



Forecasting and Trading in the Crude Tanker FFA Market

*Forecasting and applying trading strategies on Crude Tanker
Forward Freight Agreements using Neural Networks and AIS-data*

Camilla Helle Stefanski and Erik Tveiti

Supervisor: Roar Os Ådland

Master Thesis

MSc in Economics and Business Administration

Financial Economics and Business Analytics

NORWEGIAN SCHOOL OF ECONOMICS

This thesis was written as a part of the Master of Science in Economics and Business Administration at NHH. Please note that neither the institution nor the examiners are responsible – through the approval of this thesis – for the theories and methods used, or results and conclusions drawn in this work.

Preface

This master thesis is conducted as a part of the finalization of our Master of Science degree in Economics and Business Administration at the Norwegian School of Economics (NHH). The thesis is a part of the specialization in the study programs Financial Economics and Business Analytics.

We would like to express our sincere gratitude to our supervisor, Roar Os Ådland, from the Department of Business and Management Science at the Norwegian School of Economics. We are grateful for his valuable support and guidance throughout the writing process, and for sharing his knowledge within the field of shipping economics.

We extend the gratitude to the Signal Ocean Group for contributing to the thesis in terms of advice and providing cross-border data on voyage charter contracts. Also, we would like to thank the Centre for Applied Research at the Norwegian School of Economics for giving access to the Automatic Identification System data. Furthermore, we acknowledge the Baltic Exchange for sharing maritime market information on the trading of derivative contracts, Clarkson Shipping Intelligence Network for giving access to the World Fleet Register databases, and lastly, Bloomberg, for providing financial data and publications through the Bloomberg Terminal.

Abstract

The objective of this thesis is to forecast derivative prices of Forward Freight Agreements (FFAs) using machine learning techniques and investigate the profitability of implementing quantitative trading strategies. The thesis concentrates on two dirty tanker routes: TD3C, transporting oil from Ras Tenura in the Middle East to Ningbo in the Far East, and TD20, transporting oil from Nigeria in West Africa to Rotterdam in Europe.

The machine learning model predicts the future daily price movements of the individual FFA contracts, and the daily price spread of the FFA pair using a Long-Short-Term Memory (LSTM) Neural Network (NN) machine learning methodology. The model benefits from Automatic Identification System (AIS) and voyage contracts data when constructing proxies for supply, demand and geographical distribution. To capture economic development, the model utilizes macroeconomic and financial data. With a forecasting horizon of one day, the findings suggest that the LSTM model outperforms Vector Autoregressive (VAR) and Random Walk (RW) benchmark models.

To generate profitable trading signals, the forecasted individual routes and the directly forecasted price spread make use of two quantitative trading strategies: A Simple Long Short strategy and a Bollinger Band strategy. The strategies compare trading signals generated from the VAR and LSTM model with a Buy-and-Hold benchmark strategy (B&H). The results suggest that the Simple Long Short trading signals generated from the LSTM model is profitable when implemented on individual FFAs, but not profitable when implemented on the FFA pair. Conversely, the Bollinger Bands strategy combined with LSTM model is profitable when implemented on the FFA pair, but not profitable when implemented on the individual FFAs. The LSTM model combined with the two strategies outperforms the VAR model and the B&H benchmark.

1	INTRODUCTION	4
2	LITERATURE	6
3	DATA	8
3.1	DATA FOUNDATION.....	8
3.2	AIS-DERIVED FEATURES.....	10
3.3	NON AIS-DERIVED FEATURES.....	13
3.4	FINANCIAL VARIABLES.....	15
4	THEORY & METHODOLOGY	18
4.1	MACHINE LEARNING THEORY & METHODOLOGY.....	18
4.1.1	<i>Supervised Machine Learning</i>	18
4.1.2	<i>Artificial Neural Networks</i>	19
4.1.3	<i>Recurrent Neural Networks</i>	19
4.1.4	<i>Long Short-Term Memory</i>	20
4.1.5	<i>Data Pre Processing</i>	22
4.1.6	<i>Feature Selection</i>	23
4.1.7	<i>Model Evaluation</i>	27
4.2	TRADING THEORY & METHODOLOGY.....	29
4.2.1	<i>Co-Integration Approach</i>	30
4.2.2	<i>Bollinger Bands</i>	32
4.2.3	<i>Simple Long Short</i>	34
4.2.4	<i>Strategy performance</i>	35
5	RESULTS	36
5.1	FORECASTING RESULTS.....	36
5.1.1	<i>Feature selection and hyperparameter results</i>	36
5.1.2	<i>Prediction results</i>	39
5.2	TRADING RESULTS.....	41
5.2.1	<i>Individual Routes Forecasting Results</i>	41
5.2.2	<i>Spreads Trading Results</i>	44
6	DISCUSSION	49
6.1	DISCUSSION OF FORECASTING RESULTS.....	49
6.2	DISCUSSION OF TRADING RESULTS.....	50
7	CONCLUSION	51

1 Introduction

Participants of the financial markets have the potential of generating lucrative profits if future directional movements of financial assets are precisely predicted and profitable trading strategies implemented. Forward Freight Agreements (FFAs) are defined as commodity derivatives, which derive from the underlying physical shipping markets in terms of spot freight rates on major shipping routes (Baltic Exchange, 2020). As financial derivative instruments, FFAs are mainly used by shipping companies, financial institutions and traders for speculation and hedging purposes.

The first part of the thesis concentrates on predicting FFA prices in the crude tanker market, using Machine Learning methodology. Machine Learning and Artificial Intelligence has for many decades been used for computational statistics and forecasting purposes. Computational power has increased with greater technology and expanded the possibilities with stronger models able to capture non-linear relationships. With constant improving quality of marine trafficking data, the popularity of applying these model frameworks on the complex structure of this data has increased. One of the models used in this thesis is the Long Short-Term Memory Neural Network model, a feedback connection model able to learn complex relationships. The model framework is utilized to forecast the one-day-ahead price and the spread of FFA contracts applying feature engineering on AIS-and trade-data to derive relevant features. This thesis shows an improvement in the prediction compared to a Vector Autoregressive and Random Walk benchmark.

The second part investigates the profitability of trading strategies implemented on the forecasted individual routes and the directly forecasted spread of the paired FFA contracts. The derivatives market plays an important role in transferring risks from those wanting to get rid of it to those willing to bear the risk in return for profits. As trading in the FFA derivative market has become more popular, possibilities for investors to trade across different shipping routes and maturity contracts has increased. In this thesis, three trading strategies has been applied to the above-mentioned forecasting models; A Simple Long Short Strategy, a Bollinger Band Strategy, and a Buy and Hold Strategy. Utilizing the relative mispricing in the FFA contracts, the thesis shows that it is possible to take long-short positions depending on the future expectations of the price movements, and hence generate profits.

Research on the FFA derivative market and the use of AIS data in shipping markets have been covered extensively in previous literature. There are a vast amount of existing studies covering quantitative trading strategies in the derivative market. However, the literature regarding forecasting and trading in the crude tanker FFA market is considered somewhat limited, especially concerning the FFA derivatives of TD3C and TD20. The study contributes to previous research by the use of AIS data in combination with an LSTM neural network model to forecast of the price movements and price spreads of FFA derivatives in the crude tanker market. The study takes the research further by providing insights into possible methods of using quantitative strategies in the trading of the forecasted FFA prices and the directly forecasted price spread. The research covers the complexity of FFA derivative market, while still making it possible for other researchers and parakeet participants to implement the same forecasting and trading techniques.

The thesis is organized as follows: Chapter 1 presents a review of relevant literature. Chapter 2 describes the variables used in the prediction and the data gathering process. Chapter 3 covers the machine learning theory and methodology. Chapter 4 includes trading theory and methodology. Chapter 5 presents the forecasting results and the profits achieved from the trading strategies. Chapter 6 provides a discussion of the results and final conclusions.

2 Literature

There have been several studies applying machine learning techniques and methods on AIS data. Regli and Nomikos (2019) utilized feature engineering to derive measures of capacity for the VLCC fleet to capture some of the development in the weekly TD3C tanker freight rate. They found that parts of the short-term freight rate could be explained by the measured changes in supply from AIS-data. Prochazka, Adland and Wolff (2019) investigated the contracting behaviour in the tanker spot market by utilizing AIS- and fixture-data to analyse the positions of tankers at the time of fixture. They found indications of fixtures being set earlier during strong market conditions to secure oil tonnage for oil buyers. Kaluza et. al (2010) presented a network of cargo vessel movements in seaborne trade for various vessel classes using AIS.

In a similar way, Cheng et. Al (2018) mapped movement of oil tanker trajectories to identify busy routes and areas surrounding the “Maritime Silk Road” using AIS-derived data, where they found AIS data to be successful in providing an analysis of the movements. Another study, conducted by Kumaraand and Heymann (2020), applied machine learning methods utilizing classification with an Artificial Neural Network (ANN) model to detect anomalies in maritime navigation using AIS data with promising results.

Adland et. al (2017) showed that data derived from AIS could be used to obtain an estimate of seaborne crude exports, obtaining similar results to the aggregated crude export based on customs-reports. Tham (2008) investigated the TD3C tanker spot rates using ordered logistic regression and found that Brent-Dubai spreads, fleet utilization rate and availability of Very Large Crude Carriers (VLCC) were one of the primary drivers of the price.

Kavussanos and Nomikos (2003) applied forecasting models on the relationship between spot and the freight futures market, finding a long-run relationship between them. They also found that freight futures could provide additional information when forecasting spot-rates but not necessarily the other way around. Additional discoveries were indications of the shipping derivative market capturing new information before the underlying physical market due to informed investors prioritizing the derivatives market. Assman (2020) investigated both classification- and regression-methods to predict freight derivatives for directional and point-forecasting, respectively. She found that significant profits could be obtained applying trading rules for both methods and argued that the findings could indicate inefficient FFA markets.

There is a vast amount of literature and research covering spread trading strategies and other strategies implemented in the FFA derivative markets, in the tanker market, the stock market and the commodity markets. One paper, written by Alizadeh and Nomikos (2006), investigates the profitability of trading strategies in the tanker market. Based on a cointegration relationship between earnings and price, and statistical tests using a bootstrap approach, they implemented earnings-price ratio strategies. Their results suggested that the trading strategies significantly outperformed the buy and hold trading benchmark in the tanker market.

In a paper written by Gatev, Goetzmann and Rouwenhorst (1999), the profitability of pairs trading strategies was investigated using a standard deviation strategy on equities. The stocks were paired based on normalized historical co-integrated prices. In line with their expectations, the results showed that excess returns obtained from temporary mispricing existed, achieving an annualized excess return of about 11% for the top-pair portfolios. In another paper, written by Elliott, van der Hoek and Malcolm (2005), a mean-reverting Gaussian Markov chain model is proposed to be implemented as a spread trading strategy. They compared the model predictions to subsequent observations of the spread between stocks to determine investment decisions. They showed that the model had potential to generate profits when their asset prices in the financial markets deviate from the underlying equilibrium.

Kavussanos, et al (2010 and 2014a) investigated information flows and spill-over effects between the commodity futures markets and the freight derivative markets. Their findings suggest that there exist significant information flows between the two, and specifically from the commodity futures to the freight derivatives markets. Thus, investors can utilize the fluctuations in commodity futures and take appropriate positions in the FFA markets.

Chang and Geman (2016) modelled spreads between different stock pairs of oil companies and implemented a high-frequency intraday spread trading strategy. Their results obtained were remarkable, achieving a Sharpe ratio of 7.2, even after accounting for transaction costs. A more recent work written by Engmark and Haugland (2018) examined an intraday spread trading strategy based on a stochastic process model using Brent Crude oil futures contracts. They constructed different calendar spreads for trading and achieved a maximum Sharpe Ratio of 4.3. Their findings were less optimistic, and under conservative assumptions, they concluded that the stochastic model was not profitable.

3 Data

3.1 Data Foundation

In this study, selected routes for the corresponding 2nd month FFA contracts, defined by the Baltic Dirty Tanker Index (BDTI) are investigated. The choice is based on a consideration of market liquidity, the variability in data of daily prices and time horizon of the data available for investigation. The two routes are TD3C, which is operated by VLCCs, transporting crude oil from Ras Tenura in the Middle East to Ningbo in the Far East, and TD20, a Suezmax route transporting oil from the port of Bony in Nigeria, West Africa, to Rotterdam on the European continent (Baltic Exchange, 2020).

Studying the shape and behaviour of the supply and demand curves that characterize the shipping market is essential for the variable selection when predicting the price movements of the FFA contracts. The variables selection and variable extraction make use of a data foundation consisting of an Automatic Identification System (AIS) dataset from Vesseltracker GmbH provided by the Centre for Applied Research at NHH, a voyage dataset provided by Signal Ocean Group, as well as open-source financial and macroeconomic data.

The variables are selected on a geographical regional and global basis and related to the vessel type of the respective routes. The regional level is meant to capture route specific changes in local demand and supply, while the global level is supposed to capture the current market conditions for the shipping industry and the economy as a whole. Additionally, some of the variables are selected based on reported destinations to account demand and vessels heading to specific ports. Concerning vessel types, the data on VLCC tankers are linked to TD3C, while the data on Suezmax tankers are linked to TD20, as these vessel classes are the most central and prominent for the respective routes (Baltic Exchange, 2020). The chosen time horizon of the data is set to the 1st of January 2014 to the 31st of July 2019.

AIS Data

Automatic Identification System (AIS) is an automatic vessel tracking system which keeps track of vessels at sea through transceivers on a global case. The main objective of the system is to provide an overview of marine traffic both locally and globally, resulting in improved navigation, security and control in marine trafficking (Navigation Center, 2020). The system was first enforced upon voyaging vessels with a carrying capacity on more than 300 DWT by

the International Marine Organization in 2002 and has expanded in use since. The data is gathered by satellites surrounding the Earth, solidifying the quality and frequency of AIS signals. The signals vary in length, content and frequency, and in 2019 there were 27 different message types (Navigation Center, 2020). The messages include information of ship type, ship ID, positions in longitude and latitude, timestamps, speed, course, draught, destination and other ship specifications. In this study the variables are created using the AIS information on longitude, latitude, speed, course, draught, destination and vessel class. An overview of the information provided by the AIS dataset utilized in this thesis is presented in Table 1:

Table 1: AIS provided Data

AIS Data	Description
<i>Imo</i>	Unique ship identifier
<i>Longitude</i>	Longitude of the reported position
<i>Latitude</i>	Latitude of the reported position
<i>Speed</i>	Sailing speed (knots)
<i>Course</i>	Sailing course
<i>Draught</i>	Draught in meters
<i>Destination</i>	Reported destination by crew
<i>Vessel Class</i>	i.e. Tanker- or bulk carrier vessel

The IMO number is a unique vessel identifier provided for each ship type by the International Marine Organization. The IMO number is used to merge satellite data from the AIS dataset with the additional shipping data from the Clarkson World Fleet Register Database. The additional data from the Clarkson Database included in this study is vessel type, loading capacity, maximum draught and design speed. In all datasets, the information provided is filtered to only contain VLCC and Suezmax crude oil tankers, as these are the respective vessel classes for the selected routes.

An issue regarding the AIS-data is partly or entirely missing observations. Missing IMO numbers in the AIS dataset are obtained from the Clarkson Fleet Register by matching vessel names. The data observations are removed when both the IMO number and the vessel names are missing. Also, observations on non-business days and holidays are excluded, as there has been no trade taken place on these days. Information on speed, draught and destination is manually reported by the vessel crew, leaving room for error (Jia et al., 2015). Hence, if data for unique vessels are missing, interpolation is used to fill in the missing values.

Voyages Dataset

The data provided by Signal Ocean spans from 2014 until August 2020, but only data up until 31.07.2019 is used to match the AIS dataset. Like the AIS data, the subset of vessel types consists of crude tankers most prevalent on the routes of interest, namely VLCC and Suezmax. Each observation in the dataset contains a completed voyage from receiving a port call to the date of a new port call. The dataset offers voyage specifications such as dates for fixtures, loading and discharging of cargo in addition to coordinates for all ports and is involved in deriving variables for demand and operational status.

