ECONOMETRIC MODELING OF OCEAN FREIGHT RATES: THE CASE OF TANKERS

by

DIMITRIS AYVATOGLU

B.Sc. MECHANICAL ENGINEERING – BROWN UNIVERSITY, 1999
B.A. BUSINESS ECONOMICS – BROWN UNIVERSITY, 1999

SUBMITTED TO THE DEPARTMENT OF OCEAN ENGINEERING IN PARTIAL
FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF

MASTER OF SCIENCE IN OCEAN SYSTEMS MANAGEMENT
AT THE
MASSACHUSETTS INSTITUTE OF TECHNOLOGY

JUNE 2001

© 2001 Massachusetts Institute of Technology. All rights reserved.

The author hereby grants to MIT permission to reproduce and to distribute publicly paper and
electronic copies of this thesis document in whole or in part.

Signature of Author:.............................................

Department of Ocean Engineering
May 11, 2001

Certified by:.............................................

Professor Henry S. Marcus
Professor of Marine Systems
Thesis Supervisor

Accepted by:

Professor Henrik Schmidt
Professor of Ocean Engineering
Chairman, Department Committee on Graduate Students
Econometric Modeling of Ocean Freight Rates: The Case of Tankers

by

Dimitris Ayvatoglu

Submitted to the Department of Ocean Engineering on May 11, 2001
in Partial Fulfillment of the Requirements for the Degree of
Master of Science in Ocean Systems Management

Abstract

The aim of this thesis is to investigate the efficiency of the tanker freight market, and at the same time illustrate shortcomings in the prevailing econometric methodology. The analysis is based on three different vessels: a product tanker, a suezmax tanker, and a VLCC tanker. The first part of the paper gives an overview of the maritime industry, focusing on the economic forces behind the cyclicality of the freight market. The second part of the paper applies the Box-Jenkins methodology to model the earnings time series of the above-mentioned vessels using an ARIMA-type model. The results show evidence against an efficient market for the product tanker, whereas they are inconclusive as far as the suezmax and VLCC tankers are concerned. The third part of the paper identifies possible shortcomings of the ARIMA class of models, as they are applied to the maritime industry. To correct for those, an R/S analysis is performed on the same earnings time series. This second approach verifies that there are inefficiencies in the product tanker freight market. Moreover, it provides evidence against an efficient VLCC freight market. On the other hand, the suezmax earnings time series is shown to be perfectly random, hence rendering this particular freight market efficient.

Thesis Supervisor: Henry S. Marcus
Title: Professor of Marine Systems
Biography of Author

Dimitris Ayvatoglu was born in Athens, Greece in 1976. He attended Athens College from which he received his High School Diploma in June of 1995. In September of 1995, Mr. Ayvatoglu started his undergraduate studies at Brown University of Rhode Island. He graduated in 1999 with a dual degree in Mechanical Engineering and Business Economics, with a Magna Cum Laude distinction. Mr. Ayvatoglu continued his academic career at the Massachusetts Institute of Technology. He joined the Ocean Engineering Department and concentrated in Ocean Systems Management. Mr. Ayvatoglu is expected to graduate from MIT in June of 2001 with a Master of Science degree in Ocean Systems Management.

Mr. Ayvatoglu has been following the maritime industry closely for the past several years. During the summer recessions he worked part-time in various shipping-related firms in Piraeus, Greece.

Mr. Ayvatoglu is fluent in Greek and English.

Acknowledgements

I would like to start by thanking my advisor and thesis supervisor, Dr. Henry Marcus, for his continuous support during the past two years. His enthusiastic and easygoing character was an essential ingredient of this graduate thesis. Furthermore, this thesis could not have been written without Marsoft, Inc.’s kind gesture of providing the author with the necessary data set. My sincere thanks also go to the whole Ocean Engineering faculty of MIT for their never-ending struggle for the bestowment of a better education. I would also like to thank my friends for their support, whether inside or outside the classroom. Finally, I would like to thank those who, in one way or another, have assisted me in successfully completing this project, by giving me advice and putting up with my temper through the long process of preparing this work. In particular, I would like to express my gratitude to Anastasia, Anthony, and Athina for their assistance in editing the text.

Dimitris Ayvatoglu
May 11, 2001
This work is dedicated to Vassili, Irini and Gianni, who have guided me, encouraged me, inspired me and sustained me all these years. Most of all, I am grateful to them because they have believed in me, through good times and bad.
Table of Contents

Abstract ............................................................................................................................ 3
Biography of Author ........................................................................................................ 5
Acknowledgements ......................................................................................................... 5
Table of Contents ........................................................................................................... 9
List of Figures .................................................................................................................. 12
List of Tables ................................................................................................................... 13

Chapter 1  Introduction ................................................................................................. 15
  1.1 Scope of Thesis ....................................................................................................... 16
  1.2 Overview of Subsequent Chapters ........................................................................ 17

Chapter 2  The Maritime Industry .................................................................................. 19
  2.1 The Cost of Transport .......................................................................................... 19
  2.2 The Tanker Sector ............................................................................................... 22
  2.3 Supply vs. Demand .............................................................................................. 23
  2.4 Market Cycles ....................................................................................................... 26
    2.4.1 Short Cycles .................................................................................................... 27
    2.4.2 Long Cycles .................................................................................................... 28
  2.5 Historical Overview .............................................................................................. 29

Chapter 3  Literature Survey and Motivation ................................................................. 31
  3.1 Supply – Demand Models .................................................................................... 31
  3.2 The Mean Reverting Nature of Freight Rates ....................................................... 33
    3.2.1 Introduction to Stochastic Variables .............................................................. 34
    3.2.2 Mean Reversion ............................................................................................ 34
    3.3.3 Concerns ........................................................................................................ 36

Chapter 4  Time Series Models ..................................................................................... 39
  4.1 Common Time Series Features .............................................................................. 40
    4.1.1 Trends ............................................................................................................ 40
4.1.2 Seasonality, Conditional Heteroskedasticity, and Non-Linearity ................................................................. 42

4.2 ARIMA Modeling of Time Series ........................................ 43
   4.2.1 The Autoregressive (AR) Model .................................. 43
   4.2.2 The Moving Average (MA) Model ............................. 44
   4.2.3 The Autoregressive Integrated Moving Average Model
         (ARIMA) ..................................................................... 45

4.3 The Box - Jenkins Methodology ........................................ 46
   4.3.1 Identification ........................................................... 46
   4.3.2 Estimation ............................................................... 48
   4.3.3 Diagnostics .............................................................. 49
   4.3.4 Forecasting ............................................................... 49
   4.3.5 AR, MA and VAR in the Literature ............................ 50

Chapter 5 ARIMA Modeling of the Tanker Sector .................... 51

5.1 Description of Data ....................................................... 51
5.2 Modeling Procedure ...................................................... 53
5.3 Results ........................................................................ 53
   5.3.1 Product Tanker Model ............................................. 53
   5.3.2 Suezmax Tanker Model 1 ......................................... 55
   5.3.3 Suezmax Tanker Model 2 ......................................... 57
   5.3.4 VLCC Tanker Model 1 ............................................. 58
   5.3.5 VLCC Tanker Model 2 ............................................. 60

5.4 Comments .................................................................. 61

Chapter 6 A Non-Gaussian Approach to Modeling ................... 63

6.1 Random Walks and Efficient Markets ............................... 63
6.2 R/S Analysis of the Tanker Sector ................................. 66
6.3 Comments .................................................................. 67

Chapter 7 Concluding Remarks ............................................. 69

7.1 Suggestions for Further Research ................................... 69
Appendix A - Statistical Summary of Time Series................................. 71
Appendix B - Product Tanker Model....................................................... 75
Appendix C - Suezmax Tanker Model...................................................... 81
Appendix D - VLCC Tanker Model .......................................................... 87
Appendix E - R/S Analysis .................................................................... 93
References ............................................................................................. 99
List of Figures

Figure 1 - Fleet Innovations 1860-1995 ................................................................. 20
Figure 2 - Transport cost of oil and coal 1947-95 .................................................. 21
Figure 3 - Average size of tankers 1900-1997 ......................................................... 21
Figure 4 - The shipping market model ................................................................. 25
Figure 5 - Single voyage tanker rates 1947-1997 ..................................................... 26
Figure 6 - The short shipping cycle ................................................................. 28
Figure 7 – A generalized Wiener process ........................................................ 35
Figure 8 - Mean reversion at different speeds $\eta$ ............................................. 35
Figure 9 – Deterministic and stochastic trends .................................................. 41
Figure 10 - Realizations of an AR(1) process with different $\varphi_1$ ..................... 44
Figure 11 – Realizations of a MA(1) process with different $\theta_1$ ....................... 45
Figure 12 - ACF and PACF of an ARIMA(1,0,0) process .................................. 47
Figure 13 - ACF and PACF of a ARIMA(0,0,1) process ..................................... 48
Figure 14 - ACF of a non-stationary time series .................................................... 48
Figure 15 - Daily Earnings for product, suezmax and VLCC tankers .................. 52
Figure 16 - Forecast for product tanker using ARIMA(3,0,0) ......................... 55
Figure 17 - Forecast for suezmax tanker using ARIMA(1,0,0) ......................... 57
Figure 18 - Forecast for suezmax tanker using ARIMA(0,1,0) ......................... 58
Figure 19 - Forecast for VLCC tanker using ARIMA(1,0,0) ......................... 60
Figure 20 - Forecast for suezmax tanker using ARIMA(0,1,0) ......................... 61
List of Tables

Table 1 - Statistical characteristics of product, suezmax and VLCC tanker earnings .... 52
Table 2 - Augmented Dickey-Fuller test for unit roots ............................................ 54
Table 3 - Product tanker ARIMA(3,0,0) model .......................................................... 54
Table 4 - Augmented Dickey-Fuller test for unit roots ............................................. 56
Table 5 - Augmented Dickey-Fuller test for unit roots ............................................. 56
Table 6 - Suezmax tanker ARIMA(1,0,0) model ....................................................... 56
Table 7 - Augmented Dickey-Fuller test for unit roots ............................................. 59
Table 8 - Augmented Dickey-Fuller test for unit roots ............................................. 59
Table 9 - VLCC tanker ARIMA(1,0,0) model ............................................................ 59
Table 10 - Summary of AIC statistic ............................................................................ 61
Chapter 1  Introduction

The times when people looked at the future as simply being at the whim of God do not belong to the distant past. On the contrary, the notion of risk management owes its roots to people like Bernoulli, Fermat, Fibonacci, Pascal, Poincaré, Cardano, and others who lived no more than five hundred years ago. It might seem surprising that the problem of dealing with future uncertainty was not addressed by ancient civilizations such as the Egyptians or the Greeks; civilizations that were troubled with far more challenging ideas, such as the establishment of the scientific proof through the mastery of logic. However, this last sentence almost perfectly describes the Greeks’ nemesis in dealing with probability. Aristotle declared that a "mathematician who argues from probabilities in geometry is not worth an ace" (1). However, that does not imply that games whose roots were purely probabilistic were not common in the ancient world. Nevertheless, their outcomes were simply recorded and not studied, something that could mainly be attributed to the difficult-to-handle numbering system that prevailed.

The introduction of the Arabic numbering system enabled enlightened minds to dig through records of games, as well as real-life incidents, and discover shared patterns. The applications have varied from the prediction of the outcome of throwing a die to the likelihood that a man of a certain age will die within the next number of years. It is most often the case that new theories are developed in light of a potential economic gain. In the case of probability theory, the driving force was insurance policies. Although the companies engaged in this field initially failed to recognize the importance of sampling so as to get a sense of the probable future outcomes, academics have dealt with the problem extensively.

Probability and shipping merged in a very early stage in the game. The coffeehouse that Edward Lloyd opened in 1687 near the Thames River became the favorite place for ship operators and shippers. It served much like a contemporary brokerage house, where people would get together and arrange for the transportation of goods. Each transaction was recorded and was followed until either delivery took or an accident took place. This provided the perfect framework under

---

1 Bernstein (1999), p.17
which shipping insurance would be developed. Interesting for us is not so much the development of the shipping insurance industry, but rather the probabilistic approach undertaken by insurance companies to evaluate the risk of insuring a certain cargo going on a certain route, based on past accident records for that particular route. Nevertheless, the decision-making tools in this industry failed to keep up with the pace. Although the industry adopted the relatively simple discounting rules for analyzing projects, it failed to follow recent developments in time series analyses, such as unit root tests, cointegration, GARCH, non-linearity, seasonality, and others\(^{(2)}\). This became evident in the late 1980s and ‘90s, when these new technologies ceased to be abstract academic formulations and were successfully adopted by different markets around the world.

### 1.1 Scope of Thesis

Statistics, and later on econometrics, emerged from the solid foundations that probability theory had already established. Statistics is defined as “a branch of mathematics dealing with the collection, analysis, interpretation, and presentation of masses of numerical data”\(^{(3)}\). The application of mathematical statistics to economic data with a goal of providing empirical support to models constructed by mathematical economics later gave rise to the field of econometrics. Time series analysis, as employed in the context of econometric analysis, is concerned with the detection of non-random\(^{(4)}\) patterns in a sequence of measurements of economic variables observed over time\(^{(5)}\).

The variety of available time series models will be discussed in Chapter 4. However, due to time and size limitations, this work will mainly focus on the application of univariate time series analysis and more specifically on the autoregressive integrated moving average (ARIMA) methodology. The variable under examination will be a variant of freight rates, namely the ‘rent’ that a ship owner earns in exchange for providing transportation services. The area of shipping

---

2. These statistical notions are extensively discussed in Chapter 4  
4. The topic of how one should define randomness is extensively discussed in Chapter 6  
that will be investigated using the proposed methodology will be the tanker sector and more specifically 40,000 dwt (\textsuperscript{6}) product carriers, suzemax and VLCC crude oil carriers.

