Erasmus University Rotterdam

MSc in Maritime Economics and Logistics

2018/2019

Evaluation of Forecasting Approaches for Crude Oil Tanker Freight Rates

By

Epameinondas-Nektarios Diakodimitris

Acknowledgements

I would like to thank my parents for their continuous financial and emotional support throughout this year. I would also like to thank my supervisor, Professor Albert Veenstra for his counseling and support during my thesis procedure, as well as his contribution to making this research possible. I would like to thank everyone working in the MEL office for supporting us emotionally, as well as always being there to help us with all practical matter we needed help with throughout this academic year. Finally, I would like to thank my friends and classmates, for making this year enjoyable and unforgettable. Thanks for everything.

Abstract

Crude oil freight rates is point of great interest within the shipping industry. A great number of parties within the maritime industry is interested in the prediction of the future regarding these freight rates.
A lot of different approaches have been applied in order to efficiently produce forecasts of shipping freight rates. These approaches differ in their efficiency and level of complexity. This thesis, based on the assumption of market efficiency, plans to uncover if crude oil tanker timecharter freight rates, as well as Worldscale rates(an index used for calculating tanker spot freight rates) can be forecasted efficiently by using a range of econometric models to produce forecasts of twelve months ahead, without the use of other exogenous variables which based on existing literature, affect freight rates.
The research is done for four vessel sizes in the crude tanker industry, Panamax, Aframax, Suezmax, VLCC and ULCC. Five different forecasting approaches were employed. An autoregressive process of order 1 was used as a naïve forecasting approach. An ARIMA process, as described by the methodology of Box-Jenkins, an ARMA-GARCH approach, as well as multivariate VAR and VECM approaches were the chosen approaches. The reasoning behind the choice of approaches is analyzed in the methodology chapter of this thesis.
The performance of these different forecasting approaches was evaluated based on three formal indicators of forecasting performance, MAE, RMSE and MAPE.
The analysis was done using eight different time series of freight rates, information about these times series is also presented in the methodology chapter of this thesis.
The results of the analysis were that ARIMA models produced better forecasts for almost every vessel size regarding the crude oil spot freight rates, while VAR-VECM models outperformed the other models for every vessel size in the case of crude oil time charter freight rates.
It was concluded that forecasts by the chosen approaches led to large forecasting errors with a minimum MAPE of 7.34%. These large forecasting errors were attributed to the fact that time series used had great variability as all the time series had high coefficients of variation, according to descriptive statistics of the data. It was also argued that the use other exogenous variables, such as economic shocks, or other exogenous variables that according to existing literature can affect freight rates in general, might be able to increase the efficiency of the proposed models.

Table of Contents

List of figures

List of Tables

List of abbreviations

ACF	Auto Correlation Function
ADF	Augmented Dickey Fuller
ANN	Artificial Neural Networks
AR	Auto Regression
ARCH	Auto Regressive Conditional Heteroscedasticity
ARIMA	Auto Regressive Integrated Moving Average
ARMA	Auto Regressive Moving Average
GARCH	Generalized Auto Regressive Conditional Heteroscedasticity
KPSS	Kwiatkowski– Philips–Schmidt–Shin
MAE	Mean Absolute Error
MAPE	Mean Absolute Percentage Error
RMSE	Root Mean Squared Error
PACF	Partial Auto Correlation Function
ULCC	Ultra Large Crude Carrier
VAR	Vector Auto Regression
VECM	Vector Error Correction Model
VLCC	Very Large Crude Carrier
WNN	Wavelet Neural Networks

1. Chapter 1. Introduction

1.1. Background

The beginning of seaborne oil trade happened during late 19th century. The original source producer was USA, while major consumers were European countries followed by Far East. Many things have changed since the first shipment in 1861, but the tanker market and generally charter shipping markets made the first steps to an era of globalized economy. The tanker market is still dependent on the global socioeconomic environment, and the state of the market is established by a vast number of variables and shocks. Undeniably, there is great level of difficulty in revealing all the qualitative and quantitative variables and their corresponding impact on freight rate determination (Wijnolst, 1999).

Forecasting the future with absolute certainty is one of mankind's oldest dreams, and naturally, it is one of the impossibilities of life, especially in the long term. In science, forecasting is attempted using "suitable" econometric methods and models.

Tanker freight rates is a topic of great interest within the shipping industry. Forecasting these freight rates can be done by using several different stochastic modeling approaches. Stochastic modeling is closely related to the use of probability theory, with the aim of modeling a range of technology and natural sciences phenomena. Until today, a considerable amount of work has been done in the field of tanker shipping and in the field of determining and forecasting shipping freight rates with the use of stochastic and econometric explanation modeling techniques.

Forecasting is even more difficult in case of shipping industry as it is one of the most stochastic and unpredictable economic environments, since there are important aspects of the future of the maritime industry that are not predictable.

Future freight rates are dependent on the number of ships which are ordered, a behavioral variable which at the is totally unpredictable in case of extreme shipping cycles, and developments in the world economy which, with its business cycles and crises, and its level of complexity is very high to be predicted by human beings. In such circumstances, even the most sophisticated scientific forecasting method will be of limited success. In addition, it must be mentioned, that forecasting is not about the future, it is about obtaining and analyzing the right information about the present. The right information is not always easy to come by, but it is crucial (Stopford M., 2009).

The focus of my research will be to research the parameters, benefits and drawbacks of different forecasting approaches with the aim of evaluating their accuracy and reaching a conclusion of the most accurate and attractive method of forecasting spot and timecharter freight rates of different size categories tankers by using only the actual freight rates as variables.

1.2. Problem Statement and research objectives

Until today, a lot of researchers have developed different models in order to forecast freight rates within the shipping industry. Models such as AR(Autoregressive model), ARMA(Autoregressive moving average model), ARIMA(Auto regressive integrated moving average model), ARCH(Autoregressive conditional heteroscedasticity), VAR(Vector autoregression), VECM(Vector error correction model), as well as ANN(Artificial Neural

Networks) and WNN(Wavelet Neural networks) have been used with the aim of predicting freight rates.

The conclusions of these various researches have been drawn based on the data used in each specific research. As I mentioned above, the goal of this thesis is to determine which of these approaches is best suited to forecast tanker freight rates, by using the same data for all the different models. Conclusions on the forecasting performance for spot and timecharter freight rates for different vessels sizes within the tanker industry will be drawn. In order to achieve that objective, it is necessary for a main research question to be formed:

“What is the most efficient model for forecasting crude oil tanker freight rates?”

This main research question shall help in understanding the benefits and drawbacks of each model, as well as their forecasting accuracy for crude tanker freight rates, as they will be tested using the same data. Of course, it is crucial for sub-research questions, which will help in answering the main research question, to be formulated:

I. “What are the main variables that affect tanker freight rates?”
II. “What are the possible forecasting approaches?”
III. “How can the quality of these forecasts be evaluated?”

1.3. Thesis structure

It is fundamental and crucial for this research to be appropriately segregated for the final objective to be reached. At this point, the content of the rest of the chapters of this thesis will be discussed.

Chapter 2 will include the literature review that will be needed in order to answer the main and the sub research questions. Past researches and the forecasting models used in them and their results, as well as the factors that directly or indirectly affect freight rates will be presented.

In chapter 3, firstly the choice of forecasting approaches, as well as the data that will be used will be presented, secondly, the specification of the forecasting models which will be compared in terms of their forecasting performance will be discussed and analysed.

In chapter 4, the analysis of the data in chapter 3 will be presented, by using the chosen forecasting models. After that, the results after running these models with the same data, as well as their forecasting performance for the different vessel sizes will be presented.

Finally, the conclusions of this research, as well as recommendation for further research that can be done in the subject will be discussed in Chapter 5.

1.4. Difficulties Encountered

At this point, some of the difficulties that were encountered during this research will be presented in short. The first problem that had to be solved, was that there was a lack of data for some specific freight rates series that were the were the main variables in this thesis. Such series were that of Panamax timecharter rates, as well as those of ULCC Worldscale and timecharter rates. For that reason, I decided to use a series of modern products tanker instead of dirty products tanker in the case of Panamax timecharter series, while the VLCC and ULCC were merged into the same category series, using data from VLCC vessels. This can also be seen in the data specification section in Chapter 3. The second difficulty I had to face was that the econometric knowledge that was required was beyond my expectations. Countless hours of self-study were required just to identify the correct methodology for the more sophisticated models that were used in this thesis. Last but not least, after identifying these methodologies, I run against many difficulties when I attempted to run these models using the R-programming language software. In order to ensure that every step of the way was done correctly, which was crucial for some statistically valid results to be derived, I had to spend countless hours just to understand and apply this knowledge of R programming language in order to solve those specific problems and produced valid forecasts.

2. Literature Review

2.1. The freight market

The Baltic Shipping Exchange, which was the original freight market, first started to trade as a commodity and shipping exchange in the mid-nineteenth century. Nowadays, the freight market remains a marketplace in which sea transport is being bought and sold, but the business is mainly transacted by telephone, e-mail and messaging services instead of being conducted on the Baltic. Today, there is a single international freight market but, just as there are separate sections for different products, there are separate markets for different ships in the freight market. In the short term the freight rates for tankers, bulk carriers, containerships, gas tankers, and chemical tankers behave differently, but because it is the same broad group of traders, what happens in one sector eventually also happens to the others. The freight market has two different types of transactions. The first one is the time charter under which the ship is hired by the day. The second one is the freight contract (also known as spot freight rate) in which the shipper buys transport from the shipowner at a fixed price per ton of cargo. The first choice is suitable for experienced ship operators who prefer to manage the transport themselves. On the other hand, freight contracts are preferred by shippers who want to pay an agreed sum and have the shipowner manage the transport of the cargo (Stopford, 2009).

2.2. Time Charter Rates

A time charter gives the charterer control in the operation of the ships which carry his cargo, while the shipowner manages and owns the vessel. The length of the charter can be the time taken to complete a single voyage, also known as trip charter, or a longer period of months or years, known as period charter. The reasons time charter can be attractive are three. Firstly, the shipper doesn't wish to become a shipowner, but he requires the use of a ship under his own control. Secondly, the time charter may be cheaper than buying a vessel, especially if the owner has very low costs, because of larger fleet and low overheads. Lastly, the charterer may be a speculator and anticipates a change in the market. Time charter rates are generally measured in thousands of dollars per day. Time charterer rates are commonly reported for a single trip voyage, 6 months, 1 year or 3 years. (Stopford, 2009).

2.3. The Worldscale Rates

The tanker industry uses this freight rates index as a more convenient way of negotiating the freight rate per barrel of oil transported on many different routes. This concept was developed during the Second World War when the British government introduced a schedule of official freight rates as a basis for paying the owners of requisitioned tankers. The Worldscale index is published in a book which is used as the basis for the calculation of tanker spot rates. The book shows the cost of transporting a ton of cargo using the standard vessel on a round voyage for different routes worldwide. This cost is known as 'Worldscale 100'. Each year the Worldscale

Panel meets in New York and London to update this book. The Worldscale system makes it easier for shipowners and charterers to compare the earnings of their vessels on different routes (Stopford M., 2009). Worldscale is used to calculate freight rates for oil tankers and product carriers. It is a tanker chartering tool, which is used to provide the ship-owner with the same net return per day regardless of voyage performed for the Worldscale Standard Vessel at WS100. These rates (Worldscale 100) are based upon a 75,000 ton total capacity vessel performing a round voyage and expressed in dollars per ton of cargo (WorldScale Association (London) Limited [GB]).

2.4. Freight Rate Mechanism

Tanker freight rates are the point of interest in this research. In this chapter the mechanism around freight rates will be briefly explained. The freight market is depicted in Figure 1. This mechanism links supply and demand for freight transport, which is crude oil in the case of the crude oil tanker market. The way of operation of this mechanism is quite straightforward. Shippers negotiate with shipowners in order to establish a suitable freight rate, which will reflect the available cargoes and balance of ships in the market. Hence, the freight rate will be low if there is oversupply of ships in the market, and high if the ships in the market are limited (Stopford, 2009).

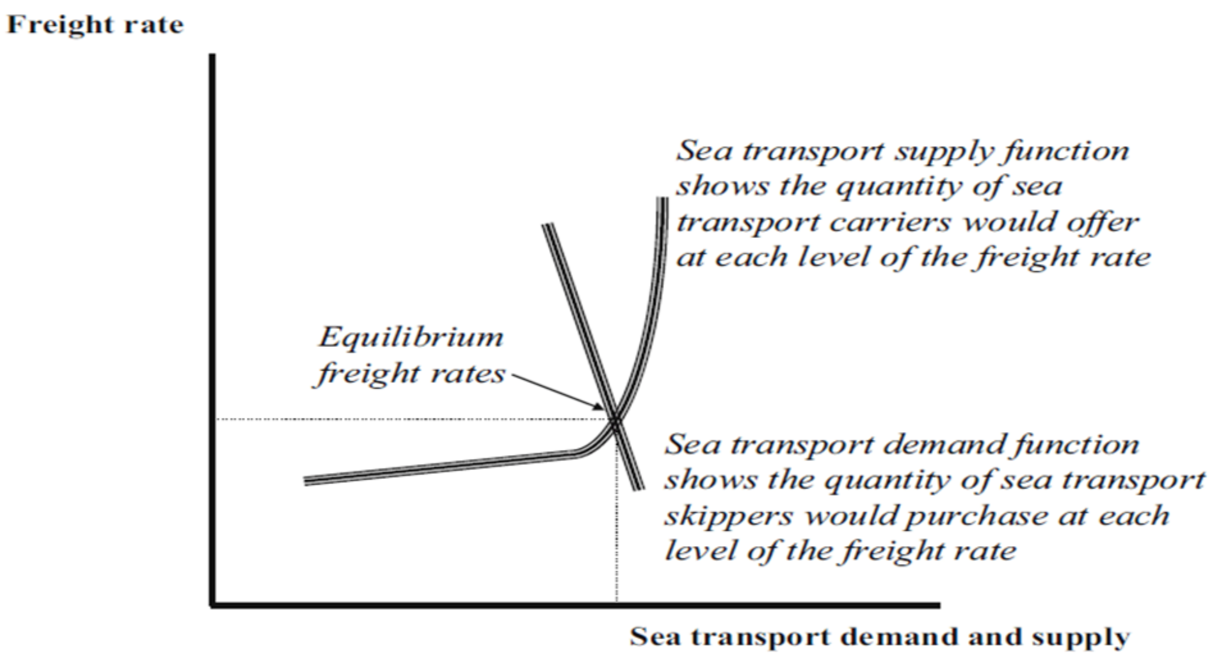


Figure 1. Freight rate mechanism.
Source: Lun et al., 2010, page 28.

