NHH

Scanning the Horizon

Forecasting and trading on Forward Freight Agreements using Long Short-Term Memory Neural Networks and AIS-derived features

Herman Farbrot and Sindre Kalvik

Supervisor: Roar Os Ådland

Master thesis

MSc in Economics and Business Administration, Business Analytics and Finance

NORWEGIAN SCHOOL OF ECONOMICS

Acknowledgements

Abstract

The purpose of this study has been to predict Forward Freight Agreement (FFA) prices using machine learning techniques, investigate the additional forecasting power of Automatic Identification System (AIS) derived features, and to evaluate the profitability of applying forecasted directional movements to trading strategies.

A Long-Short-Term Memory (LSTM) neural network is used to predict price movements for the two closest quarterly, and the closest calendar year Capesize 5 Time Charter (5TC) FFAs. We have derived features from AIS data to generate proxies for supply, demand and geographical distribution for a subset of Capesize vessels. Additionally, we have included commodity prices and macroeconomic variables. The forecasting horizon investigated has been one week, two weeks, and one month ahead. To benchmark the LSTM model, we have included Vector Autoregressive (VAR) models, Autoregressive Integrated Moving Average (ARIMA) models and a Random Walk.

The VAR models were found to be superior at forecasting FFA prices, and the results showed that the LSTM neural network and VAR show potential for predicting directional movements of prices. The results further indicate that AIS data holds predictive capabilities regarding directional movements of prices. Lastly, the trading results give implications of increased profitability compared to buy-and-hold and trend-following benchmarks, by utilizing the trading signals from the models.

Contents

1   Introduction

Historically, the dry bulk shipping market has been extremely volatile. The large freight rate fluctuations create opportunities for generating substantial returns. On the other hand, these large fluctuations are a source of great risk for the operators. Foresight of the future developments in rates is therefore invaluable for speculators and operators, where derivatives such as Forward Freight Agreements (FFAs) may be used for both reducing and increasing freight rate exposure.

FFAs are financially settled forward contracts, where the underlying asset is one of the freight rate indices published by The Baltic Exchange. The settling price for an FFA contract is calculated using the average spot price over the maturity period for the underlying indices. The contracts trade in a competitive over-the-counter market and go through clearinghouses (Baltic Exchange, 2019). The clearing of contracts eliminates the default risk, but price risk and cash flow risk is still inherent, due to movements in the underlying rates and margin requirements. The use of brokers as intermediaries affects the execution speed of trades, as well as liquidity and transaction costs. In combination with relatively sizeable minimum trading requirements, it is difficult for individual investors to access the market (Wilson, 2013).

FFAs were initially intended as a risk management tool for operators in the market, but there has been substantial speculative interest from investment banks and hedge funds (Zheng & Chen, 2018). Shipowners have a long exposure to physical freight, and benefit from increased rates, while charterers have a short exposure, as they are obliged to pay for freight services. These parties may take opposite positions in FFA contracts to hedge against unfavorable price movements to stabilize earnings and cash flow. Precise forecasts of FFA prices may be particularly helpful regarding contractual decisions and development of hedging strategies. For speculative third parties, information about future FFA prices may be used to form effective trading strategies, without participating in the underlying market.

Nomikos and Doctor (2013) applied technical trading rules for the FFA market, where the superior trading strategies generated substantial excess returns compared to a buy-and-hold benchmark. Their results gave implications against a weak form of market efficiency in the FFA market. The efficient market hypothesis states that information should be reflected in the prices to a degree, where the additional gains to be made by acting on available information do not exceed marginal costs (Fama, 1991). However, we want to take a different approach than Nomikos and Doctor, as we aim to include relevant market features in addition to

geospatial data to support our predictions of movements. The utilization of Automatic Identification System (AIS) data has in recent years emerged as a new way of supplementing research on maritime economics. The vast amount of available AIS data and possibilities for the extraction of features offers a wide variety of new opportunities within shipping analytics. AIS data, in combination with programming tools, enables calculation of several metrics for selected vessels across time and space, in addition to the tracking of vessels and fleets in real-time. The increased availability of computing power and applications for machine learning techniques has, simultaneously, created new possibilities for analyzing complex data sets.

The objective of this paper is to forecast FFA prices using machine learning techniques. Our contributions to the literature are threefold: Firstly, we will apply and evaluate the performance of an LSTM neural network to predict movements of Capesize 5 Time Charter (5TC) FFA prices. Secondly, we will create new AIS-derived features and evaluate their predictive powers on FFAs. Thirdly, we will evaluate the forecasting models' ability to generate profitable trading signals, by utilizing them in simple trading strategies. The first quarterly contract (1Q), the second closest quarterly contract (2Q), and the closest calendar year contract (1CAL) will be used for forecasting and trading, where the forecasting horizon will be one week, two weeks, and one month ahead.

The remainder of this paper is structured as follows: firstly, we present a literature review that serves as a foundation for the work conducted in this paper, which culminates in our contributions to the current body of literature. Secondly, the data used will be presented. Thirdly, we go through the process of creating features. Fourthly, the methodology for feature selection and machine learning will be presented. Lastly, we present the results, before rounding off with some concluding remarks and recommendations for further work.

2   Literature Review

The literature review for this is paper is extensive, as the paper revolves around several topics. First, relevant FFA market research is covered, followed by research on applications of AIS, before introducing research covering the predictive capabilities of machine learning techniques in relation to shipping, commodity and financial markets.

There is a large body of literature covering the relationships between spot and FFA prices, forecasting, and hedging performance. Kavussanos et al. (2004) investigated the impact of FFA trading on spot market volatility, where they found that FFAs have a reducing effect on the spot freight rate volatility. Kavussanos et al. (2004b) studied the unbiasedness of FFA prices, where they found FFA prices one and two months before maturity to be unbiased predictors of the spot prices for Panamax routes. Bessler et al. (2008) also found evidence of a cointegrated relationship between spot and forward rates for Panamax bulk carriers. Zhang et al (2014) studied the relation between spot and Time Charter (TC) rates, as well as spot and FFA rates. Their results gave evidence of cointegration between spot and FFA rates, and for TC rates and FFA rates. Adland and Alizadeh (2018) studied TC rates and FFA prices, and also found evidence of cointegration, but that TC rates overall are priced higher than FFAs. A convenience yield, and the additional risk related to physical freight contracts, among other reasons, were pointed out as explanations for the price differences. Kavussanos and Visvikis (2004) found FFA contracts to discover market information faster than spot prices, which was pointed out to originate from lower transaction costs in the forward market, compared to the spot market. Additionally, they found a bi-directional relationship between the FFA and spot markets. Kasimati and Veraros (2017) found that FFAs have limited prediction power for prediction of future freight rates, but that FFA prices were useful for directional predictions. Further, Yin et al. (2017) found mean-reverting tendencies for both FFA and spot prices.

Regarding the hedging performance of FFAs, Alizadeth et al. (2015) found that the hedging performance of FFAs for tankers is worse compared to futures in other commodities and financial markets. One reason for this was identified as the absence of a cost-of-carry relationship between spot and forward prices, due to freight being a non-storable commodity. Alexandridis et al. (2017b) found that freight rate risk can be reduced by 48% by holding a diversified portfolio of freight rates, and that additional risk can be reduced by hedging with forward contracts.

Regarding the topic of forecasting, Batchelor et al. (2007) tested the performance of time series models for predicting spot and forward rates in the dry bulk shipping market, and found that

Autoregressive Integrated Moving Average (ARIMA) models and Vector Autoregressive (VAR) models were superior to Vector Error Correction Models (VECM) for predicting forward rates. Further, the study gave evidence of forward rates providing additional information for spot rates in the future, but that spot rates were unhelpful for predicting forward prices. Lyridis et al. (2004) applied neural networks for forecasting FFA Prices. The main findings were that neural networks performed well at forecasting future prices, but that connectionist models overall held superior predictive performance. Kavussanos et al. (2014a) investigated spillover effects between dry bulk FFAs and commodities, and found agricultural commodity futures to lead freight markets. Kavussanos et al. (2010) further found that spillover effects of return and volatility generally are one-directional from commodity futures to FFAs. Regarding the topic of FFA trading, Nomikos and Doctor (2013) conducted a comprehensive study of quantitative trading strategies for Capesize, Panamax, and Supramax FFAs across different maturities. They applied trend, momentum, and volatility-based strategies, and evaluated these against a buy-and-hold benchmark. The trend-following strategies were superior among the simple strategies based on mean returns and Sharpe ratio, while complex learning strategies provided the highest average outperformance in terms of Sharpe Ratio, compared to the benchmark. Their best active trading strategies generated significant excess returns compared to the buy-and-hold benchmark, which implies inefficiency the FFA market, as prices do not reflect all available information.

Several studies covering AIS utilization revolve around shipping network detection, demand estimations, and trade patterns. Kaluza et al. (2010) studied the trade patterns for the different ship classes. The study interpreted the global movements of cargo as a network with a high level of complexity. Spiliopoulos et al. (2017) present a methodology for converting AIS data to be used effectively for understanding the shipping patterns in relation to global trading patterns. Wu et al. (2017) used AIS data for mapping vessel density and traffic density, to reveal the distribution of ships and traffic. Vessel density was defined as the number of vessels per unit area, and traffic density was defined as the average number of vessels crossing a region per unit area per unit time. Vector and grid-based methods were applied for traffic density calculations, while vessel density calculations were based on geofencing. Geofencing is a method of extracting data, based on geographical boundaries.

Jia et al. (2015) investigated the reliability of reported draught in AIS data for estimating vessel utilization, in the dry bulk freight market. Due to AIS messages lacking info on cargo type and volume, they present different models for estimation of cargo size, mainly based on draught. They found that AIS data alone is insufficient for precisely tracking seaborne trade. Adland et al. (2017) compared the accuracy of AIS-derived trade statistics for the crude oil market to

official customs data. Their results revealed that AIS-derived data for seaborne crude exports align well with official export numbers in aggregate, but that there are several challenges related to the aggregation of micro-level data. Some key challenges pointed out were the usage of pipelines in parts of the supply chain, in addition to countries and regions operating as storage and transshipment hubs. They further state that any maritime research which covers market fundamentals, could draw benefits from AIS-derived tonne-mile demand data, if the cargo is observable and homogenous. This is to a large extent the case for the dry bulk and tanker markets.

Adland (2019) presents a framework for utilizing AIS data for dry bulk market analysis. He presents algorithms for generating data for tonne-mile demand, proxies for operational efficiency, and counting of unemployed ships. He argues that freight rates between regions tend to move synchronized in the long run, but that there may be differences in the short run due to local supply and demand imbalances. Regarding idle ships, he shows an inverse relationship between Capesize earnings and idle ships waiting in open sea. There are drawbacks to the metrics presented due to limitations of the information from AIS. However, he states that enriching the AIS data with information from other sources, such as vessel characteristics, contractual information, and bill of ladings, can lead to better results.

Regli and Nomikos (2019) studied the effect of tanker supply for the TD3 tanker route between Ras-Tanura and Japan. They created a proxy for short-term supply in the voyage charter market, where vessels were classified as available or unavailable based on geographical restrictions, self-reported destination, loading condition, and employment status. They found their AIS-derived supply measures to partially explain freight rate movements, where other more traditional supply measures, such as fleet size, were ineffective. Also, the study gave evidence of a lagged relationship between ballast sailing speeds and short-term freight rate movements.

Machine learning techniques as a prediction tool have been covered extensively for various stock, commodity and shipping markets. Herrera et al. (2019) examined forecasting of long-term prices for crude oil, coal and gas by applying neural networks, Random Forest and hybrid models, which were compared to a Random Walk benchmark. The results showed that Random Forest were superior. Huang and Wu (2018) applied Deep Multiple Kernel Learning for forecasting energy commodity prices. Their model included information from oil, gold, and currency markets, and was found systematically superior for forecasting crude oil prices, compared to traditional neural networks and regression models. Fischer and Krauss (2018) applied LSTM neural networks for predicting directional movements of the constituent stocks

of the S&P500. LSTM neural networks outperformed memory-free classification methods, such as Random Forest, logistic regression, and memory-free neural networks. The model was able to generate excess returns compared to the market portfolio from 1992 to 2009, but from 2010, the model was not able to yield excess returns after transaction costs. Their findings give evidence of the market becoming increasingly mature.

Lyridis et al. (2004) applied neural networks for forecasting monthly VLCC spot freight rates from 1979 to 2012. The results gave evidence of neural networks providing valuable forecasts, especially in volatile periods. Further, they found that crude oil price spreads, and Capesize rates, improved the forecasting performance. Fan et al. (2013) utilized wavelet neural networks for predicting the Baltic Dirty Tanker Index. Among the features included in their model, was the Dow Jones Industry Average and the AMEX Oil Index. The results showed that their model was unable to predict rates more accurately than an ARIMA benchmark on short horizons, but showed signs of superiority on longer forecasting horizons.

There are two recent and relevant studies that cover machine learning methods with the utilization of AIS-extracted features, for predicting freight rates. Næss (2018) investigated whether multivariate machine learning methods with the inclusion of AIS-derived features, improved predictions of short-term rates in the LPG freight market. The thesis gave evidence of favoring multivariate machine learning models over a VAR model, where a Multi-Layer Perceptron neural network and a LSTM neural network were tested. The LSTM model yielded the best prediction power, and both machine-learning models predicted short-term freight rates more accurately when including AIS-derived features. Salen and Århus (2018) also applied LSTM neural networks with the inclusion of AIS-derived features, for predicting freight rate movements for the route between Ras Tanura, Saudi-Arabia and Singapore. The forecasting horizons were one, five, and ten days ahead. The model performed best on the ten days ahead forecast horizon, compared to a multivariate linear regression benchmark. The additional variables derived from the AIS data did not improve the model significantly. However, they state in the paper that more recent AIS data, and improved optimization of hyperparameters, could have improved the results.

FFA market research, the applicability of AIS data, and the predictive powers of machine learning, is covered to a great extent in previous literature. However, there has not been carried out comprehensive studies regarding the use of AIS features in combination with machine learning techniques, for forecasting movements in FFA prices. Næss (2018) and Salen and Århus (2018) applied machine learning techniques in combination with AIS data, for forecasting spot prices for selected routes, where we aim to predict the FFA prices for a

composition of routes for the Capesize segment. Thus, our approach is on a more global scale. Further, we create additional AIS-derived features, among them, a more global tonne-mile demand estimation, as well as several new approximations for unemployment. In addition, our study is perhaps more applicable in practical terms, as FFAs allow for more dynamic adjustments to freight rate exposure, in addition to our forecasting horizon being longer.

3   Data Foundation

This Section presents the AIS and price data that is utilized in this paper. Due to the increased quality in AIS data from 2014, and the change from the Capesize 4TC to the Capesize 5TC index, our study period will be from May 2014 (Skauen, 2015).

3.1   AIS Data

AIS is an automated system used in the maritime space for tracking and exchange of navigational information for vessels. It was mainly developed to prevent collisions and assist port authorities in controlling marine traffic more efficiently. Signals from AIS transponders are transmitted using Very High Frequency radio waves. Messages include both dynamic information, such as speed, positioning, and course, as well as static information, such as International Maritime Organization (IMO) number. (Marine Traffic, 2018).

We have been granted AIS data by the Center for Applied Research (SNF), which contains AIS messages for all bulk carriers from May 2014 to December 2018. We have separated the data into files based on the IMO numbers. The AIS messages do not contain information about vessel specifications, such as DWT. Hence, we have matched the IMO numbers from the AIS messages with fleet information from Clarksons World Fleet Register, and filtered the complete fleet list to only keep vessels above 150.000 DWT. This subset represents the most relevant vessels for the contracts that we are predicting. Figure 3.1 shows a sample message after separation, for each IMO number. Table 3.1 explains the message components.

timestamp_position,lon,lat,course,speed,draught,destination

2018-03-31 19:29:14,-45.49054,-26.026875,277.5,14.8,7.6,PARANAGUA BRZL

Figure 3.1 AIS sample message.

Table 3.1 AIS message components.

Message Component	Meaning
"timestamp_position"	date and time for the position
"lon"	longitude of the position
"lat"	latitude of the position
"course"	Sailing Course
"speed"	speed in knots
"draught"	draught in meters
"destination"	destination text as sent by the ship

Figure 3.2 shows the deadweight tonnage (DWT) distribution for the fleet subset. The average DWT in our sample is 202.472, with a clear separation in the distribution between Capesize vessels (150.000 DWT to 320.000 DWT), and Valemaxes of around 400.000 DWT.
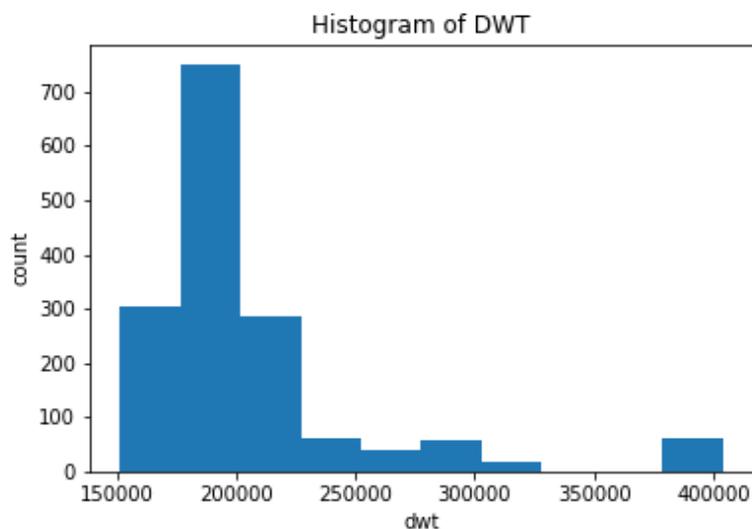
Figure 3.2 Histogram of DWT for the fleet subset.