The aim of this thesis is to test whether the ARIMA class of models can accurately depict patterns that emerge in freight rate time series. If this is shown to be true, then the theory that states that the shipping market is efficient (\textsuperscript{7}) - in the sense of the Efficient Market Hypothesis (EMH) - will be questioned. This could have significant implications for everyone involved in the shipping industry, since the relatively simple ARIMA methodology could then be used for hedging and/or speculative purposes.

1.2 Overview of Subsequent Chapters

Chapter 2 serves as a general introduction to the maritime industry. It presents a brief history of events from the early 1880’s until the late 1990’s. Although the belief that history repeats itself is very popular among hoi polloi, academics most often tend to discard it as unscientific. However, the intuition that can be gained from studying past events is, in the view of the author, invaluable. Hence, a relatively extensive presentation of the market cycles and the micro- and macro-economic forces that have been affecting the seaborne shipping sector will be given.

Chapter 3 summarizes the literature on econometric modeling of vital shipping variables. Both univariate and multivariate models are presented, as well as a generic approach to stochastic processes.

Chapter 4 presents the theoretical framework on which the proposed ARIMA model is built. The pros and cons of univariate versus multivariate analysis are explained. Issues such as stationarity, heteroskedasticity, trends, non-linearity, and seasonality of time series are also discussed. Finally, the seminal work by Box and Jenkins on the methodology of model building for autoregressive models is laid out in some detail.

\footnote{Dwt stands for deadweight tonnage, i.e. the weight of cargo, needed to sink the ship to her design waterline.}
Chapter 5 starts by describing the data set used to test the author’s hypothesis that shipping freight rates follow an ARIMA-type process. It goes on by explaining how a specific ARIMA model that is intended to fit the data is identified, estimated, diagnosed, and eventually used for forecasting purposes. The results of the model will show whether the theory of J.J. Evans on the efficiency of the shipping market is valid and to what extent forecasting is possible.

Chapter 6 discusses the major shortcomings of conventional econometric modeling. It then introduces a newly developed theory for analyzing time series data. This theory was built upon the pioneering work of H.E. Hurst on biased random walks, and Benoit Mandelbrot on fractal time series. A simple example of rescaled range analysis (R/S analysis) will be presented, as it relates to the shipping industry. The methods presented in this chapter should serve as a guide for further research in maritime time series, as they appear to be very useful in identifying non-periodic cycles, a phenomenon vividly present in the shipping market.

Chapter 7 provides the reader with concluding remarks and suggestions for further research, including extensions to conventional univariate modeling, vector autoregressive models (VAR) and R/S analysis.

---

7 Evans (1994)
Chapter 2  The Maritime Industry

Efficient sea transportation was made possible by the technologies developed during and after the Industrial Revolution. Since then, the maritime industry has played a central role in the transformation of the national systems that existed in the 19th century into the single global economy of the late 20th century. However, the contribution of seaborne transport to the world economy should not be viewed as a one-way transaction; to the contrary, these two entities evolved together. As nations realized the benefits of extensive trade, ships were given a prime role in the new status quo, enabling ship owners to benefit from the explosive trends in trade patterns and, in turn, further facilitate the movement of cargo over long distances at surprisingly low prices.

2.1 The Cost of Transport

The low cost of sea transport has been one of the main factors behind the evolution of the shipping industry. This has been made possible by continuous technical revolutions (Figure 1) that have enabled ship owners to keep their costs at very low levels and hence offer their services to shippers at acceptable prices. It is worth noting that the cost of transporting a barrel of oil in the 1960s was approximately 50 cents, while in the 1990s it has averaged to about $1 (Figure 2). To put this into perspective, one has to consider that the price of crude oil averaged only $3 in 1960, with a ten-fold increase over the same number of years. The only transportation sector that has managed to keep costs from increasing throughout its history is the airline industry, constituting one reason why the shipping industry has failed to capture the long haul passenger transport business. However, in absolute terms, airborne fares are high compared to seaborne ones, hence limiting the use of airfreight to urgent or high value cargo.

From ancient times, ship captains relied on muscle power or wind to propel their ships and on their navigation instincts to find their way. With the advent of the Industrial Revolution, things changed drastically. Steam power was introduced and eventually the radio was invented, allowing for more efficient ships to be built. With the passage of time, breakthroughs in materials science and thermodynamics allowed ships to be built with steel and be fitted with
internal combustion engines, further increasing their capacity (Figure 3) and range. The next big step in the evolution of the shipping industry was the introduction of specialized vessels in the market. Given an adequate demand for the transportation of a certain type of cargo, a specialized vessel was economically viable, allowing for more efficient loading/unloading and stowage of cargo \(^8\). This was vital for keeping freight rates low at times when demand for sea transportation was increasing by leaps and bounds.

---

**Figure 1 - Fleet Innovations 1860-1995\(^9\)**

---

\(^8\) Before the introduction of the tanker, oil was transported in barrels by general cargo ships, leading to decreased stowage and increased weight.

\(^9\) Stopford (1997), p.20
Figure 2 - Transport cost of oil and coal 1947-95

Cost of oil transport

Cost of coal transport

Figure 3 - Average size of tankers 1900-1997

Average size of tanker (000 Dwt)

The average size of tanker remained below 20,000 dwt until 1960 when the closure of the Suez Canal in 1956 and growing Japanese oil imports triggered an escalation in size to over 300,000 dwt. The average size peaked in 1980 and then remained relatively stable.

Ibid, p5

Ibid, p.22
2.2 The Tanker Sector

In 1995, the tanker sector accounted for approximately 30% of the gross tonnage available for sea transportation. At the same time, crude oil and its products constituted more than 40% of the total tonnage shipped around the world, with the majority originating from the OPEC countries going to Western Europe, the USA and Japan. The average yearly haul of oil during the 1980s and 90s has been fairly constant to around 5,000 miles. However, this has not always been the case. After the Suez Canal had been closed in 1967, the average haul rose to more than 7,000 miles, decreasing the effective supply of ton-miles (12) and hence dramatically increasing freight rates to the benefit of the ship owner. The immediate reaction of the shipping community was to engage in a shipbuilding rush, in view of expectations that the canal would remain closed for a considerable amount of time. However, the combination of the canal’s re-opening with the delivery of the newly built vessels led to a dramatic downfall of freight rates. The downfall was further amplified in 1973 with the decision of OPEC to implement the well-known Arab Oil Embargo, decreasing long-haul demand.

It is evident by the events described above that the maritime industry is closely related to political events. After all, oil is a valuable strategic commodity that is mainly produced by only a limited number of nations, which hold similar political views. These nations use oil production as a means to make their political statements heard around the world. Fortunately, during the past year, OPEC itself committed to keeping the oil price at around $30 per barrel, giving assurances the levels of production. This has positively affected tanker owners, who can plan their newbuilding and scrapping policies on relatively safe assumptions about future prospects.

The 1973 Arab Oil Embargo had another significant impact on the transport policy of oil. Up until then, the major oil companies controlled the sea transport of oil by either holding a self-owned fleet or by engaging in long-term time charters. Consequently, such a highly structured environment minimized the spot market effect, which is widely known to be a major source of

12 Although intuitively it makes more sense to measure cargo-carrying capacity in tons, the “ton-mile” is a more accurate measure of the availability of capacity. To understand why, consider what would happen if all ships
volatility. However, after 1973 the oil companies, faced with uncertainty over trade volume, started relying more heavily on the spot market, causing a wide market fragmentation to occur. It is worth mentioning that the spot market’s share rose from 10% before 1973 to around 50% in the 1990s\(^{(13)}\).

### 2.3 Supply vs. Demand

From the viewpoint of the shipper, the ship owner provides a service that needs to be reliable, safe, and speedy, as well as offered at a competitive price. However, the reader should not be tricked into believing that the ship owner has the power to set the price for the provided service, namely the freight rate. On the contrary, the shipping market is a very good example of perfect competition where the ship owner is a price taker. On the other hand, it should be noted that for most cargoes the cost of transportation is a very small fraction of the cost of the production and hence, the equilibrium freight rate is most often set at the inelastic region of the demand function. In the past, ship owners, being aware of that latter fact, have tried to set prices by forming organizations\(^{(14)}\) where they pool their ships together and act as a monopoly. However, prospective speculators who obtain market share by undercutting prices have easily undermined the efforts of these artificial monopolies. After all, there are almost non-existent barriers to entry, apart from the need of a substantial initial investment. Consequently, pooling has only been successful in manipulating freight rates for specialized cargoes and routes and only for a limited period of time.

In general, the ship owner does not make all of his profit by operating his vessels. Instead, enormous fortunes have been made in the sales and purchase market (S&P market). This is made possible by the nature of the shipping market, which inherently moves in cycles and therefore makes it theoretically possible to buy low and sell high. However, playing the cycles is not only

\(\text{\footnotesize Stopford (1997), p.124}\)

\(\text{\footnotesize In the liner business, such pooling of ships goes under the name of “conference”.}\)
a matter of accurate prediction, but also of encountering another player with the opposite beliefs. In general, consensus about market movements does not create profit opportunities.

A theoretical approach to understanding market cycles is the equilibrium supply and demand model. On the one hand, the world economy, the average haul, political stability, and transport costs affect demand, while fleet productivity, shipbuilding production, and tonnage retirement affect supply. The way that these variables equilibrate themselves is through the freight rate market, as shown in Figure 4. For instance, when ships are in short supply, freight rates move higher and ship owners order ships. When these ships are delivered, freight rates depress, leading to increased lay up and scrapping, causing the supply of vessels to decrease and freight rates to increase once again. However, this process is not instantaneous, especially due to the time lag between the placement of the order and the delivery, which usually takes between 18 and 24 months. As a result, the market cycle amplitude is magnified.

The supply of vessels is determined by the interplay between four shipping markets: the newbuilding market, the freight market, the S&P market, and the scrap market. During a market trough, the newbuilding market stagnates and the scrap market flourishes. The least efficient vessels are laid up or scrapped and the rest of the fleet slow steams. At the other end of the spectrum, when the market is at its peak, the newbuilding market is most active, while the scrap market eases. All ships are operational and running at full speed. Under such conditions, the mismatch in supply and demand can only be settled in the long run, when the newly built vessels are delivered. The S&P market is usually active during both times, energized by speculators who play their game of seeking fast profit. A peculiarity of the shipping industry is that often expectations will appear to be ex post rational while in reality they are irrational. In other words, “tankship markets influence the actual price to the extent that it confirms their expectations” (15). Ultimately, what happens tomorrow depends on today.

---

(15) Zannetos (1966), p.10
There is little to be said about the demand side. It is usually viewed as an exogenous variable with many non-linear effects taking place. Under normal circumstances, the demand for sea transport, and in particular for oil, exhibits inelastic characteristics and hence is slightly affected by movements in the freight rates, especially in the short run. In other words, most of the imbalance is settled within the boundaries of the supply of vessels, as discussed above.

---

16 Ibid, p.116
2.4 Market Cycles

If one undertakes the simple task of plotting freight rates since the turn of the century, it will immediately become evident that a cyclical pattern dominates all other trends (Figure 5). To an outsider, this could suggest that fluctuations in the shipping market could be partially predicted. However, a closer look would reveal that the peaks and troughs occur at non-periodic times. To make matters even worse, simple econometric analyses that test for mean reverting patterns\(^{(17)}\) in freight rates yield contradictory results, depending on the time frame and route examined. Consequently, people who are entrusted with the Herculean task of forecasting the shipping markets are caught between the Scylla of shipowners who demand the development of an accurate model, and the Charybdis of randomness. Nevertheless, a recent approach to tackling randomness in the financial markets has shown promising signs. The theory of R/S analysis is by no means new, as it was developed in the early 20\(^{th}\) century by a hydrologist by the name of Hurst. However, there are no previous studies on the application of R/S analysis to the maritime industry, to the extent of the author’s knowledge, and thus this project will try to motivate new research, rather than complement existent ones.

\[ \text{\textit{Figure 5 – 1 year T/C tanker rates 1947-1997}} \]^\(18\)

\(^{17}\) The notion of mean reversion will be discussed in Chapter 3 
\(^{18}\) Stopford (1997), p.59
2.4.1 Short Cycles

As trade became a dominant force in the development of national economies, people identified the shipping industry as a potential source of high profits. Shipping being an international business, was faced with very few regulatory restrictions, making it a perfectly competitive market. As a result, during high profitable times, speculators would enter the market, forcing rates to decrease. Rates would rarely stabilize at their equilibrium point, however, as crowd sentiment would dominate and lead to oversupply, causing ships to eventually be run at a loss. Nevertheless, even the downturn of a cycle has its purpose as it drives the least efficient operators out of the market, allowing for an eventual recovery.

Dr. Martin Stopford has identified the stages of a cycle in a very illustrative graph, which is presented below (Figure 6). At stage one, freight rates are low, leading ship owners to operate at a loss. Short of cash, ship owners are forced either to lay-up or scrap their vessels. As the supply of ships drops, freight rates increase and the market enters stage two. Market sentiment still remains uncertain as to whether a full-blown recovery will develop. As freight rates climb even higher, optimism spreads and ships return to the market from their laid-up status. More significantly, orders for newbuildings are placed, which often exceed what the market can absorb in the long run. Consequently, when these ships start being delivered, supply once again exceeds demand and freight rates drop, leading the market to enter stage four. At this point it should be evident that the freight and S&P markets are in complete phase with each other, while both of them are 180 degrees out of phase with the scrap market. Finally, the growth rate of the fleet, which reflects the difference between shipbuilding and scrapping, also moves periodically, but tends to lag behind rates.