2.5. Variables that affect tanker freight rates

The tanker freight market is characterized by the interaction of supply and demand for tanker shipping services (Koopmans, 1939, Zannetos, 1966, Beenstock and Vergottis, 1993). The demand for tanker services is a derived demand, in the sense that a lot of other variables affect it. Such variables are the international trade in oil and oil products, which in turn depends on world economic activity and imports and consumption of energy commodities (Stopford, 2009). On the other hand, the supply of tanker shipping services, is affected by other variables, such as the size of the tanker fleet, tanker shipbuilding activities, bunker fuel prices, the tonnage available for trading, the scrapping rate of the tanker fleet, and the productivity of the tanker fleet at any point in time (Alizadeh et.al, 2011). Many studies have been conducted with the aim of empirically investigating the determination of shipping freight rates and the price for sea transportation, through the interaction between supply and demand for sea transportation. It was found that factors such as world economic activity, growth in industrial production, seaborne trade in commodities, oil prices, availability of vessel tonnage, new vessel buildings on order, tanker shipbuilding deliveries and the scrapping rates determine freight rates for sea transportation (Hawdon, 1978 , Strandenes, 1984 , Beenstock and Vergottis, 1989, Beenstock and Vergottis, 1993).

2.6. Vessel sizes in the crude tanker market

This research focuses on forecasting methods for crude oil tanker freight rates. Therefore, it is crucial for the different size of tanker vessels to be analyzed for a deeper understanding of the problem to be reached. A categorization of crude oil tankers based on their DWT will be made. (Information about the vessels sizes were gathered in http://maritime-connector.com/)

2.6.1. Panamax

Panamax tankers operate in the region of Panama and are designed to cross the Panama Canal. Their size regulations are set by the Panama Canal Authority. They have length of 950 feet, width of 105 feet and depth of 39.5 feet. Their DWT ranges from 60,000 to 80,000.

2.6.2. Aframax

Aframax are medium-sized tankers with DTW between 80,000 and 120,000. It has a carrying capacity of 70,000 – 100,000 metric tons (750,000 barrels) and it can serve most ports of the world because of their favorable size. They are widely used for medium haul crude oil transportation.

2.6.3. Suezmax

Suezmax tankers are mid-sized oil tankers with DWT ranging from 120,000 to 200,000. Their design allows them to be able to enter most of ports worldwide. They are the largest tankers which meet the restrictions of the Suez Canal. A typical Suezmax vessel would have length of 900 feet, width of 157 feet and draught of 53 feet, corresponding to about 150,000 DWT.

2.6.4. VLCC (Very Large Crude Carrier)

The size of a VLCC ranges between 180,000 to 320,000 DWT. Due to the face that they can pass through the Suez Canal in Egypt, they are used extensively around the North Sea, Mediterranean and West Africa. Their length is up to 1,540 feet, their beam is up to 200 feet and their draught is up to 66 feet. These vessels are very flexible in using terminals and can operate in ports with some depth limitations.

2.6.5. ULCC (Ultra Large Crude Carrier)

ULCCs are the largest vessels in the world with DWT ranging between 320,000 and 500,000 and dimensions of 415 meters length, 63 meters width and 35 meters draught. Because of their size, they usually need custom made terminal in order to operate. They are primarily used for very long-haul crude oil transportation from the Persian Gulf to Europe, Asia and North America.

2.7. Forecasting models used in shipping

The mechanism of freight rates is one of the most crucial points in the shipping industry. The nature of the industry raises a point to carry out numerous studies and researches. The uncertainty and implied volatility of freight prices are the main motives, which lead the researchers to discover more appropriate quantitative methods, with the goal of deciphering the market (Geomelos et al., 2014).

New studies are focused on dynamic systems which exploit the development of econometrics. The dynamic nature of models interprets shipping markets in better way, and it helps to understand the mechanism of markets by producing more accurate forecasts (Hawdon, 1978).

Many economists have laid the foundation of dynamic analysis of shipping markets and especially that of forecasting the freight market using techniques of econometric analysis. The studies that have been done in this field will be now be presented in in this chapter.

One of the first approaches was that of Beenstock and Vergottis in 1989. They applied a model in which secondhand ships are considers capital assets and freight market and shipping markets are not dependent, to the world for dry cargo market. In the same year, they also applied this model empirically to the world tanker market.
They estimated an aggregated econometric model of the tanker market based on the model M. Beenstock proposed in 1985, in which it was assumed based on rational expectations, that tanker prices and freight rates were dynamically and jointly determined. They concluded that the dry bulk and tanker markets are independent (Beenstock, and Vergottis**,** 1989).

In 1992, Kevin Cullinane developed a model through the application of the Box—Jenkins approach to time series analysis and forecasting, with the aim of applying it to speculate Baltic International Freight Futures Exchange. This model was evaluated based on its forecasting performance compared to alternative forecasting models.
He concluded that ARIMA models are put to best use when it comes to production of short-term forecasts, and that given this argument ARIMA models must be responsive to short-term market movements. Furthermore, he argued that, as any highly specified model is handicapped by the number of its parameters and less specified models are limited by lesser accuracy, there is an optimum level of specification and that Box-Jenkins approach provides an effective method of estimating that level. That level ensures that the selected ARIMA model is sufficiently specified, that the forecasts it produces replicate reasonably well the characteristics of the market, but, at the same time, is not over specified and thus hindered by the number of its parameters. Lastly, he concluded that the use of univariate time-series analysis provides a cost-effective and efficient technique for developing forecasts as the basis of a strategy for speculating on BIFFEX (Cullinane, 1992)

ARCH-type models have also been used in shipping. Some researches that used arch-type models will now be presented.

In 1996, Kavussanos, M. G. utilized an autoregressive conditional heteroskedasticity (Arch) model to model and compare time varying risks between different size tankers. His conclusions were that the monthly price returns on larger vessels were more volatile than those of smaller ones and the volatilities varied depending on the size of the tanker. (Kavussanos, 1996).

In 2002, D. Bessler, S. Jonnala and S. Fuller used directed graphs and autoregressive conditional heteroskedastic error processes (GARCH model) in order to specify and estimate an ocean grain rate equation. The results of their research showed that voyage distance, ship size, contract terms, flag and season are important determinants of these rates. Their findings also suggested that efficient port infrastructure and its ability to accommodate the increasingly large, more efficient bulk carriers is of outmost importance in maintaining exporting countries' competitiveness in world grain markets (Bessler et al.,2002)

In 2018, K. Gavriilidis et al., examined if the inclusion of oil price shocks of different origin as exogenous variables in a wide set of GARCH-X models improves the accuracy of their volatility forecasts for spot and 1-year time-charter tanker freight rates. Their results, which appeared robust across freight rate contracts and vessel sizes such as VLCC, Suezmax, Aframax and MR. indicated that the inclusion of aggregate oil demand shocks and precautionary oil-specific demand shocks significantly improves the accuracy of the volatility forecasts drawn, while the inclusion of oil supply shock leads to weak improvements in forecasting volatility in the tanker freight markets (Gavriilidis et al., 2018)

Some papers focusing on the use of multivariate models in order to forecast freight rate will now be presented.

In 1997, based on formal statistical tests and the assumption that there are five cointegration relations between the six series, Veenstra, A. W. and Franses, P. H. formulated the VEC model. This model doesn't include other endogenous or exogenous variables, because the authors consider the spot market as efficient, thus the first differences of the six freight rate series are explained by simple differences between the freight rate series. This hypothesis led to large forecast errors, because of the existence of common stochastic trend among the six different time series of routes. The results indicated that an economically meaningful structure exists in a set of ocean dry bulk freight rates and that there are stable long-run relationships between such freight rates. Based on this structure, the stochastic trend behind the freight rates was uncovered. They concluded out that a substantial part of the movement in freight rates is stochastic in nature. Even if there are long-run relationships between freight rates, it was derived from their research, that such relationships do not result in improved forecasts (Veenstra and Franses, 1997).

In 2018, Taib C. and Mohtar, Z. I., employed an applied econometric study concerning forecasting the spot freight rates based on Forward Freight Agreement and Time Charter contracts. Empirical analysis contained investigation of the relationship between the spot freight rates with FFA and TC. After all the time series were checked for stationarity, the calibration of vector error correction model (VECM) was carried out using ordinary least square method. Later, the VECM was used to forecast the spot rates. MAD and RMSE were used to analyze the forecasting performance of the two models. They concluded that, FFA was the more suitable managing the volatility of the spot freight market (Taib and Mohtar, 2018).

Artificial Neural Networks have also been used in the past in order to produce freight rates forecasts.

In 2004, Lyridis, D. V. *et al.* attempted to uncover the benefits of using Artificial Neural Networks (ANNs) with the aim of forecasting VLCC spot freight rates. Some of their findings were that ANNs are performing better when the tanker market is volatile and that the use of informative variables such as the arbitrage between types of crude oil as well as freight rates of Capesize tankers can improve ANN performance (Lyridis et al., 2004).

Following, some papers that focused on the evaluation of different forecasting approaches including neural networks in the field of shipping will be presented.

In 2006, R. Adland and K. Cullinane, analyzed the possibilities of using Wavelet Neural Networks to forecast the Baltic Exchange Dirty Tanker Index (BDTI). Their results show only for longer periods the WNN model shows superiority over ARIMA, because it offers reasonable non-linear forecasts about the BDTI movements. In terms of forecasting accuracy, WNN decreases as forecasting times increase. WNN model did not perfectly fulfill the forecasting tasks in challenging situations, but it was able to predict and capture some useful information about trends and movements of the BDTI index. (Adland et al., 2006).

In 2007, R. Bachelor et al. tested the performance of popular time series models in predicting spot and forward rates on major seaborne freight routes. Some of their conclusion were that vector error correction models (VECM) give the best in-sample fit, but they also suggested that forward rates converge on spot rates. They also found that in the case of forward prices, ARIMA or VAR models forecast way better than VECM models (Bachelor et al., 2007).

In 2013*,* S. Fan et al. examined non-linearity and non-stationary features of the BDTI and elaborated WNN model building procedures. Their results outlined that even though there seemed to be no significant difference in performance between WNN and ARIMA model in BDTI short term forecasting, for longer periods, the WNN model shows superiority over ARIMA, because it offers reasonable non-linear forecasts about the BDTI movements. Furthermore, in terms of forecasting accuracy, WNN performance seemed to decrease as forecasting times increased (Fan et al., 2013).

Recently, in 2017, Geomelos and Xideas employed an applied econometric study concerning forecasting spot prices in bulk shipping in both markets of tankers and bulk carriers in a disaggregated level. The uncertainty about the future development of spot prices could be reduced by using estimates of ex-post and ex-ante forecasts. The focus of their analysis was the comparison multivariate time series models (VAR and VECM) and univariate time series models (ARIMA, GARCH and E-GARCH) with the goal of deriving the best predicting model for each ship type. They also tried to yield better forecasts by combining forecasting approaches. They concluded that the forecasting errors could be reduced even more by using that combining methodology (Geomelos and Xideas, 2017).

This range of papers contributed to my knowledge and understanding of the econometric models and approaches which have been used to forecast and analyze different phenomena within the maritime sector, but most significantly in the forecasting of different types of freight rates. It also helped me decide on the approaches I am going to include in my analysis.

2.8. Definition of Time Series

A short definition of time series is crucial, as all the data in this thesis are expressed in time series. A time series is a set of observations x_t, each one being recorded at a specific time t. A discrete-time time series is one in which the set of times at which observations are made is a discrete set and observations are made at fixed time intervals. Continuous time time series are obtained when observations are recorded continuously over some time interval (Brockwell and Davis, 2001). In this research discrete time series of freight rates will be used.

2.9. Time series components

Four components may be present in a time series. The first one is the trend. The trend, which is also known as secular trend, is a relatively smooth, long term pattern or direction which is visible in the series. Its duration is more than a year and it is not always linear. The second component is cyclical variation (or cyclicality), which is a wavelike pattern that describes a long-term trend that is, in general, apparent over a finite number of years and results in a cyclical effect. This variation has a duration of more than a year, by definition. Some representative examples of such variation are business cycles which include periods of economic recession and inflation, demand cycles of long-term products, as well as cycles in monetary and financial sectors. Nevertheless, for reasons of practicality, this type of variation is most of the times ignored, as cyclical patterns than are predictable and consistent are extremely rare. The third component is seasonal variation (or seasonality). It also has a duration of more than a year, which refers to cycles occurring over repetitive calendar periods. The term of seasonal variation can refer to the four traditional seasons or to a different seasonal pattern of a month, a week or a day. The last component is random variation. This type of variation is caused by random and unpredictable changes in a time series and has nothing to do the other previously mentioned components. It exists in almost every time series and it usually covers the existence of more predictable components(Keller, 2015).

3. Methodology (Choice of forecasting models, data specification and analysis of the chosen models)

The main justification for modelling freight rates in a univariate and multivariate time series models is that shipping markets are assumed to be approximately efficient. In fact, the conditions for perfect and competitive markets all hold to a considerable extent in shipping (Evans, 1994). Assuming that the market is efficient, it can be argued that the relevant freight rates contain all publicly available information (Fama, 1970, Nicholson ,1989). Thus, it can be assumed that no extra variables beyond freight rates are necessary for model building. Given the obvious linkage of freight rates over time, one might expect that the freight rates of different parts of the shipping industry may be correlated, but that will be not be researched in this thesis. The purpose of this thesis will be to examine different forecasting approaches using only the freight rate times series as variables and reach a conclusion on the most efficient approach for each freight rate time series. The reason this approach was chosen is the fact that forecasting crude oil tanker freight rate series using only the freight rate as variables can be less complex and less time consuming for the interested parties. Next, the reasoning behind the choice of models for this task will be analyzed.