3.2   FFA Price Data

We have obtained FFA prices from the Baltic Exchange. The data contains prices from May 2014 to December 2018. We will be looking into the contracts for the two nearest quarters, in addition to the nearest calendar year. A quarter consists of a basket of three monthly contracts settled on a rolling basis, while a calendar contract consists of 12 monthly contracts. The Capesize 5TC basket is comprised of route C8, C9, C10, C14, and C16, where a weighted average of the underlying routes is used for calculating the 5TC price. A brief description of the routes and the weights, are presented in Table 3.2.

Table 3.2 5TC description (Schmitz, 2016).

Route Code	Route description	Delivery	Duration	Weight
C8	Transatlantic round voyage	Gibraltar/Hamburg	30-45 days	25%
C9	Fronthaul	Amsterdam/ Rotterdam	About 65 days	12.5%
C10	Transpacific round voyage	China/Japan	30-40 days	25%
C14	China-Brazil round voyage	Qingdao	80-90 days	25%
C16	Revised backhaul	North China/ South Japan	About 65 days	12.5%

To create a continuous series of prices, we have sorted the contracts by maturity, and created continuous time series containing the contracts that are closest to maturity, but have not entered the settling period. Due to the structure of the series, the price may jump when rolling between contracts. According to Masteika et al. (2012), a proportional back-adjustment is a suitable for backtesting purposes. The adjustment ratio is calculated by dividing the price of the first day of the new contract by the price of the last day of the old contract. The price series will later be normalized, in order to keep the trends. See Figure 3.3 for a chart showing the actual FFA price series, and Figure 3.4 for a chart showing the synthetic FFA prices. See appendix A.2 for descriptive stats for the FFA prices before and after proportionally back-adjusting.

FFA Prices Average Daily
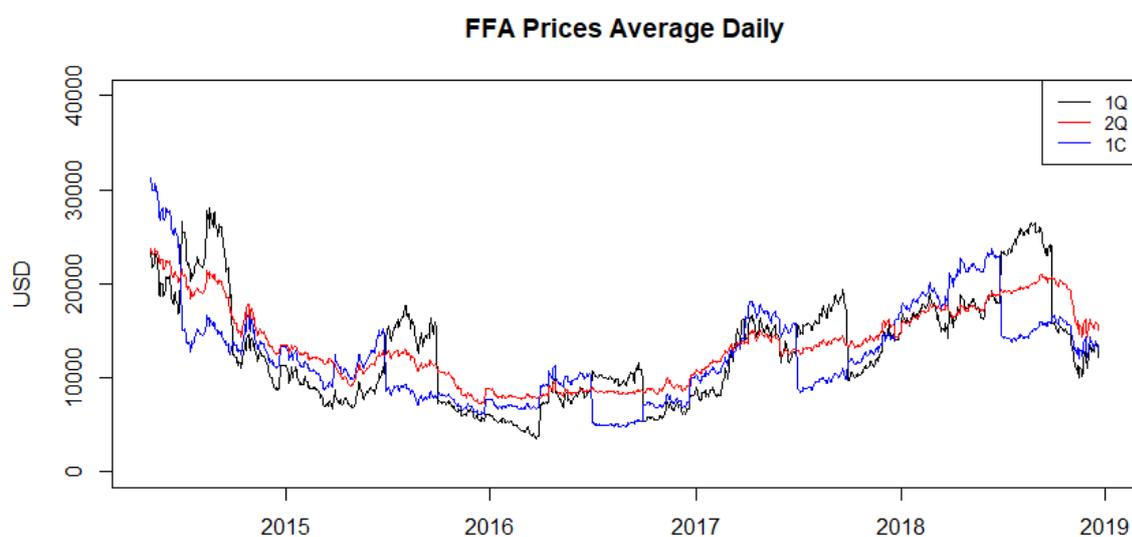
Figure 3.3 Actual FFA Prices.

Synthetic FFA Prices
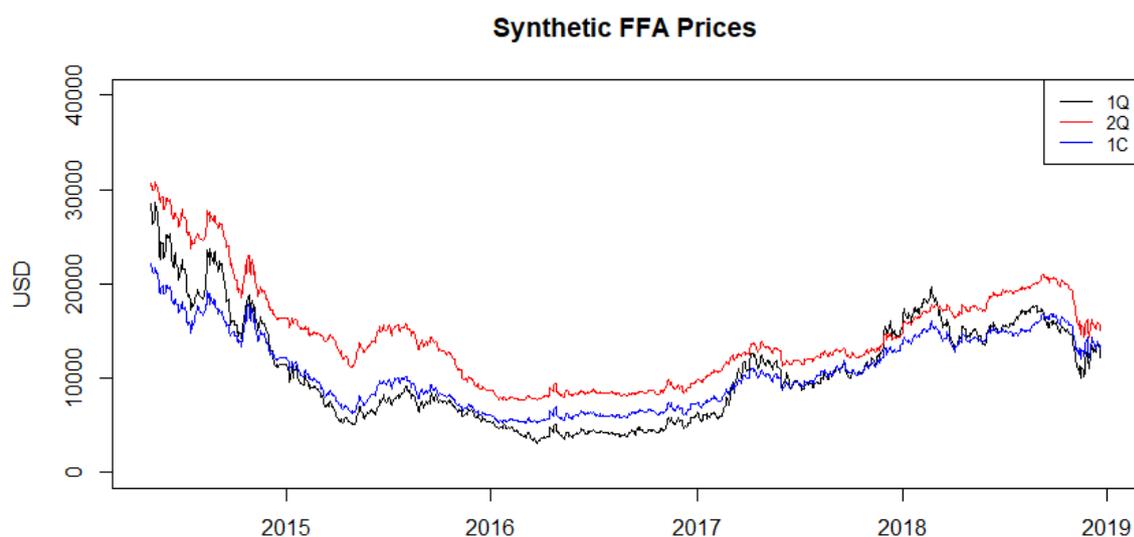
Figure 3.4 Proportionally back-adjusted FFA prices.

4   Feature Extraction

The process of identifying relevant features to extract, has been based on data exploration and the studies presented in the literature review. To capture the geographical positioning of the fleet, we will divide the world map into different world regions. By generating a plot showing the density of AIS signals from our vessel subset, we can identify the main sailing patterns for the fleet. The density plot helps to better visualize the general patterns, in contrast to a plot visualizing all observed patterns equally visible. This is pointed out in the study by Næss (2018). Figure 4.1 show a density plot for our fleet subset.
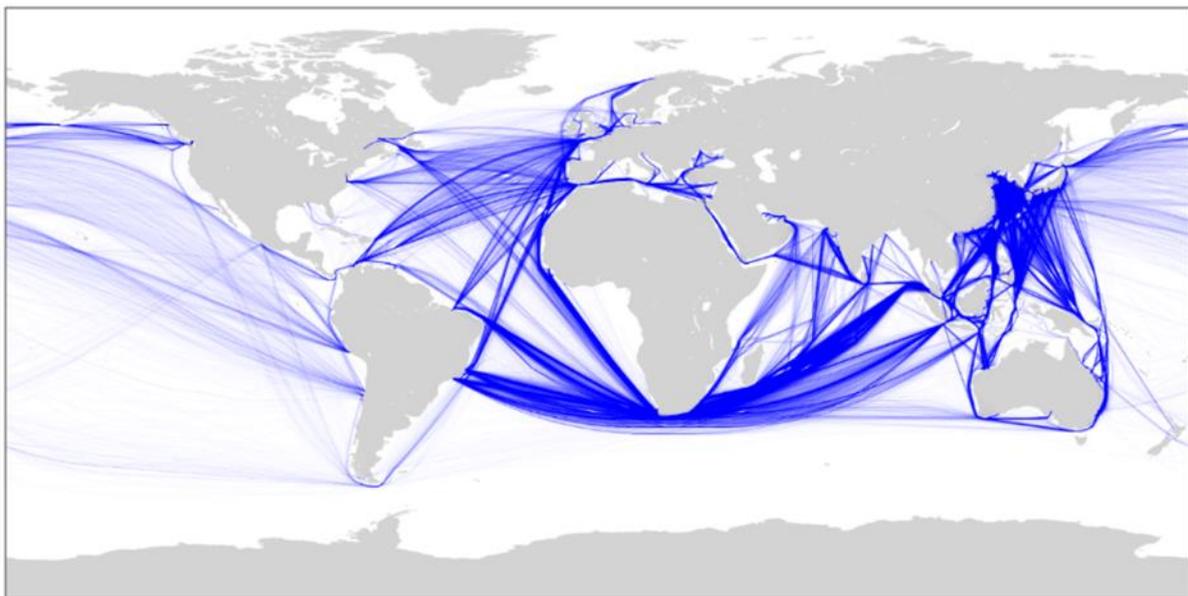
Figure 4.1 Density plot of the Capesize fleet.

Based on visual inspection of the density plot, in combination with export and import data from Clarksons Shipping Intelligence Network, we have divided the world into a set of polygons. The purpose of dividing the map into polygons, is to isolate regions with different characteristics concerning the trading pattern for Capesize vessels, and meant to capture movements between export and import regions. Figure 4.2 shows the world map divided into world regions, and Table 4.1 shows the world region names. To create daily time series concerning different regions, we have used the ray casting algorithm, which is a common method for determining which polygon a longitude/latitude pair is inside (Narkawicz & Hagen, 2016). The use of a ray casting algorithm is also suggested in the work of Næss (2018).
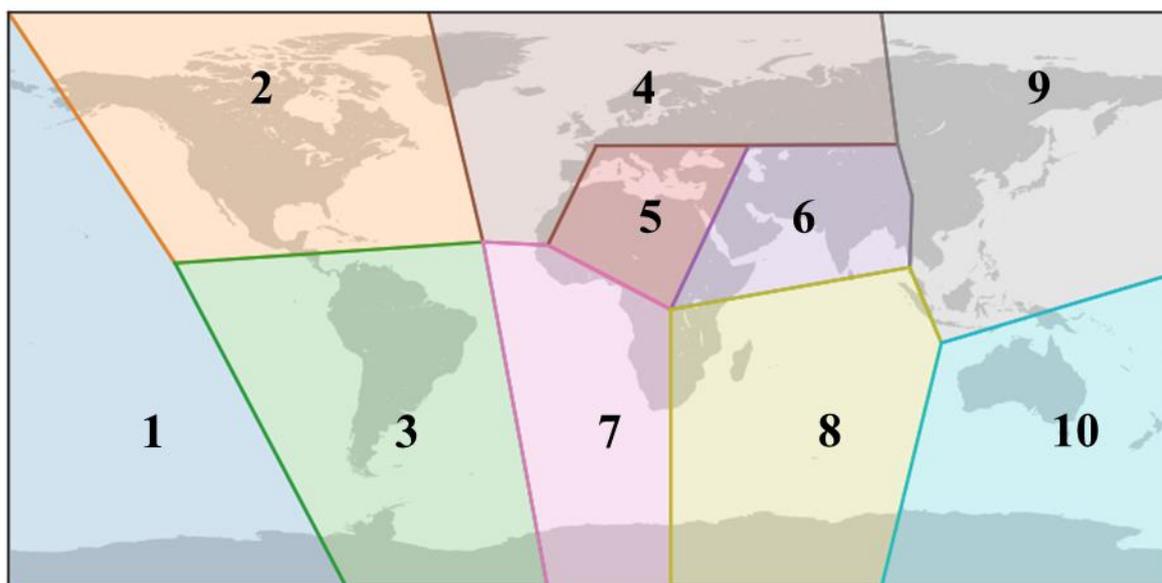
Figure 4.2 World regions.

Table 4.1 World region names and corresponding number.

Number	Name
1	pacific_ocean
2	north_america
3	south_america
4	europe
5	med_sea
6	arabian_gulf
7	south_africa
8	indian_ocean
9	asia
10	aus

To calculate sailing distances between positions, we utilize an open-source distance calculator called, "python ports distance calculator" (Witsung, 2019). The method makes use of a pixelated world map, where all land areas are marked as unavailable for travelling through. In turn, the pixelated world map is transformed to an array. The algorithm finds the route between two sea coordinates passing the minimal amount of points in the map array. Finally, the distance in nautical miles is calculated between each point in the identified least cost route, using Vincenty's formula. This formula calculates the distance on the surface of the earth, assuming the shape of the earth is an oblate ellipsoid (Scheucher, 2016). Table 4.2 shows

examples of the calculated distances between two ports, compared to the ones listed on sea-distances.org. As can be seen, there are some minor differences for the selected routes, but it seems like an acceptable approximation.

Table 4.2 Distance comparison for selected routes.

	Port Hedland to Qingdao	Rotterdam to Qingdao
Sea-distances.org	3583 NM	10751 NM
Distance calculator	3531.60 NM	11218.8 NM
Difference in nautical miles	51.4	467.8
Difference in % of seadistance.org	-1.43%	4.35%

When processing the draught data, we assume that an average draught status below 70% of a vessel's maximum observed draught, implies that the ship is sailing ballast. Figure 4.3 shows the distribution of draught ratios on a given day for the Capesize fleet. The bimodal shape of the distribution implies that this threshold is reasonable.
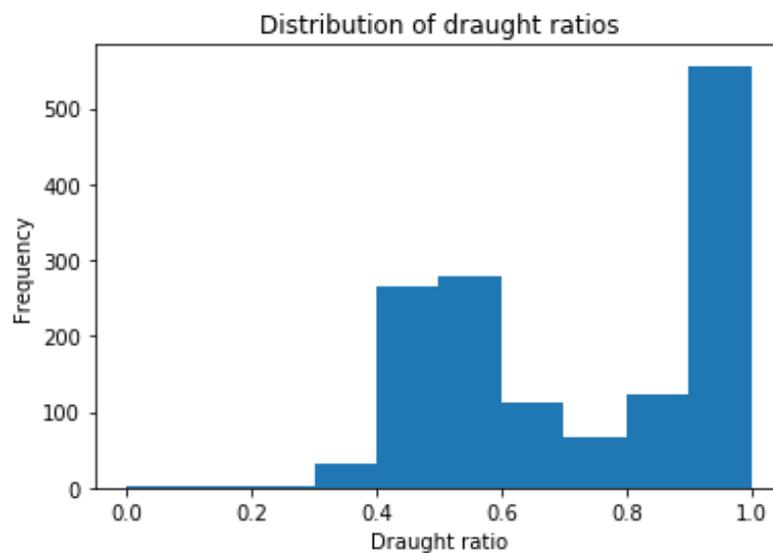
Figure 4.3 Distribution of draught ratios.

4.1 Count of Vessels and Capacity

The total fleet capacity reflects the total supply (Stopford, 2009), hence we will count the total number of vessels and capacity globally. Further, we will create features for each of the world regions, that count the number of vessels, total freight capacity, and the relative capacity distribution. These features may capture regional imbalances that are relevant for the development of freight rates (Regli & Nomikos, 2019; Næss, 2018). The freight capacity for each vessel is assumed to be 95% of its DWT. The capacity in each world region is then calculated by aggregating the capacity for each vessel within a world region. Thus, this measurement does not consider whether a vessel is ballast or laden, or whether it is contracted or not. The count of vessels is simply measured as the total number of vessels within a world region, and the relative capacity is calculated by dividing the capacity in a world region, by the total capacity of the entire fleet, on a given day. The relative count of vessels is also calculated similarly. Figure 4.4 shows regional capacity and relative capacity for selected world regions.
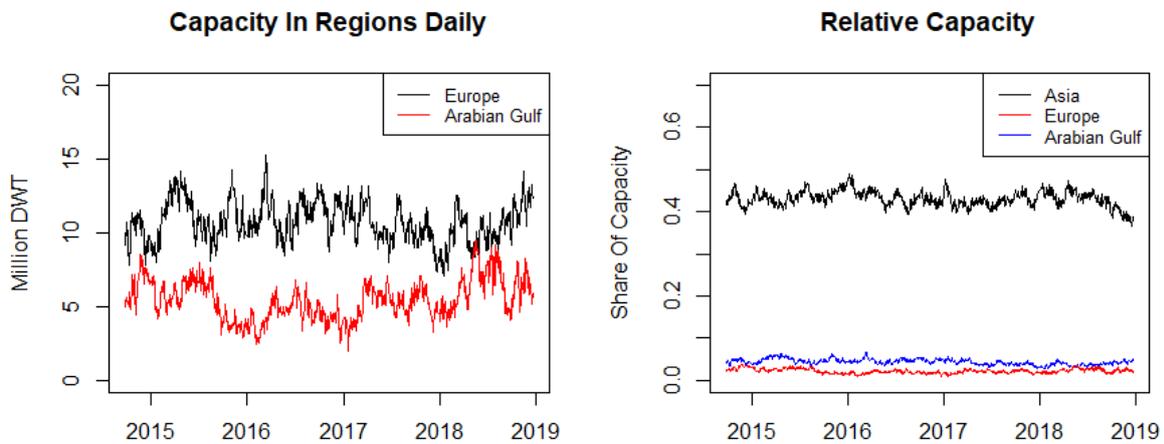
Figure 4.4 Capacity and relative capacity for selected world regions.

