UNIVERSITY OF PIRAEUS
DEPARTMENT OF MARITIME STUDIES

HELLENIC NAVY ACADEMY
DEPARTMENT OF NAVAL SCIENCES

Inter Institutional

M.Sc. in Marine Science and Technology Management

Thesis

“Freight Rates Forecasting in the Dry Handy Sector”

“Dionysia Sourmpi”

“MNSND22050”

Supervisor: Dr. Ioannis Lagoudis

Piraeus

“April” “2024”

THREE-MEMBER EXAMINATION COMMITTEE:

MEMBER A΄: Ioannis Lagoudis

MEMBER B΄: Ioannis Theotokas

MEMBER C΄: Dionysis Polemis

“Acknowledgments or dedication”

I would like to dedicate this thesis to the unwavering support of my parents, whose love, guidance, and encouragement have been my foundation throughout this academic journey. Their belief in my abilities has been a constant source of motivation, and I am deeply grateful for their sacrifices and unwavering support.

I am also indebted to my friends, whose understanding, and encouragement have enriched my life beyond measure.

Additionally, I extend my sincere appreciation to my professor for their guidance, mentorship, and expertise throughout this research endeavor. His invaluable insights, constructive feedback, and dedication to academic excellence have shaped my scholarly growth and contributed significantly to the development of this thesis.

To my parents, friends, and professor, I offer my heartfelt gratitude for their belief in me, their support, and their encouragement every step of the way. This work reflects your collective influence, and I am profoundly grateful for your presence in my life.

Abstract

The purpose of this dissertation is the presentation of the forecasting freight rates with well-known methods Markov Chain which uses time series from Clarckson's, with two routes and applications in the statistical package, R. The need to forecast charter revenue for shipping companies is examined. The concept of bulk shipping and the importance of ship sizes are introduced. Various freight rates forecasting methods are mentioned such as ARMA, with an emphasis on the Markov Chain Method. The statistical analysis of data for two shipping routes to understand the tendance of both time series. The freight rates forecasting using three-dimensional transition matrices and the apply in Markov Chain Model, is studied. Also, limitations of the research are mentioned, such as the uncertainty in determining the demand and geopolitical factors. The thesis concludes with ideas for future development, including reducing uncertainty in forecasts and evolving the Markov model into a Hidden Markov Model (HMM) for more detailed forecasting.

Key - words

Freight rate forecasting, Markov Chains, Statistics, Maritime, Handysize

Contents

Abstract v
Contents vi
1. Introduction 1
1.1 Bulk Carrier Types 1
1.2 Shipping Cycles in Bulk Carrier Market 2
1.3 Chartering 3
1.4 Freight Rate Indices 5
1.4.1 Baltic Dry Index (BDI) 5
2. Freight Rates Forecasting (Literature Review) 7
3. Methodology – Markov Chains 15
3.1 Markov Chain 15
3.2 Data 19
3.3 Program R 20
4. Results 22
4.1 Historical Data and Descriptive Statistics 22
4.2 Forecasting with Markov Chains Method 30
4.2.1 Route 1 (97210) 31
4.2.2 Route 2 (97218) 32
4.2.3 Comparing the two routes 34
5. Conclusions 36
Appendix 38
References 41

1. Introduction

1.1 Bulk Carrier Types

In this thesis, our purpose is to analyze the freight rate of Handysize Bulkers. However, first, it is important to understand the Shipping Market of Bulkers and especially about the exact type that is studied. Dry cargo ships are used for transport of dry solids goods, as ores, coal, steel products, forestry products and grain, they have greater tolerance to heat and cold. These ships are equipped with cranes on it deck and others loading and unloading mechanisms. (UNCTAD,2020). A bulk carrier or bulker is a merchant ship specially designed to transport unpackaged bulk cargo, such as grain, coal, ore, steel coils, and cement, in its cargo holds. Today's bulk carriers are specially designed to maximize capacity, safety, efficiency and durability. Handysize and Handymax vessels are characterized by their versatile nature. Together, these categories constitute 71% of all bulk carriers exceeding 10,000 DWT and exhibit the most rapid growth rate. This trend is attributed in part to emerging regulations that impose stricter limitations on the construction of larger ships. This is partly due to new regulations coming into effect which put greater constraints on the building of larger vessels. Global bulk shipping transport has reached huge proportions: in 2005, 1.7 billion metric tons of coal, iron ore, grain, bauxite, and phosphate were transported by ship. Today, the world's bulk carrier fleet includes 6,225 ships of over 10,000 DWT and represents 40% of all ships in terms of tonnage and 39.4% in terms of vessels. Including smaller ships, bulk carriers have a total combined capacity of nearly 346 million DWT. Combined carriers constitute a minor fraction of the fleet, accounting for less than 3% of its capacity. Moreover, by 2005, the typical age of a bulk carrier stood slightly above 13 years. The countries with the largest fleets are Greece, Japan, and China, the top three owners of bulk carriers, with 1,326, 1,041, and 979 vessels respectively. (U.S. Department of Transportation Maritime Administration, 2006)

Asian companies dominate the construction of bulk carriers, almost 62% were built in Japan by shipyards such as Oshima Shipbuilding and Sanoyas Hishino Meisho. Also, South Korea, with notable shipyards Daewoo and Hyundai Heavy Industries, ranked second among builders.

The personnel aboard a bulk carrier usually ranges from 20 to 30 individuals, although smaller vessels can be managed by just 8. This team comprises the captain or master, the deck crew, the engine crew, and the steward's team. A mini-bulk carrier carries two to three deck officers, while larger Handysize bulk carriers carry four.

The voyages undertaken by bulk carriers are influenced by market dynamics, leading to frequent variations in both routes and cargo types. A ship may engage in the grain trade during the harvest season and later move on to carry other cargoes or work on a different route. On a coastal carrier engaged in the tramp trade, the crew frequently remains unaware of the next destination port until the cargo loading process is completed. Since discharging bulk cargo is so challenging, bulk carriers spend more time in port than other ships. (UNCTAD, 2006)

1.2 Shipping Cycles in Bulk Carrier Market

As Martin Stopford analyzes in the period 1945 – 2008, almost half of a century the average of the cycle is about 6.7 years. The cycles have similarities and differences as well. However, the cycles are defined, and the higher spots have a duration bigger than other shipping markets. This leads to the conclusion that the market is more stable, especially if we are comparing it with the wet market (tankers). Concerning cargo specialization, it led to larger types of ships. Moreover, the market was adversely affected by several political developments, in the 50's Korea had a war, in '67 Suez Canal was closed for the second time, Yom Kippur's War in '73 and the crisis about petrol in '79 (the second one as well), gulf war in the '90 and invasion to Iraq in 2003

During the cycle from 2003 to 2008, freight rates for bulkers and tankers notably surged, maintaining elevated levels for the subsequent four years despite some fluctuations. This was concurrent with China's adoption of an open market economy, attracting significant inward investment. China's intensive infrastructure development from early 2003 spurred a substantial demand for raw materials. Notably, between 2002 and 2007, China's annual steel production skyrocketed from 144 million tons to 468 million tons, rivaling the combined output of Europe, Japan, and South Korea.

Several factors influence the market, some of them are the demand and the offer of ships and cargo, the political conditions, the voyages (distance, routes) and the size of ships such as their age nowadays. The major factors though, are demand and offer. In this half - century, these factors have had lots of changes. The table below will briefly show this course of the two factors both individually and comparatively.

1.2.1.1.1.1 Table 1_Market Analysis

PERIOD	OFFER	DEMAND	MARKET
1998-2008	Deficit	Super fast	Prosperity
1988-1997	Expanding	Slow	Competitive
1973-1988	Exaggerating	Deficient	Recessional
1956-1973	Expanding	Super fast	Competitive
1945-1956	Deficit	Super fast	Prosperity
1930-1945	Exaggerating	Deficient	Recessional
1920-1930	Exaggerating	Fast	Weak

Source: Author based on Stopford, 2009

It is evident that the supply and capacity of the shipbuilding market have a role to play in setting the tone for a decade, but they're not the only ones. Supply-side management is an area in which shipping economics has a role to play. The challenge is to remind the shipping industry of the past and prepare it for the future. To make this possible, the clarity of our messages should be improved through better information, improved analysis, clearer presentation and greater relevance to the decisions made in the commercial shipping market, but above all, with an open mind. (Stopford, 2009)

1.3 Chartering

Another peculiarity relates to the volatile character of the bulk shipping markets. Bulk shipping due to its dependence on unpredictable factors does not provide the same predictability compared to other industries. The company cannot estimate the long-term market with the same precision. (Theotokas, 2019) Thus, accurately predicting its income is limited. Choosing the right ship (size, year of construction) is also important. In other words, the function of selecting and acquiring ships includes decisions regarding the composition of the fleet and the shipping markets that will move, the purchase of new or secondhand ships or the sale of existing ones as well as the choice of the time of implementation of the decision. Most of the shipping companies who manage bulk carriers are addressed to Shipping Brokers for both types; S&P (Sale and Purchase) and chartering with cargo.

A charterer has to be updated on the latest news on a worldwide scale. Moreover, the person involved in chartering is expected to be able to deal with the ever-changing conditions that affect the international freight market day by day. Charterers' target is to pay the lowest possible price or hire (cost of transportation). To acquire accurate and updated information concerning the market condition. Finally, brokers are professional experts in shipping who act as intermediates between ship owners seeking employment for their ship and starters looking for the services of a ship.

From a chartering perspective, a cargo is specified by the quality (type of cargo), the quantity (weight and volume of the cargo to be transported) and the packaging (bagged, bulk, palletized etc.)

Time chartering and spot chartering are two common methods used in the shipping and maritime industry for hiring vessels for transporting goods. They differ in terms of the duration of the charter and the associated financial arrangements.

