Erasmus University Rotterdam

Maritime Economics and logistics

2018/2019

Statistical Analysis and Modeling of the Baltic Clean Tanker Index (BCTI) to Predict the Future Rates of the Baltic Clean Tanker Index

By

Abbas Karkhaneh

Rotterdam, August, 2019

Acknowledgment:

Acknowledgment:

The final part of my studies in Maritime Economics and Logistics (MEL) at the Erasmus University of Rotterdam is this thesis that has been carried out under the supervision of Assistant Professor Ted Welten. His encouragement and trust plus his academic guidelines helped me to overcome the difficulties I had while conducting this thesis. Without his support, my effort would not be of academic standard and quality offered by this thesis. I would also like to thank the other members of my supervisory committee.

I would like to appreciate efforts of the people from the MEL office, namely Ms. Felicia, Ms. Renee and Ms. Martha who has been there every day and supported us by their advice on our studies and life issues. And I want to thank lecturers who shared their professional knowledge and working perspective with us.

My special thanks to my family for supporting me all the way through my study.

Abstract:

This thesis scope is to construct models simulating Baltic Clean Tanker Index (BCTI) by using independent economical and industrial indices and indicators by performing statistical analysis on selected indices against BCTI. Therefore, modelling the feasibility of BCTI by using major independent economical and industrial indices has been examined. Methodology to achieve this goal was divided into two parts. At first statistical calculations of key characteristics of each indicator performed to find out their mean and Standard deviation. Then correlation of each index against BCTI and also against other indexes tested by calculation of correlation coefficient. Based on this finding a set of indices were selected to construct three models. On the second part, these models were used against two sets of data one set as input data and another set as output data (BCTI). Every model was tested against training and all data of their data sets to determine its predictive ability. The key finding of this thesis is to show the existence of a correlation between some of major microeconomic and macroeconomic indices as well as constructing a model based on these findings. The conclusion of this thesis is that BCTI can be modelled by related indexes with good fitting and predictive and forecasting characteristics.

Table of contents

Table of Contents

List of figures

List of tables

List of abbreviations

BCTI	Baltic Clean Tanker Index
BDTI	Baltic Dry Tank Index
BDI	Baltic Dry Index
ICS	International Chamber of Shipping
N4SID	Numerical algorithm for Subspace State Space System Identification
VLCC	Very Large Crude Carrier
ULCC	Ultra Large Crude Carrier
AFRA	Average Freight Rate Assessment
LNG	Liquefied Natural Gas
LPG	Liquefied Petroleum Gas
UNCTAD	United Nations Conference on Trade and Development
MARPOL	International Convention for the Prevention of Pollution from Ships
ABS	American Bureau of Shipping
OPEC	Organization of The Petroleum Exporting Countries
WTI	West Texas Intermediate
OECD	Organization for Economic Co-operation and Development
HFO	Heavy Fuel Oil
NEER	Nominal Effective Exchange Rate
LIBOR	London Interbank Offered Rate
MA	Moving Average
ARMA	Autoregressive Moving Average
ARIMA	Autoregressive Integrated Moving Average
ANN	Artificial Neural Network
MAPE	Mean Absolute Percent Error
MVLR	Multi-Variable Linear Regression
FFA	Forward Freight Agreement
WNN	Wavelet Neural Network
RSI	Relative Strength Index
MFI	Money Flow Index

1 Introduction

1.1 Introduction to Shipping

Shipping has been one of the best methods of the transport of the different types of cargo. The international shipping industry is responsible for the carriage of around 90% of world trade. (ICS, 2019)The growth in seaborne trade in the last century has led to the expansion of the shipping industry and its related businesses and markets such as shipbuilding, ship-broking, insurance, together with shipping finance and investment. (Alizadeh, 2011)To know that shipping has such a huge role in the international economy encourages those who are engaged in trade or economy to study shipping. There are more than 50000 oceangoing and coastal vessels around the world that are responsible for the transfer of the goods and energy as a form of cargo around the world (ICS, 2019)The shipping industry is divided to different form of transport, there are mainly container vessels, tankers and gas carriers, bulk carriers and general cargo and project carriers however global maritime transport is not limited to above categories. Tankers and gas carriers carry most of the world energy as a form of liquid gas or oil. Raw materials to feed industry is carried mainly by Bulk carries and to less extent by container vessels and container vessels mainly carry final products to customers worldwide. General cargo vessels mainly carry cargoes that cannot be carried by containers such as pipes and cargo with irregular size or shape. One of the most important types of ships is the tanker vessel. Tankers are responsible for the carriage of the oil and liquid chemicals that are main energy and industry derivers. This thesis focus is therefore on the oil industry since this industry is affected by many external factors predictions are a real challenge. Here I try to facilitate predictions by defining a model.

1.2 Research Objectives

The main objective of this thesis is to statistically conclude whether and to what extent several economic indexes are related to the Baltic Clean Tanker Index (BCTI). This index has high importance, especially for the tanker industry. The next step is to check the feasibility of predicting the Baltic Clean Tanker Index (BCTI) by observing targeted industrial and economical indexes by modelling the Baltic Clean Tanker Index (BCTI) and testing the model against historical data of the Baltic Clean Tanker Index (BCTI).

This model has a great interest for shipowners, charterers, investors, and academics involved in the maritime economy and transport and people involved in the maritime industry and interested in the global economy and oil industry as well.

1.3 Research Questions and Sub-Questions

As explained before the objective of this thesis is to test to what extent does the Baltic Clean Tanker Index (BCTI) related to macro and microeconomic indicators and based on that, test feasibility of developing a predictive model for Baltic Clean Tanker Index (BCTI). Therefore, the main research question is as below.

1.3.1 Main research question

Is Baltic Clean Tanker Index (BCTI) related to main macro and microeconomics indexes and is there a possibility of modelling of Baltic Clean Tanker Index (BCTI) to predict its future rate based on that indexes?

1.3.1.1 Sub questions

- To what extent tested economical indexes are correlated and related to the Baltic Clean Tanker Index (BCTI)

- Is this feasible to build a model using correlated indexes to model the Baltic Clean Tanker Index (BCTI) and predict the Baltic Clean Tanker Index (BCTI)?

- What will be the accuracy of the model according to the given data?

- What will be the respond of the model to the extreme conditions such as a war or economic recessions?

Baltic Clean Tanker Index (BCTI) is an indicator that is calculated by collecting this index data from different sources and is an important indicator for chartering of the tankers.

1.4 Research Design and Methodology

Since for prediction of Baltic Clean Tanker Index (BCTI), we shall develop a numeric model, therefore, this research will be quantitative. This research will start with statistical analysis of various economic and industrial indexes to find the correlation between those indexes and Baltic Clean Tanker Index (BCTI) and to determine if there is any correlation between selected indexes by this research with Baltic Clean Tanker Index (BCTI).

The final results will be a mathematical model to predict Baltic Clean Tanker Index (BCTI). The predicted freight rates will be tested against real data to determine their accuracy. The indexes that are to be tested against the Baltic Clean Tanker Index (BCTI) for their correlation are selected during the literature review. They will be explained and their background and how these indexes are provided will be discussed and their relation to Baltic Clean Tanker Index (BCTI) will be explained.

The relation between indexes and Baltic Clean Tanker Index (BCTI) will be statistically analyzed and the correlation between them will be shown. Those indexes that are more correlated to the Baltic Clean Tanker Index (BCTI) will be used as input for the predictive model. The results assume the indexes as input and Baltic Clean Tanker Index (BCTI) as output. For the purpose of this thesis, data will be provided from reliable sources such as Clarkson research service.

1.5 Structure of thesis

Chapter 1 - Introduction

The purpose of this chapter is to provide an overview of the topic.

Chapter 2 - Literature Overview

This chapter aims to explain the Baltic Clean Tanker Index (BCTI) in more details. It provides reader about the market, how it works and explains the indexes of this industry. Also, different methods of forecasting briefly will be explained

Chapter 3 - Literature Review

This chapter explains the work that has been done regarding forecasting of Baltic Clean Tanker Index (BCTI) till now and examines the relation of the indexes to the Baltic Clean Tanker Index (BCTI) to choose the most related ones for the modelling of BCTI.

Chapter 4 – Methodology

This chapter provides the reader with a more detailed theoretical background. This chapter is divided into two parts the first part is about the statistical approach to the subject and the tools that are used for examining the data and the second part is about methodology under which the predictive model is identified. The N4SID method as presented by Favoreel et al. (2000) will be presented in the second part.

Chapter 5 – Results

This chapter provides the reader with the results of this thesis. This chapter also is divided into two parts. The first part is about the statistical analyses of data and modelling results of the data. The results obtained from the model is then tested against the real data. Fitting of each model will be compared with other models. Based on that a model will be selected for refining and refined model will be used to forecast a 12-months range to test its accuracy.

Chapter 6 – Conclusions

This chapter presents the conclusion of the research, suggestion for future work and the limitations faced when conducting this research.

2 Theoretical overview

2.1 Introduction

Oil and oil products have been used by people since the prehistoric era. Commercial transfer of the oil through the sea in our time started around the 1850s when oil was transfer from Upper Burma by earthenware vessels to the river bank where it was loaded to boats holds for transfer to Britain. Breakbulk boats and barges were used to transport Pennsylvania oil in 40 gallons (150 l) wooden barrels. (Mike Ratcliffe, 1985)barrels were heavy and their weight was 20% of the total weight of the cargo (29 kg). They were expensive and they could not be used for several times since they tend to leak after the first use. In 1863 two sailboats designed to carry oil as oil tankers. Following their design on 1873 the first oil tanker steamer was built for Belgian owner. (Shrivastav Anuj, 2016) Modern oil tankers designs formed during late 1880s by design of the tankers with several cargo tanks that were divided in port and starboard sides (National Academy Press Washington, 1991) that eliminated problem of free surface effect and facilitated large parcel of the oil cargo to be carried over the ocean routes (Grace's Guide, 2016) After further developments in tankers and the transport of the oil by the sea during the first world war this industry became so important that after loss of tankers due to submarine attacks Georges Clemenceau wrote to US president Wilson "*Gasoline is as vital as blood in the coming battles…a failure in the supply of gasoline would cause the immediate paralysis of our armies*" (Timothy C. Winegard, 2016).

During the 2nd world war block construction of vessels invented by Americans because German warships were sinking more boats than it could be replaced by traditional methods. This invention plus welding greatly improved the shipbuilding industry. Following that invention, T2 tanker was introduced because they could be built relatively fast and in large numbers to replace sunken ships. T2 tankers introduced a new class and easier way of transport of oil by the sea. During the Suez crisis of 1956 companies forced to move around the Cape of Good Hope, this new condition removed restrictions imposed by Suez Canal construction and showed companies that bigger tankers will help more efficient transport of the oil.

After 2nd world war size of tanker vessels changed significantly. Political developments in the middle east and nationalization of the oil industry in middle east plus industrial development in west and more demand for energy fueled ship owner's competition to build larger tankers. Product tankers were replaced by crude oil tankers due to the demand of the market. Because of economies of scale, it was cheaper to carry oil over larger ships. Soon Panamax tankers were built and followed by that AFRAmax and later Suezmax tankers. VLCC and ULCC introduced to the industry in the late 60s and early 70s however oil crisis of 1976 due to the Iranian revolution and sudden increase of the oil price made the production of large tankers very risky due to a decrease of demand in the oil market. After the 2008 crisis a few large tankers are made and since product market developed while crude oil market relatively declined due to oil-producing country investment in refinery and chemical industry. Market changes caused more demand on AFRAmax size of vessels due to their suitable size in relation to routes they can use and cargo parcels they can accept.

Different types of liquid bulk cargo need a different type of tanker to carry the cargo, therefore, freight markets are different. The main two markets are Dirty Tanker and Clean Tanker markets. To measure market condition and values there are indices defined by market players. The Baltic exchange publishes the most commonly used indices in the tanker industry. Baltic Dirty Tanker Index rates show the market condition for dirty liquid cargo freight and Baltic Clean Tanker Index that is subject of this thesis is used to indicate freight rates in clean bulk liquid cargo.

Since these indices are of interest of many sectors in economy, industry and shipping, their relationship with different microeconomics and macroeconomics indices have been studied extensively by researchers. Some of these indicators and indexes that are more related to this thesis and the methods that have been used to show their relations will be explained in chapter 2 and chapter 3 this thesis.

2.2 Liquid bulk cargoes

The shipping industry is divided into different sectors that each of them is specialized to carry a particular kind of cargo. The segment of the industry that is dedicated to carrying liquids in bulk is characterized as “liquid bulk” since cargo is carried in large tanks that are a part of the ship, types of the vessel which are carrying these categories of cargo are called “tanker”.

There are different categories of liquid bulk cargoes which are mainly Hydrocarbon products such as oil, liquefied petroleum gas (LPG), and liquefied natural gas (LNG) followings are other main liquid bulk cargoes;

- Chemicals, such as ammonia, chlorine, and styrene monomer
- Freshwater
- Wine
- Molasses
- Citrus juice

Our focus in this paper is on indices related to the carriage of Hydrocarbon products except for LNG and LPG that are liquid gases.

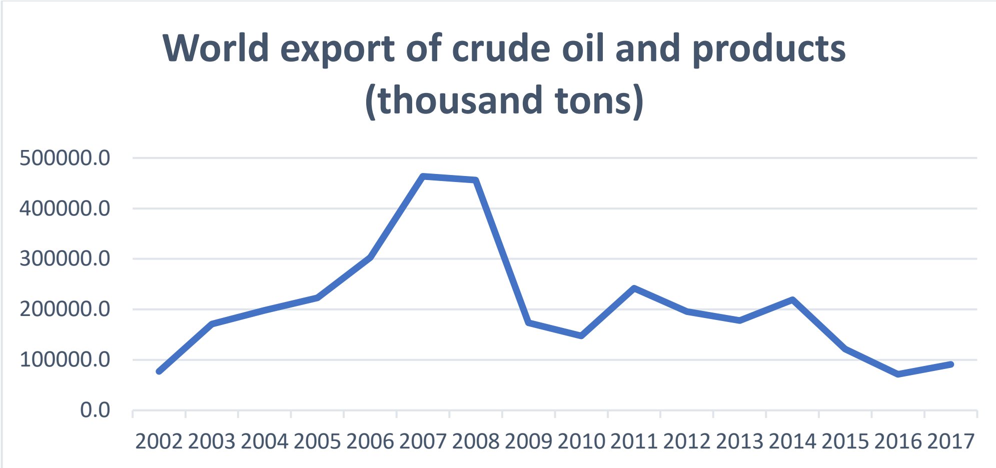


Figure 2.1 World export of crude oil and products (thousand tons) Source : (UNCTAD, 2019)

BCTI is dealing with tankers that carry oil products by sea. Carriage of the oil cargo by the sea is an indicator of consumption of the energy and initial products used in industry to produce goods and therefore, is an indication of the trends of the global industry. Figure 2.1 shows, a sharp drop in the quantity of carriage of the oil and oil product by the sea during the recession from 2008 to 2010 period. At this period the world economy experienced an economic recession.

2.3 Tanker vessel types

Oil tankers are mostly categorized by their size. Definition of an oil tanker as per MARPOL" ***Oil tanker*** *means a ship constructed or adapted primarily to carry oil in bulk in its cargo spaces and includes combination carriers,..., when carrying a cargo or part cargo of oil in bulk.* "(MARPOL Annex I reg. 1.5)

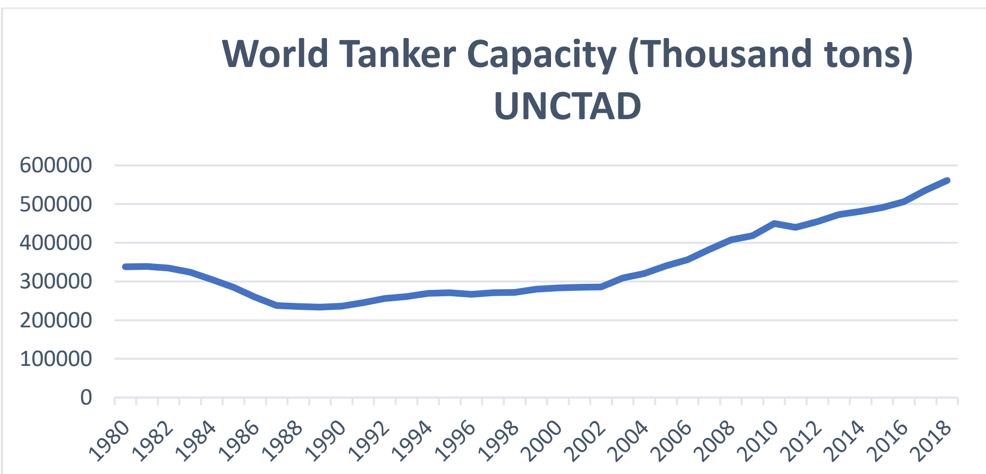


Figure 2.2 World Tanker capacity Thousand tons Source: (UNCTAD, 2019)

Tanker ships fleet capacity is constantly growing since the mid-80s. According to data from (Sin.Clarksons.net, 2019) tanker shipping market in 2018 was about 29.16% of total world shipping tonnage with the capacity of 56079 thousand tons carrying capacity and a total number of about 10420 tanker vessels trading worldwide (Sin.Clarksons.net, 2019).In 1954 SHELL developed the average freight rate assessment (AFRA) which classify tankers of different size. These categories were used mainly in tanker market for the definition of freight rates and market indices due to later development of the tanker market and increase of demand and change of the size of the oil tankers also a flexible market scale developed that presently used as the standard of the measurements of oil tanker sizes in the market (ABS, 2002).

The main categories of oil tankers according to ABS are as follow (ABS, 2002)

- Product tankers their size is between 10,000 to 60,000 dwt. These vessels are used in areas with more restrictions for draft and for cargoes with smaller parcel size that are mainly oil products. They are used in coastal waters and restricted lower depth waters such as the Caribbean Sea.
- Panamax tankers their size is between 60,000 to 80,000 dwt. They are designed to overcome the restriction imposed by the Panama Canal due to its width and depth of available water. Due to the expansion of the Panama Canal after 2016 a new design of the ship is introduced that is called new- Panamax or post-Panamax design.
- Aframax tankers their size is between 80,000 to 120,000 dwt. This category introduced during the 70s as a new category due to the size increase of the ships for measuring the size of the oil tankers. Aframax size is a size suitable for both product and crude oil cargo parcels.
- Suezmax tankers their size is between 120,000 to 200,000 dwt. The size represents tankers that can transit the Suez Canal while fully loaded. They are mainly used for the carriage of the Crude oil
- VLCC tankers are between 200,000 to 315,000 dwt. New Suez Canal allows passage of this type of the vessels but they may need to be lightened for transit trough Canal. This size is relatively a new design and not many of them are made due to the difficulty to find suitable cargo parcel for them.
- ULCC tankers are between 315,000 to 550,000 dwt not many of this type of the ship is made due to the same problem of difficulty to find cargo for them and their restriction in relation to available depth and width of navigable water for them.

Table 2.1: Tanker vessels market scales (ABS, 2002)

Scaling the tanker market			
Fixed AFRA Scale		**Flexible Market Scale**	
General Purpose (GP)	10,000-24,999 dwt	Product Tanker	10,000-60,000 dwt
Medium Range (MR)	25,000- 44,999 dwt	Panamax	60,000- 80,000 dwt
Long Range 1 (LR1)	45,000- 79,000 dwt	Aframax	80,000- 120,000 dwt
Long Range 2 (LR2)	80,000- 159,000 dwt	Suezmax	120,000- 200,000 dwt
VLCC	160,000- 319,999 dwt	VLCC	200,000- 315,000 dwt
ULCC	320,000- 549,999 dwt	ULCC	315,000- 550,000 dwt

Since AFRA scale is relatively obsolete only flexible market scale explained here.

2.4 Tanker vessel routes

Global energy perspectives show that demand for oil and oil products will continue to rise until 2030 (McKinsey, 2019).Shipping companies continue to support transport for the global market. This market is not evenly spread over the globe and its spread is based on the demand and supply in areas that cargo exist. Crude oil mainly comes from oil-producing countries such as Saudi Arabia and Persian Gulf areas and going to other parts of the world where it will be changed to other products or consumed. These areas are normally industrial countries. Refined and changed products will be exported from developed countries to other countries. The different area needs different products and extra products will be exported to where there is a demand this demand and supply showed in by Figure 2.4. Oil-producing countries send their products mainly over the sea by tanker vessels. As Figure 2.3 shows oil is sent over different continents which are separated by the sea. Figure 2.3 illustrates the movements of the oil around the world.

Table 2.2: Main trade routes of Clean tankers

Range	Name- origin and destination
Long range 2 (LR2)	**TC1-** Ras Tanura, Saudi Arabia, to Yokohama, Japan.
Long range 1 (LR1)	**TC5** – Ras Tanura to Yokohama **TC8 –** Jubail, Saudi Arabia, to Rotterdam
Medium range (MR) and Handysize	**TC2** – Rotterdam to New York **TC3 –** Caribbean to US Atlantic coast **TC4** – Singapore to Japan **TC6 –** Skikda, Algeria, to Marseilles, France **TC7** – Singapore to Sydney **TC10 –** Yeosu, South Korea, to Los Angeles **TC11** – Yeosu to Singapore **TC12 –** Sikka, India, to Chiba, Japan **TC14** – Houston to Amsterdam

Several routes, therefore, are shaped to support global trades. These routes are categorized to long-range and short-range and medium-range routes. Also, routes are different for dirty products and clean products as well. Some of the main routes of the clean tanker market are shown in Table 2.2.

Tanker routes and their development and demolishment impacted by industrial players and oil producers. Extensive explanation and research to explain routes and their change or their change of products carried by those routes is not the aim of this research, however, tanker routes will be explained in more detail in chapter 2.6 of this thesis.

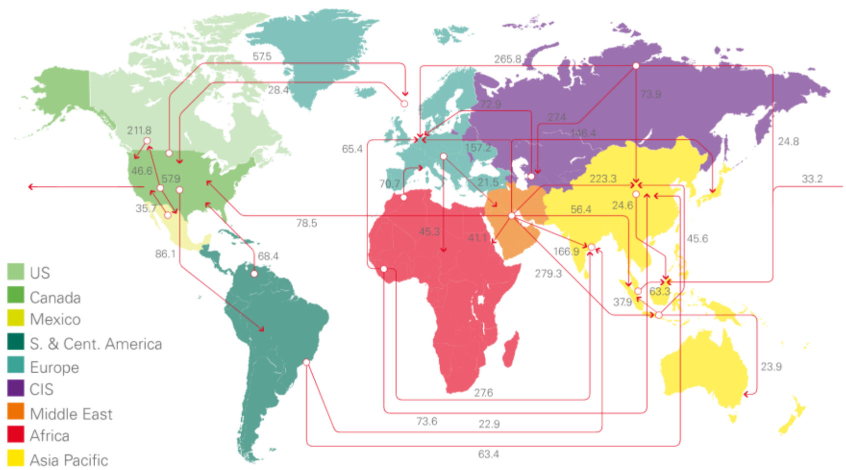


Figure 2.3: Oil major trade flows worldwide (Million tons)Source: (BP, 2019)

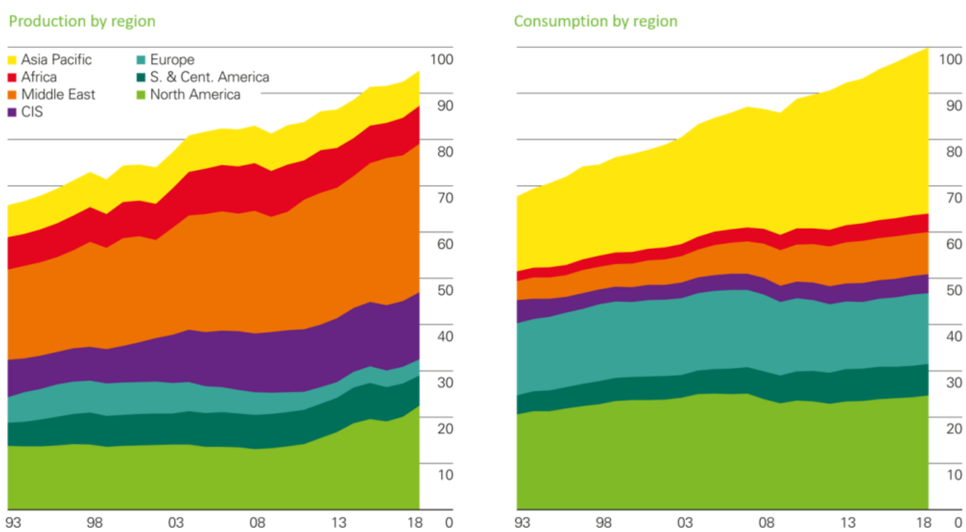


Figure 2.4: Oil production/consumption by region (Million barrels daily)Source: (BP, 2019)

2.5 Clean tanker and clean liquid bulk cargoes (Definitions)

A general definition of the clean tanker is "*A clean product tanker is a tanker that is dedicated to moving finished petroleum products so as to maintain their quality. This is in contrast to a dirty tanker that moves crude oil and fuel oil with less concern about maintaining tight quality levels. A clean product tanker will typically carry product in multiple tanks, allowing multiple products and product grades to be moved simultaneously while maintaining product segregation*". (MacKiensy, 2019)

Clean tanker products are normally Gasoline, Jet Fuel, Naphtha and Clean Condensates carried primarily on clean petroleum tankers, which include MR, LR1 and LR2 tankers. (Fenwick, 2007)

2.6 The Baltic exchange

Baltic exchange operates as a membership organization with over 600 members. It is an independent source of maritime information for trading and settlement of physical and derivative shipping contracts (Baltic Exchange, 2019). The Baltic exchange publishes a wide range of market reports and indicators on daily and weekly bases. The Baltic Exchange publishes a series of rates for assessment of the market rates for dry and wet bulk market and Forward Freight Assessment (FFA) market and its associated markets (Baltic Exchange, 2019). The Baltic production of Its indices is based on the confidential provision of assessment by panellists. The assessment shows a professional assessment of the panellists for the shipping route. Indexes are calculated by measuring trades over certain routes that will be discussed later in this thesis. The criteria for selecting routes are a. trade volume b. transparency c. standard terms.

To find each index a basket of routes on main geographical locations and trade routes is made with the average weighting of each route included in the basket. Routes that are used for the calculations are shown in chapter 2.6.1 of this thesis.

Baltic exchange index are named Dry Index (BDI), the Baltic Exchange Capesize Index (BCI), the Baltic Exchange, Panamax Index (BPI), the Baltic Exchange Supramax Index (BSI), the Baltic Exchange Handysize Index (BHSI), the Baltic Exchange Dirty Tanker Index (BDTI) and the Baltic Exchange Clean Tanker Index (BCTI) (Baltic Exchange, 2019).

Worldscale is tanker nominal freight scale that shows average freight rate for more than 320000 tanker voyages during a year and widely used for freight calculation in freight tanker market (WorldScale Association, 2019)

2.6.1 Main tanker indices (Baltic Exchange, 2019)

Followings are the main indices for tanker ship freight rates which are published daily by the Baltic Exchange. Trade routed defines the routinely contracts traded in the tanker freight market. (Baltic Exchange, 2019). Other tanker indices such as Clean Tankers Asia (BITR Asia) and Time charter Equivalents (TCE) can be found in Baltic exchange website.

All Worldscale and TCE rates include any costs where burning Low Sulphur fuel is mandated e.g. in an Emission Control Area (ECA). (Baltic Exchange, 2019)

2.6.1.1 Baltic Exchange Dirty Tanker Index[1] (BDTI)

Individually traded Dirty Tanker routes (Baltic Exchange, 2019)

- *TD1:* 280,000mt. Middle East Gulf to US Gulf (Ras Tanura to Louisiana Offshore Oil Port (LOOP)). Laydays/cancelling 20/30 days from index date. Age max 15 yrs. 2.5% total commission.
- *TD2:* 270,000mt. Middle East Gulf to Singapore (Ras Tanura to Singapore). Laydays/cancelling 20/30 days from index date. Age max 15 yrs. 2.5% total commission.
- *TD3C:* 270,000mt. Middle East Gulf to China (Ras Tanura to Ningbo).Laydays/cancelling 15/30 days from index date. Age max 15 yrs. 3.75% total commission.
- *TD6:* 135,000mt. Black Sea to Mediterranean (Novorossiyk to Augusta). Laydays/cancelling 10/15 days from index date. Age max 15 yrs. 2.5% total commission.
- *TD7:* 80,000mt. North Sea to Continent (Hound Point to Wilhelmshaven). Laydays/cancelling 7/14 days from index date. Age max 15 yrs. 2.5% total commission.
- *TD8:* 80,000mt crude and/or DPP, heat 135F. Kuwait to Singapore (Mena al Ahmadi to Singapore). Laydays/cancelling 20/25 days from index date. Double hull, age max 15 yrs. 2.5% total commission.
- *TD9:* 70,000mt. Caribbean to US Gulf (Covenas to Corpus Christi). Laydays/cancelling 7/14 days from index date. Age max 15 yrs. Assessment basis Oil Pollution Act premium paid. 2.5% total commission.
- *TD12:* 55,000mt fuel oil. Amsterdam-Rotterdam-Antwerp range to US Gulf (Antwerp to Houston). Laydays/cancelling 15/20 days from index date. Double hull, age max 15 yrs. 2.5% total commission.
- *TD14:* 80,000mt. South East Asia to east coast Australia (Seria to Sydney). Laydays/cancelling 21/25 days from index date. Double hull, age max 15 yrs. 2.5% total commission.
- *TD15:* 260,000mt. West Africa to China (Serpentina FPSO and Bonny Offshore Terminal to Ningbo). Laydays/cancelling 20/30 days from index date. Double hull, age max 15 yrs. 2.5% total commission.
- *TD17:* 100,000mt. Baltic to UK-Cont (Primorsk to Wilhelmshaven), Great Belt laden/ballast. Laydays/cancelling 10/20 days from index date. Double hull, age max 15 yrs. 2.5% total commission.
- *TD18:* 30,000mt fuel oil. Baltic to UK-Cont (Tallinn to Amsterdam). Laydays/cancelling 10/15 days from index date. Double hull, age max 15 yrs. 2.5% total commission.
- *TD19:* 80,000mt. Cross Mediterranean (Ceyhan to Lavera). Laydays/cancelling 10/15 days from index date. Age max 15 yrs. 2.5% total commission.