4.2 Net Flow of Vessels in World Regions

In addition to counting vessels within world regions, we will calculate the net flow of vessels for each world region. This is done by recording when a vessel travels from one world region to another. The sum of incoming and outgoing vessels for a world region is then calculated. Figure 4.5 shows the net flow of vessels for Asia.
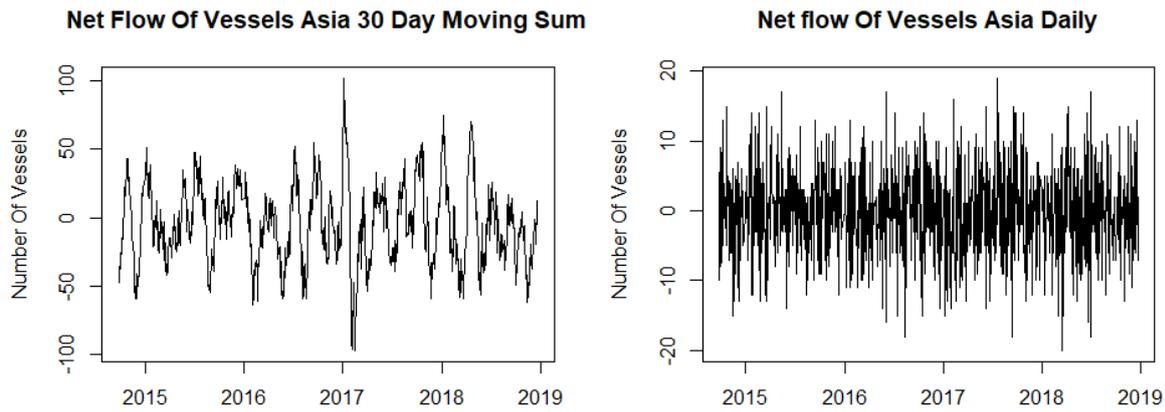
Figure 4.3 Net flow of vessels for Asia.

4.3 Fleet Sailing Speed and Standard Deviation

The speed of the fleet affects operational efficiency (Stopford, 2009), and we will generate several speed features. The average speed for a vessel is calculated by looking at the first and last observation for a ship for a single day, and calculating the distance. The distance traveled is then divided by the time difference between the first and last observations. We will calculate average speeds for the whole fleet, and average speeds within each world region. Further, we will distinguish between vessels classified as sailing laden or ballast. First, we will calculate all the speed features without including stationary vessels (vessels with a daily average speed below 2 knots), as this subset better represents the speed of the fleet actually sailing. Second, we will create the same features with the whole fleet, including stationary vessels, as this may capture some additional information. Additionally, we will include the standard deviation for each speed feature as this may provide information regarding the variation in operational efficiency. The calculations for the speed features are shown in Equations 4.1 and 4.2. Speed plots are shown in figure 4.6.

$$\text{Average Speed} = \frac{\sum_{n \, \varepsilon \, N} \frac{\text{Distance sailed}}{\text{Sailing time}}}{\text{N bulkers}} \tag{4.1}$$

$$\text{Speed standard deviation} = \frac{1}{N}\sqrt{\sum_{n\,\varepsilon\,N}\left(\frac{\text{Distance sailed}}{\text{Sailing time}} - Average\ speed\right)^2} \quad (4.2)$$
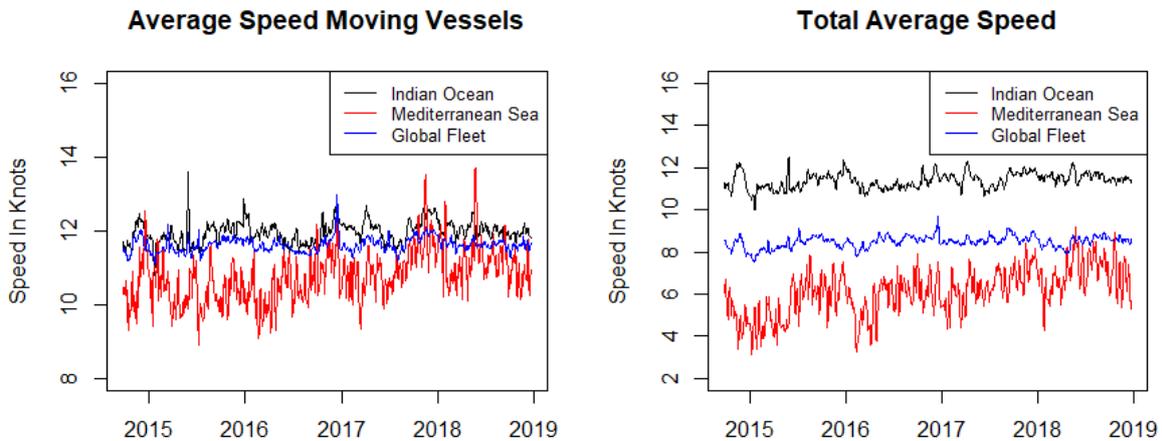
Figure 4.4 Average speed moving vessels, and total average speed.

As we can see from the two speed plots in Figure 4.6, the inclusion of stationary vessels makes a big difference for the average speed for the Mediterranean Sea and the global fleet. However, the differences are not notable for the Indian Ocean, which seems reasonable due this world region covering open sea for vessels passing Africa, where all vessels are expected to be sailing in normal speeds. An additional note is that the average speed for the Indian Ocean is consistently higher than for the global fleet, which also seems reasonable.

4.4   Tonne-Mile Demand

The real demand of freight is calculated on a tonne-mile basis, as it includes both the volume of cargo and the distance (Stopford, 2009). We will therefore create a proxy for tonne-mile demand for our fleet. Based on visual inspection of AIS data, investigation of export and import data from Clarksons Shipping Intelligence Network, and a list of Capesize ports, we have created the port area polygons shown in Figure 4.7. Each port area is labeled as either an import or export port area. The rationale for this, is that we want a system for estimating tonne-mile demand without relying on draught. See appendix A.1 for the labeling of port areas.
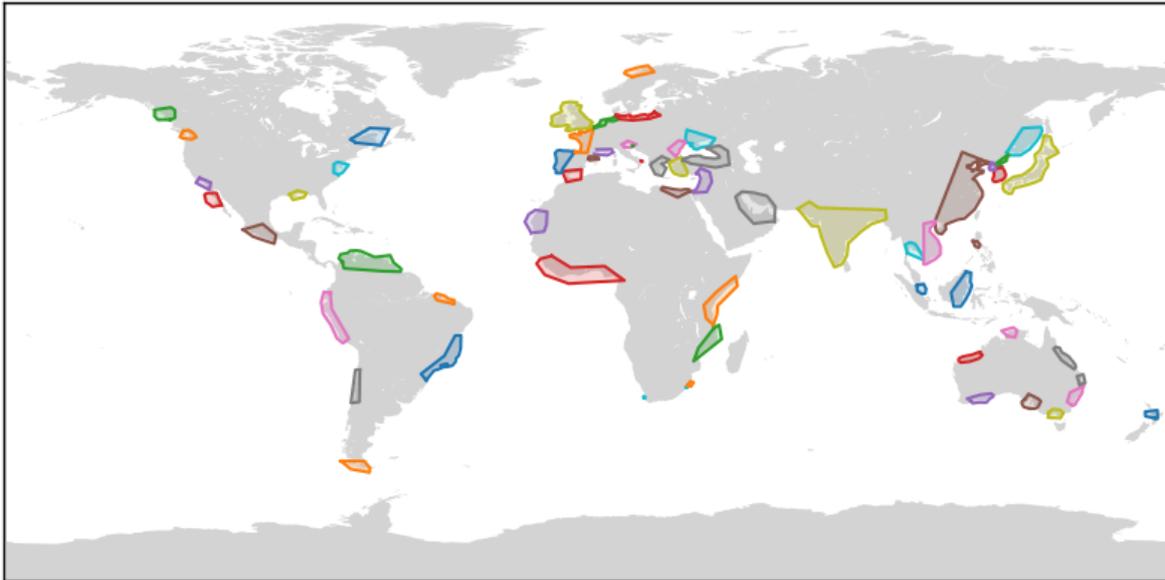
Figure 4.7 Port areas.

An estimate of realized tonne-mile demand is created through the following pseudo-algorithm:

- All positional observations for each vessel are processed, and stationarity in either a defined import or export port area longer than 12 hours is classified as a port call for loading or discharge of cargo.
- Each classified unload port call is matched with the previous load port call.
- If there are multiple loading port calls registered in a row for a vessel, the last will be registered, and if there are registered multiple discharge port calls, the first will be registered. Hence, there will only be a matching of two ports for distance calculations.
- For each pair of loading and discharge port calls, the distance is calculated and multiplied by 95% of the vessel's DWT, thus assuming all cargo is transported in full shiploads.

A drawback with this estimate, is that the tonne-mile demand only will include the international part of a multi-port voyage. An additional drawback is that the tonne-mile demand will be observable on the day of discharge. Thus, it is expected to be lagged four to eight weeks from when the demand was actually realized, i.e. when the ship was fixed. The tonne-mile demand estimate for a given day is calculated as shown in Equation 4.3. Figure 4.8 shows moving sums for the tonne-mile demand estimate.

$$\text{Tonne} - \text{mile demand} = \sum_{n\,\varepsilon\,N} \text{DWT} \cdot 95\% \cdot \text{distance} \qquad (4.3)$$
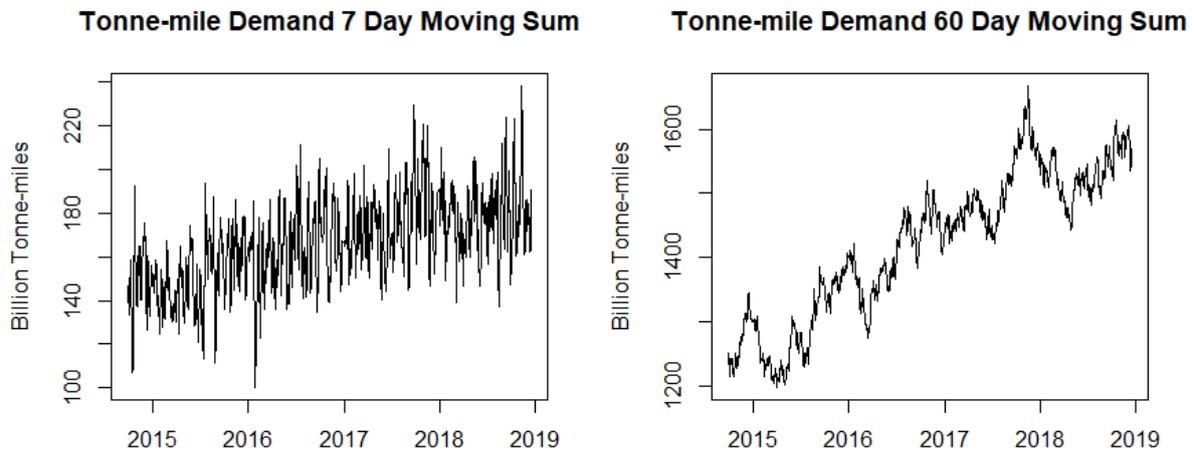
Figure 4.8 Moving sums of the tonne-mile demand estimate.

When calculating the realized tonne-mile demand for 2017 based on our proxy, we end up at 9.148 billion tonne-miles for our fleet. According to Clarksons Research (2019), the total demand for iron ore and coal was 13.430 billion tonne-miles, while the demand for all major bulks (iron ore, coal and grain) was 16 852 Billion tonne-miles in 2017. As we have excluded the small and mid-size vessels in the dry bulk fleet, it is hard to compare directly, but it seems like our tonne-mile estimate might be able to capture some dynamics.

4.5 Load Factor and Loading Status

The load factor for the fleet is calculated by taking the daily average draught for each vessel, both globally and in each world region. Subsequently, all the average regional draught ratios are then averaged for each world region. Because the ship crew manually enters this data, it is prone to errors according to Jia et al., (2015). In addition, the method of calculating a mean of means could also be affected by an uneven distribution of observations, among the vessels in a world region. Equation 4.4 provides the calculation of the average load factor.

$$\text{Average load factor} = \frac{\sum_{n\,\varepsilon\,N} \frac{\text{Current Draught}}{\text{Maximum Draught}}}{\text{N bulkers}} \qquad (4.4)$$

The observed vessel count and capacity sailing with a draught less than 70% of its observed maximum, are also aggregated, providing a proxy for the number of vessels sailing ballast. Similarly, we have made a proxy for vessels and capacity sailing laden, by aggregating the count of vessels with a draught ratio above 70%. As draught is manually set, we have also included a feature counting vessels leaving export port areas and a feature counting vessels leaving import port areas. Figure 4.9 shows the average load factor and the relative laden share of the fleet.
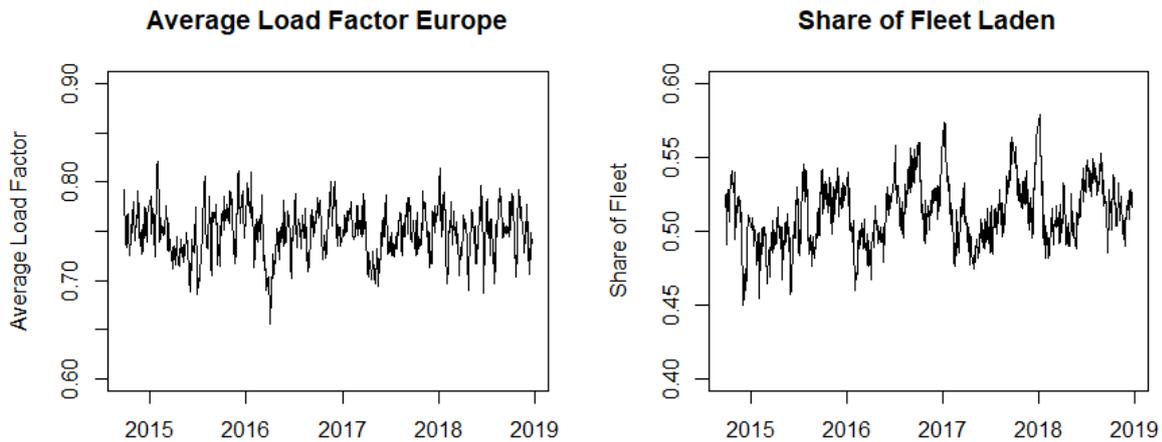
Figure 4.9 Average load factor and laden share of the fleet.

4.6 Operational Status

According to Adland (2019), the number of unemployed ships may provide information regarding the short-term balance of supply and demand. Unemployed vessels represent an oversupply, hence there is an inverse relationship between unemployed ships and freight rates. We will therefore create proxies for unemployed ships, total idle non-laden ships, and stand-by capacity.

A proxy for unemployed ships is made by aggregating stationary vessels that are non-laden based on AIS-reported draught (<70%), and are outside of discharge ports, or in areas not defined as port areas. An extension of this feature is also created, which is called "total idle non-laden" ships, which aggregates non-laden, stationary ships, regardless of location. These measures are included, as stationary status, or non-laden status, by themselves are insufficient for determining the contractual status of a vessel.

Further, we have created a measure for stand-by capacity, which measures the number of stationary vessels in the major exporting world regions (North, America, South America and Australia), as it may capture some information regarding vessels awaiting loading operations. In addition, the proxy may capture some information regarding vessels waiting to get a contract, or waiting for a contract to commence.

Additionally, features cumulating the stationary time for vessels either categorized as "unemployed" or "total idle non-laden" are also created. These features represent unused supply, providing measures of shipdays.

Both the proxy for unemployed and total idle non-laden ships may underestimate the actual values, due to inactivity of AIS transmitters. Even though there are drawbacks with these measures, we believe they may provide useful information over time. Figure 4.10 shows the unemployed share of the fleet, and the total idle non-laden share of the fleet, based on our calculations.
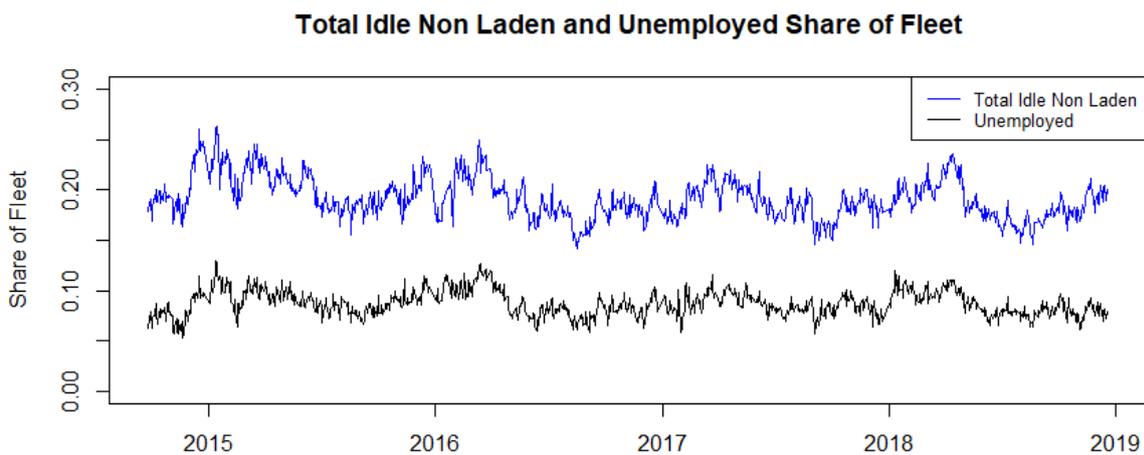
Figure 4.10 Total idle non-laden and unemployed share of the fleet.