Time chartering involves the hiring of a vessel for a specific period, typically ranging from a few months to several years. The charterer (the party hiring the vessel) pays a regular, fixed rate daily or monthly fee to the shipowner, irrespective of the cargo carried or the market conditions. The charterer is responsible for covering operating expenses, including fuel, crew, maintenance, and insurance during the charter period. Time chartering is often preferred for long-term commitments and when the charterer seeks to establish a consistent shipping schedule or needs vessel capacity over an extended period.

In contrast, *Spot Chartering* (also known as Voyage Chartering), involves the hiring of a vessel for a single voyage or a short-term period, usually a single cargo shipment. The charterer pays a one-time, lump-sum charter rate, which is negotiated based on market conditions, demand, and the specific voyage's characteristics. The shipowner retains more control over the vessel's scheduling and routing in spot chartering, as the charterer typically has limited influence over these aspects. The shipowner is responsible for the vessel's operating expenses, and these costs are factored into the negotiated charter rate. Spot chartering is often preferred for short-term needs, irregular cargo shipments, or when market conditions are uncertain, as it provides more flexibility and cost control for the charterer. (Pentheroudakis & Pagonis, 2019)

The comparison of the above markets shows us that, time chartering provides stability and predictability in terms of costs and scheduling, making it suitable for long-term commitments and establishing regular shipping services. In contrast, spot chartering is more flexible and suitable for short-term or one-off shipping needs. In time chartering, the charterer takes on the responsibility of vessel operating expenses, while in spot chartering, these expenses are typically included in the negotiated charter rate. The charter rate in time chartering is typically fixed and agreed upon in advance, while the charter rate in spot chartering is subject to market fluctuations and negotiations. Time chartering is associated with a longer-term commitment and often involves a closer relationship between the charterer and shipowner. Spot chartering is more transactional and based on short-term needs. Finally, the decision to choose between time chartering and spot chartering depends on factors such as the duration of shipping needs, market conditions, cost considerations, and the level of control desired by the charterer.

1.4 Freight Rate Indices

Freight rate indices play a crucial role in the maritime industry by providing a barometer for assessing market conditions and predicting trends in shipping costs. This section focuses on various freight rate indices, with special attention given to the Baltic Dry Index (BDI) and other relevant indicators.

1.4.1 Baltic Dry Index (BDI)

The Baltic Dry Index (BDI) is one of the most prominent and widely recognized indicators in the shipping industry. It tracks the average daily earnings of dry bulk cargo vessels across various routes, such as iron ore, coal, and grain. The BDI serves as an essential tool for gauging the demand and supply dynamics within the dry bulk shipping sector.

The BDI consists of four sub-indices, each reflecting different vessel types, BDI this component represents Capesize vessels, BPI the Panamax Index accounts for Panamax vessels, BSI reflects the performance of Supramax vessels, BHSI the Handysize Index.

The BDI serves as an invaluable tool for industry stakeholders, including shipowners, traders, and investors, in understanding market trends and making informed decisions. It is a forward-looking indicator, often used to assess future shipping rates, given its sensitivity to changes in supply and demand for dry bulk shipping services. Understanding these indices is essential for anyone involved in the maritime industry, as they offer valuable insights into market conditions and help in making informed decisions. (Stopford, 2009)

1.4.1.1.1 Image 1_The BDI trajectory over the period 1988 – 2010

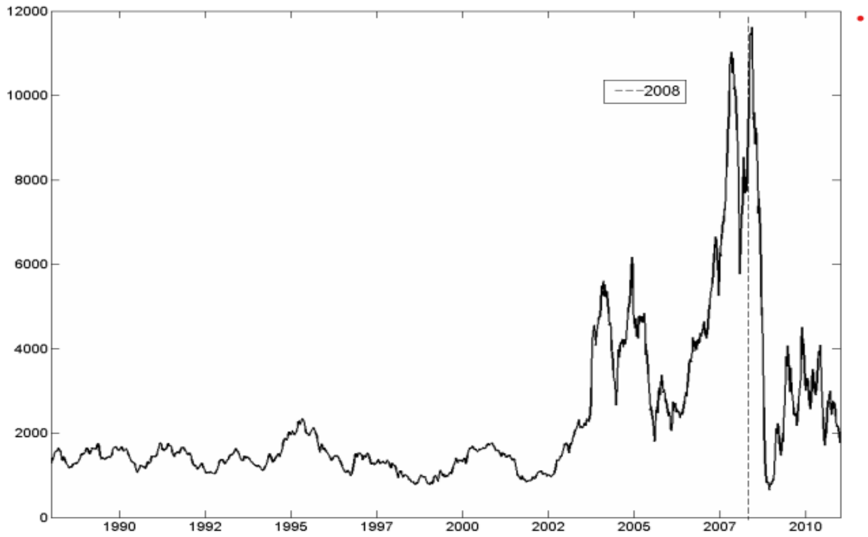


Exhibit 1: The BDI trajectory over the period 1988 - 2010

Source: Stopford, 2009

2. Freight Rates Forecasting (Literature Review)

The issue of predicting freight rates is always a major factor in the company's progress. Lots of researchers have dealt with the specific issue, others with different methods and others studying different periods. Moreover, shipping freight rates induce seasonal cyclic effects. Within the time of year, the rates fluctuate between the highs and lows of the seasonal demand for the cargoes they carry (Kavusssanos and Alizadeh, 2001). Their methodology is that the freight rate is calculated in US dollars per TEU that includes ocean freight rate (base price) and surcharges such as fuel or bunker price, exchange rate, container or equipment repositioning and other charges related to the operational cost (like terminal handling, space-booking and document charges). Static forecasts are the common forecasting method used to predict freight rates for practical shipping business operations. As explained by Stopford (2009), momentary equilibrium in shipping is about negotiating deals within hours, days or weeks. Hence, short-term predictions of freight rates are of great interest to charterers and shipbrokers. Although forecasting can be performed using simple methods such as the native or moving averages. (Vemuri and Munim, 2023). The evaluation of each forecasting model has been done by training and testing sample forecast accuracy. The need for doing such is to validate estimated forecast methods. In essence, a forecasting method that has a lower error is considered the most accurate. Three types of forecast accuracy measures are calculated including root mean squared error (RMSE), mean absolute percentage error (MAPE) and autocorrelations at lag 1 (ACF1). While RMSE and MAPE are widely used, the ACF1 is considered as a forecast accuracy measure because autocorrelations are a definitive measure showcasing the magnitude of the effect of the previous values over the current predicted ones.

The pricing structure within ocean freight rates is intricate, often contingent upon multifaceted elements such as cargo weight, cargo characteristics, transportation distance, and various other pertinent factors. Despite efforts to narrow the gap between supply and demand, there remains a lack of substantial recovery in freight rates. A comprehensive study conducted by Totakura et al. (2020) delved deeply into the examination of seasonal patterns' impact and prevalence in container freight rates. This meticulous research also encompassed rigorous seasonal unit root tests and meticulous scrutiny of deterministic seasonality across a spectrum of freight rate indices. The observed seasonality inherent in containerized cargo rates was attributed to intricate trade patterns, evolving demands for commodities, fluctuations in supply capacity, adaptability in service provisions, geopolitical influences, and a myriad of other factors contributing to this observed cyclicality.

The cyclicality within the dry bulk shipping market (DBSM), serving as a representation of spot and period charter rates in dry bulk shipping, was investigated thoroughly to formulate strategies about investment timing (asset play) and fleet trading (chartering strategy). Spectral analysis emerged as a crucial numerical approach employed to decipher significant cyclicality, thus aiding in the formulation of effective trading strategies. Instead of singularly focusing on a standalone dataset (univariate), an encompassing systemic approach was adopted to holistically comprehend the intricate shipping market cycle within its multifaceted context. Furthermore, a system dynamics design was implemented to unearth cyclical trends within the Dry market and its specific industrial context. This approach exhibited competitive forecasting accuracy in comparison to univariate time series models and artificial neural networks (ANNs), showcasing superior forecasting performance metrics such as mean absolute scale error (MASE), root mean square error (RMSE), and mean absolute percentage error (MAPE).

Seasonality identifies distinct monthly patterns, such as reduced rates during February, December, and September (Christmas and Chinese New Year), increased demand before Christmas in October and November, and disruptions in May due to monsoons. Specific months consistently exhibit seasonality across freight indices, except for the European Export Freight Index during the Christmas and Chinese New Year periods. Carriers adjust capacity during low-demand phases and maximize revenue through charter markets during peaks. supply chain requires a lead time before the period when specific products are needed in the market; hence, it's not at its peak demand in months like December. This also impacts freight rates.

Shipping entities grapple with critical decision-making processes involving long-term asset management guided by asset prices (e.g., ship prices) and countercyclical investment opportunities, alongside short-term liquidity management focused on working capital. The utilization of artificial intelligence techniques in forecasting, particularly studies exploring the potential of Forward Freight Agreements (FFAs) to enhance forecast accuracy, has been an area of significant exploration. Comparative analysis between two dynamic ANN models—NARNET and NARXNET—for diverse forecast horizons revealed the superior performance of NARXNET, underscoring the valuable insights derived from FFAs in improving forecast accuracy. Such insights aid in enhancing forecasts and rationalizing chartering decisions (Yang & Mehmed, 2020).

Shipowners tend to adjust fleet sizes based on cargo availability, with the return on such investments hinging on trade volume. Failure to increase fleet size during periods of growing trade volumes can strain sea transport due to a shortage of available shipping vessels. Conversely, an augmented fleet size amidst stagnant trade can result in inefficiencies and costly lay-ups. Short-term fluctuations in the freight market do not necessarily imply expected returns on ship ownership or operation. However, an uptick in demand for shipping services typically leads to increased freight market prices, prompting shipowners to invest in new vessels equipped with technology that adheres to international "green" shipping policy standards.

Studies examining the relationship between the Forward Freight Agreement (FFA) market and relevant spot freight markets have demonstrated the influence of the FFA market on spot freight markets. There has been substantial interest in investigation lead-lag relationships in returns and volatilities between spot and futures markets using VECM and GARCH models. Furthermore, an improved support vector machine (SVM) model has been effectively deployed in forecasting the Baltic Dry Index (BDI) with satisfactory results.