[1] Time-Charter Equivalent (TCE) assessments for dirty tankers are made up of average dollar pricing conversions for Very Large Crude Carriers (VLCCs), Suezmax and Aframax tankers derived from a range of routes already reported on by the Baltic Exchange and expressed in Worldscale. In addition, TCE assessments for individual routes are also reported. (Baltic Exchange, 2019)

- *TD20:* 130,000mt. West Africa to UK-Continent (offshore terminal Bony) to Rotterdam. Laydays/cancelling 15-20 days from the index date. Age max 15 years. 82,000grt. 2.5% total commission.
- *TD21:* 50,000mt fuel oil, Caribbean to US Gulf (Mamonal to Houston), laydays/cancelling 7/14 days from index date. Double hull, age max 15 yrs. 2.5% total commission. (This route does not contribute to the BDTI calculation)
- *TD22:* 270,000mt. USG/China (Galveston O/S lightering area to Ningbo), laydays/cancelling 25/35 days from Index date. 3.75% total commission. (This route does not contribute to the BDTI calculation)
- *TD23:* 140,000mt. AG/Med (Basrah to Lavera), laydays/cancelling 20/30 days from Index date, 2.5% total commission.
- TD25: 70,000mt. USG/Med (Corpus to Trieste), laydays/cancelling 10/20 days from Index date. 2.5% total commission.

2.6.1.2 Baltic Exchange Clean Tanker Index[2] (BCTI)

Individually traded clean Tanker routes (Baltic Exchange, 2019)

- *TC1:* 75,000mt CPP/naphtha condensate. Middle East Gulf to Japan (Ras Tanura to Yokohama). Laydays/cancelling 30/35 days from index date. Age max 15 yrs. 3.75% total commission.
- *TC2_37:*37,000mt CPP/UNL. Continent to US Atlantic coast (Rotterdam to New York). Laydays/cancelling 10/14 days from index date. Age max 15 yrs. 3.75% total commission.
- *TC5:* 55,000mt CPP/UNL naphtha condensate. Middle East Gulf to Japan (Ras Tanura to Yokohama). Laydays cancelling 30/35 days from index date. Age max 15 yrs. 3.75% total commission.
- *TC6:*30,000mt CPP/UNL. Algeria to European Mediterranean (Skikda to Lavera). Laydays cancelling 7/14 days from index date. Age max 15 yrs. 3.75% total commission.
- *TC8:* 65,000mt CPP/UNL middle distillate. Middle East Gulf to UK-Cont (Jubail to Rotterdam). Laydays/cancelling 20/30 days from index date. Double hull, age max 15 yrs. This route to be reported as US$ per mt. 3.75% total commission.
- *TC9:* 30,000mt CPP/UNL/ULSD middle distillate. Baltic to UK-Cont (Primorsk to Le Havre). Laydays/cancelling 5/10 days from index date. Double hull, age max 15 yrs. 3.75% total commission.
- *TC14:* 38,000mt CPP/UNL/diesel. US Gulf to Continent (Houston to Amsterdam). Laydays/cancelling 6/12 days from index date. Age max 15 yrs. 3.75% total commission.
- *TC15:* 80,000mt naphtha. Med / Far East (Skikda to Chiba). Laydays/cancelling 15/25 days from index date. This route to be reported on a US$ lumpsum basis. Age max 15 yrs. 2.5% total commission.
- *TC16:* 60,000mt CPP. A-R-A / West Africa (Amsterdam to offshore Lome). Laydays/cancelling 10/14 days from index date. Age max 15 years, 2.5% total commission.

[2] Time-Charter Equivalent (TCE) assessments for clean tankers are made up of average dollar pricing conversions for Medium Range (MR) Product tankers derived from a range of routes already reported on by the Baltic Exchange and expressed in Worldscale. In addition, TCE assessments for individual routes are also reported. (Baltic Exchange, 2019)

Baltic Exchange International Tanker Routes[3] – Asia

Individually traded International Tanker routes (Baltic Exchange, 2019)

- *TC7:* 35,000mt CPP. Singapore to east coast Australia (Singapore to Sydney). Laydays/cancelling 17/23 days from index date. Double hull, age max 15 yrs. 3.75% total commission.
- *TC10:* 40,000mt CPP/UNL. South Korea to west coast North Pacific (South Korea to Vancouver – Rosarito range). Laydays/cancelling 14/21 days from index date. Double hull, age max 15 yrs. 3.75% total commission.
- *TC11:* 40,000mt CPP. South Korea to Singapore. Laydays/cancelling 10/17 days from index date. Double hull, age max 15 yrs. 3.75% total commission.
- *TC12:* 35,000mt naphtha condensate. West coast India to Japan (Sikka (Jamnagar) to Chiba). Laydays/cancelling 7/14 days from index date. Double hull, age max 15 yrs. 3.75% total commission.
- *TD24:* 100,000mt. Russian Pacific/China (Kozmino to Qingdao), Laydays/cancelling 10/20 days from Index date. 2.5% total commission.

The difficulty to calculate exact freight rate used for benchmarking the Baltic indices is due to the fact that to accurately calculate results every transaction in the market to be accessed that is impossible in practice. (Baltic Exchange, 2019)

2.6.2 Baltic clean tanker index

BCTI is published by the Baltic exchange every Monday to Friday at 13:00 UTC and is an important price index for product liquid bulk cargo shipping worldwide. This index is determined by the Baltic exchange from standard information provided by participant and panellists.

Oil product cargo demand depends on Gross Domestic Product growth, technological progress, and fuel oil price and also demand from new markets such as the Republic of China and India. Freight rate depends on the available cargo space index. Freight rates are determined from information from shipbrokers shipowners and charterers. Since real demand and real supply on the routes are affecting BCTI, manipulation is not possible by the method of calculation of BCTI (Baltic Exchange, 2019).

2.7 Economy derivers and important indicators influencing BCTI

Many indices are influencing BCTI. They can be categorized into industry, economy, oil demand and supply, vessel demand and supply and routes and operation costs. Based on explanations of this thesis, at the end of this chapter, one can conclude that these are the most important drivers of BCTI. Some of the most important indices for this thesis are explained later

[3] The TCEs are calculated using a variable feed of bunker prices supplied by Bunkerworld. Variable exchange rates supplied under license by XE.com are also used for adjusting port costs. Port charges used in this calculation are provided under license by Cory Brothers. (Baltic Exchange, 2019)

in this chapter. From these indices, those that are used later in chapter 5 for statistical analysis are shown in Figure 2.5.

2.7.1 Global oil productions

This index is a general indicator of the amount of oil produced during a certain period of the time. This index is calculated monthly. It is a good indication of the trends in industry and economy since it can be an indication of the energy and initial material consumption.

2.7.2 Crude oil price

"*Oil Price refers to the spot price of one barrel of the benchmark crude oil. The price depends upon its grade, location and the content of sulfur present in it. The price of oil can be determined with the help of balance between its demand and supply."* (Petropedia, 2019). As the definition of oil price shows crude oil price is location-related and defined by demand and supply from the market since this price directly affects energy market and therefore transport market is highly relevant to The Baltic exchange indices. Below are the main market and indicators of crude oil price.

2.7.2.1 Brent crude

Brent crude is a class of light and sweet crude oil and is leading price benchmark for the area of Atlantic basin crude oil. It is light because of its relatively low density and sweet since it has lower Sulphur content in comparison with other major crude oil supplies.

2.7.2.2 OPEC

Organization of the petroleum exporting countries (OPEC) is consisting of the 14 countries of major oil-exporting nations. OPEC is holding approx. 82% of the world oil reserve. (Investopedia, 2019). OPEC traditionally has a great influence over the oil price and is the largest policymaker in oil and natural gas market.

2.7.2.3 WTI (West Texas Intermediate)

WTI crude oil is another important benchmark of oil. WTI is produced and consumed in North America and therefore has less effect over the freight rates but it is still a main indicator of oil price. (Investopedia, 2019)

2.7.3 World fleet carriage capacity

According to UNCTAD world tanker capacity has been increasing during the last decade. (UNCTAD, 2019). It can be an indicator of the global economy growth. However, fleet capacity growth is related to few factors one of them is new building orders the other is scarping of the old ships and both of above are mainly related to the freight rates and capacity utilization of the current fleet. Formula 2.1 explains how this capacity is calculated. It is the total capacity of the world fleet plus new orders mines capacity demolished mines capacity lost due to accidents.

Equation 2.1

$$\sum \mathit{Tanker\ Carriage\ Capacity} = \sum \mathit{Present\ capacity} + \sum \mathit{New\ orders\ capacity} - \sum \mathit{Demolition\ capacity} - \sum \mathit{Loses\ \ capacity}$$

2.7.3.1 Capacity utilization

Total utilization of shipping capacity plays a critical rule on freight rates. Excessive unused capacity of the fleet enforces lower freight rates. In some routes, incoming freight rate is higher than out-going freight rates for the vessels. Routes of North American fluid bulk are some of these routes since incoming demand is not equal to outgoing demand for capacity. This reason enforces different incoming and outgoing freight rates on the same port or area.

Also, world capacity utilization has a direct effect on freight rates. Higher demands on a relatively short time cause a surge in freight rates. Also, the oversupply of capacity lowers demand and therefore causes drop of the freight rates.

2.7.3.2 Ship building orders

New orders normally show expectation of the future market. Shipowners order new ships when demand for the capacity is high or when they expect future demand will be high. Also, to replace the old fleet and reduce OPEX costs new orders are placed. New orders can be an indication of what shipowners are expecting from the market and new capacity will directly influence freight rates.

2.7.3.3 Ship scraping

The scrapping of the old fleet is a way of balancing the market capacity. Total carriage capacity of the world fleet is a function of scarped capacity since present capacity is equal to present capacity plus new orders minus scarped capacity minus loses.

2.7.4 Industrial production

Industrial production is the number of products produced at a particular time by industrial activity. It includes energy; therefore, industrial production can drive the shipping industry by the number of products or amount of the energy that is required to be transferred by the sea. Industrial production indicators and indices analyzes can show the future trend for the shipping industry as well.

2.7.4.1 Industrial production OECD

The Organization for Economic Co-operation and Development or OECD is an organization consisting of 37 countries engaging in intergovernmental economic activities. It is a forum of countries describing themselves as committed to democracy and the market economy. OECD is providing a platform to compare policy experiences, seek answers to common problems, identify good practices and coordinate domestic and international policies of its members. (OECD.org, 2019). OECD countries and key partners represent about 80% of world trade and investment. OECD industrial production index shows the production of 37 countries that among them most of the largest economic power such as Japan exist. Therefore, this thesis selects this index also to test it against BCTI and its possibility to be used as one of the variables of the models.

2.7.4.2 US

The US alone produces 24.4% of world GDP (Mark J. Perry, 2018) that is enough to be considered as a main leading country influencing shipping industry. US government decision regarding industrial developments and import/export regulations have a huge impact on the future of the shipping industry. Hence the US Industrial production index is one of the drivers of the shipping industry that can be used to forecast the future of the freight rates.

2.7.4.3 China

China produces 19% of the world GDP and therefore, as mentioned earlier about US industrial index, can be used for forecasting the future of the shipping market demand and freight rates as well.

2.7.5 Dow jones 30

The Dow Jones Industrial Average (DJIA), or simply the Dow, is a stock market index that indicates the value of 30 large, publicly owned companies based in the United States, and how they have traded in the stock market during various periods. This index is not a weighted average of the value of those companies but it is rather an average of the share prices of those companies and therefore is not an indication of the value of those companies or an indication of US market or total market.

This index has been selected for this thesis since the change of this index is a good indicator of the expected market trends. Dow is an independent index from BCTI and this is another reason to select it for development of models.

2.7.6 Bunker prices

Bunker price directly influences freight price. Since bunker cost is the most important cost of every voyage bunker price shall be considered for the forecast of the future freight rates. Main bunker type used on board of the ships is HFO 380. This type of fuel is heavy fuel oil with very high viscosity and a high octane number.

Vessels are mostly receiving bunker during their course of the voyage and in locations that are close to their routes. These locations are not many and three most famous locations are Fujairah (UAE), Singapore Eastern Ground (Singapore), Rotterdam (Netherlands). Bunker prices are different from location to location but the difference is not very much.

This thesis uses average HFO 380 price in these three locations as bunker price when it is necessary.

2.7.7 USD index

The international value of the US Dollar is measured by USD index. This index measures the value of the USD against the top 6 currencies in a basket. EUR, JPY, GBP, CHF, CAD and SEK are currencies making the basket to measure the relative value of the USD. this indicator is important since this can show the flow of import and export to the US that is very important for supply and demand for shipping freight.

2.7.8 Gold price

Gold traditionally is a medium to preserve the value. This gives a unique characteristic to this commodity since it has shown over ages that it is a reliable place to store value. When sudden changes in economy or market condition are expected gold is a relatively safe investment. Gold price is a good indicator of the general condition of the world economy. as demand for this metal raises the price also rise and this shows uncertainty in other markets and global economy. Therefore, it is an independent indicator from BCTI that can be used for further analysis of BCTI in this thesis.

2.7.9 Exchange rates Euro index

“The nominal effective exchange rate (NEER) of the euro is a weighted average of nominal bilateral rates between the euro and a basket of foreign currencies. It is an indicator of the external values of the euro vis-à-vis the currencies of selected euro area’s trading partners”. (European Central Bank, 2019)

Euro exchange rates are an indicator of EC monitory and economy policy and condition. Since Europe is a big player in the oil products market this index can be a potential indicator of the clean tanker condition. This index also will be analyzed later to test the possibility of use of this indicator for modelling.

2.7.10 Interest rates

Interest rates are the cost of borrowing money and represent what creditors earn for lending money. Monetary Policy and Interest Rates will drive the economy and market. Major economies such as Japan, US and Europe interest rates detect future expectation of the market. Also, it shows if these governments are following an expansionary and contractionary monetary policy and also shows aggregate demands. Interest rates also have an impact on inflation rates. Major economies interest rates will be tested against BCTI later in chapter 5 of this thesis to detect if they are correlated with BCTI and if they are suitable variable for models in chapter 5.

2.7.11 ClarkSea index

This index is defined as “A weighted average index of earnings for the main vessel types where the weighting is based on the number of vessels in each fleet sector.” (Sin.Clarksons.net, 2019). This index is very similar to Baltic indices but calculated by Clacksons Researches.

2.7.12 LIBOR

London Inter-Bank Offred Rate is an index of interest rate that is used to calculate most of the loans. Shipping companies also using the loan for many of their activities and also for new orders. LIBOR interest rate affects the Capex of shipping companies. LIBOR also is used vastly in industry and economy for interest rate calculations. LIBOR, therefore, is an index that can influence future freight rates.

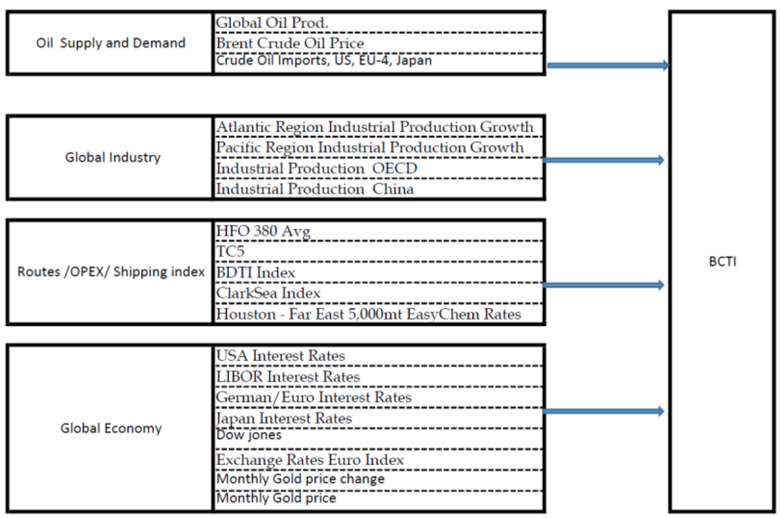


Figure 2.5 BCTI determination factors discussed in this thesis and their categories. Source: Author

2.8 Quantitative & Forecasting methods

Forecasting methods are categorizing in two major categories of the methods, quantitative and qualitative methods. Qualitative methods are mostly used for forecasting the data that has no historical records and also for long-term decisions. Delphi method, market research and historical life-cycle analogy are some of them. This thesis focus is on quantitative use of time series to predict the Baltic Clean Tanker Index. Quantitative forecasting models are used to forecast future data by analyzing and modelling the past data. Their approach is by analyzing the past data a pattern or series of the pattern are found and then predict future figures or rates based on the past data patterns. It needs a reasonable past numerical data to be accurate enough to forecast future outcome. Based on nature and length and volatility of data different methods may be chosen for forecasting. below are the most common types of forecasting methods.

2.8.1 Average approach

This approach future value is predicted by calculating the mean of all the previous values and putting the mean as the predicted value. This approach is common when previous values are available. This method is used when certain data are not available on a series of data and to produce those missing fields this method may be used.

2.8.2 Time series methods

2.8.2.1 Moving average

Is the average of the last n number of known data in a time series. it is normally used to smooth the original data series and make the trend clearer. the use of this method is very common in finance and stock price predictions.

As Anderson et al (2012) stated in their book the moving averages method uses the average of the most recent k data values in the time series as the forecast for the next period. A moving average forecast of order k is defined mathematically as follows:

Equation 2.2

$$F_{t+1} = \frac{\sum(most\ recent\ k\ data\ values)}{k} = \frac{Y_1 + Y_{t-1} + \cdots + Y_{t-k+1}}{k}$$

Where

F_{t+1} =forecast of the time series for period t + 1

The term moving average is used because every time a new observation becomes available for the time series, it replaces the oldest observation in the equation and a new average is computed. As a result, the average will move, as new observations become available. (David R. Anderson, 2012)

2.8.2.2 Weighted moving average

In the moving averages method, each observation in the moving average calculation receives the same weight. the weighted moving average is an average that gives different weight to different data in a time series depending on their position in the data series. In an n day moving average the last day has weight of n and the one before that has weight of n-1 down to 1. (David R. Anderson, 2012)

2.8.2.3 Exponential smoothing

In moving average past observations are weighted equally but exponential smoothing is a technique for smoothing time series data giving previous data an exponential decreasing weight over time. This is used normally to remove noise from high frequency data.

Equation 2.3

$$F_{t+1}= aY_t+(1-a)\ F_t$$

Where

F_{t+1}=forecast of time series for period t+1

Y_t=actual value of the time series in period t

F_t=forecast of the time series for period t

a =smoothing constant ($0 \le a \le 1$)

the exponential smoothing forecast for any period is actually a weighted average of *all the previous actual values* of the time series. (David R. Anderson, 2012)

2.8.2.4 Autoregressive Moving Average (ARMA)

ARMA model is a tool for predicting and understanding of future data in a time series. Auto Regression part (AR) is used for regression of the data on its own lagged past data and Moving Average (MA)part is to model errors at various time in past. The model is referred to as ARMA (p,q). p refers to the order of AR and q refer to the order of MA part. The equation is given by:

Equation 2.4

$$X_t = c + \varepsilon_t + \sum_{i=1}^{p} \varphi_i X_{t-i} + \sum_{i=1}^{q} \theta_i \varepsilon_{t-i}$$

Where

φ= the autoregressive model's parameters

θ= the moving average model's parameters

c= a constant

ε= error terms

(Petris, 2009)

2.8.2.5 Autoregressive integrated moving average (ARIMA)

Auto Regression Integrated Moving Average (ARIMA) is a generalization of ARMA model. when data are non-stationary ARIMA model is used. ARMA and ARIMA are apart because of differencing. An ARMA model is a stationary model; If your model isn't stationary, then you can achieve stationarity by taking a series of differences. The "I" in the ARIMA model stands for integrated; It is a measure of how many nonseasonal differences are needed to achieve stationarity. If no differencing is involved in the model, then it becomes simply an ARMA (Petris, 2009) Another similar model is ARIMAX, which is just an ARIMA with additional explanatory variables.

2.8.2.6 Extrapolation

Extrapolation in mathematics is where the future value of a variable is calculated by its relation to known previous data beyond the original observation. It is similar to interpolation but risk and uncertainty are more in extrapolation.

2.8.2.7 Linear prediction

LP is a mathematical operation where future values are estimated as a linear function of previous samples.

2.8.2.8 Trend estimation

This method is used to describe the behaviour of observed data without explaining it. This method is mostly used to estimate the tendency of the data in a time series by relating the data to the time at which they occurred. It distinguishes the behaviour of data from random behaviour of the data and detects decreasing or increasing tendency of the data.

2.8.3 Casual economic forecasting method

Sometimes it is possible to forecast a variable by observing changes in another variable. For example, by observing weather data we can predict the sale of umbrella since on rainy days it is more probable to sell umbrella. Some forecasts examine past relation between variables. If a variable has a linear relation to the other variable, extrapolation of this relation in future is an option.

2.8.4 Artificial intelligence methods

2.8.4.1 Artificial neural network

The artificial neural network is developed to simulate the function of the biological model by taking inputs and process them and use the output as an input for another network to finally get desired results as output. ANN is a network of artificial neurons that receives input, change their state according to input this process is called activation. Then produce output depending on the input and activation.

The process of activation can be modified by a process called learning which is governed by learning rules.

2.8.4.2 Group method of data handling

GMDH is a family of inductive algorithms for mathematical models which are computer-based for multi-parametric datasets. GMDH is used in such fields as data mining, knowledge discovery, prediction, complex systems modelling, optimization and pattern recognition.

2.8.4.3 Machine learning

MA Is a subset of artificial intelligence (AI) closely related to computational statistics using computers to make predictions. It uses algorithms and statistical models to perform a specific task effectively without using instructions, relying on patterns and inference instead. In order to make predictions or decisions without being programmed for the task ML algorithms build a mathematical model based on sample data known as training data. Machine learning algorithms are used in a wide variety of applications, such as email filtering, and computer vision, where it is infeasible to develop an algorithm of specific instructions for performing the task.

2.8.5 Forecasting accuracy

Forecasting accuracy used to test predicted data. It is the difference between the actual value and the forecast value for the corresponding period.

Mean Absolute Percent Error (MAPE)

Statistically, MAPE is defined as the average of percentage errors. The MAPE formula consists of two parts M and APE. A_t is the actual value and F_t is the forecast value. mean of all percentage errors for the given period of time can be calculated by using formula (2.5). The formula for MAPE is as follow

Equation 2.5

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

Mean Absolute Deviation or MAD

Another common way to find forecast error is to calculate the Mean Absolute Deviation (MAD). Mean absolute deviation (MAD) of a data set is the average distance between each data value and the mean of that data set. Mean absolute deviation used to describe variation in a data set. MAD helps to get a sense of how "spread out" the values in a data set are. (KhanAcademy.org, 2019)

2.9 Conclusion

Now the reader of this thesis is more familiar with the liquid bulk market, parts of the global industry and economy that may affect this market and some of the methods used to predict future data in time series. In previous parts of this paper different routes, also different types of oil tankers and its categories explained to the reader. The Baltic exchange tanker indices and routes these indices are calculated based on contracts made on them are also known now. Factors changing BCTI were categorized and explained in this chapter to make a reader ready to easier understand the literature review. Previous work which is done in this field will be discussed in Chapter 3 of this thesis to better connect above explained with their actual use to make a model for forecasting of the BCTI.

3 Literature Review

3.1 Introduction

This chapter presents previous works and papers related to this thesis topic i.e. literature. This literature review on the first part explain and examine data that influence Baltic Clean Tanker Index (BCTI) and oil tanker chartering market and later presents various modelling/ regression attempts that have been done to examine and explain the connection of different market indicators and variables to Baltic Clean Tanker Index (BCTI).

3.2 Data selection and Modeling methods

The oil shipping market is very important in terms of its importance for economy transport and industry. Its indicators such as Baltic dirty tanker index and Baltic clean tanker indexes are daily monitored by economists and investors. Baltic Clean Tanker index defined as the daily weighted average freight price to ship oil and products across the world. Therefore, this index incorporates information about future economic activity and thus have potentials of being one of the important leading economic indicators. Prior knowledge of relationships and importance of these relationships with trends of the market and economy affect expectations from the shipping market and therefore, chartering out and chartering in prices. These relationships also define general risk factors of the market for the investors and ship owners. Various studies have been done over the time to study relations and influence of main economic indicators to BDI and BDTI and BCTI.

As the study of Yuying Yang et al (2015) shows the oil tanker market can be regarded as an extension to the oil market since oil price has a direct effect over the tanker market indexes. The volatility of the oil price and tanker freight makes it a challenging and difficult business for the oil companies (Yuying Yang, 2015). Yuying Yang et al (2015) paper reveals an interesting fact that not all oil markets have the same effect on tanker freight prices. VAR-BEKK-GARCH method was used to test the spillover effect from the crude oil market to the tanker shipping market. Brent oil market influence is stronger than WTI influence over oil tanker shipping market. That could be due to their different role on the international oil markets. Oil tanker shipping market is more influenced by oil market shocks. It is most probably due to the difficulty of reallocating the ships due to high demand for oil and therefore it will be more difficult to reach an equilibrium between demand for ships and supply of oil during the shock periods (Yuying Yang, 2015)

Baltic dirty tanker exchange is an index of high importance in the dirty tanker shipping industry of the world (Shuangrui Fan, 2013). Shuangrui fan et al (2013) used Wavelet Neural Networks (WNN) method to forecast BDTI. They used a hybrid AI method for their forecast. By using BDTI calculation formula. Then determinants of BDTI identified then six factors were chosen to represent externalities and applied in the WNN model. To determine the factors, they first divided internal factors that affect BDTI then they extend them to the external factors which impact internal factors, then, variables were selected based on external determinants ((Shuangrui Fan, 2013). Final results showed that WNN offers a good prediction of long term future trends that is valuable information for many sectors of the shipping industry such as business negotiations decision making and financial budgeting (Shuangrui Fan, 2013).

Another attempt to forecast freight rates in tanker shipping market is done by Xiaolei et al (2014). Crude oil price is used to predict freight rates (Xiaolei Sun, 2014) .IMF method was used to analysis data. Results of modelling showed that BDTI and oil price exhibit different multiscale properties, it means that different patterns are identified in long term trend and short-term trends. (Xiaolei Sun, 2014).

The relation between FFA rates and spot rates studied by Batchelor et al (2007). They examined if the forward market is liquid enough to embody future spot rate into the prices, they should observe cointegration between spot and forward rates. They found this cointegration by their research. But according to the results forward, rates were adjusted to fill the gap between forward and spot rate. They used time series modelling to get these results. (Roy Batchelor, 2007)

Previously a set of variables was used by researchers to determine the demand for shipping services. The economic variables such as – West Texas intermediate spot crude oil price -S&P GSCI agriculture index – FTSE ST Maritime index - One-month LIBOR interbank – Dirty Tanker DT3 route (Goulas, 2010). Goulas used 11 variables to forecast freight rates. These data were used to forecast spot rates and future rates and Baltic indices. According to his findings, strong predictable patterns are found on spot freight rates and all Baltic indices however forecasting results for future rates were not as successful. (Goulas, 2010).

To analyze data the first thing to do is to analyze them statically and prepare them for later modelling. In statistics, linear regression is a linear approach to model relation between a dependent variable and one or more explanatory or independent variables. When there is one explanatory variable it is called simple regression and in case of multiple explanatory variables, it is called multiple linear regression. Linear regression analyses have been used vastly in many areas of economics and science including its usage in the shipping industry. (Chantzaras, 2018) .