Table 4.3 shows all AIS features previously described. We aim to capture several aspects affecting freight rates, as we both include measures for supply, demand, operational efficiency and operational status. The AIS data quality improved significantly from 2014 because of improved satellite coverage (Skauen, 2015). However, there are still signal gaps, giving an uneven distribution of AIS messages. Thus, there may be inconsistencies in the feature values. To account for this, we have calculated moving averages of seven, thirty and sixty days for all features, except for tonne-mile demand and net flow of vessels, where a moving sum is used.

Table 4.3 All described AIS features and what they measure.

Feature	Measure
Global capacity	Total supply
Global vessel count	
Global count of ballast vessels	Operational Status
Global count of laden vessels	and utilized supply
Global load factor	
Regional load factor	
Regional vessel count	Regional supply
Regional capacity	and distribution
Regional relative capacity	of supply
Regional relative vessel count	
Regional net flow of vessels	Regional changes
Global count of unemployed vessels	Excess supply and
Global unemployed capacity	Stand-by capacity
Global count of idle-non-laden vessels	
Global cumulative sum of shipdays unemployed	
Global cumulative sum of shipdays idle non-laden	
Stationary vessels in export world regions	
Global average speed	Operational efficiency
Global average speed moving vessels	
Regional average speed	
Regional average speed moving vessels	
Global average speed laden vessels	
Global average speed laden vessels moving	
Global average speed ballast vessels	
Global average speed ballast vessels moving	
Global average speed standard deviation	Changes in operational
Regional average speed standard deviation	efficiency
Ballast speed standard deviation	
Laden speed standard deviation	
Global average speed for moving vessels standard deviation	
Tonne-mile demand	Total demand

4.7 Non-AIS features

In addition to the AIS-derived features, we will include price and market information shown in Table 4.4.

The Capesize 5TC spot rate, Baltic Capesize index and Baltic Dry Index will be included, as Kavussanos and Visvikis (2004), among others, have found spot rates to have an effect on FFA prices. The exchange rate of EUR to USD and Yuan to USD will additionally be included, as fluctuations in exchange rates has an impact on revenue and costs for operators in the market (Kavussanos & Visvikis, 2006).

Bunker prices affect the cost of operating vessels. In classical literature, vessels adjust the speed corresponding to changes in bunker prices, causing changes in the operational efficiency of vessels (Stopford, 2009). We assume our subset of vessels use 380Cst marine heavy fuel oil, and we will include the price from Bunkerindex.com. The price is calculated in dollars per metric ton, based on the average prices for all 380Cst port prices. We will also include Brent Crude oil prices, due to having several applications, among them transportation (Tsioumas, 2016).

Due to the findings of Tsioumas and Papadimitriou (2018) implying a bi-directional relationship between the Baltic Capesize Index (BCI) and the prices for iron ore and coal, we will include the spot index for iron ore 62% (ISIX62IU), and Rotterdam Coal futures (API21MON). In a setting with increased demand for iron ore, the price for iron will increase, also causing a rise in the demand for transportation. On the other hand, a positive shock in freight rates, may cause operators in the market to consider other transportation options, or store more commodities as inventory. This will effectively reduce the supply, which leads to increased commodity prices. A weakness of including these commodity prices, is that the effect on freight rates is dependent on whether the change in a commodity price is driven by supply or demand factors. A sudden fall in demand for major bulks will usually lead to a fall in commodity prices, and lead to reduced freight rates. On the other hand, a negative supply shock will usually lead to increased commodity prices, but decreased freight rates (Tsioumas & Papadimitriou, 2018).

In addition to the features above, we will include the S&P500, US 10 year government bond yields, and US 3 month Libor yields. They may hold information about the development in the economy, as well as future funding rates, and general expectations for the future, according to Da et al. (2015).

Table 4.4 Non-AIS features.

Feature	Description
Baltic Capesize Index	Spot index for Capesize vessels
5TC spot rate	Spot index for 5TC basket
Baltic Dry Index	Spot index for dry bulk
Euro to USD exchange rate	Exchange rate
Yuan to USD exchange rate	Exchange rate
Average 380Cst bunker prices	Bunker price average
Brent crude oil price	Brent Crude Spot Price
ISIX62IU	Iron Ore Spot Price Index
API21MON	Rotterdam Coal Futures Price
S&P500	S&P500 Index
US 10 year government bonds	Bond yield
US 3 month LIBOR	Bond yield

The total number of features available is 623 after including moving averages and moving sums. Descriptive statistics of the features used in the final models in this study, are presented in appendix A.2.

5   Methodology

5.1   Data Preparation

The process of preparing the data consists of data transformation, data normalization, and splitting of the data into training and test sets. Data transformation is the process of differencing the data to get it in a stationary form. To check for stationarity, we will be performing an Augmented Dickey-Fuller test. See appendix A.3 for test results showing evidence of stationarity after first differencing.

After the data is transformed into stationary form, the next step is normalization. We utilize the min-max scaling method for normalization as shown in Equation 5.1. Normalization could prove necessary when the scale of features differs, and when the ranges of values are large. The former because features with larger scales will have a greater impact on the predicted output (Angelov & Gu, 2019). The latter because it could cause slow learning and convergence for the neural network (Brownlee, 2019). The min-max scaling gives each observation a value between 0 and 1, which is appropriate in the context of a neural network (Brownlee, 2019). Although the min-max scaling method is commonly used in practice, it does not handle outliers well. If outliers are present, they will highly influence the results (Angelov & Gu, 2019).

$$x_{norm} = \frac{x - X_{min}}{X_{max} - X_{min}} \tag{5.1}$$

5.2   Walk Forward Validation

When training and evaluating the prediction models, we perform walk forward validation, also called expanding window cross-validation. This procedure makes use of a series of test sets, each consisting of a single observation H steps ahead, where H denotes the forecasting horizon. Every corresponding training set consists of all observations that are at least H steps prior to each test observation, in order to avoid look-ahead bias. The concept of updating the predictive model at each time step improves the models opportunity of making good predictions, due to continuously receiving new information and patterns to be included in retraining (Brownlee, 2016). As the forecasts will not be reliable if they are based on a small training sets, the series of test sets do not start before the last 15% of the available data, from May 2018. The accuracies of the forecasts are obtained by averaging the results in the test set

the over the entire forecasting period. The nature of this procedure involves successive testing on the same data and could be a source of overfitting. Figure 5.1 shows the principle of the walk forward validation method for predictions $H = 5$ steps ahead. Here, the blue series represent the training sets, while the yellow fields represent the test sets. (Hyndman & Athanasopoulos, 2013).
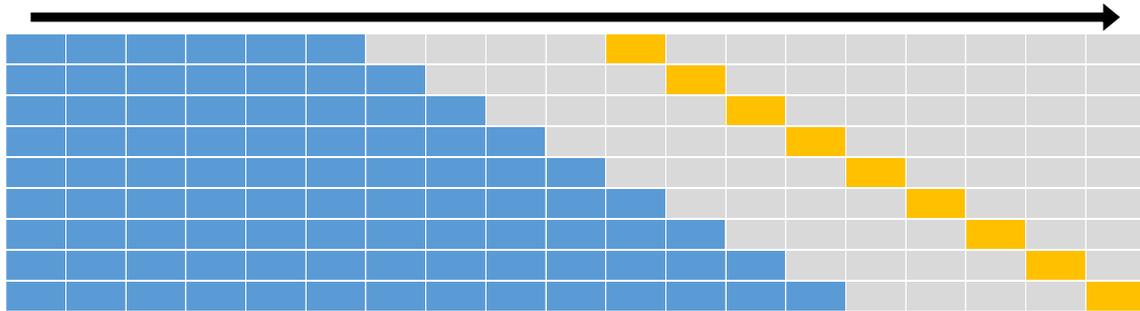
Figure 5.1 Walk forward validation.

5.3   Machine Learning Methodology

Machine learning techniques represent a set of algorithms that enables a learning process from a data set, without being directly programmed. When working with supervised learning and regression tasks, the goal of the machine learning model is to receive input data, adjust parameters, and produce an output that is as close as possible to the actual value. The process of adjusting the parameters in the model is done by training on historic observations according to James et al. (2017)

Neural networks belong to a class of machine learning models that are capable of adding increased complexity, and is able to comprehend non-linear relationships (Haykin, 2009). Neural networks consist of layers of neurons and weighted connections. The first layer of a neural network is the input layer, which is passed the independent variables. The network further consists of hidden layers and an output layer. Figure 5.2 shows a simplified structure of a neural network. (Haykin, 2009).

Input layer   Hidden layer   Output layer
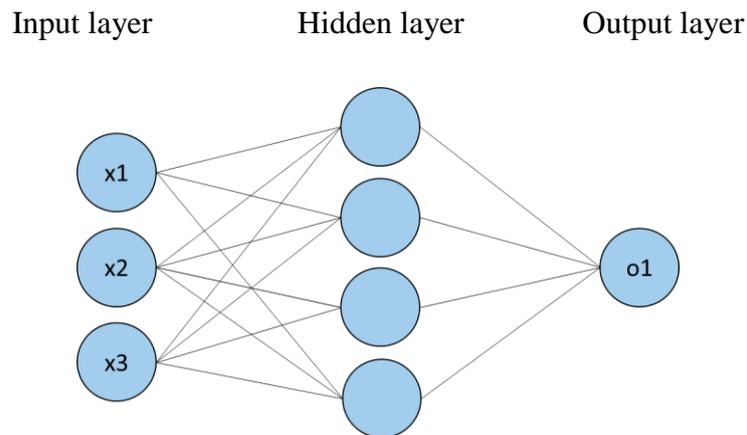
Figure 5.2 inspired by Haykin (2009).

The information in the network flows from the input nodes, through the nodes in the hidden layers, before finally calculating an output. The connections between the nodes has a weight w, which regulates the information flow between nodes. The neuron values in the hidden layers and the output layer is calculated as the sum of the products of every incoming neuron and the connecting weights, and additionally adjusting for the bias b. (Haykin, 2009).

Equation 5.2 shows the calculation of the value of a neuron $x_{l,f}$, based on the weights and neuron values from the previous layer, $w_{l-1,f}, x_{l-1,f}$, as well as the bias, $b_{l,f}$.

$$x_{l,f} = \sum_{f=1} (w_{l-1,f} \cdot x_{l-1,f}) + b_{l,f} \tag{5.2}$$

The full process of calculating predicted values based on the independent variables is called a forward propagation. When a forward propagation is completed, the predicted value is compared to the actual value, and a loss function is computed. The loss function expresses the accuracy of the predictions, and the learning process for the network is based on minimizing the loss function by adjusting the parameters (Haykin, 2009). The contribution for each parameter to the loss function is calculated and adjusted between every forward propagation.

Equation 5.3 expresses a given loss function, where x represents the input values, θ represents the parameters (weights and biases), and y represents the actual output value.

$$L(\hat{y}, y) = L(f(x, \theta), y) \qquad (5.3)$$

Recurrent neural networks are looped, which enables the passing of information between the steps in the network, thus enabling information to persist. Olah (2015) states that recurrent neural networks may be thought of as multiple copies of the same network, where each network passes on information. Figure 5.3 shows the principle of the passing of information between consecutive steps in a.recurrent neural network.
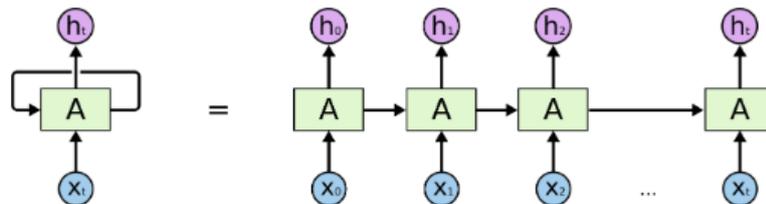
Figure 5.3 Illustration of the chained structure of a recurrent neural network (Olah, 2015).

Recurrent neural networks generally perform well on short-term dependencies, but struggle to perform well when the duration of the dependencies increases. The reason for performing poorly on long-term dependencies is due to exploding or vanishing gradients, according to Bengio et al. (1994). Vanishing gradients shrink exponentially and make it difficult for a model to learn. Exploding gradients grow exponentially and impairs learning, and can cause instability and crash the model. LSTM networks are a subgroup of recurrent neural networks that are capable of learning long-term dependencies, as well as overcoming the problems of exploding or vanishing gradients (Hochreiter et al., 2001). See appendix A.8 for a more in-depth introduction to LSTM.

5.4   Hyperparameters

The internal parameters are set by training the neural network, while hyperparameters on the other hand, are determined by the researcher. The hyperparameters are set before the training process of a network commences. There are numerous configurations for the hyperparameters, and the optimal values are dependent on the problem to be solved (Leoni, 2019). The hyperparameters considered for adjustment in our LSTM neural network are presented below.

- Number of hidden layers
- Hidden nodes
- Learning rate
- Batch size
- Epochs
- Window size
- Regularization

The number of *hidden layers* determines how many layers there are between the input and output layers. When adding hidden layers, they essentially form new combinations of the previous learned representation of the problem to be solved. (Brownlee, 2017). The number of *hidden nodes* determines the number of units in each hidden layer. When training the model, the learning rate determines how much the model changes the weights based on the estimated error (Brownlee, 2019b). The *batch size*, on the other hand, determines the number of samples of data to be processed before adjusting the model parameters, where a sample represents the input sequence for one timestep (Brownlee, 2018). The process of going through the entire training set and adjusting the weights is known as an *epoch*, and the number of epochs determines how many times this process is repeated (Brownlee, 2018). Each sample includes an amount of previous observations, determined by the *window size*. Finally, the *regularization* is the inclusion of constraints to a model, which helps to reduce overfitting, and increase out of sample performance (Brownlee, 2017b).

5.5   Benchmark Models

To create grounds for comparison for the LSTM models, we will include a Random Walk model, Autoregressive Integrated Moving Average (ARIMA) models, and Vector Autoregressive (VAR) models.

Random Walk

The Random Walk is a standard model for benchmarking in forecasting. This model is also known as the naive model and takes the last actual value as the forecast (Hyndman & Athanasopoulos, 2013). The Random Walk model can be formulated as shown in Equation 5.4.

$$\hat{y}_{t+H} = y_t \tag{5.4}$$

Where H is the forecasting horizon

ARIMA

ARIMA is a univariate forecasting technique which uses the past lags and errors of dependent variable Y. It is a common forecasting tool when working with time series, and capable of capturing trends The AR term refers to the use of previous observations of dependent variable Y as features, and the number of lags included is determined by the parameter p. The integrated term is defined by a parameter d, which determines the order of differencing. The MA term refers to the use of past error terms e_t, where parameter q determines the number of error terms to include. Equation 5.5 shows the generalized combinations of the AR and MA terms depending on the values of p and q. (Hyndman & Athanasopoulos, 2013).

$$y_t = a + \beta_1 y_{t-1} + \ldots + \beta_p y_{t-p} + \phi_1 \varepsilon_{t-1} + \ldots + \phi_q \varepsilon_{t-q} \tag{5.5}$$

VAR

VAR is a multivariate forecasting technique that facilitates the inclusion of previous values of features and predictions, and has proven to be a powerful forecasting tool when working with financial time series (Zivot & Wang, 2006). VAR is a system of equations where every variable is calculated as linear combinations of past values of all the variables. The difference between traditional models, such as linear regression where predictor variables only affect the dependent variable, is that the variables influence each other. An example system with two variables and one lag can be expressed as shown in Equations 5.6 and 5.7. (Hyndman & Athanasopoulos, 2013)

$$Y_{1,t} = c_1 + \emptyset_{11} Y_{1,t-1} + \emptyset_{12,1} Y_{2,t-1} + \varepsilon_{1,t} \tag{5.6}$$

$$Y_{2,t} = c_2 + \emptyset_{21}Y_{1,t-1} + \emptyset_{22,1}Y_{2,t-1} + \varepsilon_{2,t} \tag{5.7}$$

Where $\varepsilon_{1,t}$ and $\varepsilon_{2,t}$ represents white noise processes, $\emptyset_{ii,\ell}$ denotes the influence of the ℓth lag of variable Y_i on itself, while $\emptyset_{ij,\ell}$ denotes the influence of the ℓth lag of the variable Y_j on Y_i.

5.6 Feature Selection

Feature selection is the process of choosing the features to be utilized in the prediction model. The goal is to remove irrelevant and redundant features, and avoid getting an overfitted model. According to James et al., (2013), effective feature selection increases out of sample prediction accuracy, in addition to making the model easier to comprehend. Ideally, we would test the model with all combinations of the available features, but this will be too computationally expensive.

We will perform the filter methods presented in Table 5.1, where the resulting metric of feature importance is scaled to the range [0, 1]. Subsequently, we will calculate a mean importance score based on the performance from all the filter methods. Among the filter methods considered are both univariate and multivariate, as both could provide useful insight as to the feature's importance. This scheme of creating a mean feature importance score is inspired by the work of Næss (2018). In the following are brief descriptions of the filter methods considered.

Table 5.1 Filter methods for feature selection.