3.2 AIS-Derived Features

Fleet Productivity

Speed is included as a supply-driven feature, aiming to capture the fleet's productivity. Shipowners have the tendency to increase sailing speed as a response to increased excess demand and high freight rates (Tsioumas, 2016). In theory, this incentivises an increase in speed up until the marginal profit from being able to complete more voyage charters at a fixed day-rate equals the marginal bunker fuel cost (Stopford, 2009). However, speed is also limited to the design speed, which varies for each individual vessel (Clarkson, 2020). Classic shipping theory suggests that higher sailing speed increases productivity and therefore supply (Stopford, 2009). The opposite will apply for a reduction in sailing speed. Capturing a shift in supply through fleet speed can provide information regarding the underlying shipping market when forecasting the FFAs. The standard deviation of the speed is also included as a feature to capture speed volatility.

The daily speed is derived by measuring the distance sailed per hour elapsed each day. The elapsed time difference and distance travelled are based on the first and last observation for each vessel on a given day. A haversine formula is implemented, which calculates the distance in nautical miles sailed given the longitude and latitude, and accounts for spherical curvature. Further, a raster map is used to provide mesh grid points for the surrounding land area, avoiding occurrences where the vector provides a straight line over land as designated path. The daily mean speed for each vessel is calculated by dividing the distance sailed by the time difference, then the aggregated average speed is divided by the number of tankers sailed for the respective region. The calculation is shown in Equation 1.

$$\text{Average Sailing Speed} = \frac{\sum_{i=1}^n \frac{\text{Distance sailed}}{\text{Time difference}}}{N \text{ tankers}}$$

Equation 1: Average Sailing Speed

For the subsequent features, a dummy variable is implemented to distinguish moving and stationary ships. A threshold for a ship moving is set to 6 knots. Observations where vessel-specific speed exceeds design speed of 17.2 and 18 for the Suezmax and VLCC class, respectively are filtered out. In cases where missing data leads to zero daily observations, interpolation of the previous day is used. In situations with one daily vessel observation available, the last observed position for the previous day is used.

A map of the VLCC fleet sailing pattern through 2018 is illustrated in Figure 1:

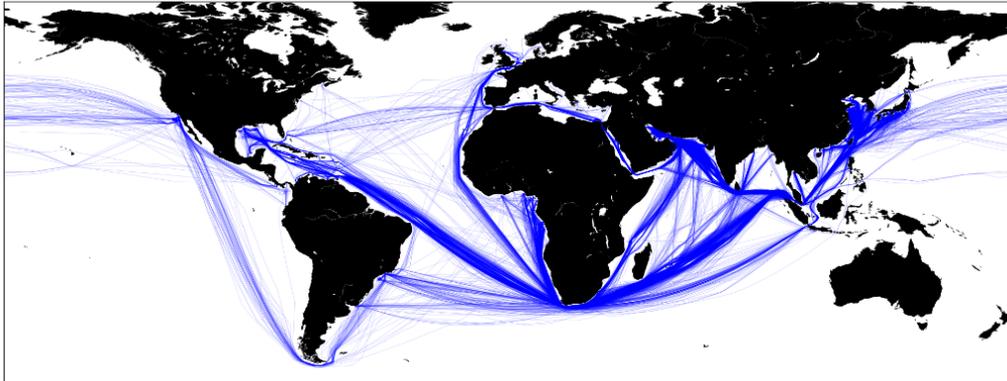


Figure 1: VLCC Fleet Sailing Pattern 2018

Capacity Utilization

Features concerning the average fleet load factor are included to further account for the fleet productivity and capacity utilization. Adland et. al (2016) showed that low freight rates lead to a state where vessels will take on suboptimal contracts where the capacity is not fully utilized. Hence, the vessel load factor provides information regarding supply and market conditions, potentially affecting FFA prices depending on the market view. Shipment quantity for each ship signal is not provided in the AIS-data. Inspired by Adland et. Al (2018), a proxy to measure the load factor is created. The reported draught is divided by the maximum draught level designed for each vessel and is then used to make an estimate of the loaded quantity at the time of the signal. The calculation of the load factor is given in Equation 2:

$$\text{Average Load Factor} = \frac{\sum_{i=1}^n \frac{\text{Draught}}{\text{Maximum draught}}}{N \text{ tankers}}$$

Equation 2: Average Load Factor

In addition to the load factor, a feature measuring the share of vessels sailing ballast is created as an indicator of the fleet capacity utilization. According to Stopford (2009), non-cargo items such as fuel, crew, and water ballast account for approximately 5% of the total DWT, leaving 95% of the remaining DWT to cargo. To account for voyages on sub-optimal contracts, the threshold of a vessel sailing ballast is set to equal or below 70% of its maximum draught. The number of vessels sailing under this threshold is calculated, giving a proxy regarding the share of vessels sailing ballast each day.

Three features are included to describe the fleet capacity. The first feature counts the number of vessels in each region of interest, capturing the capacity allocation and the supply status, indirectly accounting for new additions and demolitions of each fleet. The second capacity-feature captures the capacity available by obtaining an aggregate estimate for the total DWT of the fleet in each region. The last capacity-feature counts the in- and outflow count of vessels for each region to account for changes in local capacities.

To derive the capacity features, a world grid map is divided into several polygons by using a Ray casting algorithm featured in several previous studies, among them Narkawicz and Hagen (2016). The algorithm determines whether an observation is inside a predetermined area by testing the number of times a ray intersects the polygons given a starting point. If the ray intersects the predetermined geo-fence an odd number of times, the signal is inside the polygon. Conversely, if the ray intersects an even number of times, the signal is outside the polygon. For each latitude- and longitude-pair, a dummy variable is created and labelled for each region. In the study, the world grid map is divided into the following polygons: *North Atlantic*, *South Atlantic*, *North-Western Europe*, *Mediterranean Ocean*, *Arabian Gulf*, *Indian Ocean*, *Pacific Ocean* and *Asia* continent. The polygons are illustrated in Figure 2.

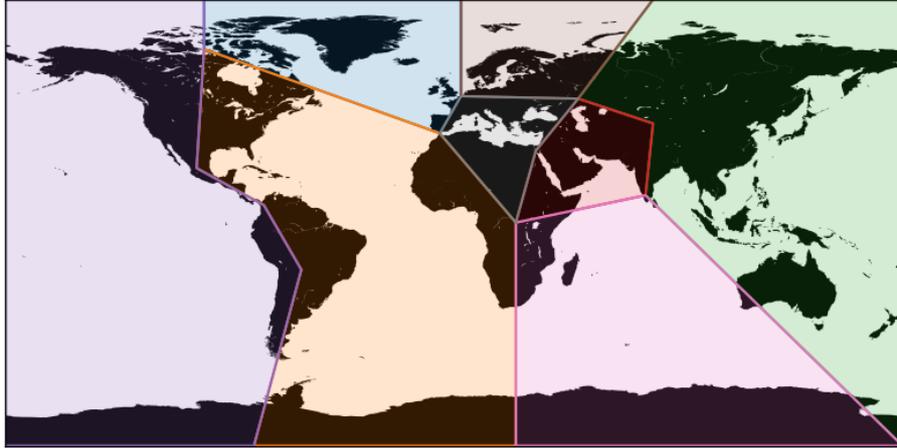


Figure 2: World Raster Map Divided in Eight Polygons

After determining in which polygons the observations are located, a count of VLCC and Suezmax vessels in each polygon is derived. Then, the daily DWT capacity available is summarized, obtaining a measure of the total supply capacity within each polygon. Finally, the in- and outflow of vessels is derived by calculating the number of times vessels cross a particular polygon.

3.3 Non AIS-Derived Features

Tonne-Mile Demand

Tanker demand is a function of the distance required to complete a voyage and the cargo quantity transported. The distance component of is denoted as the average haul of the trade, which varies with the trade locations. To incorporate average haul, the demand is measured on a “tonne-mile” basis (Stopford, 2009). Tonne-mile demand (TMD) is derived by multiplying the quantity by the distance sailed for voyages sailing laden. An increase in crude tonne-mile demand drive freight rates up as more tankers are required, which positively affects FFA prices. To create a proxy for tonne-mile demand based on time-specified fixtures in the Voyages dataset, the load- and discharge dates available for each voyage are utilized. A potential issue when determining TMD precisely, is obtaining the date for which the demand arose. In this study it is assumed that the demand arises on the loading arrival date. There is a weakness to this assumption, as the destination might not be visible to the forecasting model at the time of loading, especially if there are delivery options in place. Additionally, most fixtures are arranged long before arrival to the loading port (Prochazka et. Al, 2019).

To calculate the voyage distance, the ‘Python-Ports-Distance-Calculator’, an algorithm by Huang (2017), is used. This algorithm uses a raster map to find the shortest path from one port to another without crossing any land areas. When calculating the average haul, the loading and discharge ports are the ports of interest. For most observations, the quantity is not reported, and an assumption of laden vessels sailing at 90% of its DWT capacity is made.

To get the final estimation of TMD, the distance is multiplied by the DWT and an assumed load factor of 90% for each laden vessel. Equation 3 illustrates the calculation. The TMD is calculated globally and for each region related to the routes selected for the Suezmax and VLCC fleet. A feature measuring the cumulative sum of TMD is also created.

$$Tonne - mile\ demand = \sum_{n \in N} 90\% * DWT * Distance$$

Equation 3: Tonne-Mile Demand

Figure 3 presents the aggregated estimate of TMD per year compared to the true world seaborne trade of crude oil based on values from Clarkson Research Portal (2020). The estimated values seemingly match the true values well, deviating by approximately 1 billion tonne-miles per year except 2014 and 2019.

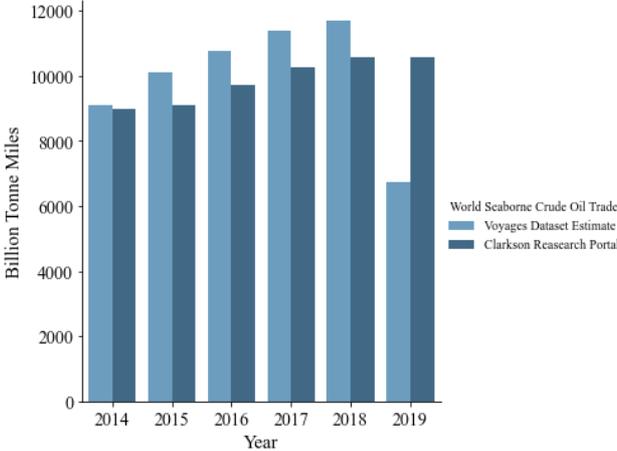


Figure 3: World Seaborne Crude Oil Trade vs. Estimated Tonne-Mile Demand

Unemployed Days

A feature counting the days between voyages is created to account for fleet operational status, measuring the time vessels remain unemployed after last cargo discharge. Adland (2019) shows an inverse relationship between idle ships and earnings in the Capesize 16 market. If this relationship is similar for the crude tanker market, one could argue that a feature capturing this

dynamic either directly or indirectly can provide useful information of the operational status. This estimate of unemployed days is calculated by obtaining the daily time difference between the last discharge and the next voyage departure for each vessel. The number of days between voyages for each individual vessel is aggregated both globally and regionally based on the area of the original port call. Calculation is done at the time a new fixture is known when sailing to a new loading port to avoid forward-looking bias. The time series of unemployed days for the Suezmax and VLCC fleet in West Africa and Asia are illustrated in figure 4.

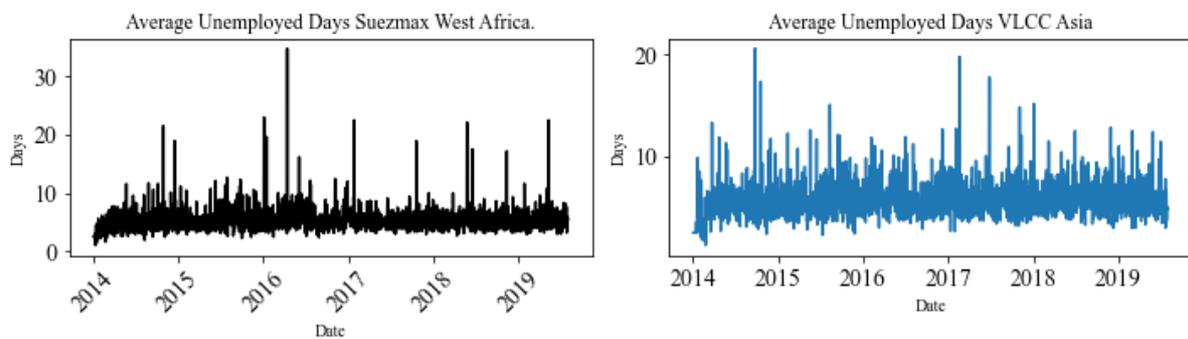


Figure 4: Unemployed Days for Suezmaxes in West Africa (left) and VLCCs in Asia (right)

According to Stopford (2009) there is a significant lag between decisions and implementations in the shipping markets. This time-lag dynamic is partly solved through the mechanics of the machine learning models, as they allow the model to look back to previous time steps when forecasting.

3.4 Financial Variables

Exchange Rates

Exchange rates are defined as the price of a nation’s currency expressed in the terms of another currency (Investopedia, 2020). Kavussanos and Visvikis (2006) argues that capturing exchange rate movements will reflect the fluctuating relationship between economies. According to Stopford (2009) the US dollar is considered the main currency in the shipping industry. Consequently, exchange rates for currencies of the routes are included, measured against the US dollar. It is assumed that a strengthening of the USD compared to the other currencies have a positive impact on the derivative prices.

Time Charter Rates

In addition to freight rates, the Baltic Exchange produces time-charter rates. One-year Time Charter Rates for Suezmax and VLCC are included to capture future spot rate expectations and period-to-period changes in revenue performance and voyages costs. The time charter rates are calculated by taking voyage revenues, subtracting voyage expenses, including canal, bunker and port costs, and then dividing the total by the round-trip voyage duration in days. (Investopedia, 2020). The VLCC Time Charter Rate uses TD1 and TD3C, while the Suezmax Time Charter Rate uses TD6 and TD20.

One can argue that the physical time charter market does not contain additional information about the future market, and that the information is already reflected in FFA prices. However, the variables are still included to investigate the relationship with the FFA prices. Köhn (2008) argues that positive changes in time-charter rates reflect market expectations of improving spot rates. Stopford (2009) argues that higher transportation costs or bunker prices may affect the shipping demand negatively. Hence, it is expected that the time charter rates are both positively and negatively correlated with FFA prices. The correlation with the FFA prices further depends on which variable is changing and the degree of the change.

Interest rates

Interest rates exhibit information about the future expectations of economic activity. Also, interest rates stimulate the economy by impacting the cost of capital and investors willingness to invest (Da et al). The LIBOR (London Inter-Bank Offered) index is an interest rate base often used in the financing of loans in the shipping market. A decline in the LIBOR rate increases the purchasing power of investors and impacts the Capex of shipping companies, leading to higher company values. The opposite is true when the market observes a rise in the interest rate. Therefore, the LIBOR-rate quoted in USD is considered important for the prediction and the rate is assumed to have a negative relationship with the FFA values.

Stock Index

A selection of stock exchange indices linked to the countries of the respective routes are included in the prediction. These are Tadawul (SA), Shanghai SE Composite Index (CH), GSEINDEX (NIG), AEX INDEX (NL) and S&P500 (US), including the VIX index. S&P500 measures the value of the 500 largest companies listed on the New York Stock Exchange, while the VIX index captures the market's and S&P500's volatility expectations. A positive

relationship between the stock indices and the FFA prices is hypothesized, as improving market conditions leads to optimism and higher trade activity both on the stock and the derivative market. A negative relationship between the FFA price and the VIX is expected, as activity is expected to decrease when fear among investors increases.

The Baltic Dirty Tanker Index

The Baltic Exchange is among the world leading providers of freight market information, contributing to the development of well-functioning derivative markets. Several indices have been constructed over the years to meet the needs of the market participants trading in the different segments. The Baltic Dirty Tanker Index (BDTI), which is this thesis most appropriate index, reflects dirty cargo voyages transporting crude oil and lower distillates of oil refineries. The final calculation of the BDTI is an equally weighted average of each individual route, quoted in WorldScale (WS) points and \$/mt. (Alizadeh and Nomikos, 2009). The BDTI is constructed using data and information served by a panel of independent shipbrokers. Every business day, the published freight rates and assessments are based on negotiations, the shipbrokers' perceptions of each trading routes' value, considerations of supply and demand, and other available market information. Consequently, the prices of the FFA contracts are mostly expected to be positively correlated with the Baltic Dirty Tanker Index.