Multi-Variable Linear Regression is one of the methods that has been used to calculate occupational risk in the shipping industry. MVLR used to define a model to calculate occupational risk from occupational accident data. By comparing predicted values with actual data, the effectiveness of the model was tested and it concluded that its predictions are reliable and acceptable and can be used to efficiently predict occupational risks. (Vasilios Tsoukalas, 2016)

In statistics, signal processing and economics Auto Regression model is used to represent certain types of random time-varying processes in nature and economics. Autoregressive model defines that output variable is linearly dependent on its previous values. Model is in the form of stochastic (imperfectly predictable) difference equation. together with Moving Average (MA), it forms another type of model ARMA and ARIMA. In a recent paper ARMA and VAR along with two other models used for prediction of the spot and FFA freight markets. VAR model performance in the short term for the spot market rates is more accurate in comparison with other more complicated methods. ARIMA model, however, presents more accurate results when used to forecast FFA rates. Batchelor et al suggested ARIMA and VAR model as better models for forecast (Roy Batchelor, 2007).

Shaunagrui fan et al chooses to model their work by Wavelet Neural Network WNN. The concept of WNN is based on wavelet transformation theory. The main feature of WNN is that it combines the time-frequency property with learning nature of the neural network. (Shuangrui Fan, 2013).

They also used ARIMA model by using stationary data. they used Moving Average to smooth short-term fluctuations. No long-term trend, however, has been found. ARMIA model used as a reference model. ARIMA model accuracy is acceptable for 40 days trend and is within 8% error but when used for longer periods of 60- and 120-days errors are significant and not acceptable. Shaunagrui fan et al found the ARIMA model is not a suitable model for prediction of BDTI for medium- and long-term trend. They used 6 variables that may externally affect the BDTI and one day delay for BDTI as input and BDTI value used as output. WNN model shows superiority over the ARIMA model in mid-term and long-term forecasts however when it is more than 120 days model is not able to deliver acceptable results (Shuangrui Fan, 2013). Since WNN can be effective in forecasting non-linear non-stationary data it can be used as a relevant tool for other sectors of the shipping business.

Goulas et al used models to forecast freight rates. VAR, ARMA, GHARCH-family models and Economic Variable Models were used to obtain results. (Goulas, 2010). The economic variable model tries to forecast using certain economic variables. Goulas et al used a Beenstock and Vergotties framework to construct their model. This framework is a supply and demand model. generalized autoregressive conditional heteroskedasticity (GARCH) [4] model is commonly employed in modelling financial time series that exhibit time-varying volatility (Chris Brooks, 2014). Since Goulas et al used their model to predict short term output, in general, all selected models returned acceptable results within the chosen period.

All the above can lead us to the conclusion that the model is different from the system. Since the system is the reality and produces the real data however the model tries to reproduce the same data and explain the output data (Christian Lyzell, 2009).

Chantazaras (2018) used Subspace State Space System IDentification (N4SID) algorithm to define three models for forecasting BDI. He tested his models against each other and found a model with fewer variables presented more accurate results and his models also predicted sudden changes in trend just before it was going to happen. Fewer variables in the model will give more flexibility to the model resulting in faster response to the change of market condition (Chantzaras, 2018).

Sahin et al by using Artificial Neural Network attempt to forecast BDI. First, they used ANN to forecast BDI then they added COP as an input parameter. Three models then constructed by this method and each model tested against the other two models. The volatility of oil price showed a considerable impact over the results. They concluded that it is generally hard to forecast the BDI because it is volatile, complex, and cyclic (Sahin, 2017).

Batrinca et al used two ARIMA models for forecasting BDI. The statistical output of the models then checked by analyzing the residuals and they concluded that the ARIMA model returns acceptable results for forecasting and prediction of the BDI (Batrinca, 2013).

Papailias et al (2017) used statistical analysis to determine related commodity and indexes to forecast BDI. They investigate cyclical properties of the Baltic Dry Index (BDI) and their implications for forecasting performance. Later they showed changes in the BDI can lead to permanent shocks to trade of major exporting economies. They showed that commodities and trigonometric regression can lead to improved predictions and they can be used for improved risk management in the freight sector. (Papailias, 2017).

[4] If an autoregressive moving average model (ARMA) model is assumed for the error variance, the model is a generalized autoregressive conditional heteroskedasticity (GARCH), model.

Chen & Dyke (2015) used MACD, RSI, MFI and ATR for prediction of a stock price. Their idea is that the stock price dynamics is an unknown stochastic dynamic system to be identified. The stock price then treated as the system output and the technique analysis data such as MACD, RSI, MFI and ATR are treated as the system inputs. They used system identification techniques Extended Least Squares (ELS) method identify the system parameters (Chen, 2015).

3.3 Conclusion

In chapter 2 an overview of tanker ships and its related markets and economy derivers presented and later in chapter 3 some of the previous works that have been done which are related to the topic of this thesis was explained .By examining purposed papers for data and models used to predict various indices related to the shipping and Baltic indices we can conclude some various variables that are not dependent on the BCTI. Some of these identified variables are such as LIBOR and other interest rates, Gold price, BRENT crude oil price, OPEC oil price, Industrial Production of China, indexes related to transport and EUR and USD rates. These data will be examined for their dependencies against BCTI in Chapter 5 of this thesis.

In chapter 3.2 models and methods of prediction of Baltic indices was reviewed. Most of the methods used were based on regression analysis. Due to the concentration of regression methods to predict the Baltic indices and BCTI and lack of bibliography to identify the model of BCTI with any of existing methods, Numerical algorithm for Subspace State Space System Identification (N4SID) will be used to determine the model that fits best to the BCTI from an input-output dataset. This method is a relatively new method and has shown its capacity for modelling in other sectors such as signal processing and other input-output systems.

4 Methodology

4.1 Introduction

To answer the main question of this thesis a series of statistical and modeling analysis has to be done. This chapter will provide information on where and how to collect the data and how to perform all statistical analysis and modeling. This thesis will provide information on collection of the data on Ch.4.2 later on Ch.4.3 statistical analysis methods used in this thesis will be explained and correlation of the data with BCTI and with other data will be examined and main statistical characteristics of this data will be explained. On Ch.4.4 the methodological approach to the modeling of BCTI is explained.

4.2 Data collection

For the purpose of this thesis, secondary data that are already collected by research and studies will be used. Reliable data sources will be used to obtain the required data for this thesis. These data sources are including (but not limited to) Sin.Clarcksons.net, WorldBank.

- Sin.Clarcksons.net: Shipping related data such as BDI, BCTI, BDTI will be obtained only from this database. The advantage of this database is reliable and due to that, their data is vastly used by the experts in the shipping industry (Sin.Clarksons.net, 2019). Another advantage is that data presented by them is up to date and the most recent data. The disadvantage of this data is that frequency of observation of the data is sometimes a long duration such as one month that causes fewer observations during a fixed period of data. This database contains all the data that is required by this thesis. The database will be used for the following data. Global Oil Production, Brent Crude Oil Price, Atlantic Region Industrial Production Growth, Pacific Region Industrial Production Growth, Industrial Production OECD, RouteTC5 Rates, USA Interest Rates, LIBOR Interest Rates, BDTI Index, ClarkSea Index, HFO 380 Avg, Houston - Far East 5,000mt EasyChem Rates, Industrial Production China, German/Euro Interest Rates, Japan Interest Rates, Crude Oil Imports, US, EU-4, Japan, Dow Jones index, Exchange Rates Euro Index and BCTI.

- WorldBank: Growth factors will be extracted from this database in addition to the same data from Clarkson for comparison. Advantage of data from world bank is that they are free and reliable since most of them are directly provided by governments. The disadvantage is that some data are outdated and yearly published

4.3 Statistical analysis

Statistics is “the science of using information discovered from collecting, organizing, and studying numbers” (Cambrige Dictionary, 2019)it is also defined as “a way to get information from data” (Keller, 2014) In general, statistics is a branch of mathematics working with collection, correction, organization, analysis and interpretation of data to present or solve a scientific or industrial or social problem. Statistics deals with every aspect of data. These data obtained from a sample of a population or a whole population. A variable is some characteristic of a population or sample. (Keller, 2014), the value of the variables is a possible observation of the variables, data are observed values of the variables (Keller, 2014). Interval data are real numbers such as weight, value, distance these data are sometimes referred to as quantitative or numerical data. When numerical data presented over a timeline, they are time-series data. This thesis deal with time series to model BCTI.

Statistics is divided in two basic areas Descriptive statistics and inferential statistics (Keller, 2014) , descriptive statistics involves arranging, summarizing and presentation of a set of data in such a way that useful information is produced. Its methods make use of graphical technics and numerical descriptive measures to summarize and present the data. (Keller, 2014).

Following subsections are based on random variable vector[5] A_k (k=1, 2, ..., K) made up of N scalar observation in a matrix form (Chantzaras, 2018).

Equation 4.1

$$A_k = \begin{pmatrix} A_{1k} \\ A_{2k} \\ \vdots \\ A_{Nk} \end{pmatrix}$$

Where N is number of observations (i.e. the sample size is N).

And where all variable sectors A_k together can form an external observation matrix A, as:

Equation 4.2

$$A = [\, A_1, A_2, A_3, \ldots, A_K]$$

(Chantzaras, 2018)

4.3.1 Measures of mean values and standard deviations

Average or mean value

[5] **random vector** is a list of mathematical **variables** each of whose value is unknown, either because the value has not yet occurred or because there is imperfect knowledge of its value.

To describe the center of the data there are three different measures Mean or arithmetic mean that is average of the data, median and mode. The mean is the most common measure of center of numerical data computed summing of all the observations and dividing by number of observations.

For a random variable vector A_k made up of N scalar observations, the mean (μ) value is mathematically calculated as:

Equation 4.3

$$\mu_k = \frac{\sum_{i=1}^{N} A_{ik}}{N}$$

Variance and standard deviation

The variance and its related measure, the standard deviation, are the most important statistics (Keller, 2014). They are used to measure variability since they play a vital role in all statistical inference procedures (Keller, 2014). Population Variance is calculated as:

Equation 4.4

$$\sigma^2 = \frac{\sum_{i=1}^{N} (x_i - \mu)^2}{N}$$

And from that standard deviation of the population is calculated as (MathWorks, 2019):

Equation 4.5

$$\sigma = \sqrt{\sigma^2} = \sqrt{\frac{\sum_{i=1}^{N} (x_i - \mu)^2}{N}}$$

For a random variable vector A_k made up of N scalar observations, the standard deviation (S_k) is defined as:

Equation 4.6

$$S_k = \sqrt{\frac{1}{N-1}\sum_{i=1}^{N} |A_{ik} - \mu_k|^2}$$

Where μ_k is mean of variable A_k (MathWorks, 2019). Standard deviation is used to measure variations. mean is used to measure the center of the data. Standard deviation is the square root of the variance (S_k^2) and always a positive number.

4.3.2 Correlation analysis

Correlation coefficient

Correlation is used to test relations between quantitative or categorical variables. The study of how variables are correlated is called correlation analysis. Correlations are useful since they show how related are variables and based on that, it can be used for the future forecast of behaviour of the variables.

Scatter graphs are used to visually determine the relationship between variables. The Value of one variable will be on x-axis and the value of the other variable will appear on the y-axis. Plotting of two variables makes a pattern that can reveal the correlation between two variables. The correlation measured by a number between +1 and -1 where +- 1 indicates a perfect degree of association between two variables. And zero shows no relation between the two variables. A negative sing shows a negative relation between variables and a positive sign shows a positive relation between variables. (Keller, 2014) If the two variables are negatively correlated when one variable increases the other one will decrease. If the correlation between two variables is positive when one variable increases the other variable will increase as well.

Generally, the closeness of the points in a scatterplot shows the degree of correlation. If the form of the relation between two variables is linear then the plot shows a straight-line pattern, cured and cluster patterns are also other types of patterns. Usually, when data are presented, they are grouped by frequency or relative frequency of each class interval. Grouped data are presented by histograms.

To make inference from results of the correlation coefficient, results must be carefully tested against other results, for example, a plot of the monthly deaths from heart disease against the monthly sale of ice-cream would show a negative association. Here a third factor that is environmental temperature which is positively related to the sale of ice-cream. This example shows that when interpreting the results special attention shall be made to other factors affecting the outcome of the analysis.

"*The correlation coefficient is a statistical measure that calculates the strength of the relationship between the relative movements of two variables."* (Ganti, 2019). Scatterplot shows the visual correlation of the variables. Numerical relation of variables is calculated by population coefficient correlation for a whole population as below (Keller, 2014):

Equation 4.7

$$\rho = \frac{\sigma_{xy}}{\sigma_x \sigma_y}$$

Where:

$$\sigma_{xy=} \frac{\sum_{i=1}^{N}(x_i - \mu_x)(y_i - \mu_y)}{N}$$

and sample coefficient correlation is calculated by below formula (Keller, 2014):

Equation 4.8

$$r = \frac{S_{xy}}{\sqrt{S_{xx}}\sqrt{S_{yy}}}$$

Where:

$$S_{xx} = \sum_{i=1}^{N}(A_{ik} - \mu_k)^2$$

$$S_{yy} = \sum_{i=1}^{N}(A_{ij} - \mu_j)^2$$

$$S_{xy} = \sum_{i=1}^{N}(A_{ij} - \mu_j)(A_{ik} - \mu_k)$$

4.4 Modeling

This section provides theoretical fundamentals regarding the procedures that have to be followed for the model identification. A set of input-output data will be used for the identification of the models[6].

System identification is a methodology for building mathematical models of dynamic systems using measurements of the system's input and output signals. The process of system identification requires that:

[6] Chapter 4.4.1 will be the mathematical procedure of identification. This chapter is as been published by the Favoreel et al. (2000).

- Measuring the input and output signals from your system in time or frequency domain.

- Selecting a model structure.

- Applying an estimation method to estimate value for the adjustable parameters in the candidate model structure.

- Evaluate the estimated model to see if the model is adequate for your application needs.

The main objective of the subspace identification methods is to estimate a linear time invariant state space model, directly from data obtained from input-output measurements. Their application to estimate model parameters of industrial processes became recently highly developed. The first main implemented algorithm used is the N4SID (Hachicha, 2014).

4.4.1 Numerical algorithm for subspace state space system Identification (N4SID)

The N4SID is a type of linear system identification algorithm as described by Favoreel et al. (2000). Below mathematical procedure of identification is as been published by the Favoreel et al. (2000).

Favoreel et al. (2000). state that Linear subspace identification methods are concerned with systems and models of following form (state space model of linear system):

Equation 4.9

$$x_{k+1} = A_{x_k} + B_{u_k} + w_k$$

$$y_k = C_{x_k} + D_{u_k} + v_k$$

With

Equation 4.10

$$\boldsymbol{E}\left[\begin{pmatrix} w_p \\ v_p \end{pmatrix} \begin{pmatrix} w_q^T & v_q^T \end{pmatrix}\right] = \begin{pmatrix} Q & S \\ S^T & R \end{pmatrix} \delta_{pq} \geq 0$$

The vector $u_k \in \mathbb{R}^{m \times 1}$ and $y_k \in \mathbb{R}^{l \times 1}$ are the measurements at time instant k of the m inputs and l outputs of the process, whereas E shows expectations. The vector $x_k$is the state of the process at discrete instant k, $v_k \in \mathbb{R}^{l \times 1}$ and $w_k \in \mathbb{R}^{n \times 1}$ are unobserved vector signals, v_k is called measurement noise and w_k process noise accordingly. It is assumed that they are zero mean, stationary white noise vector sequences and uncorrelated with inputs u_k. $A \in \mathbb{R}^{n \times n}$, $B \in \mathbb{R}^{n \times m}$, $C \in \mathbb{R}^{l \times n}$, $D \in \mathbb{R}^{l \times m}$ are the matrices as described above, while $Q \in \mathbb{R}^{n \times n}$, $S \in \mathbb{R}^{n \times l}$, and $R \in \mathbb{R}^{l \times l}$ are the covariance matrices of the noise vectors $\mathrm{w_k}$ and v_k.

In subspace identification it is typically assumed that the number of available data points goes to infinity and that the data is ergodic. In order to state the problem treated it is assumed a large number of measurements of the input u_k and the input y_k generated by unknown system, the system matrices A,B,C,D up to within a similarity transformation and estimate of the matrices Q,S, R (Favoreel W., 2000)

Subspace identification algorithms consist of two steps. The first step makes a projection of certain subspaces generated from the data, to find an estimate of the extended observability matrix and/or an estimate of the states of the unknown system. The second step then retrieves the system matrices from either this extended observability matrix or the estimated states.

The following input-output matrix equation 4.11, played an important role in the development of subspace identification:

Equation 4.11

$$Y_f = \Gamma_i X_i + H_i^d M_f + N_f$$

Where

- The extended observation matrix Γ_i

Equation 4.12

$$\Gamma_i \stackrel{\text{def}}{=} \begin{pmatrix} C \\ CA \\ CA^2 \\ \vdots \\ CA^{i-1} \end{pmatrix}$$

- The deterministic lower block triangular Toeplitz matrix H_i^d

Equation 4.13

$$H_i^d \stackrel{\text{def}}{=} \begin{pmatrix} D & 0 & 0 & \dots & 0 \\ CB & D & 0 & \dots & 0 \\ CAB & CB & D & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ CA^{i-2}B & CA^{i-3}B & CA^{i-4}B & \dots & D \end{pmatrix}$$

The stochastic lower block triangular Toeplitz matrix H_i^s

Equation 4.14

$$\mathrm{H_i^s} \stackrel{\text{def}}{=} \begin{pmatrix} 0 & 0 & 0 & \dots & 0 \\ \mathrm{C} & 0 & 0 & \dots & 0 \\ \mathrm{CA} & \mathrm{C} & \mathrm{D} & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \mathrm{CA^{i-2}} & \mathrm{CA^{i-3}} & \mathrm{CA^{i-4}} & \dots & 0 \end{pmatrix}$$

The input and output block Hankel matrices are defined as

Equation 4.15

$$U_{0|i-1} \stackrel{\text{def}}{=} \begin{pmatrix} \mathrm{u}_0 & \mathrm{u}_1 & \mathrm{u}_2 & \cdots & \mathrm{u}_{j-1} \\ \mathrm{u}_1 & \mathrm{u}_2 & \mathrm{u}_3 & \cdots & \mathrm{u}_j \\ \mathrm{u}_2 & \mathrm{u}_3 & \mathrm{u}_4 & \cdots & \mathrm{u}_{j+1} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \mathrm{u}_{i-1} & \mathrm{u}_i & \mathrm{u}_{i+1} & \cdots & \mathrm{u}_{i+j-2} \end{pmatrix}$$

Equation 4.16

$$Y_{0|i-1} \stackrel{\text{def}}{=} \begin{pmatrix} y_0 & y_1 & y_2 & \cdots & y_{j-1} \\ y_1 & y_2 & y_3 & \cdots & y_j \\ y_2 & y_3 & y_4 & \cdots & y_{j+1} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ y_{i-1} & y_i & y_{i+1} & \cdots & y_{i+j-2} \end{pmatrix}$$

where it is assumed for stochastic reasons that $J \to \infty$. For convenience and shorthand notation, we call (Favoreel W., 2000)

$$U_P \stackrel{\text{def}}{=} U_{0|i-1}, U_f \stackrel{\text{def}}{=} U_{i|2i-1}, Y_P \stackrel{\text{def}}{=} Y_{0|i-1}, Y_f \stackrel{\text{def}}{=} Y_{i|2i-1}$$

Where subscript p and f denote, respectively, the past and future. The matrix containing the inputs U_p and the output Y_p will be called W_p

$$W_p \stackrel{\text{def}}{=} \begin{pmatrix} Y_p \\ U_p \end{pmatrix}$$

The block Hankel matrix Eq. (4.15) and (4.16) formed with the process noise w_k and measurement noise v_k are defined, respectively, as $M_{0|i-1}$ and $N_{0|i-1}$ in the same way. Once again, we define for shorthand notation

$$M_P \stackrel{\text{def}}{=} M_{0|i-1}, M_f \stackrel{\text{def}}{=} M_{i|2i-1}, N_P \stackrel{\text{def}}{=} N_{0|i-1}, N_f \stackrel{\text{def}}{=} N_{i|2i-1}$$

We finally denote the state sequence X_i as:

Equation 4.17

$$X_i \overset{\text{def}}{=} (x_i \quad x_{i+1} \quad x_{i+2} \quad \cdots \quad x_{i+j-1})$$

later we will use the matrices $\mathcal{A} \in \mathbb{R}^{p\times j}$ and $\mathcal{B} \in \mathbb{R}^{q\times j}$.

Definition (orthogonal projections)

The orthogonal projection of the row space of $\mathcal{A}$ into the row space of $\mathcal{B}$ is denoted by $\mathcal{A}/\mathcal{B}$ and defined as:

$\mathcal{A}/\mathcal{B} = \mathcal{A}\mathcal{B}^{\dagger}\mathcal{B}$

$\mathcal{A}/\mathcal{B}^{\perp}$ is the projection of the row space of $\mathcal{A}$ into $\mathcal{B}^{\perp}$, the orthogonal complement of row space of $\mathcal{B}$, for which we have $\mathcal{A}/\mathcal{B}^{\perp} = \mathcal{A} - \mathcal{A}/\mathcal{B}$.

The two basic steps in subspace identification

As previously mentioned, all subspace algorithms consist of two main steps (see Figure 4.1: Main steps in subspace algorithms (Source: Favoreel et al. (2000)).). The first step always performs a weighted projection of the row space of the previously defined data Hankel matrices. From this projection, the observability matrix Γ_i and/or an estimate X_i of the state sequence $\widetilde{X}_i$ can be retrived. In the second step, the system matrices A, B, C, D and Q, S, R are determined. A clear distinction can be made between the algorithms that use the extended observability matrix Γ_i to obtain the state space matrices, and those using the estimated state sequence $\widetilde{X}_i$. In addition, subspace algorithms are considered non-iterative, as opposed to least squares and prediction error methods which are iterative.

First step: Finding the state sequence and/or the extended observability matrix

All subspace methods start from the previously presented matrix input-output Eq. (4.11). It states that the block Hankel matrix containing the future outputs Y_f is related in a linear way to the future input block Hankel matrix U_f and the future state sequence X_i. The basic idea of subspace identification now is to recover the $\Gamma_i X_i$ -term of this equation. This is a particularly interesting term since either the knowledge of X_i or Γ_i leads to the system parameters. Moreover, $\Gamma_i X_i$ is a rank deficient term (of rank n, i.e. the system order) which means that once $\Gamma_i X_i$ is knownΓ_i, X_i and the order n can be simply found from a Singular Value Decomposition (SVD).

How can an estimate of $\Gamma_i X_i$ be extracted from the equation (4.11)? For this we need the previously defined notion of orthogonal projection. By projecting the row space of Y_f into the orthogonal complement $U_f^{\perp}$of the row space of U_f we find.

$$\mathrm{Y_f}/U_f^{\perp} = \Gamma_\mathrm{i}\mathrm{X_i}/U_f^{\perp} + \mathrm{H_i^s}\mathrm{M_f}/U_f^{\perp} + \mathrm{N_f}/U_f^{\perp}$$

Since it is assumed that the noise is uncorrelated with the inputs, we have that

$$\mathrm{M_f}/U_f^{\perp} = \mathrm{M_f},\ \mathrm{N_f}/U_f^{\perp} = \mathrm{N_f}$$

Therefore

$$\mathrm{Y_f}/U_f^{\perp} = \Gamma_\mathrm{i}\mathrm{X_i}/U_f^{\perp} + \mathrm{H_i^s M_f} + \mathrm{N_f}$$

The following step consists in weighting this projection to the left and to the right with some matrices W1 and W2

$$W_1\,\mathrm{Y_f}/U_j^{\perp}\,W_2 = \underbrace{W_1\Gamma_i}_{1.}\,\underbrace{\mathrm{X_i}/U_f^{\perp}\,W_2}_{2.} + \underbrace{W_1(\mathrm{H_i^s M_f} + \mathrm{N_f})W_2}_{3.}.$$

Of course, the inputs U_f and the weighting matrices W_1 and W_2 cannot be chosen arbitrarily but they should satisfy the following three conditions:

Equation 4.18

1. $$rank\,(W_i\Gamma_i) = rank\;\Gamma_i$$

Equation 4.19

2. $$rank\left(X_i/U_f^{\perp}\,W_2\right) = rank\;X_i$$

Equation 4.20

3. $$W_1\left(H_i^s M_f + N_f\right)W_2 = 0$$

The first two conditions guarantee that the rank-n property of $\Gamma_\mathrm{i}\mathrm{X_i}$ is preserved after projection onto $U_f^{\perp}$ and weighting by W_1 and W_2. The third condition expresses that W_2 should be uncorrelated with the noise sequences w_k and v_k. If these three conditions are satisfied, we have that:

Equation 4.21

$$\mathcal{O}_i \overset{\mathrm{def}}{=} W_1\,\mathrm{Y_f}/U_j^{\perp}\,W_2 = W_1\Gamma_i\,\mathrm{X_i}/U_f^{\perp}\,W_2$$

With SVD

$$\mathcal{O}_i = (U_1 \quad U_2)\begin{pmatrix} S_1 & 0 \\ 0 & 0 \end{pmatrix}\begin{pmatrix} V_1^T \\ V_2^T \end{pmatrix}$$

The following important properties can now be stated

$$rank\;\mathcal{O}_i = n$$

$$W_i\Gamma_i = U_1 S_1^{1/2}$$

$$\mathrm{X_i}/U_f^{\perp}\,W_2 = S^{1/2}V_2^T$$

Obviously, the singular value decomposition of the matrix $W_1 Y_f = U_j^{\perp} W_2$ delivers the order n of the system. Moreover, from the left singular vectors corresponding to the non-zero singular values the extended observability matrix Γ_i can be found (up to a similarity transformation) whereas the right singular vectors contain information about the states X_i. For an appropriate choice of the weighting matrixw_2, the matrix

Equation 4.22

$$\widetilde{X}_i \stackrel{\text{def}}{=} X_i/U_f\,W_2$$

can indeed be considered as an estimate of the state sequence $\mathrm{X_i}$. It was shown from (Van Overschee, 2012)that, for a particular choice of W_2, $\widetilde{\mathrm{X}}_\mathrm{i}$is a Kalman filter estimate of $\mathrm{X_i}$. By choosing appropriate weighting matrices W_1 and W_2, all subspace algorithms for LTI systems can be interpreted in the above framework, including N4SID algorithm (Van Overschee, 1994). It should be noted that for the basic-4SID algorithm, condition 4.20 is not satisfied which implies that in general this method is not consistent. (Favoreel W., 2000)

Second step: finding the state space model N4SID algorithm uses the state estimates $\widetilde{\mathbf{X}}_\mathbf{i}$ (the right singular vectors) to find the state space model. If the weights W_1 and W_2 correspond to those of the N4SID algorithm ((Van Overschee, 2012)) and (Van Overschee, 1994)):

Equation 4.23

$$\mathrm{W}_1 = I_{li}$$

Equation 4.24

$$\mathrm{W}_1 = \left(W_p/U_f^{\perp}\right)^{\dagger} W_p$$

The estimated state sequence $\widetilde{\mathrm{X}}_\mathrm{i}$ can be interpreted as the solution of a bank of Kalman filters, working in parallel on each of the columns of the matrix Wp. Besides $\widetilde{\mathrm{X}}_\mathrm{i}$, we also need the state sequence $\tilde{X}_{i+1}$. This sequence can be obtained from a projection and new weights $\overline{W}_1$, $\overline{W}_2$ in Eq.(4.21) based on $W_{0|i}$, $Y_{i+1|2i-1}$ and $U_{i+1|2i-1}$. This leads to the sequence $\mathcal{O}_{i+1}$ and the Kalman filter states $\tilde{X}_{i+1}$:

$$\mathcal{O}_{i+1} = \overline{W}_1 Y_{i+1|2i-1}/U_{i+1|2i-1}^{\perp}\,\overline{W}_2 = \overline{W}_1 \Gamma_{i-1} \tilde{X}_{i+1} \overline{W}_2$$

System model: The state space matrices A; B; C; D can now be found by solving a simple set of over determined equations in a least squares sense((Van Overschee, 2012)) and (Van Overschee, 1994)):

Equation 4.25

$$\begin{pmatrix} \tilde{X}_{i+1} \\ Y_{i|i} \end{pmatrix} = \begin{pmatrix} A & B \\ C & D \end{pmatrix} \begin{pmatrix} \tilde{X}_i \\ U_{i|i} \end{pmatrix} + \begin{pmatrix} \rho_w \\ \rho_v \end{pmatrix}$$

with obvious definitions for ρ_w and ρ_v as residual matrices. This reduces to:

$$\min_{A,B,C,D} \left\| \begin{pmatrix} \tilde{X}_{i+1} \\ Y_{i|i} \end{pmatrix} - \begin{pmatrix} A & B \\ C & D \end{pmatrix} \begin{pmatrix} \tilde{X}_i \\ U_{i|i} \end{pmatrix} \right\|_F^2$$

Noise model: The noise covariances Q; S and R can be estimated from the residuals ρ_w and ρ_v as:

$$\begin{pmatrix} Q & S \\ S^T & R \end{pmatrix}_i = \frac{1}{j} \left[\begin{pmatrix} \rho_w \\ \rho_v \end{pmatrix} \begin{pmatrix} \rho_w^T & \rho_s^T \end{pmatrix} \right] \geq 0$$

where the index i denotes a bias induced for finite i, which disappears as $i \to \infty$. As is obvious by construction, this matrix is guaranteed to be positive semi-definite. This is an important feature since only positive definite covariances can lead to a physically realizable noise model. (Favoreel W., 2000).

