Erasmus University of Rotterdam

MSc in Maritime Economics and Logistics

2017/2018

Statistical Analysis and Modeling of the Baltic Dry Index (BDI)

by

Alexandros Chantzaras

Rotterdam, August 29, 2018

Acknowledgement

This thesis has been carried out as the final part of my Master of Science studies in Maritime Economics and Logistics (MEL) at the Erasmus University of Rotterdam (EUR), under the supervision of Assistant Professor Ted Welten. Without his trust and academic guidelines, I would not be able to conduct this thesis at the highest of standards, especially when taking into consideration the scarce amount of time. I would also like to thank the other members of my supervisors committee.

At this point I would like to express my thanks to the people from the MEL office, namely Ms. Felicia, Ms. Renee and especially Ms. Martha who has been there every single day for us, giving a great effort to help us with everyday life issues, studies and even future career. For sure, the excellence of the MEL programme would not be achievable without the commitment of the lecturers who shared their professional knowledge and working perspective with us.

I am also sincerely grateful to my family for the unceasing encouragement, support and motivation throughout all the years of my studies.

I thank you all.

Abstract

The scope of this thesis is to perform statistical analysis of several economic indexes against the Baltic Dry Index (BDI) and finally, based on the findings of this analysis, construct models simulating the BDI. Therefore the relationship of the BDI with major macroeconomic and / or microeconomic indexes and the modeling feasibility by them has been examined. The followed methodology for achieving that was divided into two parts. The first pillar was the statistical analysis, i.e. the calculation of key statistical characteristics of each variable, such as mean values and standard deviations, as well as the correlation coefficients both between the several indexes and the BDI, and among each other. The second pillar was based on the findings of the statistical analysis, where selected datasets have been arranged in input (selected indexes) / output (BDI) set-ups as to identify the examined models. The resulted models were also plotted against a portion of data not used for the identification process. The key results obtained from this thesis are the existence of correlation between the BDI and several of the examined indexes, as well as the ability to construct a model of the BDI using as inputs some of those indexes with an accepted predictive accuracy. All the above conclude to the fact that BDI can be modelled by related indexes with good predictive and fitting characteristics.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

"God must have been a shipowner. He placed the raw materials far from where they were needed and covered two thirds of the earth with water." – Erling Naess

1.1 Introduction to Shipping

As Stopford (2013) states in the introduction of his book *"Shipping is a fascinating business"*. This phrase reveals the whole perspective of shipping, a worldwide economic and logistic industry that accommodates international trade around the globe for 5000 years. The international shipping industry is responsible for the carriage of around 90% of world trade (Ics-shipping.org 2018). A fact that makes shipping one of the most important factors in the global economy, and especially merchant shipping, as it is the transport facilitator of most of the global trade. The absence of shipping would mean that, the intercontinental trade would be impossible (Ics-shipping.org 2018). According to Ics-shipping.org (2018), the existence of more than 50000 merchant ships transporting every kind of cargo is the *"life blood of the global economy"*, with the world fleet been registered in about 150 nations, and manned by more than a million seafarers of almost every nationality, clear signs of the global character of shipping.

The "fascinating" part of the diversified shipping industry according to Stopford can be summed up in the following quote:

"Shipping is a complex industry and the conditions which govern its operations in one sector do not necessarily apply to another; it might even, for some purposes, be better regarded as a group of related industries. Its main assets, the ships themselves, vary widely in size and type; they provide the whole range of services for a variety of goods, whether over shorter or longer distances. Although one can, for analytical purposes, usefully isolate sectors of the industry providing particular types of service, there is usually some interchange at the margin which cannot be ignored" (Stopford 2013).

Shipping generally can be divided in different groups of industries, based on the differences in the transported cargoes, the different kind of vessels, etc. One of the main categories of shipping is the bulk. Bulk shipping is one of the simplest and most straightforward sectors in shipping, as the main focus, according to Stopford (2013), is the provision of low-cost transport of as many goods as possible in simply designed vessels. The dry bulk ships usually accommodate the trade of a wide spectrum of bulk cargoes including grain, phosphate rock, iron ore, coal, parcels of chemicals, etc. with attention to economy and flexibility (Stopford 2013). The bulk shipping economy would therefore be the scope of this thesis, not due to the simplicity of the sector, but due to the difficulty of predicting the seemingly straightforward.

1.2 Research Objectives

The main objective of this thesis, i.e. the expected results, is to statistically conclude of whether and to what extend several economic indexes are related to the Baltic Dry Index (BDI), an index of high importance for the bulk industry. Following that the feasibility of creating an improved and revised model of the BDI with a predictive ability. The aim is to compare the results of the derived model against real time series data of the BDI throughout the years and see the predictive ability of the proposed model and its responsiveness to given economical and political shocks that have been occurred in the past years.

Therefore, this work is mainly of the interest of investors, ship-owners, charterers, shipping carriers, stakeholders, people generally involved in the maritime industry and interested in the global economic growth, in terms of this is affected by the dry cargo shipping industry.

1.3 Research Questions and Sub-Questions

As stated above, the main objective of the thesis, i.e. the main research question, is to examine to what extend does the Baltic Dry Index (BDI), i.e. generally the dry bulk shipping industry, is related with macroeconomic and / or microeconomic indicators and further examine the feasibility of building a model based on the related factors. Therefore, the main research question is as follows.

- **Main Research Question**
  Does the Baltic Dry Index (BDI) relate with major macroeconomic and/or microeconomic indexes and is there a modeling feasibility of the BDI by them?

In order to answer the general main research question a series of sub−research questions have been posed. These sub−research question are as follows.

- **Sub-questions**
  - To what extent are the examined indexes correlated to each other and the BDI?
  - Is it feasible to perform a system identification of the BDI and eventually build a model connecting the correlated indexes as to model the BDI with predictive abilities?
  - What will be the fitting percentage of the given model based on existing time-series data?
  - How will the models respond to potential economic shocks, i.e. the economic collapse of 2008 etc.?

BDI, which is the index of interest of this thesis, is a daily calculated index by the Balticexchange.com (2018) as a weighted average of freight rates reported from ship-related professionals around the globe, covering only specific vessels carrying dry cargo.

Hence, this index is an indicator of global economic activity that is been derived by the supply and the demand in the dry cargo shipping industry. Hence, the maritime community focuses to correctly identify the economic factors (both microeconomic and macroeconomic ones) that influence the BDI. So, an analysis of the factors that may affect the BDI, as well as the build of a model for the above-mentioned index is always of great importance.

1.4 Research Design and Methodology

The research that is going to be conducted is basically quantitative, since a mathematical model would be derived from numerical data. The correlation analysis of the data will both show the feasibility of that intention and will define whether the examined economic indexes are related and to what extent. The aim finally is to conclude whether there is a correlation between the proposed data and the BDI time series data and to finally produce an estimation model as to achieve a prediction / modeling of the BDI.

The expected results are, therefore, statistical conclusions of the extend that several economic indexes are related to the BDI, and an improved up-to-date model compared to the existing ones in the bibliography.

As to achieve that, the research that is going to be conducted can be roughly divided in two main parts. The first part is qualitative and includes the empirical investigation and identification, based on the literature review, of the main economical parameters that potentially affect the BDI. In this part the selection of the potentially independent variables, i.e. the inputs of the model, are going to be selected. Furthermore, a theoretical background of those indexes will be also provided, i.e. how they are derived etc., when necessary.

Secondly, the examination of the above-mentioned parameters is going to be conducted. So, in the quantitative part, the proposed parameters would be further examined in what extend they truly affect the BDI. This examination will be accomplished by performing correlation analysis between the selected parameters. This analysis would also reveal the correlation between the inputs (independent variable) and the output (dependent variable, i.e. the BDI), since the correlation among them and the BDI will be calculated. Afterwards the identification of the independent variables the extraction of models will follow, which would be validated against the original data of the BDI. Through statistical indicators the level of fitting will be calculated. So, the second part of the research is purely quantitative, since a mathematical model is proposed to be derived from numerical data with the simultaneous examination of the relationship between the data, as mentioned above.

Note that the models' training data will not be the whole set of data available, but a portion of them, since the rest of the real data are going to be used for evaluating the predictive ability of the proposed model against known data (not only through comparison with the training data).

Regarding the collection of the data, secondary ones will be used, since they will be obtained by studies and sources, meaning that databases, like Sin.clarksons.net (2018), and other sources of professional literature, are going to be used.

1.5 Structure of Thesis

Chapter 1 - Introduction

The purpose of this chapter is to provide general information of the thesis, to introduce the reader to the topic and to justify the selection of this topic on behalf of the writer.

Chapter 2 - Literature Overview

This chapter aims to provide the reader with the main characteristics of the dry bulk shipping industry, to give a brief overview of how the market functions and to provide information on the several indexes of this industry, as those are calculated by the Balticexchange.com (2018).

Chapter 3 - Literature Review

This chapter aims to examine the till now related literature review on the topic of forecasting the BDI and the examination of whether or not the BDI can be related with other economic indexes.

Chapter 4 - Methodology

The aim of this chapter is to provide the reader the required theoretical background as to understand how the final results have been obtained. This chapter is therefore been divided into two main subsection, the first one is the presentation of the required statistical tools and the second one is the methodology under which the identification of the desired models have been done, i.e. the N4SID method as presented by Favoreel et al. (2000).

Chapter 5 - Results

This is the chapter of interest of this thesis, as all the final results acquired from the analysis of the obtained data is been presented. As the methodological Ch. 4, this chapter is also divided in two parts, i.e. one referring to the statistical analysis of the data, and one devoted to the modelling results of the BDI. Finally a comparison between the models and a final choice among them is included.

Chapter 6 - Conclusion

In this chapter the final conclusions of this thesis are presented accompanied by some suggestions for future work based on the findings and the limitations faced while conducting the research.