The dynamics described above are present in almost all short cycles. However, the story does not end here. To the shipowner’s despair, there are other factors that magnify the amplitude and length of those cycles. For instance, a war, an embargo or a sudden increase in the price of oil by OPEC would lead to a significant increase in uncertainty, which in market terms translates into increased volatility and hence higher risk. Accordingly, the ship owner should always keep in
mind that he is not alone in this game and therefore he should prudently plan his actions, considering that there are hundreds of other ship operators going after the same source of profit.

2.4.2 Long Cycles

In contrast to short cycles, long cycles have much less effect on the everyday operations of a shipping firm, if any at all. However, cumulatively, their presence has been shown to have a significant effect on the industry as a whole. Kondratieff identified three periods of expansion and contraction in the world’s economic activity between 1790 and 1916 (20). It has been argued that Kondratieff’s cycles have coincided with periods of technological innovation, such as the harnessing of steam, the railway boom, and the invention of electricity. In recent years, important technological innovations that could support Kondratieff’s theory are the development of the chemical industry, the introduction of the aircraft as a cost-effective means of travel, and the telecommunications revolution.

---

19 Ibid, p.43
20 Ibid, p.44
2.5 Historical Overview

The first shipment of oil via a sea route was in 1861. Since then, the tanker sector has come a long way. However, in understanding current market affairs, one has to first study the history of oil transport. A logical division of epochs would consist of three periods. The pre-World War I era, the interwar period and the post-World War II period. The two major landmarks between 1861 and 1913 were the introduction of the steam engine and the opening of the Suez Canal to tankers. Oil trade was sluggish at the beginning, as coal was the dominant source of power. However, the efficiencies associated with the production of energy from oil, transformed oil into a strategic commodity by the turn of the century.

The interwar period was characterized by a World War and a market crash. Modern war operations generate a big demand for oil. Moreover, after the end of the war, armored with optimism about future world affairs, shipowners went on a shipbuilding spree. Luckily, world demand for seaborne trade increased at an astonishing pace during the 1920s, supporting high freight rates. The 1930s saw the tanker sector experiencing one of its biggest depressions. Partly due to the extensive shipbuilding programs and partly due to the effects of the Great Depression, demand for oil transport fell sharply, leaving shipowners in despair. As a reaction to falling rates, tanker owners formed the first ship cartel in 1934. By 1937, the market boomed due to commercial activities but only for a short period of time, as the coming of World War II once again disrupted trade flows.

World War II was a great period for tanker owners. Since oil was behind every major production operation, and since desperate governments wanted to ensure the flow of oil into their military machines, ship owners could effectively set prices at astronomical levels. Once again shipbuilding capacity reached its peak, leading to an oversupply of vessels after the end of the war. Eventually, the Middle East emerged as a major producer of crude oil, increasing the average haul of trade and hence effectively decreasing the supply of tonnage, supporting rates at acceptable levels. Needless to say, the post-World War II era was not without surprises. The Korean War in 1951, the closing of the Suez Canal twice in 1956-57 and in 1967, the formation
of OPEC in 1960, the Yom Kippur War in 1973, and the Gulf War in 1991 had a definite influence on the volatility of the market.

However, probably the greatest depression of the post-World War Two era was a consequence of OPEC’s decision to raise the price of oil in 1973, when expectations were at their peak. The factors that lead to the collapse of the tanker sector in the mid-seventies were threefold. First, the tanker sector was leaving its flourishing during the late 1960s and early 1970s as a result of the Suez Canal closing. Moreover, technological improvements had allowed the production of ships with a capacity of over 300,000 tons, increasing the effective supply of tonnage per ship built. Finally, commercial banks \(^{(21)}\) were lending to ship owners with little or no collateral, reinforcing the oversupply momentum that was carried through the early 1970s. One can imagine the effect that OPEC’s decision had, as expectations failed to materialize. Phenomena like VLCC size vessels going directly from the shipbuilder to the scrap yard were not fictional creations, but actual market events. It is worth mentioning that the market recovered as late as 1986, when OPEC raised its production quotas.

Entering the 1990s the tanker market was once again faced with a shipbuilding boom. This time the rationale behind it was quite different. The aging tankers of the 1970s construction boom were reaching their scrapping age. Moreover, regulations as to operating restrictions set forth by the IMO, such as the double hull and segregated ballast tanks, were expected to render a significant portion of the fleet as non-operational. Finally, an increase in oil demand was foreseen, and with the Middle East being the only producer that could provide for it via long-haul transport, market sentiment started improving.

Evidently, the short cycle mechanics described in Section 2.4.1 manifest themselves over and over again during the 140 years of tanker history presented above. Additionally, there are political events that have further destabilized the market, by not allowing expectations to materialize. By now, the reader should have a clear view of the complexity of the shipping market and the parameters that could potentially form the basis of a complete equilibrium model.

---

\(^{(21)}\) Refer to Stokes (1997) for an excellent treatment of the subject
Chapter 3  Literature Survey and Motivation

The shipping industry is dominated by a handful of traditional maritime nations. In most cases, shipping companies are family businesses operated based on gut feeling. This means that sophisticated forecasting tools, such as the ones occupying this work, are rarely encountered in the market. This, in turn, discourages people from further engaging in econometric analyses of ocean freight rates, as it is almost certain that most shipping firms will disregard them as being too complicated or unrealistic. Therefore, a forecasting model for the maritime industry will initially be developed by the academia. However, high competition and the absence of market control by individual firms, dictates that even the slightest competitive advantage will guarantee high earnings and hence, a successful model will not pass unnoticed. Last but not least, shipping companies are starting to discover funding through equity and debt as an alternative to bank lending. As a consequence, the interaction of the shipping companies with the stock market and the highly sophisticated investment bank houses will motivate a change in the way things currently work in the industry.

The aim of this chapter is to present the evolution of econometric analysis related to the maritime industry, as it evolved during the century. Our survey will start with the work of Tinbergen and Koopmans in the late 1930s, and lead its way up to the work of Beenstock and Vergotis in the early 1990s. Finally, the notion of stochastic variables will be introduced, as it relates to the choice of modeling undertaken by the author. This is another approach to modeling the shipping market, moving away from the previous theories that investigate the relationship between supply and demand, based on economic theories.

3.1 Supply – Demand Models

As early as the 1934, Tinbergen investigated the sensitivity of freight rates to the level of demand, the level of available tonnage, and the price of bunker. Like most early models, he assumed that the level of demand was fixed at a constant level, while freight rates responded to shifts in the supply of tonnage. A few years later, Koopmans performed a theoretical analysis of the equilibrium condition between supply and demand. Most of his efforts were focused on
modeling supply, while the level of demand was exogenously specified. He postulated that at high freight rates, the supply of tonnage is very inelastic, as most ships are out of lay-up and the fleet is steaming at high speeds. As a consequence, even the slightest changes in demand, cause wide fluctuations in the freight market. At low freight rates, the supply of tonnage is almost completely elastic, as ships go in and out of lay-up freely, absorbing any excess demand. This result agrees with reality where we observe high volatility at high freight rates and vice versa.

In 1966, Zannetos set forth to develop a theory on how tankship rates move. He distinguished between the short and the long run, as well as between a spot and a time charter market. In addition, he devoted an extensive part of his work in discussing the mechanics behind the formulation of expectations. Like all of the models that preceded Zannetos, he failed to tie demand into his model, and eventually ended up considering it as an exogenous parameter. The issue of modeling demand was not addressed until the pioneering work of Eriksen in 1977, who considered the pattern of trade via sea routes, for a variety of dry cargos. His results showed that demand responds negatively to freight rates. This may generally be true for dry cargoes, but it is not necessarily the case for oil, as there are very few readily available substitutes for energy production. However, there is evidence to support the theory that as freight rates increase, trade is undertaken in short-haul routes. Once again, due to the particular nature of oil, which is mostly produced in the Middle East, the extent to which short-haul substitution can occur is limited.

The first serious attempt to examine the dry bulk and tanker sectors in a common context was first performed by the Center for Applied Research of the Norwegian School of Economics and Business Administration. The revolutionary aspect of the so-called Norship model was the integration of the shipbuilding and scrapping markets, as well as the effect that combined carriers have on the market equilibrium. The presence of these three parameters acted as a conductor of information between the dry bulk and tanker markets, causing shocks in one to also be felt by the operators in the other market.
By the late 1970s, the majority of models examined the tanker sector by considering the interaction between the freight market, the shipbuilding market, and the scrap market. Each new modeling attempt had a slightly different view of things; nevertheless, it still considered the same fundamental markets. In 1978, Hawdon attempted to model the differences between the short and the long run. In 1984, Rander examined the distinctive features of vessels that operate in the spot market versus those that operate under time charters, finding that spot market vessels have a more elastic reaction to changes in the freight rates. In 1981, Charemza and Gronicki developed a model with the distinctive feature that the markets did not necessarily clear, while in the late 1980s, Martinet formulated probably the most disaggregate model, considering many different tanker sizes and flags of registration.

In 1993, Beenstock and Vergotis constructed the most complete econometric model (22) that I came across in my research. In their model they try to resolve two issues that are ill-treated by the rest of the literature. Firstly, the interaction of market conditions with the demand for new ships, which is one of the fundamental causes of the cyclical behavior; and secondly, the treatment of future market expectations, which are endogenized, so as to agree with the results of the model. Their approach was to investigate the freight, second-hand, shipbuilding, and scrap markets and come up with the effect that anticipated and unanticipated shocks have, both in the short and the long run. Eventually, they constructed an integrated shipping model, which, like the Norship model, used the spillover effects of the combined fleet, the shipbuilding market and the scrap market to examine the dry bulk and tanker sectors simultaneously.

3.2 The Mean Reverting Nature of Freight Rates

Section 2.4 discussed the theoretical forces behind the cyclical behavior of the shipping markets, while Section 2.5 indicated how these forces manifest themselves in reality. Logically, all the models discussed above predict a cyclical behavior in accordance with the spirit of the actual market movements. However, it is not enough for a model to forecast cyclicality, but rather to capture the regularity with which the cycles occur, as well as their intensity and persistence. An

---

22 Beenstock et al (1993)
alternative tool to the supply – demand approach discussed thus far is the approach offered by stochastic processes and in particular the mean reverting process. Before we start discussing how stochastic processes can be used as a basis for a forecasting model, let us introduce some basic definitions for the unaccustomed reader.

3.2.1 Introduction to Stochastic Variables

A stochastic variable is one whose value changes over time in an uncertain way. A Markov process is a particular type of stochastic processes where only the present value of the variable is relevant in predicting the future. Of particular interest is a Wiener process. A random variable that follows a Wiener process has changes in its value over small intervals of time of the form

\[ \Delta z = \varepsilon (\Delta t)^{1/2} \]

where \( \varepsilon \) is a random drawing from a standardized normal distribution. The mean of such a process is zero and its variance is \( \Delta t \). Such a random variable would model a white noise signal.

A generalized Wiener process is nothing more than a Wiener process with the addition of a drift term. This process is more widely known as a geometric Brownian process and is formulated as

\[ dx = a dt + b \, dz \]

where \( dz \) is a Wiener process in continuous time, while \( a \) and \( b \) are constants. The mean of such a process is \( a dt \) and the variance is \( b^2 dt \). Figure 7 breaks down the process into its components, as a function of time.

3.2.2 Mean Reversion

A mean-reverting process is one that exhibits diffusive characteristics over the short run, but returns to some mean value in the long run. A good example of a commodity that should follow a mean reverting process is oil. In the short run, oil is subject to many destabilizing forces,

\[ ^{23} \text{A white noise series is distributed normally with mean zero, i.e. it follows a random walk.} \]
making its price volatile. However, over the long run it ought to return towards its marginal cost of production. Such a behavior can be modeled by

\[ dx = \eta(\beta-x) \, dt + \sigma \, dz \]

This form of mean-reverting process is called an Ornstein-Uhlnbeck process. Here \( \eta \) is the speed of reversion and \( \beta \) is the long-term equilibrium level. Figure 8 graphs an Ornstein – Uhlnbeck process for four different \( \eta \). Notice that the greater the speed of reversion, the faster the path returns towards its long run equilibrium level.

**Figure 7** – A generalized Wiener process

**Figure 8** – Mean reversion at different speeds \( \eta \)
The most interesting characteristic of a mean-reverting process is that it exhibits a cyclical behavior. Short-term destabilizing forces, coupled with a long-term restricting force is a scenario that should sound familiar, since this is how shipping cycles have been described. In the short-term, there is a limit as to how much the supply of tonnage can change, while in the long run, it is allowed to adjust freely, providing a balance force. In most cases, the long-term balance force acts a bit too strong and the long-term equilibrium level is overshot. From a cursory look, the shipping market exhibit mean reverting characteristics. However, looking like a mean reverting process is a long way from being one. Luckily for us, there are certain econometric techniques that allow to test our hypothesis.

A mean reverting process is a continuous time process, while in reality the data that we gather is discrete-time data. This not a problem, as there exists a discrete-time equivalent to the mean reverting process, namely an autoregressive process of order 1 (denoted by AR(1) for short). In the case of an Ornstein-Uhlnbeck process, the equivalent AR process is

\[ x_t - x_{t-1} = \beta(1 - e^{-\eta}) + (e^{-\eta} - 1)x_{t-1} + \varepsilon_t \]

where \( \varepsilon_t \) is normally distributed with mean zero and standard deviation \( \sigma \). In other words, an AR(1) process maintains that the evolution of variable \( x_t \) depends on its past plus a normally distributed error term. Testing whether the shipping market follows an autoregressive process is therefore equivalent to testing for mean reversion.