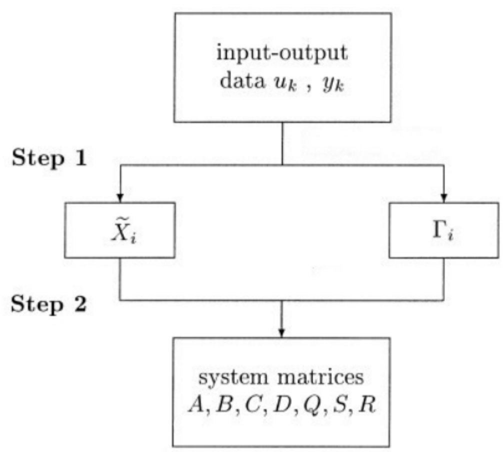


Figure 4.1: Main steps in subspace algorithms (Source: Favoreel et al. (2000)).

4.5 Conclusion

Earlier in this chapter different statistical tools explained that by utilizing them important statistical relations such as correlation coefficient and mean and standard deviation of data will be obtained. These statistical characteristics are used to determine which data set is suitable to make a model for prediction and forecast of the final data set that is BCTI. The final model obtained by constructing these data in an input-output model with the system identification method N4SID will be used for forecasting the BCTI. Examination of potential model and results of the statistical analysis will be explained in the next chapter. The simplified procedure is graphically presented by Figure 4.2 where the procedures is illustrated step by step.

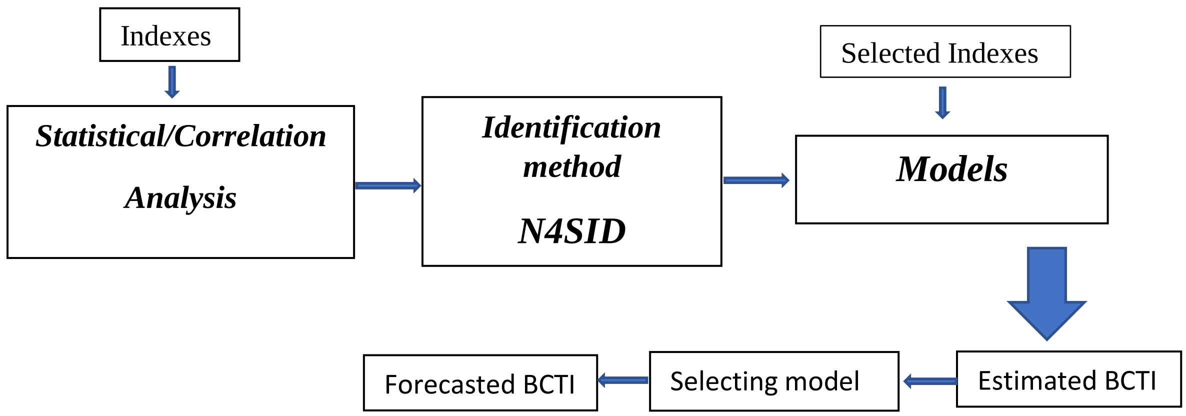


Figure 4.2: The methodology Flowchart Source: Author

5 Results

5.1 Introduction

This chapter will present the results and analysis that are obtained from the data of this thesis. This chapter is divided into four chapters. the first chapter will discuss the data source and the software which is used for analysis and results. Later statistical analysis of the data will be presented to the reader. In the statistical part of this chapter relation between data will be tested and based on that the data will be selected for the future tests and development of models. At the end of this chapter, the modelling results will be presented.

5.2 Data, data range, software

For the purpose of analysis and making the graphs and calculation MATLAB R2019a has been used, due to its ease of use and wide range of apps and tools available for this software. To change type and duration of data R programing software has been used for this thesis.

As mentioned before all the data used for analysis and prediction in this thesis has been collected and validated by *Clarkson Research*™. As explained in *the Clarkson Research*™ website their data is collected from reliable sources and vastly used by researchers and analysts in the shipping industry. The reader of this thesis can find a list of examined data and their abbreviations in table 5.1. This table also provides mean, standard deviation, correlation coefficient analysis of data against BCTI. Selection of these indexes has been explained in chapter 3.2 of this thesis during the literature review.

From now, readers of this thesis shall familiar their self to the abbreviation in table 5.1 to better understand modelling and prediction modelling that will be presented in this chapter and the next chapter. Observation range of all data in this chapter is from 01-01-2000 to 01-02-2019. This range is selected so that results can be tested against future real figures from 01-02-2019 to the latest available data. BCTI will be referred to as **var 20** where it has been forced by the limitations of the software.

5.3 Statistical analysis

5.3.1 Mean and standard deviation

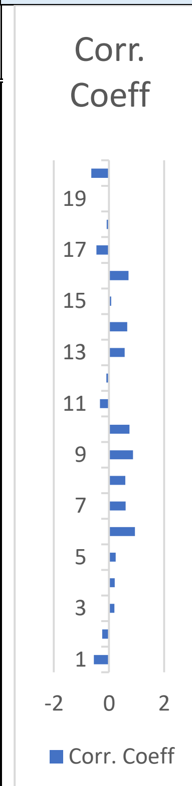
In chapter 4.3 methodological analysis of mean, standard deviation and correlation coefficient was explained. These relations have been presented in table 5.1 to show numerical relations of selected indices with BCTI. Correlation coefficient values will be further discussed in this chapter to analyze and chose the best and most relevant indexes for developing a prediction model. Some of the below indices will not be used to develop the model even they are well related to the BCTI.

Using methods explained in chapter 4.3, mean and standard deviation for the BCTI (var 20) will be obtained as μ_{BCTI}= 835.02 and standard deviation is σ_{BCTI}= 311.81 for the selected date range that is from 01-01-2000 to 01-02-2019.

A visual demonstration of the correlation coefficients also has been presented in table 5.1.

Table 5.1: List of examined data with their abbreviations, units and their related statistical analysis including standard deviation and mean and correlation coefficient. A visual demonstration of correlation coefficients is placed on the right side of the table.

Row	*Data*	*Unit*	*Abbreviation*	*Standard Deviation (σ)*	*Mean(μ)*	*BCTI correlation coefficient*
1	Global Oil Prod.	mbpd	var1	7.17	87.45	-0.55
2	Brent Crude Oil Price	$/bbl	var2	30.69	64.55	-0.25
3	Atlantic Region Industrial Production Growth	% Yr/Yr	var3	4.00	1.02	0.20
4	Pacific Region Industrial Production Growth	% Yr/Yr	var4	5.08	5.13	0.21
5	Industrial Production OECD	% Yr/Yr	var5	4.69	1.10	0.25
6	TC5	ws	var6	96.50	143.62	0.94
7	USA Interest Rates	%	var7	2.08	5.25	0.60
8	LIBOR Interest Rates	%	var8	2.06	2.38	0.59
9	BDTI Index	Index	var9	440.16	1038.50	0.87
10	ClarkSea Index	$/day	var10	8945.25	17271.39	0.75
11	HFO 380 Avg	$ / MT	var11	195.50	383.50	-0.33
12	Houston - Far East 5,000mt EasyChem Rates	$/Tonne	var12	15.31	58.70	-0.09
13	Industrial Production China	% Yr/Yr	var13	4.50	11.76	0.57
14	German/Euro Interest Rates	%	var14	1.58	1.79	0.66
15	Japan Interest Rates	%	var15	0.00	0.01	0.08
16	Crude Oil Imports, US, EU-4, Japan	mbpd	var16	2.74	15.59	0.71
17	Dow jones	index	var17	4551.85	13458.60	-0.46
18	Exchange Rates Euro Index	Index	var18	16.98	119.36	-0.08
19	Monthly Gold price change	%	var19	3.77	0.74	0.01
20	Monthly Gold price	$	var21	12.95	279.09	-0.64



5.3.2 Correlation analysis

Previously correlation coefficient figures of all data sets against BCTI was presented in table 5.1. Figure 5.1 is presenting the same information for all data sets against each other and against BCTI (var 20) plus their scatter plots and each data set's histogram. Graphical representation of the statistical datasets reveals more information for a better choice of related variables later in this thesis. A well-scattered data near the line of relationship shows a valid data set for prediction of BCTI.

Figure 5.3 shows the stacked plot of all datasets over BCTI to visualize their relation with this index. Most of the selected indexes show the same trend during the troughs and crests over the plot waves. Relation of critical economical condition with the change of trends can be seen in this graph. A crisis in 2008 is visible around 110 in graph 5.3 over most of the datasets. More detailed plots of each data set against BCTI can be found in Appendix A of this thesis.

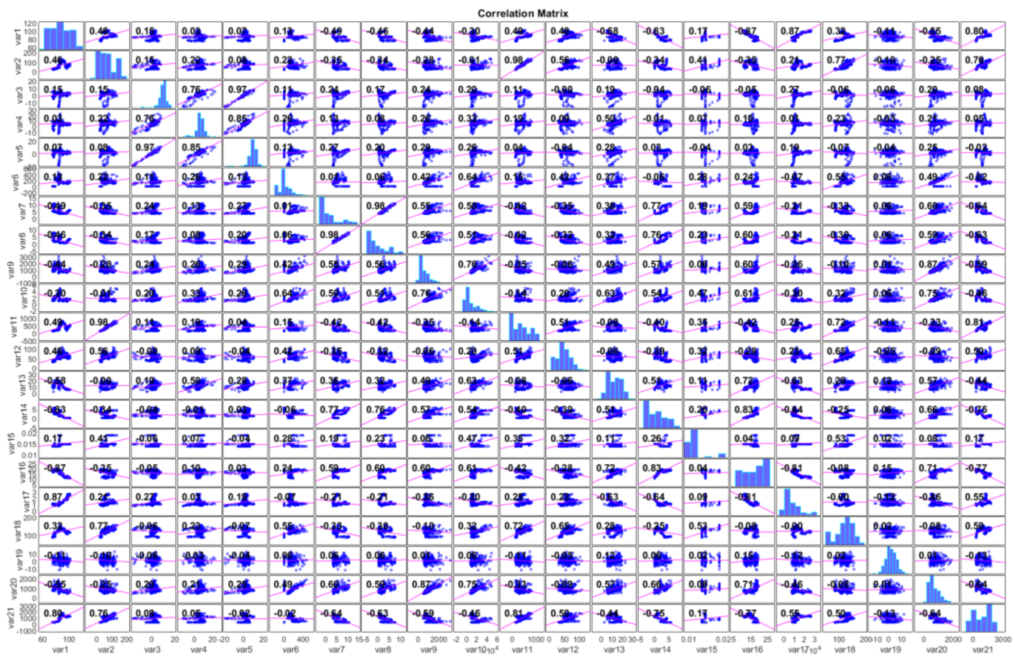


Figure 5.1: Graph indicating the coefficient of correlation plus scattered plot of examined datasets against each other and the BCTI together with their histograms (Source: Clarkson Research™ *(2019))*

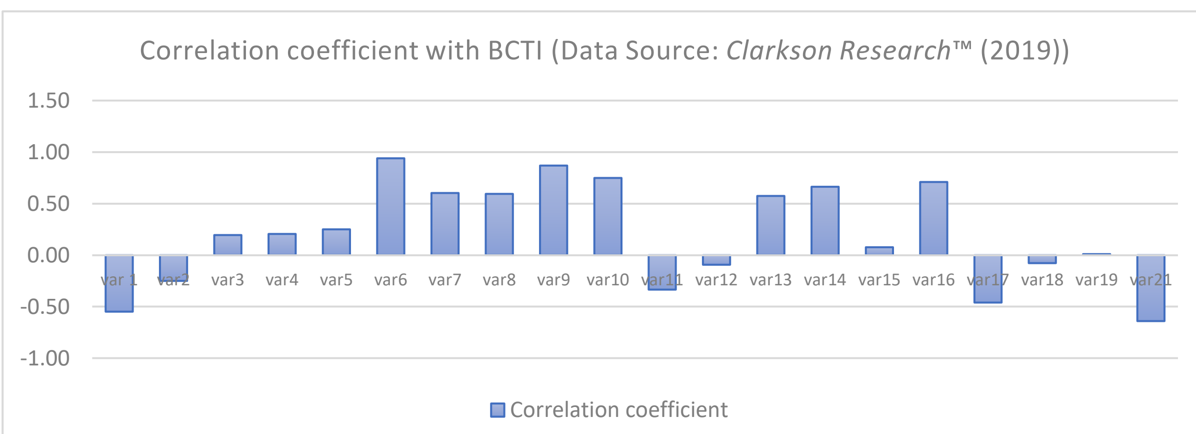


Figure 5.2: Correlation coefficient with BCTI (Data Source: Clarkson Research™ (2019))

Figure 5.4 represent the relation between monthly Global Oil Production and monthly BCTI and their histograms. The correlation coefficient of these two variables is -0.55 as per table 5.1 that means a relatively good negative relation is abiding. From Figure A.2 it can be observed that global oil production is constantly growing and it can be seen that up to around 110 or mid-2008 these two indices were correlated but after 2008 global oil production continued to grow while BCTI followed a different pattern. therefore, monthly global oil production is no longer a good independent variable which is affecting BCTI.

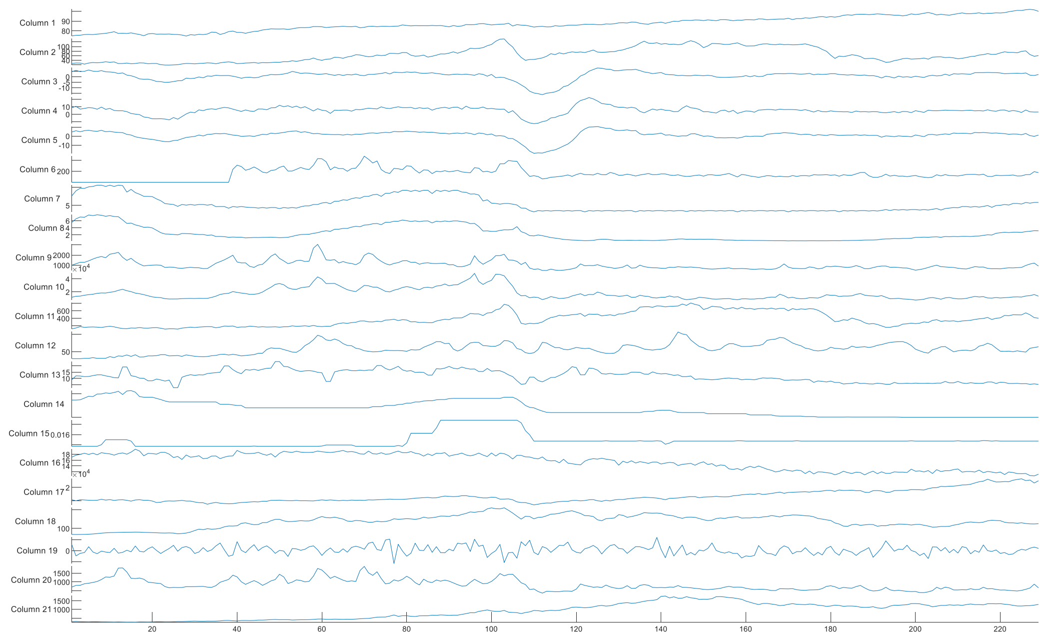


Figure 5.3: Graph that shows a stacked plot of all data sets over the BCTI. Each column represents a variable with the same number and order as presented in Table 5.1 earlier in this chapter (Source: Clarkson Research™ *(2019)).*

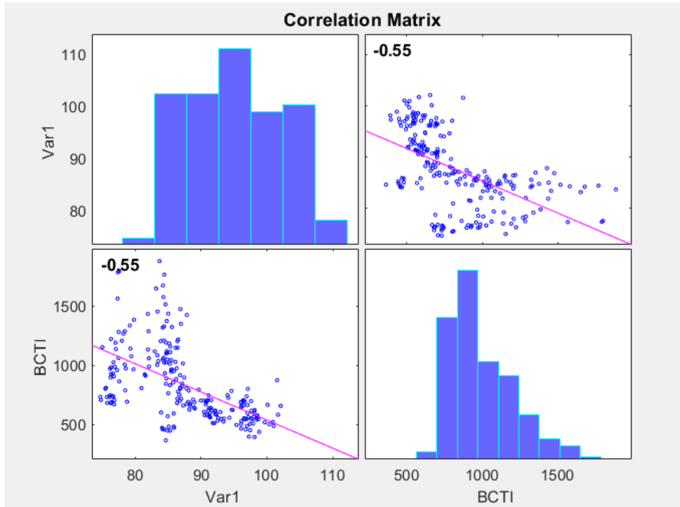


Figure 5.4: Graph that illustrates the correlation between monthly Global Oil Production and monthly BCTI and their histograms (Data source : (Sin.Clarksons.net, 2019)).

From figure 5.5 it can be seen that Brent crude oil price has a negative correlation with BCTI. this correlation is not a very strong correlation as scatter plot also shows. But since this variable plot is well scattered over the relation line it will be used as an independent variable form BCTI that has an impact over the mentioned index.

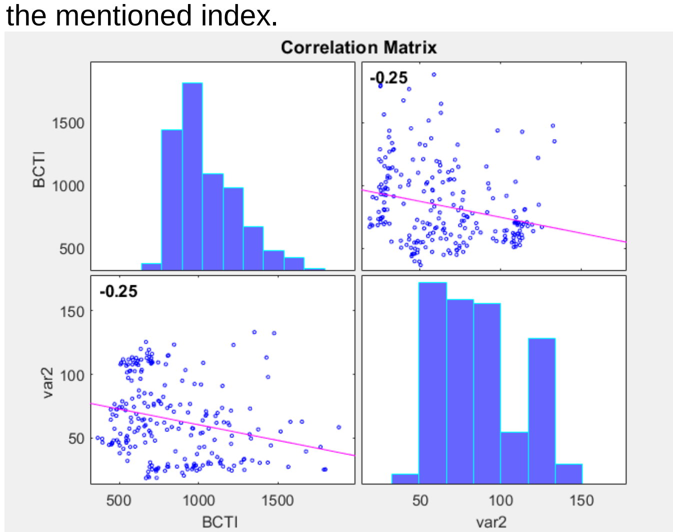


Figure 5.5: Graph that illustrates the correlation between monthly Brent Crude Oil Price and monthly BCTI and their histograms (Data source : (Sin.Clarksons.net, 2019)).

Figure 5.6 illustrates monthly Atlantic Region Industrial Production Growth and monthly BCTI. This index shows similar characteristics as the next index monthly Pacific Region Industrial Production scatter plot that is shown in figure 5.7 and index next to that is the correlation between monthly Industrial Production OECD and monthly BCTI. From Figures A.4, A.5, A.6 one can observe the similar shape of the plot of these three indexes. Since they are correlated to the BCTI but they are very similar to each other and have the same characteristics the most correlated index will be used for further development of the models. In this case, most correlated index is monthly Industrial Production OECD. This index scatter plot and its histogram is shown in Figure 5.8.

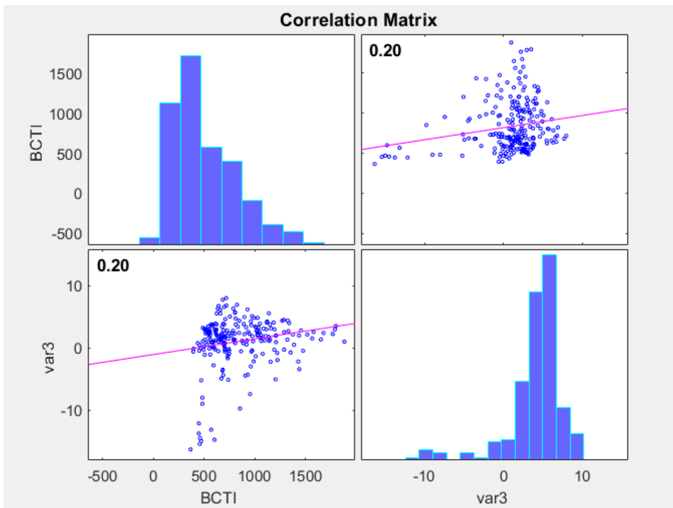


Figure 5.6: : Graph that illustrates the correlation between monthly Atlantic Region Industrial Production Growth and monthly BCTI and their histograms (Data source : (Sin.Clarksons.net, 2019)).

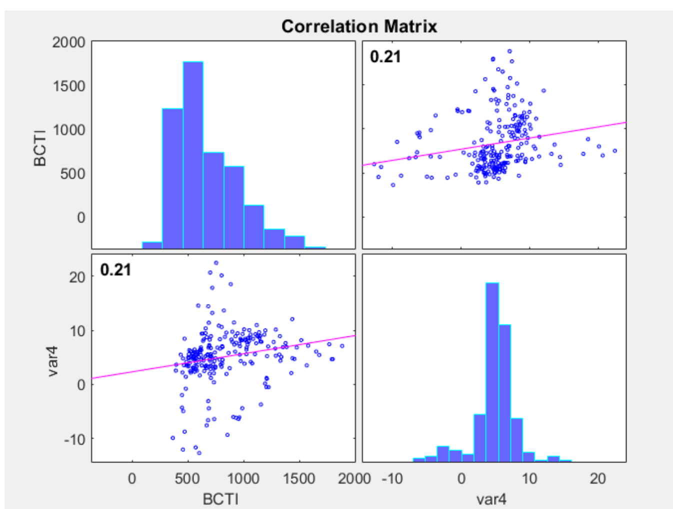


Figure 5.7: Graph that illustrates the correlation between monthly Pacific Region Industrial Production Growth and monthly BCTI and their histograms (Data source : (Sin.Clarksons.net, 2019)).

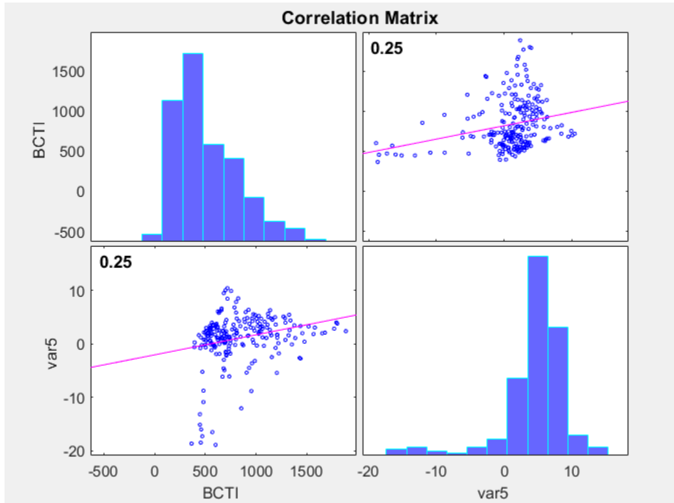


Figure 5.8 :Graph that illustrates the correlation between monthly Industrial Production OECD and monthly BCTI and their histograms (Data source : (Sin.Clarksons.net, 2019)).

Figure 5.9 illustrates the correlation between route TC5 monthly rates and monthly BCTI. Route TC5 rates are selected to represent one of the routes that BCTI is calculated from it. Both variables show very strong relationship over the relation line but since BCTI is calculated from routs weighted values, therefore, this index is not independent from BCTI and is not a good variable for the final chose of variables.

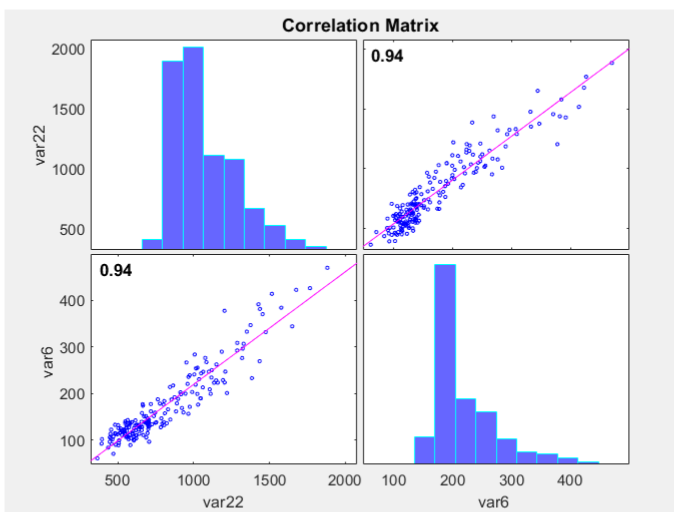


Figure 5.9: Graph that illustrates the correlation between monthly route TC5 and monthly BCTI and their histograms (Data source : (Sin.Clarksons.net, 2019)).

Figure A.8 shows that after 2008 monthly USA Interest Rates is kept constant. From Figure 5.10 however, a relatively strong correlation can be found between monthly USA interest rates and monthly BCTI. Since for a significantly long period, monthly USA Interest Rates were constant and as scatter plot and histogram shows a significant amount of the data is linearly concentrated in one area this variable is rejected and will not be used in the development of the models later on this thesis.

Figure 5.11 illustrates the correlation between monthly LIBOR Interest Rates and monthly BCTI and their histograms. From this figure, a well-scattered set of data with a good correlation between variables can be found out. LIBOR rates as the reader can observe in Figure A.9 and Figure 5.11 is correlated and independent from the BCTI. Therefore, this variable will be used for the construction of the models later in this thesis.

BCTI and BDTI are showing a strong linear relation with each other in Figure 5.12. However, these indexes are independent but they are affected by the same market that is freight and oil market, therefore, this is not a good choice of the variable for BCTI modelling.

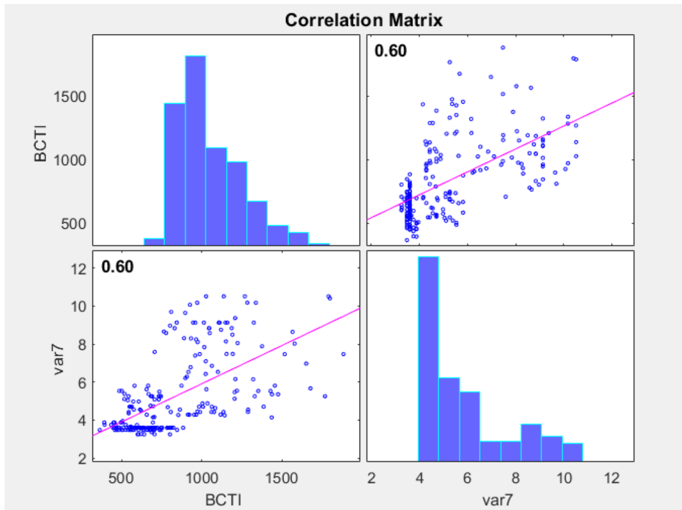


Figure 5.10: Graph that illustrates the correlation between monthly USA Interest Rates and monthly BCTI and their histograms (Data source : (Sin.Clarksons.net, 2019)).

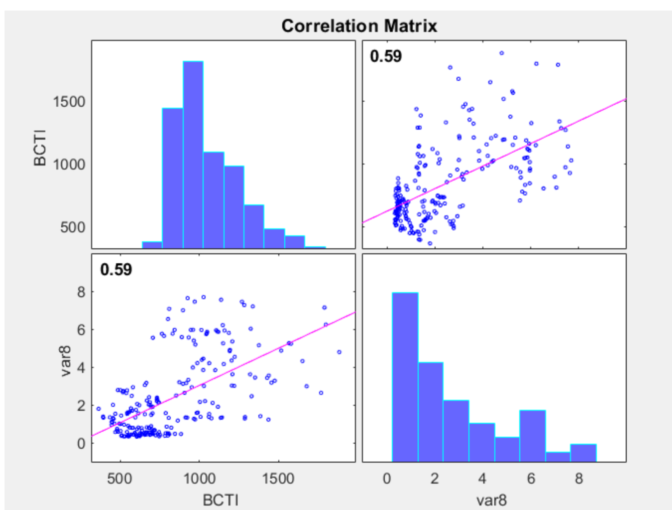


Figure 5.11: Graph that illustrates the correlation between monthly LIBOR Interest Rates and monthly BCTI and their histograms (Data source : (Sin.Clarksons.net, 2019)).