Chapter 2

The Dry Bulk Shipping Industry

2.1 Introduction

As stated in Stopford (2013) the shipping industry is serving the world economy for more than 5000 years, with the today's shipping to be a sophisticated transport service to every part of the globe. In other words, the maritime industry is the connecting node in the international trade with the exploitation of vessels, which connects efficiently the edges of the world by transporting in large volumes and amounts goods and commodities.

Veenstra (1999) mentions that there are a lot of markets in the shipping industry for the different types of transportation services, the ships traded and ship related services, i.e. financing, insuring, bunkering, also from the port perspective. Therefore the simultaneous examination of the shipping industry as a whole is of limited implementation, especially in the frame of this thesis. As a result the main concentration of this thesis will be in the dry bulk shipping, with several definitions and segments of this market been analysed in this chapter.

Moreover, it is a common, generally accepted, fact that the shipping industry, and especially the dry bulk shipping industry, can be characterized by high volatilities (Stopford 2013). This fact is an illustration of the global scope of the shipping industry, meaning that political as well as economical shocks influence the demand of goods around the globe, in combination with seasonal related factors, in the case for instance od agricultural trades (Stopford 2013).

Generally, the bulk shipping can be defined as the service of transporting homogeneous dry bulk parcels (such as iron ore, coal and grain) by the usage of bulk vessels based on an irregular scheduled line (see Veenstra (1999) and Stopford (2013)). So, both voyage charters[1] and time charters[2] are utilized in practise as for chartering contracts in dry bulk shipping (Kavussanos & Visvikis 2016).

The purpose of this chapter therefore is to introduce the reader into the dry bulk

[1]According to Stopford (2013): A transportation contract under which the *"ship earns freight per ton of cargo transported on terms set out in the charter-party which specifies the precise nature and volume of cargo, the port(s) of loading and discharge and the laytime and demurrage. All costs paid by the shipowner"*.

[2]According to Stopford (2013): *"A transportation contract under which the charterer has the use of the vessel for a specific period. A fixed daily or monthly payment is made for the hire of the vessel. Under this arrangement, the owner manages the day-to-day running of the ships, and pays the operating and capital costs. The charterer pays fuel, port charges, loading/discharging fees and other cargo-related costs, and directs the ship operations"*.

shipping industry and the main indicators of it. On top of that the reader will also be familiarised with the various vessel types of this segment of the shipping industry, the types of cargoes been transported by them and the main trading routes of this type of the vessels.

2.2 Dry Bulk Cargoes

The segment of the shipping industry that is dedicated in transporting commodities characterized as "dry bulk", due to the fact that the goods are been shipped in "bulk", i.e. the ship is filled with commodities directly in her hulls in large volumes, and the goods are dry, opposed to liquid bulk cargoes, is the dry bulk shipping industry. The main dry bulk cargoes are categorised into two groups, namely the major bulks and/or minor bulks.

The major bulks are namely five, iron ore, grain, coal, phosphates and bauxite, i.e. the main homogeneous commodities been transported by such vessels (Stopford 2013). On the other hand, as minor bulk cargoes can be characterised commodities like agricultural products, sugar, metals and mineral cargoes, steel products, forest products and even grain, i.e. the many other commodities that travel in shiploads (Stopford 2013).

According to database of Sin.clarksons.net (2018) between the years 2010 and 2017 the total seaborne trade for the main major bulks and in total for the minor bulks is as presented in Table 2.1 below.

Table 2.1: The total world seaborne trade of bulk cargoes between 2010 and 2017.

Date	World Seaborne Iron Ore Trade	World Seaborne Coking Coal Trade	World Seaborne Steam Coal Trade	World Seaborne Bauxite Trade	World Seaborne Phosphate Rock Trade	World Seaborne Minor Bulk Trade
	Million Tonnes	Million Tonnes	Million Tonnes	Million Tonnes	Million Tonnes	Million Tonnes
2010	989.53	228.29	697.58	53.53	23.40	1,577.94
2011	1,049.63	218.28	779.84	60.60	29.20	1,682.37
2012	1,107.12	225.17	885.81	76.00	29.50	1,736.24
2013	1,187.65	258.87	923.56	107.60	28.10	1,823.00
2014	1,339.88	256.16	960.23	72.30	29.50	1,844.35
2015	1,363.56	245.34	892.41	93.99	29.80	1,872.10
2016	1,418.13	245.77	896.11	81.90	28.78	1,874.29
2017	1,472.73	255.99	949.35	97.53	32.48	1,919.54

As can be derived, major bulks consist of more than 60% of the world's dry seaborne trade throughout the years.

In 2017, the latest year which can be examined, the total bulk cargo seaborne trade exceeded 4500 million tonnes (Sin.clarksons.net 2018). The tendency of the growth of world sea dry bulk trade and its high correlation with the average growth of the industrial production can be observed in Fig.2.1.

2.3 Dry Bulk Vessel Types

According to the International Association of Classification Societies (IACS) *"bulk carriers are ships which are constructed generally with single deck, double bottom, hopper side tanks and topside tanks and with single or double side skin construction in cargo hold region and intended primarily to carry dry cargoes in bulk"* (Iacs.org.uk 2018).

While, according to Stopford (2013), bulk carrier vessels are, again, single-deck ships with the capability of carrying dry cargoes such as ore, coal, sugar or cereals,

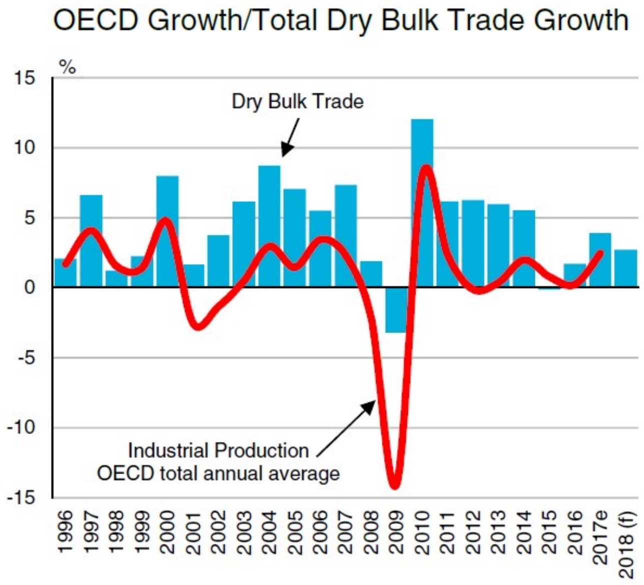

Figure 2.1: The annual world seaborne trade growth of bulk cargoes against the Industrial Production annual average growth (Source: Sin.clarksons.net (2018)).

with the potentiality of smaller vessels to have their own cranes, where larger sizes rely on shore based equipment.

The bulk shipping market in 2017 was compromised of about 11100 vessels with a total tonnage of 817 million dwt, according to data from the Sin.clarksons.net (2018), which according to Kavussanos & Visvikis (2016) corresponds to approximately the 35% of the total merchant fleet. The main categories of dry bulk carriers according to Stopford (2013) and Kavussanos & Visvikis (2016) are:

- Very large ore/bulk carrier (VLOC/VLBC): Bulk carriers very large of size with $200000 - 400000$ tonnes deadweight (dwt)[3]. A special category which can also be accounted as part of the Capesize class.

- Capesize: Bulk carrier too wide to transit the Panama Canal. Usually over 100000 tonnes dwt, but size increases over time.

- Panamax: Bulk carrier which can transit Panama Canal where the lock width of 32.5 m is the limiting factor. Vessels of $60000 - 80000$ tonnes dwt fall into this category.

- Handymax/Handy: Bulk carrier at the smaller end of the range of sizes associated with this type of ship, typically up to $30000 - 60000$ tonnes dwt. Most have their own cargo-handling gear.

- Handysize: The smaller bulk carrier of the range of sizes with $10000 - 30000$ tonnes dwt

[3]Deadweight is the weight a ship can load until the maximum allowable submersion is reached. A contant and unique for every ship weight (Van Dokkum 2003).

The total number of vessel in 2017, according to data collected from Sin.clarksons.net (2018), for the aforementioned classes can be seen in Fig.2.2. As well as the fleet development over the period of years in Fig.2.3.

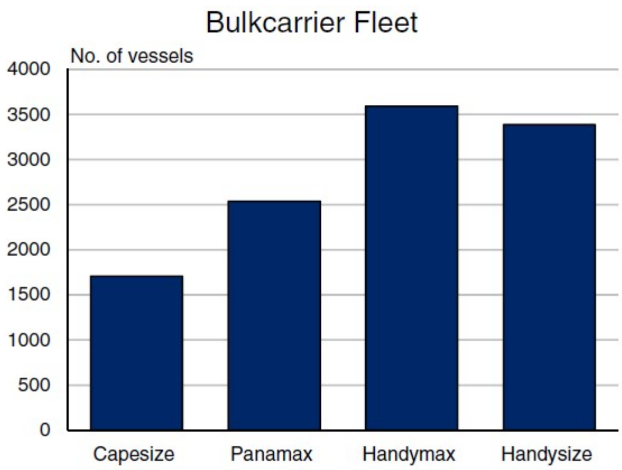

Figure 2.2: The distribution of the bulk carrier fleet in 2017 according to their class (Source: Sin.clarksons.net (2018)).

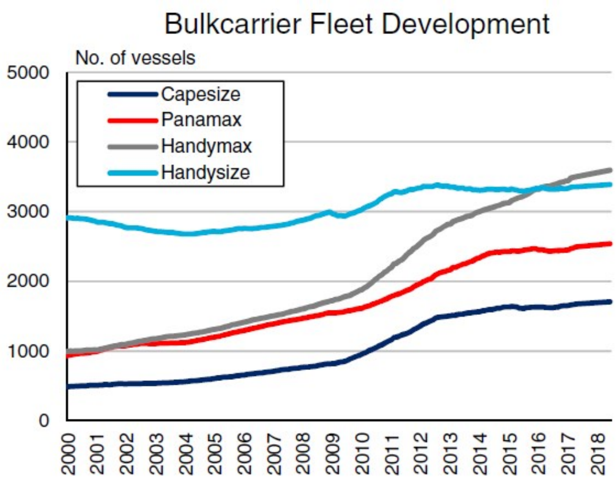

Figure 2.3: The development of the bulk carrier fleet according to their class (Source: Sin.clarksons.net (2018)).