Crude Oil Price

The Brent Oil is one of the most common price benchmarks for crude oil. It is considered to be easily transported and refined, and according to ICE (2013), the price of the Brent Oil is the origin of the pricing of 60% of the oil traded globally. Hence, the variable included is the Arab Emirates Dubai Brent oil and is supposed to capture the world demand for crude oil. Poulakidis and Joutz (2009) argue that a rise in the oil prices give an upward pressure on spot rates. The price mechanisms involved in transportation and the production of oil rather complicates the forecasting of directional movements of the oil price on the price of the FFA derivative contracts. Increased oil demand can be the source of a higher Brent price, but also increased bunker costs can put an upward pressure on the Brent price. According to Clarkson (2020), the dynamic of the oil price can have both of positive and negative effects on shipping, and hence it depends on the source of the price increase.

High Yield Spread

To capture the investor's willingness to invest money in the market, a high yield bond spread is included in the model. The index captures the market's expectations the future activity in the global economy. When investors become less risky, a tightening of the high-yield spread is expected to occur, as well as increased investments in the economy. Conversely, when investors become more risk averse, is reflected in an expansion of the high yield spread, which in turn leads to less economic activity. Westgard et al (2017) investigated the relationship between freight and high yield bond spreads and found a negative relationship in relation to the oil prices. In relation to FFA derivatives.

OSX & OVX

Two indexes related to the oil sector to capture future expectations of the shipping and derivative market. The indexes are The Phil Oil Service Sector Index (OSX) and the Crude Oil Volatility Index (OVX). The Phil Oil Service Sector Index (OSX) is a weighed market index consisting of companies that are involved in the Oil Service Sector and tracks the performance of the oil companies' share price (Investopedia, 2020). The Cboe Crude Oil Volatility Index (OVX) is an estimate of the expected 30-day volatility of crude oil as priced by the United States Oil Fund. (Investopedia, 2020). Westgaard et al (2017) used the OSX as an oil price indicator when studying oil price movements and concluded that the OSX index served as an appropriate indicator for crude oil prices. It is expected that the OSX is positively correlated with the FFA prices, while the OVX can is expected to be negatively correlated with the oil prices, as increased volatility usually reflect increasing fear among investors.

4 Theory & Methodology

4.1 Machine Learning Theory & Methodology

4.1.1 Supervised Machine Learning

Machine learning is often divided into two subcategories, unsupervised and supervised learning. Supervised machine learning is the process where an algorithm is trained on data with labelled inputs in order to learn the mapping function between an input X and output \hat{Y} . When learning the model has access to previous examples of the X and Y combinations in terms of training data, eventually returning a prediction of the out-of-sample data based on what is learned. In unsupervised machine learning, the model is presented with data in which

the data label is unknown. The objective of an unsupervised machine learning model is to learn the patterns and clusters for the unknown labelled inputs. This thesis deals with labelled data that is already known, trying to optimize the mapping function and obtain a prediction, making it a supervised machine learning study.

4.1.2 Artificial Neural Networks

Artificial Neural Networks (ANN) are computational systems based on the principles behind biological neural networks. Neural networks consist of neurons able to communicate through electrical signals and are structured in different layers. The bottom layer consists of the predictors (inputs to the model), while the outputs (predictions) make up the upper layer. Additionally, there could be layers in between these two layers, often denoted as hidden layers. A NN without hidden layers resembles a linear regression. When the hidden layers are added, the model becomes non-linear. Based on the type of neural network, these hidden layers can have different properties. An example of this is a Recurrent Neural Network.

4.1.3 Recurrent Neural Networks

Recurrent neural networks, in contrast to normal neural networks, can use previously iterated time steps, and subsequently inform later time steps by preserving memory. An illustration of a recurrent neural network is shown in Figure 5.

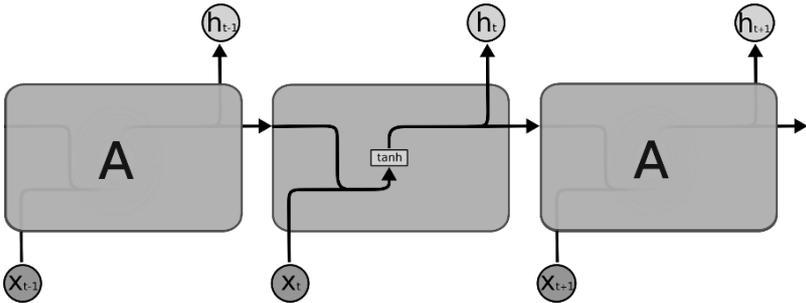


Figure 5: A simple RNN (Figure inspired by Olah., 2015)

At each cell A, the model returns to previous information when evaluation input X_t X_t through the activation of the hidden layer, h_{t-1} h_{t-1} , to form an output, (prediction) y_{t+1} y_{t+1} . A simple RNN can preserve this information to an extent. However, as the number of time steps increase, the context depreciates and connecting the same dots might not be easy or beneficial to the model, as this leads to the model not learning properly. What happens is that the values

propagated forward explode or vanish, often referred to as the vanishing and exploding gradient problem (Hochreiter et al., 2001). During training of a model, the gradients in terms of weights early in the model can be very small. For each update of these weights throughout the model, they remain small. As this carry through the model, it essentially learns less over time and brings forward information poorly. The opposite is the case for an exploding gradient, as they grow exponentially and heavily influences the model negatively as time goes on. A framework able to account for the vanishing gradient problem is the LSTM model.

4.1.4 Long Short-Term Memory

Long short-term memory networks are variants of RNN, able to capture and learn long-term dependencies due to its cell structure, hence the name. An LSTM-chain is shown in Figure 6:

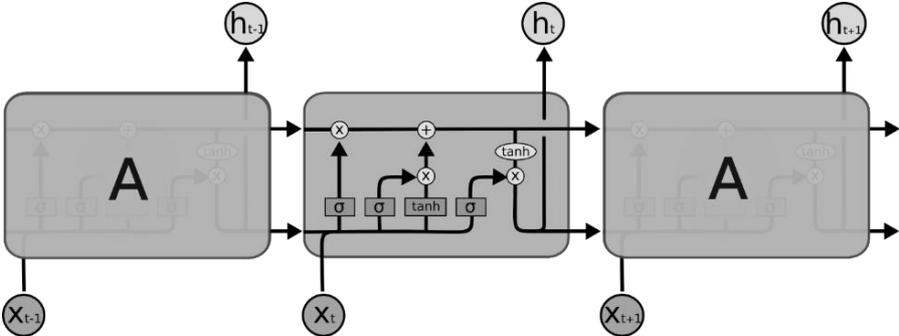


Figure 6: LSTM Neutral Network (Figure inspired by Olah, 2015)

In addition to the hidden layer activation from the RNN, there is an extra connection between each cell denoted as the cell state vector Ct between each LSTM cell. At each time step, the LSTM cell has the alternatives to either read from the vector, write to it, or reset the cell using the mechanics of gating. Within an LSTM cell there are three gates with different functions affecting the cell state vector: *Forget gate layer*, *input gate layer*, and *output gate layer*. All these cells have a *Sigmoid* activation, denoted as σ , which is there to form smooth curves in the range of zero to one and the model stays differentiable. Some gates have a *tahn* activation, a mechanism that distributes the gradients well, and allows the information in the cell state vector to contain information longer without vanishing or exploding. The gates are explained one by one, starting from the far left in figure 6. The forget gate layer (function 7) is the first layer, which looks at the previous output h_{t-1} and the current input x_t . If the sigmoid function returns “0” for the cell state, it implies that it is forgotten. If a “1” is returned, it means that it is completely kept moving forward.

$$f_t = \sigma(W_f * [h_{t-1}, x_t] + b_f)$$

Figure 7: Forget Gate Layer

The *input gate layer* consists of a sigmoid activation layer as well as a following *tahn* activation layer. Firstly, the sigmoid layer decides which values should be updated. In parallel, the *tahn* layer forms a vector of new values that are candidates to be added to the cell state

$$i_t = \sigma(W_i * [h_{t-1}, x_t] + b_i)$$

Figure 8: Input Gate Layer

$$\tilde{C}_t = \tanh(W_C * [h_{t-1}, x_t] + b_C)$$

Figure 9: Cell State Vector

What happens next is combining the functionality of the *forget gate layer* and *input gate layer* to update the previous cell state C_{t-1} to C_t . Here, the old cell state is multiplied by the function where what should be forgotten is decided, f_t . The next step is adding new candidate values multiplied by the function determining to which degree each state value is updated. Finally, the output of the cell needs to be decided in the output layer gate. First, a sigmoid layer is run, deciding which parts of the cell states to be included. Second, the cell state goes through the *tahn* layer and scales the values between -1 and 1. Lastly, output from the sigmoid layer is multiplied with the output from the *tahn* layer, resulting in the final output for that time step. The cell state vector is then passed on to the next time step, repeating the same process.

Hyperparameters

Two types of parameters are present when training machine learning models. One is where the model configures and adjusts the internal weights during training based on what the model learns. Second, there are parameters that are set manually before training by the researcher. The aim is to optimize these parameters to make the model training as efficient as possible (Yang and Shami, 2020). Hyperparameter-tuning can either be done manually on a validation partition of the dataset or through various machine learning methods. For this thesis, a grid search optimizer algorithm is performed on the validation data (Autonomio Talos, 2019). The grid search is based on a parameter boundary for each hyperparameter, where combinations up to a threshold are subsequently tested one by one on the validation data only, separate from the remaining out-of-sample test data. The hyperparameter combination that returns the lowest Root Mean Square Error is selected to be applied on the testing data.

The first hyperparameters considered in this thesis are the *numbers of hidden layers* and the *number of units* within each layer, where the number of layers determines how many hidden

layers there are between each input and output. The number of units for each hidden layer determines the dimension of each cell within a layer, in practice the units that the activations travel through before going to the next time step. *Learning rate* is also set, and a low learning rate converges easily but take longer to learn, while a high learning rate learns fast but converges slowly. The *window size* is the amount of time steps the model uses as “look-back” time to predict the next time step. In addition to these, a parameter accounting for *regularization* is added, ensuring that some of the recurrent information in the network gets dropped to avoid overfitting. Lastly, the *batch* and *epoch* size are set. The batch size is denoted by the number of samples that are worked through before the internal parameters of the model are updated. The epoch size defines the number of times that the learning algorithm works through the training dataset.

Forecasting Horizon

Selecting the right forecasting horizon is essential in determining what to forecast and the time horizon the data should be extracted for. In this thesis the data is sampled daily and aims to forecast using the sliding window method. The sliding window method utilizes a fixed lag of features p to predict the next time step, sliding one forward each time. For instance, for an input of feature p ten steps lagged $x_{p,t-10}$ to predict \hat{Y}_{t+1} , this would roll forward as $x_{p,t-10}$ for that same feature when predicting \hat{Y}_{t+2} . As previously mentioned, this could contribute to the prediction because of the time-lag dynamics in the shipping market. Since this thesis aims to utilize trading strategies on the predicted values, a one-day-ahead forecasting horizon has been chosen, where \hat{Y}_{t+1} is predicted for each time step.

4.1.5 Data Pre Processing

Pre-processing of the data is done before the models were implemented. One concern is whether the features show signs of stationarity or not. If a time series is non-stationary, trends and seasonality is present, affecting the values of the time series at various points in time (Hyndman and Athanaopoulos, 2018). On the other hand, if the time series do not depend on the time at which the series is observed, it is stationary. To test for stationarity, an Augmented Dickey Fuller-test is performed on all features. The test checks whether a unit root is present in the time series, with the null hypothesis of a unit root being present and an alternative hypothesis that the time series is stationary. For some features the null hypothesis cannot be rejected at a 5% level and an attempt to make the time series for the features stationary

is done by differencing (Hyndman and Athanaopoulos, 2018). The Dickey-Fuller test after differencing for features connected to each route is presented in the Appendix.

The second step in pre-processing the data for supervised learning is splitting the data into a training, validation and test set. The split for this thesis is 60%, 10% and 30% respectively. Doing this split is important to train the model on a subset of the data without having knowledge of the subsequent observations, as the objective is to feed the model unseen data and use the inputs for each feature to make a prediction. The reasoning behind the split selection is to have enough data to train and validate the model architecture, while still maintaining sufficient data to perform a prediction and trade on. Test data is then compared to the predicted values and model performance is evaluated. The choice of splitting into training, validation, and test before applying a scaler to normalize the data is done to prevent as much data leakage as possible. After the data is split it is ready for normalization.

Artificial neural network models tend to make assumptions regarding the distribution of the data, in some cases making scaling unnecessary. However, due to the varying nature of the variables, it is expected that the training of the model will be slower and not as optimal if the data is not scaled. By normalizing the data before applying it to machine learning models, the estimation error can be reduced in addition to the calculation time when training the model (Sola and Sevilla, 2020). When normalizing, the scaler is fit and transformed on the training data by utilizing Equation 4. Next, the scaler is applied on the test data with the same scaler fit on the training data. In that way, data leakage from the training to the test data is prevented.

$$X_{norm} = \frac{X - X_{min}}{X_{max} - X_{min}}$$

Equation 4: Normalizing Formula

4.1.6 Feature Selection

Feature selection is the process of finding a subset of features that give the best predictions among a selection of features (James et al. 2013). The reasoning behind selecting only certain features is because some features do not contribute or are irrelevant to the prediction. For high-dimensional datasets, an increasing number of features also increases training time for the model and in some scenarios leads to an increasing risk of overfitting. Feature selection also simplifies the model, reducing computational time and increases accuracy due to less misleading data. Lagged features are also included in the feature selection to capture time-

lagged effects. Feature selection methods are often divided into three categories: filter, wrapper, and embedded methods. Wrapper methods require significantly more computational power, as they essentially use cross-validation and one by one attempts to learn which predictors contribute the most to the performance when added. Filter methods have shown to be slower, less accurate and more prone to overfitting, and could perform worse when dealing with too many features (Sánchez-Marono, 2020). Embedded methods combine some of the qualities of filter and wrapper methods, where the algorithm responsible for feature selection is an integrated part of the learning algorithm. As a result, primarily embedded methods are used in this thesis, except adding a linear correlation and multiple regression, as they might contribute additional information in cases of linear relations between features. Inspired by Næss (2018), a mean score is attributed to each feature based on the rank assigned by all the feature selection methods for a specific feature. The coefficients for all feature selection methods are scaled in the range [0,1] for better comparison.

Multiple linear Regression

A multiple linear regression model is a statistical method used to predict a dependent variable based on several independent variables, where the objective is to model the linear relationship between the dependent and independent variables. The formula for a multiple linear regression is denoted in Equation 5. (James et. al, 2013).

$$y_i = \beta_0 + \beta_1 x_{i1} + \dots + \beta_p x_{ip} + \epsilon$$

Equation 5: Multiple Linear Regression Formula

Where y_i is the dependent variable, β_0 is the y-intercept, β_p is the coefficients determining the slope for each independent variable, and x_{ip} is the independent variable value for each i for every feature p . As with a regular Ordinary Least Squares model, the objective is to estimate the coefficients β_p that minimize the residual sum of squares (RSS) given in Equation 6.

$$RSS = \sum_{i=1}^n (y_i - \hat{y}_i)^2$$

Equation 6: Residual Sum of Squares

Linear regression assumes linear relationships between the independent and dependent variables, as well as independence between observations. This might not always be the case for features presented in this study but is included under the presumption of information gain regarding the feature selection.

Lasso Regression

The Lasso regression method is a feature selection method which the framework of the ordinary least squares is used when fitting the model. In addition, this method incorporates a penalizing factor $\lambda \sum_{k=1}^p |\beta_k|$ which shrinks the coefficient estimates towards zero, as seen in Equation 7 (James et.al 2013).

$$\sum_{i=1}^n \left(y_i - \beta_0 - \sum_{j=1}^p \beta_j x_{ij} \right)^2 + \lambda \sum_{k=1}^p |\beta_k|$$

Equation 7: Lasso Regression Formula

Where n is the number of observations, p is the number of features/variables and λ is the penalty coefficient. With an increasing penalty coefficient λ , some coefficients β_k shrink to 0 as a result, which rules out this coefficient and therefore feature. When λ is equal to 0 the remainder is just the ordinary least squares.

Random Forest

Random Forest (RF) is a tree-based method which can be used for prediction and feature selection among other uses. RF is an embedded selection method and used to select features of greatest importance to the model. RF consists of several hundred trees that all are “grown” based on randomly selected subsets and features from the dataset. The trees do not have information regarding the other branches, preventing overfitting and correlation-issues. Random forest utilizes what is called bagging. Bagging is the process of creating multiple copies of the training data using bootstrap, subsequently fitting each copy with a unique decision tree and then combining all these trees resulting in a final tree (James et al., 2013). At the end, the parts having the least variance increases the importance for that feature. At each node of the tree, the tree divides into two parts determining the predictability of a feature.