Multivariate	Univariate
Linear Regression	Correlation coefficient
Lasso Regression	Maximal Information Coefficient (MIC)
Ridge Regression	
Random Forest	

Pearson's Correlation Coefficient

Pearson's correlation coefficient expresses the linear relationship between two variables and can be calculated as shown in Equation 5.8. The output for a correlation coefficient is in the range of [-1, 1], and by taking the absolute value, the coefficient will be in the range of [0, 1] The correlation coefficient is useful for indicating linear relationships but could be misleading if the there exists a non-linear relationship.

$$\rho X, Y = \frac{cov\ (X, Y)}{\sigma_X \sigma_Y} \qquad (5.8)$$

Maximal Information Coefficient (MIC)

MIC, first presented by Reshef et al., (2011) is a metric that is able to discover a wide variety of relationships between two features, linear and non-linear. The metric possesses the property of *equitability,* meaning that it returns the same score for equally noisy relationships independent of the type of relationship (e.g linear, polynomial etc.) (Reshef, et al., 2011). The MIC takes a value in the range of [0, 1] and is calculated as shown in Equation 5.9. (Kinney & Atwal, 2014)

$$MIC(X, Y) = max \left\{ \frac{I(X, Y)}{\log_2 min\{n_X, n_Y\}} \right\} \qquad (5.9)$$

Where $I(X, Y)$ is the mutual information, and can be defined as shown in Equation 5.10 (Kinney & Atwal, 2014).

$$I(X, Y) = \sum_{x, y} p(x, y) \log_2 \frac{p(x, y)}{p(x)p(y)} \qquad (5.10)$$

The MIC represents the mutual information between random variables X and Y, normalized based on the minimum joint entropy between the two given random variables (Wint, 2019).

Multiple Regression

Multiple regression can also give an indication of the importance of variables through the magnitude of the coefficients. The multiple regression linear model between the dependent variable Y and the independent variables X, can be expressed as in Equation 5.11, according to James et al. (2017).

$$y_t = \beta_0 + \beta_1 x_{1,t} + \ldots + \beta_p x_{p,t} + \varepsilon_t \qquad (5.11)$$

Where $x_{1,t}, \ldots, x_{p,t}$ represents the value of a feature at time t, $\beta_1, \ldots, \beta_p$ are the coefficients for the features in regards to Y, β_0 is the intercept, and ε is the error term. The regression coefficients, β, are found by minimizing the residual sum of squares, shown in Equation 5.12.

$$RSS = \sum_{t=1}^{N} (y_t - \hat{y}_t)^2 \qquad (5.12)$$

Fitting a linear regression model with all available features will probably lead to overfitting. The upcoming methods of Lasso and Ridge regularization could be able to counter this drawback. When fitting an Ordinary Least Squares model with many predictors, the model will likely suffer from multicollinearity. However, the method is simple and can provide insights with regards to feature importance. The coefficients from the linear regression model are scaled to the range of [0, 1] for comparability with the other methods.

Lasso Regression

The Lasso regression is a regularization technique that applies shrinkage to a linear regression model. The Lasso adds a penalty correspondingly to the absolute value of the magnitude of the coefficients. This has the effect of shrinking less important coefficients features toward zero, where some coefficients may even shrink all the way to zero. Hence, the Lasso procedure results in sparse models that reduce problems with overfitting and multicollinearity. λ is the parameter that determines the impact of the regularization. The cost function for lasso can be expressed as shown in Equation 5.13. (James et al. 2017).

$$\sum_{t=1}^{n} \left(y_t - \beta_0 - \sum_{j} \beta_j x_{t,j} \right)^2 + \lambda \sum_{j=1}^{p} |\beta_j| \qquad (5.13)$$

Ridge Regression

The Ridge regression behaves similarly to the Lasso regression, but the added penalty is here equivalent to the square of the magnitude of the coefficients. This causes all coefficients to be shrunken by the same factor, and will not eliminate any coefficients from the model, unlike the Lasso regression. The cost function to be minimized can be expressed as shown in Equation 5.14. (James et al. 2017).

$$\sum_{t=1}^{n} \left(y_i - \beta_0 - \sum_{j} \beta_j x_{t,j} \right)^2 + \lambda \sum_{j=1}^{p} \beta_j^2 \qquad (5.14)$$

Random Forest

Random Forest consists of several regression trees, a machine learning method that splits the predictor space (features) into distinct and non-overlapping regions (branches). When a regression tree predicts an observation, it locates the region that corresponds with the input features and takes the average of all observations in that region. The final predicted value for a Random Forest model is the average of all the predicted values from the decision trees. (James et al. 2017).

Random Forest has a built-in method for measuring feature importance. The condition on which the predictor space is divided in a regression tree is the standard deviation, called impurity. The Random Forest calculates how much the features contribute to reducing the weighted impurity, thereby providing a measure of feature importance. (James et al. 2017).

The Random Forest method is useful, being a multivariate method and being able to detect non-linear relationships. However, Random Forest may struggle with identifying the importance of correlated variables. Considering two correlated variables, X and Y, where both account for the same decrease in impurity, only the first of the variables used to perform a split would be associated with the decrease in impurity. The method could perform splits on the features on different levels in the different trees, thus misinterpreting the importance of each feature. (Simon, 2015).

In order to capture lead-lag relationships between the FFA price and the predictors, we will include lags from t-1 to t-20 in the filter methods. We will also investigate time lagged cross-correlation between the variables and the response variable, as another method for identifying lead-lag relationships, proposed by Cheong (2019).

The results of the filter method scheme and the time lagged cross-correlations will form the basis for different combinations to be tested with the LSTM and VAR methods. Additionally, features that intuitively could have explanatory power on the FFA prices will be included. Final testing of combinations with the LSTM and VAR methods ensures that the final subset of features provide the best predictions, corresponding to the method applied. The implication is that the filter methods might be unable to capture the same dependencies as the machine learning methods. The combinations tested with the highest predictive accuracy will constitute the final prediction models.

6   Results

6.1   Feature Selection and Final Hyperparameters

Included in appendix A.4 are the results for the 60 best performing features from the feature selection scheme introduced in Section 5.6. Appendix A.5 shows the results for the highest time lagged correlations, where the predictors have a lead-lag relationship with the response variable. The mean feature importance scores and the time lagged correlations provided a reference point when considering which features to include in the prediction models. Through testing numerous different feature combinations, well performing subsets were identified for both the VAR and LSTM models for the different forecasting horizons. Table 6.2 shows the features included in the best VAR models for the different horizons, and Table 6.3 shows the features included in the best LSTM models for the different horizons. As there are a lot of features in this study, Table 6.1 provides guidelines for interpreting the feature names.

Table 6.1 Explanation of feature name components.

Feature name component	Explanation
"count"	Count of vessels
"cap"	Capacity in DWT
"rel"	Relative share
"ex_regions"	Exporting world regions
"MA"	Moving average
"MS"	Moving sum
"cum"	Cumulative
"idle_nladen"	Idle non-laden
"stat"	Stationary
"std"	Standard deviation
"7", "30", "60"	Number of days used for MA or SM
No number at the end	Daily values

The name at the start of a feature refers to the world region, as shown in Figure 4.2, and features with no world region name refer to the global fleet.

Table 6.2 Subsets of Non-AIS, and AIS and Non-AIS features to be used in the final VAR models, for the different forecasting horizons.

One week forecasting horizon	Two weeks forecasting horizon	One month forecasting horizon
	Non-AIS features	
Iron Ore Spot Price Index	Iron Ore Spot Price Index	Iron Ore Spot Price Index
Brent Crude Oil	Brent Crude Oil	Brent Crude Oil
Gov 10Y yield	Gov 10Y yield	Gov 10Y yield
Average Spot 5TC	Average Spot 5TC	Average Spot 5TC
EUR/USD	EUR/USD	EUR/USD
	AIS and Non-AIS features	
Average Spot 5TC	Average Spot 5TC	Average Spot 5TC
3Month USD LIBOR	3Month USD LIBOR	3Month USD LIBOR
total_idle_nladen _MA_30	total_idle_nladen _MA_30	total_idle_nladen MA_60
stat_in_ex_regions_MA_60	stat_in_ex_regions_MA_60	arabian_gulf_rel_cap_MA_60
europe_rel_cap_MA_30	europe_rel_cap_MA_60	europe_rel_cap_MA_60
tonne_miles_demand_MS_30	tonne_miles_demand_MS_60	tonne_miles_demand_MS_60
indian_ocean_speed_MA_30	indian_ocean_speed_MA_30	indian_ocean_speed_MA_60
europe_speed_MA_60	europe_speed_MA_60	europe_speed_MA_60
laden_speed_std_MA_60	laden_speed_std_MA_60	cum_idle_nladen_time_MA_60
	med_sea _speed_MA_60	asia_draught_new_MA_60
		laden_speed_std_MA_60

Table 6.3 Subsets of Non-AIS, and AIS and Non-AIS to be used in the final LSTM models, for the different forecasting horizons.

One week forecasting horizon	Two weeks forecasting horizon	One month forecasting horizon
	Non-AIS features	
Iron Ore Spot Price Index	Iron Ore Spot Price Index	Iron Ore Spot Price Index
Brent Crude Oil	Brent Crude Oil	Brent Crude Oil
Gov 10Y yield	Gov 10Y yield	Gov 10Y yield
Average Spot 5TC	Average Spot 5TC	Average Spot 5TC
EUR/USD	EUR/USD	EUR/USD
	AIS and Non-AIS features	
Average Spot 5TC	Average Spot 5TC	Average Spot 5TC
3Month USD LIBOR	ballast_speed_std_MA_60	ballast_speed_std_MA_60
total_idle_nladen_MA_30	arabian_gulf_rel_cap_MA_60	cum_idle_nladen_time_MA_60
tonne_miles_demand_MS_30	europe_rel_cap_MA_60	arabian_gulf_rel_cap_MA_60
indian_ocean_speed_MA_30	tonne_miles_demand_MS_60	europe_rel_cap_MA_60
europe_speed_MA_60	indian_ocean_speed_MA_30	tonne_miles_demand_MS_60
laden_speed_std_MA_60	europe_speed_MA_60	indian_ocean_speed_MA_60
	laden_speed_std_MA_60	europe_speed_MA_60
	cum_idle_nladen_time_MA_7	asia_draught_MA_60
		laden_speed_std_MA_60

When looking at the final subsets of features, the VAR and LSTM models share a lot of the same features. The measures of relative capacity seem to give the most valuable information regarding the geographical distribution. Further, speed features also seem relevant, and it is interesting that the speed features including all vessels are used, instead of the speed features for moving vessels only. The speed and capacity features for Europe, the Indian Ocean and the Arabian Gulf seem to yield the best results. Regarding the proxies for operational status, total idle non-laden vessels, as well as stationary vessels in exporting world regions, seem to hold more predictive powers than the proxy named unemployed vessels. Tonne-mile demand is also present in all subsets, which may indicate that it provides useful information regarding the demand. An additional note is that several of the non-AIS features are removed in the subsets containing all available features, which may imply that the AIS-derived features are able to capture similar information. The inclusion of moving sums and moving averages seem to be helpful, and we note that the moving sums and moving averages have longer periods as the forecasting horizon increases.

The hyperparameters for the LSTM model are set on a trial and error basis. Multiple values and combinations were tested until we were unable to improve performance notably. The final model specifications are presented in Table 6.4.

Table 6.4 Final hyperparameters.

Hyperparameter	Value
Learning rate	0.01
Number of hidden layers	1
Number of neurons:	46
Number of Epochs	200
Rolling window size:	50
Batch size	200
Regularization	30% dropout regularization

6.2   Forecasting Results

Table 6.5 to 6.7 summarizes the results from the prediction models with and without AIS-derived features, for one week, two weeks, and one month, for all three maturities on the synthetic FFA price series. The rationale for using the synthetic price series for predictions, is that the actual prices jumps when the contracts roll, which will create noise for this purpose.

The Random Walk model should be considered as a lower bound benchmark in terms of predictive accuracy. Regarding directional accuracy, our lower bound benchmark is a coin toss, which theoretically will have a 50% chance of predicting the correct direction.

In general, we observe that some of the models perform well at predicting directions of price movements, but often struggle with the magnitude of the movements, when compared to the Random Walk benchmark model. We note that the VAR models with all features perform superior overall, particularly regarding the direction of movements. The results from the models with all features indicate that AIS data have predictive powers regarding the direction of price movements, when considering the DAR values. However, the VAR models with all features are only marginally better than the Random Walk when comparing RMSE values.

Seeing as the VAR models with all features are generally superior, we have investigated whether the improvements of the predictions are statistically significant. We have conducted a Diebold-Mariano test to see whether the seemingly improved prediction errors of the VAR models with all features compared to the Random Walk benchmark, are significant. To test whether the directional accuracies for the VAR models are better than a coin toss, we have performed a z-test, where we assume that the coin toss realizes its theoretical directional accuracy of 50%. In addition, we have carried out a z-test to see whether the directional accuracy is improved significantly when including AIS-derived features in the VAR models.

A brief introduction to the Diebold-Mariano test, followed by the test results are included in Appendix A.7. Similarly, a brief introduction to the z-test, followed by the results are included in Appendix A.6.

Table 6.5 forecasting results for the 1Q contract.

Horizon	Statistic	Random Walk	ARIMA	VAR Non-AIS	VAR All	LSTM Non-AIS	LSTM All
1 week	RMSE	864.48	1154.92	839.41	**828.14**	863.20	837.64
	MAE	640.00	903.91	635.00	**618.12**	647.86	634.25
	MAPE	4.59%	6.29%	4.52%	**4.39%**	4.64%	4.55%
	DAR	-	55.21 %	62.58 %	**65.64 %**	55.56 %	58.28 %
2 weeks	RMSE	1281.49	2082.26	1248.32	**1199.7**	1307.42	1238.3
	MAE	933.33	1647.96	934.86	**887.98**	971.72	906.53
	MAPE	6.82 %	11.61 %	6.81 %	**6.49 %**	7.12 %	6.68 %
	DAR	-	50.00 %	62.66 %	**69.62 %**	59.24 %	65.19 %
1 month	RMSE	1858.63	4106.9	1839.7	**1650.53**	2059.79	1737.99
	MAE	1488.12	3280	1496.13	**1268.6**	1652.84	1331.96
	MAPE	10.96 %	28.97 %	10.94 %	**9.37 %**	12.17 %	9.92 %
	DAR	-	49.32 %	54.79 %	**73.29 %**	40.69 %	65.75 %

Note: Boldface indicates the best result for each horizon

Table 6.6 forecasting results for the 2Q contract.

Horizon	Statistic	Random Walk	ARIMA	VAR Non-AIS	VAR All	LSTM Non-AIS	LSTM All
1 week	RMSE	636.07	911.34	610.96	**595.37**	636.67	613.67
	MAE	441.3	693.21	441.93	**423.2**	448.57	438.19
	MAPE	3.07 %	4.65 %	3.05 %	**2.93 %**	3.11 %	3.04 %
	DAR	-	49.08 %	63.19 %	**67.48 %**	56.79 %	58.28 %
2 weeks	RMSE	906.13	1654.59	901.18	**850.75**	931.31	893.2
	MAE	635.3	1287.71	625.58	**573.35**	639.01	597.81
	MAPE	4.45 %	8.52 %	4.36 %	**4.02 %**	4.48 %	4.20 %
	DAR	-	43.67 %	61.39 %	**67.72 %**	61.15 %	63.92 %
1 month	RMSE	1363.34	3187.23	1391.2	**1255.64**	2059.79	1303.8
	MAE	1006.75	2539.33	1053.93	**930.56**	1652.84	958.45
	MAPE	7.05 %	17.27 %	7.34 %	**6.52 %**	12.17 %	6.74 %
	DAR	-	40.41 %	51.37 %	**63.01 %**	46.21 %	60.96 %

Note: Boldface indicates the best result for each horizon

Table 6.7 forecasting results for the 1CAL contract.