### 3.3.3 Concerns

So far, the discussion has focused on introducing the basic building blocks of the most commonly used stochastic process. The two basic elements that were introduced are the drift term (i.e. \( dt \)) and the diffusion term (i.e. \( dz \)). It has been noted that uncertainty is introduced to the model through the latter term. However, \( dz \) imposes a couple of limitations on the real processes that can be modeled. Firstly, it only allows for small changes in the value of the asset during each increment of time. It is well observed that certain events cause jumps in the value of

---

24 Hull (1997), p. 214  
25 Dixit et al. (1994), p. 75
assets and therefore assets that answer to such events cannot be accurately modeled with the technology developed up to this point. A possible fix would be the introduction of a dq term that would follow a Poisson type process, which would account for path discontinuities. Nevertheless, the greatest shortcoming that accompanies the use of dz is that it implicitly assumes that the variable under examination follows the ‘physics’ of Gaussian statistics \(^{(27)}\), whereas in reality this is rarely true. This fact should not puzzle the reader, as Gaussian statistics are mathematically traceable and hence highly valued. Even for the stock market, which is the most examined time series and is agreed not to follow Gaussian statistics, very few attempts have been made to model stock prices otherwise \(^{(28)}\), since the broader class of traceable distributions, namely the Paretian distributions \(^{(29)}\), are significantly more complex.

---

26 Ibid, p.76  
27 Gaussian statistics assume that the variables under examination are normally distributed.  
28 Refer to Peters (1994) and Peters (1996) for an excellent treatment of the subject  
29 We will return to that topic in Chapter 6
Chapter 4    Time Series Models

This thesis investigates the presence of patterns in maritime time series. Time series is defined as a set of discrete-time observations of a variable, as it evolves over time. This latter aspect of time series is what significantly complicates things. In contrast to the physical sciences, where a theory can be tested by holding certain control parameters constant and running the experiment over and over again, in time series analysis, the researcher must realize that all the available information is jumbled together in the time series. In other words, the researcher cannot replicate market conditions in order to test his hypothesis, but rather he has to formulate and test his hypothesis over the same data sample.

There are two dominant approaches to analyzing time series: spectral or frequency-domain analysis and time-domain analysis. The former is used to find various kinds of periodic behavior in series. It decomposes the data into sine and cosine waves of different frequencies, much like Fourier decomposition. Unfortunately, when a time series is analyzed using spectral methods, the researcher plays a much more involved role in identifying patterns. In other words, it takes much more experience to develop a model based on spectral analysis, indirectly making this approach more subjective to the views of the researcher. Moreover, any stationary process has an equivalent time-domain representation and hence, any information revealed in the frequency-domain can also be identified in the time-domain.

Any time series can be viewed as the discrete-time realization of a stochastic process. The ultimate goal is to build a model for the underlying stochastic process. This is directly analogous to cross-sectional modeling, where we use a sample to estimate the statistical characteristics of a population. Most modeling techniques rely on a property called stationarity to analyze the data. Broadly speaking, “a stochastic process is said to be stationary if its mean and variance are constant over time and the value of covariance between two time periods depends only on the distance or lag between the two periods and not on the actual time at which the covariance is computed” \(^{(30)}\). A most common cause of non-stationarity is the presence of a trend in the time

\(^{30}\) Gujarati (1995), p.713
series. At each point in time we can compute the mean of a trending time series, however, for different sub-sets the mean will be different.

4.1 Common Time Series Features

The most commonly encountered features of economic time series are trends, seasonality, conditional heteroskedasticity, and non-linearity. A first step in identifying the presence of one or more of the above-mentioned features in a specific data set is the construction of a time plot. Most often, a particular data set will incorporate a combination of these features, making it difficult to say exactly what is happening. Identifying these features on hand will only come with experience. Fortunately, there are certain techniques developed that allow the researcher to make a rigorous identification approach.

4.1.1 Trends

As explained above, trends are the most common source of non-stationarity for a time series and therefore they need to be addressed with special care. In particular there are two types of trends, namely a deterministic trend and a stochastic trend. The mechanisms that generates them is the following:

\[
\text{Stochastic Trend: } y_t = \alpha + \varphi y_{t-1} + \varepsilon_t \\
\text{Deterministic Trend: } y_t = \alpha t + \varepsilon_t
\]

where \( \varepsilon_t \sim N(0, \sigma) \). Figure 9 plots one realization of such trends for \( \alpha = 0.2 \), \( \varphi = 1 \), \( \mu = 0 \) and \( \sigma = 1 \). The key difference between the two trends is that the stochastic trend deviates from the average trend significantly, while the deterministic trend deviates from the average trend only to the extent of the error term \( \varepsilon_t \). In other words, for a stochastic trend model, shocks have a longer memory effect. In particular, because \( \varphi = 1 \), the effect of shocks does not dissipate, whereas if \( |\varphi| < 1 \), memory would decay exponentially, effectively deleting the effect of the shock after a certain number of time periods (Figure 10). This is a very significant aspect of a time series that needs to be addressed before a model can be specified. Moreover, a stochastic trend is in reality a random walk, since \( y_t \) is nothing more than \( y_{t-1} \) superimposed over a constant trend \( \alpha \) plus a
normally distributed error term (31). This feature alone is what introduces the non-stationarity in the time series.

![Figure 9 – Deterministic and stochastic trends](image)

To proceed with the model identification, the econometrician needs to remove such a trend in order for the time series to become stationary. Fortunately, this is easily done by differencing the series (33). In other words, instead of investigating for patterns in $y_t$, series $x_t$ is analyzed, where

$$x_t = y_t - y_{t-1}$$

Differencing once removes trends that have a constant rate of change. If a trend itself is changing over time, then we need to difference more than once. A non-stationary series that becomes stationary after it is differenced once is called integrated of order 1, or I(1). If we need to difference twice to make the series stationary, then the series is called integrated of order 2, or I(2), and so on and so forth.

---

31 Refer to our discussion of the geometric Brownian motion in Chapter 3.2.1
32 Figure produced by the author
33 The proof of why differencing effectively transforms a series into a stationary one is beyond the scope of this work.
A rigorous method for testing whether a time series is stationary was proposed by Dickey and Fuller in 1976. To illustrate this, consider the following model:

$$y_t = \varphi y_{t-1} + \varepsilon_t$$

The Dickey-Fuller test is nothing more than a t-test on the null hypothesis that $$\varphi = 1$$, which corresponds to a so-called unit root\textsuperscript{34}. If the t-test does not reject the null hypothesis, then we say that the time series has a unit root and hence it is non-stationary. In such a case the series needs to be integrated once and tested for stationarity again. Luckily for us, the Dickey-Fuller methodology is embedded in most statistical packages. The author tested for the presence of unit roots using STATA v.6.

4.1.2 Seasonality, Conditional Heteroskedasticity, and Non-Linearity

Seasonality is present in time series when “observations in certain seasons display strikingly different features to those in other seasons” (35). In general, seasonal fluctuations need to be removed before testing for unit roots. Similar to trend unit roots, one can test for the presence of seasonal unit roots. A proposed methodology for such tests was developed by Dickey, Fuller and Hasna in 1984. As far as maritime time series are concerned, there is no economic rationale as to why seasonality would enter the picture (36) and hence we do not test for its presence. On the other hand, one should expect freight rates to exhibit both conditional heteroskedasticity and non-linearity.

Conditional heteroskedasticity is a case of volatility clustering. This means that, within a specific time frame, the observed variable exhibits increased volatility when compared to the long-term norm. A possible cause for this, would be the arrival of important information to the market. For instance, even without looking at the freight market, one would expect to see increased volatility after the Suez Canal was closed in 1967, or after the OPEC raised the price of oil in 1973. In other words, conditional heteroskedasticity is a sign of market uncertainty.

\textsuperscript{34} Saying that a series has a unit root is equivalent to saying that the time series is non-stationary

\textsuperscript{35} Franses (1998), p.15

\textsuperscript{36} Among other things, oil is used for the production of energy that goes into heating. Consequently, we expect greater demand for oil during the winter months. However, the difference in demand is small when compared to the total tonnage of oil carried and hence it has little or no effect on the overall pattern of the freight rates.
Closely related to this, is the notion of non-linearity. The latter indicates that the reaction of a time series to certain exogenous factors does not depend on those factors alone, but also on the current state of the series itself. Unfortunately, rigorous treatment of these features is still under development and hence it is the opinion of the author that they should be left untouched in order not to complicate our model. Ideally however, one should investigate for the presence of all of the above-mentioned features when considering the best modeling approach.

4.2 ARIMA Modeling of Time Series

There are four basic approaches to economic forecasting. These involve the use of single-equation regression models, simultaneous-equation regression models, autoregressive integrated moving average (ARIMA) models, and vector autoregressive models (VAR). The latter two are closely related to the work of the author. They emerged from the seminal work of Box and Jenkins in the early 1970s. Their distinctive characteristic is that they are not based on any economic theory, but rather exploit patterns present in the data in order to make forecasts. Unlike traditional regression models, Box and Jenkins analyze the stochastic properties of economic time series by examining past, or otherwise called, lagged values of the series, as well as lagged values of other explanatory time series and their stochastic error terms. In the special case where a time series $y_t$ is analyzed using only its own lagged values $y_{t-p}$, that a univariate model is developed. If more than one time series are used in the model, then a multivariate model is in use.

4.2.1 The Autoregressive (AR) Model

An autoregressive model of order $p$ - denoted by AR($p$) from now on - is nothing more than a process for which $y_t$ depends on a weighted sum of its $p$-lagged values, $y_{t-1}, y_{t-2}, \ldots, y_{t-p}$, plus a normally distributed error term. In other words,

$$y_t = \varphi_1 y_{t-1} + \varphi_2 y_{t-2} + \ldots + \varphi_p y_{t-p} + \epsilon_t$$

where $\varphi_i$ are unknown constant parameters and $\epsilon_t \sim N(0,\sigma)$. The value of $\varphi_i$ determines how fast the effect of $y_t$ decays with time. Clearly, if any $\varphi_i > 1$, then the series is explosive. Moreover, in the case that $|\varphi_i| < 1$, the closer it is to one the longer a shock $\epsilon_t$ is felt. To illustrate this, 199
realizations of an AR(1) process for three different values of $\phi_1$ are plotted (Figure 10). All error terms are drawn from the standard normal distribution, except for $\varepsilon_{49}$, which is arbitrarily set equal to 10. Notice that after the unexpectedly large shock occurs, the series with $\phi_1 < 1$ returns to its normal level, i.e. after a certain number of observations the shock is not felt any more. However, for the process with $\phi_1 = 1$, the shock is felt indefinitely. Generally speaking, $|\phi_1| < 1$ models a mean reverting process, while $\phi_1 = 1$ models a random walk. In the case of $\phi_1 = 1$, it might be necessary to difference the series in order to remove the infinite memory effect, as discussed previously in the context of stationarity.

![Figure 10 - Realizations of an AR(1) process with different $\phi_1$](image)

**4.2.2 The Moving Average (MA) Model**

A moving average model of order $q$ - denoted by MA($q$) from now on - is nothing more than a process for which $y_t$ depends on a weighted sum of the $q$-lagged values of the error term, $\varepsilon_{t-1}, \varepsilon_{t-2}, \ldots, \varepsilon_{t-q}$, plus the current error term. In other words,

$$y_t = \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \ldots + \theta_q \varepsilon_{t-q}$$

---

37 Figure produced by the author
where $\theta_i$ are constant unknown parameters and $\varepsilon_i \sim N(0, \sigma)$. In short, an MA process is nothing more than a linear combination of lagged error terms. Similarly to the AR(1) model, the realizations of an MA(1) model for $\theta_1 = 0.1$ and $\theta_1 = 1$ are plotted, where the shocks $\varepsilon_i$ are drawn from the standard normal distribution (Figure 11). Again, we arbitrarily set $\varepsilon_{49} = 5$ to check for the memory of the process. It is quite obvious from the graphs that the shock is immediately absorbed by both series. Does this mean that the MA process has no memory? Indeed, it turns out that the effect of large shocks is of little importance when examined in the context of a MA model. Moreover, notice that the series with the greater $\theta_1$ exhibits greater variance over the whole sample.

![Figure 11 - Realizations of a MA(1) process with different $\theta_1$](image)

4.2.3 The Autoregressive Integrated Moving Average Model (ARIMA)

It is most often the case that a time series exhibits both AR and MA behavior. In such case a model of the form

$$y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + \ldots + \phi_p y_{t-p} + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \ldots + \theta_p \varepsilon_{t-q}$$

needs to be considered. This model is denoted as ARMA(p,q), since we consider $p$ lags of $y_t$ for the AR part and $q$ lags of $\varepsilon_t$ for the MA part. Such a model has $p + q$ unknown parameters that
need to be estimated explicitly. Auspiciously, routines that estimate these parameters are incorporated in most statistical packages. The author performs his analysis using SPSS v.8.

Before we end our introduction on autoregressive moving averages, we need to also incorporate the notion of integration. As discussed above, a non-stationary series can be made stationary by applying the differencing operator. A process that is I(2), for instance, and is modeled using an ARMA (p,q) model is said to follow an ARIMA(p,2,q) model. In general, an ARIMA(p,d,q) model involves p autoregressive lags, q moving average lags and d differencing operations to make the series stationary. Finally, special attention needs to be drawn to the fact that ARIMA-type models can only capture short-term memory relations, as discussed in Sections 4.2.1 and 4.2.2.

4.3 The Box - Jenkins Methodology

The choice of ARIMA as a possible model for freight rates is not accidental. In Section 3.2 we saw the relationship between mean reversion and AR-type models. Moreover, the extent to which a time series can be represented by an AR, MA or ARIMA model can be recognized by specific features of the data set itself. Such features are present in the data that we will examine in Chapter 5, and hence the choice of ARIMA is justified. G.P.E Box and G.M. Jenkins introduced the process that identifies these features in a time series. However, the theory they developed is much more than that. It is a complete guide from beginning to end on how a researcher should approach ARMA-type econometric modeling. They proposed a method that consists of four steps: identification, estimation, diagnostics, and forecasting.

