2073	1801.7	117%	117%	115%	1.0%	20.59
Oct-14	1939	1872	1721.6	115%	115%	109%	6.9%	133.82
Nov-14	1920	1653	1752.2	109%	109%	94%	2.5%	48.27
Dec-14	1962	2340	1977.8	94%	94%	118%	15.7%	308.99
Jan-15	1987	1453	1772.3	118%	118%	82%	17.8%	352.95
Mar-15	1763	1667	2031.6	82%	82%	82%	17.6%	309.73
Apr-15	1613	1640	1960.9	84%	84%	84%	3.3%	53.95
May-15	1512	1549	1861.6	83%	83%	83%	8.4%	127.71
Jun-15	1288	1370	1644.4	83%	83%	83%	20.3%	261.28
Jul-15	1165	1272	1385.7	92%	92%	92%	17.6%	204.64
Aug-15	1563	1614	1376.9	114%	114%	117%	18.6%	290.77
Sep-15	1363	1414	1229.2	117%	117%	115%	18.4%	251.29
Oct-15	1206	1201	1104.5	115%	115%	109%	17.3%	208.15
Nov-15	970	904	958.2	109%	109%	94%	23.8%	230.87
Dec-15	775	1022	864.0	94%	94%	118%	16.6%	129.00
Jan-16	1413	1135	1091.6	118%	118%	104%	27.7%	390.79
Feb-16	1206	934	1138.5	104%	104%	82%	5.9%	70.73
Mar-16	765	833	996.5	82%	82%	84%	22.1%	169.12
Apr-16	761	780	936.9	84%	84%	83%	9.5%	72.25
May-16	800	1118	953.7	83%	83%	117%	2.5%	20.27
Sep-16	1700	1491	1295.6	117%	117%	115%	34.2%	581.84
Oct-16	1781	1598	1469.6	115%	115%	109%	16.3%	290.42
Nov-16	1663	1425	1510.9	109%	109%	94%	3.9%	65.27
Dec-16	1420	1783	1507.0	94%	94%	118%	0.4%	5.42
Jan-17	2100	1759	1691.7	118%	118%	104%	15.1%	317.12
Feb-17	1813	1417	1727.2	104%	104%	82%	3.0%	53.68
Mar-17	1313	1371	1639.7	82%	82%	84%	7.9%	104.19
Apr-17	1306	1320	1586.0	84%	84%	83%	5.0%	65.17
May-17	1350	1342	1610.9	83%	83%	83%	2.2%	30.07
Jun-17	1163	1343	1463.1	83%	83%	92%	15.4%	178.72
Jul-17	1312	1639	1439.6	92%	92%	114%	2.4%	31.27
Aug-17	1565	1635	1394.8	114%	114%	117%	4.7%	73.98
Sep-17	1415	1456	1265.3	117%	117%	115%	15.6%	220.37
Oct-17	1300	1274	1172.1	115%	115%	109%	12.0%	155.73

Nov-17	1238	1084	1149.1	109%	109%	94%	2.9%	36.30
Dec-17	1065	1343	1135.2	94%	94%	118%	1.8%	19.04
Jan-18	1361	1191	1145.6	118%	118%	104%	1.3%	18.00
Feb-18	1400	1053	1283.8	104%	104%	82%	14.9%	208.53
Mar-18	1006	1040	1244.0	82%	82%	84%	4.7%	47.36
Apr-18	1062	1050	1261.9	84%	84%	83%	2.0%	21.72
May-18	1265	1199	1439.8	83%	83%	83%	17.0%	214.80
Jun-18	1213	1332	1451.2	83%	83%	92%	1.1%	13.84
Jul-18	1523	1815	1594.3	92%	92%	114%	12.5%	190.63
Aug-18	2015	2011	1715.3	114%	114%	117%	9.9%	199.95
Sep-18	2223	2117	1839.8	117%	117%	115%	9.5%	211.94
Oct-18	2415	2195	2018.5	115%	115%	109%	12.4%	298.34
Nov-18	2375	2012	2132.9	109%	109%	94%	7.6%	180.45
Dec-18	1865	2396	2025.4	94%	94%	118%	7.9%	147.16
Jan-19	2025	1882	1809.2	118%	118%	104%	18.3%	371.21
Feb-19	1843	1463	1783.6	104%	104%	82%	2.1%	38.57
Mar-19	1315	1387	1659.0	82%	82%	84%	11.3%	148.49
Apr-19	1541	1486	1785.6	84%	84%	83%	10.0%	153.75
May-19	1324	1375	1651.5	83%	83%	83%	12.2%	162.05
Jun-19	1340	1489	1622.1	83%	83%	92%	2.6%	35.50
Jul-19	1515	1869	1641.4	92%	92%	114%	1.7%	25.72
Aug-19	1340	1549	1321.4	114%	114%	117%	39.5%	528.72