Goulielmos and Psifia (2011) utilized rescaled range analysis (R/S) in their paper as a non-parametric method to analyze time series in shipping economics. This method's efficacy, particularly in non-Gaussian distributions characterized by heavy kurtosis and skewness, has been highlighted. Their study revealed deviations in shipping freight rate distributions from normal distributions, indicating variance-associated risks. Additionally, a short-term shipping cycle lasting approximately four years was identified through the V-statistic derived from a generalization of Einstein's equation for a random walk.

Finally, Abouarghoub et al. (2016) analyze a Two-State Markov-Switching Distinctive Conditional Variance model to meticulously analyze freight returns. A correct risk measurement, crucial for freight transport, relies on an appropriate conditional volatility measure, especially in estimating Value-at-Risk (VaR). This study investigates the regime dependence of freight return's conditional variance using a two-state Markov regime-switching conditional variance model on the Baltic Dirty Tanker Index (BDTI). Analyzing daily BDTI returns from May 30, 1998, to October 30, 2009, it identifies distinct freight volatility regimes. The data reveals non-normality, indicating potential volatility changes attributed to the shipping industry's response to the financial crisis. VaR estimates from distinctive conditional variance models suggest a better capture of freight dynamics. Results show two distinct volatility regimes with normal and fractional integrated conditional variance frameworks best fitting each state.

The study by Chen, Liu and Wang (2020) employs decomposition and component reconstruction, identifying three distinct timescales: short, medium, and long-term fluctuations, each with unique characteristics. Short-term fluctuations are predicted using traditional ARMA time series analysis. Medium and long-term predictions utilize the grey wave forecasting method, known for modeling periodic features effectively. Unequal-interval contour lines by Chen et al. (2016) aid in contour moment sequence determination due to varied oscillation ranges. Contour time sequence selection is employed for out-of-sample forecasting based on proximity to the forecasted observation's serial number.

One more research compares forecasting outcomes between the proposed method, random walk, and ARMA(1,1) models. It also examines the performance of using ARMA to predict individual components (EMD–ARMA). Evaluating the forecasting performance over one-month and two-month periods, the study avoids small sample problems by conducting predictions ten times with rolling samples. Assessment metrics include RMSE, MAPE, Theil’s inequality coefficient, and correct direction prediction (CDP). Lower RMSE, MAPE, and Theil’s inequality coefficient denote better model performance, while CDP assesses correct directional predictions. Results demonstrate that the hybrid model

(EMD–ARMA/grey wave) consistently outperforms ARMA and random walk models across both periods for both four-weeks-ahead and eight-weeks-ahead predictions, exhibiting smaller error metrics and higher correct direction prediction percentages.

The hybrid model stands out with the lowest RMSE, MAPE, and Theil's inequality coefficient, showcasing enhanced predictive accuracy compared to traditional models. Additionally, it consistently outperforms others in correctly predicting directional movements in all cases.

For reference presents the forecasting performance comparison across various models and periods, highlighting the superiority of the hybrid EMD–ARMA/grey wave model in CCFI forecasting.

Forecasting shipbuilding orders effectively involves various internal and external factors. This study's literature review outlines the shipbuilding market's components, identifying a moderate cyclical trend in bulk shipbuilding order forecasting. The study focuses on the period 2008–2017, utilizing the ARIMA technique to assess an econophysics-based forecasting model for bulk shipbuilding orders. Among shipbuilding segments, forecasting bulk shipbuilding orders is valuable, reflecting global resource trade dynamics, offering insight into global production, and serving as a marker for the global economy.

Innovative forecasting methods enhance research and enable comparison with other techniques. This study introduces an econophysics approach for forecasting bulk shipbuilding orders, ideal for intricate shipbuilding predictions involving numerous variables. The econophysics model shows a relatively stable fit, adaptable to refine predictive variables, although enhancements in maritime data quality are necessary.

External uncertainties, like global maritime regulations and operational environments, significantly influence shipbuilding forecasting, urging investors to attentively monitor these factors. Interpretation and predictive capability of these external factors vary among investors, encouraging the use of this flexible econophysics-based model for bulk shipbuilding order prediction. (Hirata et al., 2020)

Zheng and Chen (2018) delve into the determinants affecting FFA trading volume across various markets. Initially, a theoretical model is constructed to outline these determinants and their influence on FFA trading volume within market equilibrium. Subsequently, a two-step model is employed to empirically test these theoretical assertions across the Capesize, Panamax, and Supramax sectors.

The findings are consistent with the theoretical model in the Capsize sector, offering potential insights for FFA traders to construct effective hedging strategies. The study suggests that the buyer's optimal FFA procurement quantity involves the expected spot demand multiplied by the quantity premium, derived from factors like demand covariance, riskiness, and volatility terms.

Furthermore, the study highlights the impacts of spot demand volatility and covariance between spot demand and spot rate on FFA trading activities, offering valuable insights to traders. Future research extensions might involve incorporating constraints like budget constraints in the theoretical model for forward freight procurement and examining their influence on FFA trading volume. Additionally, exploring cross-market linkages between forward freight markets and derivative markets of commodities transported by dry bulk vessels could unveil further determinants affecting FFA trading volume. This could include examining the trading volumes and volatilities of derivative prices in commodity markets to discern their impacts on FFA trading activities.

Other researchers, proposed to analyze a model about the relationship among several ocean bulk freight rate series. Through statistical examination, evidence of non-stationarity and cointegration relations between the series was found. Five cointegration relations among six series were identified using formal statistical tests, leading to the formulation of a VEC model. This model effectively explains the differences in the first derivatives of the freight rate series by simple variations between them. Using this model, two applications were formulated: estimating the underlying stochastic trend and forecasting freight rate series. The stochastic trend was remarkably similar across all six series, indicating a predominantly stochastic nature in the general pattern of freight rate movements, posing challenges for accurate forecasting. Despite establishing stable long-run relationships among ocean dry bulk freight rates, these relationships did not improve forecasting accuracy. The results suggest an economically meaningful structure exists within these freight rates, emphasizing stochastic trends as a substantial contributor to rate movements. While these findings do not contradict the efficient market hypothesis concerning ocean freight rates, limitations in obtaining reliable forecasts with this model are apparent. Further research on models with exogenous explanatory variables is warranted to enhance predictive capabilities. (Veenstra & Franses, 1996)

The last paper provided in this study evaluates the performance of ARIMA, ARIMAX, and VAR models by incorporating soft data in the form of sentiments, perceptions, or confidence regarding past, present, or future market activities as exogenous variables to predict CCFI and SCFI. CCFI and SCFI serve as established container freight rate indexes for the trade route from the Far East to Northern Europe. It's established in existing literature that ARIMA models perform well in predicting container shipping freight rates (Munim & Schramm, 2017 & 2020), despite being a somewhat myopic forecasting technique.

The study demonstrates that utilizing an ARIMAX approach to integrate soft data like LCI into ARIMA models significantly enhances forecast accuracy compared to a conventional ARIMA model. This approach might prove beneficial in predicting container freight rates across various trade routes. Despite VAR models being consistently outperformed by the base case ARIMA model with LCI, and CPLI being statistically insignificant in M5, M6, and M7, it appears that market activity influences future forecasts. Additionally, M1 incorporating the two LCI indexes from Transport Intelligence, assessing current volume development and a 6-month outlook at the trade lane level, exhibited superior forecast performance. The findings have substantial implications for shipping industry literature, providing evidence of the importance of soft data, such as sentiments, perceptions, or confidence. The inclusion of soft data improves the overall forecast performance of freight rates compared to univariate modeling. Consequently, these indices should continue to be standard practice. Since this study mainly estimated next-month freight rate forecasts within a recursive horizon, shipowners and cargo owners can implement the most effective models for decision-making within a monthly timeframe. CCFI and SCFI indexes are frequently used in forward freight agreements, and the party with superior forecasts is likely to face minimal risk.

Unfortunately, data is unavailable post-August 2017, possibly due to a discrepancy with the General Data Protection Regulation (EU) 2016/679. However, considering the ongoing trend of digitalization in the maritime shipping industry, future research employing real-time data from container booking platforms, terminal operation systems, as well as constant tracking of container vessels via their automatic identification system (AIS) transceivers or container shipment movements via their ISO 6346:1995 identification system, may further enhance the predictability of container shipping freight rates. Additionally, due to the increasing impact of container freight indices in hedging shipping risks, future studies should analyze not only the returns but also the volatility of container freight rates.

While reviewing the literature, it was noted that the methods followed were the usual ones. In fact, the markets they analyzed were either liner shipping, i.e. containerships, or wet bulkers (tanker market). very few forecasting articles talk about dry buying and even fewer use Markov chains. For this reason, research will be done on the dry market, using Markov chains.

3. Methodology – Markov Chains

3.1 Markov Chain

For the purposes of this study, a mathematical approach has been adopted in order to develop a model, we will apply the Markov Chain theory. This model is stochastic in nature, delineating a series of potential events where the likelihood of each event is contingent solely upon the state achieved in the preceding event. When the chain transitions states at discrete time intervals, it yields a discrete-time Markov chain (DTMC), constituting a countably infinite sequence. Conversely, a continuous-time process is termed a continuous-time Markov chain (CTMC). These chains are named in honor of the Russian mathematician Andrey Markov and find extensive utility as statistical frameworks for real-world phenomena. Applications range from examining cruise control systems in automobiles to analyzing queues of customers at airports, currency exchange rates and animal population dynamics.

Markov chains are used in finance and economics to model a variety of different phenomena, including the distribution of income, the size distribution of firms, asset prices and market crashes. The random walk was later seen as evidence in favor of the efficient market hypothesis based on this methodology. Dynamic macroeconomics makes heavy use of Markov chains. An example is using Markov chains to exogenously model prices of equity (stock) in a general equilibrium setting. Markov chains have been used for forecasting in several areas: for example, price trends, wind power and solar irradiance. The Markov chain forecasting models utilize a variety of settings, from discretizing the time series, to hidden Markov models combined with wavelets and the Markov chain mixture distribution model (MCM). (Chalidias, 2016)

When modeling real systems, it is very common that the rule describing the dynamics of the system depends only on the current state of the system and not on how the system got there. Stochastic systems that have this property are characterized as Markov Chains. Even most board games are Markov Systems.