In practise further sub-segments can be found, for instance bulk carriers named after the infrastructural limitations and their regional trade (e.g. Kamsarmax with length up to 229 meters, Newcastlemax with beam up to 47 meters etc.), as well as cargo owners (e.g. Valemax with 400000 dwt) (Kavussanos & Visvikis 2016). Vessels can further differentiate according to their equipment, two main categories are geared and gearless vessels.

As can be derived from the list above there are five (5) in total categories of bulk carrier vessels, with three (3) of them been the main categories among the ships. The first main class size is Capesize vessels, vessels which in terms of deadweight are the ships with over 80,000 dwt. But the introduction of the new-Panamax, which are 82000 dwt to 85000 dwt with the ability to transit the Panama Canal fully loaded, changed this category, with a more up-to-dated definition of a Capesize to be vessels of over 100000 dwt. Table 2.2 reveals that the Capezize sector is specialized on long haul trips of iron ore and coal. The disadvantage of this category, according to Chen (2011), is that due to the relatively big size of those vessels there is a lack of ports able to accommodate their birth.

The second main class size is the Panamax, which consists from vessels between 60000 and 80000 dwt, with maximum width (beam) of 32.2 meters, i.e. the maximum width allowed to transit the Panama canal. Those vessels, see also Table 2.2, usually carry iron ore, coal, grain, and sometimes minor bulks.

The third main class size is the Handymax/Handy vessels with a deadweight of 30000 to 60000 and with usual cargoes, see also Table 2.2, grain and minor bulks. In this category one can also find geared ships, e.g. vessels with in-board cranes, which make it possible for those vessels to birth into ports with limited infrastructures, enabling them also to have more spread out trading routes. The advantage of this factor can also be visualised in Fig.2.3, where the ascending tendency of those class vessels is observed.

Handysize vessels, i.e. vessels up to 30000 dwt, are mainly for the carriage of minor bulks (Chen 2011). The small size of those vessels make them suitable for relatively smaller ports with lack in infrastructure, due to the existence of cranes at those kind of vessels and to the draft and length restrictions.

From the supply/demand perspective of the dry bulk industry, as every other shipping industry, the supply is a function of the deliveries of new vessels and the simultaneous loss of the existing capacity through scrapping. In addition, as also stated by Chen (2011), not only the absolute number of ships worldwide affects the supply/demand balance, but also the operating efficiency, e.g. port congestion can work as a balancing force in the case of oversupply of tonnage. The demand side is been calibrated through the world's demand for goods and commodities, in other worlds the world's economic growth.

2.4 Dry Bulk Trading Routes

Among other sources, according to Grammenos (2013) and Chen (2011), the main trading routes given the class of the dry bulk vessel, as well as the proportion of carriage worldwide of the major bulks, can be seen in the Table 2.2. A more detailed analysis of the specific trade routes and the respective cargoes been transported in each one of them is beyond the scope of this thesis, however the reader is advised to see the database of Sin.clarksons.net (2018).

Moreover, the major dry ports around the globe according to the Sin.clarksons.net (2018) are illustrated in Fig.2.4.

2.5 The Baltic Exchange

The Baltic Exchange is a unique independent source of information for the maritime market for trading and settlement of physical and derivative shipping contracts (Balticexchange.com 2018). As part of the daily operations the Baltic Exchange publishes daily indexes regarding the maritime world. Namely those indexes are the Baltic

Table 2.2: Different types of bulk carriers with their main cargoes and trading routes.

Dry Bulk Carriers		Commodities		
		Iron Ore	Coal	Grain
Capesize	Percentage	70-80%	30-40%	0-5%
	Main Trading Routes	West Australia to Western Europe West Australia to Far East Brazil to Western Europe Brazil to the Far East	East Australia to Western Europe East Australia to Far East South Africa to Western Europe South Africa to Far East	Argentina to Near East and East Europe
Panamax	Percentage	10-20%	40-50%	40-50%
	Main Trading Routes	West Australia to Western Europe West Australia to Far East Brazil to Western Europe Brazil to Far East	North America to Western Europe North America to Far East South Africa to Western Europe South Africa to Far East	North America to Western Europe North America to Far East North America to Near East
Handymax/Handy	Percentage	10%	10-20%	45-55%
	Main Trading Routes	India to Far East North America to Far East West Africa to Western Europe	South Africa to Far East South Africa to Europe	North America to Western Europe North America to Africa Australia to Far East

Exchange 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) (Balticexchange.com 2018).

The calculation of those indexes is been made from a number of components, i.e. rates assumed as benchmarks for the international shipping routes. Totally the dry bulk market consists of thirty (30) routes, as opposed to the tanker industry where thirty-one (31) routes can be identified (Balticexchange.com 2018). Baltcic Exchange is using in total about twenty-five (25) different freight derivative contracts, each one for a number of period, for settling the Baltic benchmarks (Balticexchange.com 2018).

Additionally, as part of the responsibilities of the Balticexchange.com (2018), a weekly benchmarking of sale and purchase of ships and of ship recycling prices is been made by the Baltic Exchange, with six (6) reference prices for sale and purchase and another six (6) for the recycling of the vessels.

The unique factor of all those indexes is the fact that all the benchmark prices and data used as components are from shipbroking members of the Baltic Exchange (Balticexchange.com 2018). As stated by the Balticexchange.com (2018): *"We provide independent daily shipping market information; maintain professional shipbroking standards and resolve disputes"*.

All those benchmarks are mainly applicable for the interest of owners and charters as indicators for the market profitability and trends of contracts, like charterparties, as well as forward freight agreements (FFAs). Baltic benchmarks are also of the interest of vessels' operators. According to the Balticexchange.com (2018) the majority of all those interesting parties are members of the Baltic Exchange.

2.5.1 Main Dry Bulk Indexes

The following information is as presented by the Balticexchange.com (2018). The routinely contracts traded in the freight market and for which the Baltic Exchange provides daily pricing and at the end of each month a settlement price are presented below.

Especially for the Dry Bulk Contracts, which is the main interest of this thesis we have the following indexes:

Figure 2.4: Major Dry ports around the globe (Source: Sin.clarksons.net (2018)).

- **Baltic Exchange Capesize Index (BCI)**[4]
  - ***Individually traded Cape Routes***
    - Route C2: Tubarao/Rotterdam 160000 long tons 10 per cent iron ore, 6 days Sundays holidays included, 6 hours turn time at loading port and 6 hours turn time at discharge port, 0.5% in lieu of weighing. Laydays 20 days forward from date of index, cancelling maximum 35 days forward from date of index. Vessel's age maximum 18 years. Freight based on long tons. 3.75% total commission. Nominal Weighting = 10%.
    - Route C3: Tubarao/Qingdao, 160000 mt 10 per cent iron ore, 18 m salt water arrival draft, scale load/30000 mt. Sundays holidays included discharge. 6 hours turn time at loading port, 24 hours at discharge port. Laydays 20 days forward from date of index, cancelling maximum 35 days forward from date of index. Vessel's age maximum 18 years. Freight based on metric tonnes. 3.75% total commission. Nominal Weighting = 15%.
    - Route C4: Richards Bay/Rotterdam, 150000 mt 10 per cent coal, scale load/25000 mt. Sundays holidays included discharge. 18 hours turn time at loading port and 12 hours at discharge port. Laydays 25 days forward from date of index, cancelling 40 days forward from date of index. Vessel's age maximum 15 years. Freight based on metric tonnes. 3.75% total commission. Nominal Weighting = 5%.
    - Route C5: W. Australia/Qingdao, 160000 mt 10 per cent iron ore, 18 m salt water arrival draft, scale load/30000 mt. Sundays holidays included discharge, 6 hours turn time at loading port and 24 hours at discharge

[4] Capesize 4 Time Charter Average: A straight average of routes C8_03, C9_03, C10_03, C11_03.

port. Laydays 20 days forward from date of index, cancelling max. 35 days forward from date of index. Vessel's age max. 18 years. Freight based on metric tonnes. 3.75% total commission. Nominal Weighting = 15%.

* Route C7: Bolivar/Rotterdam 150000 mt 10% coal, 50000 mt Sundays holidays included loading/25000 mt. Sundays holidays included discharge, 12 hours turn time at loading port and 12 hours turn time at discharge port. Laydays 20 days forward from date of index, cancelling maximum 35 days forward from date of index. Vessel's age maximum 15 years. 3.75% total commission. Nominal Weighting = 5%.
* Route C8_03: Delivery Gibraltar-Hamburg range, 5-15 days ahead of the index date, transAtlantic round voyage duration 30-45 days, redelivery Gibraltar-Hamburg range. 3.75% total commission. Based on a Baltic capesize of the following specifications: 172000 mt dwt, not over 10 years of age, 190000 cbm grain, max length overall (L_{OA}) 289 m, max beam 45 m, draft 17.75 m, 14.5 knots laden, 15.0 knots ballast on 56 mts fuel oil, no diesel at sea. Nominal Weighting = 10%.
* Route C9_03: Delivery Amsterdam-Rotterdam-Antwerp range or passing Passero, 5-15 days ahead of the index date, redelivery China-Japan range, duration about 65 days. 3.75% total commission. Based on a Baltic capesize of the following specifications: 172000 mt dwt, not over 10 years of age, 190000 cbm grain, max L_{OA} 289m, max beam 45 m, draft 17.75 m, 14.5 knots laden, 15.0 knots ballast on 56 mts fuel oil, no diesel at sea. Nominal Weighting = 5%.
* Route C10 03 Delivery China-Japan range, 5-15 days ahead of the index date, round voyage duration 30-40 days, redelivery China-Japan range. 3.75% total commission. Based on a Baltic capesize of the following specifications: 172000 mt dwt, not over 10 years of age, 190,000 cbm grain, max L_{OA} 289 m, max beam 45 m, draft 17.75 m, 14.5 knots laden, 15.0 knots ballast on 56 mts fuel oil, no diesel at sea. Nominal Weighting = 20%.
* Route C11 03 Delivery China-Japan range, 5-15 days ahead of the index date, redelivery Amsterdam-Rotterdam-Antwerp range or passing Passero, duration about 65 days. 3.75% total commission. Based on a Baltic capesize of the following specifications: 172000 mt dwt, not over 10 years of age, 190,000 cbm grain, max L_{OA} 289 m, max beam 45 m, draft 17.75 m, 14.5 knots laden, 15.0 knots ballast on 56 mts fuel oil, no diesel at sea. Nominal Weighting = 15%.