Before proceeding with the rest of the methodology, it is crucial to justify the choice of models made in this thesis. The first model that was chosen was a simple Autoregressive process of order one (AR(1)). This approach was considered a naïve forecasting approach because the model was based on the hypotheses that the value of a freight rate in a certain point is only dependent on the previous value plus a random error. The second model was an ARIMA-approach (Auto Regressive Integrated Moving Average) as defined by the Box-Jenkins methodology. The results of the naïve model and the fact that the data were not stationary as well as the fact that the ARIMA approach has been efficient in producing forecasts that replicated reasonably well the characteristics of the market (Cullinane, 1992), justifies the choice of the ARIMA model. The third approach will be an ARCH-type (Auto Regressive Conditional Heteroskedasticity) model. From the literature review, it can be derived that arch type models are usually used to model the volatility of financial time series. The choice of using an ARMA-GARCH approach, which is an extension of the GARCH model was based on the fact that the approach of modelling not only the volatility but also the conditional mean of the time series using an ARMA process was considered an approach more fitted for the data used in this thesis. As a last approach, a multivariate VAR (Vector Auto Regression) model will be used to produce forecasts. Previous attempts to use such model without using other exogenous variables, did not have great success (Veenstra and Franses, 1997), but that was the case for dry bulk spot freight rates. I was curious in discovering its forecasting performance in the case of crude oil tanker spot and timecharter freight rates. The choice of the exact form of the model (VAR or VECM) will be decided based on the statistical properties of the data and appropriate statistical tests. It is also important to be mentioned that it was decided not to use Wavelet Neural Networks, because in terms of forecasting accuracy, WNN performance seemed to decrease as forecasting times increase(Fan, 2013), it did not perfectly fulfill the forecasting tasks in challenging situations(Cullinane, 2006), its level of complexity is higher than the rest of the models and they usually require a larger data set to produce reliable forecasts.

After this discussion on the choice of models which will be used, a short discussion about the benefits and drawbacks of each model will now be presented. Starting with the univariate models which were used, ARIMA and ARMA-GARCH, it should be mentioned that univariate time series models attempt to model and to predict financial variables using only information about their own past values and their own error terms. These models are a-theoritical, meaning that they don't require any theoretical relationship between the variables in order to be constructed and used, while they also attempt to capture empirically relevant features of the observed data. Specifically, ARIMA models are used to model the conditional mean of past values. On the other hand, an Arch type model is used to model the conditional variance of the data. Arch type models have benefits such as that it can produce volatility clusters, meaning that large changes tend to be followed by large changes and small changes tend to be followed by small changes (Mandelbrot, 1963). One more advantage of these models is that the distribution of the models has heavy tails. These models also have some drawbacks. Firstly, these models assume that an increase and a decrease in the price of the variable have the same effect on its volatility. Secondly, Arch-type models are restrictive in the choice of their parameters. Lastly, they merely provide a mechanical way to describe the behavior of the conditional variance, meaning that it doesn't give any indication about why such behavior occurs (Tsay, 2014). Moving to multivariate models, VAR models, which are also a-theoritical models have some advantages and disadvantages compared to univariate time series models. One of its benefits is that the model itself is a generalization of a simple AR model and can be extended so that the model includes first difference terms and cointegration relationships between the variables. Another benefit of a VAR model is that it allows a variable to depend not only on its own lags but also on the lags of other variables. Lastly, the forecast generated by a VAR process are often better than traditional structural models, as it has been argued that large-scale structural models perform badly when it comes to out of sample forecasting (Sims, 1980). On the other hand, its disadvantages are that, firstly there is difficulty in deciding the optimal lag length for such models, and secondly, the number of the parameters used in such models is huge. For example, for a VAR model with 3 variables and 3 lags, there will be 30 parameters to be estimated. At this point, it should also be mentioned that univariate time series models can be expressed as a restricted version of a VAR model. (Brooks, 2019)

3.1. Data Specification

The data used for this thesis were retrieved by Clarksons Shipping Intelligence Network. One of Clarksons functions is that it collects, processes and maintains much data on the shipping industry, including an extensive number of freight rate series over a large span of time. Its research team is respected worldwide as the most authoritative provider of intelligence on global shipping and the offshore sector. Four vessel categories were selected. VLCC and ULCC were put into the same category due to the very limited data available for the ULCC series. Only for the Panamax Timecharter series, a rate of modern product tankers was used, due to lack of data of dirty products within the Clarksons database.
For each of the vessel categories time series with monthly data from January 2004 to December 2015 were used. The actual data from January 2015 to December 2015 will be left out of the analysis in order to be compared with the forecasts of the models. In the following table the choice of time series for time charter rates and Worldscale rates will be presented. The data in the time charter times series are measured in $/day, while the data in the spot rates time series are measured in Worldscale rates, which is sort of linear transformation of spot rates in $/ton. In the following table the choice of time series for time charter rates and Worldscale rates can be seen in Table 1.

Category	Time series
Panamax Worldscale	Antwerp-Houston Dirty 55K Worldscale Rates
Pananax Time Charter	1-year timecharter rate 74K modern products tanker
Aframax Worldscale	Azrew-Philadelphia 80K Worldscale Rates
Aframax Time Charter	1-year timecharter rate Aframax (Long Run Historical Series)
Suezmax Worldscale	Bonny off – Philadelphia Suezmax 130K Worldscale Rates
Suezmax Time Charter	1-year timecharter rate Suezmax (Long Run Historical Series)
VLCC-ULCC Worldscale	Ras Tanura – Rotterdam VLCC 280K Worldscale Rates
VLCC-ULCC Time Charter	1-year timecharter rate VLCC (Long Run Historical Series)

Table 1. Time Series Used for each freight rate category

In Figure 2 and Figure 3, the time series plots of Worldscale Rates and Timecharter rates for each type of vessel can be observed. After plotting the data, no unusual observations were identified, while a sudden drop in the values is identified during the year 2008, probably because of the economic crisis that happened during that time.

At this point, some descriptive statistics of the time series which will be used in this thesis will be presented. This will be helpful in identifying the mean value, the minimum and maximum values, as well as the standard deviation and coefficient of variation of each of the times series that will be used. The coefficient of variation shows the extent of variability in relation to the mean of the sample and is equal to the standard deviation of the sample divided by the mean value of the sample. Despite the fact the other four descriptive statistics of Worldscale series and timecharter series are on different scales, because the first is measured in dollars per ton while the second is measured in dollars per day, they can be compared based on the coefficient of variation, as it is expressed as a percentage. The Panamax Worldscale series has a mean value of 183.8623, a minimum value of 70, a maximum value of 481.25, a standard deviation of 85.3809, while its coefficient of variation is 46.43%. The Aframax Worldscale series has a mean of 145.039, a minimum value of 67.5, a maximum value of 377.5, a standard deviation of 63.7395, while its coefficient of variation is 43.94%. The Suezmax Worldscale series has a mean value of 114.0135, a minimum value of 46.5625, a maximum value of 350.625, a standard deviation of 57.5065, while its coefficient of variation is 5.43%. The VLCC-ULCC Worldscale series has a mean value of 59.8729, a minimum value of 17.5, a maximum value of 222.5, a standard deviation of 37.9844, while its coefficient of variation is 63.44%. Moving to the Timecharter series, the Panamax timecharter series has a mean value of 21,867.9924, a minimum value of 12,500, a maximum value of 37,100, a standard deviation of 7,526.3129, while its coefficient of variation is 34.41%. The Aframax timecharter series has a mean value of 24,200.6628, a minimum value of 13,000, a maximum value of 43,500, a standard deviation of 9,226.9599, while its coefficient of variation is 38.12%. The Suezmax timecharter series has a mean value of 32,136.6174, a minimum value of 15,246, a maximum value of 58,750, a standard deviation of 12,224.2424, while its coefficient of variation is 38.03%. Lastly, the VLCC-ULCC timecharter series has a mean value of 43,078.8825, a minimum value of 18,000, a maximum value of 90,000, a standard deviation of 18,685.3504, while its coefficient of variation is 43.37%. These descriptive statistics can also be seen in Table 2.

Time series	Mean Value	Minimum Value	Maximum Value	Standard Deviation	Coefficient of Variance
Panamax Worldscale	183.8623	70	481.25	85.3809	46.43%
Pananax Time Charter	21,867.9924	125,000	37,100	7,526.3129	34.41%
Aframax Worldscale	145.039	67.5	377.5	63.7395	43.94%
Aframax Time Charter	24,200.6628	13,000	43,500	9,226.9599	38.12%
Suezmax Worldscale	114.0135	46.5625	350.625	57.5065	50.43%
Suezmax Time Charter	32,136.6174	15,246	58,750	12,224.2424	38.03%
VLCC-ULCC Worldscale	59.8729	17.5	222.5	37.9844	63.44%
VLCC-ULCC Time Charter	43,078.8825	18,000	90,000	18,685.3504	43.37%

Table 2. Descriptive statistics of the times series (January 2004 – December 2014)

3.2. Seasonality

The examination of seasonality in the time series is a crucial parameter for this research. Seasonality can be identified by observing regular peaks in the autocorrelation function (ACF). Additional information about the seasonality of the time series can also be provided by the partial autocorrelation function (PACF). In the autocorrelation function, it can be defined how data points in a time series are related, on average, to the preceding data points (Box, Jenkins, & Reinsel, 1994). A partial autocorrelation function shows a summary of the relationship between an observation in a time series with observations of its lags with the relationships of intervening observations removed. In Appendix 3 and 4 the ACF and PACF (correlogram) at 24 lags of the 8 different times series can be observed. From a quick look at the ACF plots, the TC rates show no signs of seasonality, while the WS rates could have some form of very weak seasonality. The ACF and PACF were produced by using the statistical software gretl.

3.3. Stationarity

Stationarity implies that the variances and autocovariances are finite and independent of time or in other words that the distribution of the variable in question does not depend on time. It is crucial for the analysis to test the order of integration of the time series. For univariate time series models, the time-charter rates and the Worldscale rates must be stationary or integrated of order zero (0,0). There are three most used statistical tests according to Lutkepohl and Kratzig (2004). The first test, is the Augmented Dickey-Fuller (ADF) which tests the pair of hypotheses
H0: γ = 0 versus H1: γ < 0 and estimates the following regression:

$$\Delta Yt = a + \beta_t + \gamma Y_{t-1} + \delta_1 \Delta Y_{t-j} + \cdots + \delta_{p-1} \Delta Y_{t-p+1} + \varepsilon_t$$

where α is a constant, β is the coefficient of time trend and p is the lag order of the autoregressive process. This lag order for the test needs to be determined when running the test. In this research the Akaike information criterion (AIC) is used for determining the lag order. ADF test is based on t-statistic, while the critical values have been obtained by Davidson and MacKinnon (1993).

The second test is known as Philips–Perron (PP), and it is an alternative to the aforementioned ADF test. The PP test is used for series which have structural breaks (Perron, 1989). The reason this test will not be used in this research is that Davidson and MacKinnon (2004) reported that the PP test performs worse than ADF test in the case of finite samples.

The third test is known as Kwiatkowski– Philips–Schmidt–Shin (KPSS) test. This test examines as null hypothesis that the data generating process is stationary
[H0: Y_t is integrated of order 0] against the alternative that is
[H1: Y_t is integrated of order 1]. KPSS test is not appropriate for very large samples and its application is questioned in models with a very large number of observations as Caner and Kilian (2001), Kuo and Tsong (2004) and Geomelos and Xideas (2014) notice. The data used in this thesis consist of time series of 132 observations, thus the samples are not considered to be large.

Forecasting procedure and forecasting performance indicators

Using the approaches, that will be analyzed later in this chapter, a forecast for every of the 8 times series used in this research will be produced. Firstly, the models will be chosen for the data sample. The data sample which will be used will be the observations from January 2004 to December 2014. Then, the values from January 2015 to December 2015 will be. These produced forecasts will then be compared with the actual values of these periods (in-sample forecasting). The forecasting performance of different models will be compared using some formal indicators. These indicators, as well as some information about them will be now presented.

The first indicator will be Mean Absolute Error (MAE). It measures the average importance of the errors in a set of predictions, without taking into consideration if they are positive or negative. It’s the average over the test sample of the absolute differences between the forecasted and the actual observations where each difference has the same weight. The equation of this metric is:

$$MAE = \frac{1}{n}\sum_{j=1}^{n}\left|y_j - \widehat{y_j}\right|$$

Where n is the number of forecasted observations, y_j are the real values at period j and $\widehat{y_j}$ are the forecasted values at period j.

The second indicator will be Root Mean Square Error (RMSE).It is a quadratic error metric indicator that also measures the average importance of the error. It’s the square root of the average of squared differences between forecasted and actual observation. The equation of this metric is:

$$RMSE = \sqrt{\frac{1}{n}\sum_{j=1}^{n}\left(y_j - \widehat{y_j}\right)^2}$$

Where n is the number of forecasted observations, y_j are the real values at period j and $\widehat{y_j}$ are the forecasted values at period j. RMSE differs from MAE in the sense that the errors are squared before they are averaged. Thus, the RMSE gives higher weight to large errors, which means that it is more useful when large errors are particularly undesirable. This exactly the case, when it comes to freight rate prediction, as huge amount of money is invested in the procedure. This means that, the larger the forecast error is, the greater the exposure of the parties interested in the freight rate forecast to financial damage.

The third indicator will be Mean Absolute Error Percentage (MAPE). As its name suggests, this indicator is very similar to MAE, but it measures the percentage of error between the actual variables and the forecasted ones, in other words, it expresses accuracy as a percentage of the error. Because this number is a percentage, it is easier than the other two metrics to understand and makes the comparison of forecasts of different time series possible. The equation of this metric is:

$$MAPE = \frac{100\%}{n}\sum_{j=1}^{n}\left|\frac{y_j - \widehat{y_j}}{y_j}\right|$$

Where n is the number of forecasted observations, y_j are the real values at period j and $\widehat{y_j}$ are the forecasted values at period j.

3.4. Naïve Forecasting Approach

The first forecasting approach that will be used in this thesis, is a naïve forecasting approach. An autoregressive process of order 1 [AR(1)] is going to be used. As this is a naïve approach, the assumptions that there is no seasonality in the data and that all times series are stationary will be made. The equation of this model is:

$$Y_t = a + bY_{t-1} + e_t$$

Where $\boldsymbol{Y_t}$ the value at period t , a is a constant, b is the coefficient of the value at period t-1 (lag 1 of dependent variable $\boldsymbol{Y_t}$) and $\boldsymbol{e_t}$ is white noise.
The forecasting performance of this approach will later be compared to the results of more sophisticated forecasting models that will be used in this paper.
The models will be run in the programming language R.