Example 1: The various tests on real data have proven that the Markovian dependence works well in modeling some patterns weather. The dependence of today's weather on yesterday's may be minimal or non-existent in some climates. In others, the dependency is larger but suitable to fit with a Markov model. (Tsapiri Evanthia, 2013)

Example 2: Game theory uses Markov chains, games like snake, where the player is in a state and probabilistically goes to a second situation (2x2 matrices).

Definition: A stochastic process $\{X_n\}n\in\mathbb{N}$ is called a Markov chain, if, for each $n\in\mathbb{N}$, the bounded distribution of X_{n+1} given $(X_0, \ldots, X_n)$, is identical to the bounded distribution of X_{n+1} with only X_n given. Therefore, $\{X_n\}n\in\mathbb{N}$ with values in an enumerable space of states $\mathbb{X}$ is a Markov chain if, for every $n\in\mathbb{N}$ and every $v_0, \ldots, v_{n-1}$, $x, y \in\mathbb{X}$, so this is the mathematical equation that will be used :

$$\mathbb{P} = [\, X_{n+1} = y \mid X_0 = v_0, \ldots, X_{n-1} = v_{n-1}, X_n = x \,] = \mathbb{P}\,[\, X_{n+1} = y \mid X_n = x \,] \quad (1)$$

From Bayes theory of probabilities, $P(B|A) = \frac{P(AB)}{P(A)}$ where $P(A)\neq 0$. In most Markov Chains of interest, the rule $\mathbb{P}\,[\, Xn+1 = y \mid Xn = x \,]$ which describes the evolution of the chain, does not depend on the time parameter n. We then say that the chain is time homogeneous, define the transition probabilities of the chain. (Ross, 2010)

$$p : \mathbb{X} \times \mathbb{X} \rightarrow [0,1], \text{ with } p(x,y) = \mathbb{P}\,[\, X_{n+1} = y \mid X_n = x \,] \quad (2)$$

The collection $P = \{p(x, y)\}_{x,y\in \mathbb{X}}$ is called the transition matrix of the α chain. The terminology comes from the case that the state space X is finite,
so, we can describe a function with domain X × X as a *square matrix*.

Specifically, if $\mathbb{X} = \{v_1, \ldots, v_N\}$, we have

$$P = \begin{pmatrix} p(v_1, v_1) & p(v_1, v_2) & \cdots & p(v_1, v_N) \\ p(v_2, v_1) & p(v_2, v_2) & \cdots & p(v_2, v_N) \\ \vdots & \vdots & \ddots & \vdots \\ p(v_N, v_1) & p(v_N, v_2) & \cdots & p(v_N, v_N) \end{pmatrix}.$$

signify by π_0 the initial distribution of the chain,

$\pi_0 : \mathbb{X} \rightarrow [0,1]$ with $\pi_0 (x) = \mathbb{P} [X_0 = x]$. (3)

As a function mass probability (f.m.p.) of the random variable X_0, π_0 satisfies the following conditions:

$$\pi_0(x) \geq 0 \text{ for every } x \in \mathbb{X} \text{ and } \sum_{x \in \mathbb{X}} \pi 0(x) = 1 \quad (4)$$

Respectively, the probability p(x,~), which is Conditional Probability of f.m.p. of X_{n+1} given that X_n=x, so probability transition of the Markov chain $\{X_n\}$ n∈ ℕ satisfy the following conditions:

$$p(x,y) \geq 0 \text{ for every } x, y \in \mathbb{X} \text{ and } \sum_{y \in \mathbb{X}} p(x,y) = 1 \text{ for every } x \in \mathbb{X} \ (5)$$

The matrix P={p(x,y)} x, y ∈ 𝕏 it's a stochastic matrix if its elements satisfy the above conditions. So, this stochastic matrix P and the π_0 fully define the statistical properties of the Markov chain $\{X_n\}$ n∈ ℕ. For every n∈ ℕ and for every $(x_0, \ldots, x_n) \in \mathbb{X}^{n+1}$ then:

$$\mathbb{P}[X_0 = x_0, \ldots, X_n = x_n] = \pi_0 (x_0) p(x_0, x_1) \ldots p(x_{n-1}, x_n) \quad (6)$$

So, if the position x ∈ 𝕏 of a Markov chain at some time is known time n∈ ℕ the chain will evolve as a Markov chain starting at x, has the same probabilities and it is independent of what has gone before. This property continues to hold if we replace the fixed time n with any strategy interrupt T that cannot use information from the future. these are the strategic stop times and the property of a Markov chain to be renewed at these times, a strong Markov property. (Loulakis, 2019)

Markov chains can be used to predict the likely evolution of freights based on past observations. It is important, however, to consider the complexities of the shipping market and consider the suitability of Markov chains for the particular problem at hand., this method also has several positive points.

- *Simplicity*: The mathematical structure of Markov chains is usually simple, and the implementation of the model can be done with relatively few lines of code. This makes the method feasible to implement and understand.
- *Flexibility:* Markov chains can be adapted to model various situations and conditions in the shipping market.
- *Detailed Forecasting:* If the model is well parameterized and trained, it can provide detailed forecasts of the future states of the shipping market.
- *Supervision:* When properly implemented and trained, Markov chains are transparent and allow users to understand the predictions and decisions made by the model.
- *Dealing with Uncertainty:* The Markovian approach allows dealing with uncertainty in predictions as it can incorporate probabilities for various situations.

However, some possible problems that may arise are the following:

- *Markovian Property Assumption:* The Markovian Property Assumption means that the future state depends only on the present state and not on the history of how we got there. If this assumption does not hold, your model may not fully represent the structure of the data.
- *Model Complexity:* If the model is too complex, it may overfit the training data and not generalize well to new data.
- *Parameter Estimation:* Model parameter estimation may require careful selection and fitting. Miscalculations can lead to poor performance.

Finally, some other methods are regression, Markov switching multifractal (MSMF) techniques for modeling volatility evolution and a Hidden Markov model (HMM) is a statistical Markov model in which the system being modeled is assumed to be a Markov process with unobserved (hidden) states. An HMM can be considered as the simplest dynamic Bayesian network. HMM models are widely used in speech recognition, for translating a time series of spoken words into text.

3.2 Data

In Mathematics, and more specifically in Statistics, all the information is obtained from the sample. Sampling methods are very important for arriving at a correct and valid assessment. The data used in this research are Timeseries from Clarkson's in the period September 2009 until December 2023. It is of great importance to examine the progress of the freight rates, as well as the tension of the time series. Using descriptive statistics, such as mean, median, var the progress of the freights up to the last available date will be tracked. Descriptive statistics provide simple summaries of the sample and about the observations that have been made. These summaries can serve as the foundational depiction of data either as part of a broader statistical examination or independently for a specific investigation. Descriptive statistics play a pivotal role in the business realm, offering a concise overview of various data types. For instance, investors and brokers often leverage historical return data to conduct empirical and analytical assessments of their investments, aiding them in making informed decisions for future investments. In the context of handysize vessels (~28,000 dwt) timecharter for multiple routes, descriptive statistics can provide valuable insights for decision-making. We will apply the Markov chain method to forecasting freight rates. Time series analysis involves techniques for examining time series data to derive significant statistics and other data traits. Time series forecasting employs models to anticipate future values based on past observations. Stochastic models for time series typically acknowledge that observations closer in time tend to exhibit stronger relationships compared to those further apart. Moreover, time series models commonly utilize the inherent chronological order of time, whereby values for a specific period are understood to be influenced by past values rather than future ones. Because of Markov chain is a memoryless method, the use of this method does not require a period so long as this, so the study focuses on the third quarter of 2023 (09/23-12/23).

1.4.1.1.2 Image 2_Timeseries about freights per day

Shipping Intelligence Network Timeseries Created 12 December 2023 06:33						
	97210	97214	97218	97222	530380	530384
	Handysize 28,000 dwt Tripcharter Rate, Cont - ECSA	Handysize 28,000 dwt Tripcharter Rate, Cont - USEC	Handysize 28,000 dwt Tripcharter Rate, ECSA - Cont	Handysize 28,000 dwt Tripcharter Rate, US Gulf - Cont	Handysize 28,000 dwt Tripcharter Rate, SE Asia - Australia R/V	Handysize 28,000 dwt Tripcharter Rate, Transpacific R/V via NOPAC
Date	$/day	$/day	$/day	$/day	$/day	$/day
18/9/2009	13.500	13.500	12.500	22.000		
25/9/2009	15.000	13.750	14.000	21.500		
2/10/2009	16.000	13.750	14.750	21.750		
9/10/2009	15.750	13.250	15.250	21.250		
16/10/2009	15.250	12.500	14.750	20.000		
23/10/2009	14.000	11.500	13.750	18.000		
30/10/2009	13.500	11.000	14.000	17.000		
6/11/2009	13.500	10.750	15.250	17.500		
13/11/2009	14.000	11.750	17.250	19.250		
20/11/2009	14.750	12.500	19.750	23.000		
27/11/2009	15.000	12.750	20.750	23.750		
4/12/2009	15.250	13.500	21.500	24.250		
11/12/2009	15.500	14.250	23.750	26.750		
18/12/2009	14.000	13.750	23.500	26.500		
25/12/2009	13.500	13.000	23.000	26.000		
1/1/2010	13.500	13.000	23.000	26.000		
8/1/2010	13.500	13.000	23.500	26.250		
15/1/2010	13.750	13.250	24.250	26.500		

Source: Export by Clarckson’s

3.3 Program R

Notably, what statistical program is used in this research. Preferably by the researcher is the R language. The reasons the language has been chosen over other statistical packages for implementation. Statistical methods are plenty. In the first place, it is an Open Source software which is upgraded and enriched very often and offers the user many possibilities. It can be used either with direct existing commands, or with programs that the user can create to solve more complex statistical problems. Also, the user can use ready-made programs, which are integrated into libraries, which are also freely available. The variety of such programs is huge. Finally, one of the most attractive aspects of R is its ability to produce high-quality statistical charts. Its latest versions are the result of a collective effort with contributions from researchers from all over the world and for this reason make it a very popular language. It is a programming language mainly used for data and application of various classical and modern statistical techniques. In this research, R version 4.2.1 is used in Windows operating system. This practice is about data from an R library (MASS). The commands we will use for analyzing the historical data are as follows.