- **Baltic Exchange Panamax Index (BEP)**[5]

  - ***Individually traded Panamax Routes***

    * Route P1a_03: Basis a Baltic panamax 74000 mt dwt not over 12 years, 89000 cbm grain, max L_{OA} 225 m, draft 13.95 m, 14.0 knots on 32 mts fuel oil laden, 28mts fuel oil ballast and no diesel at sea. For a transAtlantic (including ECSA) round of 45/60 days on the basis of delivery and redelivery Skaw-Gibraltar range. Loading 15-20 days ahead

[5] Panamax 4 Time Charter Average: A straight average of routes P1a_03, P2a_03, P3a_03, P4_03)

in the loading area. Cargo basis grain, ore, coal, or similar. 3.75% total commission. Nominal Weighting = 25%.

  * Route P2a_03: Basis a Baltic panamax 74000 mt dwt not over 12 years of age, 89000 cbm grain, max L_{OA} 225 m, draft 13.95 m, 14.0 knots on 32 mts fuel oil laden, 28 mts fuel oil ballast and no diesel at sea, basis delivery Skaw-Gibraltar range, for a trip to the Far East, redelivery Taiwan-Japan range, duration 60/65 days. Loading 15-20 days ahead in the loading area. Cargo basis grain, ore, coal, or similar. 3.75% total commission. Nominal Weighting = 25%.
  * Route P3a_03: Basis a Baltic panamax 74000 mt dwt not over 12 years of age, 89,000 cbm grain, max L_{OA} 225 m, draft 13.95 m, 14.0 knots on 32 mts fuel oil laden, 28 mts fuel oil ballast and no diesel at sea, for a trans Pacific round of 35/50 days either via Australia or Pacific (but not including short rounds such as Vostochny/Japan), delivery and redelivery Japan/South Korea range. Loading 15-20 days ahead in the loading area. Cargo basis grain, ore, coal or similar. 3.75% total commission. Nominal Weighting = 25%.
  * Route P4_03: Basis a Baltic panamax 74000 mt dwt not over 12 years of age, 89000 cbm grain, max L_{OA} 225m, draft 13.95 m, 14.0 knots on 32 mts fuel oil laden, 28 mts fuel oil ballast and no diesel at sea, delivery Japan-South Korea range for a trip via US West Coast-British Columbia range or Australia, redelivery Skaw-Passero range, duration 50/60 days. Loading 15/20 days ahead in the loading area. Cargo basis grain, petroleum coke, coal or similar. 3.75% total commission. Nominal Weighting = 25%.

- **Baltic Exchange Supramax Index (BES)**[6]
  - ***Individually traded Supramax Routes***
    * Route 1A: Delivery Antwerp/Skaw range for a trip of 60/65 days redelivery Singapore/Japan range including China 5% total commission. Laycan 5/10 days in advance. Nominal Weighting = 12.5%.
    * Route 1B: Delivery passing Canakkale for a trip of 50/55 days redelivery Singapore/Japan range including China 5% total commission. Laycan 5/10 days in advance. Nominal Weighting = 12.5%.
    * Route 2: Delivery South Korea/Japan range for 1 Australian or trans Pacific round voyage, for a 35/40 day trip, redelivery South Korea/Japan range 5% total commission. Laycan 5/10 days in advance. Nominal Weighting = 25%.
    * Route 3: Delivery South Korea/Japan range for a trip of 60/65 days redelivery Gibraltar/Skaw range 5% total commission. Laycan 5/10 days in advance. Nominal Weighting = 25%.
    * Route 4A: Delivery US Gulf for a trip about 30 days, redelivery Skaw-Passero range, 5% total commission. Laycan 5/10 days in advance. Nominal Weighting = 12.5%.

[6] The Baltic Exchange Supramax Index is based on the following description: *"Standard "Tess 52" type vessel with grabs as follows: 52454 mt dwt self trimming single deck bulkcarrier on 12.02 m ssw, 189.99 m L_{OA} 32.26 m, Beam 5ho/ha, 67756 cum.grain 65600 cbm bale, 14L /14.5B on 30 mt (380 cst) no mdo at sea, Cr 4 x 30 mt with 12 cbm grabs, maximum age 10 years"*.

    * Route 4B: Delivery Skaw-Passero range for a trip about 30 days, redelivery US Gulf, 5% total commission. Laycan 5/10 days in advance. Nominal Weighting = 12.5%.
    * Route 5: Delivery Dakar/Douala range for a trip via East Coast South America of 60/65 days, redelivery Singapore/Japan range, 5 per cent total commission. Laycan 5/10 days in advance. (Route S5 does not contribute towards the BSI or TC Average).
    * Route 9: Delivery Dakar/Douala range for a trip via east coast South America of about 45 days, Redelivery Skaw-Cape Passero range, 5% total commission, Laycan 5/10 days in advance. (Route S9 does not contribute towards the BSI or TC Average).

- **Baltic Exchange Supramax Asia Index (BES Asia)**[7]
  - ***Individually traded Supramax Asia Routes***
    * Route 6: Delivery South Korea/Japan range for a trip via Australia or 50/55 days, redelivery India, 5% total commission. Laycan 5/10 days in advance. (Route S6 does not contribute towards the BSI or TC Average)
    * Route 7: Delivery Cape Comorin / Haldia range including Sri Lanka for a trip of 20/30 days redelivery China. Cargo basis iron ore or similar with 5 per cent total commission. Laycan 5/10 days in advance (Route S7 does not contribute towards the BSI or TC Average)
    * Route 8: Delivery China Hong Kong/Shanghai range including Taiwan for a trip via Indonesia with coal of 20-25 days redelivery East Coast India Chennai/ Paradip range, with 5% total commission. Laycan 5/10 days in advance. (Route S8 does not contribute towards the BSI or TC Average)

- **Baltic Exchange Handysize Index (BHSI)**[8]
  - ***Individually traded Handysize Routes***
    * Route 1: Dely Skaw-Passero for a trip about 35/45 days, redelivery Recalada-Rio de Janeiro range. 5% total commission. Laycan 5/10 days in advance. Nominal Weighting = 12.5%.
    * Route 2: Dely Skaw-Passero range for a trip about 35/45 days, redelivery Boston-Galveston range. 5% total commission. Laycan 5/10 days in advance. Nominal Weighting = 12.5%.
    * Route 3: Dely Recalada-Rio de Janeiro for a trip about 35/45 days, redelivery Skaw-Passero range. 5% total commission. Laycan 5/10 days in advance. Nominal Weighting = 12.5%.
    * Route 4: Dely US Gulf for a trip about 35/45 days, via US Gulf or NC South America, redelivery Skaw-Passero range. 5% total commission. Laycan 5/10 days in advance. Nominal Weighting = 12.5%.

[7] The Baltic Exchange Supramax Asia Index is based on the same description as the Supramax description.

[8] The Baltic Exchange Handysize Vessel Description: *"28000 mt dwt self trimming single deck bulk-carrier on 9.78 m ssw, 169 m L_{OA}, 27 m beam, 5 holds/5 hatches, 37523 cbm grain, 35762 cbm bale, 14 knots average laden/ballast on 22 mt ifo (380) no diesel at sea, 4 x 30 t cranes, Maximum age 15 years"*.

* Route 5: Dely SE Asia for a trip via Australia, about 25/30 days, redelivery Singapore-Japan range including China. 5% total commission. Laycan 5/10 days in advance Nominal Weighting = 25%.
* Route 6: Dely S Korea-Japan range for a trip via Nopac of about 40/45 days, redelivery Singapore-Japan range including China. 5% total commission. Laycan 5/10 days in advance Nominal Weighting = 25%.

According to Balticexchange.com (2018) the difficultly of calculating the exact value of transactions in the freight derivatives market used for benchmarking the Baltic indices is due to the fact that for an accurate result an access to every transaction is necessary. An access though which is impossible in practise.

2.5.2 Baltic Dry Index (BDI)

As previously mentioned in the beginning of this section the Baltic Exchange as part of the daily operations publishes daily indexes regarding the maritime world. One of the main indexes, among the others presented in the previous subsection, is namely the Baltic Exchange Dry Index (BDI). BDI is an index of high interest for the dry bulk shipping market and for the scope of this thesis. This is due to the fact that, as presented in Eq.2.1, BDI is a weighted combination of other indexes, providing a more holistic overview of the dry bulk shipping industry. Hence, the BDI is an indicator of global trade and shipping, providing information regarding the cost of maritime transportation of goods around the world. Therefore the BDI is one of the most closely followed indexes in the maritime sector.

According to the Balticexchange.com (2018) we can find that the Baltic Dry Index (BDI) is the successor to the Baltic Freight Index (BFI) and came into operation on 1 November 1999. Since 1 July 2009, the index has been a composite of the Dry Bulk Timecharter Averages[9].