3.5. ARMA models and Box-Jenkins Methodology

Wold (1938), who combined both AR(Auto Regression) and MA(Moving Average) schemes and showed that ARMA processes can be used to model a large class of stationary time series as long as the appropriate order of p, the number of AR terms, and q, the number of MA terms, was appropriately specified. This means that a general series x_t can be modelled as a combination of past x_t values and/or past e_t errors. The following equation is the form of an ARMA model.

$$x_t = \varphi_1 \chi_{t-1} + \varphi_2 \chi_{t-2} + \cdots + \varphi_p \chi_{t-p} + e_t - \theta_1 e_{t-1} - \theta_2 e_{t-2} - \cdots \theta_q e_{t-q} \qquad (3.5.1)$$

The following approach proposed by Box and Jenkins came to be known as the Box Jenkins methodology to ARIMA models, where the letter `I', between AR and MA, stood for the `Integrated' and rejected the need for differencing to make the series stationary.
The equation of a non-seasonal ARIMA(p,d,q) process is the following

$$\varphi(B)\left(1 - B^d\right)y_t = c + \theta(B)e_t \qquad (3.5.2)$$

where e_t is a white noise process with mean zero and variance σ^2, B is the backshift operator, and φ(z) and θ(z) are polynomials of order p and q respectively. To ensure causality and invertibility, it is assumed that φ(z) and θ(z) have no roots for |z| < 1 (Brockwell and Davis 1991). If $c \neq 0$, there is an implied polynomial of order d in the forecast function.
The equation of seasonal ARIMA(p,d,q)(P,D,Q)m process is given by

$$\Phi(B^m)\varphi(B)(1 - B^m)D(1 - B)dy_t = c + \Theta(B^m)\theta(B)e_t \qquad (3.5.3)$$

where Φ(z) and Θ(z) are polynomials of orders P and Q respectively, each containing no roots inside the unit circle. If $c \neq 0$, there is an implied polynomial of order d + D in the forecast function.

Next, the four steps of Box and Jenkins methodology will be presented.

Firstly, before equation (3.5.1) can be used, the series should be stationary in its mean and variance. The Box-Jenkins methodology suggests short and seasonal (long) differencing to achieve stationarity in the mean, and logarithmic or power transformation to achieve stationarity in the variance. If the series are seasonal, the Box and Jenkins methodology proposes multiplicative seasonal models coupled with long-term differencing, if necessary, to achieve stationarity in the mean. The difficulty with such an approach is that there is practically never enough data available to determine the appropriate level of the seasonal ARIMA model with confidence. Users therefore proceed through trial and error in both identifying an appropriate seasonal model and in selecting the correct long-term (seasonal) differencing.

Secondly, the order of the ARMA model can be found by examining the autocorrelations and partial autocorrelations of the stationary series. Box and Jenkins provided both a theoretical framework and practical rules for determining appropriate values for p and q as well as their seasonal counterparts P and Q. The possible difficulty is that often more than one model could be considered, requiring the user to choose one of them without any knowledge of the implications of his or her choice on post-sample forecasting accuracy. In that case, they did recommend the principle of parsimony, which means that a simpler model with fewer parameters should be selected if more than one model is possible.

Thirdly, the coefficients of the parameters of the model must be estimated. This part is quite straightforward, as the non-linear optimization procedure, based on the method of steepest descent (Marquardt, 1963) is used to estimate the coefficient for each parameter of p and/or q as well as their seasonal equivalent P and/or Q if necessary. Except for occasional problems when there is no convergence (in which case another model is used) the estimation provides no special difficulties except for its inability to guarantee a global optimum (a common problem of all non-linear algorithms). The estimation is completely automated, as it requires no judgmental inputs, as all computer programs use the same algorithm in applying this optimization procedure of Marquardt.

Lastly, after a model has been decided, and the parameters of this model have been estimated, some diagnostics checks need to be conducted, because the residuals of the actual values minus those of the forecasted values need to be examined according to Box-Jenkins methodology. If these residuals are random, the model is assumed to be appropriate. If they are not random, then another model must be considered, and the same procedure must be conducted again. Several tests have been suggested by Box and Pierce (1970), in order to determine if the residuals of the model are random. (Maridakis et al, 1997)

The way of implementation of this methodology in this paper is through the use of the auto.arima() function in the programming language R. This function uses the Hyndman-Khandakar algorithm for automatic ARIMA(p, d, q)(P,D,Q) modelling (Hyndman & Khandakar, 2008), where p is the order of the autoregressive process, d is order of difference until the data become stationary and q is the order of the moving average process.
Next, in auto.arima() the number of differences, which is the order of d($0 \leq d \leq 2$), is determined using repeated KPSS tests, until the data become stationary. Then, it examines the ACF (to estimate the order of q) and PACF (to estimate the order of p) and chooses the appropriate model(s) for the data. It tries the chosen model(s) and uses the AIC to search for a better

model. The command stepwise=False was also added in the function, so that a vast number of different models would be tried, before the model with the lowest AIC value was chosen. Lastly, the function also compares the chosen models with and without seasonality parameters and chooses the best according to AIC values.
Then, the residuals from your chosen models must be checked manually by plotting the ACF of the residuals, and by doing a Ljung-Box test. The null hypothesis H0 of this test is that the residuals are independently distributed, which means they are white noise, while the alternative hypothesis H1 is that they are not. If they do not look like white noise, a modified model will be tried. Once the residuals look like white noise, the forecasts from January 2015 to December 2015 will be calculated using the forecast() function in R. (Hyndman & Athanasopoulos, 2018)

3.6. ARCH, GARCH and the ARMA-GARCH approach

Engle (1982) was the first to propose the autoregressive conditional heteroscedasticity (ARCH) model for modeling the changing variance of a time series. Let's assume that the return series of a financial asset denoted by r_t is often a serially uncorrelated sequence with zero mean, even as it exhibits volatility clustering. This suggests that the conditional variance of r_t given past returns is not constant. The conditional variance (also known as conditional volatility) of r_t will be denoted by ${\sigma^2}_{t|t-1}$, with the subscript t − 1 signifying that the conditioning is upon returns through time t − 1. When r_t is available, the squared return r_t^2 provides an unbiased estimator of ${\sigma^2}_{t|t-1}$. A series of large squared returns may indicate a relatively volatile period. On the other hand, a series of small squared returns may indicate a relatively quiet period. The ARCH model is a regression model with the conditional volatility as the response variable and the past lags of the squared return as the variables. For example, the ARCH(1) model assumes that the return series r_t is generated with the following equation:

$$r_t = \sigma_{t|t-1}\varepsilon_t \quad (3.6.4)$$

where,

$${\sigma^2}_{t|t-1} = \omega + \alpha r_{t-1}^2 \quad (3.6.5)$$

where α and ω are unknown parameters. ε_t is a sequence of identically and independently distributed random variables each with zero mean and unit variance (known as innovations). ε_t is also independent of r_{t-j} , $j = 1,2, \ldots$. ε_t is presumed to have unit variance so that the conditional variance of r_t is equal to ${\sigma^2}_{t|t-1}$.
The ARCH model is mainly used to predict the future conditional variances. For instance, in order to forecast the h-step-ahead conditional variance

$${\sigma^2}_{t+h|t} = E({r^2}_{t+h}|r_t, r_{t-1}, \ldots) \quad (3.6.6)$$

for h=1 (ARCH(1) model), equation (6) can be expressed as:

$${\sigma^2}_{t+1|t} = \omega + \alpha r_t^2 = (1-a)\sigma^2 + \alpha r_t^2 \quad (3.6.7)$$

which is a weighted average of the long-run variance and the current squared return.

In practice, it can be expected that the improved forecasting accuracy can be achieved by the inclusion of all past squared returns with lesser weight for more distant volatilities. One approach is to include further lagged squared returns in the model. The ARCH(q) model, proposed by Engle (1982), generalizes Equation (3.6.5), by specifying that

$$\sigma^2{}_{t|t-1} = \omega + \alpha_1 r_{t-1}^2 + \alpha_2 r_{t-2}^2 + \cdots + \alpha_q r_{t-q}^2 \qquad (3.6.8)$$

where q is the order of the ARCH process.
Bollerslev (1986) and Taylor (1986) proposed another approach, where p lags of the conditional variance are introduced in the model, where p is referred to as the GARCH order. The combination of these two approaches formed a model which is called the generalized autoregressive conditional heteroscedasticity, GARCH(p,q) model. The equation of this model is:

$$\sigma^2{}_{t|t-1} = \omega + \beta_1 \sigma^2{}_{t-1|t-2} + \cdots + \beta_p \sigma^2{}_{t-p|t-p-1} + \alpha_1 r_{t-1}^2 + \cdots + \alpha_q r_{t-q}^2 \qquad (3.6.9)$$

In terms of the backshift operator B equation (9) can be expressed as follows:

$$\left(1 - \beta_1 B - \cdots - \beta_p B\right)\sigma^2{}_{t|t-1} = \omega + (\alpha_1 B + \cdots + \alpha_q B^q) r_t \qquad (3.6.10)$$

At this point, it is crucial to mention the fact that in some of the literature, the notation GARCH(p,q) is written as GARCH(q,p), which means that, the orders are switched. It can be rather confusing but true that the two different sets of conventions are used in different software. In this paper the convention which will be used by the R-language will be analyzed before fitting or interpreting a GARCH model.
Another important detail that should be outlined for a GARCH(p,q) model, is that because conditional variances must be non-negative, the coefficients in a GARCH model are often constrained to be non-negative. However, the non-negative parameter constraints are not necessary for a GARCH model to have non-negative conditional variances with probability 1 (Nelson and Cao, 1992 & Tsai and Chan, 2006).

A GARCH model can be generalized in numerous ways. An assumption for the GARCH model is that the conditional mean of the time series is zero, but even for financial time series, such as freight rates, this strong assumption doesn’t need to always hold.

In this thesis, an extension of the GARCH model is going to be used to forecast the future freight rate values of 2015. The reason of choosing this approach was based on the fact that the portmanteau(known also as Ljung-Box) diagnostic tests for a simple GARCH approach were proven to be inconsistent and rendered the model not valid for all the time series. Thus, the approach of choice will be an ARMA-GARCH approach, where the conditional mean structure will be modeled by an ARMA(u,v) model, while the white noise term of the ARMA model will be modeled by some GARCH(p,q) model. If Y_t are the freight rate time series, then it will be given by the equation:

$$Y_t = \varphi_1 Y_{t-1} + \varphi_2 Y_{t-2} + \cdots + \varphi_u Y_{t-u} + \theta_0 + e_t - \theta_1 e_{t-1} - \theta_2 e_{t-2} - \cdots \theta_v e_{t-v} \qquad (3.6.11)$$

where,
$e_t = \sigma_{t|t-1}\varepsilon_t$
and
$\sigma^2{}_{t|t-1} = \omega + \alpha_1 e_{t-1}^2 + \cdots + \alpha_q e_{t-q}^2 + \beta_1 \sigma^2{}_{t-1|t-2} + \cdots + \beta_p \sigma^2{}_{t-p|t-p-1}$

The orders of the ARMA will be identified based on the freight rate time series Y_t ,as described in chapter 3.6, whereas the GARCH orders will be identified based on the squared residuals of the fitted ARMA model. Once the orders have been identified, full maximum likelihood estimation for the ARMA-GARCH model will be carried out by maximizing the likelihood function (Brooks, 2019).
This approach was also implemented in R programming language using the statistical package "rugarch", which uses a garch(q,p) approach in which the q denotes the ARCH order, while p denotes the GARCH order. The ARMA models will be modeled, and the residuals of these models will be modeled using a GARCH approach. The orders of the ARMA models will be decided based on the ARIMA approach described in the previous section (as an ARIMA(u,1,v) can be expressed as ARMA(u+1, v). Differencing the data was avoided, because this led to huge forecasting errors.
A general error distribution was assumed for the data, since the data were not normally distributed according to Shapiro-Wilk test. Then, the order for the GARCH approach were chosen after a lot of trials up to order (3,3) based on the AIC values of the tried models.
After choosing the orders of the GARCH models, diagnostic checks will be run for every model. The first check is Ljung-Box test, which checks if the residuals are independently distributed, and Kolmogorov-Smirnov test to check the if the errors also followed a general error distribution. If the p-value of this test is greater than 0.05, then it means the assumption about the general error distribution for the data was correct.
As a final step the forecasts for 2015 were made, using the function ugarchforecast(). The analysis and results of this approach will be shown in the next chapter.

3.7. VAR-VECM approach

In this section, the basic vector autoregressive and error correction models, neglecting deterministic terms and exogenous variables, are going to be introduced. For a set of K time series variables $y_t = (y_{1t,} \dots, y_{Kt})'$, a VAR (Vector Auto Regression) model captures their dynamic interactions. The basic model of order p (VAR(p)), without deterministic trends and exogenous variables is shown in the following equation:

$$y_t = A_1 y_{t-1} + \cdots + A_p y_{t-p} + u_t \qquad (3.8.1)$$

where the Ai's are (K ×K) coefficient matrices $u_t = (u_{1t,} \dots, u_{Kt})'$ is an unobservable error, which is usually assumed to be a zero-mean independent white noise process with time-invariant, positive definite covariance matrix $E(u_t u'_t) = \Sigma_u$. This means that the u_t are independent stochastic vectors with $u_t \sim (0, \Sigma_u)$.
The process can be considered stable on the condition that

$$\det\left(I_K - A_1 z - \cdots - A_p z^p\right) \neq 0 \; for \; |z| \leq 1 \qquad (3.8.2)$$

which means that the polynomial defined by the determinant of the autoregressive operator has no roots in and on the complex unit circle. Given the assumption that the process has been initiated in the infinite past $t = 0, \pm 1, \dots$), it generates stationary time series that have time-invariant means, variances and covariance structure. On the other hand, if the polynomial (3.8.2) has a unit root (the determinant is 0 for z=1), it can be derived that some or all the variables are integrated (it will be assumed that they are at most I(1) for convenience reasons). In case that the variables have a common stochastic trend, there is a possibility that linear combination of them which can be integrated of order 0, exist. If that is the case, the variables are cointegrated. In other words, a set of I(1) variables is called cointegrated if a linear combination exists that is I(0). Most of the times, it is convenient to consider systems with both I(1) and I(0) variables. Thereby the concept of cointegration is extended by calling any linear combination that is I(0) a cointegration relation, even if this terminology is not in line with the original definition because it could be concluded that a linear combination of I(0) variables is called a cointegration relation.