- ❖ Mean (mean(x))
- ❖ smallest and largest observation (min(x) & max(x))
- ❖ median (median(x))

- variance (var(x))
- standard deviation (sd(x))
- as well as the coefficient of variation
- histograms (hist(x))
- boxplots (boxplot(x))

4. Results

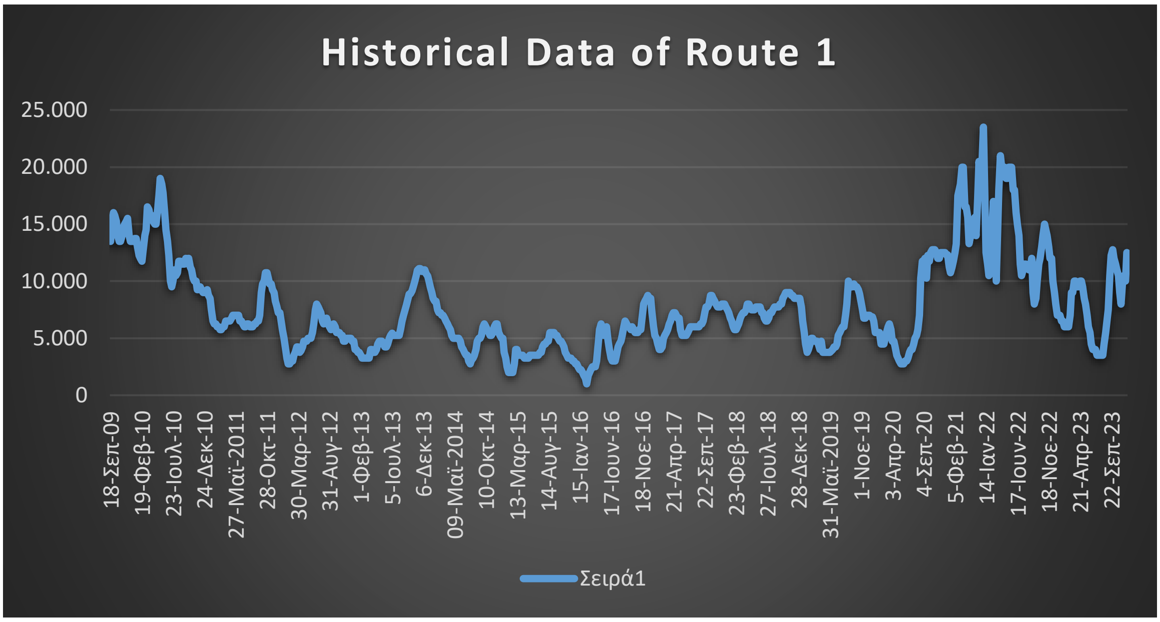
4.1 Historical Data and Descriptive Statistics

Below is the descriptive analysis of the variable which reflects the data for the specific routes of handysize vessels (28,000 dwt) the daily rates from September 2009 to December 2023. *Route number 1* is Cont-ECSA, henceforth we will call it route 1. In the transition period, the price range is in the interval [1000,23500] dollars per day, that is, its range has the price of 22500 dollars.

Their mean is 7726.69$ while the median is 6500$. The variance is 16270455 while the standard deviation is 4033.665, the mode is 6000. From these first indications, we get the following information. We first observe that the following inequality holds mode<median<mean, which gives the first information that the distribution of our variable is positive asymmetrical. Also, through the standard deviation information we realize that most information, quantitatively is in a period which creates the average value if we add and remove σ. In mathematical form $x \in [7726.69 \pm 4033.665] = [3693.025, 11760.355]$ which includes mode (the value that appears most times in the sample) but also the median which divides our sample into two.

Proceeding to the analysis with more descriptive measures is done control for quartiles and interquartile range. So, we have Q1=4750 which translates to ¼ of the data (i.e. 25%) is below $4750, ¾ below Q3= $10000 while the remaining 25% gets prices above $10000. The intra-quartile range has the value of 5250 which has been right because Q3-Q2=5250. (Appendix)

Figure 1: Tendence of timeseries for Route 1



Source: Author based on Clarckson’s

Furthermore, from the diagram resulting from the historical course of the freight rates, the tension of the time series can be observed. In general, from 18/9/2009 until 15/1/2016 there was a descent in prices, with the price ranging from $1000 to $17000. in the period of 15/6/2016 - 5/2/2021 there was an increase in freights. From that April until the summer of 2023 there was a downward trend. And then an upward trend again.

However, since the above diagram is condensed, a diagram that only concerns the year 2023 is also listed.

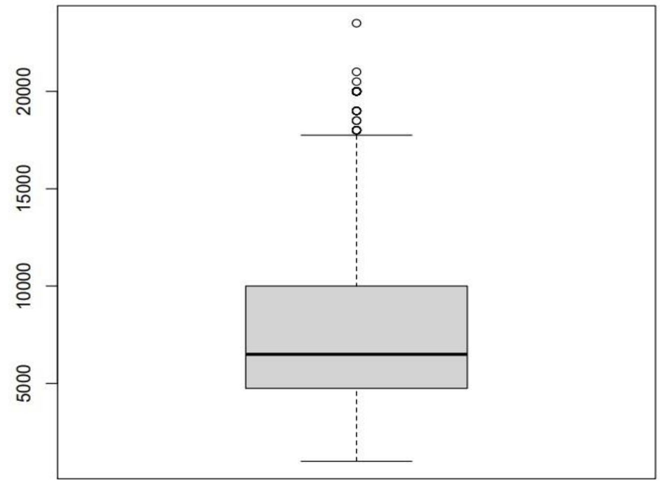
Figure 2: Tendence of freights per day for 2023

Source: Author based on Clarckson's

It will be split into two smaller periods, Jan-2023 until Aug-2023 when the lowest freight is below 4000$, which shows the trend of the new time series. in the first period we have a downward voltage while in the second an upward tendency.

For a better understanding of the above statistics, some more diagrams will follow which will be analyzed further. Firstly, the boxplot. From the Boxplots the extremes are 25% and 75% of the observations. The lower end corresponds to 25% of the observations and as we had observed in the descriptive measures, we analyzed previously for this variable have a match. Likewise, the upper end is 75% and the red line which shows the expected value (mean). The fences [Q3+1.5(Q3-Q1)] where they extend quite a bit, but we see they are there extremely high prices. It is mainly observed in the higher freights. The diagram fully agrees with the above descriptive findings.

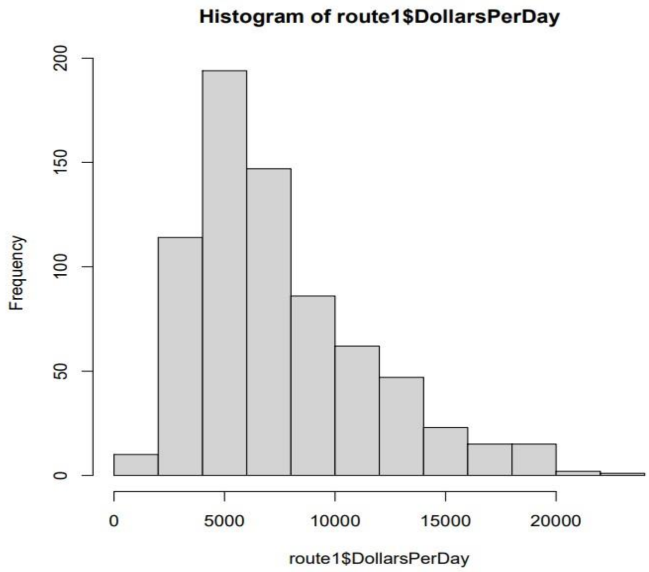
Figure 3: Boxplot of Route 1



Source: Author based on Clarckson's

Another diagram that will help in understanding time series is the histogram.

Figure 4: Histogram of Route 1



Source: Author based on Clarckson's

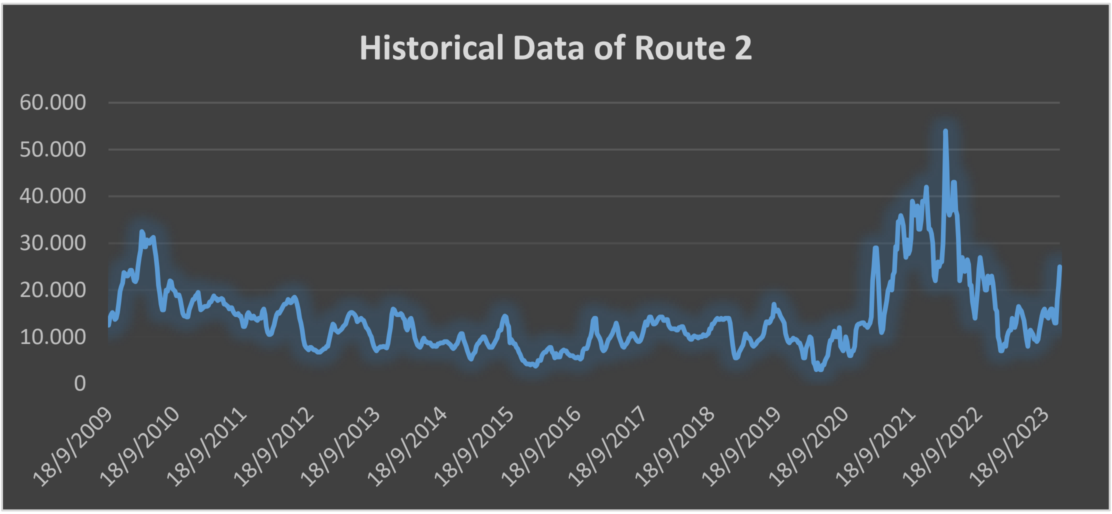
From this figure, the distribution of prices and the positive symmetry mentioned above are clearly visible. The kurtosis is 0.9291 and the skewness is 1,1174. (with point reference 0) that is, it is positively asymmetric and because 3>1.1174 it is slightly convex. Finally, the coefficient of variation is 52.20% (or 0.522). Because the cv is greater than 0.1 or 10%, it means that there is no reliability in predicting this variable.