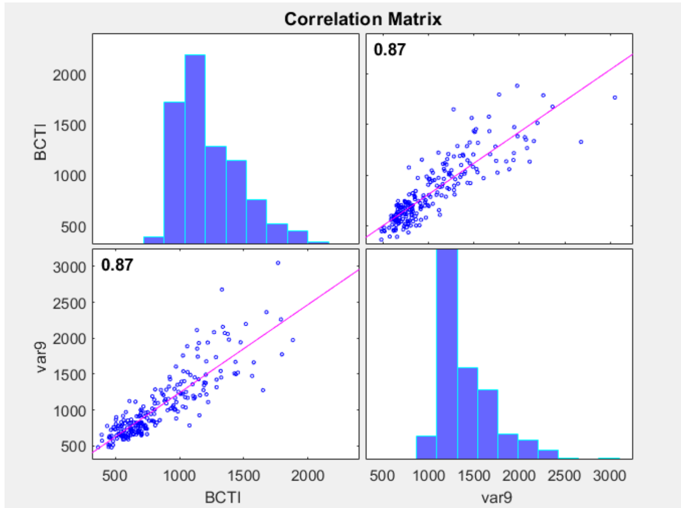


Figure 5.12: Graph that illustrates the correlation between monthly BDTI Index and monthly BCTI and their histograms (Data source : (Sin.Clarksons.net, 2019)).

The next index is ClarkSea index. This index is also showing a very strong correlation with BCTI but monthly ClarkSea index and monthly BCTI showing freight rates for a different area of the shipping operations and basically are showing the same thing, therefore, ClarkSea index also is rejected and will not be used for later development of the models.

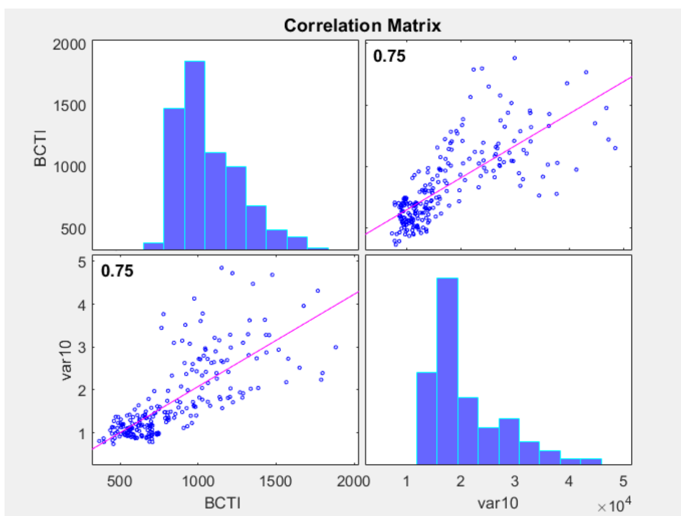


Figure 5.13: Graph that illustrates the correlation between monthly ClarkSea Index and monthly BCTI and their histograms (Data source : (Sin.Clarksons.net, 2019)).

Correlation between monthly HFO 380 Average price and monthly BCTI and their histograms are shown in Fig 5.14. A negative correlation between the two indices shows that with the increase of HFO 380 price BCTI will decrease. HFO 380 is the main type of the bunker used by ocean-going vessels. Average HFO 380 price here is calculated by adding up prices of this fuel in Singapore, Rotterdam and Fujairah and dividing the result to three. Singapore, Rotterdam and Fujairah are three main bunker suppliers and therefore representing bunker price in their area. This index is showing independence from BCTI and shows a good negative relation with the BCTI. Therefore, this index will be used as input for developing the models.

Monthly Houston - Far East 5,000mt EasyChem Rates has been selected to find the relation between freight price in a single route of chemical tankers and BCTI. As Figure 5.15 shows there is a weak negative correlation with BCTI and therefore this index is also rejected

Correlation between monthly Industrial Production China and monthly BCTI and their histograms are illustrated in Figure 5.16. the scattered plot shows a good scatter of data over the relation line and correlation coefficient also shows a good correlation between both indices. Therefore, the industrial production growth of china causes an increase of BCTI. This index is independent of BCTI but correlated to it and therefore is a good choice for modelling of BCTI.

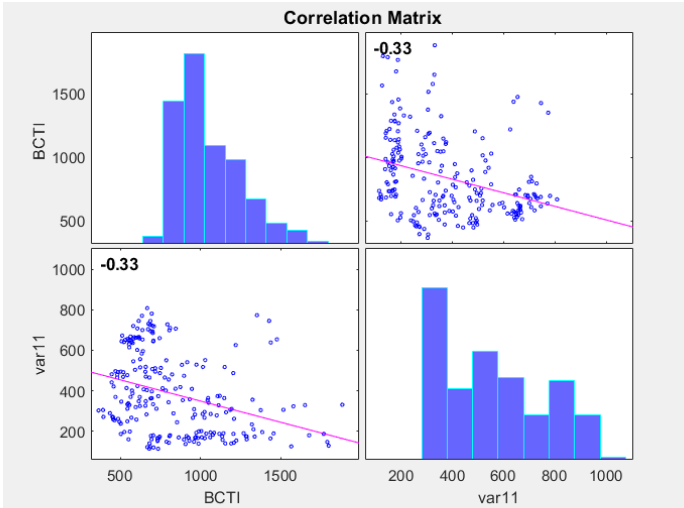


Figure 5.14: Graph that illustrates the correlation between monthly HFO 380 Average price and monthly BCTI and their histograms (Data source : (Sin.Clarksons.net, 2019)).

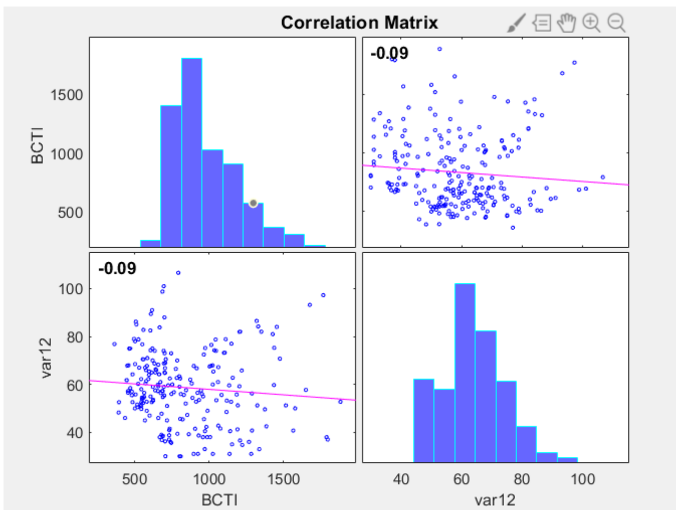


Figure 5.15: Graph that illustrates the correlation between monthly Houston - Far East 5,000mt EasyChem Rates and monthly BCTI and their histograms (Data source : (Sin.Clarksons.net, 2019)).

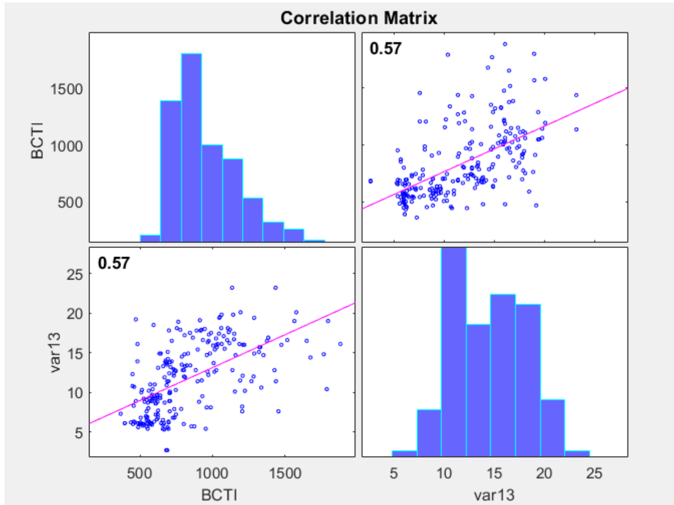


Figure 5.16: Graph that illustrates the correlation between monthly Industrial Production China and monthly BCTI and their histograms (Data source : (Sin.Clarksons.net, 2019)).

Monthly German/Euro Interest Rates and monthly BCTI are Correlated as per Figure. 5.17. Since data is scattered over the correlation line and correction coefficient also shows a good correlation to the BCTI it seems a good choice for the purpose of modeling but as the scattered plot shows the data are not homogeneously divided over the correlation line. This index will be used for modelling in this thesis since as figure 5.3 shows German/Euro Interest Rates change shows the same trend as BCTI. This variable will be omitted from two of the models later to better study the result of the outcome.

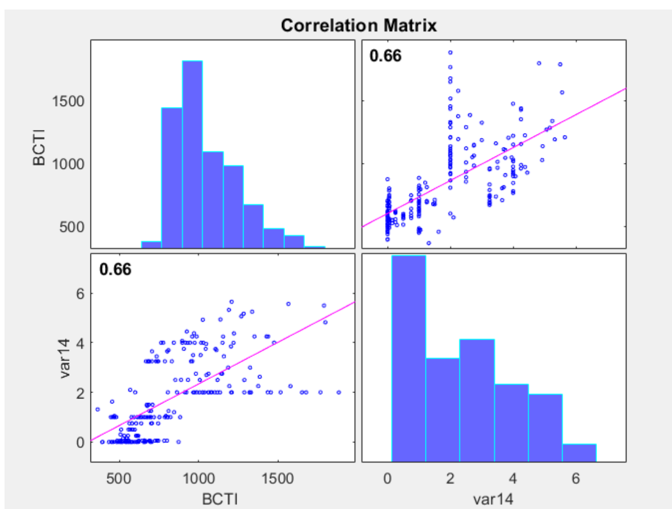


Figure 5.17: Graph that illustrates the correlation between monthly German/Euro Interest Rates and monthly BCTI and their histograms (Data source : (Sin.Clarksons.net, 2019)).

Figure 5.18 illustrates the correlation between monthly Japan Interest Rates and monthly BCTI and their histograms. It is clear from the graph that Japan Interest Rates is not a good choice of data and will be rejected.

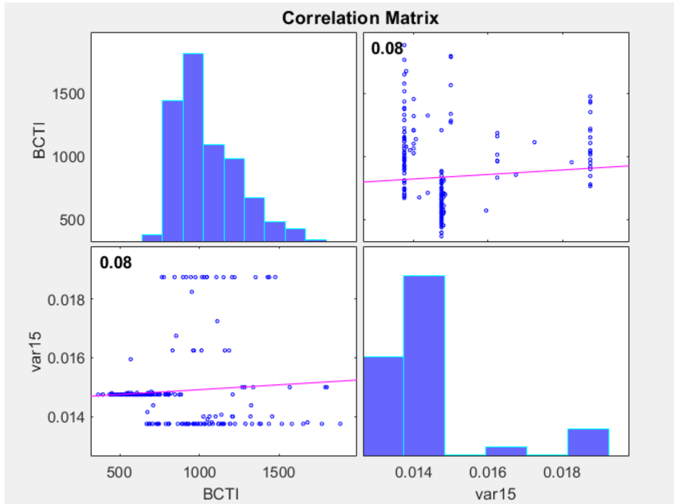


Figure 5.18: Graph that illustrates the correlation between monthly Japan Interest Rates and monthly BCTI and their histograms (Data source : (Sin.Clarksons.net, 2019)).

Figure 5.19 illustrates the correlation between monthly Crude Oil Imports, US, EU-4, Japan and monthly BCTI and their histograms. Monthly Crude Oil Imports, US, EU-4, Japan is independent of BCTI but has got a good correlation with the BCTI which makes this index a good choice for later modelling of the BCTI. As figure 5.19 shows data scattered over the correlation line and shape of the histogram also shows a good spread of the data.

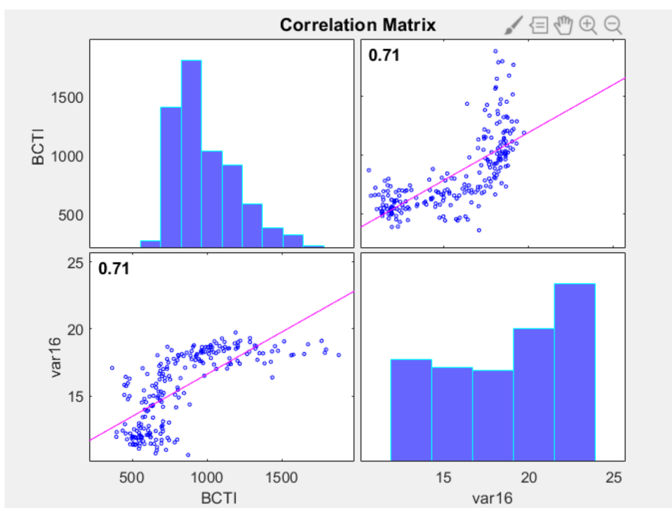


Figure 5.19: Graph that illustrates the correlation between monthly Crude Oil Imports, US, EU-4, Japan and monthly BCTI and their histograms (Data source : (Sin.Clarksons.net, 2019)).

Dow Jones industrial average (DOW) index is a stock market index that indicates the value of 30 large, publicly owned companies based in the United States, and how they have traded in the stock market during various periods of time. Correlation of This index with BCTI shows how value of 30 large us companies is related to BCTI. Since DOW actually can be an indication of the general world industrial growth is independent of BCTI. Figure 5.20 shows a relatively good negative correlation between the two indexes. DOW is then will be used for modelling of BCTI in this thesis since it is independent of BCTI and also has an acceptable negative relation with BCTI.

Gold is a medium for reserving of the value of money. Therefore, when the extreme economical condition is expected gold price increases or decreases according to market demand. Change of gold price can be an indication of expectations. Figure 5.22 illustrates the correlation between monthly Gold Price Change in percent and monthly BCTI and their histograms. The relation is very weak and therefore this index is rejected. Figure 5.23 indicates the correlation between monthly Gold Price and monthly BCTI and their histograms. Data is scattered over the correlation line and there is a strong negative correlation between two indexes. Gold price in independent from BCTI with good negative correlation and therefore is a good variable to be used to construct a model.

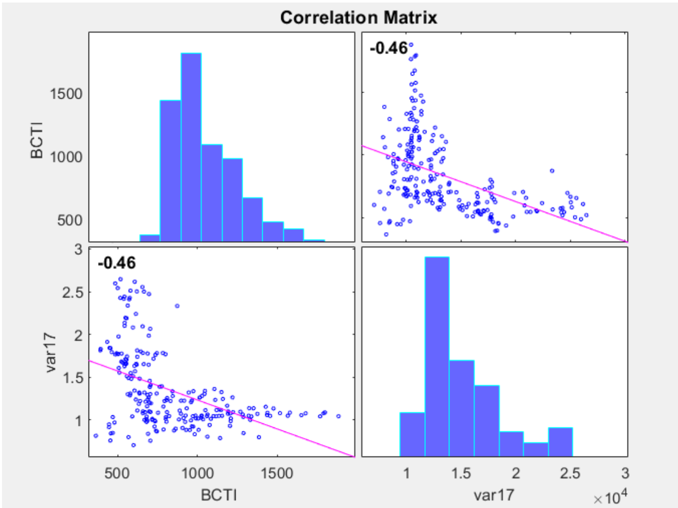


Figure 5.20: Graph that illustrates the correlation between monthly DOW index and monthly BCTI and their histograms (Data source : (Sin.Clarksons.net, 2019)).

Figure 5.21 illustrates the correlation between monthly Exchange Rates Euro Index that is a very weak relation and therefore this index is rejected and will not be used for Modeling purpose of this thesis.

Figure 5.22 shows correlation between monthly change of gold price with BCTI its relation is very week and this variable also is rejected.

Figure 5.23 shows a very negative strong correlation between the monthly gold price and the monthly BCTI. Since this variable is correlated but independent from the BCTI it will be used as a variable for the models of this thesis.

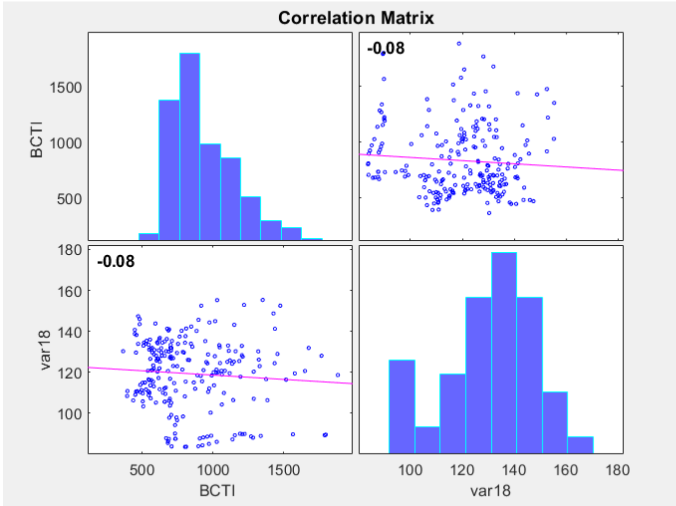


Figure 5.21: Graph that illustrates the correlation between monthly Exchange Rates Euro Index and monthly BCTI and their histograms (Data source : (Sin.Clarksons.net, 2019)).

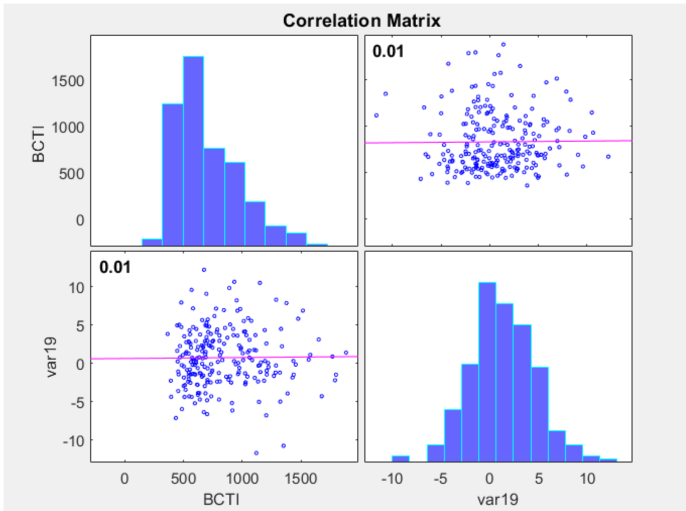


Figure 5.22: Graph that illustrates the correlation between monthly Gold Price Change in percent and monthly BCTI and their histograms *(Data source : (Goldprice.org, 2019)).*

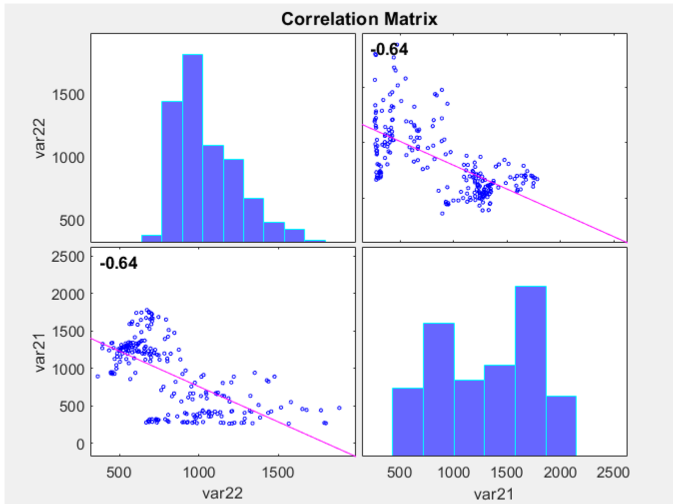


Figure 5.23 :Graph that illustrates the correlation between monthly Gold Price and monthly BCTI and their histograms *(Data source : (Goldprice.org, 2019)).*

5.4 Modeling

From the statistical analysis that was discussed earlier in chapter 5.3 nine indexes have been selected to construct models for prediction of the BCTI. These indexes as per table 5.1 are var2 or Monthly Brent crude oil price , var5 or Monthly industrial production OECD, var8 or Monthly LIBOR interest rates, var11 or Monthly HFO 380 Average price, var13 Monthly industrial production China, var 14 or Monthly German /Euro interest rates, var16 or Monthly crude oil imports, US , EU-4, Japan, var17 or Monthly DOW Jones index and var 21 or Monthly gold price.

Now that those modeling variables are known next step is to define models. The idea is to utilize the finding of correlation analysis in chapter 5.3 to define models with the best independent variables of BCTI. N4SID method as was explained in chapter 4 will be used to construct models based on input-output datasets.

The concept of each model is that BCTI is a dependent variable defined by a set of input data sets ($Index_1$, $Index_{2,}$ $Index_{3\ldots}$) which are independent form the output and BCTI will be the output of the model:

Equation 5.1

$$BCTI = f(\ Index_1,\ Index_{2,}\ Index_3\ldots)$$

All variables used for modelling are monthly variable from 01-01-2000 to 01-02-2019 as mentioned in chapter 5.3.1 therefore 229 monthly observation for each index. 223 observations will be used as training data to develop a model with the different variables as is defined later in chapter 5.4.1, 5.4.2 and 5.4.3. Obtained modes then will be tested against the last 6 observations to test accuracy to forecast the next 6 months. Then the best model will be selected and will be refined in chapter 5.4.4. The refined model will be tested against the next 12 observations to forecast the next one years.

5.4.1 First model

Following data set was selected earlier in chapter 4.3.3 after statistical and correlation analysis for the development of the models var2, var5, var8, var11, var13, var14, var16, var17, var21.

Refer to equation 5.1 this model is based on the idea of using all the selected indices by correlation analysis to create a model. following is the model:

Equation 5.2

$$\text{BCTI} = f(Var2, var5, var8, var11, var13, var14, var16, var17, var21)$$

With replacing the variables with their corresponding dataset names model 5.2 will be

Equation 5.3

$$\begin{gathered}\text{BCTI} = f(Brent\ Crudeoil\ Price, Industrial\ Production\ OECD,\\ LIBOR\ Interest\ Rates, HFO\ 380\ Avarage\ Price, Industrial\ Production\ China,\\ German/Euro\ \ Interest\ Rtaes, Crudeoil\ Imports\ (US, EU-4, Japan),\\ DowJones\ Index, Gold\ Price\ )\end{gathered}$$

Datasets suggested by the statistical analysis is used as input and represented by function f of the model and the BCTI is a dependent variable.

Figure 5.24 represent the stacked plot of all datasets used for this model. On the left are all available time series data and on the right training data have been plotted. The last 6 months of time series of the data have been neglected, as to have a percentage of unbiased available data for further evaluation reasons.

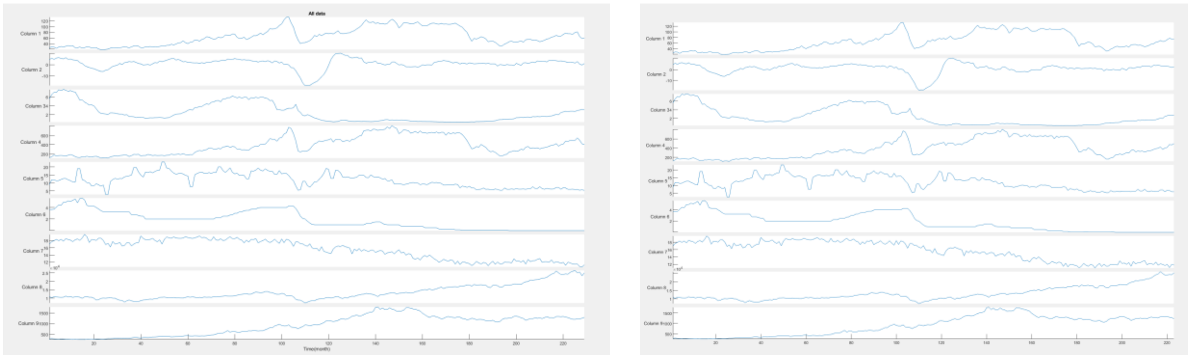


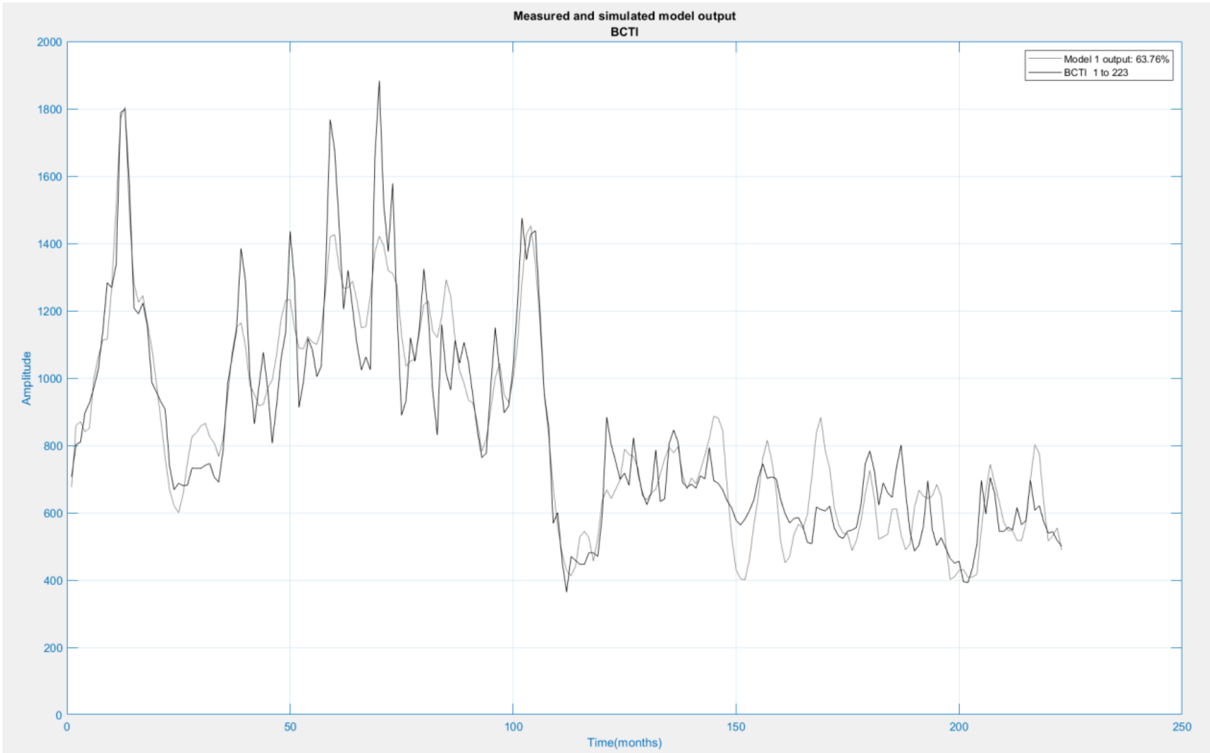
Figure 5.24: Stack plot of datasets used for model 1, the plot of all data on the left and plot of training data on the right. (each column respectively represents an index in the model)

Model response to the training data can be seen in Figure 5.25a. The model response to the given data is well. Model fitting to the system training data is 63.76%. the fitting of the model to the training data is acceptable to continue testing the model over all data.

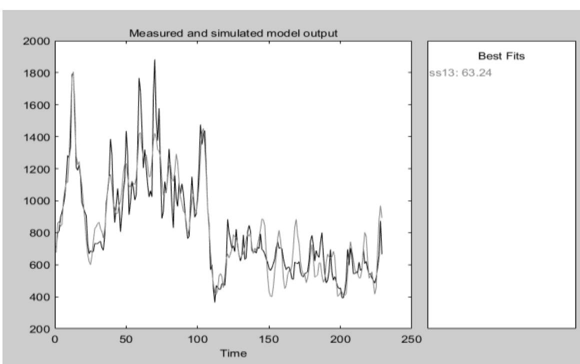
In Figure 5.25 bottom left (b) the result of the test of the model fitting against all data can be seen. Model fitting to all data is 63.24% which is only 0.52% less than the fitting to training data. This result is surprisingly well and shows that the model selection is rather a good choice. However, this model will be refined later in chapter 5.4.4 to present even better results.

This model has to respond well during the shocks and difficult situations. In figure 5.25a there is a sudden rise and fall from 90 to 110 and this range have been selected to test the response of the model in sudden changes. In figure 5.25c the result of such a test against a shock is shown. As one can see the fitting is 79.45% that is a very well result and shows that response of the model to shocks can be a reliable result.

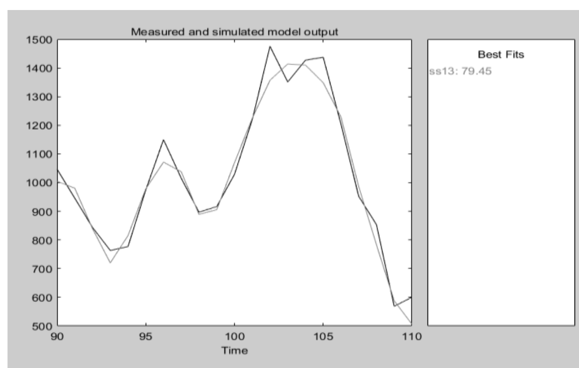
As one can see the model response to whole data is very well especially between 223 to 229. Since there is a rise and fall in this area and the model followed the real data plot pattern that was not used in training data, this model is a good model for forecasting the future trends and values of BCTI.



a



b



c

Figure 5.25:" a" is the model response to training data. "b" is the model response to all dada. "c" is the model response to a shock.