4.3.1 Identification

During this step the appropriate values of p, q, and d are found. The primary tools used in this step are the autocorrelation function (ACF) and partial autocorrelation function (PACF). The ACF is defined as

---

38 Figure produced by the author
\[ \rho_k = \frac{\gamma_k}{\gamma_0} \]

where \( \gamma_k \) is the \( k \)th order autocovariance of the time series \( y_t \), i.e.

\[ \gamma_k = E[(y_t - \mu)(y_{t-k} - \mu)], \quad \text{for } k = \ldots, -1, 0, 1, \ldots \]

For example, the white noise series \( \varepsilon_t \) has \( \rho_k = 0 \) for all \( k \) different from zero. The PACF is defined along the same lines, only this time correlation is measured after controlling for correlations of intermediate lags \( k \). Interestingly enough, it turns out that an AR\( (p) \) process has an exponentially decaying ACF, and a PACF that is zero \(^{(39)}\) after the \( p \)th lag (Figure 12). Similarly, an MA\( (q) \) process has an exponentially decaying PACF, and an ACF that is zero after lag \( q \) (Figure 13). Eureka! Based on the ACF and PACF alone one can theoretically determine the values of \( p \) and \( q \). Nevertheless, things get a bit tricky in reality since, when a process follows an ARMA process, the features of the ACF and PACF get jumbled together. Another interesting feature of the autocorrelation function is that is can identify a non-stationary time series. As already mentioned, an exponentially decaying ACF is an indication of an AR process. However, if the decay is slow then we might suspect that the series under examination is non-stationary (Figure 14) and therefore we have a second tool, apart from the Dickey-Fuller test, to check for stationarity.

![Diagram](ACF_PACF_ARIMA.png)

**Figure 12 - ACF and PACF of an ARIMA(1,0,0) process** \(^{(40)}\)

\(^{(39)}\) A cautionary note: When we say that the ACF or the PACF is zero at a certain lag \( k \), we do not necessarily mean that they are identically equal to zero, but rather sufficiently close to zero to be rendered insignificant. The critical value is \( \frac{2}{\sqrt{n}} \), where \( n \) is the number of observations in the data set.

\(^{(40)}\) SPSS, Inc (194), p339
4.3.2 Estimation

Having identified the appropriate values for $p$, $d$, and $q$, the next step is to actually estimate $p$ $\phi$'s and $q$ $\theta$'s, or in other words the coefficients of the ARIMA model. Take caution in that the estimation must be done on the differenced series, if $d$ is found to be different than zero. Like most statistical packages, SPSS v.8 has the necessary routines that will allow us to estimate the required parameters. It is therefore beyond the scope of this paper to lay out the actual estimation procedure.

41 Ibid, p.338
42 Ibid, p.340
4.3.3 Diagnostics

After the model parameters are estimated and the model is fully specified, we need to verify that some fundamental assumptions about the nature of the model are satisfied. Probably the most important requirement is that the residuals from our model, namely the difference between what the model predicts and what really happened, are approximately white noise. If this is not the case, then we most probably missed a dynamic structure in $y_t$, since the presence of significant autocorrelations in the residual means that the residual still has predictive power. There are three ways to check the residual. We can either invoke the Q-statistic or the Ljung-Box statistic to verify that the residual series is normally distributed \(^{(43)}\). A third test involves plotting the ACF and verifying that it is insignificant for all lags greater than zero.

4.3.4 Forecasting

The ARIMA model owes its success to its forecasting power, especially in the short-run. In order to be able to check the accuracy of the forecast, it is common practice to reserve a part of the sample just for forecasting purposes. This is called out-of-sample forecasting. Once the model is estimated and diagnosed, a sample forecast is prepared for the time span of the reserved series. The two series are compared to see how well our model performed. It is common practice to report the 95% or 99% confidence intervals \(^{(44)}\) of the model, to see if the reserved series remains within these boundaries 95% or 99% of the times, respectively. If this is not the case, then the model is most probably underestimating the variance of the time series $y_t$. Apart from visual inspection, there are certain statistics that show how close the forecast is to the reserved series. The two most widely used statistics are the Akaike information criterion (AIC) and the standard prediction error (SPE) \(^{(45)}\).

---

\(^{42}\) Gujarati (1995), p.717
\(^{44}\) A 95% confidence interval gives you the model’s estimate of where the actual series will be with 95% confidence.
4.3.5 AR, MA and VAR in the Literature

Theoretically, the Box-Jenkins methodology can be applied to any time series, since it is not built on a particular economic theory. It is therefore a good candidate for examining maritime time series. However, in his review of the literature the author came across a handful of related articles. Kevin Cullinane (1992) uses an ARIMA(3,1,0) model to forecast the BIFFEX. Veenstra and Franses (1997) use an extension of the Box-Jenkins methodology to multivariate models to examine the relationship between various dry bulk freight rates. Finally, Berg-Andreassen (1997) investigates the validity of several theories for the relationship between period and spot rates. To the extent of the author’s knowledge ARIMA modeling of tanker freight rates is unique and should serve as a motivation for further research.
Chapter 5  ARIMA Modeling of the Tanker Sector

In the previous chapter we discussed the main features of a time series that can be described with an ARIMA-type model. Moreover, we saw how in theory we can identify the proper ARIMA model using the ACF and PACF of the time series. However, even before analyzing the time series using the ACF and PACF, we suspect that in principle a model with an AR component will be a good description of the freight market. The rationale behind that hypothesis is the cyclical, or otherwise mean reverting, nature of the shipping market, which in discrete time is equivalent to an AR process. Before we start the analysis, it is necessary to introduce the data that will be used to test our hypothesis.

5.1 Description of Data

The data set used in this work is a complement of Marsoft, Inc., a shipping consulting and forecasting firm based in Boston, Massachusetts. The time series involve quarterly data covering the first quarter of 1980 through the fourth quarter of 1998. In particular, daily earnings of three different ships operating on the spot market will be analyzed. These are the following (Figure 15):

1) A 40,000 dwt product tanker operating on the Carib/USAC route
2) A 140,000 dwt suzmax operating on the Bonny/USG route
3) A 280,000 dwt VLCC operating on the AG/East route

The choice of modeling earnings, rather than spot rates, is made on the basis that the shipowner is more interested in profit than revenue. Similar studies on freight rates often assume costs to be constant over time and hence have no impact on the decisions of the shipowner. However, this is not the case in reality and since data on earnings is available it would be naïve not to use it. Note that in calculating earnings, capital costs were not taken into account.

The author decided on the above described vessel types on the basis of their market characteristics. Product tankers are interesting because they represent a small fraction (\(^{46}\)) of the

\(^{46}\) According to Stopford (1997), product tankers carry a fifth of the total oil and oil derivatives shipped every year.
tanker fleet, operated by a relatively small number of shipowners. This favors inefficiencies and since the purpose of this paper is to examine the efficiency of the fleet, among other things, it comes as a natural choice. The suezmax and the VLCC are the most popular sizes operating in two of the biggest routes, namely Bonny/USG and AG/East, making them a logical choice, as well.

![Figure 15 - Daily Earnings for product, suezmax and VLCC tankers](image)

Selective statistics for the three vessels are presented below (Table 1):

<table>
<thead>
<tr>
<th></th>
<th>Product</th>
<th>Suezmax</th>
<th>VLCC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sample Size</td>
<td>76</td>
<td>76</td>
<td>76</td>
</tr>
<tr>
<td>Average</td>
<td>$10,789.5</td>
<td>$15,134.2</td>
<td>$21,438.6</td>
</tr>
<tr>
<td>Standard Deviation</td>
<td>3,917.2</td>
<td>5,153.5</td>
<td>9,510.2</td>
</tr>
<tr>
<td>Skewness</td>
<td>1.192</td>
<td>0.321</td>
<td>0.872</td>
</tr>
<tr>
<td>Kurtosis</td>
<td>2.788</td>
<td>-0.582</td>
<td>0.733</td>
</tr>
<tr>
<td>Minimum</td>
<td>$3,500</td>
<td>$6,200</td>
<td>$7,900</td>
</tr>
<tr>
<td>Maximum</td>
<td>$24,700</td>
<td>$27,600</td>
<td>$53,800</td>
</tr>
</tbody>
</table>

Table 1 - Statistical characteristics of product, suezmax and VLCC tanker earnings

---

47 Figure produced by the author
48 Refer also to Appendix A
5.2 Modeling Procedure

In accordance with the Box-Jenkins model building methodology described in Section 4.3, we start by checking the time series for stationarity. This is done in two independent steps. First, a qualitative examination of the ACF is performed, followed by the augmented Dickey-Fuller test for unit roots. For the time series that are found to be non-stationary, the differencing operator is applied and the new series is also tested for stationarity. Once the degree of integration of each time series is established, the value of d that is, the ACF and PACF are used to determine the values of p and q, namely the AR and MA lags that have explanatory power. This concludes the identification step of the model.

Before the model is estimated, a part of the data is reserved for evaluation of the out-of-sample forecast \(^{(50)}\). It is common practice to reserve 25% of the data for the evaluation portion. This means that in our case the models are estimated using data from Q1 1980 through Q4 1993, while the data from Q1 1994 through Q4 1998 is kept aside. Using SPSS v.8, the author performs an ARIMA(p,d,q) estimation using each of the three time series. In order to make sure that our models have taken advantage of all the information in the data, we check the residuals from each model for correlations. As described in Section 4.3.3, we simply need to verify that the residuals are normally distributed. Such a property can be verified either using the ACF, the Q-statistic or the Ljung-Box statistic. Finally, the results of the forecast are presented and evaluated.

5.3 Results

5.3.1 Product Tanker Model

The first step is to determine whether the earnings time series for the product tanker is stationary. From visual inspection of the ACF \(^{(51)}\) we conclude that the time series exhibits stationary characteristics. To verify this we perform the augmented Dickey-Fuller t-test using STATA v.6

---

\(^{(49)}\) Table produced by the author
\(^{(50)}\) Refer also to Appendix A
\(^{(51)}\) Refer also to Appendix B
(Table 2). In general, when performing a t-test, the null hypothesis is rejected if the absolute value of the t-test statistic is greater than the critical value. In the case of product tankers, the t-test suggests that the null hypothesis of a unit root can be rejected since $|\text{-5.766}| > |\text{1% critical value}|$. Therefore we conclude that the time series is stationary, namely I(0).

<table>
<thead>
<tr>
<th>t-test Statistic</th>
<th>1% Critical Value</th>
<th>5% Critical Value</th>
<th>10% Critical Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>-5.776</td>
<td>-4.097</td>
<td>-3.476</td>
<td>-3.166</td>
</tr>
</tbody>
</table>

**Table 2 - Augmented Dickey-Fuller test for unit roots**

Next, we need to reserve part of the data for the out-of-sample forecast evaluation. As explained above, the model is estimated using the data from Q1 1980 to Q4 1993. Plotting the ACF and PACF for that period, we see that the ACF is indeed exponentially decaying, while the PACF is significant only at the first three lags (53). This suggests the use of an ARIMA(3,0,0) model. Using the built-in routines of SPSS v.8, the model is estimated. The results are presented in Table 3 (54). Moreover, the normality test of the residuals is positive and hence we conclude that the model is correctly specified (55). The forecast is surprisingly close to the reserved data. Moreover, none of the reserved data is outside the 95% confidence level (Figure 16).

<table>
<thead>
<tr>
<th>Constant</th>
<th>$\Phi_1$</th>
<th>$\Phi_2$</th>
<th>$\Phi_3$</th>
<th>AIC</th>
</tr>
</thead>
<tbody>
<tr>
<td>9.2365</td>
<td>1.0024</td>
<td>-0.6249</td>
<td>0.4166</td>
<td>-4.7388</td>
</tr>
</tbody>
</table>

**Table 3 - Product tanker ARIMA(3,0,0) model**

---

52 Table produced by the author  
53 Refer also to Appendix B  
54 Refer also to Appendix B  
55 Refer also to Appendix B  
56 Table produced by the author
5.3.2 Suezmax Tanker Model 1

The first step is to determine whether the earnings time series for the suezmax tanker is stationary. From visual inspection of the ACF (58) we cannot conclude whether the series is stationary or not. In this case the augmented Dickey-Fuller test for unit roots is essential (Table 4). However, its results are also ambiguous. The t-statistic for the suezmax is between the 1% and 5% critical values and therefore we cannot reject the null hypothesis with certainty. Due to this, we difference the series and perform the test again (Table 5). Fortunately, the differenced series is stationary, and therefore our only ambiguity is whether the suezmax earnings are I(0) or I(1). Since we cannot make a certain statement about the degree of integration of the series, two models will be evaluated. Model 1 will treat the suezmax earnings as being I(0), while Model 2 as being I(1). The former modeling approach is demonstrated in the present section, while the latter is the topic of Section 5.3.3.

---

57 Figure produced by the author
58 Refer also to Appendix C
Next, we need to reserve part of the data for the out-of-sample forecast evaluation. As explained above, the model is estimated using the data from Q1 1980 to Q4 1993. Plotting the ACF and PACF for that period we see that the ACF is indeed exponentially decaying, while the PACF is significant only at the first lag \(^{61}\). This suggests the use of an ARIMA(1,0,0) model. The model is estimated using the built-in routines of SPSS v.8. The results are presented in Table 6 \(^{62}\). Moreover, the normality test of the residuals is positive and hence, we conclude that the model is correctly specified \(^{63}\). The forecast deviates from the actual data, but it does not cross the 99% confidence level (Figure 17).