Route No 2 is ECSA-Cont, henceforth we will call it route 2. Similarly, In the transition period the price range is in the interval [3000,54000] dollars per day, that is, its range has the price of 51000 dollars.

The mean is 13968$ while the median is 12250$. The variance is 59385163 while the standard deviation is 7706.177, the mode is 14000. From these first indications, we get the following information. We first observe that the following inequality holds mode>mean>median, which gives the first information that the distribution of our variable is negative asymmetrical. Also, through the standard deviation information we realize that most information quantitatively is in period which creates the average value if we add and remove σ. In mathematical form $x \in [13968 \pm 7706.177] = [6261.823, 21674.177]$ which includes mode (the value that appears most times in the sample) but also the median which divides our sample into two.

Proceeding to the analysis with more descriptive measures is done control for quartiles and interquartile range. So, we have Q1=8750 which translates to ¼ of the data (i.e. 25%) being below $8750, ¾ below Q3= $16000 while the remaining 25% gets prices above $16000. The intra-quartile range has the value of 5250 which has been right because Q3-Q2=7250.

Figure 5: Tendence of timeseries for Route 2

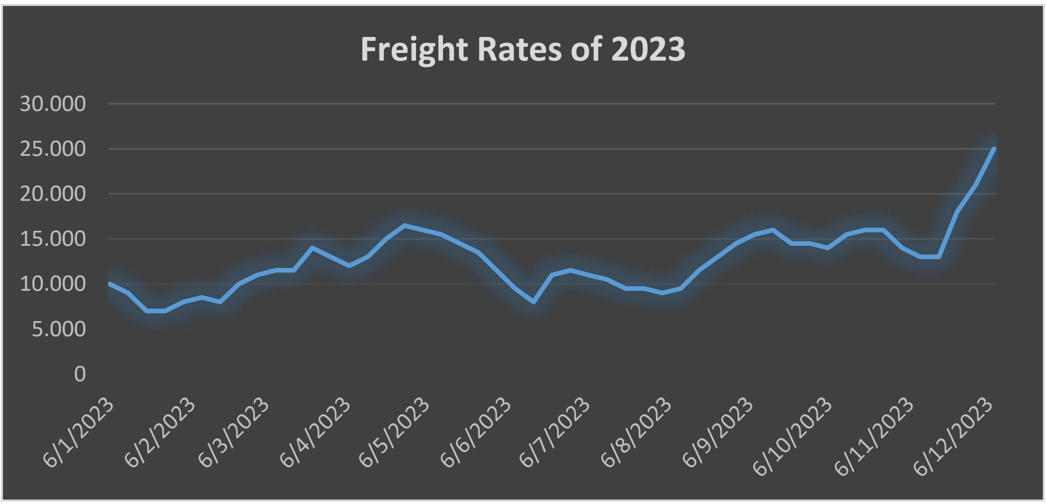


Source: Author based on Clarckson's

Furthermore, from the diagram (Figure 5) resulting from the historical course of the freight rates, the tension of the time series can be observed. In general, from 18/9/2009 until 18/9/2020 prices were stable. In the period of 18/9/2020 until early 2022 there was an increase in freights. From then until the summer of 2023 there was a downward trend. And then a slightly upward trend again.

However, since the above diagram is condensed, a diagram that only concerns the year 2023 is also listed.

Figure 6: Tendence of timeseries for Route 1

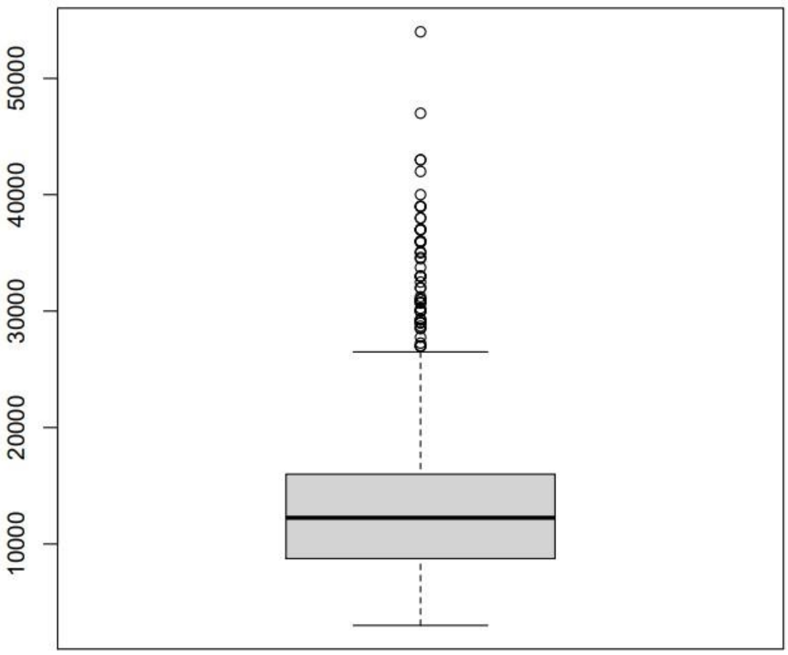


Source: Author based on Clarckson's

It will be split into two smaller periods, Jan-2023 until May-2023 when the lowest freight is above 5000$ and the highest is below 17000$. The second one is from May until the end of the year when there is a slight rise. That shows the trend of the new time series. In the first period we have a downward voltage while in the second an upward tendency.

In a more comprehensible way of the above statistics, some more diagrams will follow which will be analyzed further. Firstly, the boxplot. From the Boxplots the extremes are 25% and 75% of the observations. The lower end corresponds to 25% of the observations and as we had observed in the descriptive measures, we analyzed previously for this variable have a match. Likewise, the upper end is 75% and the red line which shows the expected value (mean). The fences [Q3+1.5(Q3-Q1)] where they extend quite a bit but we see they are there extremely high prices. It is mainly observed in the higher freights. The diagram fully agrees with the above descriptive findings.

Figure 7: Boxplot for Route 2

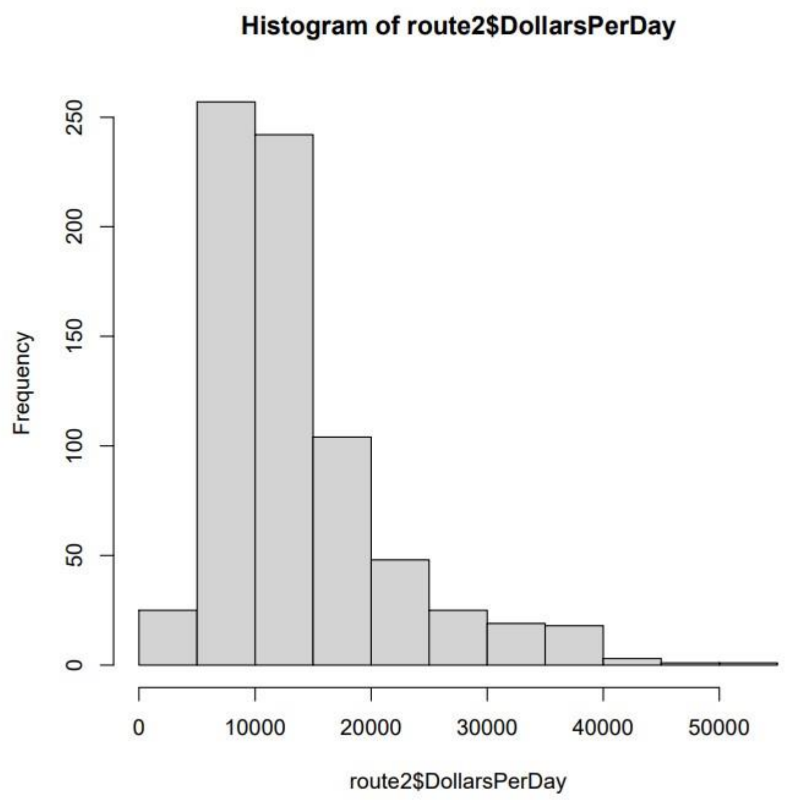


Source: Author based on Clarckson’s

For a better understanding of the above statistics about route 2, some more diagrams will follow which will be analyzed further. Firstly, the boxplot. The fences [Q3+1.5(Q3-Q1)] where they extend quite a bit but we see they are plenty of extremely high prices, many of them are between (30000,40000) US dollars. It is mainly observed in the higher freights. The diagram fully agrees with the above descriptive findings.

Another diagram that will help in understanding time series is the histogram.

Figure 8: Histogram for Route 2



Source: Author based on Clarckson’s

Likewise, the distribution, about route 2, of prices seems to have a positive symmetry. The kurtosis is 2.968 and skewness is 1,62. (with point reference 0) that is, it is positively asymmetric and because 3>1.62 it is slightly convex. Finally, coefficient of variation is 55.10% (or 0.551), because the c v is greater than 0.1 or 10%, it means that there is no reliability in predicting this variable.

These two routes are about to come and go. The first one is Cont – ESCA and the second one ECSA – Cont. Despite the fact that is almost the same trip, the freights have a remarkable difference; a great impression is given by the fact that the freights are almost twice as much for the return of the ship. Scientifically speaking the distractive statistics show that the distributions are very similar. The distributions move at the same levels of both risk and price tendency. One more of the similarities, is the uncertainty of the market for forecasting the freight rates. From this point of view, the coefficient of variability is in both routes (variable) over 50%, where the risk of this research is exposed.