At this point, it should be mentioned that even though the model (3.8.1) is general enough to accommodate variables with stochastic trends, it is not the most appropriate type of model if interest centers on the cointegration relations because they do not appear explicitly. The VECM (Vector Error Correction model) form is a more appropriate model setup when it comes to cointegration analysis. The equation of VECM is:

$$\Delta y_t = \Pi y_{t-1} + \Gamma_1 \Delta y_{t-1} + \cdots + \Gamma_p \Delta y_{t-p+1} + u_t \qquad (3.8.3)$$

where $\Pi = -(I_K - A_1 - \cdots - A_P)$ and $\Gamma_i = -(A_{t+1} - \cdots - A_P)$ for $i = 1, \ldots, p-1$

The VECM is obtained from the levels VAR form (3.8.1) by subtracting y_{t-1} from both sides and by rearranging the terms. Because Δy_t does not contain stochastic trends by our assumption that all variables can be at most I(1), Πy_{t-1} is the only term that includes I(1) variables. As a result, Πy_{t-1} must be I(0) too. Hence, it contains the cointegrating relations. $\Gamma_j\ for\ (j = 1, \ldots, p-1)$, are short-run or short-term parameters, while Πy_{t-1} is the long-run or long-term part of the model.

Suppose than rank(Π) = r < k , where k = the number of the variables in the model, then Π can be decomposed as Π = αβ' , where α and β are $(k \times r)$ matrices. The rank of Π is called the cointegration rank of the system. For example, if there are 3 variables and r = 2, then the following relationship stands:

$$\Pi y_{t-1} = \alpha \beta'^{y_{t-1}} = \begin{bmatrix} \alpha_{11} & \alpha_{12} \\ \alpha_{21} & \alpha_{22} \\ \alpha_{31} & \alpha_{32} \end{bmatrix} \begin{bmatrix} \beta_{11} & \beta_{21} & \beta_{31} \\ \beta_{12} & \beta_{22} & \beta_{32} \end{bmatrix} \begin{bmatrix} y_{1,t-1} \\ y_{2,t-1} \\ y_{3,t-1} \end{bmatrix} = \begin{bmatrix} a_{11} ec_{1,t-1} + a_{12} ec_{2,t-1} \\ a_{21} ec_{1,t-1} + a_{22} ec_{2,t-1} \\ a_{31} ec_{1,t-1} + a_{32} ec_{2,t-1} \end{bmatrix} \qquad (3.8.4)$$

where,

$$ec_{1,t-1} = \beta_{11} y_{1,t-1} + \beta_{21} y_{2,t-1} + \beta_{31} y_{3,t-1}$$

and

$$ec_{2,t-1} = \beta_{12} y_{1,t-1} + \beta_{22} y_{2,t-1} + \beta_{32} y_{3,t-1}$$

The matrix α is sometimes called the loading matrix, and it contains the weights attached to the cointegrating relations in the individual equations of the model. The matrices α and β are not unique, and thus there are many possible α and β matrices that contain the cointegrating relations or linear transformations of them. The model (3.8.3) contains several special cases that should absolutely be mentioned. If all variables are I(0), r = K and the process is stationary. If r = 0, the term Πy_{t-1} disappears in (3.8.3). In that case, Δy_t has a stable VAR representation. In other words, a stable VAR representation exists for the first differences of the variables rather than the levels variables. It is clear, that these boundary cases do not represent cointegrated systems in the usual sense of having a common trend. There are also other cases in which no cointegration in the original sense is present, although the model (3.8.3) has a cointegrating rank strictly between 0 and K. For example, if all the variables but one were I(0); then, the cointegrating rank would be K −1, although this I(1) variable would not be cointegrated with the other variables. In the same way, there could be K − r unrelated I(1) variables and r I(0) components. Generally, for each I(0) variable in the system there can be a column in the matrix β with a unit in one position and zeros elsewhere. These cases do not represent a cointegrating relation in the original sense of the term. Even if that is the case, it is convenient to include these cases in the present framework because they can be accommodated easily as far as estimation and inference are concerned (Lütkepohl et.al, 2004).

Now that these details about the VAR-VECM models have been analyzed, the procedure that will be followed to identify the models and to produce the forecasts will be presented.
As a first step, the series underwent a logarithmic transformation to stabilize the variance of the series and reduce the impact of heteroscedasticity. Of course, the forecasted data will be exponentially transformed in order to be compared with the actual data.
Next, the stationarity of the series will be examined by using KPSS tests for the data. The difference among the variables will also be checked for stationarity. The next step, which is crucial for deciding the exact form of the VAR model was to check for cointegration relations between the data. The cointegration rank of the two models was also checked by employing the maximum likelihood test for cointegration rank as described by Johansen, 1996 where the testing is conducted in a sequence and under the null hypothesis:

H_0: $r = 0$ versus H_1: $0 < r \le g$
H_0: $r = 1$ versus H_1: $1 < r \le g$
H_0: $r = 2$ versus H_1: $2 < r \le g$.
. . . .
H_0: $r = g - 1$ versus H_1: $r = g$

where r is the number of the rank of cointegration the model and g is the number of variables included in the model.
After the cointegration tests, the order for the models was decided based on AIC values and Ljung-Box diagnostic tests. If the diagnostic tests prove that the models are valid, forecasts for the year of 2015 will be produced and be compared to the actual values of that year. The procedure of calculating formal metrics for forecasting performance will be the same as the other approaches. All these processes were calculated using the R programming language and specifically the library “tsDyn”.

4. Results and analysis

In this chapter the results of the models presented in the methodology will be presented. The models which were chosen in this thesis were an autoregressive process of order 1 as naïve forecasting approach, an ARIMA approach as described by the Box-Jenkins methodology, an ARIMA-GARCH approach, as well as a VAR approach for the Worldscale rate series and a VECM approach for the Timecharter rate series.
The forecasting procedure that was followed was the following: Forecasts for every of the 8 times series used in this research were produced. Firstly, the models were chosen for the data sample. The data sample which will be used will be the observations from January 2004 to December 2014. Then, the values from January 2015 to December 2015 will be forecasted. These produced forecasts will then be compared with the actual values of these periods.(in-sample forecasting).
The quality of the forecasts will be evaluated based on formal statistical tests such as Ljung-Box tests. This procedure is also described in Chapter 3.
The forecasting performance of different models will be compared using three formal indicators for forecasting accuracy, MAE, RMSE and MAPE.

4.1. *Naïve Forecasting approach*

After utilizing the programming language R, the results of the autoregressions will be presented. The equations for each one of the time series can be seen in table 3.

Time series	AR(1) Equation	AIC
Panamax Worldscale	$\widehat{Y_t} = \mathbf{195.8971 - 0.9250Y_{t-1}}$	1319.75
Pananax Time Charter	$\hat{Y}_t = \mathbf{20,715 - 0.9784Y_{t-1}}$	2290.1
Aframax Worldscale	$\hat{Y}_t = \mathbf{149.1964 - 0.8541Y_{t-1}}$	1315.35
Aframax Time Charter	$\hat{Y}_t = \mathbf{24,328.273 - 0.9799Y_{t-1}}$	2328.15
Suezmax Worldscale	$\hat{Y}_t = \mathbf{118.6361 - 0.8524Y_{t-1}}$	1290.16
Suezmax Time Charter	$\hat{Y}_t = \mathbf{32,716.251 - 0.9725Y_{t-1}}$	2452.86
VLCC-ULCC Worldscale	$\hat{Y}_t = \mathbf{60.8704 - 0.8457Y_{t-1}}$	1175.96
VLCC-ULCC Time Charter	$\hat{Y}_t = \mathbf{42001.95 - 0.0.928Y_{t-1}}$	2611.45

Table 3. Naïve approach equations and AIC values

Following, the formal indicators for forecasting accuracy for each of the times series, mentioned in the methodology are going to be presented in Table 4.

Time series	MAE	RMSE	MAPE
Panamax Worldscale	35.76708	41.72041	32,04%
Pananax Time Charter	4156.85226	5008.22656	16,45%
Aframax Worldscale	28.39493	34.48365	31.70%
Aframax Time Charter	3598.64535	4571.59744	12.43%
Suezmax Worldscale	23.96959	27.70813	31.56%

Suezmax Time Charter	3760.26122	4755.9855	9.89%
VLCC-ULCC Worldscale	12.68741	15.03734	39.02%
VLCC-ULCC Time Charter	10728.24372	11170.1564	21.83%

Table 4. Formal indicators for forecasting performance of AR(1) models

4.2. ARIMA Approach

After utilizing the programing language R as described in the methodology, the outcomes for the 8 times series models, as well as the p-vales for the Ljung-Box test, which tests the validity of the model by checking if they residuals are independently distributed, can be summarized in table 5.

Time series	**ARIMA orders**	**Ljung-Box Test P-value**	**AIC**
Panamax Worldscale	ARIMA (1,1,2)(1,0,0)	0.8204	1293.82
Pananax Time Charter	ARIMA (1,1,2)	0.9569	2245.70
Aframax Worldscale	ARIMA (1,1,1)(2,0,0)	0.3917	-67.02
Aframax Time Charter	ARIMA (2,1,0)	0.5518	2259.26
Suezmax Worldscale	ARIMA (1,1,1) (1,0,0)	0.5151	1275.83
Suezmax Time Charter	ARIMA (0,1,1)	0.5316	2416.31
VLCC-ULCC Worldscale	ARIMA (1,1,2)	0.2972	1160.92
VLCC-ULCC Time Charter	ARIMA (0,1,1)	0.6433	2581.44

Table 5. Arima orders, Ljung-Box Test and AIC values

All the models were valid, according to the Ljung-Box test, as the p-value of the test was larger than 0.05 for all the time series, meaning that we cannot reject the null hypothesis of the test. One thing that should be mentioned is that for the time series of Aframax Worldscale rates, a logarithmic transformation of the series was used, because without being transformed the p-value of the Ljung-Box Test was 0.04818, thus the model was not appropriate within a 95% confidence interval. After the transformation, the model was valid with a p-value of 0.3918. The forecasts of this series were exponentially transformed in order to be compared with the actual values. Lastly, as expected, it can be observed that the AIC values of the ARIMA models are smaller compared to their corresponding values of the AR(1). The reason of the negative AIC value of the Aframax Worldscale rates series is that it was logarithmically transformed when the test was conducted.

Next, the formal indicators for forecasting accuracy for the ARIMA models for each of the times series will be presented in Table 6.

Time series	MAE	RMSE	MAPE
Panamax Worldscale	11.823	15.611	10.90%
Pananax Time Charter	3878.629	4813.189	15.22%
Aframax Worldscale	12.079	14.993	12.40%
Aframax Time Charter	3609.128	4375.566	12.62%
Suezmax Worldscale	7.632	9.854	8.98%
Suezmax Time Charter	3113.811	4188.849	8.11%
VLCC-ULCC Worldscale	8.724	11.141	20.94%
VLCC-ULCC Time Charter	10795.880	11367.296	21.90%

Table 6. Formal indicators for forecasting performance of ARIMA models

4.3. ARMA-GARCH approach

The results of the methodology described section 3.7, will now be presented.

Firstly, the ARMA order, the GARCH orders, as well as the AIC values of the fitted models for each of the times series will be analyzed in table 7.

Time series	ARMA orders	GARCH orders	AIC
Panamax Worldscale	(2,2)	(2,1)	9.4739
Pananax Time Charter	(2,2)	(3,1)	16.649
Aframax Worldscale	(2,1)	(1,1)	9.5855
Aframax Time Charter	(3,0)	(1,1)	16.711

Suezmax Worldscale	(2,1)	(1,1)	9.4447
Suezmax Time Charter	(1,1)	(1,1)	18.070
VLCC-ULCC Worldscale	(2,2)	(3,3)	7.8613
VLCC-ULCC Time Charter	(1,1)	(1,1)	19.063

Table 7. ARMA orders, GARCH orders and AIC values

Following the Ljung-Box tests p-values and Kolmogorov-Shmornov test p-values will be presented in table 8.

Time series	Ljung-Box Test p-value	Kolmogorov-Shmornov test p-values
Panamax Worldscale	0.09758133	0.9913758
Pananax Time Charter	0.11774441	0.9999428
Aframax Worldscale	0.26284403	0.9913758
Aframax Time Charter	0.79238903	1.0000000
Suezmax Worldscale	0.34169489	0.9913758
Suezmax Time Charter	0.89863330	1.0000000
VLCC-ULCC Worldscale	0.86539015	0.9256519
VLCC-ULCC Time Charter	0.85219195	1.0000000

Table 8. Ljung-Box tests p-values and Kolmogorov-Shmornov test p-values

Now that the orders for the ARMA and GARCH processes have been identified, and the diagnostics test have proven that the models are appropriate for each individual freight rate time series. The indicators for forecasting performance will be presented in table 9.

Time series	MAE	RMSE	MAPE
Panamax Worldscale	13.3581	16.9118	12.13%
Pananax Time Charter	3566.2525	4482.5535	13.95%
Aframax Worldscale	11.7682	15.66411	13.4%
Aframax Time Charter	3453.2582	4122.0345	12.14%
Suezmax Worldscale	12.6626	15.9489	17.09%
Suezmax Time Charter	3140.4488	4212.1700	8.18%
VLCC-ULCC Worldscale	10.43290	13.5530	24.52%
VLCC-ULCC Time Charter	10284.2366	10882.5517	20.83%

Table 9. Formal indicators for forecasting performance of ARMA-GARCH models

4.4. VAR-VECM approach

All the time series were found to be non-stationary, as the p-values for every series were less than 0.01.
Before deciding on the model form, it was crucial for cointegration checks between the variables to be conducted. Time series plots for the Worldscale and the Time-Charter series can be seen in figures 6 and 7 respectively.