<table>
<thead>
<tr>
<th>t-test Statistic</th>
<th>1% Critical Value</th>
<th>5% Critical Value</th>
<th>10% Critical Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>-3.931</td>
<td>-4.097</td>
<td>-3.476</td>
<td>-3.166</td>
</tr>
</tbody>
</table>

Table 4 - Augmented Dickey-Fuller test for unit roots \(^{59}\)

<table>
<thead>
<tr>
<th>t-test Statistic</th>
<th>1% Critical Value</th>
<th>5% Critical Value</th>
<th>10% Critical Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>-8.484</td>
<td>-4.097</td>
<td>-3.476</td>
<td>-3.166</td>
</tr>
</tbody>
</table>

Table 5 - Augmented Dickey-Fuller test for unit roots \(^{60}\)

<table>
<thead>
<tr>
<th>Constant</th>
<th>$\Phi_1$</th>
<th>AIC</th>
</tr>
</thead>
<tbody>
<tr>
<td>9.4901</td>
<td>0.6442</td>
<td>11.4801</td>
</tr>
</tbody>
</table>

Table 6 - Suezmax tanker ARIMA(1,0,0) model \(^{64}\)

\(^{59}\) Table produced by the author
\(^{60}\) Table produced by the author
\(^{61}\) Refer also to Appendix C
\(^{62}\) Refer also to Appendix C
\(^{63}\) Refer also to Appendix C
\(^{64}\) Table produced by the author
5.3.3 Suezmax Tanker Model 2

This model treats the earnings of the suezmax as being I(1). In other words, the model is estimated using the differenced series. This means that we need to identify the model from scratch and therefore, the ACF and PACF of the differenced series is plotted, keeping in mind that Q1 1994 through Q4 1998 is reserved and therefore not considered. The ACF and PACF of the differenced suezmax earnings are insignificant at all lags, suggesting that neither an AR, nor an MA component has any predictive power \(^{(66)}\). In other words, the differenced series acts much like a random walk and hence the model reduces to a simple ARIMA(0,1,0). Under the random walk hypothesis, the best future estimate is the value today, and hence the model simply forecasts constant earnings at the present level (Figure 18). The AIC measure comes out to 20.0500.

---

\(^{(65)}\) Figure produced by the author

\(^{(66)}\) Refer also to Appendix C
5.3.4 VLCC Tanker Model 1

The first step is to determine whether the earnings time series for the VLCC tanker is stationary. From visual inspection of the ACF \(^68\) we cannot conclude whether the series is stationary or not. In this case the augmented Dickey-Fuller test for unit roots is essential (Table 7). However, its results are also ambiguous. The t-statistic for the VLCC is between the 1% and 5% critical values and therefore, we cannot reject the null hypothesis with certainty. Due to this, we difference the series and perform the test again (Table 8). Fortunately, the differenced series is stationary, and therefore our only ambiguity is whether the VLCC earnings are \(I(0)\) or \(I(1)\). Since we cannot make a certain statement about the degree of integration of the series, two models will be evaluated. Model 1 will treat the VLCC earnings as being \(I(0)\), while Model 2 as being \(I(1)\). The former modeling approach is demonstrated in the present section, while the latter is the topic of Section 5.3.5.

\(^{67}\) Figure produced by the author

\(^{68}\) Refer also to Appendix D
Next, we need to reserve part of the data for the out-of-sample forecast evaluation. As explained above, the model is estimated using the data from Q1 1980 to Q4 1993. Plotting the ACF and PACF for that period we see that the ACF is indeed exponentially decaying, while the PACF is significant only at the first lag. This suggests the use of an ARIMA(1,0,0) model. The model is estimated using the built-in routines of SPSS v.8. The results are presented in Table 9. Moreover, the normality test of the residuals is positive and hence, we conclude that the model is correctly specified. The forecast deviates from the actual data, but it does not cross the 99% confidence level (Figure 19).

<table>
<thead>
<tr>
<th>Constant</th>
<th>$\Phi_1$</th>
<th>AIC</th>
</tr>
</thead>
<tbody>
<tr>
<td>9.7697</td>
<td>0.6177</td>
<td>42.6056</td>
</tr>
</tbody>
</table>

Table 9 - VLCC tanker ARIMA(1,0,0) model

---

69 Table produced by the author
70 Table produced by the author
71 Refer also to Appendix D
72 Refer also to Appendix D
73 Refer also to Appendix D
74 Table produced by the author
5.3.5 VLCC Tanker Model 2

This model treats the earnings of the VLCC as being I(1). In other words, the model is estimated using the differenced series. This means that we need to identify the model from scratch and therefore the ACF and PACF of the differenced series is plotted, keeping in mind that Q1 1994 through Q4 1998 is reserved and therefore not considered. The ACF and PACF of the differenced VLCC earnings are insignificant at all lags, suggesting that neither an AR, nor an MA component has any predictive power (76). In other words, the differenced series acts much like a random walk and hence the model reduces to a simple ARIMA(0,1,0). Under the random walk hypothesis, the best future estimate is the value today, and hence the model simply forecasts constant earnings at the present level (Figure 20). The AIC measure comes out to 52.9656.

---

75 Figure produced by the author
76 Refer also to Appendix C
5.4 Comments

In order to facilitate the reader in evaluating the above estimated models, a summary of the AIC statistic is presented in tabular form (Table 10). Evidently, the product tanker model does the best job in out-of-sample forecasting. Moreover, the ACF and PACF for the product tanker clearly suggest that a mean reverting process is at work. The fact that we can make a forecast with some accuracy is a sign that there are inefficiencies in the market. This is in conflict with the efficient market hypothesis. However, even before we analyzed the data, we suspected that product tankers are possible candidates against the efficiency hypothesis. However, this fact does not imply that the shipping freight market is inefficient overall. On the contrary, the results for the Suezmax and VLCC tankers suggest that the rest of the market acts almost like a random walk.

<table>
<thead>
<tr>
<th>Product Model</th>
<th>Suezmax Model 1</th>
<th>Suezmax Model 2</th>
<th>VLCC Model 1</th>
<th>VLCC Model 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>-4.7388</td>
<td>11.4801</td>
<td>20.0500</td>
<td>42.6056</td>
<td>52.9656</td>
</tr>
</tbody>
</table>

Table 10 - Summary of AIC statistic

---

77 Figure produced by the author
The suezmax and VLCC tankers examined have very similar characteristics. In both cases, the test for I(0) is inconclusive, while the null hypothesis for I(>1) is rejected. Once the stochastic trend is removed through differencing, both the suezmax and the VLCC behave like a random walk \(^{(79)}\), supporting the efficient market hypothesis. On the other hand, both of them receive a lower AIC value for Model 1, which means that the mean reverting model does a better job in explaining their behavior. This result is troubling, as it prohibits the author from making conclusive remarks. However, the theory developed in Chapter 6 will shed some light on this aspect of the problem.

In developing our model, we assumed that seasonality is not an issue. Moreover, conditional heteroskedasticity and non-linearity were not addressed for simplicity’s sake. However, the fact that the model does not perfectly follow the actual data in the sample indicates the presence of one or all of the neglected features (Forecast Figures). This suggests that future models should investigate the presence of such features and evaluate their impact on the forecasting accuracy of the model.

Finally, the choice of earnings as the modeling variable might estrange some readers. The main concern is that the author assumes that all shipowners face the same freight rate and cost structures. Although this might be true for the former, it is definitely not the case for the latter. Optimally, the cost structure needs to be modeled in such a way as to incorporate the flexibility that the shipowner has to reduce, for instance, maintenance expenditure in bad times. However, incorporating flexibility in the context of an ARIMA-type model is not possible. Other techniques need to be used such as the real options valuation technique \(^{(80)}\). Such an approach is beyond the scope of this thesis and can be the subject of further research.

---

\(^{(78)}\) Table produced by the author
\(^{(79)}\) Refer to Model 2 for both suezmax and VLCC tankers
\(^{(80)}\) The options approach to capital budgeting has the power of quantifying the value that arises from an active management that takes advantage of the options inherent in investment opportunities. For instance, when the market is low, the shipowner is short of cash and therefore the option of decreasing the maintenance budget to save cash and re-direct it to interest payments is more attractive. Such a scheme would alter the cost structure of the particular shipowner, making our modeling assumption, i.e. that all shipowners face the same cost structure, inaccurate.
Chapter 6  A Non-Gaussian Approach to Modeling

Econometric modeling has been dominated by linear models throughout its history. At the same time, there has been ample evidence that markets rarely react in a linear manner. Instead, they tend to overreact, a phenomenon that is repeatedly observed in the shipping industry. For example, when freight rates are high, people tend to get overoptimistic, which results in excess supply of vessels in the long run. Despite the inherent limitations of linear models, they continue to play a key role in modern econometric modeling, which can mainly be attributed to their mathematical simplicity. At the same time, new analytical methods are created, facilitating the implementation of more complex models. In the midst of all the research for the development of a realistic model, we find rescaled range analysis (R/S analysis). The beauty of this particular approach lies in its simple mathematical description and its relaxed assumptions regarding the time series under examination.

6.1 Random Walks and Efficient Markets

The efficient market hypothesis (EMH) states that “current prices reflect all available information” \(^{(81)}\) \(^{(82)}\). An immediate implication of this is that no price time series can serve as a basis for a forecasting model. In other words, prices should follow a random walk. In essence, this is a self-fulfilling argument, since any price predictability will almost immediately delete itself through arbitrage. The intuition behind the EMH theory is that there are too many players involved for the market to be wrong. As soon as a piece of information arrives, it is immediately incorporated in the prices, so that any knowledge of past prices is irrelevant to the future. This brings up the problematic treatment of time. In accordance with the EMH, most models view the markets as having short or even no memory at all. To illustrate the flaw of this approach consider the following paradigm: If in ten years all the parameters that affect a certain market variable are at their current value, the majority of the models would predict that the variable

\(^{81}\) Formally, there exist three versions of the EMH, each one interpreting the term “available information” differently. We will not go through the effort of distinguishing between them, since our discussion relates to the foundations of the theory and not the theory itself.

would also be at the current level. In other words, most forecasting mechanisms would not consider the path that the variable took during the elapsed time. However, in reality, certain events have an effect that persists over time, something that implies that markets have indeed a long memory.

A supporter of the EMH would argue that, in practice, markets are efficient, since no one can claim to have persistently outperformed the market by realizing abnormal returns. In the opinion of the author, this point of view is as true as it is false. In general, markets that involve many players are efficient, but as we will see, they are not efficient in the sense that the EMH implies. According to the efficient market hypothesis, returns are unrelated to each other and therefore independent. Under such conditions, the central limit theorem states that a large enough sample of independent observations is normally distributed, implying that the sample follows a random walk. In other words, a random walk is equivalent to saying that the EMH holds. Recent studies have shown that prices are not normally distributed, but rather exhibit leptokurtosis (83). This means that extreme events occur more often that the EMH assumes. A possible explanation might be that market agents do not react to new market conditions immediately, but rather wait for confirmation. Moreover, information does not arrive uniformly. This further encourages the accumulation of information, leading to wide fluctuations once the information is reflected on prices, causing the fat-tail effect. It is precisely this effect that hints the existence of long memory effect in the market.

In 1964, Mandelbrot suggested that the capital markets followed a family of distributions called Stable Paretian. These distributions are characterized by four parameters, namely \( \Phi, \alpha, \mu, \) and \( c. \) \( \Phi \) is called the characteristic exponent and determines the height of the extreme tails and the peakiness around the mean. Any distribution with \( \Phi < 2 \) exhibits leptokurtosis. \( \alpha \) is the index of skewness, while \( \mu \) is the \( i^{th} \) moment of the distribution. Finally, \( c \) indicates the total variation range of the distribution. A normal distribution has \( \Phi = 2, \) and \( \alpha = 0 \) (84). Although this theory was first tested in the capital markets, it eventually captured the interest of the maritime

---

83 A distribution that is similar in shape with the normal distribution, but with thinner and taller peak as well as fatter tails is said to exhibit leptokurtosis.
community. In his 1996 paper\(^{(85)}\), Berg-Andreassen examined the distribution of freight rates for ten different routes and found \(\Phi < 2\) for all of them. Although this cannot form the basis for a conclusive argument against the normality assumption, it is evidence that freight rates are distributed according to the Stable Paretian family\(^{(86)}\) and therefore exhibit long memory.

The implications of such a breakthrough are immense. Comprehensively, it is implicit that the theoretical basis of the ARIMA methodology is flawed at its foundation. To see how this is true, review our discussion of Sections 3.3.2 and 3.3.3. There, we showed that an AR process is the discrete time equivalent of a mean reverting process. However, a mean reverting process considers shocks, namely the \(dz\) term, to be normally distributed. Alas! Moreover, when investigating the properties of the ARIMA-type model, we saw that only the AR component has memory. This memory is persistent, i.e. long-term, only in the case of \(\varphi = 1\) (Figure 10). Nevertheless, this implies non-stationarity and hence cannot be directly modeled using an ARIMA-type approach. In summary, ARIMA-type models can only deal with short-term memory (depending on the number of lags used, i.e. the \(p\) and \(q\) values) of a Gaussian time series.

Up to this point we have essentially argued that a random walk is not really random. In other words, we have questioned the equivalence between the random walk and the notion of market efficiency, as it relates to forecastability, or lack of thereof. To treat this counterintuitive result, it is essential to introduce the notion of a biased random walk, also known as fractal time series\(^{(87)}\). A fractal time series is self-similar\(^{(88)}\) with respect to time, much like a geometrical fractal is self-similar with respect to space. This is also a property of the regular random walk. However, order and chaos, or otherwise put, determinism and randomness coexist in a fractal time series. In other words, there is local randomness, which prohibits us from making accurate forecasts, but

\(^{84}\) Peters (1994), p.199 - 200
\(^{85}\) Berg-Andreassen (1996)
\(^{86}\) Refer also to Appendix E
\(^{87}\) Random draws from a Stable Paretian distribution follow a biased random walk, much like random draws from a normal distribution follow a regular random walk.
\(^{88}\) Self-similarity is a property saying that any part of an object has the same characteristics as the whole. Fractals are perfect examples of self-similar structures. For example, a tree is self-similar (and fractal), the human lung is self-similar (and fractal), etc. Refer also to Jurgens et al (1992), p. 135-178
at the same time there is an underlying deterministic law that drives the series \(^{(89)}\). Most studies typically look for a deterministic law linearly superimposed by random noise. For instance, spectral analysis looks for periodic variations beneath random noise. Such a linear treatment is doomed to fail, since the deterministic law is non-linearly related to the local randomness and hence by removing the latter we distort the former.