The following formula is used to calculate the BDI:

$$BDI = ((CapesizeTC_{5avg} * 0.4) + (PanamaxTC_{avg} * 0.3) + \\ +(SupramaxTC_{avg} * 0.3)) * 0.10 \tag{2.1}$$

where: $TC_{avg} = Time\,Charter\,Average$

Note that the BDI is not a stock market index, but on the contrary it is a daily calculated index by a private owned company, the Balticexchange.com (2018), as a weighted average of freight rates reported from professionals in the shipping industry around the globe, according to Eq.2.1, covering only the vessels carrying dry cargo (oil, gas, chemicals or containers are not included). This index is thus an indicator of global economic activity providing a real-time outlook at global raw material and infrastructure demand, i.e. an indirect way of assuming the health and the future of the global economy. Due to the practical and economic difficulties to manipulate both the supply of bulk carriers and the demand of raw materials the BDI is completely devoid of speculative players (Balticexchange.com 2018). Therefore, it is an index derived directly by the supply and the demand in the dry cargo shipping industry.

In Fig. 2.5 one can see the monthly BDI's plot against time. The illustrated data are from the 01/11/1999 till the 01/04/2018 with monthly intervals. As the main data source Sin.clarksons.net (2018) has been used.

[9] See also: https://www.balticexchange.com/market-information/indices/BDI/

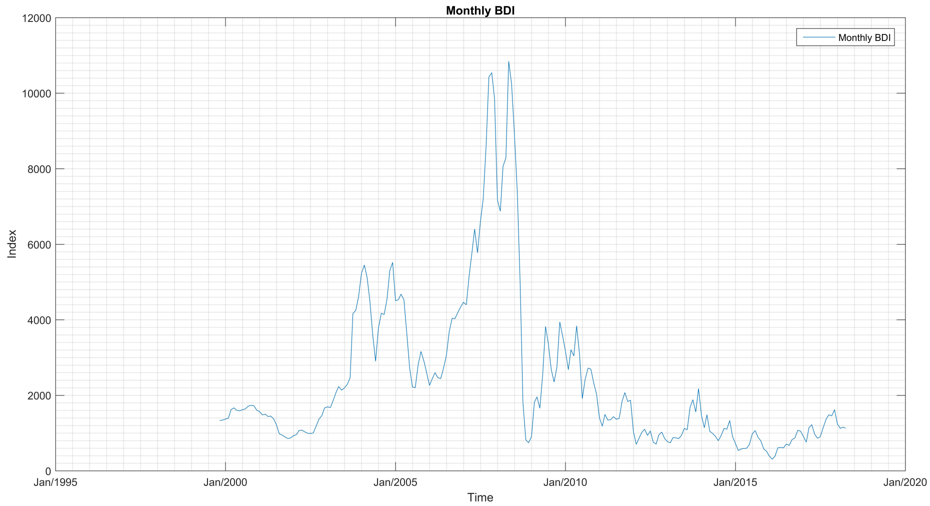

Figure 2.5: Monthly BDI plot from 01/11/1999 till 01/04/2018 (Source: Sin.clarksons.net (2018)).

2.6 Conclusion

The reader is now familiar with the main characteristics of the dry bulk shipping industry as the main information regarding this segment of shipping have been presented. The review of the dry bulk shipping industry revealed not only the main types of cargoes been transported by the various types "bulkers", i.e. the different types of bulk carriers, but also the main trading routes that those fleets are being utilized into. Furthermore, the examination of the various indexes that represent the freight rates / chartering ability of this type of market revealed that the BDI is one the most important indexes, hence it takes into consideration the main trading routes of the dry bulk market in a weighted way, as can be seen in Ch.2.5.2. Therefore, BDI reflects an overall overview of the dry bulk market and the further examination of it will be the interest of this thesis.

Chapter 3

Literature Review

3.1 Introduction

In this chapter the literature review, i.e. previous papers and works related to the topic of interest, will be presented. The literature review has been constructed in two individual parts. The first part of this chapter is the one presented in Ch.3.2 and is a bibliographical overview of the examined data that influence / affects the BDI. The second one, i.e. Ch.3.3, is presenting the various modeling / regression attempts that have been done up until now on connecting the various potential independent variables with the BDI.

3.2 Data Selection

The dry bulk shipping market is a major component of the international shipping market and it is characterized by high risk and volatility, uncertainties caused by factors such as the global economy, the volume and pattern of seaborne trade, and government policies (Chen et al. 2012). The Baltic Dry Index (BDI) is one of the most significant indexes in the dry bulk shipping industry and it is of great importance to recognize the relationship between BDI and economic indexes (Chou & Lin 2010). Based on the a priori knowledge of that relationship the trend of freight rates in the shipping market can be foreseen and influence important decisions such as charter-in, charter-out and risk-avoidance (Chou & Lin 2010). Hence, the importance of this index according to the same source is of interest of both ship-owners, charterers and generally shipping carriers. Therefore, various studies have been done over time examining the relationship of the BDI with numerous economic indicators. Based on the findings presented below has been done the selection of the examined data in this thesis.

As the study of Chou et al. (2017) revealed, BDI could be related to economic indexes including the global agriculture stock index, the international metal price index, the Organisation for Economic Co-operation and Development (OECD) combined leading index, global energy stock index and the United States (US) dollar index. The results of this work revealed that the ability of those indexes of forecasting the BDI is good, with a final Root Mean Square Percentage Error (RMSPE) of about 47%, which means that a regression model based on those independent variables is a reasonable analysis model (Chou et al. 2017).

The bulk cargo shipping industry required a hedging trading indicator, a gap that filled by the creation of the BDI (Chou et al. 2015). BDI was the replacement of the

Baltic Freight Index (BFI) in 1999 (Chou et al. 2015). As presented above in Ch.2 the dry bulk market is segmented into smaller different markets; one of those segments is the capesize market, with the BCI been the main indicator for this part of the market. According to Chou et al. (2015) the BCI, i.e. the respective BDI for only the capesize market, can be related to Asian Steel Index (ASI), i.e. the steel needs of China for production of goods. The analysis for supporting this idea by Chou et al. (2015) was the investigation of the Vector Autoregressive Moving-Average Model (VARMA), which concluded that there is a proof that ASI and BCI do influence each other.

Another attempt in the bibliography of trying to analysing the BDI has been done by G. Divya (2014). As to analyse the BDI various world economic indicators had been used, such as World GDP, World Inflation, World Exports / Imports, Gold Price, etc.. The analysis, among others, revealed that BDI influence the bond market (global bond index), GDP affected the BDI and finally that BDI is an economic indicator of the future market (G. Divya 2014).

The BDI is one of the most important indicators on the cost of shipping and on the volume of worldwide trade and manufacturing activity (Bildirici et al. 2015). Changes in the BDI lead to permanent shocks to trade of major exporting economies (Papailias et al. 2017). Hence, global factors influence the supply and demand of BDI index, since in economic growth the demand of raw materials increases together with production and investments and during economic recessions the demand of raw material decreases, a fact that reduces the utilized capacity (Bildirici et al. 2015). In Bildirici et al. (2015) the relationship between BDI and economic growth of the United States (US) has been investigated with MS-VAR Method. The results revealed that BDI can be indeed been used as an indicator of a crisis in GDP growth for the US.

In Papailias et al. (2017) the cyclicalities of the BDI and the implications for forecasting performance are investigated. More specifically Papailias et al. (2017) took into consideration commodity variables from commodities such as coal, copper, corn, cotton, iron ore, tin and wheat (as also been done by Goulas (2012)). Since those are the main bulk trading commodities, it is expected those variables to be able to predict the BDI (Papailias et al. 2017). Also crude oil prices (Brent Europe) revealed to have a positive correlation with the BDI according to the same study. For reasons of strong positive or negative correlation also a number of other economic variables took into consideration, among others the British pound/US dollar exchange rate (GBPUSD), the dollar index (DXY) and the US Treasury yields (Papailias et al. 2017). The final results of the work revealed an overall good performance of forecasting the BDI based on those indexes (Papailias et al. 2017).

Other sources, like Chen (2011), also investigates the London Interbank Offered Rate (LIBOR) as an dependent variable of the BDI with exceptional results.

3.3 Modeling Methods

In statistics, regression analysis consists of techniques for modeling the relationship between a dependent variable (response variable) and one or more independent variables (explanatory variables or predictors), more specific in linear regression the dependent variable is modelled as a linear function of a set of regression parameters and a random error, with the most commonly used criterion the least squares method (Xin Yan 2009).

Linear regression analysis has been a widely used analysis in various fields of science and economics, applied also in the shipping industry. For instance Multivariable

Linear Regression (MVLR) has been used in the development of a calculating model that calculated the occupational risk in the shipbuilding industry using occupational accidents data and by comparing the predicted values with the reported data it was demonstrated that the proposed model was an efficient way for predicting the risk of occupational injury (Tsoukalas & Fragiadakis 2016).

Therefore the paper of Chou & Lin (2010) investigates the relationship between BDI and economic indexes by linear regression technique, as mentioned also above, with the accuracy performance of this linear regression model been evaluated by the Root Mean Square Percentage Error (RMSPE), showing that the proposed model has reasonable ability of forecasting. In another paper the cyclical properties of the BDI and their implications for forecasting performance are investigated, finding that changes in the BDI can lead to shocks to trade of major exporting economies (Papailias et al. 2017). Commodities and trigonometric regression can improve predictions and forecasting results can be implemented for risk management in the freight sector (Papailias et al. 2017).

While on the contrary past researches were limited to seasonal analysis (Kavussanos & Alizadeh-M 2001), a forecasting performance evaluation has proved that a cyclical pattern of cycle duration of between 3 and 5 years can be observed (Papailias et al. 2017). The examined models perform a decent statistical forecasting ability of the BDI (Papailias et al. 2017).

In another paper from the bibliography, the validation of the results demonstrated that the development of a multivariate Vector Autoregressive model with exogenous variables (VARX) can improve the forecasting accuracy of the BDI compared to the predictive power of a univariate AutoRegressive Integrated Moving Average (ARIMA) framework, which was used as a benchmark model for purposes of comparison (Tsioumas et al. 2017). This suggests that the VARX model approach can substantially improve the accuracy of BDI forecasts, providing a useful tool for chartering decisions under uncertainty (Tsioumas et al. 2017).