4.2 Forecasting with Markov Chains Method

Freight forecasting in the shipping sector is an important and complex process, as it is influenced by many factors. Some of these factors include (except from offer and demand):

- *Economic Situation:* The economic situation at the global level affects commercial activity and thus the demand for shipping services.
- *Political Events:* Political and geopolitical events such as wars, government changes, sanctions and others can create uncertainty and affect the shipping market. (i.e. Huthi attacks on Yemen and fear of imminent closure of the Suez Canal)
- *Fuel:* Fuel prices are an important factor as they are a significant cost for charterers. Fuel prices can influence decisions about how ships are operated and the number of ships available in the market. (This factor does not concern time-charterers, since in the time charter the fuel is charged by the charterer)
- *Customer Appreciation:* Customer requirements and expectations for the future can influence the demand for shipping services. (preferential deal agreements with the aim of creating customer relationships)

To predict freights, the states to be modeled should be defined and the possible transitions between them defined. The observed data should then be used to train the model to reflect real-world conditions.

4.2.1 Route 1 (97210)

Relatively recent dates were used in this model. As mentioned, this method is memoryless. The table below shows the probabilities for the Transition matrix which is essential for Markov chains.

Table 2: Data for Transition Matrix

	97210	
	Handysize 28,000 dwt Tripcharter Rate, Cont - ECSA	Transition Matrix
Date	$/day	Probabilities
29/9/2023	12.750	0,09
6/10/2023	12.000	0,09
13/10/2023	11.500	0,09
20/10/2023	11.000	0,09
27/10/2023	10.500	0,18
3/11/2023	9.000	0,09
10/11/2023	8.000	0,09
17/11/2023	10.000	0,18
24/11/2023	10.500	
1/12/2023	10.000	
8/12/2023	12.500	0,09
Total		1,00

Source: Author based on Clarckson’s

Hence, end up with the 3x3 Transition Matrix

$$\begin{matrix} \mathbf{0.09} & \mathbf{0.09} & \mathbf{0.09} \\ \mathbf{0.09} & \mathbf{0.18} & \mathbf{0.09} \\ \mathbf{0.09} & \mathbf{0.18} & \mathbf{0.09} \end{matrix}$$

Given the facts, there is a volatility of events in global shipping. so, the following assumptions were made, there is a 20% chance that the freights will decrease, 30% that they will remain stagnant and 50% that they will increase. Those probabilities create another matrix which will be used to define the status of the Markov Chain model.

$$\left(\mathbf{0.2} \quad \mathbf{0.3} \quad \mathbf{0.5}\right)$$

Analysis was carried out for 10 steps (standard procedure) for a small future interval. Finally, the results were the following:

Table 3: Results from running Markov Chains

STEPS	SITUATION
Step 1	Situation 3
Step 2	Situation 2
Step 3	Situation 2
Step 4	Situation 2
Step 5	Situation 2
Step 6	Situation 2
Step 7	Situation 2
Step 8	Situation 2
Step 9	Situation 2
Step 10	Situation 2

Source: Author based on Clarckson's

Following the test results for the first route it seems that there is an increase in freight rates and then the freight market will stabilize. In other words, there will not be big fluctuations in terms of the daily monitoring of the freights. This does not mean that the freights will be a fixed price. Nevertheless, even though the variable dollar per day is not the same on a daily basis, the time series tension will be stable. As shown above, (Figure 1) from Dec-10 till Apr-20 there was not any upward or downward tension. This is a main pattern of bulk shipping. And this is how it will move in the near future. Due to the coefficient of variance which is over 50% we come to the conclusion that forecasting is difficult.

4.2.2 Route 2 (97218)

Similar to the first route, relatively recent dates were used in this model. As mentioned, multiple times, this method is memoryless. The table below shows the probabilities for the Transition matrix which is essential for Markov chains.

Table 4: Results from running Markov Chains

	97218	
	Handysize 28,000 dwt Tripcharter Rate, ECSA - Cont	Transition Matrix
Date	$/day	Probabilities
18/8/2023	11.500	0,06
25/8/2023	13.000	0,18
1/9/2023	14.500	0,18
8/9/2023	15.500	0,12
15/9/2023	16.000	0,18
22/9/2023	14.500	
29/9/2023	14.500	
6/10/2023	14.000	0,12
13/10/2023	15.500	
20/10/2023	16.000	
27/10/2023	16.000	
3/11/2023	14.000	
10/11/2023	13.000	
17/11/2023	13.000	
24/11/2023	18.000	0,06
1/12/2023	21.000	0,06
8/12/2023	25.000	0,06
Total		1,00

Source: Author based on Clarckson’s

Hence, end up with the 3x3 Transition Matrix. The fact that the dates are not the same is not a concern. Since, what is of interest for this research is the tendence of the freight rates and not the price of each one.

$$\begin{matrix} \mathbf{0.06} & \mathbf{0.18} & \mathbf{0.18} \\ \mathbf{0.12} & \mathbf{0.18} & \mathbf{0.12} \\ \mathbf{0.06} & \mathbf{0.06} & \mathbf{0.06} \end{matrix}$$

Given the facts, there is a volatility of events in global shipping. so, the following assumptions were made, there is a 20% chance that the freights rates will decrease, 20% that they will remain stagnant and 60% that they will increase. Those probabilities create another matrix which will be used to define the status of the Markov Chain model.

$$\left(\mathbf{0.2} \quad \mathbf{0.2} \quad \mathbf{0.6}\right)$$

Analysis was carried out for 10 steps (standard procedure) for a small future interval. Finally, the results were the following:

Table 4: Results from running Markov Chains

STEPS	SITUATION
Step 1	Situation 3
Step 2	Situation 2
Step 3	Situation 2
Step 4	Situation 2
Step 5	Situation 2
Step 6	Situation 2
Step 7	Situation 2
Step 8	Situation 2
Step 9	Situation 2
Step 10	Situation 2

Source: Author based on Clarckson’s

Following the test results for the first route it seems that there is an increase in freight rates and then the freight market will stabilize. In other words, there will not be big fluctuations in terms of the daily monitoring of the freights. This does not mean that the freights will be a fixed price. Nevertheless, even though the variable dollar per day is not the same on a daily basis, the time series tension will be stable. As shown above, (Figure 1) from Sep-10 till Sep-20, a whole decade, there was any upward or downward tension. This is a main pattern of bulk shipping. And this is how it will move soon. Due to the coefficient of variance which is over 50%, the forecasting came to the conclusion that is difficult.

4.2.3 Comparing the two routes

This specific model is the same for both routes. These two routes are opposite, that is, we have two directions and takes place in the Atlantic Ocean. The conditions were a lot. The most important one is the prices of fuel, which they will be considered fixed costs due to the study performed. The second condition is political events, on the western side of the world, even though there are no wars, nor are there any cases of piracy, the freight rates are affected by the global political situation. And after the Ukrainian issue and now with the issue with Israel and Hamas in Palestine there are impacts affecting international trade and transportation. This condition is also consistent with the financial situation. Inflation has sent the prices of products skyrocketing and the rise in freight rates is a natural progression. For this very reason, the probabilities of freight rates increasing in Markov models are high, greater than or equal to 50%.

In comparison, both routes exhibit similar transition matrices, indicating a memoryless Markov model. Assumptions about the volatility of transport prices differ slightly between the two routes. As it is also very important for the outcome. The conclusions for both paths are similar, with an initial increase in prices followed by stabilization. The coefficient of variation exceeds 50% for both routes, indicating scheduling difficulties. It appears that the two routes exhibit similar patterns in terms of freight rates changes and market stability. The Markov models and assumptions for both pathways are aligned, leading to similar prediction challenges due to the observed volatility.

In essence, the analysis highlights the intricate interplay of various factors, including geopolitical events, economic conditions, and fuel prices, that collectively shape the dynamics of freight rates in these maritime routes. The consistent trends observed in the Markov models emphasize the need for nuanced forecasting approaches in the face of the complex and ever-changing global shipping environment.

5. Conclusions

To conclude, this thesis was about freight rates forecasting with Markov Chain Method. The subject freight rate forecasting the subject has occupied a large research community. Shipping companies would pay huge amounts, if they could predict their revenue with greater certainty. At the first chapter, the concept of bulk shipping was introduced. There was a report on the sizes of the Dry Bulk Ships, what is their percentage in world shipping as well as the Baltic Dry Index. Furthermore, it was mentioned what a freight broker is, what types exist as well as how the process of chartering a ship is done.

As we have shown in chapter three, which mentions various studies by fellow researchers, the subject has been of considerable concern to the maritime community. So, we collocate 12 ways to predict. Some of them were, ARMA which analyze time series and has use only for short-term variance of freight rates, Static Forecasts that is the most common method and uses a weighted average, Spectral Analysis which is useful to predict seasonality and cyclicality. The only one literature was found with method of Markov was one; the "Two -State Markov- Switching Distinctive Conditional Variance Model", which mostly about risk value.

The approach here was a prediction of freight rates tendence with Markov Chains. Classical statistical methods were used to make a clear analysis of the data, in order to know the tension of the time series. The test was carried out for two routes in the Atlantic Ocean using the R program. The routes connect the old continent with America. With the help of three-dimensional matrices, which are Transition Matrices, and the definition of the probabilities for the possible status of the freights (downward, stable, upward), the course of the time series was checked. It was observed that the two routes had similar behavior in terms of the statistical analysis that was done, not identical numbers but they have the same motive. An important parameter was the definition of the probabilities for the possible situations. They mainly determined the result. The decision was made that the probability of increasing freight rates are 50% on *Route 1* and 60% on *Route 2*, for this reason the outcomes of the test were identical as far as the prediction with Markov chains is concerned.