Horizon	Statistic	Random Walk	ARIMA	VAR Non-AIS	VAR All	LSTM Non-AIS	LSTM All
1 week	RMSE	674.53	804.39	666.49	**648.62**	707.99	693.34
	MAE	**449.36**	594.34	470.16	454.36	468.63	464.09
	MAPE	2.57 %	3.31 %	2.65 %	**2.56 %**	2.66 %	2.63 %
	DAR	-	51.53 %	60.12 %	**62.58 %**	51.23 %	58.28 %
2 weeks	RMSE	1076.37	1427.25	1069.53	**1028.91**	1132.66	1074.92
	MAE	698.66	1055.97	726.72	672.31	748.69	**663.11**
	MAPE	4.04 %	5.81 %	4.16 %	3.87 %	4.31 %	**3.84 %**
	DAR	-	53.80 %	60.13 %	**65.19 %**	58.60 %	65.82 %
1 month	RMSE	1841.39	2719.19	1844.88	**1690.78**	2013.88	1767.95
	MAE	1295.58	2049.17	1387.47	**1186.9**	1443.54	1196.04
	MAPE	7.55 %	11.13 %	7.97 %	**6.89 %**	8.38 %	7.05 %
	DAR	-	56.85 %	42.47 %	**69.86 %**	40.00 %	69.18 %

Note: Boldface indicates the best result for each horizon

Figures 6.1 and 6.2 shows predicted and actual values for the synthetic price series, for the 1Q contract for the VAR model with all features, for the weekly and monthly forecasting horizons, respectively.
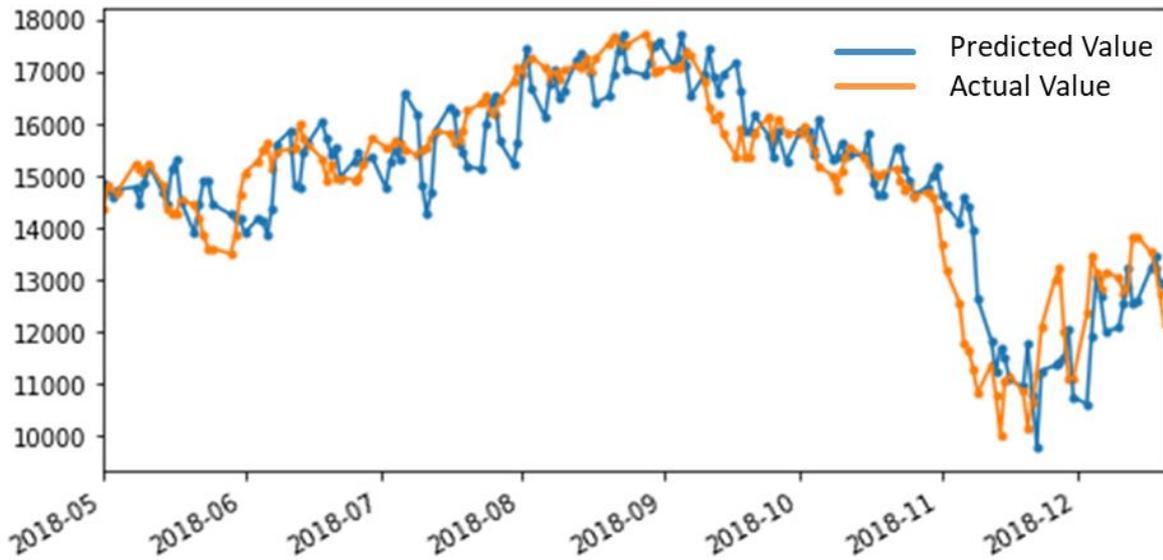
Figure 6.1 One week forecast LSTM with all features.
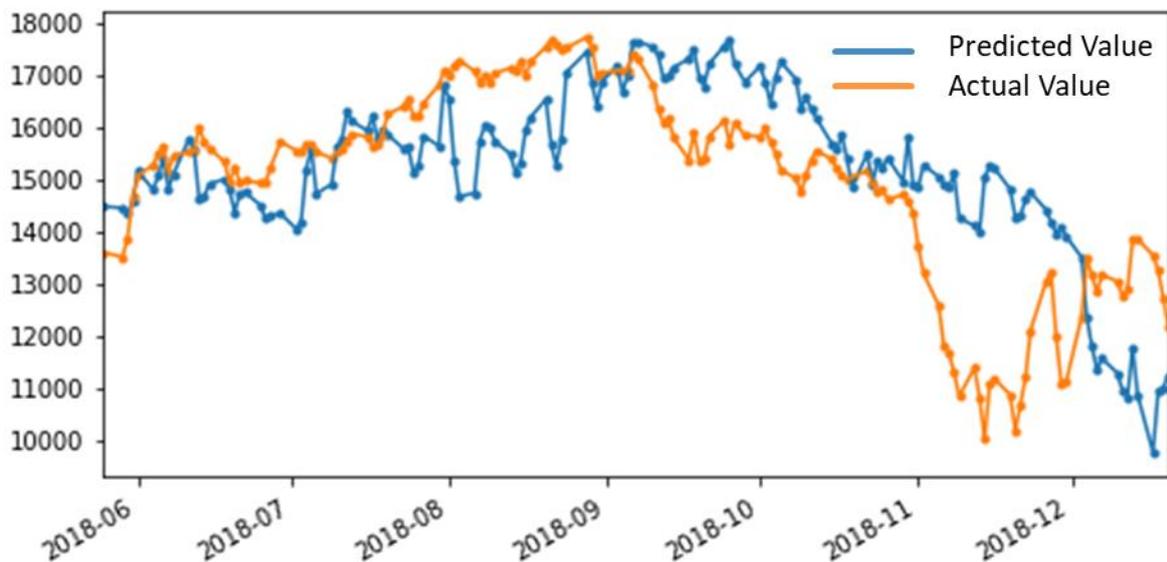
Figure 6.2 One month forecast for VAR with all features.

The VAR model without AIS features and the Random Walk model generate quite similar results regarding the magnitude across the different horizons and contracts. The results for the directional accuracy are varying, but the models have above 50% directional accuracy on eight

out of nine occasions. The VAR models with all features are superior overall, both regarding the directional accuracy and the magnitude of the predictions. When evaluating the Mean Squared Errors, they are not significantly different from the Random Walk model at the 5% level (see appendix A.7). In terms of the directional accuracy, the VAR models with all features have significantly better directional forecasts than the coin toss benchmark for all forecasting horizons and contracts, at the 5% level (see appendix A.6). The inclusion of AIS features increases the directional accuracy significantly for the all contracts on the monthly forecasting horizon (see appendix A.6).

The LSTM without AIS features performs relatively poorly regarding directional accuracy, as it fluctuates around 50% for the different contracts and horizons. Compared with the Random Walk benchmark, the LSTM without AIS features generally performs worse, particularly at the longer horizons. The inclusion of AIS features greatly increases the directional accuracy and reduces the prediction errors. However, the LSTM models with all features are still inferior to the VAR models with all features.

The ARIMA model performs poorly overall for predicting both the direction and magnitude and is consistently outperformed by the Random Walk model when comparing RMSE and MAE. Further, the directional accuracy fluctuates around 50%. The poor performance of the ARIMA model may be because the historic FFA prices alone might be a poor predictor for future developments.

The inclusion of AIS features improves the VAR and LSTM models, both regarding predictive accuracy and magnitude, especially at longer horizons. However, the results indicate that the models are generally poor at predicting the magnitude of movements, considering fairly high RMSE values. On the other hand, the results are promising regarding the directional accuracy for the LSTM and VAR, and the z-tests uncover partial evidence of the AIS features improving the directional accuracy, especially at longer horizons for the VAR models (see appendix A.6).

The price distribution for the testing period is characterized by a long positive trend, followed by a steep drop. It would be interesting to test the methodology in this study for different scenarios in the future. As the testing period is relatively short, it is difficult to test the robustness of the models sufficiently. Hence, the model framework is applied to several maturities.

The two best models, which are the LSTM and VAR with all features, will further be tested in simple trading strategies, based on the predicted direction, in the next section.

6.3 Trading Results

Based on predictions of movements on the synthetic FFA prices, we will be trading on the actual prices. We will test a long-short strategy for each forecasting horizon for all the three maturities. We perform a simulation over the testing period, where each strategy starts with the same initial cash balance. If the model predicts the prices to go up, a long position is taken, while a short position is taken if the model predicts the prices to go down. As the RMSE values for the best LSTM and VAR models are considerably high, we will not include any thresholds based on the predicted magnitude of movements. Thus, only the directional prediction is used as a basis for going long or short, and the strategies will always go long or short with all available funds. When the end of the forecasting horizon for a signal is reached, the position is liquidated, unless the current directional signal corresponds with the previously taken position. Consequently, if a long position is held at the last day of the forecasting horizon, and a new signal indicates continued price increase, the position will be held until the next horizon is reached. If the end of a long horizon is reached and the current signal indicates future price fall, the long position is first liquidated and a short position is simultaneously taken.

Each position is held throughout the duration of the forecast horizon. Thus, repositioning will take place weekly for the models with a weekly forecasting horizon, biweekly for models with a two-week forecasting horizon, and monthly for the models with a one month forecasting horizon. In addition, the positions are always netted two trading days before entering the settlement month. If a position is held during rolling of contracts, we will net out the position two days before the first contract expires and buy or sell new contracts at the first day the next contract is traded.

As benchmarks, we have included a buy-and-hold strategy, and a trend-following strategy (TF in the return table), supported by naive directional forecasts. The trend-following strategy takes positions at the same time as the other strategies, but trading signals are generated by the observed directional price movements in the period of the previous forecast horizon. For example, for the weekly repositioning scheme, a long position will be taken if there was an increase in prices in the last week. Similarly, the strategy looks at the historic price difference during the last two weeks for the biweekly repositioning scheme. Thus, this strategy simply follows the trend for the last period, and assumes this will continue for the next period.

We assume that broker commissions are fixed at 25 basis point and that we can buy and sell at the daily average prices. We will not take initial margin or variation margin into account.

Also, we will assume that earnings received as a result of mark-to-market in the clearing account are unavailable for trading. The calculations of returns are based on the initial nominal value and the nominal value at the end of the testing period. The trading results for the different trading strategies are shown in Table 6.8.

Table 6.8 Trading results.

1Q	Buy-and-Hold	Reposition weekly			Reposition biweekly			Reposition monthly		
		VAR	LSTM	TF	VAR	LSTM	TF	VAR	LSTM	TF
Annualized returns	-26.8%	**188.7%**	40.7%	-31.4%	71.8%	-37.4%	44.7%	140.3%	97.1%	74.7%
Annualized Std.	0.46	0.65	0.57	0.59	0.61	0.48	0.39	0.34	0.36	0.31
Sharpe ratio	-0.64	2.86	0.67	-0.58	1.14	-0.84	1.09	**4.06**	2.66	2.31
2Q										
Annualized returns	-11.81%	**108.2%**	12.4%	-43.7%	40.1%	-15.0%	25.2%	35.6%	-38.1%	-5.3%
Annualized Std.	0.34	0.52	0.43	0.45	0.50	0.37	0.31	0.30	0.30	0.30
Sharpe ratio	-0.43	**2.02**	0.22	-1.02	0.75	-0.48	0.72	1.09	-1.38	-0.27
1CAL										
Annualized returns	-20.9%	49.1%	-10.4%	-21.9%	36.8%	-25.8%	28.9%	**57.3%**	-25.5%	11.8%
Annualized Std.	0.27	0.45	0.32	0.46	0.42	0.30	0.28	0.20	0.26	0.28
Sharpe ratio	-0.87	1.03	-0.41	-0.54	0.82	-0.94	0.93	**2.76**	-1.09	0.32

Note:Boldface indicates best result for each contract

As can be seen in Table 6.8, there are vast differences in returns, volatility, and risk-adjusted returns for the different models across the different maturities and repositioning schedules. We note that the VAR model manages to generate excess returns compared to the benchmark models, and generates a positive return for all contracts for all repositioning schedules. It is especially impressive when repositioning weekly for the 1Q contract, as it generates an annualized return of 188.7%.

The LSTM model beats the buy-and-hold and trend-following benchmarks in four out of nine scenarios. The results indicate that the LSTM model works best when repositioning weekly. However, the LSTM model is consistently outperformed by the VAR model.

The trend-following strategy beats the buy-and-hold benchmark six out of nine times, and generates positive returns five out of nine times. It is interesting that it generally performs better than the LSTM model. However, it seems to struggle at shorter horizons where the returns are consistently worse than the buy-and-hold benchmark.

Overall, the results indicate that the VAR model is superior to both the buy-and-hold and trend-following benchmarks. It further shows superiority compared to the LSTM model for trading applications. The results are quite interesting, due to the VAR model consistently generating excess returns compared to the benchmarks. Further, it is interesting that the performance of the LSTM and VAR are so different, when the directional accuracies for both models were relatively decent for all contracts and horizons, as shown in Section 6.1. This may be due to the penalty of a wrongful directional prediction having a significant impact on the returns. One additional thing to point out regarding this issue, is that the models essentially generates daily directional forecasts, while the trading strategies are forced to hold the positions for the entire forecast period.

The results of Nomikos and Doctor (2013) gave implications against a weak form of market efficiency in the FFA market. Their outperforming strategies for the Capesize 1Q, 2Q and 1CAL FFA contracts, generated annualized returns of 327%, 254% and 199% respectively, with Sharpe ratios of 6.98, 6.72 and 6.13. These results are beyond our results for the similar contracts. However, it is hard to compare these results directly, due to the study horizon in their paper being much longer. Moreover, one should be careful to compare results from different time periods, as they have different characteristics. All of the buy-and-hold strategies yields negative returns in our case, but are positive in the study period of Nomikos & Doctor (2013). A final note regarding this issue, is that they applied strategies with dynamic holding periods based on a continuous input of signals, while our holding periods are rigid. However, our results also give implications against a weak form of market efficiency in the Capesize FFA market, and that there exist opportunities for generating excess returns by applying active trading strategies.

7   Concluding Remarks

The results from this study show that AIS-derived features have some predictive power regarding the directional movements of Capesize 5TC FFA prices. The increased directional accuracies for the VAR models with AIS features are statistically significant for the longer periods, while not significant for the shorter periods, compared to the VAR models without AIS data. When compared to the coin toss benchmark, the VAR models with AIS data have a significantly superior directional accuracy, whilst the RMSE values are not significantly different. These results imply that there are opportunities for forecasting future directional movements of FFA prices.

Among the features considered in this study, AIS features representing tonne-mile demand, geographical distribution of capacity, speed and idle non-laden ships, stand out as the most prominent features when forecasting future prices.

When using the predictions from the best VAR models in simple trading strategies, it outperformed both benchmarks for all three repositioning schedules, across the three maturities. However, the LSTM neural network also showed promising results in some cases. As discussed in Section 6.2, the results give implications against a weak form of efficiency in the FFA market, and that there are opportunities for generating excess returns by applying active trading strategies.

Regarding the performance of the LSTM, the results could possibly have been improved with more careful calibration of hyperparameters. One could also be more aware of the suitableness of the features used, with this kind of non-linear modeling framework. Further, both the LSTM and VAR models could have been improved by more scientifically determined geofencing. Several of the features are calculated based on the world regions and may therefore be sub-optimal.

Furthermore, the port areas cover relatively large areas, which may impair the classification of port calls, and in turn the features that rely on the port areas. The predictive powers of the created features are also dependent on the data quality. Differences in satellite coverage may for instance cause gaps in the data during the time period studied. There may also be other concerns regarding the AIS data quality, such as failures in the automatic message system, instances of incorrectly reported positions, as well as inconsistencies in manually reported data. Regarding feature selection, there is always the possibility that different combinations would have been better able to capture the underlying structure in the data.

The period for which our study covers is a little less than four years, and the number of observations may not be sufficient for the LSTM neural network models to learn the general underlying patterns. Furthermore, the freight market is highly cyclical, and the typical shipping cycle lasts about 7-8 years (Stopford, 2009). Ideally, the study period should have been longer. However, the AIS data quality has become increasingly better in recent years, and the period before and after the financial crisis has historically been an outlier for the freight market. Additionally, there are several alternatives regarding structuring and adjustment techniques of the FFA price data, which is used to develop the continuous time series.

Even though our results indicate that AIS-derived features are able to forecast directional movements of FFA prices to some extent, our forecasting period is too short to provide any robust conclusions regarding the long-term benefits and profitability. The trading strategies presented in this study are relatively simple, and the trading results could have been improved by using more sophisticated strategies. Also, the trading assumptions are simplified, which may give an unrealistic view of the potential profitability. We therefore recommend investigating the profitability on an extended horizon while making the trading conditions as realistic as possible. We also recommend further work on AIS data in relation to maritime economics.

8 Bibliography

Adland, R. &. (2018). Transportation Research Part E: Logistics and Transportation Review, Elsevier, vol. 118(C). *Explaining price differences between physical and derivative freight contracts*, 20-33.

Adland, R. (2019). Shipping Economics and analytics. *forthcoming chapter in Maritime Informatics, edited by D. Zissis and A. Artikis, Springer*.

Adland, R., Jia, H., & and Strandenes, S. (2017). Are aisbased trade volume estimates reliable? the case of crude oil exports. *Maritime Policy & Management, 44(5)*, 657-665.

Alexandridis, G., Sahoo, S., Song, D.-W., & Visvikis, I. D. (2017b). Shipping risk management practice revisited: A new portfolio approach. *Transport. Res. Part A: Policy*.

Alizadeh, A., Huang, C., & Dellen, S. (2015). A regime switching approach for hedging tanker shipping freight rates. *Energy Economics 49*, 44-59.

Angelov, P. P., & Gu, X. (2019). *Empirical Approach to Machine Learning*. Cham: Springer.

Baltic Exchange. (2019). *www.balticexchange.com*. Retrieved 12 17, 2019, from What is an FFA?: https://www.balticexchange.com/ffa/what-is-an-ffa/

Batchelor, R., Alizadeh, A., & Visvikis, I. (2007). Forecasting spot and forward prices in the. *International Journal of Forecasting 23* , 101-114.

Bengio, Y., Simard, P., & Frasconi, P. (1994). Learning Long-Term Dependencies with Gradient Descent is Difficult. *IEEE: Transactions On Neural Networks 5(2)*, 157-166.

Bessler, W., Drobetz, W., & Seidel, J. (2008). Ship Funds as a New Asset Class: An Empirical Analysis of the Relationship between Spot and Forward Prices in Freight Markets. *Journal of Asset Management 9*, 102–120.

Brownlee, J. (2016, 12 19). *How To Backtest Machine Learning Models for Time Series Forecasting*. Retrieved from machinelearningmastery.com: https://machinelearningmastery.com/backtest-machine-learning-models-time-series-forecasting/

Brownlee, J. (2017, 08 18). *Stacked Long Short-Term Memory Networks*. Retrieved from machinelearningmastery.com/: https://machinelearningmastery.com/stacked-long-short-term-memory-networks/

Brownlee, J. (2017b, 05 05). *How to Use Weight Regularization with LSTM Networks for Time Series Forecasting*. Retrieved from machinelearningmastery.com: https://machinelearningmastery.com/use-weight-regularization-lstm-networks-time-series-forecasting/

Brownlee, J. (2018, 06 20). *Difference Between a Batch and an Epoch in a Neural Network*. Retrieved from https://machinelearningmastery.com/: https://machinelearningmastery.com/difference-between-a-batch-and-an-epoch/

Brownlee, J. (2019, 08 05). *How to Scale Data for Long Short-Term Memory Networks in Python*. Retrieved from machinelearningmastery.com: https://machinelearningmastery.com/how-to-scale-data-for-long-short-term-memory-networks-in-python/

Brownlee, J. (2019b, 01 25). *Understand the Impact of Learning Rate on Neural Network Performance*. Retrieved from machinelearningmastery.com: https://machinelearningmastery.com/understand-the-dynamics-of-learning-rate-on-deep-learning-neural-networks/

Bunker Index. (n.d.). *Bunkerindex.com*. Retrieved from https://www.bunkerindex.com/prices/indices.php?indices_id=1

Cheong, J. H. (2019, 05 13). *Four ways to quantify synchrony between time series data*. Retrieved from towardsdatascience.com: https://towardsdatascience.com/four-ways-to-quantify-synchrony-between-time-series-data-b99136c4a9c9

Clarksons Research. (2019). London: Shipping Intelligence Network.