Gradient Boosting (XGBoost)

Gradient boosting is also a tree-based method frequently used for embedded feature selection, with resemblance to the random forest framework. However, instead of fitting several unique

trees, the trees are grown sequentially, essentially utilizing information from previous grown trees to grow the next (James et al., 2013). There are three options for measuring the feature importance in the XGBoost framework, namely weight, cover and gain. The weight is used in this instance and is a measurement of how many times a certain feature is used when splitting the data across the range of trees grown. The importance of a certain feature is reflected in the importance score based on how weighted a particular feature is.

Linear Correlation

Predictors can correlate with the target of the prediction. The most used measure for correlation is the Pearson correlation coefficient, given in Equation 8. In terms of feature selection, a large positive correlation score indicates that the predictor and target feature move in the same direction, giving added value to that feature.

$$\rho_{X,Y} = \frac{cov(X,Y)}{\sigma_X \sigma_Y}$$

Equation 8: Correlation Coefficient Formula

FFA Price Methodology

The FFA prices of TD3C and TD20 are log-transformed in addition to the difference- and normalizing-process described earlier in the methodology in order to make the inputs as indistinguishable compared to the real values as possible during training. The TD3C-TD20 spread is precalculated by deriving the natural logarithmic value of both TD3C and TD20 before subtracting them as TD3C-TD20. TD3C and TD20 are predicted separately, while the TD3C-TD20 spread is predicted directly. Features included when performing feature selection for TD3C-TD20 directly are the same features based on the individual routes TD3C and TD20.

4.1.7 Model Evaluation

Random Walk Benchmark

Random walk is a commonly used benchmark model primarily based in financial theory when predicting the stock market. The intuition behind the model is that when forecasting one step ahead, the subsequent movement of a security is random and cannot be predicted, and therefore the prediction for tomorrow \hat{Y}_{t+1} given a forecasting horizon of 1 day can simply be stated as the value today, Y_t , plus an error term ϵ_t (Hyndman and Athanasopoulos, 2018).

Vector Autoregressive Benchmark

A Vector Autoregressive (VAR) model is a commonly used alternative for multivariate time series forecasting due to its strong bidirectional capabilities. In comparison to a univariate time series model such as ARIMA, VAR can utilize the relationship between the predictors and the variable to be forecasted when training. In practice this implies that every variable affects the others as well as the independent variable (Hyndman and Athanasopoulos, 2018). The mechanics of an autoregressive vector is illustrated below in Equation 9 and 10, where a simple example of two variables with a lag of 1 is presented.

$$Y_{1,t} = c_1 + \theta_{11,1}y_{1,t-1} + \theta_{12,1}y_{2,t-1} + e_{1,t}$$

Equation 9: Vector Autoregressive Model variable 1

$$Y_{2,t} = c_2 + \theta_{21,1}y_{1,t-1} + \theta_{22,1}y_{2,t-1} + e_{2,t}$$

Equation 10: Vector Autoregressive Model variable 2

Where c_i is a k -vector of constants acting as the intercept of the model. $e_{i,t}$ denotes the k -vector of error terms. The first coefficient $\theta_{ii,l}y_{i,t-1}$ accounts for the impact that the l th lag of variable y_i has on itself, while $\theta_{ij,l}y_{j,t-1}$ considers the impact of the l th lag of variable y_j on y_i . The optimal number of lags for the model is chosen based on the Akaike Information Criterion (AIC), where the quality of each model according to the estimator is measured and the best model is returned (the one with l lags for each feature).

Performance Metrics

Several metrics are considered when evaluating the performance of the predictions conducted on the test data. Root Mean Square Error (RMSE) is one of the most common metrics of performance within machine learning. RMSE denotes the square root of the average difference between the predicted and true values squared.

$$RMSE = \sqrt{\frac{\sum_{i=1}^N (Y_i - \hat{Y}_i)^2}{N}}$$

Equation 11: Root Mean Square Error

Another performance metric widely used for forecasting is Mean Absolute Percentage Error (MAPE). This performance metric penalizes to a larger degree negative error compared to positive errors but has the advantage of being unit-free (Hyndman and Athanasopoulos, 2018) MAPE is shown in Equation 12 below.

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

Equation 12: Mean Absolute Percentage Error

Where A is the actual value, F is the forecasted value and n is the number of observations. Mean Absolute Error is also used as a metric of performance, which measures the sum of the absolute difference between the true y_t and predicted value \hat{y}_t divided by the number of observations n (Hyndman and Athanasopoulos, 2018)..

$$MAE = \frac{1}{n} \sum_{j=1}^n |y_j - \hat{y}_j|$$

Equation 13: Mean Absolute Error

4.2 Trading Theory & Methodology

One part of the trading strategy implementation investigates the profitability of applying trading strategies to the forecasted FFA individual price time series. A Simple Long Short Strategy is compared to two benchmarks consisting of a Bollinger Bands and a Buy and Hold Strategy. The other part investigates the same trading strategies implemented on the directly forecasted spreads of the FFA contracts, utilizing the difference in the future price movements. Considering the price spreads, it is possible to profit from the convergence or divergence in prices by holding simultaneously long-short positions in the FFA derivatives.

In spread trading, a long signal means buying the relatively expensive contract and selling the relatively cheap contract. A short signal means selling the relatively expensive contract and buying the relatively cheap contract. Whether the FFA contracts for the respective routes are considered relatively cheap or expensive is in this case determined by the \$/mt-value of the FFA contracts. Hence, the route with the highest \$/mt value is considered the most expensive, while the route with the lowest \$/mt value is considered the cheapest. Another, perhaps more precisely way to determine which contracts are cheap and expensive, is to also take route distance into consideration. The WorldScale measurements quoted by the Baltic Exchanges could be used to determine relatively expensive and cheap routes, and afterwards convert the World Scale points it into a USD-values. But as mentioned, in this case only the prices of the contracts quoted in \$/mt by the Baltic Exchange are taken into consideration.

The Simple Long Short and The Bollinger Bands trading strategies rely on forces of supply and demand in the market to correct the mispricing of the FFA contracts, both in the case of individual routes or the relative value of the spreads. According to Gatev et al (2006), short-term liquidity shocks occur and cause prices to converge or diverge. Both entering individual positions and combining opposite position of the FFA securities are ways to obtain different risk-exposures to the market. In contrast to trading individual routes, combining simultaneously opposite positions has an element of diversification, and is hence expected to lead to a more neutral risk-exposure to the market.

4.2.1 Co-Integration Approach

A co-integration approach is chosen to investigate the short- and long-term statistical relationship between the predicted spreads. The cointegration method states that two time series that are integrated of order d , for example in the case where $d = b = 1$, and can be linearly combined to produce a single time series that is integrated of order $d - b$, where $b > 0$. The linear combined time series is then said to be stationary mean reverting. (Engle and Granger, 1983).

$$\ln FFA_{A,t} = \mu + \kappa * \ln FFA_{B,t} + \epsilon_t$$

Equation 14: Co-Integration Method Formula

$$Y(t) = \ln FFA_{A,t} - \kappa * \ln FFA_{B,t} = \mu + \epsilon_t \sim N(\mu, \sigma)$$

Equation 15: Co-Integration Method Formula

In Equation 14 and 15, γ is the cointegration coefficient, ϵ is the residual of the regression, κ is the proportion of asset B for each unit of asset A and μ reflects the mean spread. In pairs trading, it is desirable to investigate if the spread between two pairs are cointegrated (mean reverting) by checking the stationarity of the residuals in the cointegration regression. If cointegration is found to be present, then a spread trading strategy is normally applied to generate long or short trading signals. In this thesis, the spread trading strategies are applied to the route combination. According to Kavussanos and Visvikis (2006), a necessary condition for successful and profitable spread trading is high long-term correlation and low short-term correlation between the two spread pairs. The long-term correlation measures the degree of co-movement and correlation between price levels in the long run. The short-term correlation reflects how much prices diverge in the short term and measures the degree of co-movements between the price changes.

To evaluate the short-term statistical relationship, the chosen pairs are initially tested through their price returns correlation coefficients, using a testing range from January 3rd, 2014 to November 11th 2017, and a trading horizon of approximately one and a half year, from November 11th 2017 to June 31st 2019. The correlation coefficient measures the degree to which paired asset prices return move together, taking into consideration their standard deviation. The formula used for evaluating short-term relationships is presented in Equation 16:

$$\rho(y_r, x_r) = \frac{\sum_{t=1}^n (y_{r,t} - \mu y_r)(x_{r,t} - \mu x_r)}{\sqrt{\sum_{t=1}^n (y_{r,t} - \mu y_r)^2 \sum_{t=1}^n (x_{r,t} - \mu x_r)^2}}$$

Equation 16: Correlation Coefficient Formula

The correlation coefficient of the testing range always takes a value between zero and one. When the coefficient is equal to zero ($\rho(y_r, x_r)=0$), there is no relationship between movements of the price returns. When the correlation coefficient is more than zero ($\rho(y_r, x_r)>0$), the price returns of the FFA contracts move in the same directions. When the correlation coefficient is less than zero ($\rho(y_r, x_r)<0$), the asset price returns of the FFAs move in the opposite directions. To examine the long-term statistical relationship, an Engle granger cointegration test is used, where the stationarity of the paired asset prices is tested using augmented Dickey-Fuller unit root test. The test is checking whether the individual price time series are non-stationary and then, if individual price time series difference are stationary.

The Augmented Dickey-Fuller test evaluates whether paired FFA assets price spreads, individual asset prices and individual asset price difference time series are mean stationary.

$$\Delta s_t = \alpha + \beta_t + \gamma s_{t-1} + \delta \Delta s_{t-1} + e_t$$

Equation 17: Mean Stationarity Formula

The null hypothesis of the Augmented Dickey-Fuller test states that the time series contains a unit root and are not stationary. The p-values of the γ -coefficient are the ones that are tested. If the p-value is less than 0.05 ($p < 0.05$), the time series are mean stationary, and the null hypothesis is rejected with 95% statistical confidence. If the p-value of the γ -coefficients is greater than 0.05 ($p > 0.05$), then the null hypothesis is not rejected, and with 95% statistical confidence, the series are not mean stationary.

To make the monetary contract sizes of the spreads comparable, the FFA prices of each pair are regressed to find a κ -value of the FFA contract. The value is computed by a regression based on the historical price time series. The rationale for including the κ is that for every position taken in one contract, the κ -value tells how many positions to take in the second contract of the pair. For example, when 10 000 contracts are bought of one FFA, the value tells how many contracts to buy of the other FFA in the same pair. After the monetary sizes are made comparable, different trading strategies are implemented to the FFA contracts. In addition to contract size, transactions costs are considered. The transaction cost is set to one cent per trade.

Furthermore, for all strategies, it has to be noted that the positions in the FFA contracts are closed when the contract rolls-over to a new period. For monthly contracts, the trading period is around 1 month.

4.2.2 Bollinger Bands

The Bollinger Bands strategy involves constructing upper and lower bands following the forecasted spread time series and individual price time series. The bands of Bollinger Bands expand and contract, representing volatility measures, where the lower bands will increase during periods of high market volatilities and narrow during periods of low volatilities. To construct the Bollinger Band strategy, an exponential moving average (EMA) is calculated over a window of n days of the forecasted time series values. The bands are then calculated by multiplying the standard deviations over the same window as the EMA, which is then multiplied by a factor, alpha. The upper band is constructed by adding the calculated standard deviation value to the EMA series, and the lower band is constructed by subtracting the standard deviation value from the EMA series. The formulas are presented below, where $\widehat{SFFA}_{t+1|t}^{upper}$ and $\widehat{FFA}_{t+1|t}^{upper}$ are the upper bands of the spread and FFA price, and $\widehat{SFFA}_{t+1|t}^{lower}$ and $\widehat{FFA}_{t+1|t}^{lower}$ are the lower bands. The α is in this case set to be equal to one. The window of which the EMA and the standard deviations are calculated is set to twenty days. The formulas below show the Bollinger Bands applied to the *forecasted spreads time series* of the FFA contracts:

$$\widehat{SFFA}_{t+1|t}^{upper} = EMA_SFFA_{t+1|t} + \alpha * \sigma_SFFA_{t+1|t}$$

Equation 18: Upper Bollinger Band of predicted FFA Spread

$$\widehat{SFFA}_{t+1|t}^{lower} = EMA_SFFA_{t+1|t} - \alpha * \sigma_SFFA_{t+1|t}$$

Equation 19: Lower Bollinger Band of predicted FFA Spread

$$\sigma_SFFA_{t+1|t} = \sqrt{\frac{\sum_{i=t-n}^t (\widehat{SFFA}_{t+1|t} - \widehat{SFFA}_{t+1|t})^2}{n-1}}$$

Equation 20: Standard deviation of predicted FFA Spread

Where $\widehat{SFFA}_{t+1|t}^{upper}$ is the upper band of the forecasted FFA spread, $\widehat{SFFA}_{t+1|t}^{lower}$ is the lower band of the forecasted FFA spread, $EMA_SFFA_{t+1|t}$ is the exponential moving average of the predicted FFA spreads over the last twenty days, α is the factor multiplied with the standard

deviation, and $\sigma_{\widehat{SFFA}_{t+1|t}}$ is the standard deviation of the predicted values over the last twenty days.

The formulas below show the Bollinger Bands applied to the *forecasted individual price time series* of the FFA contracts:

$$\widehat{FFA}_{t+1|t}^{upper} = EMA_{\widehat{FFA}_{t+1|t}} + \alpha * \sigma_{\widehat{FFA}_{t+1|t}}$$

Equation 21: Upper Bollinger Band of predicted FFA Price

$$\widehat{FFA}_{t+1|t}^{lower} = EMA_{\widehat{FFA}_{t+1|t}} - \alpha * \sigma_{\widehat{FFA}_{t+1|t}}$$

Equation 22: Lower Bollinger Band of predicted FFA Price

$$\sigma_{\widehat{FFA}_{t+1|t}} = \sqrt{\frac{\sum_{i=t-n}^t (\widehat{FFA}_{t+1|t} - \overline{\widehat{FFA}_{t+1|t}})^2}{n-1}}$$

Equation 23: Standard Deviation of Predicted FFA Price

Where $\widehat{FFA}_{t+1|t}^{upper}$ is the upper band of the forecasted FFA price, $\widehat{FFA}_{t+1|t}^{lower}$ is the lower band of the forecasted FFA price, $EMA_{\widehat{FFA}_{t+1|t}}$ is the exponential moving average of the predicted FFA price time series over the last twenty days, α is the factor multiplied with the standard deviation, and $\sigma_{\widehat{FFA}_{t+1|t}}$ is the standard deviation of the predicted values over the last twenty days.

Signal generators are activated when one of the bands are reached. When the real value of the FFA spread (or real individual FFA price) is lower than the lower threshold, $SFFA_t < \widehat{SFFA}_{t+1|t}^{lower}$ (or $FFA_t < \widehat{FFA}_{t+1|t}^{lower}$), after being greater than the lower threshold, $SFFA_t > \widehat{SFFA}_{t+1|t}^{lower}$ (or $FFA_t > \widehat{FFA}_{t+1|t}^{lower}$), a long position is entered. In the case of spreads, the short position implies going short in the relatively expensive route and long in the cheapest route, while a long position implies going short in the relatively cheap route and long in the relatively expensive route. When the real spread (or individual FFA price) reverts to the forecasted value, and the real spread is larger than the forecasted value, $SFFA_t > \widehat{SFFA}_{t+1|t}$ (or $FFA_t > \widehat{FFA}_{t+1|t}$), after being smaller than the lower bound $SFFA_t < \widehat{SFFA}_{t+1|t}^{lower}$ (or $FFA_t < \widehat{FFA}_{t+1|t}^{lower}$), then the long signal is closed, and profits are taken.