5.4.2 2nd model

Refer to equation 5.1 this model is based on the idea of using selected indices from correlation analysis of data which have a stronger correlation with BCTI and two of the variables that were used to construct model one are removed here. This model simplifies model one as presented in formula 5.2 and 5.3 to create another model. following is the model:

Equation 5.4

$$\text{BCTI} = f(v\text{ar2, var8, var11, var13, var16, var17, var21})$$

With replacing the variables with their corresponding dataset names formula 5.4 will be

Equation 5.5

$$\text{BCTI} = f(Brent\ Crudeoil\ Price, LIBOR\ Interest\ Rates, HFO\ 380\ Avarage\ Price, Industrial\ Production\ China, Crudeoil\ Imports\ (US, EU-4, Japan), DowJones\ Index, Gold\ Price\ )$$

Datasets suggested by statistical analysis is used as input and represented by function f of the model and the BCTI is a dependent variable.

To simplify the model one, two variables have been removed to construct a new model. These variables are ,$Industrial\ Production\ OECD$ and $German/Euro\ Interest\ Rtaes$ the first one is removed due to its relatively weak correlation with BCTI and the second one is removed due to its constant value over a considerable number of observations as can be seen in Figure A.15.

Figure 5.26 represents the stacked plot of selected datasets used for this model. On the left are selected available time-series data and on the right training data have been plotted. The last 6 months of time series of the data have been neglected, as to have a percentage of unbiased available data for further evaluation reasons.

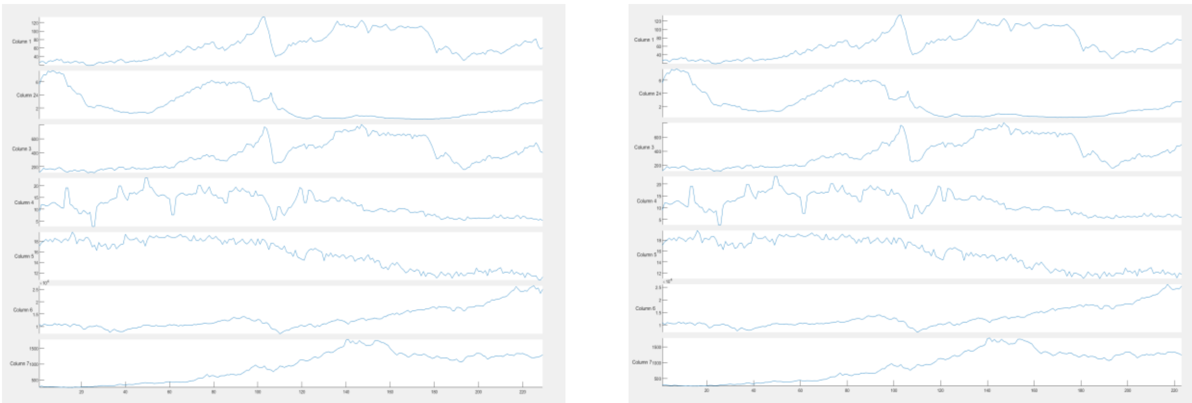
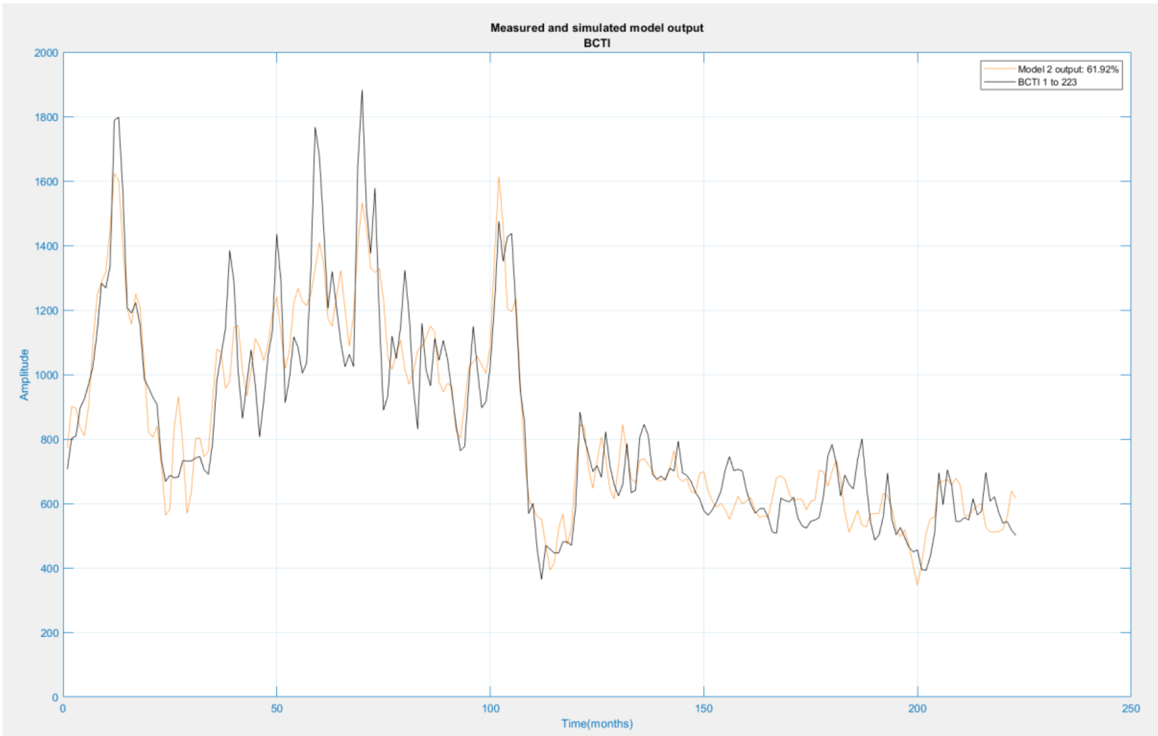


Figure 5.26: Stack plot of datasets used for model 2, the plot of all data on the left and plot of training data on the right. (each column respectively represents an index in the model)

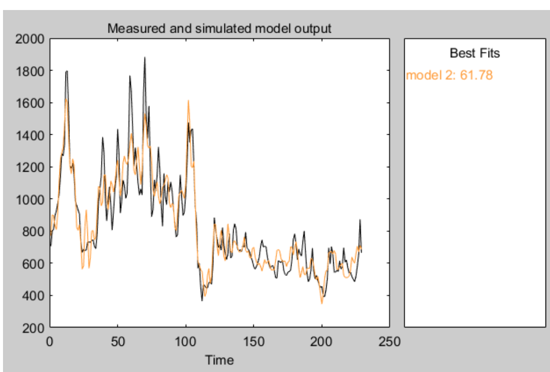
Model response to the training data can be seen in figure 5.27a. The model response to the given data is well however in comparison to the model one fitting to the training data results are less accurate. Model fitting to the system training data is 61.92%. the fitting of the model to the training data is acceptable to continue testing the model overall data.

Figure 5.27b is the result of the test of the model fitting against all data. Model fitting to all data is 61.78% which is only 0.14% less than the fitting to training data. This result is very good and shows a less change in data fitting in comparison with the model one however as represented by figure 5.26b model fitting to the trends is not very good and therefore this is not a good model to use for the forecast of the trend and future data.

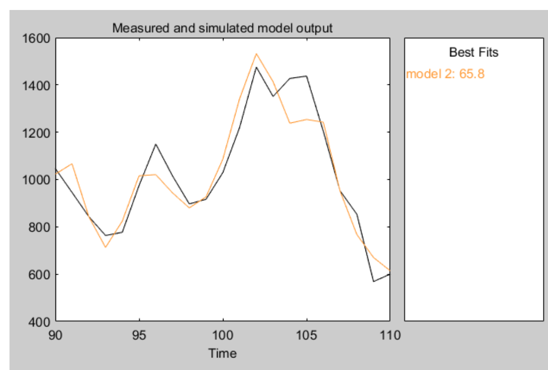
This model has to respond well during the shocks and difficult situations. In figure 5.27a there is a sudden rise and fall from 90 to 110 and this range have been selected to test the response of the model in shock condition. In figure 5.27c the result of such a test against a shock is shown. As one can see the fitting is 65.8% that returns a better result in compare to the fitting of the model output over the whole data. It shows that response of the model to shocks can be a reliable respond.



a



b



c

Figure 5.27:" a" is the model response to training data. "b" is the model response to all dada. "c" is the model response to a shock.

5.4.3 3rd model

Refer to equation 5.1 this model is based on the idea of using selected indices by correlation analysis of them against BCTI to create a model. following is the model:

Equation 5.6

$$\text{BCTI} = f(\text{Var2, var8, var11, var13, var21})$$

With replacing the variables with their corresponding dataset names model 5.6 will be

Equation 5.7

$$\text{BCTI} = f(Brent\ Crudeoil\ Price, LIBOR\ Interest\ Rates, HFO\ 380\ Avarage\ Price,$$
$$Industrial\ Production\ China, Gold\ Price\ )$$

Fewer variables suggested by the statistical analysis is used as input (in compare with the other two models) and represented by function f of the model and the BCTI is a dependent variable. The idea is to remove datasets with weak correlation coefficient and to remove data with more constant or very turbulent condition to develop a new model. Therefore model 3 is based on the most relevant variables selected by correlation coefficient and statistical analysis. Earlier in chapter 3.2 during the literature review, we mentioned that previously some of the researchers state to have less variable to develop a model suggest more flexibility and therefore a better fitting to the tested data. To some extent, this model is to test aforesaid idea regarding fewer variables also. Figure 5.28 represent stacked plot of selected datasets used for model 3. On the left are all available time series data and on the right training data have been plotted. The last 6 months of time series of the data have been neglected, as to have a percentage of unbiased available data for further evaluation reasons.

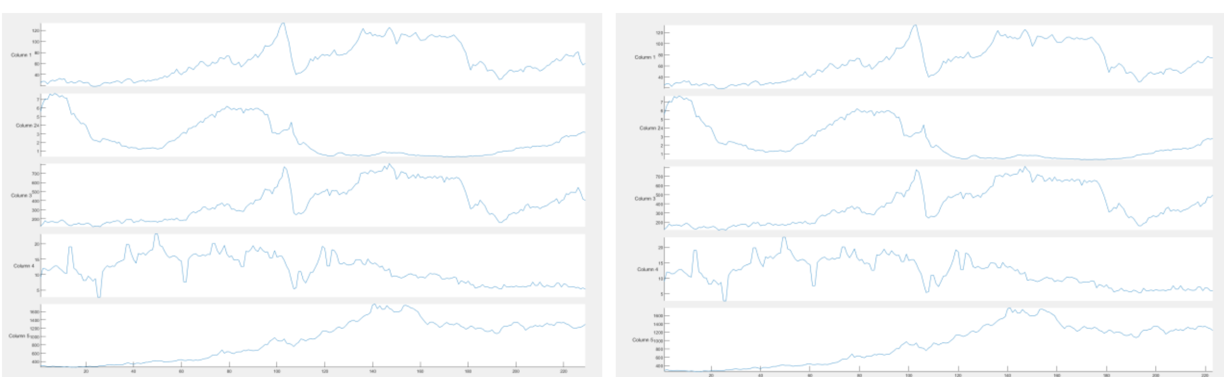
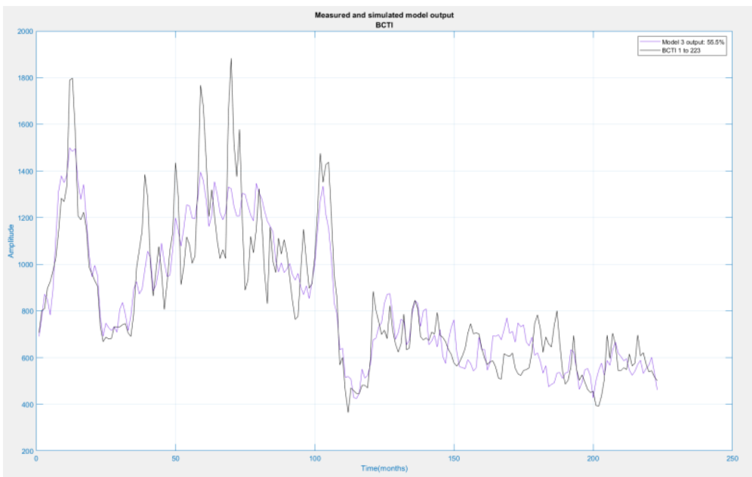
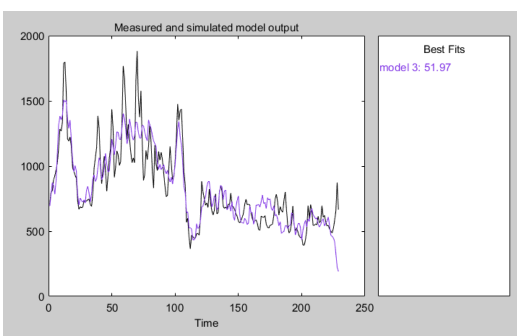


Figure 5.28: Stack plot of datasets used for model 3, plot of all data on the left and plot of training data on the right. (each column respectively represents an index in model)

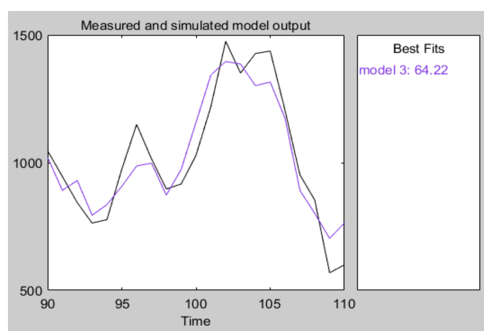
Model response to the training data can be seen in Figure 5.29a. The model response to the given data is not very good and in comparison, to the model one and model two, fitting to the training data results are less accurate. Model fitting to the system training data is 55.5 %. the fitting of the model to the training data is acceptable to continue testing the model over all data. In Figure 5.29b the result of the test of the model fitting against all data can be seen. Model fitting to all data is 51.97 % which is only 3.53% less than the fitting to training data. This result is not good in comparison with the first two models but still is an acceptable result. This model has to respond well during the shocks and difficult situations. in figure 5.29a there is a sudden rise and fall from 90 to 110 and this range have been selected to test the response of the model in case of a shock. In figure 5.28c the result of such a test against a shock is shown. Fitting of the model is 64.22 % that returns a better result in compare to the fitting of the model output over the whole data. This result is surprisingly good in comparison with the results obtained from the test of the model over the whole data. Results of the fitting of the model 3 over the test conditions to some extent reject the hypothesis of having fewer variables results in the construction of a better model as mentioned before.



a



b

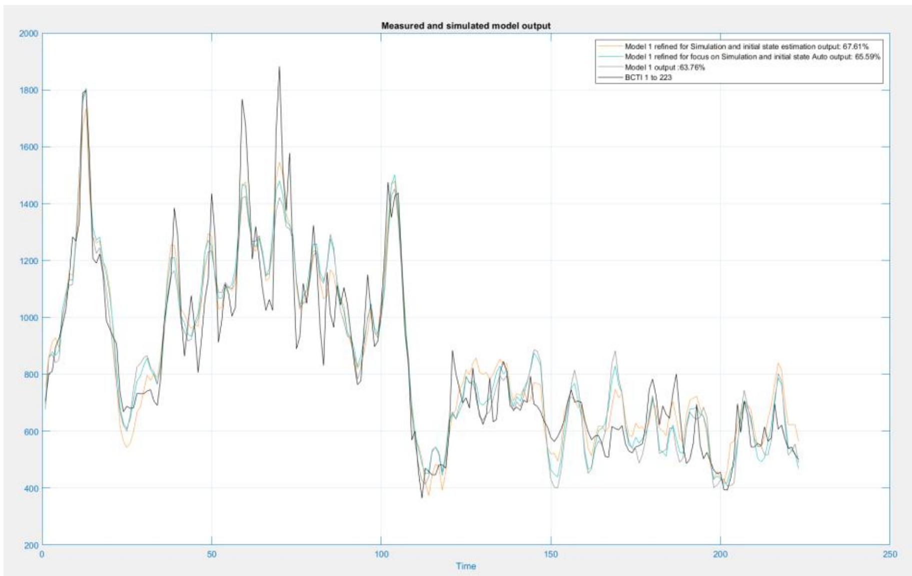


c

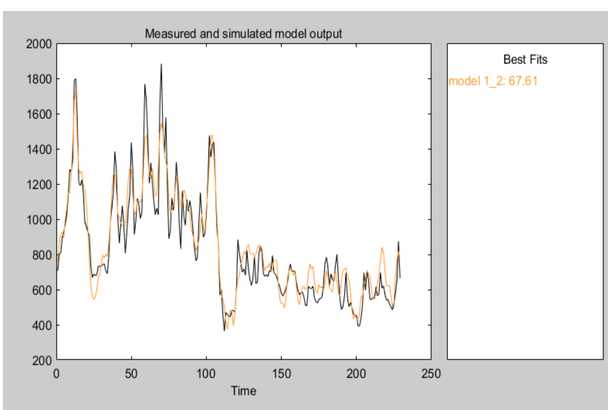
Figure 5.29 : The top model is the response to training data. Bottom left model is the response to all dada. The bottom right model is the response to a shock.

5.4.4 Refining the best model

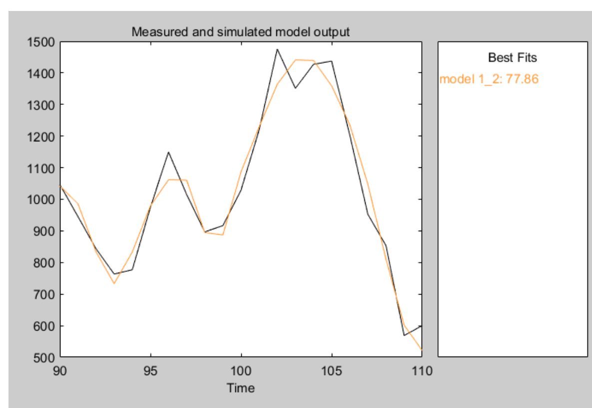
Previously in chapter 5.4 three models constructed based on different hypothesis and those models were checked against real data in different conditions. Model one fitting to data is the best amongst three according to the Table 5.2. Therefore, this model will be used for further refining of the model to find out the best model for forecasting of the BCTI. Since N4SID calculations are very complicated, without the help of the computer, it takes a lot of time. Models fittings against the data in chapter 5.4.1, 5.4.2 and 5.4.3 were tested automatically for the preselection of N4SID input settings. MATLAB™ does these calculations automatically but let the user change settings for a better result. Based on this information model one was manually refined by the change of initial condition of N4SID system identification method in MATLAB™ software. The result suggests a better output for the model one as can be seen in Figure 5.30



a



b



c

Figure 5.30: Top, model response to training data. Bottom left, model response to all dada. Bottom right, model response to a shock.

Model response to the training data can be seen in figure 5.30a. The model response to the given data is very well and shows a better fitting in comparison with the base condition. model fitting to the training data is 67.61% that show a significant improvement from the base condition output that is 63.76%. The fitting of the model to the training data is acceptable to continue testing the model over all data.

In figure 5.30b the result of the test of the model fitting against all data can be seen. Model fitting to all data is 67.61% and shows no change from the fitting of the refined model 1 over the training data. This result is the best result that was achieved in this thesis until now.

This model has to respond well during the shocks and difficult situations. In figure 5.30a there is a sudden rise and fall from 90 to 110 and this range has been selected to test the response of the model in harsh condition. In figure 5.30c the result of such a test against a shock is shown. As one can see the fitting is 77.86 % that returns a less fitting result in compare to the fitting of the model one output over the shock period of the data. This model initial setting returns a better result for forecast but a less favourable result during the shocks and sudden changes of BCTI.

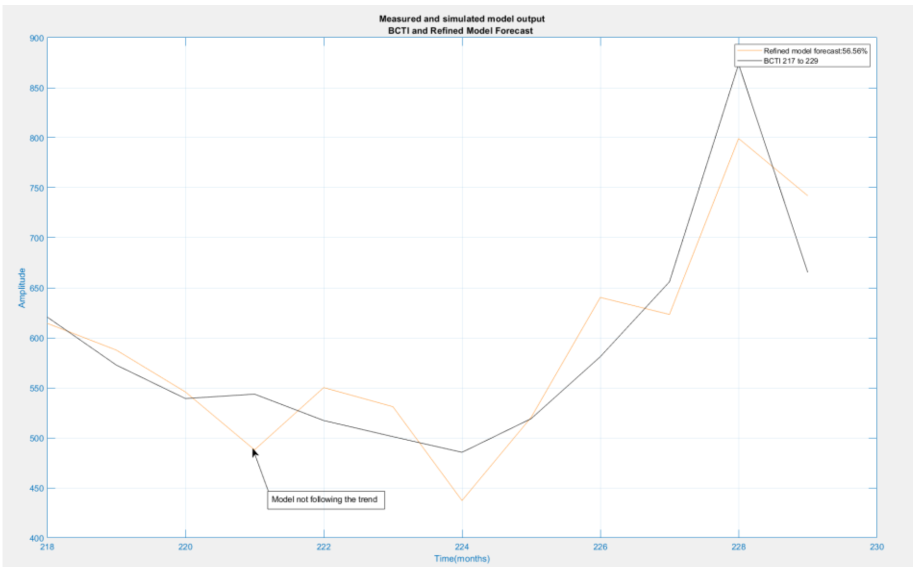


Figure 5.31: Refined model 12 months forecast

Now that the first model is refined it will be used to forecast a 12-months period from 217 to 229 that is from 01 February 2018 to 01 February 2019. The initial condition of the N4SID system for the refined model was set and training data have been set to 217 observations. Refined model output can be found in figure 5.31. As one can see model fitting is 56.56% to the real data and model follows the trends very well. As it has been marked over the plot of in figure 5.31 model in one point delivers a different trend from real observations that means model fitting to the data in one month out of 12 months, made a wrong assessment about the trend. Therefore, the outcome of this model is acceptable for a one-year forecast.

5.5 Choice of model

Models in this thesis are based on indices selected by statistical analysis. Therefore, indices selected for the purpose of the modelling in this thesis are independent of BCTI but still correlated to its rates. Chantzaras. A (2018) on his thesis suggests that even fitting of the model that is not based on statistical analysis may be better to the actual data but in the certain condition such as shocks, their behaviour is not reliable.

For a better understanding of the work of this thesis regarding modelling, all the models are compared in Table 5.2.

All models showed very similar fitting when they are used against both training and all data especially in case of the refined model 1 the results are exactly the same, that is a surprisingly good result. Model 2 showed a good fitting results however in term of following the trend patterns in the plot its response to the changes of training data was not as well as model one and was not good in general. Model 3 followed the trends for training data but as we see in Figure 5.29b model showed a strange behaviour at the end of the time series and followed a completely different pattern in compare with the plot of the actual data and therefore its trend fitting is not acceptable for all data.

Model 2 and 3 performances are good in general but since their trend pattern is not very well fitted to the pattern of the BCTI real data we leave them here and proceed with Model 1.

Model 1 showed a good fitting and trend follow up and, therefore, the author decided to try to refine N4SID parameters to get a better fitting to the data for this model. And later refined model tested to forecast a one-year period or 12 observation and checked the outcome against BCTI rates. However, refined model fitting was better both in case of the training data and all data in comparison with model one but the model outcome was less accurate in case of the defined shock period.

All of the above models are successful to an acceptable level of fitting of prediction with actual observations. BCTI receives impact from other areas that have not been discussed in this thesis, therefore, these models can't predict 100% fitting to the actual data. But when suggesting a model, the model that has best fitting to the trends and has the best outcome during the shocks is a better model for the forecast of the market in general, and BCTI, in particular, that is subject of the discussion in this thesis.

From all above writer of this thesis suggests the refined Model 1 of this thesis since it returns a better result in case of the shocks and better fitting in general to the trends and figures. Model 1, however, returns better results in comparison with refined model 1 to the selected shock period but its outcome is not as accurate as refined model 1 in other cases. Please note that refined model 1 and model 1 are both equation 5.3 and the only difference is N4SID initial setup of the software to start its calculations.

Table 5.2: Summary of the resulted fittings of the obtained model against Shocks, the training and all the available data and trend patterns fittings.

Row	*Data*	*Training data fitting*	*All data fitting*	*Shock fitting*	*Trends training*	*Trend All data*	*Forecast fitting*
1	Model 1	63.76	63.24	79.45	Very good	Very good	N.A
2	Model 2	61.92	61.78	65.8	Not good	Good	N. A
3	Model 3	55.5	51.97	64.22	Good	Not acceptable	N. A
4	Refined Model (Model 1)	67.61	67.61	77.86	Very good	Very good	56.56

6 Conclusions and future work

6.1 Conclusions

In chapter one of this thesis after the introduction, the main question of the thesis posed followed by four sub-questions. Later in chapter 2 and 3 previous work related to the main and sub-questions studied and also different economical and commodities and markets briefly explained and their relation to the main and sub-questions of this thesis discussed. In chapter 4 methodology of this thesis explained. In chapter 5.2 of this thesis, a series of statistical analysis was carried out to find out if selected indices in chapter 2 and 3 are good choices to answer the main question. This was done by testing the correlation between selected indices against BCTI. This analysis further proved that however selected indices in chapter 2 and 3 have an impact on BCTI either directly or indirectly but only some of them have strong enough correlation with BCTI to be used for the next chapter while modelling the BCTI. Those indices that were independent and correlated to BCTI were used in chapter 5 for BCTI modelling and their behaviour and relation with real data was tested. In chapter 5.4 three models were developed by selected indices from chapter 5.3. The first model used all of the selected data sets together to shape a model and the next two models were shaped by choice of a set of most related indices. The selected indices were used as input for the models and BCTI was the output of the models. The finding of this procedure revealed the fitting of models constructed by the selected indices is chapter 5.3 with BCTI is very high. All models showed a good fitting against both training and all data and also their behaviour was tested against the shocks in BCTI rates by actual BCTI rates. Results of the models against shocks also was very good and, in all cases, even better than fitting for training and all data. Models also showed a good fitting for BCTI trend changes and only in few cases their output was not following the general trends. Later the first model refined and its forecasting ability tried out and an acceptable result was the output of that tests.

Form all above as to answer the main question of this thesis, i.e. *“Is Baltic Clean Tanker Index (BCTI) related to main macro and microeconomics indices and is there a possibility of modeling of Baltic Clean Tanker Index (BCTI) to predict its future rate based on the that indexes?”,* we can conclude from analysis and finding of this thesis that there is a relation between BCTI and economic indicators and indices that are related to clean tanker industry such as oil and bunker prices and those which are completely independent and not related to clean tanker industry such as gold price. This thesis findings also proved that there is a feasibility of constructing a model based on those indices and indicators. The model then can be used to predict and examine behaviour and rationality of the monthly BCTI. This model also can be used during the extreme economical condition to predict BCTI rates trends.

6.2 Limitations and future work

The main limitation while working on this topic was to obtain data. Reliable sources of dada mostly need a subscription and they were only a few databases accessible from the university and obtained data have not the same length. Observation duration was different for different data sets and therefore a trimming and adjustment was required to make all the datasets the same length and the same duration. Software used for this thesis needs a subscription that was provided by EUR for research purpose. BCTI rarely been investigated before, that made the author to find new relations between commodities such as **Gold Price** and **BCTI** that **have not been studied before** to the knowledge of the author.

N4SID modelling method was used before against BDI and returned accepted results but this method was not used much for modelling of shipping related indices. This method was used in several other topics and returned good results. Based on that, this method was used in the thesis.

Based on the finding of this thesis for future work it is recommended to re-examine the final results with up to date data. Finding other indices and indicators to test with these models are also advised. Testing of more models with the change of the same indicators and models of this thesis also may return different results suitable for different purposes such as risk management.

Input-output systems have not been used much for research purpose in the shipping industry and it is good to do more work in this field. Baltic Dirty Tanker Index and Baltic Dry Index have been the topic of many kinds of researches but research in the area of product and chemical tankers are fewer and this area is a relatively good area for future research.

In addition, this research was based on limited resources of time and data, therefore, more detail and in-depth research based on this work can reveal more efficient ways of modelling of shipping industry indices and indicators.

7 Appendices

Appendix A
Appendix

A.1 Graphical Representation of Used Data

In this section graphical representations of all the used data / indexes in this thesis can be find. All the data have been plotted from 01-01-2000 till the 01-02-2019 with monthly intervals.