Cover, T. M., & Thomas, J. A. (1991). *Elements of Information Theory.* New Jersey: John Wiley & Sons.

Da, Z., Jagannathan, R., & Jianfeng, S. (2015). Growth Expectations, Dividend Yields, and Future Stock Returns ∗. *School of Banking and Finance, UNSW Australia, UNSW*.

Diebold, F., & Mariano, R. (1995, 06). Comparing Predictive Accuracy. *Journal of Business & Economic Statistics, Vol. 13, No. 3,*, pp. 253-263.

Fama, F. (1991). Efficient capital market II. *Journal of Finance, 46, 5* , 1575-1617.

Fan, S., Ji, T., Gordon, W., & Rickard, a. (2013). Forecasting baltic dirty tanker index by applying wavelet neural networks. *Journal of Transportation Technologies, 3(01)*, 68-87.

Fischer, T., & Krauss, C. (2018). Deep learning with long short-term memory networks for financial market predictions. *European Journal of Operational Research, 270 (2),* , 654-669.

Haykin, S. (2009). *Neural Networks and Learning Machines Third Editio.* New Jersey: Pearson Education.

Herrera, G. P., Constantino, M., Tabak, B. M., Pistori, H., Su, J.-J., & Naranpanawa, A. (2019). Long-term forecast of energy commodities price using machine learning. *Energy Vol 179*, 214-221.

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, IEEE Press, (2001)*.

Huang, S.-C., & Wu, C.-F. (2018). Energy Commodity Price Forecasting with Deep Multiple Kernel Learning. . *Energies. 11. 3029.*

Hyndman, R., & Athanasopoulos, G. (2013). *Forecasting: Principles and Practice Chapter 3.4.* Melbourne: OTexts. Retrieved from otexts.com: https://otexts.com/fpp2/accuracy.html

James, G., Witten, D., Hasite, T., & Tibsirani, R. (2017). *An Introduction to Statistical Learning with Applications in R.* New York: Springer.

James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). *An Introduction to Statistical Learning, volume 103 of Springer Texts in Statistics.* New York: Springer New York.

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, Chios, 24-27 June.*

Kaluza, P., Kölzsch, A., Gastner, T. M., & Blasius, B. (2010). The complex network of global cargo ship movements. *Journal of the Royal Society, Interface*, , 7(48):1093–103.

Kasimati, E., & Veraros, N. (2017). Accuracy of forward freight agreements in forecasting future freight rates . *Applied Economics. 50*, 1-14.

Kavussanos M, G., Visvikis, & Bachelor, I. (2004). Over-the-counter forward contracts and spot price volatility in shipping. *Transportation Research Part E: Logistics and Transportation Review Volume 40, Issue 4*, 273-296.

Kavussanos, G. M., & Visvikis, I. (2006). Shipping freight derivatives: A survey of recent evidence. *Maritime Policy & Management 33(3)*, 233-255.

Kavussanos, M. G., Visvikis, I., & Menachof, D. (2004b). The unbiasedness hypothesis in the freight forward market: Evidence from cointegration tests. *Rev. Deriv. Res. 7 (3)*, 241–266.

Kavussanos, M., & Visvikis, I. (2004). Market interactions in returns and volatilities between spot and forward shipping freight markets. *J. Bank. Finance 28 (8)*, 2015–2049.

Kavussanos, M., Visvikis, I., & Dimitrakopoulos, D. (2010). Information linkages between Panamax freight derivatives and commodity derivatives markets. *Maritime Econ. Logist. 12 (1)*, 91–110.

Kavussanos, M., Visvikis, I., & Dimitrakopoulos, D. (2014a). Economic spillovers between related derivatives markets: the case of commodity and freight markets. *Transport. Res. Part E: Logist. Transport. Rev. 68*, 79–102.

Keller, G. (2018). *Statistics for Management and Economics.* Cengage Learning, Inc.

Kinney, J. B., & Atwal, G. S. (2014 , 01 21). Equitability, mutual information, and the maximal information coefficient. *PNAS March 4, 2014 111 (9)* , pp. 3354-3359.

Leoni, F. (2019, August 19). *Navigating the Hyperparameter Space*. Retrieved from Towards Data Science: https://towardsdatascience.com/navigating-the-hyperparameter-space-32f48be2c12d

Lyridis, D., Zacharioudakis, P., Mitrou, P., & Mylonas, A. (2004). Forecasting tanker market using artificial neural networks. *Maritime Economics & Logistics, 6(2)*, 93–108.

Marine Traffic. (2018). *What is the Automatic Identification System (AIS)?* Retrieved from Marinetraffic.com: https://help.marinetraffic.com/hc/en-us/articles/204581828-What-is-the-Automatic-Identification-System-AIS-

Masteika, S., Rutkauskas, V. A., & Alexander, J. (2012). Continuous futures data series for back testing and technical analysis. *2012 International Conference on Economics, Business and Marketing Management IPEDR vol.29.*

Narkawicz, A., & Hagen, G. (2016, 10 26). *Algorithms for Collision Detection Between a Point and a Moving Polygon, with Applications to Aircraft Weather Avoidance.* Retrieved from ntrs.nasa.gov/: https://ntrs.nasa.gov/archive/nasa/casi.ntrs.nasa.gov/20160010026.pdf

Nomikos, N., & Doctor, K. (2013). Economic significance of market timing rules in the Forward. *Transportation Research Part E 52*, 77-93.

Næss, A. P. (2018). Investigation of Multivariate Freight Rate Prediction Using Machine Learning and AIS Data. *Master Thesis, Norwegian University of Science and Technology.*

Olah, C. (2015, August 27). *Understanding LSTM Networks.* Retrieved from Colah's blog: https://colah.github.io/posts/2015-08-Understanding-LSTMs/

Regli, F., & Nomikos, N. K. (2019). The Eye in the Sky: Freight Rate Effects of Tanker Supply. *Transportation Research. Part E: Logistics and Transportation Review, 125*, 402-424.

Reshef, D., Reshef, Y., Finucane, H., Grossman, S., McVean, G., Turnbaugh, P., . . . Sabeti, P. (2011). Detecting Novel Associations in Large Data Sets. *Science (New York, N.Y.). 334*, 1518-24.

Salen, S., & Århus, G. (2018). Predicting Shipping Freight Rate Movements Using Recurrent Neural Networks and AIS Data On the tanker route between the Arabian Gulf and Singapore. *NHH Master Thesis.*

Scheucher, W. (2016, 11 11). *Vincenty's formulae (Direct & Inverse Calculation).* Retrieved from blog.solitec.com/: https://blog.solitec.com/vincentys-formulae-direct-inverse-calculation/

Schmitz, T. (2016). *European Energy Exchange.* Retrieved from EEX Group Global Commodities: http://www.eex.com/blob/69998/c66f581ab4a50f8b4b29d6dae54852bb/eex-group---global-commodities-product-portfolio-data.pdf

Simon, C. (2015, 01 25). *Feature importance in random forests when features are correlated.* Retrieved from corysimon.github.io: http://corysimon.github.io/articles/feature-importance-in-random-forests-when-features-are-correlated/

Skauen, A. N. (2015). Quantifying the tracking capability of space-based AIS systems. *Advances in Space Research 57 (2016)* , 527–542.

Spiliopoulos, G., Zissis, D., & Chatzikokolakis, K. (2017). A Big Data Driven Approach to Extracting Global Trade Patterns. In *Lecture Notes in Computer Science, vol 10731* (pp. 109-121). New York: Springer Science & Business Media. Retrieved from Researchgate.

Stopford, M. (2009). *Maritime Economics.* London: Routledge.

Tsioumas, V. (2016). *Quantitative analysis of the dry bulk freight market.* Athens: Ph.D. Thesis Department of Maritime Studies University of Piraeus.

Tsioumas, V., & Papadimitriou, S. (2018). The dynamic relationship between freight markets and commodity prices revealed. *Maritime Economics & Logistics 20(2)*, 267-279.

Wilson, R. (2013). The Principles of the Dry Bulk FFA Market. Institutional Knowledge at Singapore Management University. Research Collection Lee Kong Chian School Of. *Singapore Management University*, https://ink.library.smu.edu.sg/cgi/viewcontent.cgi?article=4570&context=lkcsb_research.

Wint, R. (2019, 01 02). *On Maximal Information Coefficient: A Modern Approach for Finding Associations in Large Data sets.* Retrieved from medium.com: https://medium.com/@rhondenewint93/on-maximal-information-coefficient-a-modern-approach-for-finding-associations-in-large-data-sets-ba8c36ebb96b

Witsung, T. H. (2019). *Witsung/Python-Ports-Distance-Calculator*. Retrieved from github.com: https://github.com/Witsung/Python-Ports-Distance-Calculator

Wu, L., Xu, Y., Wang, Q., Wang, F., & Xu, Z. (2017). Mapping Global Shipping Density from AIS Data. *Journal of Navigation, 70(01)*, 67-81.

Yin, J., Luo, M., & Fan, L. (2017). Dynamics and interactions between spot and forward freights in the dry bulk shipping market. *Maritime Policy Manage. 44 (2)*, 271–288.

Zhang, J., Zeng, Q., & Zhao, X. (2014, 10). Forecasting spot freight rates based on forward freight agreement and time charter contract. *Applied Economics 46(29)*, pp. 3639-3648.

Zheng, S., & Chen, S. (2018). Modeling the determinants of dry bulk FFA trading volume from a cross-market perspective of spot and forward. *Maritime Business Review. 3, Vol. 3*(No. 3), 256-275.

Zivot, E., & Wang, J. (2006). *Modeling Financial Time Series* . New York: Springer.

A Appendix

A.1 Port Areas
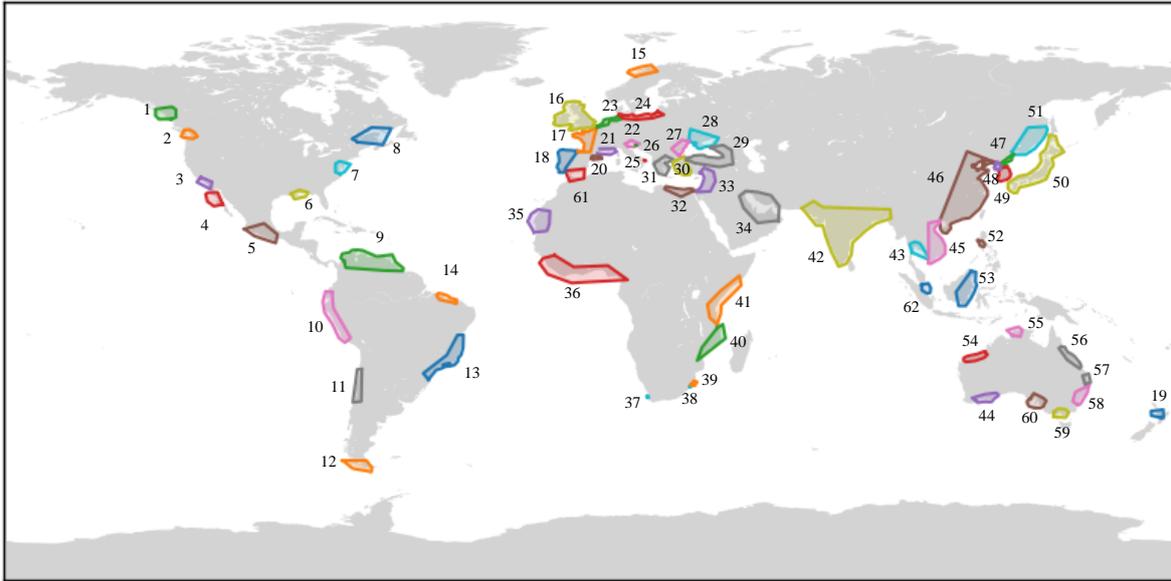
Figure A.1 Indexed port areas.

Table A.1 Labeling for defined export and import port areas.

Import		Export		Other ports
5	31	1	36	39
15	32	2	37	61
16	33	3	38	62
17	34	4	40	
18	35	6	44	
20	41	7	51	
21	42	8	52	
22	43	9	53	
23	45	10	54	
24	46	11	55	
25	47	12	56	
27	48	13	57	
29	49	14	58	
30	50	19	59	
		28	60	

A.2 Descripitive Statistics

Descriptive statistics of price data and features used for predictions

Table A.2 Synthetic FFA prices descriptive statistics.

Contract	Mean	Std.	Min	Max
1QA	10755.200	5537.768	3162.662	28624.98
2QA	10793.760	3984.674	5285.655	22103.73
1CALA	14571.470	5320.450	7701.678	30811.95

Table A.3 Actual FFA prices descriptive statistics.

	Mean	Std.	Min	Max
1Q	13033.74	5704.58	3547.00	28022.00
2Q	12466.74	5088.95	4854.00	31210.00
1CAL	13540.00	4142.19	7246.00	23805.00

Table A.4 Non-Ais features descriptive statistics.

Feature	Mean	Std.	Min	Max
Iron Ore Spot Price Index	65.888	12.990	37.5	102.1
Brent Crude Oil	61.927	18.546	27.88	115.06
Gov 10Y yield	2.316	0.4147	1.357	'3.237
3Month USD LIBOR	1.0199	0.773	0.223	2.823
Average Spot 5TC	12117.392	6215.523	1985	30475
EUR/USD	1.157	0.077	1.038	'1.392

Table A.5 AIS features descriptive statistics.

Feature	Mean	Std.	Min	Max
cum_ idle_nladen_time_MA_7	148.163	14.276	35.071	189.385
cum_ idle_nladen_time_MA_60	146.364	16.970	34.538	175.241
asia_draught_new_MA_60	0.727	0.008	0.709	0.747
laden_speed_std_MA_60	6.628	1.366	5.317	12.888
total_idle_nladen_MA_30	244.101	26.264	47.133	318.933
total_idle_nladen_MA_60	242.292	29.104	46.200	301.967
arabian_gulf_rel_cap_MA_60	0.022	0.004	0.015	0.033
europe_rel_cap_MA_60	0.045	0.007	0.032	0.083
europe_rel_cap_MA_30	0.044	0.007	0.031	0.080
tonne_miles_new_MS_30	698.082	77.907	425.268	854.348
tonne_miles_new_MS_60	1388.878	156.554	861.379	1667.321
ballast_speed_std_MA_60	6.799	0.579	6.062	9.178
indian_ocean_speed_MA_30	11.391	0.318	10.493	12.701
europe_speed_MA_60	7.762	0.505	6.576	10.494
ndian_ocean_speed_MA_60	11.391	0.273	10.771	12.197
med_sea_speed_MA_60	6.219	0.907	4.265	9.891
stat_in_ex_regions_MA_60	117.238	14.531	16.167	147.633

A.3 ADF Tests

Table A.6 ADF tests for all FFAs and features after first difference.

Feature	ADF Statistic	p-value	1 %	Accept H0 at1%
cum_ idle_nladen_time_MA_7	-13.23	9.64e-25	-3.44	FALSE
cum_ idle_nladen_time_MA_60	-6.35	2.67e-08	-3.44	FALSE
asia_draught_new_MA_60	-10.32	2.95e-18	-3.44	FALSE
laden_speed_std_MA_60	-33.75	0.0	-3.44	FALSE
total_idle_nladen_MA_30	-10.44	1.52e-18	-3.44	FALSE
total_idle_nladen_MA_60	-6.26	4.14e-08	-3.44	FALSE
3Month USD LIBOR	-4.17	0.0	-3.44	FALSE
arabian_gulf_rel_cap_MA_60	-6.05	1.29e-07	-3.44	FALSE
europe_rel_cap_MA_60	-5.47	2.45e-06	-3.44	FALSE
europe_rel_cap_MA_30	-6.85	1.74e-09	-3.44	FALSE
tonne_miles_new_MS_30	-10.20	6.01e-18	-3.44	FALSE
tonne_miles_new_MS_60	-33.78	0.0	-3.44	FALSE
ballast_speed_std_MA_60	-34.37	0.0	-3.44	FALSE
Average Spot 5TC	-17.52	4.27e-30	-3.44	FALSE
indian_ocean_speed_MA_30	-9.26	1.38e-15	-3.44	FALSE

europe_speed_MA_60	-5.74	6.41e-07	-3.44	FALSE
indian_ocean_speed_MA_60	-7.04	5.83e-10	-3.44	FALSE
med_sea_speed_MA_60	-7.01	6.83e-10	-3.44	FALSE
stat_in_ex_regions_MA_60	-6.07	1.15e-07	-3.44	FALSE
Gov 10Y yield	-13.44	3.79e-25	-3.44	FALSE
EUR/USD	-35.33	0.0	-3.44	FALSE
Brent Crude Oil	-35.73	0.0	-3.44	FALSE
Iron Ore Spot Price Index	-16.93	9.81e-30	-3.44	FALSE
1QA	-25.80	0.0	-3.44	FALSE
2QA	-25.80	0.0	-3.44	FALSE
1CALA	-15.88	8.83e-29	-3.44	FALSE

A.4 Feature Selection Results

Table A.7 Feature selection results for top 60 features in terms of the mean score.