A short position is entered when the real spread value (or real individual FFA price) is greater than the forecasted spread, $SFFA_t > \widehat{SFFA}_{t+1|t}^{upper}$ (or forecasted individual FFA price, $FFA_t >$

$\widehat{FFA}_{t+1|t}$), after being smaller than the upper threshold $SFFA_t < \widehat{SFFA}_{t+1|t}^{upper}$ (or $FFA_t < \widehat{FFA}_{t+1|t}^{upper}$). The short position is closed when the spread reverts to the forecasted FFA spread (or forecasted individual FFA price), and the spread is smaller than the forecasted FFA spread, $SFFA_t < \widehat{SFFA}_{t+1|t}$ (or forecasted $FFA_t < \widehat{FFA}_{t+1|t}$), after being larger than the upper bound, $FFA_t > \widehat{FFA}_{t+1|t}^{upper}$ (or $SFFA_t > \widehat{SFFA}_{t+1|t}^{upper}$).

To generate the timing of transactions in the model, the distances between the upper bands and real spread value (or individual FFA price), and the distances between the lower bands (and the individual FFA price) are calculated, which is shown in the formulas below:

$$\widehat{PD}_{t+1|t}^{upper} = \widehat{FFA}_{t+1|t}^{upper} - \widehat{FFA}_{t+1|t}$$

Equation 24: Distance between the Upper Band and Predicted FFA Price

$$\widehat{PD}_{t+1|t}^{lower} = \widehat{FFA}_{t+1|t}^{lower} - \widehat{FFA}_{t+1|t}$$

Equation 25: Distance between the Lower Band and the Predicted FFA Price

$$\widehat{SD}_{t+1|t}^{upper} = \widehat{SFFA}_{t+1|t}^{upper} - \widehat{SFFA}_{t+1|t}$$

Equation 26: Distance between the Lower Band and the Predicted FFA Spread

$$\widehat{SD}_{t+1|t}^{lower} = \widehat{SFFA}_{t+1|t}^{lower} - \widehat{SFFA}_{t+1|t}$$

Equation 27: Distance between the Lower Band and the Predicted FFA Spread

Where, $\widehat{PD}_{t+1|t}^{lower}$ is the upper price distance and $\widehat{PD}_{t+1|t}^{upper}$ is the lower price distance in terms of the individual routes. In terms of spreads, $\widehat{SD}_{t+1|t}^{upper}$ is the upper spread distance, and $\widehat{SD}_{t+1|t}^{lower}$ is the lower spread distance.

4.2.3 Simple Long Short

The Simple Long Short strategy consists of a short signal triggered when the forecasted LSTM or VAR values are predicting lower FFA spread values $\widehat{SFFA}_{t+1|t} < SFFA_t$ (or individual FFA price values, $\widehat{FFA}_{t+1|t} < FFA_t$). A long signal is entered when the forecasted value of the LSTM model or the VAR model is forecasting higher FFA spreads, $\widehat{SFFA}_{t+1|t} > SFFA_t$ (or individual prices, $\widehat{FFA}_{t+1|t} > FFA_t$). The position is held until the LSTM or VAR time series predict an opposite direction of the spreads (or individual price) movements to occur.

4.2.4 Strategy performance

To evaluate the performance of the implemented spread trading strategies, different performance measures are calculated. The trading strategies are compared through annualized return, annualized standard deviation, annualized Sharpe ratio and cumulative return charts.

Annualized return, as shown in Equation 28 is a performance metric that calculates the number of observations root of annually scaled cumulative product of daily returns. In the formulas, r_t is value of the return, p_t is the current day's price, r_a is the product of the daily return and n is number of days.

$$r_a = r_t \frac{p_t}{p_{t-1}} - 1$$

Equation 28: Annualized Return Metrics

$$r_a \left[\prod_{t=1}^n (r_t + 1) \right]^{252/n} - 1$$

Equation 29: Annualized Return Metrics

Annualized Standard Deviation is a metric of risk, that consists of daily standard deviations multiplied by square root of a number of periods per year, where σ_a is the standard deviation, r_t is the value of the return, and μ is the average value of the return over n number of periods.

$$\sigma_a = \sigma \sqrt{252}$$

Equation 30: Standard Deviation Formula

$$\sigma_a = \sqrt{\frac{1}{n} \sum_{t=1}^n (r_t - \mu)^2}$$

Equation 31: Annualized Standard Deviation Metrics

$$\mu = \frac{1}{n} \sum_{t=1}^n r_t$$

Equation 32: Average Value of Return

Annualized Sharpe Ratio consists of annualized excess return per unit of risk and is considered a risk-adjusted performance metric. The variable r_a is the annualized return, r_f is the annualized risk-free rate and σ_a is the annualized standard deviation. The formula is shown below:

$$SR_a = \frac{r_a - r_f}{\sigma_a}$$

Equation 33: Annualized Sharpe Ratio Metrics

5 Results

5.1 Forecasting Results

5.1.1 Feature selection and hyperparameter results

Feature Selection Results

Table 2, 3 and 4 show the final subset of features elected for TD3C, TD20 and TD3C-TD20, spread predictions based on mean scores derived from each individual feature selectin method. All features are ranked for each lag of the feature, and the rankings of features were prioritized when selecting, followed by testing combinations and dropping features until the improvements in performance no longer changed notably on the validation data. Due to having 88,77 and 153 features for TD3C, TD20 and TD3C-TD20, respectively, only the final subset of features are presented in the tables below. The top 30 feature sin terms of mean score for TD3C is found in the Appendix 1.2 to illustrate the feature selection proves.

Table 2: Feature Selection TD20

TD320	
AIS - Derived Features	Non - AIS Derived Features
<i>Speed_std_West_Africa_Suezmax_moving</i>	<i>BDTI</i>
<i>Speed_std_Global_Suezmax_moving_Rotterdam</i>	<i>3Month_spread_BrentCrude</i>
<i>LF_mean_Global_Suezmax_moving</i>	<i>NGN_USD_exchange</i>
<i>In_flowcount_Suezmax_ArabGulf</i>	<i>AEX</i>
<i>LF_std_Global_Suezmax_moving_Rotterdam</i>	
<i>Suezmax_Capacity_Asia</i>	
<i>Out_flowcount_Suezmax_IndianO</i>	
<i>LF_mean_global_Suezmax_moving_Rotterdam</i>	
<i>LF_WA_Suezmax_moving</i>	
<i>LF_mean_Global_Suezmax</i>	
<i>Suezmax_Capacity_Asia</i>	
<i>Out_flowcount_Suezmax_IndianO</i>	

Table 3: Feature Selection TD3C

TD3C	
AIS - Derived Features	Non - AIS Derived Features
<i>Count_Asia_VLCC_moving</i>	<i>Dubai_Crude_1vs2M</i>
<i>In_flowcount_VLCC_ArabGulf</i>	<i>CNY_USD_exchange</i>
<i>In_flowcount_VLCC_IndanO</i>	<i>S&P500</i>
<i>LF_std_ArabGulf_VLCC_moving</i>	<i>TCE_VLCC</i>
<i>Out_flowcount_VLCC_NWE</i>	<i>USD_SAR_exchange</i>
<i>Speed_mean_Asia_VLCC_moving</i>	<i>BDTI</i>
<i>VLCC_Capacity_ArabGulf</i>	

Table 4: Feature Selection TD3C-TD20

TD3C-TD20	
AIS - Derived Features	Non - AIS Derived Features
<i>Count_Asia_VLCC_moving</i>	<i>TCE_VLCC</i>
<i>LF_std_ArabGulf_VLCC</i>	<i>USD_SAR</i>
<i>Speed_mean_Global_Suezmax_moving_Rotterdam</i>	
<i>Speed_mean_Asia_VLCC_moving</i>	
<i>VLCC_Capacity_IndianO</i>	
<i>LF_std_Global_Suezmax_moving_Rotterdam</i>	
<i>VLCC_Capacity_NWE</i>	
<i>LF_mean_Global_Suez_moving_Rotterdam</i>	
<i>LF_std_Global_Suez_moving</i>	
<i>In_flowcount_VLCC_ArabGulf</i>	
<i>% Ballast_ArabGulf_Suezmax_moving</i>	
<i>Count_ArabGulf_VLCC_Ningbo</i>	
<i>% Ballast_Asia_VLCC</i>	

In- and outflows of vessels for regions based on the regional polygons seems to be a strong predictor overall, as it is included for all FFA contract prices. This suggests that the regional net flow in supply for route-specific regions positively contributes to explaining the FFA price. Regional capacity in terms of DWT is also included in all models, which implies that available tonnage capacity also contributes to understanding the price movements. For TD20, it appears to be changes in fleet capacity for other regions rather than route-specific that perform best in terms of predictability. Arguably, this is due to most of the oil trade occurring in areas of these features, possibly capturing the global Suezmax allocation better than the route-specific features.

Speed features are represented in all three models, although exclusively speed features with moving vessels. One thing to note is that speed-volatility features are the only speed features that appears to add to the predictability of the TD20 FFA contract. Among these features is the speed volatility of Suezmax vessels with destination Rotterdam, which implies that the change in speed for vessels heading to Rotterdam contributes more to the prediction rather than the mean speed for the same vessels. Features regarding the load factor is also present for all three models to a large degree, which is interesting considering that the proxy feature aiming to measure the share of vessels sailing ballast should be a more direct measurement of capacity utilization. Even if not present in the TD3C or TD20 model, two features measuring this is included in the spread-model. It is also intriguing that non-AIS features to a larger extent seems to show better predictability for the individual routes compared to the combined spread. The exchange rate between the local currency and USD is present for all models, which appears to overall contribute well as a predictor of price movements. BDTI is included for both TD3C and TD20 which is not surprising as it is expected and confirmed in appendix XX to be positively correlated with the FFA contract. It is also worth noting that the Dubai oil futures and Brent Oil spread are included for the TD3C and TD20 price, respectively. This could arguably show that the geographically close oil spreads hold more predictive power when it comes to predicting the price of each FFA.

Hyperparameter tuning

The final hyperparameter selection after performing the grid search for each LSTM model is shown in table 5.

Optimal hyperparameters varies between each model and does not provide much info regarding the capabilities of the models before tested on out-of-sample data. However, Epochs-sizes indicate that the TD3C-TD20 and TD3C model requires more training than TD20.

Table 5: Hyperparameter Tuning

Hyperparameter Tuning			
	TD3C	TD20	ln TD3C - ln TD20
Hidden Layers	2	2	1
Units	64,16	32,16	16,00
Learning Rate	0,01	0,10	0,10
Dropout (Regularization)	20 %	20 %	20 %
Window Size	4	8	10
Batch Size	30	40	30
Epochs	100	50	150

5.1.2 Forecasting Performance

Table 6 presents the results for the one-day-ahead prediction models for the TD3C and TD20 FFA 2-month maturity contracts in addition to direct prediction of the TD3C-TD20 spread. Due to predicting one-step-ahead, therefore being frequent timesteps, it is not expected to see large volatility in predictions from one day to another, as the previous day is taken into consideration by the models when predicting the next step. The model output for the FFA prices of TD3C, TD20 and TD3C-TD20 is the result of inverse transformed values to match the true format after prediction, which the performance metrics are based on

Table 6: Forecasting Performance

FORECASTING PERFORMANCE			
	TD3C	TD20	ln TD3C - ln TD20
LSTM			
RMSE	0.287	0.180	0.029
MAPE	1.67%	0.95%	—
MAE	0.150	0.093	0.022
VAR			
RMSE	0.301	0.187	0.024
MAPE	1.91%	1.16%	—
MAE	0.170	0.113	0.017
RW			
RMSE	0.369	0.230	0.027
MAPE	2.02%	1.23%	—
MAE	0.190	0.130	0.018

The Random Walk model and Vector Autoregressive model acts as the benchmarks to the LSTM model regarding predictive performance. The LSTM model outperforms the benchmark models in terms of RMSE and MAE for the TD3C price with a RMSE of 0.287 compared to 0.301 and 0.369 for VAR and RW, respectively. The LSTM model is also better than both benchmarks in terms of the Mean Absolute Error and MAPE for the TD3C contract, with a MAE and MAPE of 0.150 and 1.67%, respectively. The LSTM model performs well on the TD20 contract price as well, beating both benchmarks for all performance metrics included. When predicting the TD3-TD20 FFA spread the LSTM performs worse than both benchmarks with a RMSE of 0.029 compared to the 0.018 of the RW, while the VAR model performs slightly better than the RW in predicting the spread, in terms of RMSE and MAE. MAPE is not reported for the TD3C-TD20 spread for any of the models due to the true values of the price spread reaching zero.

When visually inspecting the graphs for the LSTM- and VAR-models, it is apparent that fluctuations in predictions are quite small. However, the VAR model predictions are to a greater extent volatile.

The graph of real predicted values of the LSTM and VAR for TD3C is shown in figure 10 and 11 respectively. The remaining graphs are found in appendix

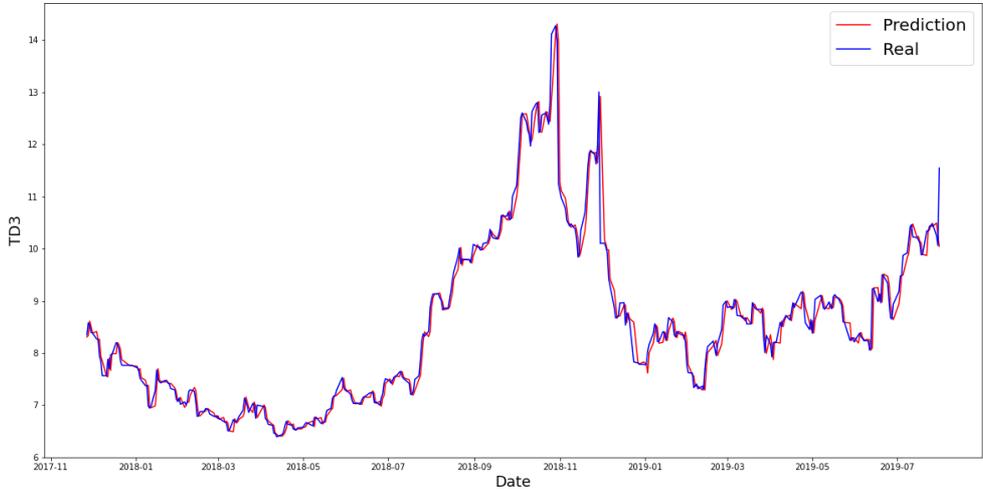


Figure 10 LSTM TD3

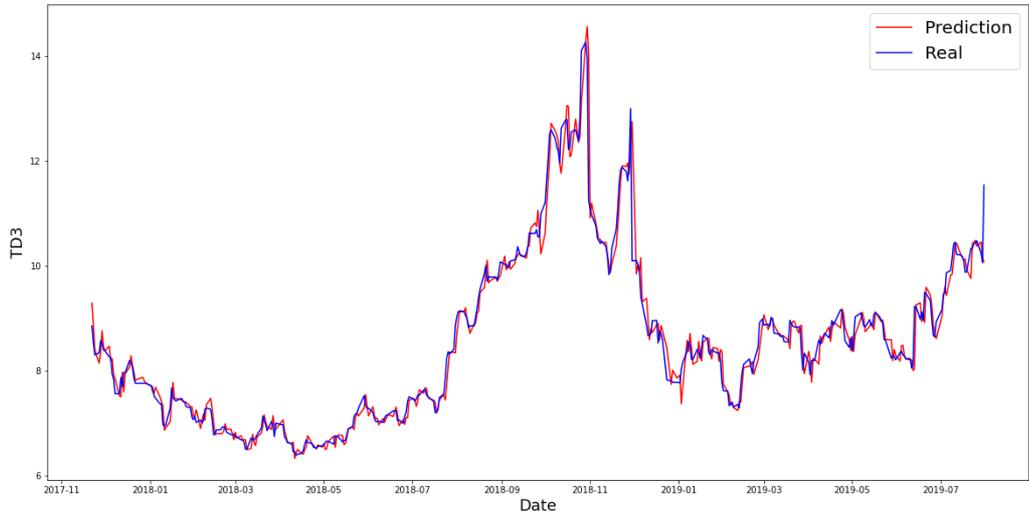


Figure 11 VAR-TD3

5.2 Trading Results

5.2.1 Individual Routes Forecasting Results

The performance metrics in Table 7 displays an overview of the trading strategies implemented on the forecasted individual tanker routes using the LSTM model and the VAR model.