Another paper examines also time series models, such as the univariate AutoRegressive Integrated Moving Average (ARIMA), univariate AutoRegressive Integrated Moving Average with exogenous variables (ARIMAX), Vector AutoRegression (VAR) and Vector AutoRegression with exogenous variables (VARX), by comparing them as of their forecasting capabilities providing useful information in the selection of superior forecasting models (Chen et al. 2012).

The findings were that forecasting models like a VAR model and a VARX model perform better on the out-of-sample forecast against the ARIMA model and ARIMAX model (Chen et al. 2012).

But it is also mentioned that the main message of the forecasting performance of various forecasting models is that caution is required by market players and / or analysts, and therefore the selection of the estimation period of forecasting models needs to be done very carefully (Chen et al. 2012).

As can be concluded from the above, various linear regressions methods have been used over the years for modeling the BDI. But it is important to differentiate the model from the system itself, since the system is in the reality what the model tries to explain (Lyzell 2009). And there lies the application of system identification, which is a subset of mathematical modeling, modeling mathematical dynamical systems from empirical data Lyzell (2009). In other words the main concept of system identification is the extraction of a mathematical model that explains the depended variable given some independent variables of the system.

Different algorithms for system identification methods can be found in the literature (such as N4SID, IV-4SID, MOESP, CVA). For this thesis the Numerical algorithm for Subspace State Space System IDentification (N4SID) was selected for the extraction of the model of the BDI based on the results of Abdelghani et al. (1998).

3.4 Conclusion

Given the resulted conclusions of the examined papers, both over the selection of data and the ability of the purposed ways of modeling, we can conclude on some various variables that can be identified as independent variables of the BDI. Such of those potential data may be idexes like the Trade Weighted Steel Production Index, the Industrial Production of China, LIBOR Interest Rates and several other interest rates, Brent Crude Oil Price and other indicators regarding oil production and oil prices, as well as indexes related to the transported by the dry bulk carries commodities. Hence, the potential connection of those data with the BDI will be further examined in the following chapters of this work.

Furthermore, due to the lack in the bibliography of an attempt to identify the model of the BDI with any of the existing methods, and the simultaneous saturation of the bibliography with typical regression methods, it has been decided that the application of the Numerical algorithm for Subspace State Space System IDentification (N4SID) as to identify the model of the BDI from an input - output dataset[1].

[1] As inputs are the various independent variables and as output the BDI.

Chapter 4

Methodology

4.1 Introduction

This chapter is aiming on providing the reader all the theoretical background on where to collect the data and how to perform all the statistical and modeling analysis as to answer the main research question of this thesis. Therefore the division of this chapter was based on that concept and in Ch.4.2 the main source been used for the gathering of the data is presented, in Ch.4.3 the main statistical tools been used as to both reveal the main characteristics of each one of the data and also the correlation among them and mainly with the BDI are presented and at the end in Ch.4.4 the methodological approach of modeling the BDI can be found.

4.2 Data Collection

Regarding the collection of the used data, secondary ones will be used. Data obtained by studies and sources will be analysed. By that sense databases, like Sin.clarksons.net (2018), and other sources of professional literature sources were used as the main data sources. The Sin.clarksons.net (2018) can be identified as a trusted database for shipping related data, since *"Clarksons Research is respected worldwide as the most authoritative provider of intelligence for global shipping"* (Sin.clarksons.net 2018).

4.3 Statistical Analysis

Statistics can identified as a very broad subject, with applications in a great number of different fields (Isotalo 2001). As stated by Keller (2013) *"Statistics is a way to get information from data"*. So in general statistics is a methodology of collecting, analysing, interpreting and drawing conclusions out of data.

A characteristic that varies from one individual member of the population[1] and / or sample[2] to another is called a variable (Isotalo 2001). The following subsections are based on assumed random variable vectors A_k (k = 1, 2, . . . , K) made up of N scalar observations organised in a matrix form (data matrix), i.e. with the following form:

[1] *"Population is the collection of all individuals or items under consideration in a statistical study"* (Weiss & Hassett 1999)

[2] *"Sample is that part of the population from which information is collected"* (Weiss & Hassett 1999).

$$A_k = \begin{pmatrix} A_{1k} \\ A_{2k} \\ \vdots \\ A_{Nk} \end{pmatrix} \quad (4.1)$$

where N is the numbers of observations (i.e. the sample size is N).

And where all variable sectors A_k together can form an extended observations matrix A, as:

$$A = \left[\; A_1, \; A_2, \; \ldots, \; A_K \; \right] \quad (4.2)$$

4.3.1 Measures of Centre and Variation

Average or Mean Value

According to Isotalo (2001) the most common measure of centre for quantitative variable is the (arithmetic) sample mean. The mean of a variable can be defined as *"the sum of observed values in a data divided by the number of observations"* (Isotalo 2001).

Therefore, for a random variable vector A_k (predefined k) made up of N scalar observations, the average or mean (μ) value is mathematically calculated as:

$$\mu_k \;=\; \frac{1}{N} \sum_{i=1}^{N} A_{ik} \quad (4.3)$$

Standard Deviation

The sample standard deviation is, according to Isotalo (2001), one of the most frequently used measure of variability. *"It can be considered as a kind of average of the absolute deviations of observed values from the mean of the variable in question"* (Isotalo 2001).

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

$$S_k \;=\; \sqrt{\frac{1}{N-1} \sum_{i=1}^{N} |A_{ik} \;-\; \mu_k|^2} \quad (4.4)$$

where μ_k is as in Eq. 4.3.

Hence, the standard deviation (S_k) is defined through the mean (μ_k) of the variable A_k. Thus, standard deviation is a measure of variation, when mean is used as the measure of centre.

Note that the standard deviation (S_k) is the square root of the variance (S_k^2) and it is always a positive number ($S_k \geq 0$) for every k.

4.3.2 Correlation Analysis

Correlation

One of the most efficient ways to reveal the relation between two variables is a scatterplot. Scatterplots illustrate the relationship between two quantitative variables since it is a graph in which the values of two variables are plotted along two axes. The values of one variable appear on the horizontal axis and the values of the other variable appear on the vertical axis. Hence, the pattern of the resulting plotting points is able to reveal any correlation among the two variables.

For the interpretation of a scatterplot, according to Isotalo (2001), the seek of an overall pattern is essential. A revealed pattern illustrates the direction, form and strength of the relationship between the two variables (Isotalo 2001).

The patterns reveal various types types of correlations according to Isotalo (2001). If the two variables are positively correlated, then as one variable increases so does the other one, and vice versa. If the two variables are negatively correlated then as one variable increases the other one decreases, and vice versa. If there is no correlation among the variables then there is no particular relationship between the variables. If the form of the relationship between the two variables is a linear relationship, then the points in the plot show a straight-line pattern Isotalo (2001). Curved relationships and clusters are also other forms (Isotalo 2001). Generally, as a measure of the strength of the relationship is the closeness of the points in the scatterplot form the identified "pattern line".

Additionally, when quantitative data are been used, then they are usually presented graphically as a histogram (Isotalo 2001). Histograms are usually formed from data been grouped, illustrating either frequencies or relative frequencies of each class interval (Isotalo 2001).

Correlation Coefficient

The visual identification of the closeness of the scattered points to a linear relation can be quantified by calculating a numerical measure, i.e. the sample correlation coefficient (Isotalo 2001).

The sample correlation coefficient can be defined as *"the measure of the strength of the linear relation between the x and y variables"* and denoted by r (or in some cases r_{xy}), and it is also called as Pearson correlation coefficient (Isotalo 2001).

The mathematical representation of the correlation coefficient r is:

$$r = \frac{S_{xy}}{\sqrt{S_{xx}}\sqrt{S_{yy}}} \tag{4.5}$$

where:

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

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

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

According to Isotalo (2001), based on r, some features of the correlation coefficient are:

1. Correlation r reveals the strength of only a linear relationship between two variables.

2. The correlation r is always between -1 and 1. Values of r near 0 reveals a very weak linear relationship. On the contrary the strength of the linear relationship increases as r moves toward either -1 or 1. Values of r close to -1 or 1 indicate that the points lie close to a straight line.

3. Positive r indicates positive correlation between variables, and negative r illustrates negative one.

4. The correlation r is unitless; it is just a number between -1 and 1.

5. Like the mean and standard deviation, the correlation is strongly affected by outlying observations.[3]

4.4 Modeling

The main objective of this section is to provide the reader the theoretical foundations regarding the procedures being followed for the model identification process. The aim was the identification of models from data in an input-output set-up[4].

As to construct a model out of data three basic entities are basically involved (Ljung 1999):

1. The data: The mathematical input - output data.

2. The identification of a set of candidate models: A set of models is obtained as to identify the most suitable.

3. A rule of thumb, under which the candidate models can be assessed by the usage of data: The determination of the most suitable model with existing data as guide.

4.4.1 Numerical algorithm for Subspace State Space System IDentification (N4SID)

The Numerical algorithm for Subspace State Space System IDentification (N4SID) is a type of linear system identification algorithm Favoreel et al. (2000). The following mathematical procedure of identification is as been published by the Favoreel et al. (2000).

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

$$x_{k+1} = Ax_k + Bu_k + w_k$$
$$y_k = Cx_k + Du_k + v_k \tag{4.6}$$

[3]So the usage of r should be with caution; especially when outliers exist on scatterplots.

[4]For this thesis as input data will be preferably used the highly correlated data, as the previous statistical analysis would suggest, and as output data the index of interest, i.e. the BDI.

[4]The Chapter 4.4.1 is the mathematical procedure of identification as been published by the Favoreel et al. (2000).

with[5]

$$\mathbf{E}\left[\begin{pmatrix} w_p \\ v_p \end{pmatrix} \begin{pmatrix} w_q^T v_q^T \end{pmatrix}\right] = \begin{pmatrix} Q & S \\ S^T & R \end{pmatrix} \delta_{pq} \geq 0. \tag{4.7}$$

The vectors $u_k \in \mathbb{R}^{m\times 1}$ 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 $\mathbf{E}$ shows the expectation. 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 the 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 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[6]. In order to state the problem treated it is assumed a large number of measurements of the input u_k and the output y_k generated by the unknown system. The task is to determine the order n of the unknown system, the system matrices A, B, C, D up to within a similarity transformation and an estimate of the matrices Q, S, R.