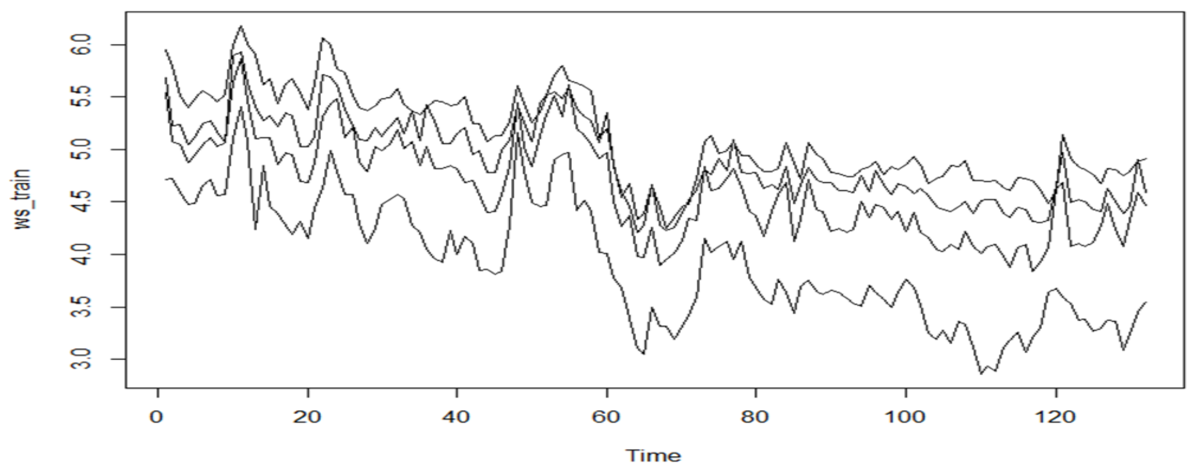


Figure 2. All four Worldscale Time Series transformed in logarithms

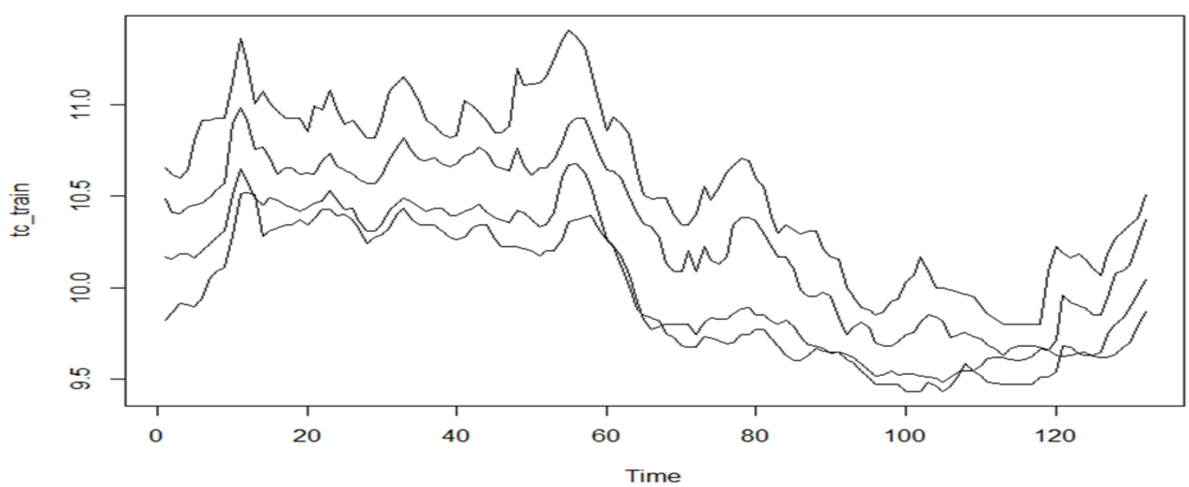


Figure 3. All four Time-Charter Time Series transformed in logarithms

From these two figures, it can be observed that the pattern of the four Worldscale time series is quite similar. This fact can justify that there may be some economic structure that ties the series together. The same can be observed for the four time-charter series. At this point, the assumption that these structures are linear will be made, for the sake of not overcomplicating the analysis. After observing these plots, it was decided to model Worldscale and timecharter series separately. To examine more deeply the relation between these series, the correlations of Worldscale series and the correlations of the timecharter series can be seen in tables 10 and 11 respectively.

	Panamax	Aframax	Suezmax	VLCC-ULCC
Panamax	1	0.9516385	0.9252270	0.9102517
Aframax		1	0.9686252	0.9220332
Suezmax			1	0.9449819
VLCC-ULCC				1

Table 10. Correlations of Worldscale series (series in logarithms)

	Panamax	Aframax	Suezmax	VLCC-ULCC
Panamax	1	0.9688947	0.9392742	0.9157590
Aframax		1	0.9762323	0.9620857
Suezmax			1	0.9783685
VLCC-ULCC				1

Table 11. Correlations of Timecharter series (series in logarithms)

Following, KPSS tests were conducted in order to check if the differences of logarithmically transformed time series were stationary. In other words, if they were cointegrated, as mentioned in the methodology. The p-values of the KPSS tests for the Worldscale and Timecharter time series can be seen in tables 12 and 13 respectively. It can be observed that most of the differences between the time series (except the differences “PanamaxWS – SuezmaxWS” and “AframaxTC-SuezmaxTC” for which p-values of KPSS test were greater than 0.1) contain a unit root, in other words they are not stationary.

	Aframax	Suezmax	VLCC-ULCC
Panamax	0.02307	>0.1	<0.01
Aframax		0.02056	<0.01
Suezmax			<0.01

Table 12. p-values for the KPSS test between the Worldscale series (series in logarithms)

	Aframax	Suezmax	VLCC-ULCC
Panamax	<0.01	0.01063	<0.01
Aframax		>0.1	0.02309
Suezmax			0.03302

Table 13. p-values for the KPSS test between the Timecharter series (series in logarithms)

Next the p-values for the cointegration rank tests as described by Johansen, 1996 will be presented in tables 14.

	0 vs 1	1 vs 2	2 vs 3	3 vs 4
Worldscale series	<0.01	<0.01	0.022	0.013
Timecharter series	<0.01	<0.01	0.516	0.144

Table 14. p-values for Cointegration test of Johansen 1996

According to table 14, it can be concluded based on the p-values of the test that the rank for the Worldscale time series is 4 and the rank for the Timecharter series is 2.
Based on the cointegration test results of the Johansen test, it was decided to use a VAR model for the Worldscale times series (because rank = 4) and a VEC model for the Timecharter time series (because rank=2). The lag order for the two models was 1 based on AIC and Ljung-Box diagnostics tests. The AIC values for the Worldscale and the Timecharter models were -2118.411 and -3167.131 respectively. The p-values of the Ljung-Box test for the models can be seen in table 15 and 16.

Time series	Ljung-Box test p-values
Panamax Worldscale	0.098
Aframax Worldscale	0.4197
Suezmax Worldscale	0.2075
VLCC-ULCC Worldscale	0.1423

Table 15. p-values of the Ljung-Box test for the Worldscale VAR model

Time series	Ljung-Box test p-values
Panamax Timecharter	0.7162
Aframax Timecharter	0.3075
Suezmax Timecharter	0.6897
VLCC-ULCC Timecharter	0.7130

Table 16. p-values of the Ljung-Box test for the Timecharter VEC model

Now that the orders for the VAR and VEC models have been identified, and the diagnostics test have proven that the models are valid. The indicators for forecasting performance will be presented in table 17.

Time series	**MAE**	**RMSE**	**MAPE**
Panamax Worldscale	17.200	22.519	16.01%
Pananax Time Charter	2175.782	2832.626	8.51%

Aframax Worldscale	13.464	18.206	15.64%
Aframax Time Charter	2827.773	3268.735	10.09%
Suezmax Worldscale	9.105	10.452	11.64%
Suezmax Time Charter	2811.237	3737.536	7.34%
VLCC-ULCC Worldscale	5.980	7.821	17.41%
VLCC-ULCC Time Charter	4451.821	5297.731	8.87%

Table 17. Formal indicators for forecasting performance of VAR-VECM models

4.5. Comparison of different forecasting approaches

At this point, after having presented the analysis and results for each individual forecasting model, a comparison of these models will be conducted in order to determine the best forecasting approach for each of the eight freight rates time series examined in this research.

4.5.1. Panamax Worldscale rates

Firstly, the formal indicators for the Panamax Worldscale rate time series for each different modeling approach will be presented in table 18.

Panamax Worldscale rates	MAE	RMSE	MAPE
Naïve	35.767	41.7204	32,04%
ARIMA	11.823	15.611	10.90%
ARMA-GARCH	13.3581	16.9118	12.13%
VAR	17.200	22.519	16.01%

Table 18. Formal indicators of forecasting performance for Panamax Worldscale rate time series

It is obvious that according to all indicators, ARIMA outperforms both Naïve and Var approach, and slightly outperforms the ARMA-GARCH approach. That means, that the choice of trying to model the error of the ARMA process using a GARCH approach proved not only not to increase the forecasting performance but also to produce bigger errors than the original ARIMA approach.

4.5.2. Aframax Worldscale rates

The formal indicators for the Aframax Worldscale rate time series for each different modeling approach will be presented in table 19.

Aframax Worldscale rates	MAE	RMSE	MAPE
Naïve	28.3949	34.4836	31.70%
ARIMA	12.079	14.993	12.40%
ARMA-GARCH	11.7682	15.66411	13.4%
VAR	13.464	18.206	15.64%

Table 19. Formal indicators of forecasting performance for Aframax Worldscale rate time series

The results in the case of Aframax Worldscale rates are almost the same with the results of the Panamax Worldscale rates, with the difference that ARIMA outperforms slightly both ARMA-GARCH and VAR models. It should also be mentioned that according to MAE, the ARMA-GARCH model slightly outperforms the ARIMA model. Of course, due to reasons discussed in the paper it was decided that the importance of RMSE was greater in the case of crude oil freight rates.

4.5.3. Suezmax Worldscale rates

The formal indicators for the Suezmax Worldscale rate time series for each different modeling approach will be presented in table 20.

Suezmax Worldscale rates	MAE	RMSE	MAPE
Naïve	23.9695	27.7081	31.56%
ARIMA	7.632	9.854	8.98%
ARMA-GARCH	12.6626	15.9489	17.09%
VAR	9.105	10.452	11.64%

Table 20. . Formal indicators of forecasting performance for Suezmax Worldscale rate series

In the case of Suezmax Worldscale rates, according all three formal indicators, the ARIMA model heavily outperforms the naïve and ARMA-GARCH models, while it slightly outperforms the VAR approach.

4.5.4. VLCC-ULCC Worldscale rates

The formal indicators for the VLCC-ULCC Worldscale rate time series for each different modeling approach will be presented in table 21.

VLCC-ULCC Worldscale rates	MAE	RMSE	MAPE
Naïve	12.6874	15.0373	39.02%
ARIMA	8.724	11.141	20.94%
ARMA-GARCH	10.43290	13.5530	24.52%
VAR	5.980	7.821	17.41%

Table 21. Formal indicators of forecasting performance for VLCC-ULCC Worldscale rate series

In contrast to the other 3 Worldscale rate series, in the case of VLCC-ULCC Worldscale rates, the VAR model heavily outperforms the rest of the models. As it can be seen though, the forecasting errors are far too huge, as the MAPE metric shows an average 17% difference between the actual and the forecasted values, which is way bigger than the forecasting percentage error that VAR produced for the rest of the vessel size categories.

Based on the results of all the Worldscale series, it could be derived that ARIMA models produce better forecasts for the categories of Panamax, Aframax and Suezmax Worldscale time series. This could mean that these three categories depend more on their lagged values than they do on the lagged values of the different vessel size categories. VLCC-ULCC on the other hand seem to produce better forecasts when modelled with the lagged values of different vessel size categories. Having presented all these results, the fact that using models that only contain freight rate data without the use of exogenous variables produce large forecasting errors. This can be also be seen in the tables, as the smallest forecasting percentage error belong to the ARIMA approach for the Suezmax freight rates with the value of 8.98%.

4.5.5. Panamax Time-Charter rates

The formal indicators for the Panamax Timecharter rate time series for each different modeling approach will be presented in table 22.

Panamax Time Charter rates	MAE	RMSE	MAPE
Naïve	4156.85226	5008.22656	16,45%
ARIMA	3878.629	4813.189	15.22%
ARMA-GARCH	3566.2525	4482.5535	13.95%
VECM	2175.782	2832.626	8.51%

Table 22. Formal indicators of forecasting performance for Panamax Timecharter rate time series

According to all three indicators, the VECM approach outperforms the rest of the approaches by at least 5% in terms of forecasting accuracy. It can also be observed that for the Panamax timecharter rates the ARMA-GARCH approach produced better forecasts than the ARIMA approach, while the naïve approach forecasts were slightly outperformed by ARIMA.

4.5.6. Aframax Time-Charter rates

The formal indicators for the Aframax Timecharter rate time series for each different modeling approach will be presented in table 23.

Aframax Time Charter rates	MAE	RMSE	MAPE
Naïve	3598.6453	4571.5974	12.43%
ARIMA	3609.128	4375.566	12.62%
ARMA-GARCH	3453.2582	4122.0345	12.14%
VECM	2827.773	3268.735	10.09%

Table 23. Formal indicators of forecasting performance for Aframax Timecharter rate time series

In the case of Aframax timecharter rates, VECM is outperforming the rest of the models but the produced forecasts only differ for 2%. The forecast accuracy of the other three models is almost the same with a MAPE of 12%.

4.5.7. Suezmax Time-Charter rates

The formal indicators for the Suezmax Timecharter rate time series for each different modeling approach will be presented in table 24.

Suezmax Time Charter rates	MAE	RMSE	MAPE
Naïve	3760.26122	4755.9855	9.89%
ARIMA	3113.811	4188.849	8.11%
ARMA-GARCH	3140.4488	4212.1700	8.18%
VECM	2811.237	3737.536	7.34%

Table 24. Formal indicators of forecasting performance for Suezmax Timecharter rate time series

For Suezmax timecharter rates, the VECM approach seem to produce the slightly better forecasts than the other three models. The forecasting performance of ARIMA and ARMA-GARCH is almost the same with a MAPE of 8%, while the naïve approach is outperformed by the other three approaches.

4.5.8. VLCC-ULCC Time-Charter rates

The formal indicators for the Suezmax Timecharter rate time series for each different modeling approach will be presented in table 25.

VLCC-ULCC Time Charter rates	MAE	RMSE	MAPE
Naïve	10728.24372	11170.1564	21.83%
ARIMA	10795.880	11367.296	21.90%
ARMA-GARCH	10284.2366	10882.5517	20.83%
VECM	4451.821	5297.731	8.87%

Table 25. Formal indicators of forecasting performance for VLCC-ULCC Timecharter rate time series

For the VLCC-ULCC timecharter rates, it is obvious that the VECM approach heavily outperforms all the other approaches. Naïve approach, ARIMA approach and ARMA-GARCH approach produce huge forecast errors with a MAPE of 21-22%, while the MAPE of VECM is almost 9%.

Based on the results of all the timecharter series, It was observed that the VECM approach was the best forecasting model for every vessel size. This could mean that when the values of timecharter rates depend more on the lagged values different vessel size categories. Even if VECM was the forecasting model for all the timecharter series without the use of exogenous variables, their MAPE values indicate that even the VECM approach produces large forecasting errors ranging between 7.3-10%.

5. Conclusions

5.1. Answers to Research and Sub-research questions

At this point of the thesis, answers to the sub-research questions as well as the main research question of this thesis will be presented.