Table 7: Trading Performance Individual Routes

TRADING PERFORMANCE					
INDIVIDUAL ROUTES	Buy and Hold	Simple Long Short		Bollinger Bands	
		LSTM	VAR	LSTM	VAR
A _n	A _n	A _n	A _n	A _n	A _n
TD_{3C}					
Overall Return	43,90 %	80,18 %	27,92 %	-10,07 %	17,86 %
Ann. Return	19,55 %	53,57 %	12,32 %	-9,89 %	4,15 %
Ann. St.Dev	38,79 %	21,80 %	30,52 %	30,36 %	35,99 %
Ann. Sharpe	0,44	2,34	0,32	-0,41	0,04
TD₂₀					
Overall Return	20,166 %	70,47 %	31,54 %	-37,19 %	-40,98 %
Ann. Return	6,881 %	43,30 %	15,89 %	-22,57 %	-23,41 %
Ann. St.Dev	32,013 %	29,18 %	26,60 %	29,00 %	24,39 %
Ann. Sharpe	0,13	1,39	0,50	-0,87	-1,07

The table shows that the best performing strategy is the Simple Long Short strategy when implemented on the LSTM forecasts of the individual tanker routes. The strategy outperforms the Bollinger Band strategy and the Buy and Hold Strategy in terms of the trading performance measurements. The implementation of the strategy on the TD3C contract results in an annualized return of 53,57%, an annualized standard deviation of 21,80% and an annualized Sharpe Ratio of 2,34. The implementation of the strategy on the TD20 contract achieved slightly lower results, with an annualized return of 43,30%, an annualized standard deviation of 29,18% and an annualized Sharpe Ratio of 1,39. The Simple Long Short strategy implemented on the VAR forecasts of the individual tanker routes also shows positive results, but slightly worse than the LSTM forecasts.

The worst performing strategy is the Bollinger Band strategy, when implemented both on the LSTM forecasts and the VAR forecasts of the individual tanker routes. The strategy obtains negative results when implemented on the LSTM forecast of TD3C, achieving an annualized return of -9,89%, an annualized standard deviation of 30,36% and an annualized Sharp Ratio of -0,41. When implemented on the LSTM forecast of the TD20, the strategy also obtains negative results, with an annualized return of -22,57%, an annualized standard deviation of 29,00%, and an annualized Sharpe Ratio of -0,87. The Bollinger Band strategy in combination with the VAR model, shows even worse results when implemented on TD20, but slightly better and positive results when implemented on TD3C.

In general, the Buy and Hold benchmark Strategy outperforms Bollinger Bands, but does not outperform the Simple Long Short strategy when implemented on the individual tanker routes. The Buy and Hold benchmark obtain annualized returns of,55% and 6,88%, and Sharpe Ratios of 0,437 and 0,13 on TD3C and TD20, respectively.

The graphs below plot the cumulative returns of the trading strategies implemented on the individual tanker routes.

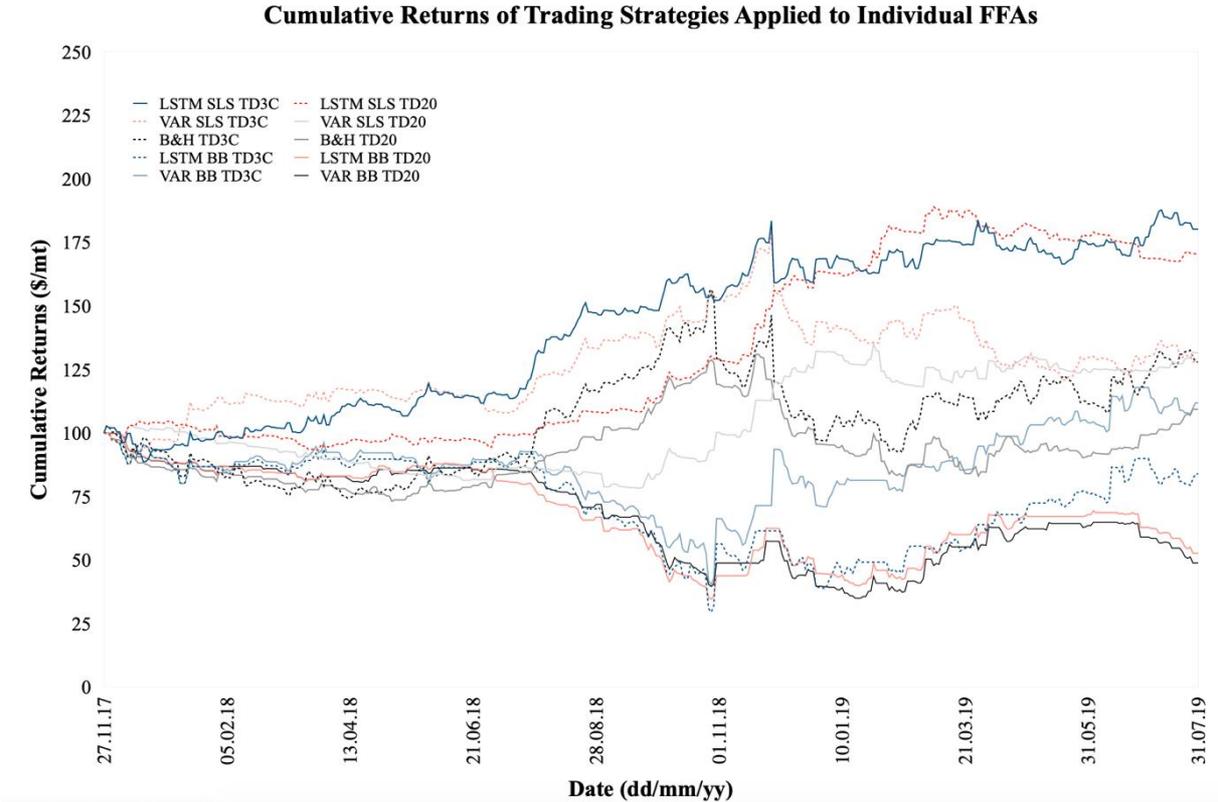


Figure 12: Cumulative Returns of Trading Strategies Applied to Individual FFAs

The graph confirms the results above, and shows that the all curves except three, obtain steady positive cumulative returns. As shown in the curves, the LSTM Simple Long Short beat the buy and hold benchmark, and obtains the highest returns, which indicate that the strategy is working when implemented on the individual tanker routes. Furthermore, the curves, show that the Bollinger Band Strategy obtain steady negative returns, and does not beat the buy & hold

Benchmark, which conversely indicate that the strategy is not working well when implemented on the individual tanker routes.

5.2.2 Spreads Trading Results

The Augmented Dickey Fuller tests suggest that the TD3C-TD20 pair does not fulfil the requirements of cointegration, as the route does not have low short-term correlation and high long-term correlation. However, to be able to make trading comparisons with the forecasted individual routes, the spread strategy is still implemented to the forecasted route combination.

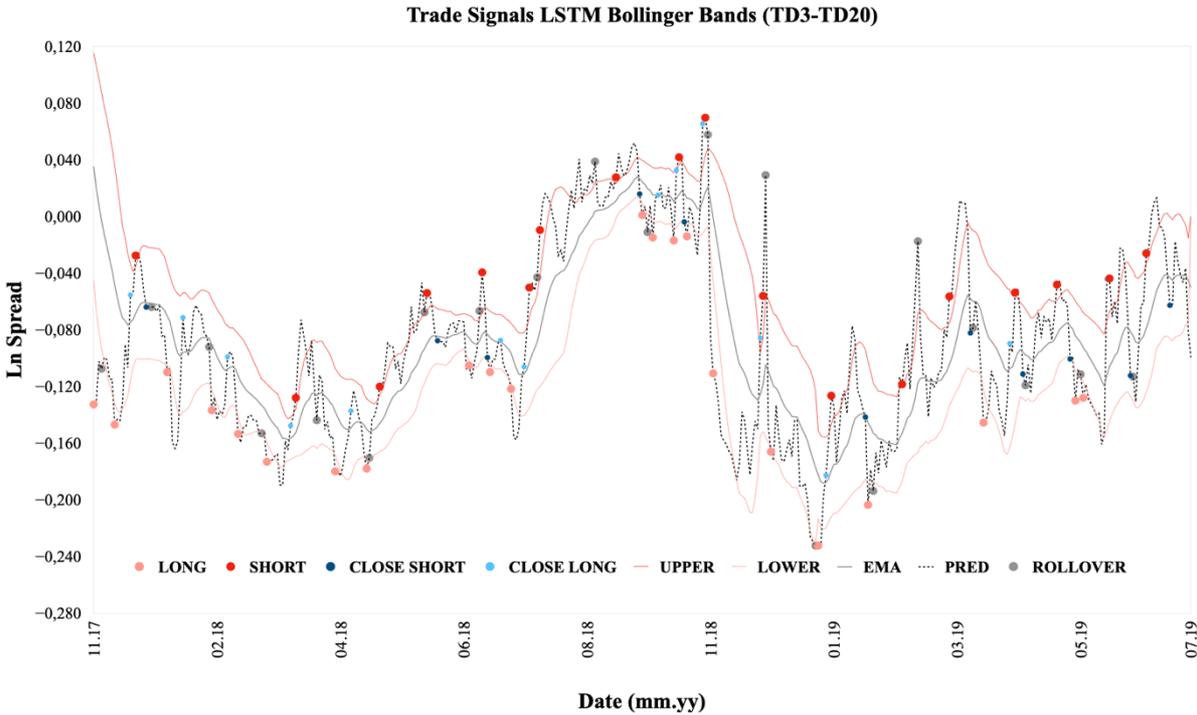
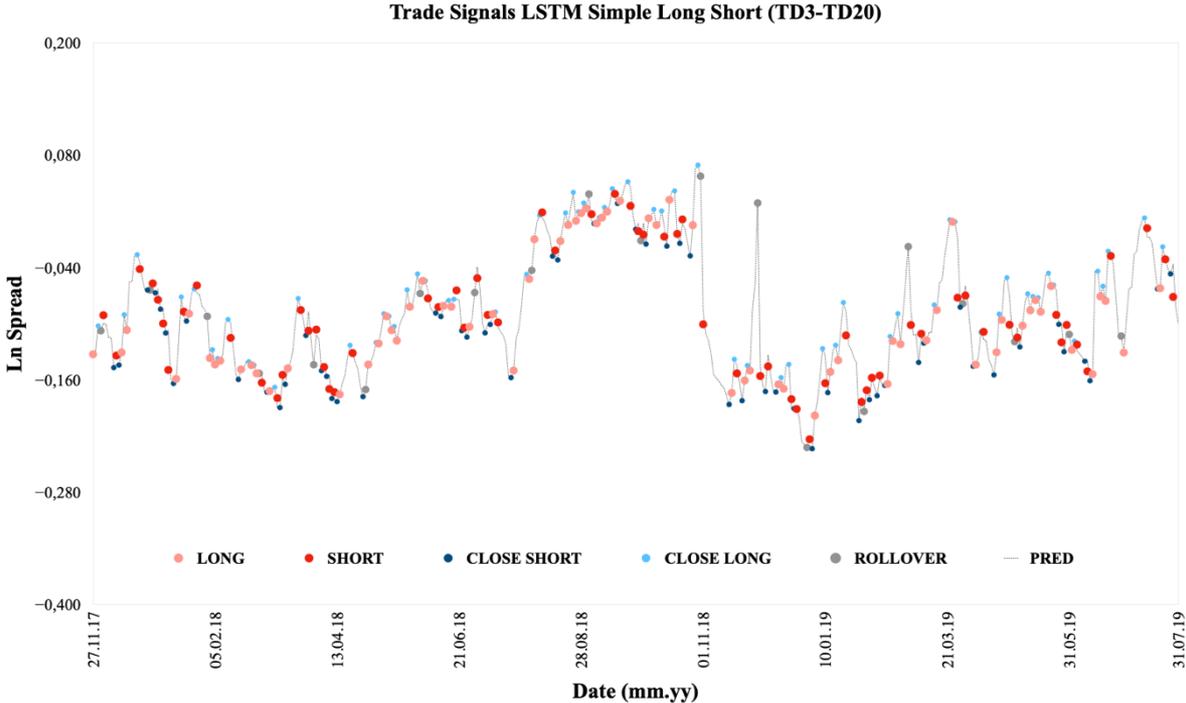


Figure 13: Trade Signals Bollinger Bands on the LSTM predicted FFA spread

The diagram above shows the trading signals activated when the Bollinger Band Strategy was implemented on the LSTM forecasted price spread of TD3C and TD20. The figure shows that the upper and lower bands are expanding and contracting around the EMA, where the red line represents the upper threshold and pink line the lower threshold. The black line represents the

LSTM predicted spread forecast, while the grey line represents the exponential moving average. The red dots show the points at where the short signals are generated at the point where the spread hit the red line (crossed the threshold from below), while the pink dots show the points at where the long signals were generated, when the spread hit the pink line (crossed the threshold from above). The close signals are generated whenever the spread reverted to a mean, where light blue dots represent the closing of long positions (cross the threshold from below) and the dark blue dots represents closing of short signals (crossed the threshold from above). The grey dots show the time of the rollover of the FFA contracts, and indicate a closing of the position, before a new 2nd month contract is entered.

The diagram below shows the trading signals that were activated when implementing the Simple Long Short strategy on the forecasted price spread. The signal dots are representative in this model as well, in the way they are described above However, the black line shows the predicted LSTM forecast, while the grey line represents the real value of the price spread.



To compare the trading signal models, the diagrams show that there are generally more trading signals generated from the Simple Long Short strategy than the Bollinger Band strategy, as the Bollinger Bands requires the thresholds to be crossed before any signals are generated. The fact that the Simple Long Short strategy generate more signals, require the trader to more often go in and out of positions, which results in slightly higher transaction costs. Furthermore, one can

argue that trading in pairs also results in higher transaction costs compared to trading in individual routes, as two positions are obtained simultaneously, instead of one. However, the transaction costs are considered to be quite low when the volume traded is high.

The performance metrics in Table 2 displays an overview of the trading strategies implemented on the directly forecasted price spreads of tanker routes using the LSTM model and the VAR model.

Table 8: Trading Performance Route Pair

TRADING PERFORMANCE				
PAIRS OF ROUTES	Simple Long Short		Bollinger Bands	
	LSTM	VAR	LSTM	VAR
$A_n - B_n$	$A_n - B_n$	$A_n - B_n$	$A_n - B_n$	$A_n - B_n$
TD_{3C} - TD₂₀				
Overall Return	-18,19 %	-2,37 %	123,82 %	122,06 %
Ann. Return	-12,72 %	-2,89 %	92,67 %	90,10 %
Ann. St.Dev	25,17 %	18,70 %	27,66 %	29,48 %
Ann. Sharpe	-0,61	-0,29	3,26	2,97

The table shows that the Simple Long Short model gained negative results when implemented on the directly forecasted FFA spreads, which is true for both the LSTM model and the VAR model. The losses obtained when using the LSTM and Var model, in terms of performance metrics, are annualized returns of -12,72% and -2,89%, annualized standard deviations of 25,17% and 18,70%, and Sharpe Ratios of -0,61 and -0,29, respectively.

The Bollinger Bands strategy obtained higher and positive trading results, both using the LSTM and the VAR model. The LSTM model achieved an annualized return of 92,67%, an annualized standard deviation of 27,66% and a Sharpe Ratio of 3,26, while the VAR model achieved an annualized return of 90,10%, an annualized standard deviation of 29,48% and a Sharpe Ratio 2,97. The graphs below plot the cumulative returns of the spread trading strategies.