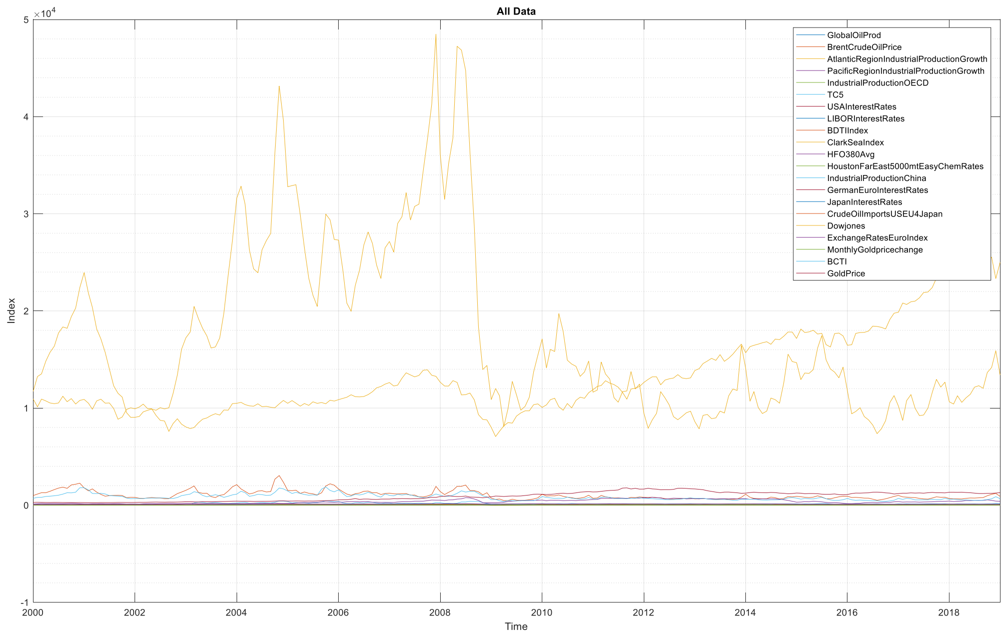


Figure A. 1: Graphical Representation of the All data (Sources: Clarkson Research™ *(2019) & Goldprice.org (2019)).*

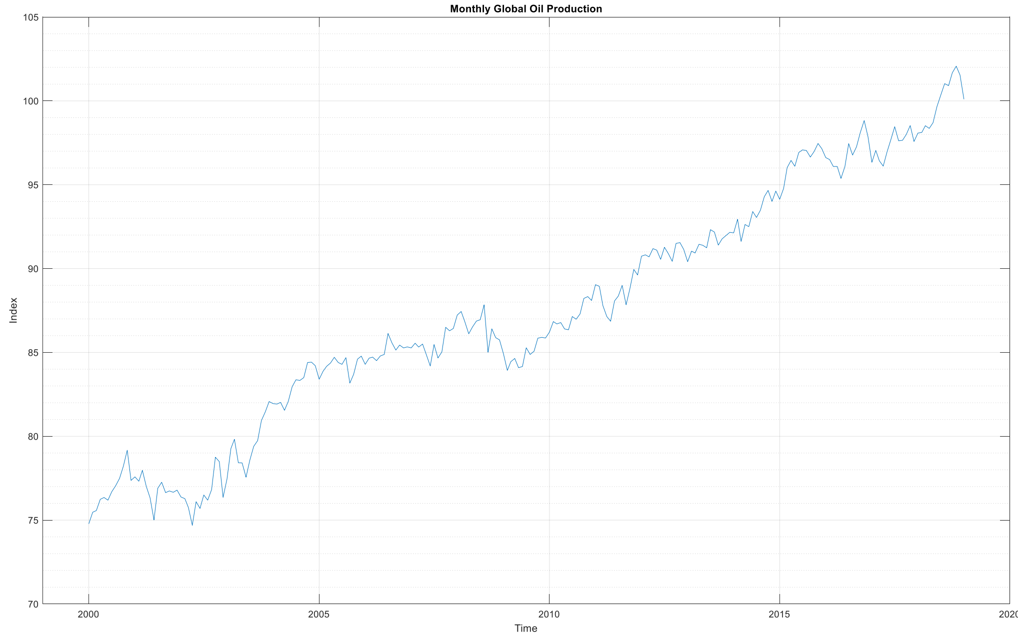


Figure A. 2: Graphical Representation of the monthly Global Oil Production (Source: Clarkson Research™ *(2019))*

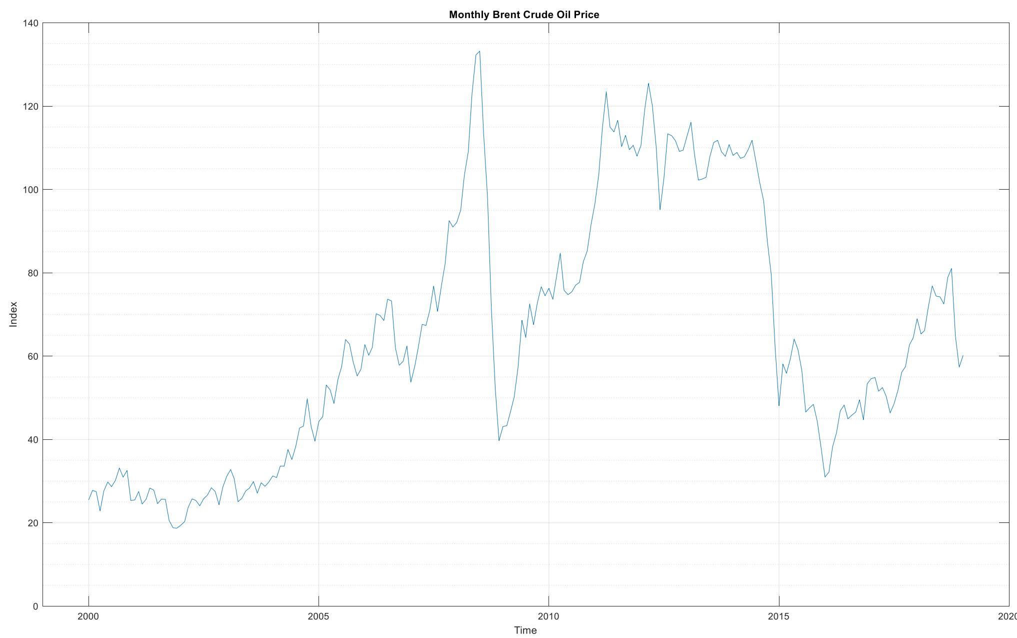


Figure A. 3: Graphical Representation of the monthly Brent Crude Oil Price (Source: Clarkson Research™ *(2019))*

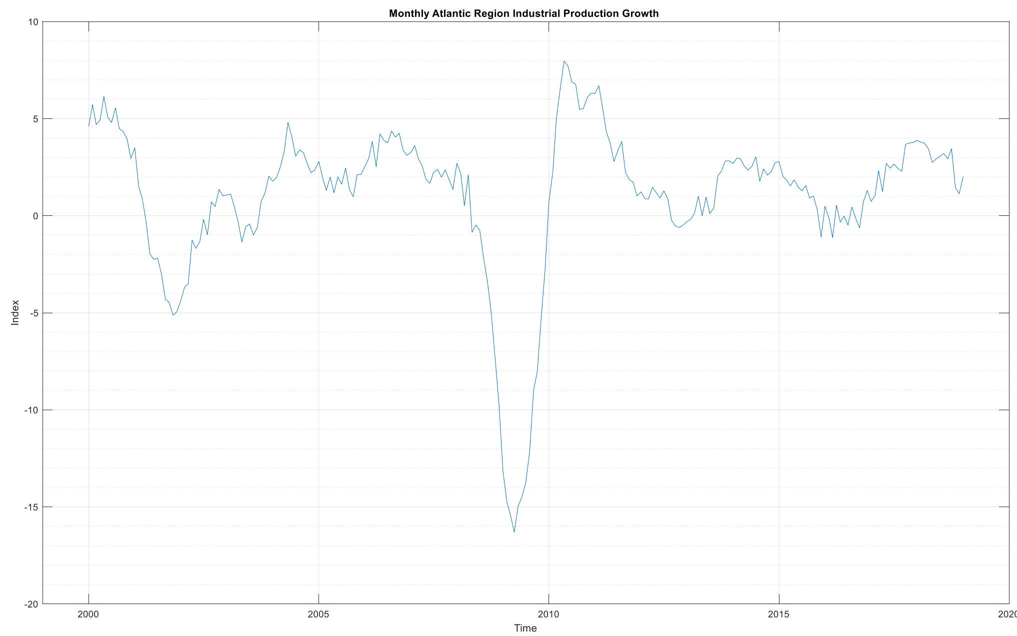


Figure A. 4: Graphical Representation of the monthly Atlantic Region Industrial Production Growth (Source: Clarkson Research™ *(2019))*

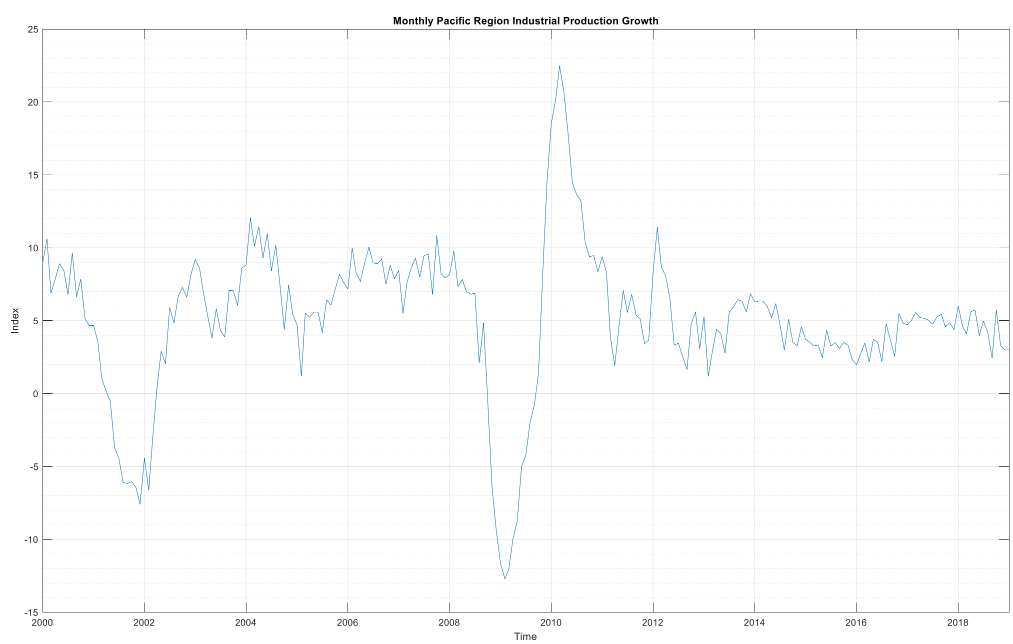


Figure A. 5: Graphical Representation of the monthly Pacific Region Industrial Production Growth (Source: Clarkson Research™ *(2019))*

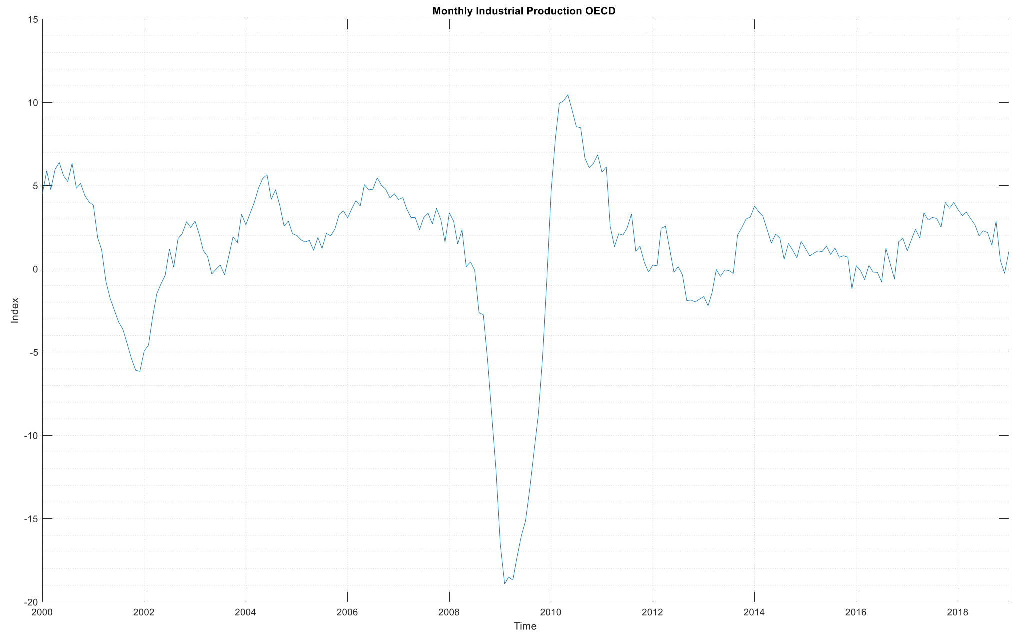


Figure A. 6: Graphical Representation of the monthly Industrial Production OECD (Source: (Source: Clarkson Research™ *(2019))*

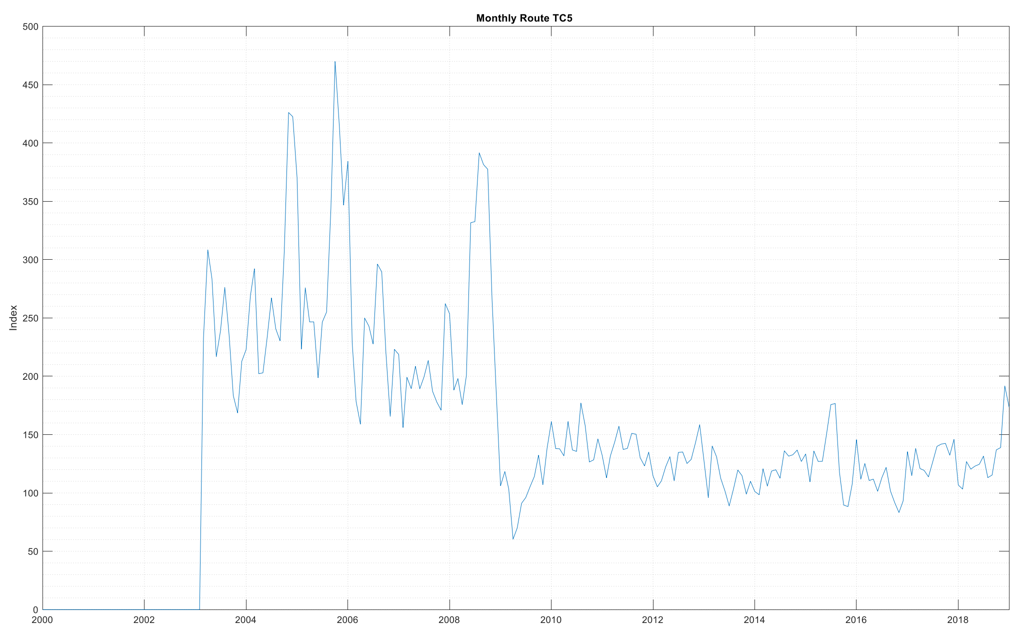


Figure A. 7: Graphical Representation of the monthly Route TC5 (Source: Clarkson Research™ *(2019))*

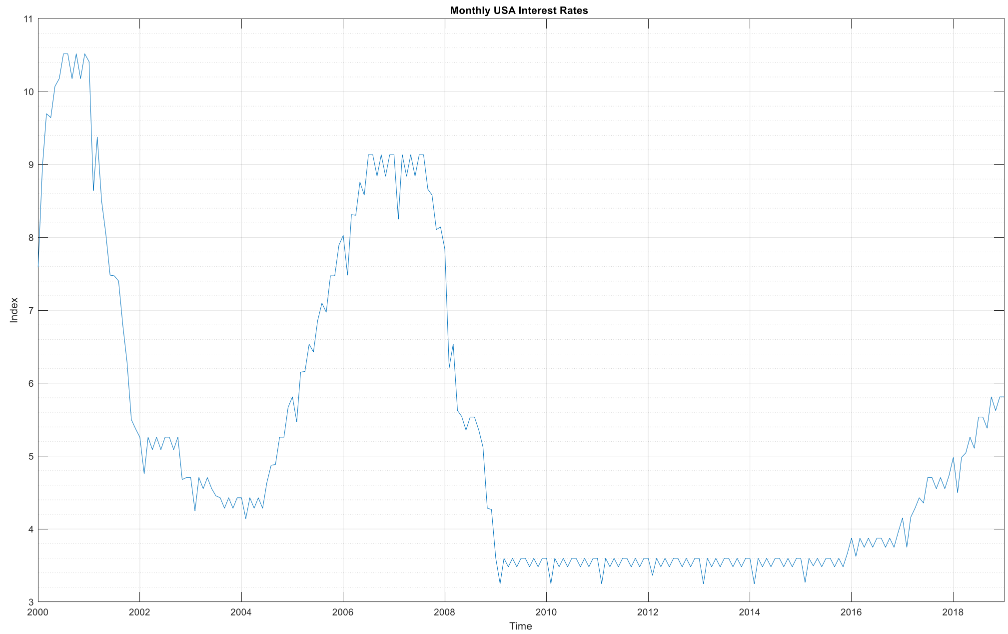


Figure A. 8: Graphical Representation of the monthly USA Interest Rates (Source: Clarkson Research™ *(2019))*

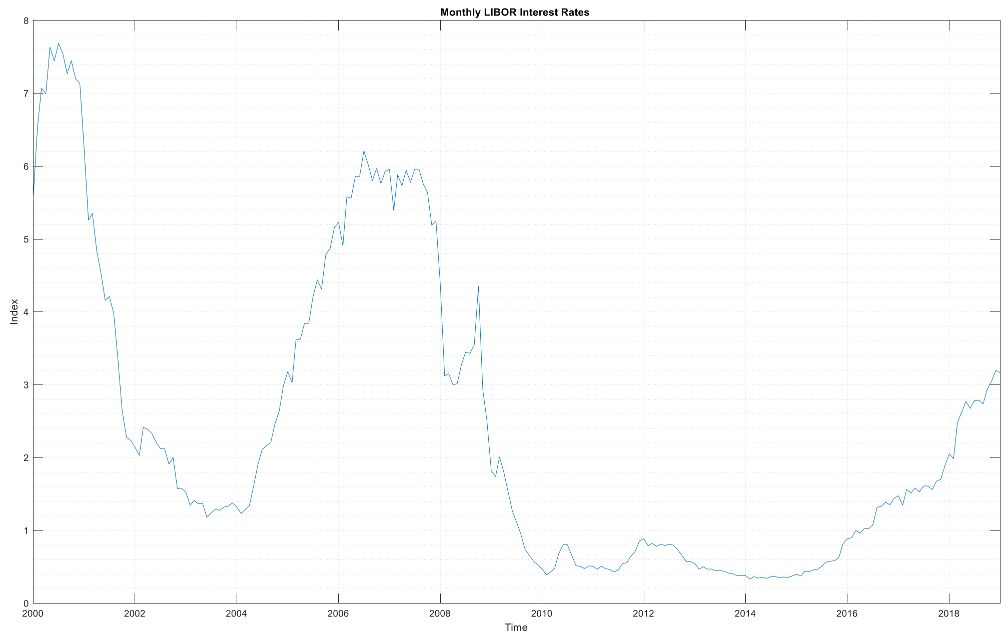


Figure A. 9: Graphical Representation of the monthly LIBOR Interest Rates (Source: Clarkson Research™ *(2019))*

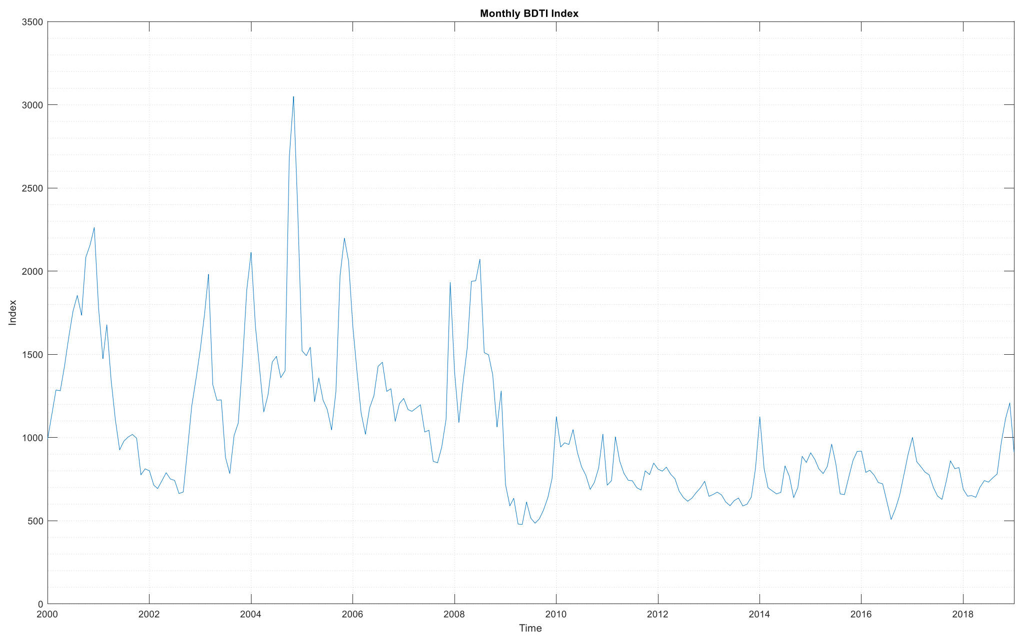


Figure A. 10: Graphical Representation of the monthly BDTI Index (Source: Clarkson Research™ *(2019))*

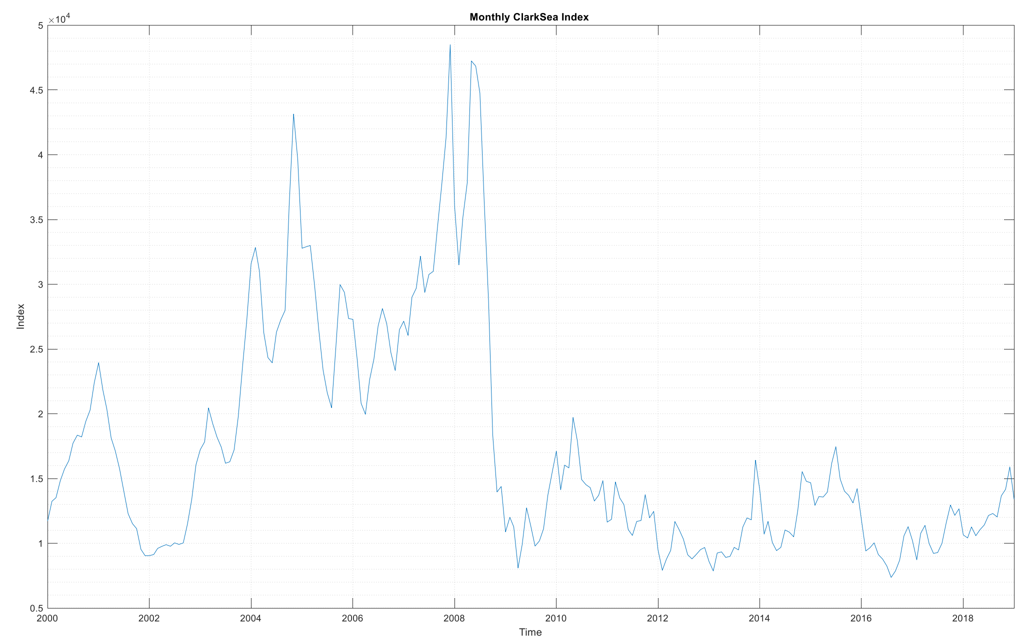


Figure A. 11: Graphical Representation of the monthly ClarkSea Index (Source: Clarkson Research™ *(2019))*

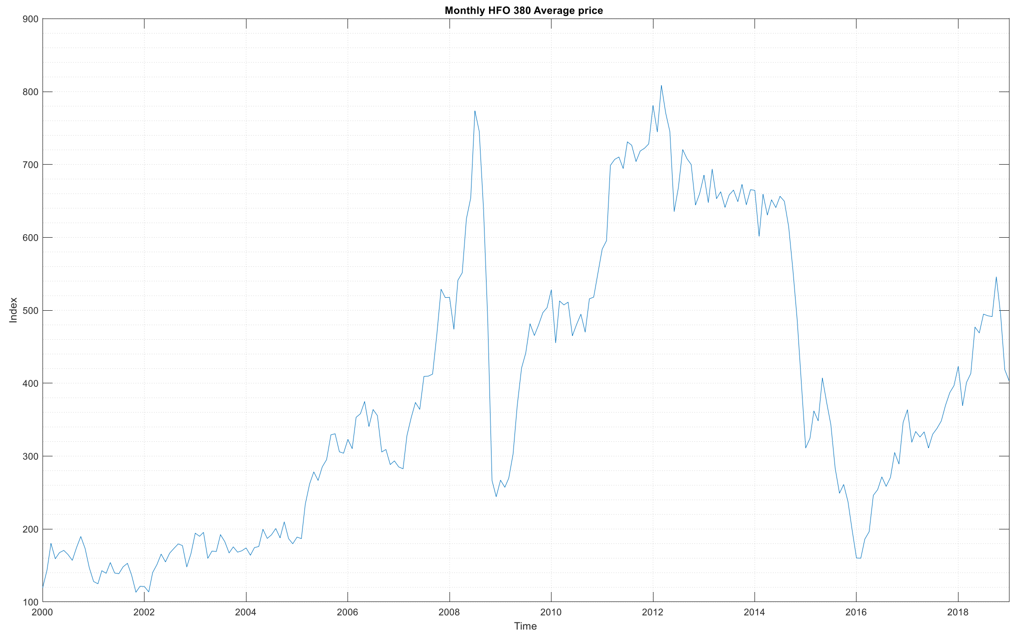


Figure A. 12: Graphical Representation of the monthly HFO 380 Average price (Source: Clarkson Research™ *(2019))*

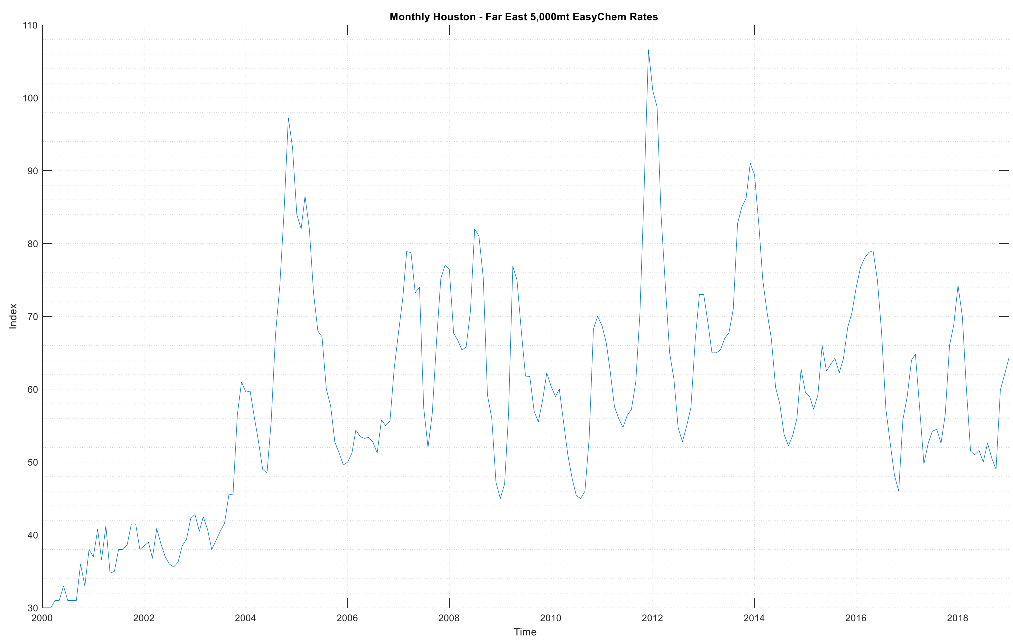


Figure A. 13: Graphical Representation of the monthly Houston - Far East 5,000mt EasyChem Rates (Source: Clarkson Research™ *(2019))*

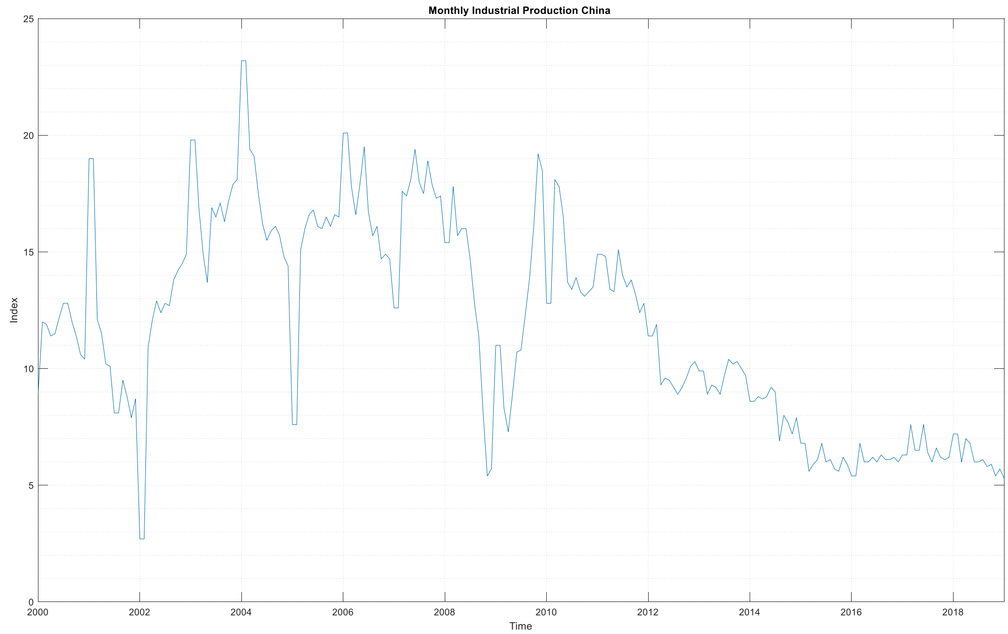


Figure A. 14: Graphical Representation of the monthly Industrial Production China (Source: Clarkson Research™ *(2019))*