### 6.2 R/S Analysis of the Tanker Sector

As mentioned above, maritime freight rates are distributed according to the Stable Paretian family of distributions. Therefore, any efficiency test should be performed within that context. The mathematical simplicity of Gaussian statistics made it the popular means of testing for efficiency, using the notion of the random walk. However, the normal distribution, being only a special case of a Stable Paretian distribution, can only test the EMH. Edgar Peters describes this approach as “putting the cart in front of the horse” \(^{(90)}\). In other words, the theory was created simply because it was the only one that could be tested. Fortunately, a methodology developed by Hurst in 1951 provides the basis for testing randomness within the broader boundaries of the Stable Paretian family of distributions.

In developing his theory, Hurst gave us a new statistic, which later took his name in honor of his valuable work. The Hurst exponent is very robust owing to the fact that it is based on a non-parametric method \(^{(91)}\). It describes a time series by a single number that varies between 0 and 1. If the Hurst exponent is equal to 0.5 then we have a truly random process, otherwise, the process has memory. This implies that the system under examination is dependent upon the path it took to get to its present condition. There are three distinct classifications of the Hurst exponent:

1) \(H = 0.5\) means that the system has no memory and the time series is truly random.

\(^{89}\) A perfect example of local randomness and global determinism is the Chaos Game. The Chaos Game draws a geometrical object based on probabilistic rules, yet the object that is eventually created is always the Sierpinski Triangle. In other words, the local randomness always reacts in a deterministic way. Refer also to Peters (1994), p.10-12

\(^{90}\) Peters (1996), p.14

\(^{91}\) Refer also to Appendix E
2) \( H < 0.5 \) means that the system exhibits negative correlations, otherwise known as mean reversion. The closer the value is to zero, the closer we are to perfect negative correlation.

3) \( H > 0.5 \) means that the series is trend re-enforcing. The closer the value is to 1, the closer we are to perfect correlation \(^{92}\).

Calculating the Hurst exponent for the product, suzmax and VLCC time series, we get some interesting results \(^{93}\). Both the product tanker and the VLCC exhibit mean reverting properties with a Hurst exponent of 0.4581 and 0.4669, respectively. On the other hand, the suzmax seems to be truly random as indicated by a Hurst exponent of 0.5002.

**6.3 Comments**

The results of the R/S analysis for the product tanker are in accordance with the results of the ARIMA model. Namely, the time series is predictably mean reverting, indicating market inefficiencies. A possible cause is the relatively small number of operators in this area, which implies that the market is not a true example of perfect competition. An argument for the VLCC can be constructed along the same lines. However, the results of the ARIMA model are not completely aligned with that of the R/S analysis. The ARIMA approach cannot safely tell whether the VLCC time series are mean reverting or random walk. Nevertheless, this is a statement about the short-term memory of the series. The R/S analysis, on the contrary, tells us that VLCC time series have a long-term mean reverting memory. Finally, the suzmax time series turn out to be the only truly random series. This can be attributed to the relatively large segmentation of this particular tanker size that favors perfect competition and allows for an efficient freight market.

\(^{92}\) Notice that the lower the value of the Hurst exponent, the greater the volatility of the time series.

\(^{93}\) Refer also to Appendix E
Chapter 7  Concluding Remarks

The aim of this work is to investigate the presence of patterns in the maritime freight market. This is done by selectively modeling earnings for three different tanker vessels. The results show that the product tanker time series exhibits short-term mean reverting memory. Along the same lines, the VLCC time series is also found to be mean reverting, but in a long-term sense. Finally, the suezmax time series is a true random walk and hence it is the only efficient freight market. A possible explanation is that the product tanker sector is smaller, allowing shipowners to be more than price takers. To the other extreme, the suezmax tanker sector is highly fragmented competing closely with at least two other tanker sectors for substitutability by size, favoring perfect competition. Somewhere in between lies the VLCC tanker sector. Although there is evidence that it is efficient in the sense of the EMH, R/S analysis suggests that there are long-term patterns embedded in its earnings time series.

The above results were derived in two steps. First, going along with traditional econometric modeling, earnings were assumed to be almost normally distributed, allowing for an ARIMA-type methodology to be implemented. The shortcoming of such an approach relates to the time horizon over which it can test for the presence of patterns. We discussed the reasons why such a model cannot test for true market efficiency, as it can only detect short-term memory effects. Then, we applied a more robust approach, namely R/S analysis, which has the twofold advantage of testing for long-term memory of non-Gaussian time series.

7.1 Suggestions for Further Research

An obvious suggestion for further research, within the context of the Box-Jenkins methodology, is the development of a multivariate model. One of the special features of univariate analysis is that it does not rely on economic theories. However, in selecting for the appropriate variables to incorporate in a multivariable model, some kind of economic theory has to be assumed. Such models are called vector autoregressive models (VAR). In principle, different versions can account for all the features discussed in Section 4.1; nevertheless, the computational complexity increases significantly.
As far as R/S analysis is concerned, a first improvement to the approach undertaken by the author would be to estimate the Hurst exponent using more extensive data. In this work, the Hurst exponent was calculated using quarterly data from 1980 to 1998. However, R/S analysis is a data intensive technique, since it investigates for the presence of long-term memory effects. Therefore, more accurate results would emerge from investigating a longer time series. Moreover, we can test the validity of the H exponent by scrambling the data, altering their order in the sample and performing the R/S analysis again. In our case, both the product and VLCC time series should yield an H very close to 0.5, while the suemax time series should give the same result as before. The rationale for this is that, by scrambling the data we destroy the order of the system. However, if scrambling actually resulted in an H-value that is further away from 0.5 than the original, then we would suspect that the fundamental assumptions for the time series are flawed and hence, an R/S analysis is not meaningful. Such a test is left to the interested reader.

The author interpreted the presence of leptokurtosis in the time series as evidence of long-term memory and hence R/S analysis was rendered as the desirable modeling technique. However, it has demonstrated that autoregressive conditional heteroskedasticity models (ARCH) can accurately model fat-tailed distributions. In this context, the autoregressive component of the model searches for short-term patterns, while the conditional heteroskedasticity routine treats the fat-tail effect. It should be noted that although ARCH-type models are non-linear models, they are not self-similar. Finally, we saw that, although an AR process can have long-term memory in the special case where $\varphi$ is very close to 1, the estimation techniques cannot handle such a phenomenon and hence the time series needs to be treated with the differencing operator prior to estimation. This limits the AR component to modeling only short-term memory. A possible fix comes through the use of fractional integration. This results in the so-called autoregressive fractionally integrated moving average (ARFIMA), which treats the fat-tail effect and at the same time is self-similar.
Appendix A - Statistical Summary of Time Series
Earnings, Product [$/day]

<table>
<thead>
<tr>
<th>Percentiles</th>
<th>Smallest</th>
</tr>
</thead>
<tbody>
<tr>
<td>1%</td>
<td>3500</td>
</tr>
<tr>
<td>5%</td>
<td>5400</td>
</tr>
<tr>
<td>10%</td>
<td>5700</td>
</tr>
<tr>
<td>25%</td>
<td>8350</td>
</tr>
<tr>
<td>50%</td>
<td>10350</td>
</tr>
<tr>
<td>75%</td>
<td>12600</td>
</tr>
<tr>
<td>90%</td>
<td>16000</td>
</tr>
<tr>
<td>95%</td>
<td>17500</td>
</tr>
<tr>
<td>99%</td>
<td>24700</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Largest</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
</tr>
<tr>
<td>Std. Dev.</td>
</tr>
<tr>
<td>Variance</td>
</tr>
<tr>
<td>Skewness</td>
</tr>
<tr>
<td>Kurtosis</td>
</tr>
</tbody>
</table>

Earnings, Suezmax [$/day]

<table>
<thead>
<tr>
<th>Percentiles</th>
<th>Smallest</th>
</tr>
</thead>
<tbody>
<tr>
<td>1%</td>
<td>6200</td>
</tr>
<tr>
<td>5%</td>
<td>7800</td>
</tr>
<tr>
<td>10%</td>
<td>8200</td>
</tr>
<tr>
<td>25%</td>
<td>10650</td>
</tr>
<tr>
<td>50%</td>
<td>15150</td>
</tr>
<tr>
<td>75%</td>
<td>18650</td>
</tr>
<tr>
<td>90%</td>
<td>22700</td>
</tr>
<tr>
<td>95%</td>
<td>25100</td>
</tr>
<tr>
<td>99%</td>
<td>27600</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Largest</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
</tr>
<tr>
<td>Std. Dev.</td>
</tr>
<tr>
<td>Variance</td>
</tr>
<tr>
<td>Skewness</td>
</tr>
<tr>
<td>Kurtosis</td>
</tr>
</tbody>
</table>

Earnings, VLCC [$/day]

<table>
<thead>
<tr>
<th>Percentiles</th>
<th>Smallest</th>
</tr>
</thead>
<tbody>
<tr>
<td>1%</td>
<td>7900</td>
</tr>
<tr>
<td>5%</td>
<td>9500</td>
</tr>
<tr>
<td>10%</td>
<td>10100</td>
</tr>
<tr>
<td>25%</td>
<td>14550</td>
</tr>
<tr>
<td>50%</td>
<td>18850</td>
</tr>
<tr>
<td>75%</td>
<td>28650</td>
</tr>
<tr>
<td>90%</td>
<td>32700</td>
</tr>
<tr>
<td>95%</td>
<td>38700</td>
</tr>
<tr>
<td>99%</td>
<td>53800</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Largest</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
</tr>
<tr>
<td>Std. Dev.</td>
</tr>
<tr>
<td>Variance</td>
</tr>
<tr>
<td>Skewness</td>
</tr>
<tr>
<td>Kurtosis</td>
</tr>
</tbody>
</table>
Appendix B - Product Tanker Model
ACF and PACF

ARIMA(3,0,0) Estimation Results

>Warning # 16445
>Since there is no seasonal component in the model, the seasonality of the
data will be ignored.

MODEL: MOD_9
Model Description:
Variable: EARNINGP
Regressors: NONE
Non-seasonal differencing: 0
No seasonal component in model.

Parameters:
AR1 < value originating from estimation >
AR2 < value originating from estimation >
AR3 < value originating from estimation >
CONSTANT < value originating from estimation >
Analysis will be applied to the natural logarithm of the data.

95.00 percent confidence intervals will be generated.

Split group number: 1 Series length: 60
No missing data.
Melard's algorithm will be used for estimation.

Termination criteria:
Parameter epsilon: .001
Maximum Marquardt constant: 1.00E+09
SSQ Percentage: .001
Maximum number of iterations: 10

Initial values:
AR1  1.00236  
AR2  -.62487  
AR3   .41663  
CONSTANT  9.23651  

Marquardt constant = .001
Adjusted sum of squares = 2.8630415

Iteration History:

<table>
<thead>
<tr>
<th>Iteration</th>
<th>Adj. Sum of Squares</th>
<th>Marquardt Constant</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2.8390698</td>
<td>.00100000</td>
</tr>
<tr>
<td>2</td>
<td>2.8388030</td>
<td>.00010000</td>
</tr>
</tbody>
</table>

Conclusion of estimation phase.
Estimation terminated at iteration number 3 because:
Sum of squares decreased by less than .001 percent.

FINAL PARAMETERS:

Number of residuals 60
Standard error .22186666
Log likelihood 6.3694032
AIC -4.7388064
SBC 3.6385719

Analysis of Variance:

<table>
<thead>
<tr>
<th>df</th>
<th>Adj. Sum of Squares</th>
<th>Residual Variance</th>
</tr>
</thead>
<tbody>
<tr>
<td>56</td>
<td>2.8387960</td>
<td>.04922481</td>
</tr>
</tbody>
</table>

Variables in the Model:

<table>
<thead>
<tr>
<th>B</th>
<th>SEB</th>
<th>T-RATIO</th>
<th>APPROX. PROB.</th>
</tr>
</thead>
<tbody>
<tr>
<td>AR1</td>
<td>1.0545811</td>
<td>.11478292</td>
<td>9.187614</td>
</tr>
<tr>
<td>AR2</td>
<td>-.6948712</td>
<td>.15646502</td>
<td>-4.441065</td>
</tr>
<tr>
<td>AR3</td>
<td>.4832456</td>
<td>.11453706</td>
<td>4.219120</td>
</tr>
<tr>
<td>CONSTANT</td>
<td>9.2572839</td>
<td>.16401507</td>
<td>56.441666</td>
</tr>
</tbody>
</table>
Covariance Matrix:

<table>
<thead>
<tr>
<th></th>
<th>AR1</th>
<th>AR2</th>
<th>AR3</th>
</tr>
</thead>
<tbody>
<tr>
<td>AR1</td>
<td>0.01317512</td>
<td>-0.01269599</td>
<td>0.00310960</td>
</tr>
<tr>
<td>AR2</td>
<td>-0.01269599</td>
<td>0.02448130</td>
<td>-0.01262423</td>
</tr>
<tr>
<td>AR3</td>
<td>0.00310960</td>
<td>-0.01262423</td>
<td>0.01311874</td>
</tr>
</tbody>
</table>

Correlation Matrix:

<table>
<thead>
<tr>
<th></th>
<th>AR1</th>
<th>AR2</th>
<th>AR3</th>
</tr>
</thead>
<tbody>
<tr>
<td>AR1</td>
<td>1.0000000</td>
<td>-0.7069231</td>
<td>0.2365276</td>
</tr>
<tr>
<td>AR2</td>
<td>-0.7069231</td>
<td>1.0000000</td>
<td>-0.7044360</td>
</tr>
<tr>
<td>AR3</td>
<td>0.2365276</td>
<td>-0.7044360</td>
<td>1.0000000</td>
</tr>
</tbody>
</table>

Regressor Covariance Matrix:

<table>
<thead>
<tr>
<th></th>
<th>CONSTANT</th>
<th>CONSTANT</th>
</tr>
</thead>
<tbody>
<tr>
<td>CONSTANT</td>
<td>0.02690094</td>
<td></td>
</tr>
</tbody>
</table>

Regressor Correlation Matrix:

<table>
<thead>
<tr>
<th></th>
<th>CONSTANT</th>
<th>CONSTANT</th>
</tr>
</thead>
<tbody>
<tr>
<td>CONSTANT</td>
<td>1.0000000</td>
<td></td>
</tr>
</tbody>
</table>

The following new variables are being created:

<table>
<thead>
<tr>
<th>Name</th>
<th>Label</th>
</tr>
</thead>
<tbody>
<tr>
<td>FIT_1</td>
<td>Fit for EARNINGP from ARIMA, MOD_9 LN CON</td>
</tr>
<tr>
<td>ERR_1</td>
<td>Error for EARNINGP from ARIMA, MOD_9 LN CON</td>
</tr>
<tr>
<td>LCL_1</td>
<td>95% LCL for EARNINGP from ARIMA, MOD_9 LN CON</td>
</tr>
<tr>
<td>UCL_1</td>
<td>95% UCL for EARNINGP from ARIMA, MOD_9 LN CON</td>
</tr>
<tr>
<td>SEP_1</td>
<td>SE of fit for EARNINGP from ARIMA, MOD_9 LN CON</td>
</tr>
</tbody>
</table>

Note: The error variable is in the log metric.