The first sub-research question was: “What are the main variables that affect tanker freight rates?”
Answer: As stated in Chapter 2 of this thesis, according to existing literature review, the tanker freight market is characterized by the interaction of supply and demand for tanker shipping services. Demand is affected by the international trade in oil and oil products, which in turn depends on world economic activity and imports and consumption of energy commodities, while the supply is affected by the size of the tanker fleet, tanker shipbuilding activities, bunker fuel prices, the tonnage available for trading, the scrapping rate of the tanker fleet, and the productivity of the tanker fleet at any point in time. Other variables, such as world economic activity, growth in industrial production, seaborne trade in commodities, oil prices, availability of vessel tonnage, new vessel buildings on order, tanker shipbuilding deliveries and the scrapping rates determine can play a role in determining freight rates for sea transportation.

The second sub-research question was: “What are the possible forecasting approaches?”

Answer: As also stated in chapter 2, the forecasting approaches which have been used in order to produce freight rate forecasts are univariate time series models, such as ARIMA and ARCH-type models, multivariate time series models, such as VAR and VECM models, as well as Neural Networks.

The third sub-research question was: “How can the quality of these forecasts be evaluated?” Answer: The quality of these forecasts, was evaluated based on formal statistical tests for each approach, described in the methodology, as well as formal indicators for forecasting accuracy. The indicators chosen for this thesis were MAE, RMSE and MAPE.

Having these sub-research questions answered the main research question can be answered. The main research question was: “What is the most efficient model for forecasting crude oil tanker freight rates?”

Answer: Following the assumption of that the market is approximately efficient (Evans, 1994) and that freight rates contain all publicly available information (Fama, 1970, Nicholson ,1989).), four modelling approaches were chosen in order to test if reliable forecasts could be produced without the use of exogenous variables, using real monthly Worldscale and timecharter rates of different crude oil vessel sizes. An autoregressive process of order 1 was chosen as naïve forecasting approach, followed by an ARIMA, ARMA-GARCH and VAR-VECM approaches. After deciding the orders for these models and running diagnostics test to test their validity, forecasts were produced for twelve periods ahead. The results indicated that ARIMA approach produced better forecasts for Panamax, Aframax and Suezmax Worldscale rates, while VAR-VECM produced better forecasts in the case of VLCC-ULCC Worldscale rates. Regarding timecharter rates, it was concluded that VECM was the most appropriate approach among the chosen approaches to model these type of times series. It should although be mentioned that even the forecasts that were considered better among all the forecasts produced in this thesis, had large forecasting errors with a minimum MAPE of 7.34%. These large forecasting errors can also be attributed to the fact that the samples used for the chosen models had great variability according to descriptive statistics presented in Table 2. For example, the coefficient of variation for the VLCC-ULCC Worldscale rates was the largest among the series with the value of 63.44%. The forecasts errors for this specific series turned out to be the largest. In the data sample used (2004-2014), there was a huge decrease in the value of the variables after 2008, the period of the economic crisis. So, it could be argued that exogenous economic variables can affect the formation of the values of the freight rates. This can mean that a large part of the freight rates is stochastic in nature and cannot be forecasted efficiently for twelve months ahead by using only freight rate data and the methods applied in this paper.

5.2. Limitations to this research

At this point, it should be mentioned that such a research can’t come with some limitations. The first of these limitations is that the research was done based on a specific sample of observations from 2004 to 2014. The same results cannot be guaranteed if another sample of observations is used.

The second limitation is that only some of all the possible forecasting approaches were used, for reasons stated in chapter 3.

The third limitation is that exogenous variables that according to existing literature affect tanker freight rates, were not used in the forecasting models used in this thesis, due to the assumption of an efficient market.

The final limitation is the evaluation of these forecasting was done based on 12-months (12-steps) ahead forecasts. Thus, the evaluation of these forecasting approaches for other type of forecasts, such as forecast only for the next period was not done in this thesis.

5.3. Recommendation for further research

As subjects for further research, Neural Networks such as Artificial Neural Networks and Wavelet Neural Networks could also be tried and compared to the models used in this thesis, using the same data.

Another possible approach without the use of exogenous variables could be a regime switching approach, because of the sudden fall in the values probably due to the economic crisis.

Another research subject could be the use of the proposed models for the forecasting of the spot and freight rates of vessels of the same size but for different markets within the shipping sector.

Finally, the effect of different exogenous variables such as price of crude oil, economic and political shocks, as well as factors that affect the supply and demand of vessels in the crude oil tanker market, in the forecasting performance of one or more of the proposed models could be a great research subject.

Bibliography

Adland, R., & Strandenes, S. (2007) 'A Discrete-Time Stochastic Partial Equilibrium Model of the Spot Freight Market. Journal of Transport Economics and Policy, 41(2), 189-218. Available at: http://www.jstor.org.eur.idm.oclc.org/stable/20054013

Alizadeh, A. H., & Talley, W. K. (2011) 'Microeconomic determinants of dry bulk shipping freight rates and contract times' Transportation, 38(3), 561-579.

Alizadeh, A. H., & Talley, W. K. (2011) 'Vessel and voyage determinants of tanker freight rates and contract times', Transport Policy, 18(5), 665-675.

Batchelor, R., Alizadeh, A. and Visvikis, I. (2007) 'Forecasting spot and forward prices in the international freight market', International Journal of Forecasting 23, pp. 2006–2008. doi: 10.1016/j.ijforecast.2006.07.004.

Beenstock, M., & Vergottis, A. (1989). 'An econometric model of the world market for dry cargo freight and shipping'. *Applied Economics*, *21*(3), 339-356.

Beenstock, M., & Vergottis, A. (1993), 'Econometric modelling of world shipping', Springer Science & Business Media.

Bessler, David & Jonnala, Sneha & Fuller, Stephen. (2002). A GARCH Approach to Modelling Ocean Grain Freight Rates. Maritime Economics and Logistics. 4. 103-125

Bollerslev, T. (1986), 'Generalized autoregressive conditional heteroskedasticity', Journal of econometrics, 31(3), 307-327.

Box, G.E.P., Jenkins, G.M. and Reinsel, G.C. (1994) 'Time Series Analysis; Forecasting and Control. 3rd Edition', Prentice Hall, Englewood Cliff, New Jersey.

Box, G. E., Jenkins, G. M., Reinsel, G. C., & Ljung, G. M. (2015). 'Time series analysis: forecasting and control'. John Wiley & Sons.

Box, G. E. P. and Pierce, D. A., `Distribution of the residual autocorrelations in autoregressive integrated moving-average time series models', Journal of the American Statistical Association, 65 (1970), 1509±26.

Brockwell P, Davis R. (2001) 'Introduction to Time Series and Forecasting', Springer Texts in Statistics.

Brooks, C. (2019), 'Introductory econometrics for finance', Cambridge university press.

Caner, M. and Kilian, L. (2001) 'Size distortions of tests of the null hypothesis of stationarity : evidence and implications for the PPP debate', 20, pp. 639–657.

Clarksons Shipping Intelligence network, viewed 20 June 2019, available at: https://sin.clarksons.net/

Davidson, R., & MacKinnon, J. G., (1993). 'Estimation and inference in econometrics', OUP Catalogue

D. Hawdon (1978) Tanker freight rates in the short and long run, Applied Economics, 10:3, 203-218, DOI: 10.1080/758527274

Engle, R. F., (1982), 'Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation', Econometrica: Journal of the Econometric Society, 987-1007.

Evans, J. J. (1994), ‘An analysis of efficiency of the bulk shipping markets’, Maritime Policy and Management 21,31 l-329.

Fama E. F. (1970), ‘Efficient capital markets: a review of theory and empirical work’, Journal of Finance 25, 383423.

Fan, S. et al. (2013) ‘Forecasting Baltic Dirty Tanker Index by Applying Wavelet Neural Networks’, Journal of Transportation Technologies, 03(01), pp. 68–87. doi: 10.4236/jtts.2013.31008.

Gavriilidis, K. et al. (2018) ‘Volatility forecasting across tanker freight rates: The role of oil price shocks’, Transportation Research Part E: Logistics and Transportation Review, 118, pp. 376–391. doi: 10.1016/j.tre.2018.08.012.

Geomelos, N. D. and Xideas, E. (2014) ‘Forecasting spot prices in bulk shipping using multivariate and univariate models multivariate and univariate models’, Cogent Economics & Finance. Cogent, 2(1), pp. 1–37. doi: 10.1080/23322039.2014.932701.

Hyndman, R.J., & Athanasopoulos, G. (2018) ‘Forecasting: principles and practice, 2nd edition’, OTexts: Melbourne, Australia. OTexts.com/fpp2., viewed 25 July 2019, <https://otexts.com/fpp2/arima-r.html>

Hyndman, R. J., & Khandakar, Y., (2007) ‘Automatic time series for forecasting: the forecast package for R’ (No. 6/07). Clayton VIC, Australia: Monash University, Department of Econometrics and Business Statistics.

Johansen, S. (1996), ‘Likelihood-based inference in cointegrated Vector Autoregressive Models’, Oxford University Press

Kavussanos, M. G. (1996) ‘Price Risk Modelling of Different Size Vessels in the Tanker Industry Using Autoregressive Conditional Heteroskedasticity (Arch) Models’

Keller, G. (2015). ‘Statistics for Management and Economics’, Abbreviated. Cengage Learning.

Kevin Cullinane (1992) A short-term adaptive forecasting model for BIFFEX speculation: a Box—Jenkins approach, Maritime Policy & Management, 19:2, 91-114, DOI: 10.1080/03088839200000018

Koopmans, T. C. (1939), ‘Tanker Freight Rates and Tankship Building: An Analysis of Cyclical Fluctuations’, by Dr. T. Koopmans (No. 27). De erven F. Bohn nv.

Kuo, H. S., Hwang, I. S., Fu, T. Y., Wu, J. Y., Chang, C. C., & Tsong, T. T. (2004). Preparation and characterization of single-atom tips. nano letters, 4(12), 2379-238

Lun, Y. V., Lai, K. H., & Cheng, T. E. (2010). ‘Shipping and logistics management’, London: Springer.

Lütkepohl, H., Krätzig, M. and Phillips, P. C. B., (2004) ‘Applied time series econometrics’ Cambridge university press.

Lyridis, D. V. et al. (2004) ‘Forecasting tanker market using artificial neural networks’, Maritime Economics and Logistics, 6(2), pp. 93–108. doi: 10.1057/palgrave.mel.9100097.

Mandelbrot, B. B., (1963), ‘The Variation of Certain Speculative Prices’, The Journal of Business 36, No. 4, 394-419

Maridakis S. et. al (1997) ‘ARMA Models and the Box Jenkins Methodology’, Journal of Forecasting, Vol. 16, 147±163

Maritime-connector, viewed 11 July 2019, http://maritime-connector.com

Marquardt, D. W., `An algorithm for least squares estimation of non-linear parameters', Soc. Indust. Appl. Math., 11 (1963), 431±41.

Nelson, D. B., & Cao, C. Q. (1992), ‘Inequality constraints in the univariate GARCH model’, Journal of Business & Economic Statistics, 10(2), 229-235.

N. Wijnolst, F. W., (1999). 'Shipping Industry Structure'. Delft: Delft University Press.

Nicholson, W. (1989), ‘Microeconomic Theory, 4th ed.’, The Dryden Press International, Chicago, Illinois.

Perron, P. (1989). The great crash, the oil price shock, and the unit root hypothesis. Econometrica: Journal of the Econometric Society, 1361-1401

Stopford, M. (2009) ‘Maritime Economics 3rd edition’. London: Routledge, https://doi-org.eur.idm.oclc.org/10.4324/9780203891742

R. Adland and K. Cullinane, (2006) “The Non-Linear Dynamics of Spot Freight Rates in Tanker Markets,” Transportation Research Part E: Logistics and Transportation Review, Vol. 42, No. 3, 2006, pp. 211-224. doi:10.1016/j.tre.2004.12.001

Sims, C. A. (1980), ‘Macroeconomics and reality’, Econometrica: journal of the Econometric Society, 1-48.

Strandenes, S. P. (1984), ‘Price determination in the time charter and secondhand markets’, Center for Applied Research, Norwegian School of Economics and Business Administration, Working Paper MU, 6, 15.

Taib, C. M. I. C., & Mohtar, Z. (2018) ‘Forecasting Spot Freight Rates using Vector Error Correction Model in the Dry Bulk Market’, MALAYSIAN JOURNAL OF MATHEMATICAL SCIENCES, 12(2), 235-254.

Tsay, R. S. (2013), ‘Multivariate time series analysis: with R and financial applications’, John Wiley & Sons.

Tsay, R. S. (2014), ‘An introduction to analysis of financial data with R’, John Wiley & Sons.