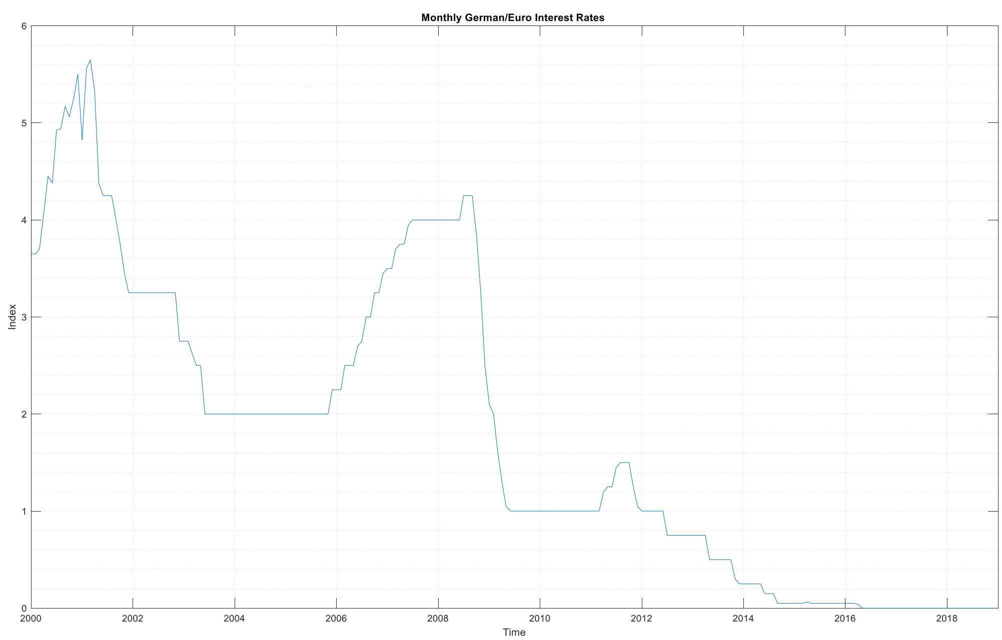


Figure A. 15:Graphical Representation of the monthly German/Euro Interest Rates (Source: Clarkson Research™ (2019))

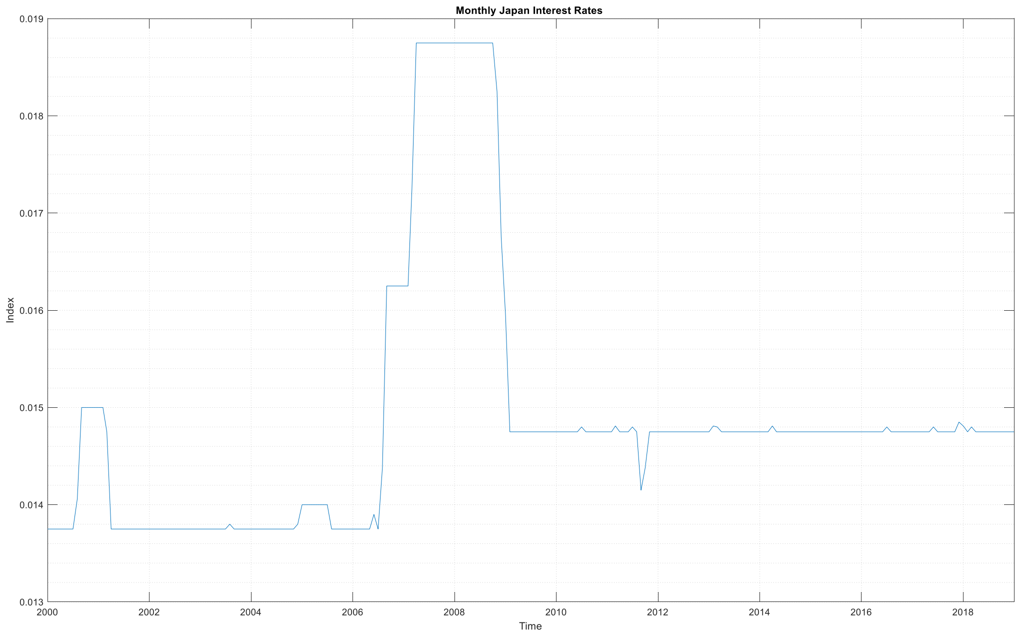


Figure A. 16: Graphical Representation of the monthly Japan Interest Rates (Source: Clarkson Research™ *(2019))*

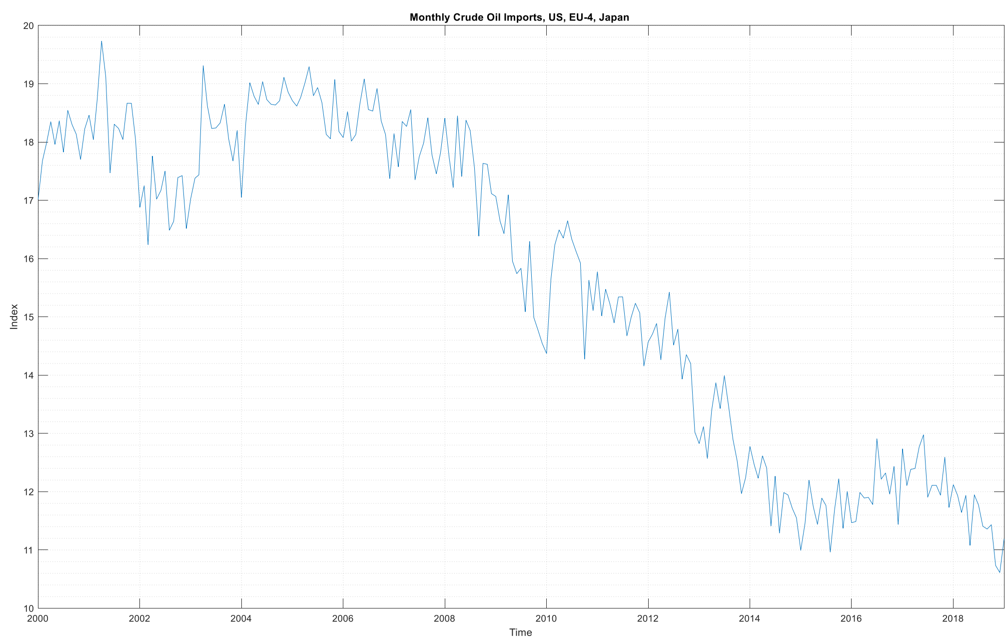


Figure A. 17: Graphical Representation of the monthly Crude Oil Imports, US, EU-4, Japan (Source: Clarkson Research™ *(2019))*

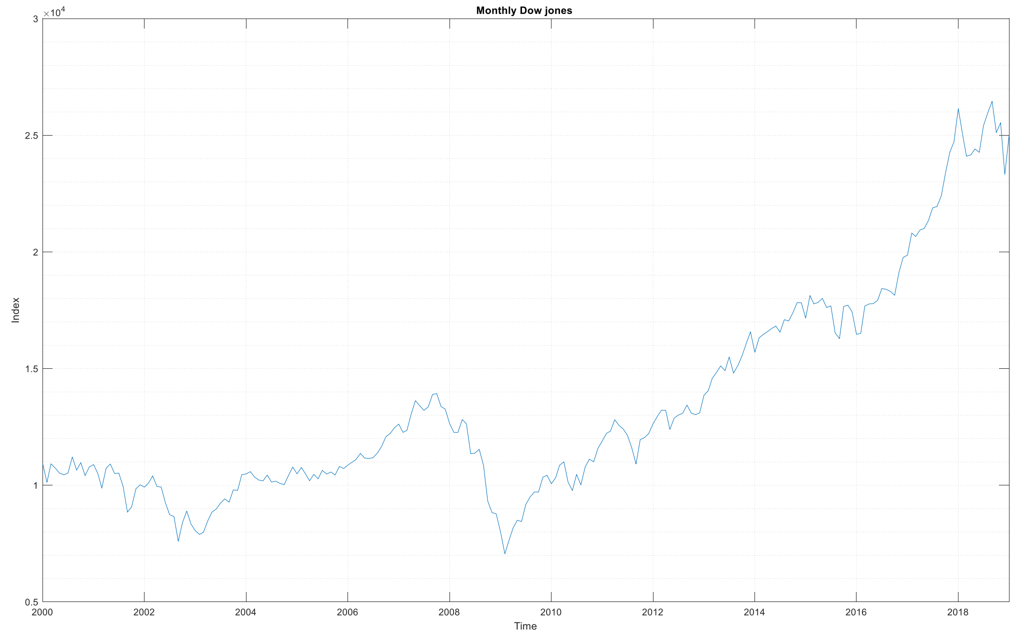


Figure A. 18: Graphical Representation of the monthly Dow jones (Source: Clarkson Research™ *(2019))*

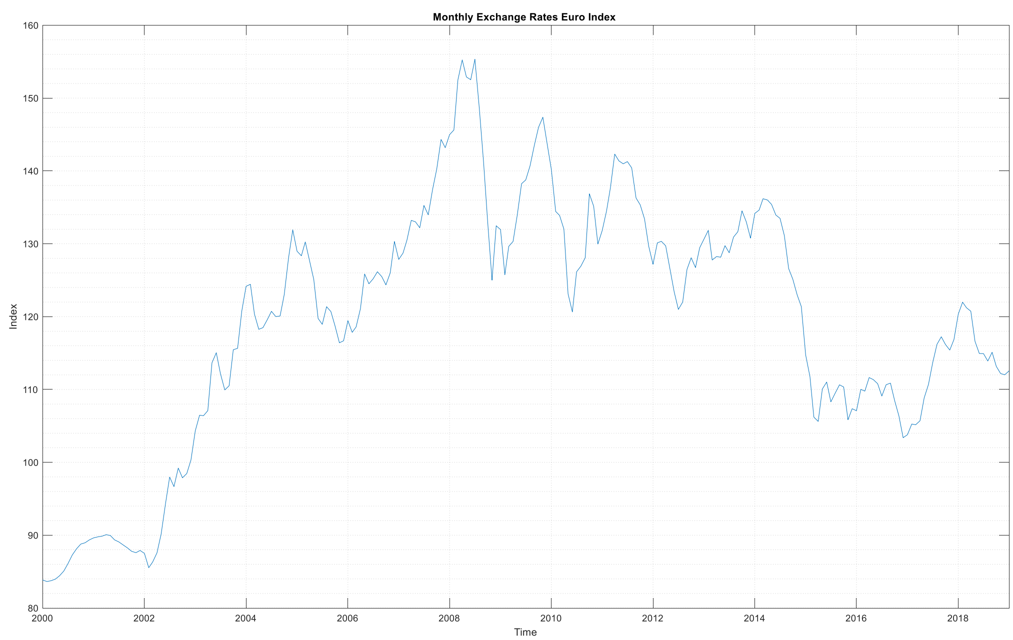


Figure A. 19: Graphical Representation of the monthly Exchange Rates Euro Index (Source: Clarkson Research™ *(2019))*

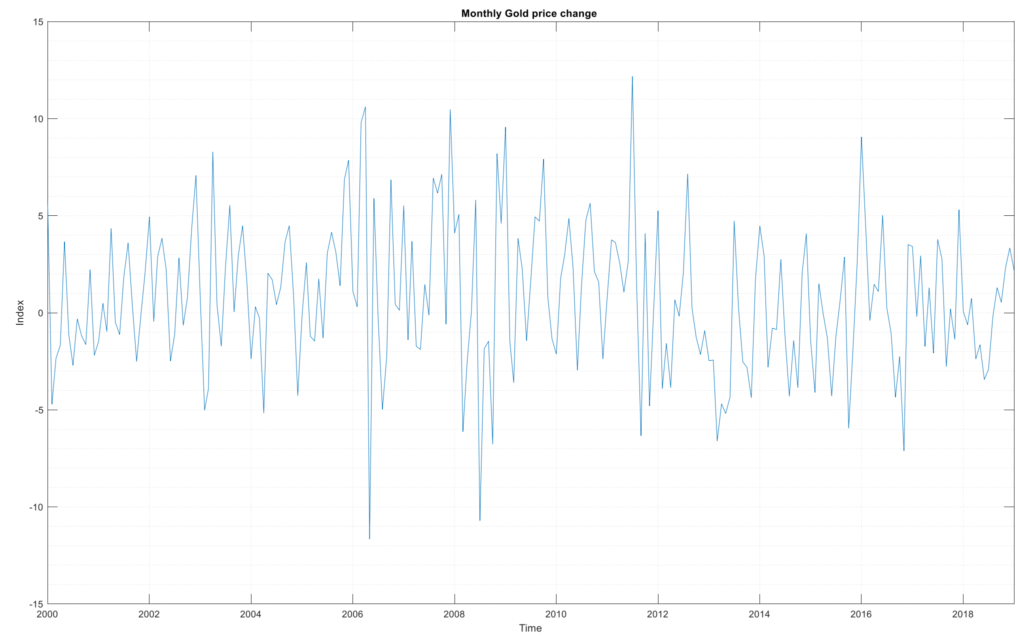


Figure A. 20: Graphical Representation of the monthly Gold price change (Source: Goldprice.org (2019)).

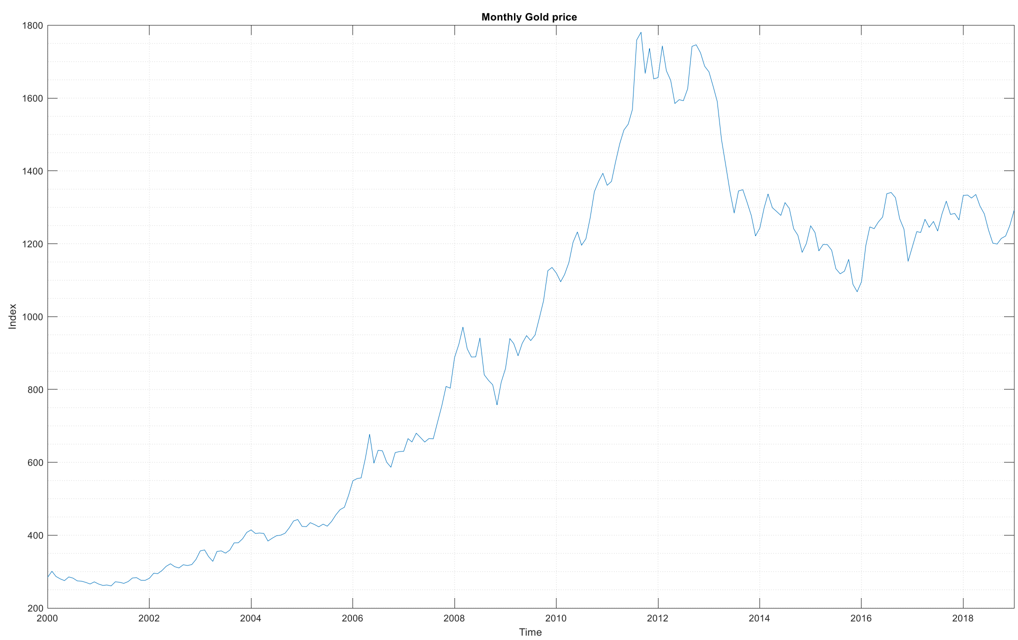


Figure A. 21: Graphical Representation of the monthly Gold price (Source: Goldprice.org (2019)).

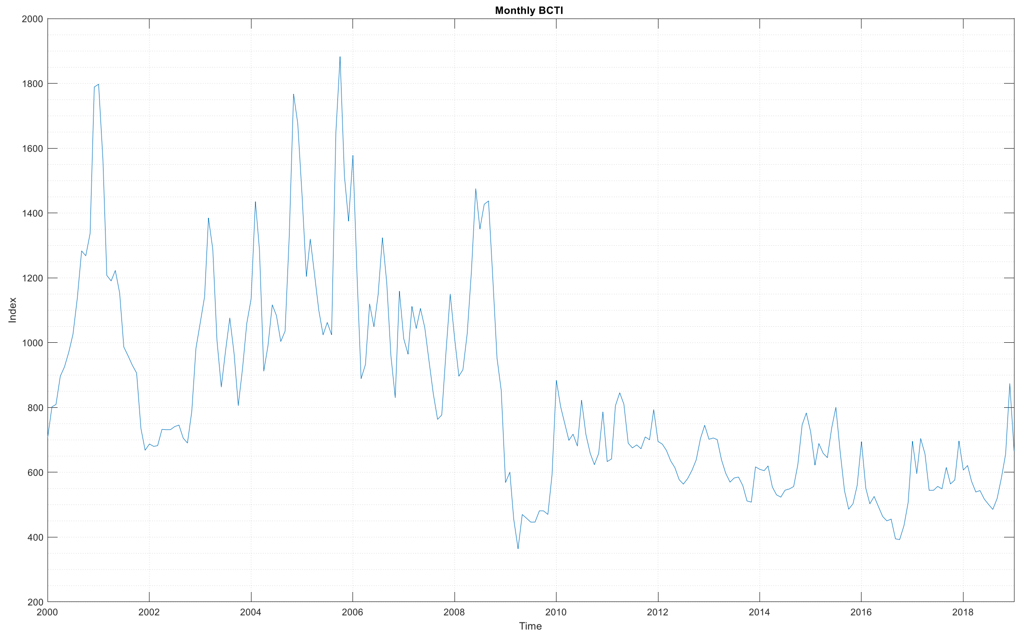


Figure A. 22: Graphical Representation of the monthly BCTI (Source: Clarkson Research™ *(2019))*

8 Bibliography

ABS, A. B. o. S., 2002. Surveyor. *winter 02*, Jan, p. 4:9.

Alizadeh, A. H. a. N. K. N., 2011. Dynamics of the term structure and volatility of shipping freight rates. *Journal of Transport Economics and Policy (JTEP),* 45(1), pp. 105-128.

Baltic Exchange, I. S. L., 2019. *Guide to Market Benchmarks.* [Online]
Available at: https://www.balticexchange.com/dyn/_assets/_forms/guide-to-market-benchmarks.shtml
[Accessed 12 July 2019].

Batrinca, G. &. C. G. &. S. I., 2013. *APPLICATION OF AUTOREGRESSIVE MODELS FOR FORECASTING THE BALTIC EXCHANGE DRY INDEX.* Constanta, Analele Universitatii Maritime Constanta.

Bloomberg a, 2019. *Content and Data.* [Online]
Available at: https://www.bloomberg.com/professional/solution/content-and-data/
[Accessed 12 Jul 2019].

Boomberg b, 2019. *Bloomberg Professional Services.* [Online]
Available at: https://www.bloomberg.com/professional/solution/bloomberg-terminal/
[Accessed 25 July 2019].

BP, 2019. *Statistical Review of World Energy / Oli.* [Online]
Available at: https://www.bp.com/en/global/corporate/energy-economics/statistical-review-of-world-energy/oil.html
[Accessed 14 August 2019].

Cambrige Dictionary, 2019. *statistics.* [Online]
Available at: https://dictionary.cambridge.org/dictionary/english/statistics
[Accessed 28 07 2019].

Chantzaras, A. (., 2018. *Statistical Analysis and Modeling of the Baltic Dry Index (BDI), MSc Thesis.* Rotterdam: Erasmus University Rotterdam.

Chen, H. &. D. P. .. 8.-8. 1., 2015. *Modelling and prediction of stock price dynamics using system identification methodology based on a popularly used data analysis technique..* s.l., s.n.

Chris Brooks, 2014. *Introductory Econometrics for Finance.* 3rd ed. Cambridge: Cambridge University Press.

Christian Lyzell, 2009. *Initialization Methods for System Identification.* [Online]
Available at: http://www.diva-portal.org/smash/get/diva2:277019/FULLTEXT02
[Accessed 10 07 2019].

DATAHUB.IO, 2019. *Gold Price.* [Online]
Available at: https://datahub.io/core/gold-prices
[Accessed 25 July 2019].

David R. Anderson, D. J. T. A. J. D. K. M., 2012. Time Series Analisis and Forecasting. In: *An Introduction to Management Science:Quantitative Approaches to Decision Making.* Mason: Cengage Learning, p. 717.

European Central Bank, 2019. *Daily nominal effective exchange rate of the euro.* [Online]
Available at:
https://www.ecb.europa.eu/stats/balance_of_payments_and_external/eer/html/index.en.html#targetText=The%20nominal%20effective%20exchange%20rates,on%20average%2C%20for%20%E2%82%AC1.
[Accessed 15 August 2019].

Favoreel W., D. M. B. V. O. P., 2000. Subspace state space system identification for industrial processes. *Journal of Process Control,* 10(2), pp. 149-155.

Fenwick, &. W. f. t. E. C., 2007. *LEGAL AND ECONOMIC ANALYSIS OF TRAMP MARITIME SERVICES - ANNEX 1 GLOSSARY OF KEY SHIPPING TERMS.* [Online]
Available at: http://ec.europa.eu/competition/sectors/transport/reports/tramp_annex1.pdf
[Accessed 15 August 2019].

Fotis Papailias, D. D. T. J. L., 2016. *The Baltic Dry Index: cyclicalities, forecasting and hedging strategies.* [Online]
Available at: https://link.springer.com/article/10.1007/s00181-016-1081-9
[Accessed 09 07 2019].

Ganti, A., 2019. *Correlation Coefficient.* [Online]
Available at: https://www.investopedia.com/terms/c/correlationcoefficient.asp
[Accessed 31 July 2019].

Goldprice.org, 2019. *Gold Price History.* [Online]
Available at: https://goldprice.org/gold-price-history.html
[Accessed 22 July 2019].

Goulas, L., 2010. *Forecastin freigt rates, MSc Thesis.* Pireaues: University of Pireaus.

Grace's Guide, 2016. *Henry Frederick Swan.* [Online]
Available at: https://www.gracesguide.co.uk/Henry_Frederick_Swan
[Accessed 07 July 2019].

Hachicha, S. &. K. M. &. C. A., 2014. N4SID and MOESP Algorithms to Highlight the Ill-conditioning into Subspace Identification.. *International Journal of Automation and Computing,* 11(1), pp. 30-38.

Holodny, E., 2015. *How oil flows in and out of every major region around the world.* [Online]
Available at: https://www.businessinsider.com/bp-map-world-oil-trade-movements-2014-2015-6?international=true&r=US&IR=T
[Accessed 14 August 2019].

ICS, I. C. o. S., 2019. *Shipping and World Trade.* [Online]
Available at: http://www.ics-shipping.org/shipping-facts/shipping-and-world-trade
[Accessed 06 May 2019].

Investopedia, 2019. *Organization Of Petroleum Exporting Countries (OPEC).* [Online]
Available at: https://www.investopedia.com/terms/o/opec.asp
[Accessed 21 07 2019].

Investopedia, 2019. *S&P 500 Index – Standard & Poor's 500 Index.* [Online]
Available at: https://www.investopedia.com/terms/s/sp500.asp
[Accessed 22 07 2019].

Investopedia, 2019. *WTI.* [Online]
Available at: https://www.investopedia.com/terms/w/wti.asp
[Accessed 21 07 2019].

Keller, G., 2014. *Statistics for management and economics.* 10 ed. s.l.:Cengage Learning.

KhanAcademy.org, 2019. *Mean absolute deviation (MAD).* [Online]
Available at: https://www.khanacademy.org/math/cc-sixth-grade-math/cc-6th-data-statistics/cc-6-mad/v/mean-absolute-deviation#targetText=Mean%20absolute%20deviation%20(MAD)%20of,in%20a%20data%20set%20are.
[Accessed 18 August 2019].

MacKiensy, &., 2019. *Clean product tanker.* [Online]
Available at: https://www.mckinseyenergyinsights.com/resources/refinery-reference-desk/clean-product-tanker/
[Accessed 14 August 2019].

Mark J. Perry, 2018. *Putting America's huge $20.5T economy into perspective by comparing US state GDPs to entire countries.* [Online]
Available at: http://www.aei.org/publication/putting-americas-enormous-20-5t-economy-into-perspective-by-comparing-us-state-gdps-to-entire-countries-2/
[Accessed 22 July 2019].

MathWorks, 2019. *std-Standard deviation.* [Online]
Available at:
https://nl.mathworks.com/help/matlab/ref/std.html#targetText=If%20A%20is%20a%20vector,standard%20deviation%20is%20a%20scalar.&targetText=By%20default%2C%20the%20standard%20deviation,is%20the%20number%20of%20observations.
[Accessed 22 July 2019].

McKinsey, &. C., 2019. *Global Energy Perspective 2019.* [Online]
Available at: https://www.mckinsey.com/industries/oil-and-gas/our-insights/global-energy-perspective-2019
[Accessed 14 August 2019].

Mike Ratcliffe, 1985. *Liquid gold ships : a history of the tanker, 1859-1984.* London: Lloyd's of London Press.

National Academy Press Washington, D., 1991. *Tanker Spills: Prevention by Design.* [Online]
Available at: https://www.nap.edu/read/1621/chapter/3#30
[Accessed 07 June 2019].

OECD.org, 2019. *About.* [Online]
Available at: http://www.oecd.org/about/
[Accessed 15 August 2019].

OECD, 2018. *Gross domestic product (GDP).* [Online]
Available at: https://data.oecd.org/gdp/gross-domestic-product-gdp.htm
[Accessed 21 07 2019].

Papailias, F. T. D. &. L., 2017. The Baltic Dry Index: cyclicalities, forecasting and hedging strategies. *Empirical Economics,* 52(1), pp. 255-282.

Petris, G. P. S. C. P., 2009. *Dynamic Linear Models with R..* s.l.:Springer Science & Business Media..

Petropedia, 2019. *Oil Price.* [Online]
Available at: https://www.petropedia.com/definition/7902/oil-price
[Accessed 21 July 2019].

Roy Batchelor, A. A. ,. V., 2007. Forecasting spot and forward prices in the international freight market. *International Journal of Forecasting,* 23(1), pp. 101-114.

Sahin, B. &. G. S. &. Ü. B. &. A. I., 2017. Forecasting Baltic Dry Index by using artificial neural network approach.. *Turkish Journal of Electrical Engineering & Computer Sciences,* 26(1), pp. 1673-1684.

Shrivastav Anuj, 2016. *INTERESTING FACTS ABOUT O&G INDUSTRY.* [Online]
Available at: https://oges.info/library/140591/DID-YOU-KNOW_-The-first-oil-tanker-steamer-was-bui
[Accessed 06 July 2019].

Shuangrui Fan, T. J. W. B., 2013. Forecasting Baltic Dirty Tanker Index by Applying Wavelet Neural Networks. *Journal of Transportation Technologies ,* Jan, 3(1), pp. 68-87.

Sin.Clarksons.net, 2019. [Online]
Available at: sin.clarksons.net
[Accessed 20 Mar 2019].

Timothy C. Winegard, 2016. *The First World Oil War P103.* NA ed. Toronto: University of Toronto Press.

UNCTAD, 2019. *Merchant Fleet.* [Online]
Available at: https://stats.unctad.org/handbook/MaritimeTransport/MerchantFleet.html
[Accessed 21 July 2019].

Van Overschee, D. M., 1994. N4SID: Subspace algorithms for the identi- fication of combined deterministic-stochastic systems. *Automatica,* 30(1), pp. 75-93.

Van Overschee, D. M., 2012. *Subspace identification for linear systems: Theory-Implementation-Applications.* s.l.:Springer Science & Business Media.

Vasilios Tsoukalas, N. F., 2016. Prediction of occupational risk in the shipbuilding industry using multivariable linear regression and genetic algorithm analysis. *Safety Science,* 83(12-22), p. Mar.

WorldScale Association, 2019. *New Worldwide Tanker Nominal Freight Scale.* [Online]
Available at: https://www.worldscale.co.uk/book/preamble-part-a#PPADefinitions
[Accessed 14 August 2019].

Xiaolei Sun, L. T. ,. Y. Y. ,. D. J. L., 2014. Identifying the dynamic relationship between tanker freight rates and oil prices: In the perspective of multiscale relevance. *Economic Modelling,* 30 Jun, 42(1), pp. 287-295.

Yuying Yang, C. L. S. L. L., 2015. *Spillover effect of international crude oil market on tanker market.* [Online]
Available at:
https://www.researchgate.net/publication/281221606_Spillover_effect_of_international_crude_oil_market_on_tanker_market
[Accessed 08 July 2019].