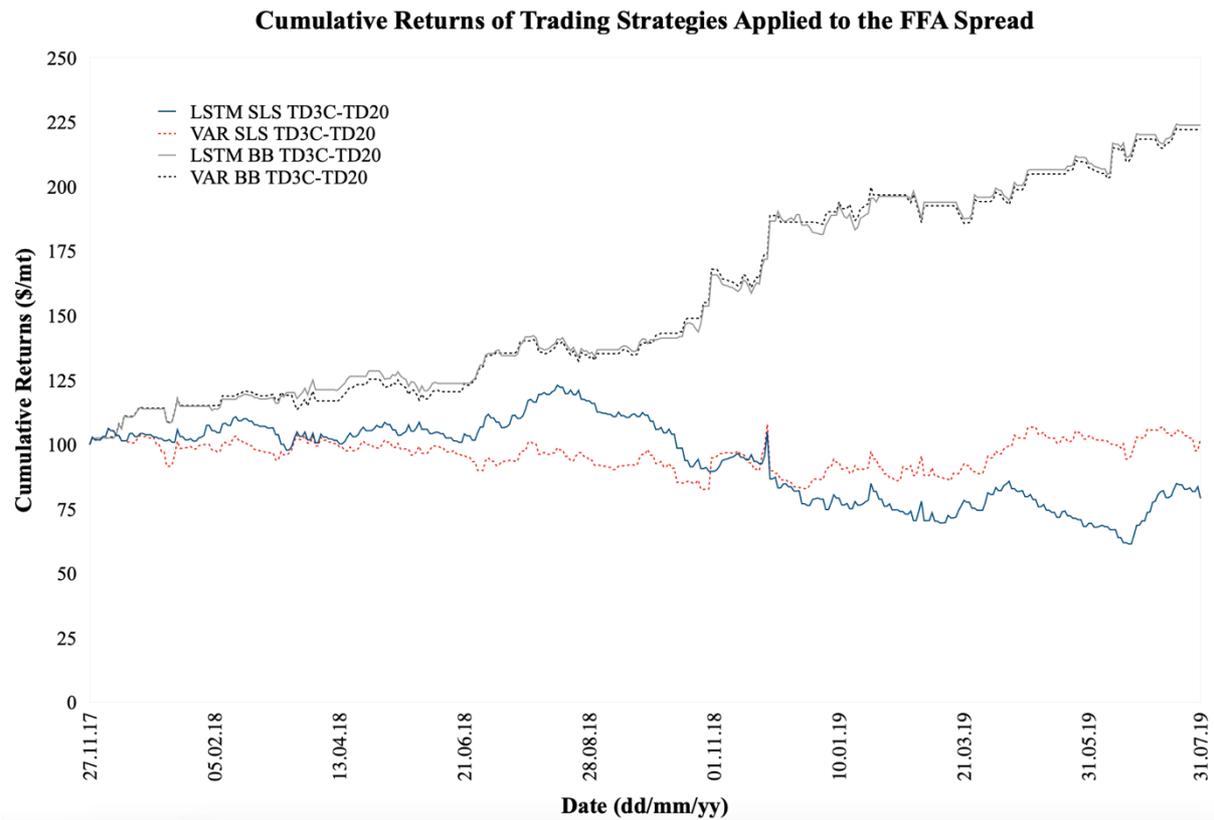


Figure 15: Cumulative Returns of Trading Strategies Applied to the FFA Spread

The diagram shows that the Simple Long Short strategy when implemented on the VAR and the LSTM spread models most of the time obtain steady negative returns. Conversely, the Bollinger Band strategy clearly outperform the Simple Long Short Strategy and earns a steady positive cumulative return when implemented on the forecasted spreads. This graph shows that the Bollinger Band strategy implemented on both the VAR model and the LSTM model almost achieve the same cumulative return pattern, which can be a result of almost similar EMA values obtained from both models.

The figure below displays the cumulative return of all strategies when implemented on the LSTM model, which means that in this graph the Buy and Hold benchmarks and the strategies implemented on the VAR model are excluded.

Cumulative Returns of Trading Strategies - Only the LSTM Model

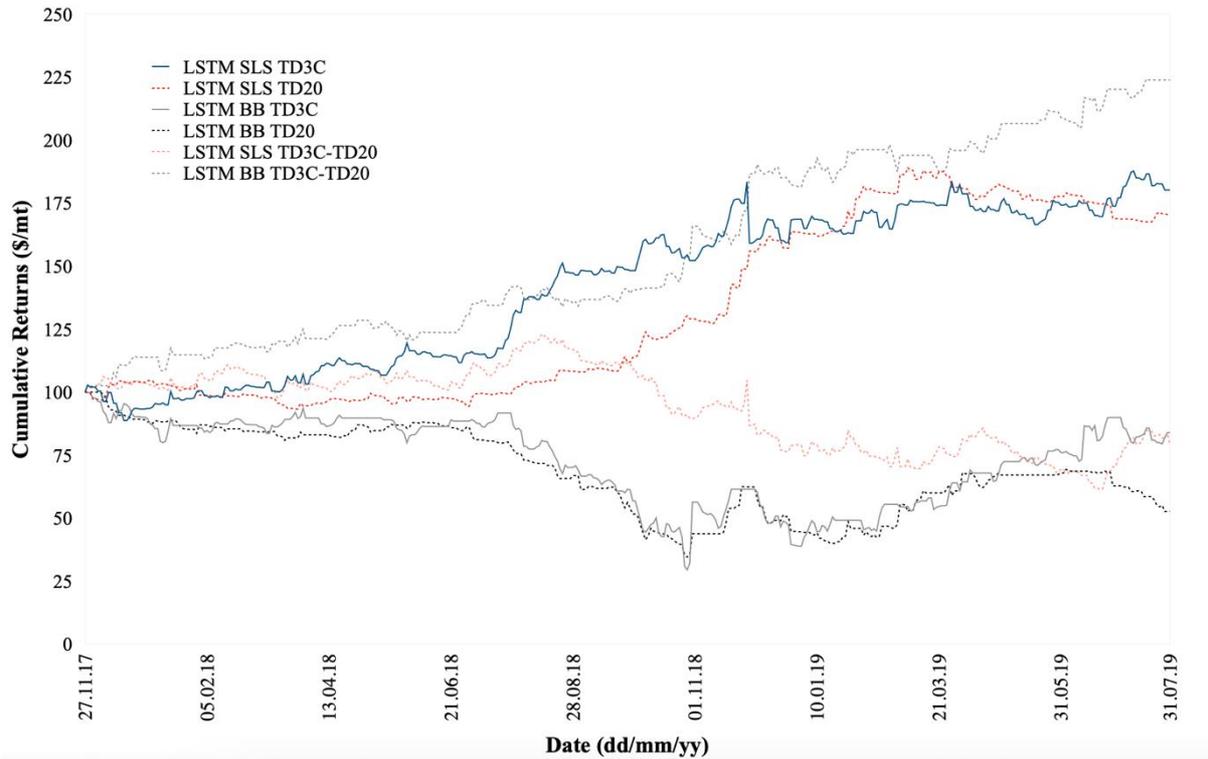


Figure 16: Cumulative Returns of Trading Strategies - Only LSTM Model

The graph shows that the spread the Bollinger Bands strategy performed best and obtained positive results when trading in the forecasted route pair, while the Simple Long Short strategy performed best and obtained positive results when trading in the individual routes. Conversely, the graph shows Simple Long Short Strategy performed worse and obtained negative results when implemented on the route pair, while the Bollinger Band strategy performed worse and obtained negative results when implemented on the individual routes. The overall best performing strategy was the Bollinger Bands on LSTM route pair of TD3C-TD20, while the secondly best performing strategy was Simple Long Short on the individual route of TD3C. To summarize the comparison of the trading strategies, it turns out that the best performing strategy for the individual routes does not necessarily works well when implemented to the route pairs, and the strategy that performs well for the route pairs, does not necessarily work well when implemented don the individual routes. Hence, which strategy is considered the best, depends on where the investor wants to take one position in individual routes, or two simultaneous positions in the route pairs.

6 Discussion

6.1 Discussion of Forecasting Results

Among the AIS-features presented in this thesis, features regarding speed, the local capacity utilization, as well as the capacity allocation of the VLCC and Suezmax fleet seem to be the most prominent in explaining the one-day-ahead price of the FFA contracts. As for the non-AIS variables, the exchange rates, BDTI, TCE and oil spreads appears to overall hold the most predictive power.

The feature selection process was performed using a combination of embedded and filter methods, where the results for the mean score of all methods constituted the final subset. One could argue that a wider range of methods when selecting the final subset could have been used, and it is possible that a different combination of features for each respective FFA contract price could net better result for the forecast.

Furthermore, the features derived in this thesis are prone to errors as they heavily depend on the quality of the data. Missing and error-signals from the AIS-data are examples of possible errors. There is also a possibility that both VAR and LSTM models could have performed better if features were based on more accurate geofencing, as many of the features are based on world regions and route-specific areas of interest. Additionally, how the daily observations are presented to the models might also influence the performance, as the model might not be able to interpret the daily fluctuations presented to the model.

Regarding the forecasting results, the LSTM outperformed both benchmarks on the individual FFA prices for all metrics included, while VAR outperformed LSTM and RW slightly when forecasting the TD3-TD20 spread. Overall, the VAR forecasts appears to be more volatile than the LSTM forecasts.

Even if the results of the LSTM model return better forecasting performance compared to Random Walk and VAR for two of the FFA contracts, the model seems to have a tendency of following the development from the previous day when predicting the next time step. There could be several reasons for this happening. Since the period of this study spans over 4 years, with only 60% being data used for training, it is possible that the model is not able to learn the time-lagged dependencies that exist in the shipping market (Stopford, 2009). As a result,

this can lead to occurrences where the input from the previous day acts as the best predictor for the next day.

Another possibility is the tuning the hyperparameters of the model. In terms of hyperparameter optimization, a more structured and extensive approach could have been chosen. Grid search hyperparameter optimization is useful when checking within certain parameter boundaries but lack depth. Using a random based grid search would require substantially more time and computational power, but would likely net better results, as a larger dimension of parameter combinations would be investigated.

Further work could be forecasting Crude tanker FFA prices using AIS-data by utilizing directional forecasts as means of obtaining a better result. As quality of AIS data is in constant improvement, more useful data will become available, and implementing the same strategy on more data could possibly improve the results.

6.2 Discussion of Trading Results

As shown in some of the results above, taking two positions can reduce the price-risk and volatility exposure, as the diversification increases, compared to only taking one position at the time. Although there are benefits, there are also disadvantages to be considered. Implementing the spread trading models to the FFA pair, some of the profits calculations result in losses, which was the case with the Simple Long Short strategy on TD3C-TD20. It turns out that when large profits are obtained selling one contract, the profits can be offset by taking the opposite position in the other contract, which can result in even higher losses. On the other hand, when the positions are closed, the gains sometimes turn out to be larger than the losses obtained, as was the case with the Bollinger Band strategy in the route pair. Hence it turns out that the strategies are not only dependent on the co-movements and correlations between the FFA contract prices, but also the strategy implemented as well as the time horizon investigated.

Another concern regarding the trading strategies, whether they are implemented on spreads or individual routes, is the liquidity of the FFA market. Some routes are generally more liquid than others, and as mentioned, TD3C and TD20 were considered as the most liquid routes concerning variability in the data. For trading and forecasting purposes, one can argue that a weakness of this study is that the route combinations and individual routes investigated are

limited. In further work, and as the liquidity in the market increase, one can investigate other routes of the BDTI. The trading and forecasting opportunities can further be increased by applying spread trading strategies across different markets, like the clean tanker market or the dry bulk market. Another possible solution is to trade on the different maturities available for the contracts by applying time spreads. One could also implement spread trading strategies on time charter equivalents and baskets of routes. These approaches were not prioritized due to time limitations but can be objects for further investigation. In further work it would also be interesting to implement other or more complicated technical trading rules than investigated in this thesis.

Bid-ask spreads are not considered in the profitability calculations. The bid-ask spread could further reduce the performance and the profitability of the strategies implemented. In the presented results, it is assumed that it is possible to entry and exit the positions before price moves occur. Concerning liquidity-related issues and bid-ask spreads, the actual performance results obtained in a real market could be more conservative or worse compared to the results presented in this thesis. Furthermore, transaction costs higher than the ones used in the models presented above, commissions and counter-party risk should be considered when trading in the real OTC FFA market. Also, the time consumed when implementing technical rules and spread trading strategies has to be taken into considerations when trading these derivatives.

7 Conclusion

The objective of this thesis has been to forecast the one-day-ahead prices of selected crude tanker FFAs using machine learning techniques and AIS-derived data in order to investigate the profitability of applying selected trading strategies on these forecasts. Specifically, this thesis has looked at the two, assumed to be, most liquid crude tanker routes (TD3C and TD20) including the route-pair TD3C-TD20 and forecasted the FFA prices. A complex Long Short-Term memory NN model is utilized, and performance is evaluated up against two benchmarks, a RW and VAR model as benchmark to measure. Further, the trading profitability is investigated by implementing a Simple Long Short, Bollinger Bands and Buy-and Hold strategy.

This thesis finds that the LSTM performs better than the RW and VAR model when predicting the one-day-ahead price for the individual FFA contracts of TD3C and TD20 but falls short to

both benchmarks when predicting the TD3C-TD20 route-pair. Even though the forecasting results has a promising outlook, we acknowledge that the study holds some weaknesses and limitations, especially regarding the predictions. The LSTM model seems to have trouble understanding the underlying mechanisms of the features presented, resulting in suboptimal predictions. A limitation of the study is the missing values in the AIS data, which could be contributing negatively when deriving the features.

When the trading strategies are implemented on the individual tanker routes and the route pair, it turns out that the Bollinger Band strategy performs best when implemented on the forecasted LSTM model of the TD3C-TD20 route pair, while the Simple Long Short Strategy performs best when implemented on the forecasted LSTM models of the individual routes of TD3C and TD20.

References

- Adland, R. O., Jia, H., & Strandenes, S. P. (2018, April). The determinants of vessel capacity utilization: The case of Brazilian iron ore exports. *Transportation Research Part A: Policy and Practice*(110), pp. 191-201.
- Adland, R., Ameln, H., & Børnes, E. A. (2020). Hedging ship price risk using freight derivatives in the drybulk market. *Journal of Shipping and Trade*(5).

- Alizadeh, A., & Nomikos, N. (2009). *Shipping Derivatives and Risk Management*. Palgrave Macmillan.
- Assman, L. M. (2020). *Forecasting, Efficiency and Speed in Maritime Shipping*. PhD, Norwegian School of Economics, Bergen.
- Baltic Exchange. (2020). Retrieved from <https://www.balticexchange.com/en/index.html>
- Cheng, L., Yan, Z., Xiao, Y., Chen, Y., Zhang, F., & Li, M. (2019). Using big data to track marine oil transportation along the 21st-century Maritime Silk Road. *Science China Technological Sciences*, 62, pp. 677-686.
- Corporate Finance Institute. (2020). Retrieved from <https://corporatefinanceinstitute.com/>
- Ducruet, C. (2019). *Advances in Shipping Data Analysis and Modeling*. Paris: Routledge.
- Elliott, R. J., Van Der Hoek, J., & Malcolm, W. P. (n.d.). Pairs trading. *Quantitative Finance*, 3(5), pp. 271-276.
- Gatev, E. G., Goetzmann, W. N., & Rouwenhorst, K. (1999). *Pairs Trading: Performance of a Relative Value Arbitrage Rule*. (Working Paper No. 7032), Yale School of Management.
- Geman, H., Chang, L., & Liu, B. (2016). Intraday pair trading strategies on high frequency data: the case of oil companies. *Quantitative Finance*, 17(1), pp. 87-100.
- Goetzmann, W., Gatev, E. G., & Rouwenhorst, K. G. (2006, February). Pairs Trading: Performance of a Relative Value Arbitrage Rule. *Review of Financial Studies*, 19(3), pp. 797-827.
- Haugland, D. V., & Abelvik-Engmark, A. (2018). *Spread Trading in Brent Crude Futures - A stochastic approach to intraday calendar spread trading*. (Master Thesis), Norwegian University of Science and Technology.
- Hochreiter, S., Bengio, Y., Frasconi, P., & Schmidhuber, J. (2001). Gradient Flow in Recurrent Nets: the Difficulty of Learning Long-Term Dependencies. *A Field Guide to Dynamical Recurrent Neural Networks*.
- Huang, T. (2017). *Python-Ports-Distance-Calculator*. Retrieved from <https://github.com/tsunghao-huang/Python-Ports-Distance-Calculator>
- Hyndman, R. J., & Athanasopoulos, G. (n.d.). *Forecasting: Principles and Practice*. Melbourne, Australia: OTexts. Retrieved from <https://otexts.com/fpp2/index.html>
- ICE. (2020). Retrieved from <https://www.theice.com/index>
- Indices, VIX. (2020). Retrieved from Cboe: <https://www.cboe.com/us/indices/dashboard/ovx/>
- James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). *An Introduction to Statistical Learning*. Springer Publishing Company.