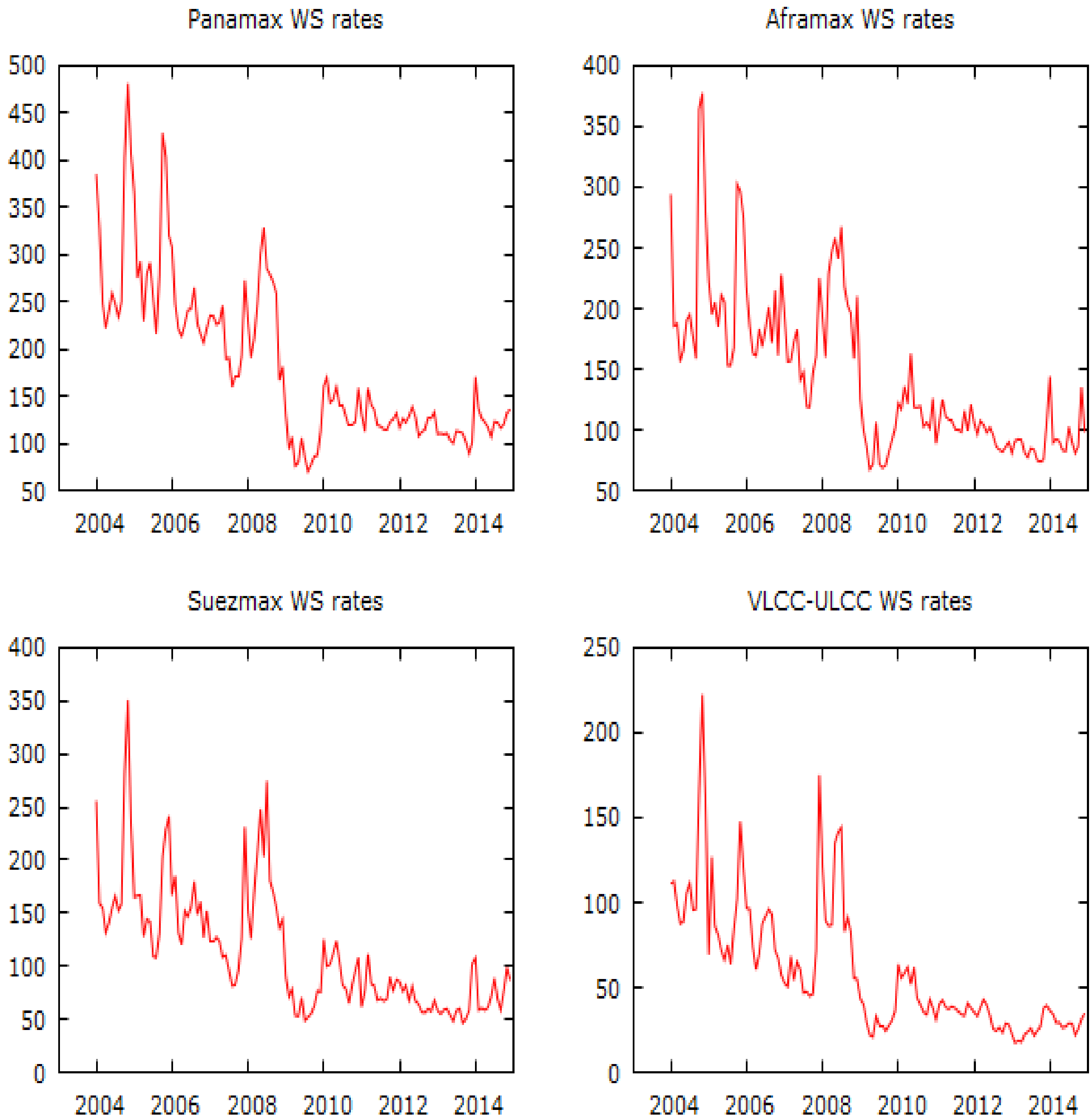
Tsai, H. and Chan, K. S. (2006), ‘A note on the non-negativity of Nelson, D. B., & Cao, C. Q. (1992), ‘Inequality constraints in the univariate GARCH model’, Journal of Business & Economic Statistics, 10(2), 229-235.

Veenstra, A. W. and Franses, P. H. (1997) ‘A co-integration approach to forecasting freight rates in the dry bulk shipping sector’, Transportation Research Part A: Policy and Practice, 31(6), pp. 447–458. doi: 10.1016/S0965-8564(97)00002-5.

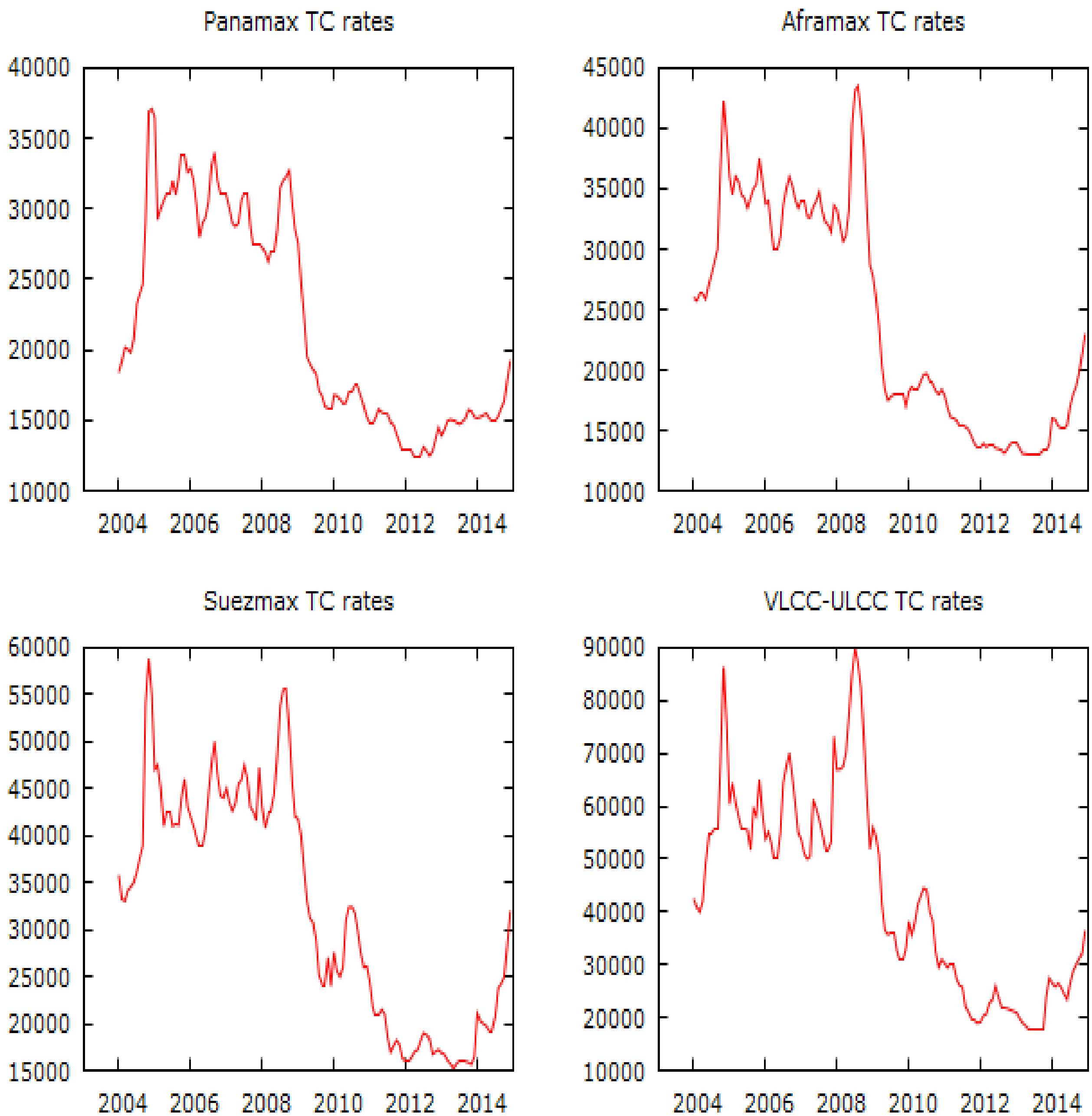
WorldScale Association (London) Limited [GB], viewed 12 July 2019, https://www.worldscale.co.uk

Zannetos, Z. (1966), ‘The Theory of Oil Tankship Rates. An Economic Analysis of Tankship Operations’.

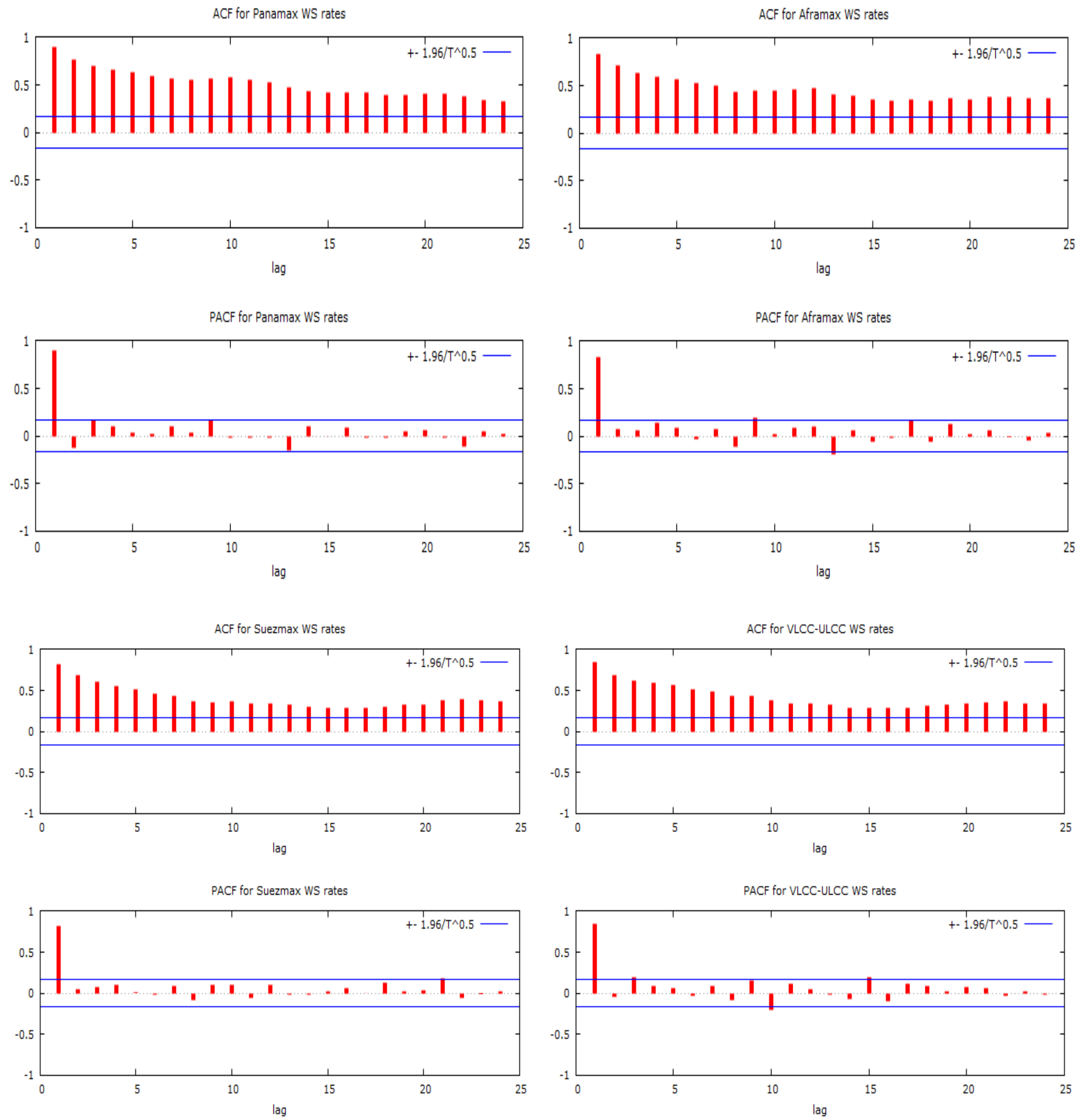
Appendix



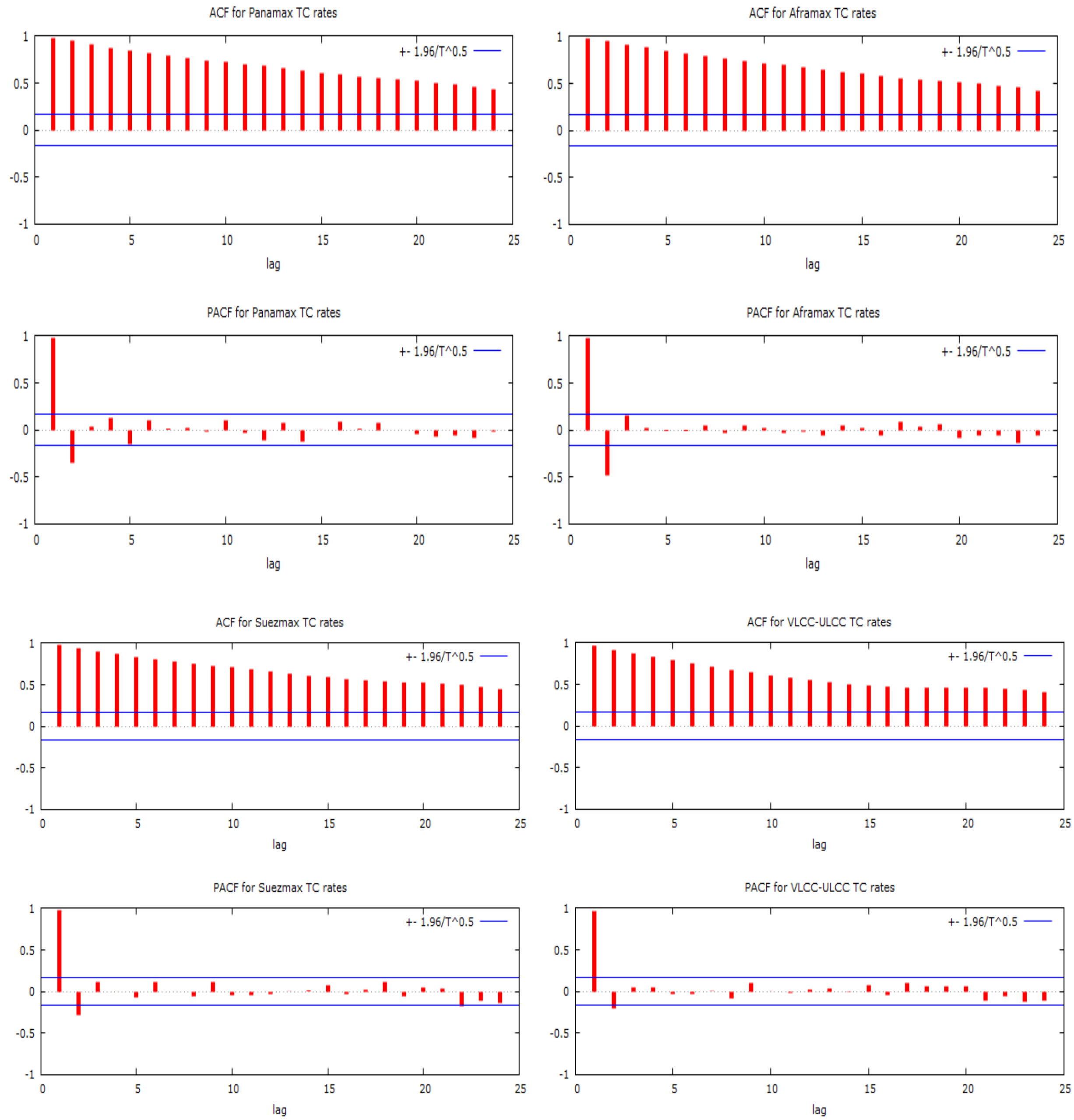
Appendix 1. Time Series plots of Worldscale Rates (January 2004 – December 2014)



Appendix 2. Time series plots of Time-Charter Rates (January 2004 – December 2014)



Appendix 3. ACF and PACF of Worldscale Rates (January 2004 – December 2014)



Appendix 4. ACF and PACF of Time-charter Rates (January 2004 – December 2014)