Table 10~~11~~. USEC Multiplicative Seasonality Exponential Smoothing

Date	USEC_freight	x^ (t, t+1)	a^	F^(t)	f^(t-p)	f^(t+T-P)	MAPE	RMSE
Aug-14	4285	4299	4031.0	106%		107%		
Sep-14	4508	4299	4173.1	107%	107%	103%	4.6%	208.61
Oct-14	4106	4094	4037.2	103%	103%	101%	4.7%	192.59
Nov-14	4196	4168	4110.2	101%	101%	101%	2.4%	101.74
Dec-14	4500	4690	4348.0	101%	101%	108%	7.4%	331.78
Jan-15	4700	4355	4354.7	108%	108%	100%	0.2%	9.91
Mar-15	4566	4439	4715.2	94%	94%	94%	4.6%	210.81
Apr-15	4007	4090	4473.0	91%	91%	91%	10.8%	431.57
May-15	4349	4297	4622.6	93%	93%	93%	6.0%	258.83
Jun-15	4096	4171	4411.5	95%	95%	95%	4.9%	200.79
Jul-15	3965	4074	4117.0	99%	99%	99%	5.2%	205.92
Aug-15	4196	4260	3993.6	106%	106%	107%	2.9%	121.52
Sep-15	4146	4034	3916.2	107%	107%	103%	2.7%	113.51
Oct-15	4016	3959	3903.5	103%	103%	101%	0.4%	17.92
Nov-15	3766	3819	3765.4	101%	101%	101%	5.1%	192.70