Feature	Linear Reg.	Ridge Reg.	Lasso	RF	Linear Corr.	MIC	Mean Score
daily_tot_laden(t)	1	1	0.03	1	1	1	0.84
asia_ballast_new(t)	0.83	0.83	1	0.22	0.34	0.47	0.61
asia_draught_new(t)	0.93	0.93	0.01	0.32	0.58	0.71	0.58
aus_ballast_new(t)	0.42	0.42	0.72	0.27	0.28	0.53	0.44
asia_laden(t)	0.68	0.68	0	0.01	0.39	0.62	0.4
south_america_ballast_new(t)	0.59	0.59	0.42	0.09	0.28	0.43	0.4
south_africa_ballast_new(t)	0.49	0.49	0.49	0	0.15	0.34	0.33
indian_ocean_ballast_new(t)	0.42	0.42	0.5	0.04	0.16	0.33	0.31
south_africa_draught_new(t)	0.49	0.49	0.01	0.21	0.18	0.26	0.27
asia_flow(t)	0.37	0.37	0	0	0.21	0.46	0.24
europe_ballast_new(t)	0.37	0.37	0.21	0	0.07	0.28	0.22
indian_ocean_draught_new(t)	0.44	0.44	0.03	0.02	0.14	0.28	0.22
south_america_draught_new(t)	0.39	0.39	0	0	0.16	0.36	0.22
rel_daily_tot_ballast_MA_7(t)	0.25	0.25	0	0.01	0.26	0.5	0.21
south_america_laden(t)	0.42	0.42	0	0	0.13	0.28	0.21
aus_flow(t)	0.31	0.31	0	0	0.17	0.39	0.2
aus_rel_count(t)	0.33	0.33	0.07	0.02	0.11	0.36	0.2
aus_rel_cap(t)	0.32	0.32	0.01	0.01	0.12	0.36	0.19

daily_tot_fleet(t)	0.18	0.18	0.56	0	0.01	0.15	0.18
aus_cap(t)	0.28	0.28	0	0	0.1	0.33	0.17
south_africa_laden(t)	0.33	0.33	0	0	0.1	0.24	0.17
aus_count(t)	0.29	0.29	0	0	0.1	0.31	0.16
aus_draught_new(t)	0.28	0.28	0	0	0.07	0.33	0.16
aus_laden(t)	0.3	0.3	0.04	0	0.05	0.26	0.16
pacific_ocean_ballast_new(t)	0.29	0.29	0.15	0	0.09	0.17	0.16
europe_draught_new(t)	0.27	0.27	0	0	0.08	0.28	0.15
north_america_ballast_new(t)	0.27	0.27	0.2	0	0.02	0.16	0.15
asia_ballast_new_MA_7(t)	0.31	0.31	0	0	0.08	0.15	0.14
pacific_ocean_draught_new(t)	0.26	0.26	0.01	0.04	0.1	0.2	0.14
aus_rel_cap(t-1)	0.17	0.17	0	0	0.08	0.35	0.13
aus_rel_count(t-1)	0.16	0.17	0.01	0	0.08	0.36	0.13
europe_laden(t)	0.29	0.29	0	0	0.04	0.16	0.13
indian_ocean_rel_count(t)	0.23	0.23	0.02	0.01	0.04	0.23	0.13
south_america_rel_count(t)	0.25	0.25	0	0	0.05	0.17	0.12
asia_draught_new_MA_7(t)	0.17	0.17	0	0	0.09	0.23	0.11
asia_rel_cap(t)	0.24	0.24	0	0.01	0	0.18	0.11
aus_ballast_new(t-1)	0.17	0.17	0	0	0.06	0.24	0.11
aus_cap(t-1)	0.17	0.17	0	0	0.07	0.26	0.11
aus_count(t-1)	0.17	0.17	0	0	0.07	0.25	0.11
indian_ocean_count(t)	0.19	0.19	0	0	0.03	0.27	0.11
indian_ocean_moving(t)	0.2	0.2	0	0	0.03	0.21	0.11
indian_ocean_rel_cap(t)	0.2	0.2	0	0	0.03	0.24	0.11
rel_total_idle_nladen (t)	0.17	0.17	0	0	0.07	0.26	0.11
south_america_flow(t)	0.22	0.22	0	0	0.04	0.18	0.11
west_past_singapore(t)	0.19	0.19	0	0	0.05	0.26	0.11
arabian_gulf_ballast_new(t)	0.19	0.19	0.16	0	0.01	0.08	0.1
asia_ballast_new_MA_30(t)	0.16	0.16	0	0	0.07	0.21	0.1
asia_flow_MS_30(t)	0.06	0.07	0	0.01	0.1	0.36	0.1
asia_rel_count(t)	0.22	0.22	0	0	0	0.17	0.1
aus_flow_MS_30(t)	0.09	0.09	0	0	0.1	0.3	0.1
daily_tot_laden(t-1)	0.14	0.14	0	0	0.03	0.28	0.1
indian_ocean_laden(t)	0.21	0.21	0	0	0.03	0.18	0.1
rel_daily_tot_ballast(t-1)	0.04	0.04	0	0.02	0.17	0.35	0.1
south_africa_rel_count(t)	0.17	0.17	0	0	0.01	0.22	0.1
south_america_count(t)	0.22	0.22	0	0	0.04	0.13	0.1
south_america_flow_MS_60(t)	0.21	0.21	0	0	0.05	0.13	0.1
south_america_rel_cap(t)	0.22	0.22	0	0	0.04	0.13	0.1
west_past_south_africa(t)	0.19	0.19	0	0	0.02	0.21	0.1
west_past_south_africa_cap(t)	0.15	0.15	0	0	0.02	0.26	0.1
EUR/USD(t-12)	0.14	0.14	0	0	0.01	0.24	0.09

A.5 Cross Correlation Results

Table A.8 Cross correlation results for best performing features leading the FFA prices.

Feature	Offset	Correlation
1QA	0	1
Brent Crude Oil	-1	0.801909
Baltic dry index	-46	0.762259
Average Spot 5TC	-49	0.705862
Gov 10Y yield	-29	0.702615
Baltic Capesize index	-46	0.697154
europe_moving_speed_MA_60	-39	0.669477
Iron Ore Spot Price Index	-28	0.65252
europe_moving_speed_MA_30	-45	0.601482
asia_draught_new_MA_60	-37	0.585185
asia_draught_new_MA_30	-35	0.53525
med_sea_speed_MA_60	-36	0.48589
med_sea_speed_MA_30	-36	0.453895
arabian_gulf_speed_std_MA_60	-8	0.44308
med_sea_speed_std_MA_60	-49	0.387417
arabian_gulf_speed_std_MA_30	-29	0.369876
med_sea_speed_MA_7	-49	0.364721
med_sea_speed_std_MA_30	-13	0.355409
pacific_ocean_ballast_MA_60	-25	0.329168
europe_speed	-12	0.32046
unloading_ports_outgoing_MA_30	-50	0.309906
europe_cap_MA_30	-2	-0.56029
europe_laden_MA_30	-3	-0.5059
europe_count_MA_7	-6	-0.49339
europe_cap_MA_7	-15	-0.48965
europe_rel_count_MA_30	-17	-0.48062
europe_rel_cap_MA_30	-23	-0.4761
europe_rel_cap_MA_60	-3	-0.47157
europe_count	-7	-0.45853
europe_rel_cap_MA_7	-32	-0.45415
europe_moving_MA_30	-3	-0.44937
pacific_ocean_laden_MA_60	-47	-0.4408
indian_ocean_speed_std_MA_60	-54	-0.43819
pacific_ocean_laden_MA_30	-53	-0.41399
med_sea_ballast_MA_60	-17	-0.39059

A.6 Z-Test for Directional Accuracy

$H_0: p_1 = p_2$

$H_a: p_1 > p_2$

$$z = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\hat{p}(1-\hat{p})(\frac{1}{n1} + \frac{1}{n2})}} \qquad \text{(A.1)}$$

Where $\hat{p}_1$ and $\hat{p}_2$ are the sample means, n_1 and n_2 are the sample sizes. The null Hypothesis is rejected if $|Z| > z_{1-\alpha}, v$, where $t_{1-\alpha}$ represents critical value of the z distribution degrees, as we are conducting a one-sided test. (Keller, 2018).

Table A.9 One-sided z-test investigating if the directional accuracy of the VAR model with all features is significantly better than the coin toss benchmark.

		10%	5%	1%
	Z statistic	1.282	1.645	2.326
1Q	2.85984143	TRUE	TRUE	TRUE
	3.55690669	TRUE	TRUE	TRUE
	4.09190063	TRUE	TRUE	TRUE
2Q	3.20632934	TRUE	TRUE	TRUE
	3.20090971	TRUE	TRUE	TRUE
	2.24285498	TRUE	TRUE	FALSE
1CAL	2.28895468	TRUE	TRUE	FALSE
	2.7319119	TRUE	TRUE	TRUE
	3.46319917	TRUE	TRUE	TRUE

Table A.10 One-sided z-test investigating if the directional accuracy of the VAR model with all features is significantly better than VAR model without AIS features.

		10%	5%	1%
	Z statistic	1.282	1.645	2.326
1Q	0.57731485	FALSE	FALSE	FALSE
	1.30758881	TRUE	FALSE	FALSE
	3.29260579	TRUE	TRUE	TRUE
2Q	0.81466267	FALSE	FALSE	FALSE
	1.17603283	FALSE	FALSE	FALSE
	2.01060661	TRUE	TRUE	FALSE
1CAL	0.45495508	FALSE	FALSE	FALSE
	0.93037906	FALSE	FALSE	FALSE
	4.71763686	TRUE	TRUE	TRUE

A.7 Diebold-Mariano Test For Predictive Accuracy

$H_0: E(d_t) = 0 \; \forall t$ (Same accuracy for the two forecasts)

$H_0: E(d_t) \neq 0$ (different level of accuracy for the two forecasts)

(Diebold & Mariano, 1995)

Table A.11 Diebold-Mariano Test evaluating if the MSE values from the VAR model with all features and from Random Walk models are significantly different.

	P value	10% level	5% level
1Q			
1 Week	0.35676	FALSE	FALSE
2 Weeks	0.07602	TRUE	FALSE
1 Month	0.08931	TRUE	FALSE
2Q			
1 Week	0.39890	FALSE	FALSE
2 Weeks	0.06895	TRUE	FALSE
1 Month	0.23285	FALSE	FALSE
1CAL			

1 Week	0.55146	FALSE	FALSE
2 Weeks	0.16440	FALSE	FALSE
1 Month	0.34990	FALSE	FALSE

A.8 LSTM

LSTM networks have a similar looped structure as recurrent neural networks, but the repeated module in a general recurrent neural network contains one layer, while the repeated module in an LSTM contains four layers that interact, visualized by the yellow rectangles in figure A.2. (Olah, 2015). The upper horizontal line of figure A.2 is the cell state C_t, where the information flow from the previous cell state is regulated by gates.
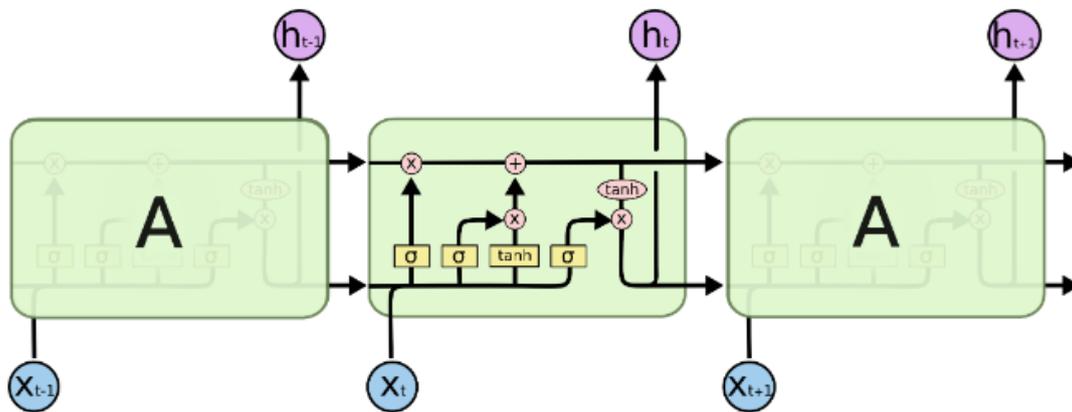
Figure A.2 illustration of the repeated module in LSTM (Olah, 2015).

An LSTM contains three gates as ways to optionally let information through. The gates are composed of a sigmoid neural net layer and a stepwise multiplication operation. The sigmoid layer outputs numbers in the range [0,1], and decides how much of each component to be let through. The first step is going through the forget gate f_t, shown in Figure A.2. The forget gate determines what information to be removed from the cell state. It has an output range between [0,1], depending on the previous inputs, maintained in the previous hidden state, h_{t-1}, and the current inputs, x_t. (Olah, 2015). The process is shown in figure A.3, and the calculation is shown in Equation A.2.
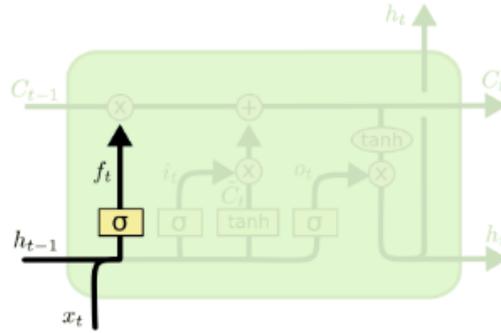
Figure A.3 Selection of information to be kept in the cell state: (Olah, 2015).

$$f_t = \sigma\,(W_f \cdot [h_{t-1}, x_t] + b_f) \tag{A.2}$$

The second step is determining which information to store in the cell state. First, a sigmoid layer, called the "input gate layer", determines which values to update, i_t. Second, a tanh layer creates a vector containing potential values for the current cell state, $\tilde{\tilde{C}}_t$. The cell state is then updated by adding $i_t \cdot \tilde{C}_t$. (Olah, 2015). The process is shown in Figure A.4 and the calculations are shown in equation A.3 and A.4. The entire process of arriving at the current cell state, C_t, is shown in Figure A.5, and the calculation is shown in Equation A.5.
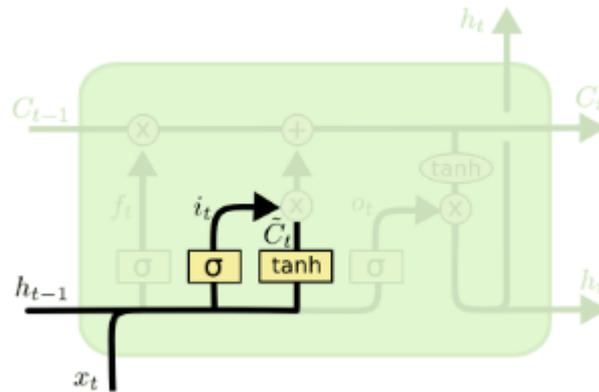

Figure.A.4 Storing of new information in the cell state (Olah, 2015).

$$i_t = \sigma\,(W_i * [h_{t-1}, x_t] + b_i) \tag{A.3}$$

$$\tilde{C}_t = \tanh(W_C * [h_{t-1}, x_t] + b_C) \tag{A.4}$$
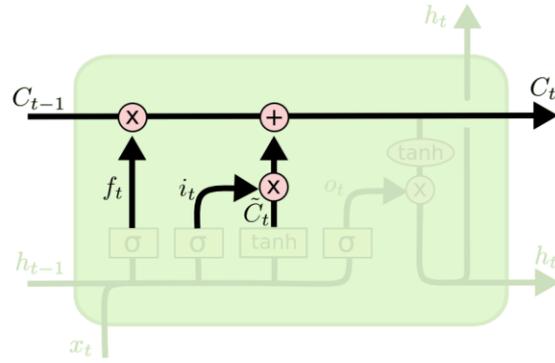
Figure.A.5 Updating the cell state (Olah, 2015).

$$C_t = (f_t * C_{t-1} + i_t * \tilde{C}_t) \qquad (A.5)$$

The third and final step is determining the output from the cell state, C_t. A sigmoid layer first decides which information from the cell state that will be used as output. Next, the cell state is passed through *tanh* (compressing the values in the range [-1,1]), and multiplied with the output from the sigmoid gate. This ensures that the values for output are the ones decided. (Olah, 2015). The process is shown in figure A.6 and the calculations are shown in equations A.6 and A.7.
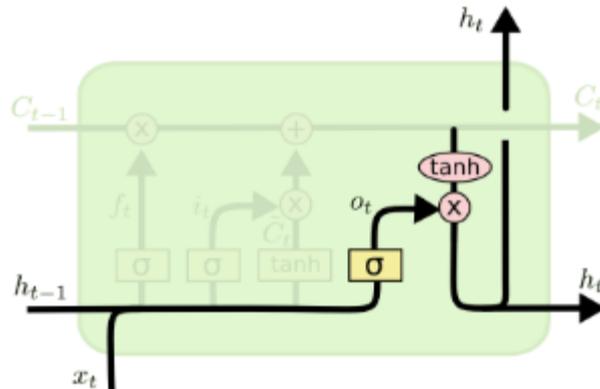

Figure A.6 Output from cell state (Olah, 2015).

$$o_t = \sigma(W_o[h_{t-1}, x_t] + b_o) \qquad (A.6)$$

$$h_t = o_t * \tanh(C_t) \qquad (A.7)$$