**ARIMA(3,0,0) Diagnostic Results**

![ACF for Residuals](image)
Appendix C - Suezmax Tanker Model
**ACF and PACF for Model 1**

![](image)

**ARIMA(1,0,0) Estimation Results**

>Warning # 16445
>Since there is no seasonal component in the model, the seasonality of the data will be ignored.

**MODEL: MOD_17**

Model Description:

Variable: EARNINGS
Regressors: NONE

Non-seasonal differencing: 0
No seasonal component in model.

Parameters:

AR1 < value originating from estimation >
CONSTANT < value originating from estimation >

Analysis will be applied to the natural logarithm of the data.
99.00 percent confidence intervals will be generated.
Split group number: 1 Series length: 60
No missing data.
Melard's algorithm will be used for estimation.

Termination criteria:
Parameter epsilon: 0.001
Maximum Marquardt constant: 1.00E+09
SSQ Percentage: 0.001
Maximum number of iterations: 10

Initial values:
AR1    0.64416
CONSTANT  9.49006

Marquardt constant = 0.001
Adjusted sum of squares = 3.9790478

Iteration History:

<table>
<thead>
<tr>
<th>Iteration</th>
<th>Adj. Sum of Squares</th>
<th>Marquardt Constant</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>3.9782312</td>
<td>0.00100000</td>
</tr>
</tbody>
</table>

Conclusion of estimation phase.
Estimation terminated at iteration number 2 because:
Sum of squares decreased by less than 0.001 percent.

FINAL PARAMETERS:

Number of residuals    60
Standard error         0.26067734
Log likelihood         -3.7400687
AIC                    11.480137
SBC                    15.668826

Analysis of Variance:

<table>
<thead>
<tr>
<th>DF</th>
<th>Adj. Sum of Squares</th>
<th>Residual Variance</th>
</tr>
</thead>
<tbody>
<tr>
<td>58</td>
<td>3.9782312</td>
<td>0.06795267</td>
</tr>
</tbody>
</table>

Variables in the Model:

<table>
<thead>
<tr>
<th>B</th>
<th>SEB</th>
<th>T-RATIO</th>
<th>APPROX. PROB.</th>
</tr>
</thead>
<tbody>
<tr>
<td>AR1</td>
<td>0.6549451</td>
<td>0.09821473</td>
<td>6.66850</td>
</tr>
<tr>
<td>CONSTANT</td>
<td>9.4909148</td>
<td>0.09458402</td>
<td>100.34374</td>
</tr>
</tbody>
</table>

Covariance Matrix:

<table>
<thead>
<tr>
<th>AR1</th>
</tr>
</thead>
<tbody>
<tr>
<td>AR1</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>AR1</th>
<th>0.0964613</th>
</tr>
</thead>
</table>
Correlation Matrix:

\[
\begin{array}{c}
\text{AR1} \\
\text{AR1} & 1.0000000 \\
\end{array}
\]

Regressor Covariance Matrix:

\[
\begin{array}{c}
\text{CONSTANT} \\
\text{CONSTANT} & 0.00894614 \\
\end{array}
\]

Regressor Correlation Matrix:

\[
\begin{array}{c}
\text{CONSTANT} \\
\text{CONSTANT} & 1.0000000 \\
\end{array}
\]

The following new variables are being created:

<table>
<thead>
<tr>
<th>Name</th>
<th>Label</th>
</tr>
</thead>
<tbody>
<tr>
<td>FIT_2</td>
<td>Fit for EARNINGS from ARIMA, MOD 17 LN CON</td>
</tr>
<tr>
<td>ERR_2</td>
<td>Error for EARNINGS from ARIMA, MOD 17 LN CON</td>
</tr>
<tr>
<td>LCL_2</td>
<td>99% LCL for EARNINGS from ARIMA, MOD 17 LN CON</td>
</tr>
<tr>
<td>UCL_2</td>
<td>99% UCL for EARNINGS from ARIMA, MOD 17 LN CON</td>
</tr>
<tr>
<td>SEP_2</td>
<td>SE of fit for EARNINGS from ARIMA, MOD 17 LN CON</td>
</tr>
</tbody>
</table>

Note: The error variable is in the log metric.

**ARIMA(1,0,0) Diagnostic Results**
Normal Q-Plot Residuals

ACF and PACF for Model 2

ACF

Lag Number
Transforms: natural log, difference (1)

Partial ACF

Lag Number
Transforms: natural log, difference (1)
Appendix D - VLCC Tanker Model
ACF and PACF for Model 1

ARIMA(1,0,0) Estimation Results

>Warning # 16445
>Since there is no seasonal component in the model, the seasonality of the data will be ignored.

MODEL: MOD_19
Model Description:
Variable: EARNINGV
Regressors: NONE
Non-seasonal differencing: 0
No seasonal component in model.
Parameters:
AR1 < value originating from estimation >
CONSTANT < value originating from estimation >
Analysis will be applied to the natural logarithm of the data.
99.00 percent confidence intervals will be generated.
Split group number: 1  Series length: 60
No missing data.
Melard's algorithm will be used for estimation.

Termination criteria:
Parameter epsilon: .001
Maximum Marquardt constant: 1.00E+09
SSQ Percentage: .001
Maximum number of iterations: 10

Initial values:
AR1  .61770
CONSTANT  9.76969

Marquardt constant = .001
Adjusted sum of squares = 6.6839258

Iteration History:

<table>
<thead>
<tr>
<th>Iteration</th>
<th>Adj. Sum of Squares</th>
<th>Marquardt Constant</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>6.6829102</td>
<td>.00100000</td>
</tr>
</tbody>
</table>

Conclusion of estimation phase.
Estimation terminated at iteration number 2 because:
Sum of squares decreased by less than .001 percent.

FINAL PARAMETERS:

Number of residuals 60
Standard error  .33814058
Log likelihood  -19.302785
AIC  42.60557
SBC  46.794259

Analysis of Variance:

<table>
<thead>
<tr>
<th>DF</th>
<th>Adj. Sum of Squares</th>
<th>Residual Variance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Residuals 58</td>
<td>6.6829090</td>
<td>.11433905</td>
</tr>
</tbody>
</table>

Variables in the Model:

<table>
<thead>
<tr>
<th>B</th>
<th>SEB</th>
<th>T-RATIO</th>
<th>APPROX. PROB.</th>
</tr>
</thead>
<tbody>
<tr>
<td>AR1</td>
<td>.6081786</td>
<td>.10286243</td>
<td>5.912544</td>
</tr>
<tr>
<td>CONSTANT</td>
<td>9.7696990</td>
<td>10863743</td>
<td>89.929399</td>
</tr>
</tbody>
</table>

Covariance Matrix:

<table>
<thead>
<tr>
<th>AR1</th>
</tr>
</thead>
<tbody>
<tr>
<td>AR1</td>
</tr>
</tbody>
</table>
Correlation Matrix:

\[
\begin{array}{cc}
AR1 & AR1 \\
& 1.0000000 \\
\end{array}
\]

Regressor Covariance Matrix:

\[
\begin{array}{cc}
CONSTANT & CONSTANT \\
& .01180209 \\
\end{array}
\]

Regressor Correlation Matrix:

\[
\begin{array}{cc}
CONSTANT & CONSTANT \\
& 1.0000000 \\
\end{array}
\]

The following new variables are being created:

<table>
<thead>
<tr>
<th>Name</th>
<th>Label</th>
</tr>
</thead>
<tbody>
<tr>
<td>FIT_3</td>
<td>Fit for EARNINGV from ARIMA, MOD_19 LN CON</td>
</tr>
<tr>
<td>ERR_3</td>
<td>Error for EARNINGV from ARIMA, MOD_19 LN CON</td>
</tr>
<tr>
<td>LCL_3</td>
<td>99% LCL for EARNINGV from ARIMA, MOD_19 LN CON</td>
</tr>
<tr>
<td>UCL_3</td>
<td>99% UCL for EARNINGV from ARIMA, MOD_19 LN CON</td>
</tr>
<tr>
<td>SEP_3</td>
<td>SE of fit for EARNINGV from ARIMA, MOD_19 LN CON</td>
</tr>
</tbody>
</table>

Note: The error variable is in the log metric.

**ARIMA(1,0,0) Diagnostic Results**

![ACF Chart](chart.png)
Normal Q-Plot for Residuals

ACF and PACF for Model 2

Transforms: natural log, difference (1)
Appendix E - R/S Analysis
Product, Suezmax, and VLCC Time Series Distributions

Product Earnings Distr vs Normal Distr.

Suezmax Earnings Distr vs. Normal Distr.

VLCC Earnings Distr vs. Normal Distr.
Step by Step Guide to R/S Analysis

1. Begin with a time series of length M. Convert this into a time series of length \( N = M - 1 \) of logarithmic ratios:

\[
N_i = \log \left( \frac{M_{i+1}}{M_i} \right), \quad I = 1, 2, 3, \ldots, (M - 1)
\]

2. Divide this time period into \( A \) contiguous subperiods of length \( n \), such that \( A \cdot n = N \). Label each subperiod \( I_a \), with \( a = 1, 2, 3, \ldots, n \). For each \( I_a \) of length \( n \), the average value is defined as:

\[
e_a = \frac{1}{n} \sum_{k=1}^{n} N_{k,a}
\]

where \( e_a = \) average value of \( N_i \) contained in the sub period \( I_a \) of length \( n \).

3. The time series of the accumulated departures (\( X_{k,a} \)) from the mean value for each subperiod \( I_a \) is defined as:

\[
X_{k,a} = \sum_{i=1}^{k} (N_{i,a} - e_a), \quad k = 1, 2, 3, \ldots n
\]

4. The range is defined as the maximum minus the minimum value of \( X_{k,a} \) within each subperiod \( I_a \):

\[
R_{I_a} = \max(X_{k,a}) - \min(X_{k,a})
\]

where \( 1 \leq k \leq n \).

5. The sample standard deviation calculated for each subperiod \( I_a \):

\[
S_{I_a} = \left[ \frac{1}{n} \sum_{k=1}^{n} (N_{k,a} - e_a)^2 \right]^{0.5}
\]

6. Each range, \( R_{I_a} \), is now normalized by dividing by the \( S_{I_a} \) corresponding to it. Therefore, the rescaled range for each \( I_a \) subperiod is equal to \( R_{I_a} / S_{I_a} \). From step 2 above, we had \( A \) contiguous subperiods of length \( n \). Therefore, the average R/S value for length \( n \) is defined as:

---

\[
\left( \frac{R}{S} \right)_n = \frac{1}{A} \sum_{a=1}^{A} \left( \frac{R_{1a}}{S_{1a}} \right)
\]

7. The length \( n \) is increased to the next higher value, and \( (M - 1)/n \) is an integer value. We use values of \( n \) that include the beginning and ending points of the time series, and steps 1 through 6 are repeated until \( n = (M - 1)/2 \). We can perform an ordinary least squares regression on \( \log(n) \) as the independent variable and \( \log(R/S)_n \) as the dependent variable. The slope of the equation is the estimate of the Hurst exponent, \( H \).

**R/S Graphs**

**Hurst Exponent - Product**

\[
y = 0.4581x - 0.3197 \\
R^2 = 0.9911
\]

**Hurst Exponent - Suezmax**

\[
y = 0.5002x - 0.4013 \\
R^2 = 0.9848
\]
Hurst Exponent

\[ y = 0.4669x - 0.4041 \]

\[ R^2 = 0.986 \]
References


Ref 6: Bodie, Zvi, Alex Kane and Alan Marcus. *Investments*. Boston, MA: Irwin, 1996


Ref 11: Evans, J. J. “Analysis of Efficiency of the Bulk Shipping Markets”. Maritime Policy and Management (1994), Vol. 21, No. 4, pp. 311-329


Ref 15: http://www.ioga.com/history_of_crude_oil_prices.html (Online), Access Date: Spring 2001