- Jia, H., Prakash, V., & Smith, T. (2015). Estimating vessel utilization in the drybulk freight market: the reliability of draught reports in AIS data feeds. *Paper presented at the ECONSHIP conference*. Chois.
- Kaluza, P., Koelzsch, A., Gastner, M. T., & Blasius, B. (2010, July). The complex network of global cargo ship movement. *Journal of The Royal Society Interface* 7(48), pp. 1093-1103.
- Kavoussanos, M., & Nomikos, N. (2003). Price Discovery, Causality and Forecasting in the Freight Futures Market. *Review of Derivatives Research*(6), pp. 203-230.
- Kavoussanos, M. G., & Visvikis, I. (2011). *Theory and Practice of Shipping Freight Derivatives*. Risk Books a Division of Incisive Media Financial Publishing Ltd.
- Kavoussanos, M. G., & Visvikis, I. D. (2016). *The International Handbook of Shipping Finance*. Palgrave Macmillan.
- Kavoussanos, M. G., Viskvikis, I. D., & Menachof, D. (2004). The Unbiasedness Hypothesis in the Freight Forward Market: Evidence from Cointegration Tests. *Review of Derivatives Research*, pp. 241-266.
- Kavoussanos, M. G., Visvikis, I. D., & Dimitrakopolous, D. N. (2014, August). Economic spillovers between related derivatives markets: The case of commodity and freight markets. *Transportation Research Part E: Logistics and Transportation Review*(68), pp. 79-102.
- Kavoussanos, M., & Visvikis, I. (2006). *Derivatives and Risk Management in Shipping*. UK: Witherbys Publishing & Seamanship International.
- Kavoussanos, M., Visvikis, I., & Dimitrakopolous, D. (2010). Information linkages between Panamax freight derivatives and commodity derivatives markets. *Maritime Economics & Logistics*(12), pp. 91-110.
- Navigation Center. (2020). *AUTOMATIC IDENTIFICATION SYSTEM OVERVIEW*. Retrieved from U.S Homeland Security: <https://www.navcen.uscg.gov/?pageName=aismain>
- Næss, P. A. (2018). *Investigation of Multivariate Freight Rate (Master Thesis)*. Norwegian University of Science and Technology.
- Nomikos, N. K., & Alizadeh, A. H. (2006). Trading strategies in the market for tankers. *Maritime Policy & Management*(33), pp. 119-140.
- Nomikos, N. K., & Doctor, K. (2013, June). Economic significance of market timing rules in the Forward Freight Agreement markets. *Transportation Research Part E: Logistics and Transportation Review*(52), pp. 77-93.

- Olah, C. (2015, August 27). *Understanding LSTM Networks*. Retrieved September 3, 2020, from <https://colah.github.io/posts/2015-08-Understanding-LSTMs/>
- PHLX Oil Service Sector (OSX)*. (2020). Retrieved from Nasdaq Global Indexes: <https://indexes.nasdaqomx.com/Index/Overview/OSX>
- Poulakidas, A., & Joutz, F. (2008, September). Exploring the link between oil prices and tanker rates. *Maritime Policy & Management*, 36(3).
- Prochazka, V., Adland, R., & Wolff, F. C. (2019, December). Contracting decisions in the crude oil transportation market: Evidence from fixtures matched with AIS data. *Transportation Research Part A: Policy and Practice*(130), pp. 37-53.
- Regli, F., & Nomikos, N. K. (2019, May). The eye in the sky – Freight rate effects of tanker supply. *Transportation Research Part E: Logistics and Transportation Review*(125), pp. 402-424.
- Sanchez-Marono, N., Alonso-Betanzos, A., & Tombilla-Sanromán, M. (2020). Filter methods for Feature Selection, A Comparative Study. *Intelligent Data Engineering and Automated Learning - IDEAL 2007*. Berlin: IDEAL 2007.
- Sherstinsky, A. (2020). Fundamentals of Recurrent Neural Network (RNN) and Long Short-Term Memory (LSTM) network. *Physica D: Nonlinear Phenomena*(404).
- Singh, S. K., & Heymann, F. (2020, July). On the Effectiveness of AI-Assisted Anomaly Detection Methods in Maritime Navigation. *2020 IEEE 23rd International Conference on Information Fusion (FUSION)* , pp. 1-7.
- Skauen, A. N. (2015). Quantifying the tracking capability of space-based AIS systems. *Advances in Space Research*, 57, pp. 527-542.
- Stopford, M. (2009). *Maritime Economics*. Routledge.
- Tsiomas, V. (2016). *Quantitative analysis of the dry bulk freight market*. (PhD. thesis), Department of Maritime Studies University of Piraeus.
- Vafaeipour, M., Rahbari, O., Rosen, M. A., Fazelpour, F., & Ansarirad, P. (2014). Application of sliding window technique for prediction of wind velocity time series. *International Journal of Energy and Environmental Engineering*, 5(105).
- Westgaard, S., Stenslet, D., Ringheim, J., & Osmundsen, P. (2017). Modeling superior predictors for crude oil prices. *Journal of Energy Markets*, 10(2), pp. 11-22.
- Yang, L., & Shami, A. (2020). On Hyperparameter Optimization of Machine Learning Algorithms: Theory and Practice. *Neurocomputing*, 415, pp. 295-316.

Appendix 1.1

Results Dickey Fuller test

TD3C

	ADF-Statistic	P-value	1%	Accept H0 at 1%
BDTI	-13.3454	5.80768e-25	-3.43503	False
In_flowcount_VLCC_IndianO	-12.7612	8.13663e-24	-3.43509	False
TCE_VLCC	-8.51221	1.15926e-13	-3.43509	False
USD_SAR_exchange	-9.50299	3.40565e-16	-3.43509	False
Out_flowcount_VLCC_NWE	-13.4334	3.97356e-25	-3.43509	False
LF_std_ArabGulf_VLCC_moving	-12.0882	2.15299e-22	-3.43509	False
CNY_USD_exchange	-37.5949	0	-3.43501	False
S&P500	-19.3427	0	-3.43502	False
Count_Asia_VLCC_moving	-13.7139	1.22292e-25	-3.43507	False
Speed_mean_Asia_VLCC_moving	-14.0193	3.59582e-26	-3.43507	False
Dubai_Crude_1vs2M	-18.0396	2.66029e-30	-3.43503	False
VLCC_Cap_AG	-16.181	4.26704e-29	-3.43506	False
In_flowcount_VLCC_AG	-13.1501	1.37265e-24	-3.43509	False
TD3C	-32.3502	0	-3.43501	False

TD20

	ADF-Statistic	P-value	1%	Accept H0 at 1%
BDTI	-15.5483	2.11616e-28	-3.43557	False
Speed_std_West_Africa_Suezmax_M_allD	-11.8597	6.89407e-22	-3.43565	False
Speed_std_Global_Suezmax_M_Rotter	-13.7304	1.14322e-25	-3.43564	False
LF_mean_Global_Suezmax_moving	-12.1629	1.47996e-22	-3.43563	False
in_flowcount_Suezmax_AG	-12.6241	1.55501e-23	-3.43566	False
3MCVBRNT Index	-8.22718	6.20108e-13	-3.43565	False
NGN_USD_ex	-34.5437	0	-3.43556	False
LF_std_Global_Suezmax_M_Rotterdam	-12.3025	7.39492e-23	-3.43565	False
Suezmax_cap_FE	-14.4473	7.19423e-27	-3.43561	False
out_flowcount_Suezmax_IndianO	-12.6328	1.49179e-23	-3.43566	False
LF_mean_Global_Suezmax_M_Rotterdam	-12.6568	1.33131e-23	-3.43565	False
LF_mean_Global_Suezmax	-11.64	2.15864e-21	-3.43564	False
TD20	-30.6859	0	-3.43556	False

lnTD3C-lnTD20

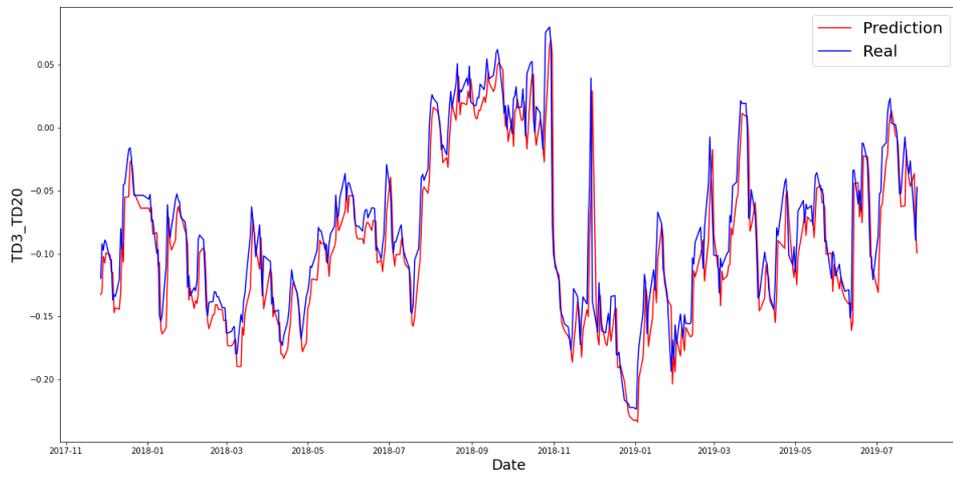
	ADF-Statistic	P-value	1%	Accept H0 at 1%
Count_Asia_VLCC_moving	-13.4339	3.96435e-25	-3.43563	False
LF_std_ArabGulf_VLCC	-14.4564	6.96295e-27	-3.43562	False
Speed_mean_Global_Suezmax_M_Rotterdam	-12.5298	2.44237e-23	-3.43566	False
USD_SAR_exchange	-10.6135	5.75169e-19	-3.43565	False
Speed_mean_Asia_VLCC_moving	-13.4688	3.41427e-25	-3.43563	False
VLCC_cap_IndianO	-12.3264	6.57508e-23	-3.43564	False
LF_std_Global_Suezmax_M_Rotterdam	-12.3025	7.39492e-23	-3.43565	False
VLCC_cap_NWE	-13.7171	1.20708e-25	-3.43563	False
LF_mean_Global_Suezmax_M_Rotterdam	-12.6568	1.33131e-23	-3.43565	False
Crude Oil(Brent)	-37.9353	0	-3.43556	False
TCE_VLCC	-14.2026	1.77705e-26	-3.43558	False
LF_std_Global_Suezmax_moving	-13.6264	1.75651e-25	-3.43562	False
In_flowcount_VLCC_AG	-12.6241	1.55501e-23	-3.43566	False
% Ballast_ArabGulf_Suezmax_moving	-12.9442	3.48309e-24	-3.43565	False
% Ballast_Asia_VLCC	-12.2771	8.38523e-23	-3.43566	False
lnTD3_InTD20	-14.9068	1.47869e-27	-3.43559	False

Appendix 1.2
Results feature importance
Top 30 features (including lag) for TD3C

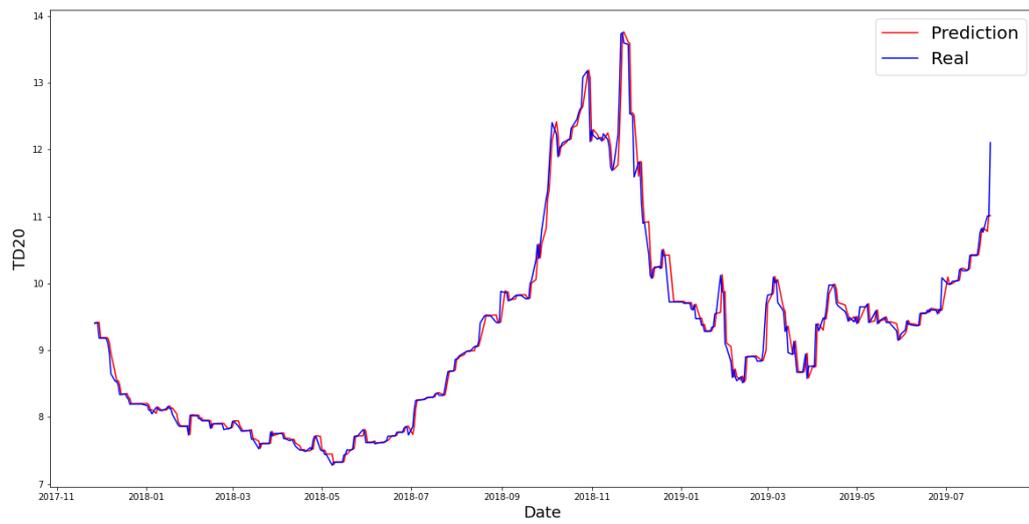
Feature (lag)	Multi Reg.	Lasso Reg.	RF	Linear Corr.	XGB	Mean Score
BDTI (t)	0.83	0	0.23	1	0.03	0.42
In_flowcount_VLCC_IndianO (t-15)	0.16	0	1	0.03	0.56	0.35
TCE_VLCC (t-1)	0.1	0	0.52	0.96	0.1	0.34
USD_SAR_exchange(t-15)	0.09	0	0.32	0.81	0.31	0.31
Out_flowcount_VLCC_NWE (t-7)	0.09	1	0.02	0.44	0	0.31
CNY_USD_exchange(t-11)	0.12	0	0.26	0.4	0.48	0.25
LF_std_ArabGulf_VLCC_moving (t-1)	0.2	0	0.03	0	1	0.25
In_flowcount_VLCC_IndianO (t-4)	0.2	0	0.27	0.38	0.36	0.24
S&P500 (t-5)	0.14	0	0.11	0.79	0.17	0.24
Count_Asia_VLCC_moving(t-15)	0.21	0	0.23	0.48	0.18	0.22
USD_SAR_exchange (t-2)	0.28	0	0.39	0.39	0.13	0.21
Speed_mean_Asia_VLCC_moving(t-9)	0.26	0	0.05	0.22	0.53	0.21
S&P500 (t-1)	0.24	0	0.69	0.01	0.09	0.21
Dubai_Crude_1vs2M (t-8)	0.75	0	0.04	0.21	0.04	0.21
VLCC_Capacity_AG (t-8)	0.54	0	0.07	0.27	0.18	0.21
Count_Asia_VLCC_moving (t-6)	0.39	0	0.01	0.44	0.14	0.20
In_flowcount_VLCC_AG (t-9)	0.09	0	0.17	0.39	0.34	0.20
Out_flowcount_VLCC_AG (t-15)	0.39	0	0.07	0	0.45	0.18
LF_mean_Asia_VLCC (t-7)	0.41	0	0.01	0.51	0	0.19
STI(t)	0.18	0	0.35	0.26	0.14	0.19
LIBOR_3M (t-6)	0.46	0	0.14	0.2	0.01	0.16
VLCC_Capacity_IndianO (t-8)	0.57	0	0.02	0.19	0	0.16
TMD_AllVessels_Global_cumsum (t-12)	0.37	0	0.03	0.08	0.21	0.14
% Ballast_Asia_VLCC_moving (t-9)	0.54	0	0.07	0.1	0	0.14
% Ballast_IndianO_VLCC (t)	0.22	0	0.01	0.37	0	0.12
TMD_AllVessels_Global_cumsum (t-8)	0.32	0	0.03	0.2	0	0.11
% Ballast_Asia_VLCC (t-7)	0.19	0	0.01	0.37	0	0.11
% Ballast_Global_VLCC (t-12)	0.51	0	0.01	0.01	0.01	0.11
SSE_index(t-8)	0.46	0	0.02	0.03	0	0.10
% Ballast_ArabG_VLCC(t-13)	0.37	0	0.01	0.06	0.02	0.09

Appendix 1.3

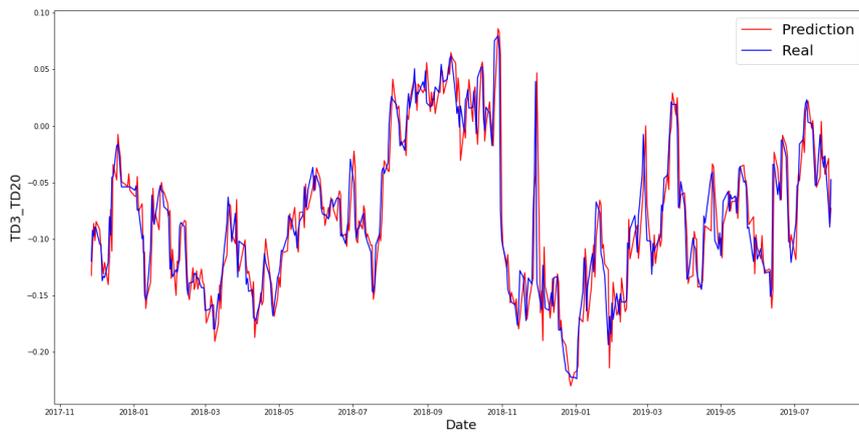
TD3C_TD20 - LSTM



TD20 - LSTM



TD3C-TD20 – VAR



TD2_VAR