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 4.8, played an important role in the development of subspace identification

$$Y_f = \Gamma_i X_i + H_i^d M_f + N_f \tag{4.8}$$

where

- The extended observability matrix Γ_i

$$\Gamma_i \overset{def}{=} \begin{pmatrix} C \\ CA \\ CA^2 \\ \vdots \\ CA^{i-1} \end{pmatrix} \tag{4.9}$$

- The deterministic lower block triangular Toeplitz matrix H_i^d

$$H_i^d \overset{def}{=} \begin{pmatrix} D & 0 & 0 & \dots & 0 \\ CB & D & 0 & \dots & 0 \\ CAB & CB & D & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ CA^{i-2}B & CA^{i-3}B & CA^{i-4}B & \dots & D \end{pmatrix} \tag{4.10}$$

[5] $\mathbf{E}$ denotes the expected value operator and δ_{pq} the Kronecker delta.

[6] a stochastic process is said to be ergodic if its statistical properties can be deduced from a single, sufficiently long, random sample of the process. The reasoning is that any collection of random samples from a process must represent the average statistical properties of the entire process.

- The stohastic lower block triangular Toeplitz matrix H_i^s

$$H_i^s \overset{def}{=} \begin{pmatrix} 0 & 0 & 0 & \dots & 0 \\ C & 0 & 0 & \dots & 0 \\ CA & C & D & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ CA^{i-2} & CA^{i-3} & CA^{i-4} & \dots & 0 \end{pmatrix} \tag{4.11}$$

- The input and output block Hankel matrices are defined as

$$U_{0|i-1} \overset{def}{=} \begin{pmatrix} u_0 & u_1 & \dots & u_{j-1} \\ u_1 & u_2 & \dots & u_j \\ \vdots & \vdots & \ddots & \vdots \\ u_{i-1} & u_i & \dots & u_{i+j-2} \end{pmatrix} \tag{4.12}$$

$$Y_{0|i-1} \overset{def}{=} \begin{pmatrix} y_0 & y_1 & \dots & y_{j-1} \\ y_1 & y_2 & \dots & y_j \\ \vdots & \vdots & \ddots & \vdots \\ y_{i-1} & y_i & \dots & y_{i+j-2} \end{pmatrix} \tag{4.13}$$

where it is assumed for stohastic reasons that $j \to \infty$. For convenience and short hand notation, we call

$$U_p \overset{def}{=} U_{0|i-1}, \quad U_f \overset{def}{=} U_{i|2i-1}, \quad Y_p \overset{def}{=} Y_{0|i-1}, \quad Y_f \overset{def}{=} Y_{i|2i-1}$$

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

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

The block Hankel matrix (Eq. (4.12) and (4.13)) formed with the process noise w_k and the 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 short hand notation

$$M_p \overset{def}{=} M_{0|i-1}, \quad M_f \overset{def}{=} M_{i|2i-1},$$

$$N_p \overset{def}{=} N_{0|i-1}, \quad N_f \overset{def}{=} N_{i|2i-1}$$

We finally denote the state sequence X_i as:

$$X_i \overset{def}{=} \begin{pmatrix} x_i & x_{i+1} & x_{i+2} & \dots & x_{i+j-1} \end{pmatrix} \tag{4.14}$$

In the following, 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[7] :

[7] $\dagger$ denotes the Moore-Penrose pseudo-inverse.

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

$\mathcal{A}/\mathcal{B}^{\perp}$ is the projection of the row space of $\mathcal{A}$ into $\mathcal{B}^{\perp}$, the orthogonal complement of the 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 Fig. 4.1). 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 $\tilde{X}_i$ of the state sequence 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 $\tilde{X}_i$. 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. (4.8). 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 Γ_i or X_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 known, Γ_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 above equation (4.8)? 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

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

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

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

Therefore

$$Y_f U_f^{\perp} = \Gamma_i X_i / U_f^{\perp} + H_i^s M_f + N_f.$$

The following step consists in weighting this projection to the left and to the right with some matrices W_1 and W_2

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

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

1.
$$rank\ (W_1 \Gamma_i) = rank\ \Gamma_i \tag{4.15}$$

2.
$$rank\ (X_i/U_f^{\perp}W_2) = rank\ X_i \tag{4.16}$$

3.
$$W_1(H_i s^s M_f + N_f)W_2 = 0 \tag{4.17}$$

The first two conditions guarantee that the rank-n property of $\Gamma_i X_i$ 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

$$\mathcal{O}_i \stackrel{def}{=} W_1 Y_f/U_f^{\perp}W_2 = W_1\Gamma_i X_i/U_f^{\perp}W_2 \tag{4.18}$$

with SVD

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

The following important properties can now be stated

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

$$W_1\Gamma_i = U_1 S_1^{1/2}$$

$$X_i/Uf^{\perp}W_2 = S^{1/2}V_2^T$$

Obviously, the singular value decomposition of the matrix $W_1Y_f = U_f^{\perp}W_2$ 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) where as the right singular vectors contain information about the states X_i. For an appropriate choice of the weighting matrix W_2, the matrix

$$\tilde{X}_i \stackrel{def}{=} X_i/U_f W_2 \tag{4.19}$$

can indeed be considered as an estimate of the state sequence X_i. It was shown from Van Overschee & De Moor (2012) that, for a particular choice of W_2, $\tilde{X}_i$ is a Kalman filter estimate of 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 & De Moor 1994). It should be noted that for the basic-4SID algorithm, condition 4.17 is not satisfied which implies that in general this method is not consistent.

Second step: finding the state space model

N4SID algorithm uses the state estimates $\tilde{X}_i$ (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 & De Moor (2012) and Van Overschee & De Moor (1994)):

$$W_1 = I_{li} \tag{4.20}$$

$$W_2 = (W_p/U_f^{\perp})^{\dagger}W_p \tag{4.21}$$

The estimated state sequence $\tilde{X}_i$ can be interpreted as the solution of a bank of Kalman filters, working in parallel on each of the columns of the matrix W_p. Besides $\tilde{X}_i$, 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.(4.18) 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 $\tilde{X}_{i+1}$:

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

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 & De Moor (2012) and Van Overschee & De Moor (1994)):

$$\begin{pmatrix} \tilde{X}_{i+1} \\ Y_{i|i} \end{pmatrix} = \begin{pmatrix} A & B \\ C & D \end{pmatrix} \begin{pmatrix} \tilde{X}_i \\ U_{i|i} \end{pmatrix} + \begin{pmatrix} \rho_w \\ \rho_v \end{pmatrix} \tag{4.22}$$

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

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

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.

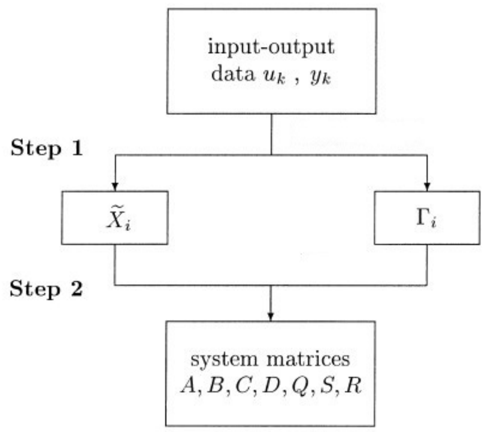

Figure 4.1: Main steps in subspace algorithms (Source: Favoreel et al. (2000)).

4.5 Conclusion

By utilizing the existing tools for statistical analysis the mean values of each dataset, as well as the standard deviation of them and their correlation can be derived. By further concluding by the results of the statistical analysis and by constructing the dataset in an input - output data with the system identification method N4SID the model of the required index can be obtained. The examination of the derived potential models, as well as the results of the statistical analysis will be presented in following chapters.

A graphical presentation of the general concept of this thesis can be found in Fig.4.2, where the whole procedure is illustrated step-wisely. Furthermore, as for the results of each step, the reader is been advised to see the following Ch.5.

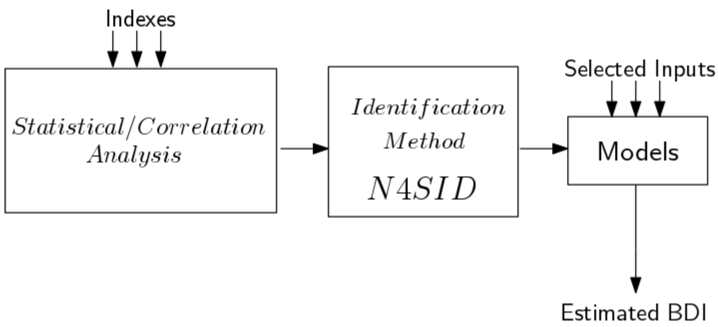

Figure 4.2: The methodology flowchart of the thesis.

Chapter 5

Results

5.1 Introduction

In this chapter all the results and the obtained from the analysis of the data conclusions are presented. The chapter is divided into three (3) main parts. The first part is dedicated to the collection of the data, i.e. the main sources used for data gathering. The second part includes the statistical analysis of the gathered data, as well as a first indication of the modeling ability of the BDI from those data. Finally the modeling results are presented.

For the analysis of the data and for the modelling of the BDI Matlab software from MathWorks © has been utilized.

5.2 Data

As previously mentioned Sin.clarksons.net (2018) was the main database been used for the collection of the examined in this thesis data. The main sets of data been used in this thesis are illustrated in Table 5.1. In this table also the abbreviations of those data can be found. So reader is been advised to seek the abbreviations of the data in the following Table, as those will be mainly been used in the following sections of this Chapter.