Due to the coefficient of variable both of the routes, from which the conclusion follows that there is uncertainty in the forecast of the freight rates, the specific research has some limitations. First of all, there can be no certainty about the demand curve and how it will evolve. So, the first step is to predict the demand. About the offer of sizes and numbers of ships there is more clear data from registries about ordering, scrapping and the ages of the ships of the global fleet. Another parameter that is uncertain is geopolitics. Usually in the east there are several states and organizations that influence world events. In such cases, the freight rates usually go up regardless of the voyage the ship follows, just in the high-risk areas there is a rapid rise. These relate to the more general subject of freight rates forecasting.

However, as far as the Markov chain method is concerned, there are few limitations. What must be accurately calculated are the probabilities of situations. In research, assumptions have been made about this matrix. But the situations will have to come to reality for it, to become applicable in Maritime Industry.

Finaly, some more ideas for the development of this thesis is to try to reduce the coefficient of variance below 10% so that there is certainty in the forecasting. At PhD level we could also develop the Markov model into Hidden Markov Model (HMM) which can be used to describe the evolution of observable events that depend on internal factors, which are not directly observable. We call the observed event a "symbol" and the invisible agent behind the observation a "state".

Appendix

```
2: >
   ^
> #descriptive statistics analysis for route 97210 Cont - ECSA
> route1<-read.csv("C:/Users/Dionysia/Desktop/Διονυσία/Test1.csv",sep=";")
> route1<-route1[-717,] #removal NAs
> route1<-route1[,-c(3,4,5,6,7,8)]
> route1
```

```
> colnames(route1)<-c("Date","DollarsPerDay")
> route1
            Date DollarsPerDay
1    18-Σεπ-2009        13.500
2    25-Σεπ-2009        15.000
3    02-Οκτ-2009        16.000
4    09-Οκτ-2009        15.750
5    16-Οκτ-2009        15.250
6    23-Οκτ-2009        14.000
7    30-Οκτ-2009        13.500
8    06-Νοε-2009        13.500
9    13-Νοε-2009        14.000
10   20-Νοε-2009        14.750
11   27-Νοε-2009        15.000
12   04-Δεκ-2009        15.250
13   11-Δεκ-2009        15.500
14   18-Δεκ-2009        14.000
15   25-Δεκ-2009        13.500
```

R Console

```
> route1$DollarsPerDay<-(route1$DollarsPerDay)*1000
> #multiply by 1000 because the program reads the point as comma
> route1
            Date DollarsPerDay
1    18-Σεπ-2009         13500
2    25-Σεπ-2009         15000
3    02-Οκτ-2009         16000
4    09-Οκτ-2009         15750
5    16-Οκτ-2009         15250
6    23-Οκτ-2009         14000
7    30-Οκτ-2009         13500
8    06-Νοε-2009         13500
9    13-Νοε-2009         14000
10   20-Νοε-2009         14750
```

```
> mean(route1$DollarsPerDay)
[1] 7726.69
> summary(route1$DollarsPerDay)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
   1000    4750    6500    7727   10000   23500
```

R Console

```
> colnames(route2)<-c("Date","DollarsPerDay")
> route2$DollarsPerDay<-(route2$DollarsPerDay)*1000
> #multiply by 1000 because the program reads the point as comma
> route2
         Date DollarsPerDay
1   18/9/2009         12500
2   25/9/2009         14000
3   2/10/2009         14750
4   9/10/2009         15250
5  16/10/2009         14750
6  23/10/2009         13750
7  30/10/2009         14000
8   6/11/2009         15250
9  13/11/2009         17250
10 20/11/2009         19750
```

```
> summary(route2$DollarsPerDay)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
   3000    8750   12250   13968   16000   54000
> var(route2$DollarsPerDay)
[1] 59385163
> sd(route2$DollarsPerDay)
[1] 7706.177
> IQR(route1$DollarsPerDay)
[1] 5250
> IQR(route2$DollarsPerDay)
[1] 7250
> cv<-sd(route2$DollarsPerDay)/mean(route2$DollarsPerDay)
> cv
[1] 0.5517157
> plot(route2$DollarsPerDay)
> hist(route1$DollarsPerDay)
> hist(route2$DollarsPerDay)
> boxplot(route2$DollarsPerDay)
```

```
MARKOV CHAINS (ROUTE1)

#Define the transition probability matrix
>transition_matrix1<-matrix(c(0.09,0.09,0.09,0.09,0.18,0.09,0.09,0.18,0.09),nrow=3,byrow=TRUE)
> transition_matrix1
     [,1] [,2] [,3]
[1,] 0.09 0.09 0.09
[2,] 0.09 0.18 0.09
[3,] 0.09 0.18 0.09
> #Defiine the initial state
> initial_state1<-c(0.2,0.3,0.5)
> initial_state1
[1] 0.2 0.3 0.5
>#Define steps
>num_steps<-10
>#Functions for Markov Chain simulation
> current_state<-initial_state1
> for(step in 1:num_steps){
+ cat("Step",step,":Situation",which.max(current_state),"\n")
+ current_state<-current_state%*%transition_matrix1
+ }
Step 1 :Situation 3
Step 2 :Situation 2
Step 3 :Situation 2
Step 4 :Situation 2
Step 5 :Situation 2
Step 6 :Situation 2
Step 7 :Situation 2
Step 8 :Situation 2
Step 9 :Situation 2
Step 10 :Situation 2
```

```
MARKOV CHAINS (ROUTE2)

>#Define the transition probability matrix
> transition_matrix2<-matrix(c(0.06,0.18,0.18,0.12,0.18,0.12,0.06,0.06,0.06),nrow=3,byrow=TRUE)
> transition_matrix2
     [,1] [,2] [,3]
[1,] 0.06 0.18 0.18
[2,] 0.12 0.18 0.12
[3,] 0.06 0.06 0.06
> #Define steps
+ num_steps<-10
> #Functions for Markov Chain simulation
> #Define initial state
> initial_state2<-c(0.2,0.2,0.6)
> initial_state2
[1] 0.2 0.2 0.6
> current_state<-initial_state2
> for(step in 1:num_steps){
+ cat("Step",step,":Situation",which.max(current_state),"\n")
+ current_state<-current_state%*%transition_matrix2
+ }
Step 1 :Situation 3
Step 2 :Situation 2
Step 3 :Situation 2
Step 4 :Situation 2
Step 5 :Situation 2
Step 6 :Situation 2
Step 7 :Situation 2
Step 8 :Situation 2
Step 9 :Situation 2
Step 10 :Situation 2
```

References

1. Jeroen Pruyn, 2017, “Are the new fuel-efficient bulkers a threat to the old fleet?”, Department of Maritime and Transport Technology, Delft University of Technology, Delft, The Netherlands.

2. U.S. Department of Transportation Maritime Administration, “World Merchant Fleet 2005”

3. Martin Stopford, 2009, 3rd Edition, “Maritime Economics”

4. UNCTAD, 2006

5. Th. Pagonis & N. Pentheroudakis,2019, “Chartering Manual By Practitioners”

6. I. Theotokas, 2019, “Organization and Management of Shipping Companies”, University of Piraeus

7. A. W. Veenstra & P. H. Franses, 1997, “A Co-intergration Approach to Forecasting FreightRates in the Dry bulk Shipping Sector”, Department of Regional Transport and Port Economics, Erasmus University, Rotterdam, The Netherlands & Department of Econometrics and Rotterdam Institute for Business Economics Studies, Erasmus University, Rotterdam, The Netherlands

8. Chen, Liu, Wang, 2020, “Analysing and forecasting China containerized freight index with a hybrid decomposition–ensemble method based on EMD, grey wave and ARMA”, School of Economics and Management, Shanghai Maritime University, Shanghai, China

9. Abouarghoub, Mariscal & Howells, 2014,” A Two-State Markov-Switching Distinctive Conditional Variance Application for Tanker Freight Returns”, Bristol Business School

10. Schramm & Munim, 2022, “Container freight rate forecasting with improved accuracy by integrating soft facts from practitioners “, Institute for Transport and Logistics Management, Vienna University of Economics and Business, Vienna, Austria Department of Operations Management, Copenhagen Business School, Copenhagen, Denmark Faculty of Technology, Natural and Maritime Sciences, University of South-Eastern Norway, Campus Vestfold, Horten, Norway

11. Duru et al, 2020, "Aneconophysicsapproachto forecast bulk shipbuilding orderbook: an application of Newton's law of gravitation", School of Civil and Environmental Engineering, Nanyang Technological University, Singapore, Singapore, School of Business and Economics, Universiti Brunei Darussalam, Bandar Seri Begawan, Brunei & Center for Mathematical and Data Sciences, Kobe University, Kobe, Japan

12. Goulielmos & Psifia, 2011, "Forecasting short-term freight rate cycles: do we have a more appropriate method than a normal distribution?"

13. Shiyuan Zheng and Shun Chen, 2018," Modeling the determinants of dry bulk FFAtradingvolumefroma cross-market perspective of spot andforward", School of Transportation, Shanghai Maritime University, Shanghai, China

14. Karaoulanis & Pelagidis, 2021, "Panamax markets behaviour: explaining volatility and expectations", University of Piraeus

15. Samhita Vemuri and Ziaul Haque Munim, 2023, "Seasonality and forecasting analysis of the South-East Asian container freight market", Faculty of Technology, Natural and Maritime Sciences, University of South-Eastern Norway, Horten, Norway

16. Totakura Bangar Raju*, Pradeep Chauhan, Saurabh Tiwari, Vishal kashav, 2020, "Seasonality in Freight Rates", Transportation Management, University of Petroleum and Energy Studies, Dehradun, India

17. Hélyette Geman & William O. Smith, 2012, "Shipping markets and freight rates: an analysis of the Baltic Dry Index"

18. D.Fouskakis, 2013,"Data Analysis Using R program", National Technical University of Athens

19. Sheldon M Ross, 2010, "A First Course in Probability"

20. Tsapiri Evanthia, 2013 "Stochastic simulation with Markov Chain Monte Carlo methods (MCMC)", National Technical University of Athens

21. UNCTAD, 2020