Dec-15	3715	3982	3691.1	101%	101%	108%	2.8%	103.61
Jan-16	3656	3611	3471.7	108%	108%	104%	8.9%	325.55
Feb-16	3565	3238	3439.9	104%	104%	94%	1.3%	45.59
Mar-16	3115	3059	3344.8	94%	94%	91%	4.0%	123.06
Apr-16	3165	3188	3429.4	91%	91%	93%	3.4%	106.44
May-16	2613	3178	2979.9	93%	93%	107%	22.0%	574.75
Sep-16	2735	2758	2677.7	107%	107%	103%	16.2%	443.35
Oct-16	2735	2699	2661.3	103%	103%	101%	0.8%	23.21
Nov-16	3011	2926	2885.0	101%	101%	101%	10.4%	312.06
Dec-16	3165	3297	3056.5	101%	101%	108%	7.6%	239.19
Jan-17	3818	3544	3407.7	108%	108%	104%	13.6%	521.00
Feb-17	3667	3289	3493.7	104%	104%	94%	3.4%	123.04
Mar-17	3115	3072	3359.5	94%	94%	91%	5.6%	173.72
Apr-17	3062	3115	3351.6	91%	91%	93%	0.3%	9.99
May-17	3513	3463	3662.6	93%	93%	95%	11.3%	397.64
Jun-17	3256	3467	3503.5	95%	95%	99%	6.4%	206.80
Jul-17	3338	3623	3408.5	99%	99%	106%	3.9%	129.40
Aug-17	3815	3775	3539.6	106%	106%	107%	5.0%	191.70
Sep-17	3815	3674	3566.6	107%	107%	103%	1.0%	39.69
Oct-17	3665	3611	3560.4	103%	103%	101%	0.2%	8.86
Nov-17	3270	3363	3316.2	101%	101%	101%	10.4%	340.70
Dec-17	3163	3422	3172.8	101%	101%	108%	6.3%	200.01
Jan-18	3415	3294	3167.8	108%	108%	104%	0.2%	7.41
Feb-18	3543	3145	3341.5	104%	104%	94%	7.0%	248.51
Mar-18	3100	3023	3306.4	94%	94%	91%	1.5%	45.48
Apr-18	2963	3029	3258.3	91%	91%	93%	2.0%	60.41
May-18	3225	3226	3411.9	93%	93%	95%	6.1%	196.28
Jun-18	3320	3448	3484.3	95%	95%	99%	2.8%	94.23
Jul-18	3712	3910	3678.0	99%	99%	106%	7.1%	263.63
Aug-18	4025	4007	3756.8	106%	106%	107%	2.9%	115.23
Sep-18	4212	4014	3896.5	107%	107%	103%	4.9%	205.04
Oct-18	4225	4103	4045.7	103%	103%	101%	5.0%	211.33
Nov-18	4216	4185	4126.8	101%	101%	101%	2.7%	113.15
Dec-18	3916	4243	3933.9	101%	101%	108%	6.9%	269.11
Jan-19	4046	3953	3800.8	108%	108%	104%	4.9%	197.40
Feb-19	3790	3471	3687.0	104%	104%	94%	4.3%	162.87
Mar-19	3240	3209	3508.8	94%	94%	91%	7.1%	230.70
Apr-19	3448	3439	3699.2	91%	91%	93%	6.9%	239.47
May-19	3219	3335	3527.5	93%	93%	95%	6.8%	219.51
Jun-19	3097	3310	3344.4	95%	95%	99%	7.7%	238.13
Jul-19	3397	3623	3408.4	99%	99%	106%	2.6%	87.09

Aug-19	3029	3202	3002.0	106%	106%	107%	19.6%	594.20

Appendix 2: ARIMA model with Exogenous Factors

Table 11~~12~~. ARIMA Model with Exogenous Indicator, Oil Price, for USWC

Model Statistics

Model	Number of Predictors	Model Fit statistics				Ljung-Box Q(18)	
		Stationary R-squared	R-squared	RMSE	MAPE	Statistics	DF
USWC_freight-Model_1	1	.002	.493	287.705	14.920	16.113	18

ARIMA Model Parameters

					Estimate	SE	t
USWC_freight-Model_1	USWC_freight	No Transformation	Constant		-73.017	180.847	-.404
			Difference		1		
	Oil_Price	No 1Transformation	Numerator	Lag 0	1.005	2.958	.340

Table 12~~13~~. ARIMA Model with Exogenous Indicator, PMI, for USWC

Model Statistics

Model	Number of Predictors	Model Fit statistics				Ljung-Box Q(18)	
		Stationary R-squared	R-squared	RMSE	MAPE	Statistics	DF
USWC_freight-Model_1	1	.002	.493	287.707	14.969	17.095	18

ARIMA Model Parameters

					Estimate	SE	t
USWC_freight-Model_1	USWC_freight	No Transformation	Constant		-207.601	575.686	-.361
			Difference		1		
	US_PMI	No Transformation	Numerator	Lag 0	3.548	10.469	.339

Table 13~~14~~. ARIMA Model with Exogenous Indicator, Exchange Rate, for USEC

Model Statistics

		Model Fit statistics				Ljung-Box Q(18)		
Model	Number of Predictors	Stationary R-squared	R-squared	RMSE	MAPE	Statistics	DF	Sig.
USEC_freight-Model_1	1	.005	.686	289.466	6.594	7.669	18	.983

ARIMA Model Parameters

					Estimate	SE	t
USEC_freight-Model_1	USEC_freight	No Transformation	Constant		-505.709	934.558	-.541
			Difference		1		
	CNY_USD	No Transformation	Numerator	Lag 0	73.382	141.782	.518

Table 14~~15~~. ARIMA Model with Exogenous Indicator, Oil Price, for USEC

Model Statistics

		Model Fit statistics				Ljung-Box Q(18)		
Model	Number of Predictors	Stationary R-squared	R-squared	RMSE	MAPE	Statistics	DF	Sig.
USEC_freight-Model_1	1	.015	.689	288.020	6.493	8.050	18	.978

ARIMA Model Parameters

					Estimate	SE	t
USEC_freight-Model_1	USEC_freight	No Transformation	Constant		-182.057	181.045	-1.006
			Difference		1		
	Oil_Price	No Transformation	Numerator	Lag 0	2.672	2.961	.902

Table 15~~16~~. ARIMA Model with Exogenous Indicator, Oil Price, for USWC with Lag = 1

Model Statistics

Model	Number of Predictors	Model Fit statistics: Stationary R-squared	R-squared	RMSE	MAPE	Ljung-Box Q(18): Statistics	DF
USWC_freight-Model_1	1	.695	.295	360.849	20.360	34.266	17

ARIMA Model Parameters

					Estimate	SE
USWC_freight-Model_1	USWC_freight	No Transformation	Constant		-1068.125	366.816
			AR	Lag 1	.524	.138
			Seasonal Difference		1	
	Oil_Price	No Transformation	Delay		1	
			Numerator	Lag 0	11.918	9.972
				Lag 1	-18.728	9.363
			Denominator, Seasonal	Lag 1	-.713	.151

Table 16~~17~~. ARIMA Model with Exogenous Indicator, PMI, for USWC with Lag = 1

Model Statistics

Model	Number of Predictors	Model Fit statistics: Stationary R-squared	R-squared	RMSE	MAPE	Ljung-Box Q(18): Statistics	DF
USWC_freight-Model_1	1	.652	.173	376.967	19.326	27.100	17

ARIMA Model Parameters

					Estimate	SE	t
USWC_freight-Model_1	USWC_freight	No Transformation	Constant		-480.236	1684.631	-.285
			AR	Lag 1	.811	.092	8.822
			Seasonal Difference		1		
	US_PMI	No Transformation	Delay		1		
			Numerator	Lag 0	6.883	30.658	.225

Table ~~18~~17. ARIMA Model with Exogenous Indicator, Exchange Rate, for USEC with Lag = 1

Model Statistics

Model	Number of Predictors	Model Fit statistics: Stationary R-squared	R-squared	RMSE	MAPE	Ljung-Box Q(18): Statistics	DF	Sig.
USEC_freight-Model_1	1	.715	.219	381.570	8.284	15.136	17	.586

ARIMA Model Parameters

					Estimate	SE	t
USEC_freight-Model_1	USEC_freight	No Transformation	Constant		-1340.995	4219.000	-.318
			AR	Lag 1	.839	.085	9.834
			Seasonal Difference		1		
	CNY_USD	No Transformation	Delay		1		
			Numerator	Lag 0	164.741	631.184	.261

Table ~~19~~18. ARIMA Model with Exogenous Indicator, Oil Price, for USEC with Lag = 1

Model Statistics

Model	Number of Predictors	Model Fit statistics: Stationary R-squared	R-squared	RMSE	MAPE	Ljung-Box Q(18): Statistics	DF	Sig.
USEC_freight-Model_1	1	.727	.251	373.649	8.188	18.462	17	.360

ARIMA Model Parameters

					Estimate	SE	t
USEC_freight-Model_1	USEC_freight	No Transformation	Constant		-1301.684	726.264	-1.792
			AR	Lag 1	.803	.096	8.360
			Seasonal Difference		1		
	Oil_Price	No Transformation	Delay		1		
			Numerator	Lag 0	18.560	11.782	1.575