As can be seen in Table 5.1 in the column *Index* the data been used for investigating the modeling ability of the BDI are in accordance to the ones suggested by the literature review been done in Ch.3. The selection of these data therefore is justified from both the suggested bibliography and the intuitive sense of the influential economical factors of the BDI.

Since the graphical representation of each of the data over time is not of interest within the main core of this thesis one can find the plots of those in the Appendix A.1.

5.3 Statistical Analysis

5.3.1 Measures of Centre and Variation

In Table 5.2 besides the list of examined data one can find the mean values of them together with the standard deviations. The methodological part of how those figures were extracted can be found in Ch.4.3. Moreover, one can also see gathered the results of the correlation analysis of the data against both the monthly BDI and the percentage

Table 5.1: List of examined data together with their abbreviations, units and examined period.

#	Abbreviation	Units	Index	Corrected Period	
				From	To
1	**Monthly_Clarksea**	Index	Monthly Clarksea Index	1/11/1999	1/4/2018
2	**Monthly_BDI**	Index	Monthly BDI	1/11/1999	1/4/2018
3	**Monthly_BDI_per**	Index	% of Change of Monthly BDI		
4	**Monthly_1m**	%Yr/Yr'	Inflation Indicator OECD Europe (excl Turkey)	1/11/1999	1/4/2018
5	**Monthly_2m**	Index	Trade Weighted Steel Production Index	1/11/1999	1/4/2018
6	**Monthly_3m**	%Yr/Yr'	Industrial Production OECD	1/11/1999	1/4/2018
7	**Monthly_4m**	%Yr/Yr'	Industrial Production China	1/11/1999	1/4/2018
8	**Monthly_5m**	%Yr/Yr'	Industrial Production South East Asia Average	1/11/1999	1/4/2018
9	**Monthly_6m**	%	USA Interest Rates	1/11/1999	1/4/2018
10	**Monthly_7m**	%	German/Euro Interest Rates	1/11/1999	1/4/2018
11	**Monthly_8m**	%	UK Interest Rates	1/11/1999	1/4/2018
12	**Monthly_9m**	%	LIBOR Interest Rates	1/11/1999	1/4/2018
13	**Monthly_10m**	$/bbl	Brent Crude Oil Price	1/11/1999	1/4/2018
14	**Monthly_11m**	$/Tonne	US Gulf Wheat Price	1/11/1999	1/4/2018
15	**Monthly_12m**	$/Tonne	US Gulf Corn Price	1/11/1999	1/4/2018
16	**Monthly_13m**	$/€	Exchange Rates Euro $/€	1/11/1999	1/4/2018
17	**Monthly_14m**	mbpd	Global Oil Prod.	1/11/1999	1/4/2018
18	**Monthly_15m**	mbpd	OPEC: Crude Oil Prod.	1/11/1999	1/4/2018
19	**Monthly_16m**	,000tonnes	World Steel Production	1/11/1999	1/4/2018

change of the monthly BDI[1] (for more details on the analysis of the correlation results see 5.3.2). For comparison reasons the mean value of the BDI is $\mu_{BDI} = 2401.9$, with a standard deviation of $\sigma_{BDI} = 2113.8$.

Table 5.2: List of examined data together with their abbreviations, units, examined period, mean values and standard deviations of them, as well as correlation percentages against the monthly BDI and the percentage change of the monthly BDI.

#	Abbreviation	Units	Index	Corrected Period		Mean Value	Standard Deviation	Correlation against	
				From	To			Monthly_BDI	Monthly_BDI_per
1	**Monthly_Clarksea**	Index	Monthly Clarksea Index	1/11/1999	1/4/2018	-	-	99%	-
2	**Monthly_BDI**	Index	Monthly BDI	1/11/1999	1/4/2018	-	-	-	
3	**Monthly_BDI_per**	Index	% of Change of Monthly BDI			-	-		
4	**Monthly_1m**	%Yr/Yr'	Inflation Indicator OECD Europe (excl Turkey)	1/11/1999	1/4/2018	0.0215	0.0086	42%	-18%
5	**Monthly_2m**	Index	Trade Weighted Steel Production Index	1/11/1999	1/4/2018	228.5025	97.123	-31%	2%
6	**Monthly_3m**	%Yr/Yr'	Industrial Production OECD	1/11/1999	1/4/2018	1.0951	4.7675	14%	-11%
7	**Monthly_4m**	%Yr/Yr'	Industrial Production China	1/11/1999	1/4/2018	12.0203	4.5664	58%	5%
8	**Monthly_5m**	%Yr/Yr'	Industrial Production South East Asia Average	1/11/1999	1/4/2018	6.7488	5.2749	45%	0%
9	**Monthly_6m**	%	USA Interest Rates	1/11/1999	1/4/2018	0.0486	0.0198	34%	-2%
10	**Monthly_7m**	%	German/Euro Interest Rates	1/11/1999	1/4/2018	0.0196	0.0196	33%	-5%
11	**Monthly_8m**	%	UK Interest Rates	1/11/1999	1/4/2018	0.036	0.0221	52%	-4%
12	**Monthly_9m**	%	LIBOR Interest Rates	1/11/1999	1/4/2018	0.0221	0.0196	36%	-3%
13	**Monthly_10m**	$/bbl	Brent Crude Oil Price	1/11/1999	1/4/2018	64.0871	31.2276	17%	-6%
14	**Monthly_11m**	$/Tonne	US Gulf Wheat Price	1/11/1999	1/4/2018	226.031	79.3499	20%	-2%
15	**Monthly_12m**	$/Tonne	US Gulf Corn Price	1/11/1999	1/4/2018	160.3117	81.8956	1%	-3%
16	**Monthly_13m**	$/€	Exchange Rates Euro $/€	1/11/1999	1/4/2018	1.2143	0.1719	51%	1%
17	**Monthly_14m**	mbpd	Global Oil Prod.	1/11/1999	1/4/2018	86.3541	8.9563	-13%	0%
18	**Monthly_15m**	mbpd	OPEC: Crude Oil Prod.	1/11/1999	1/4/2018	29.6814	2.11	7%	-6%
19	**Monthly_16m**	,000tonnes	World Steel Production	1/11/1999	1/4/2018	107440	24915.00	-13%	1%

5.3.2 Correlation Analysis

From the findings of the Table 5.2 we can find that the percentage change of the monthly BDI is not of great importance and can be neglected, since the potential independent

[1]The monthly percentage change of the BDI calculated as: $\%BDI = \frac{BDIt+1-BDIt}{BDIt}$, where t is the ascending number of month, i.e. $t = 0, 1, 2, \ldots$

variables, i.e. the examined datasets, did not revealed any strong correlation with this index. As can be seen on the same table the correlation coefficient were highly concentrated around zero percentage (0%) values. Therefore the correlation results against the BDI will be mainly presented on this subsection.

In Fig.5.1[2] the total correlation of all data been used for this thesis is presented. One can find the both the correlation coefficients among the data and the graphical representation of the linear relationships between the variables. The linear relationships between the variables can reveal multicollinearities[3]. Moreover, the histograms of class intervals of each data is been displayed.

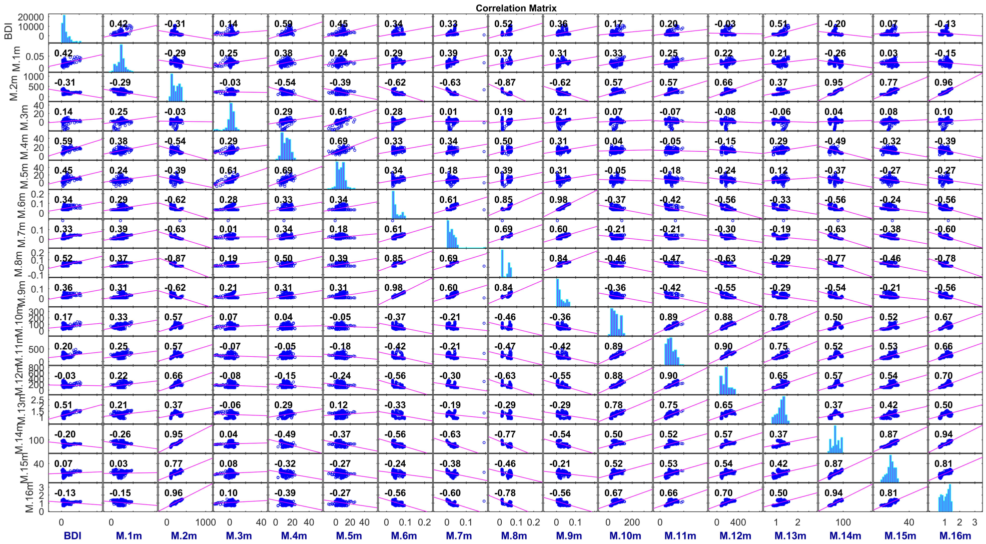

Figure 5.1: Graph that includes all the correlations among all examined datasets against each other and the BDI together with their histograms (Source: Sin.clarksons.net (2018)).

In the following figures the individual correlation of each dataset against the monthly BDI are presented.

In Fig.5.2 we can see a highly correlation between the two examined indexes, i.e. a correlation coefficient of 99%, with the scatter plot been highly concentrated around the linear relation. Therefore we can conclude that the BDI and the Clarksea Index are similarly calculated indexes. Therefore Clarksea index can not be used as an independent variable for modeling the BDI, due to the fact that they basically are a different representation of the same index.

Fig.5.3 illustrates the correlation between the monthly Inflation Indicator OECD Europe (excl. Turkey) and the monthly BDI. As we can see those two indexes are correlated with a correlation coefficient of 42% with their scatter plot revealing a good linear relationship. Therefore the monthly Inflation Indicator OECD Europe (excl. Turkey) could be a potential input variable for modeling the BDI, since the change of

[2]Note that the abbreviation M."x"m in this subsection represents the Monthly_"x"m of the Table 5.1

[3]Multicollinearity is the very high intercorrelations or inter-associations among the independent variables.

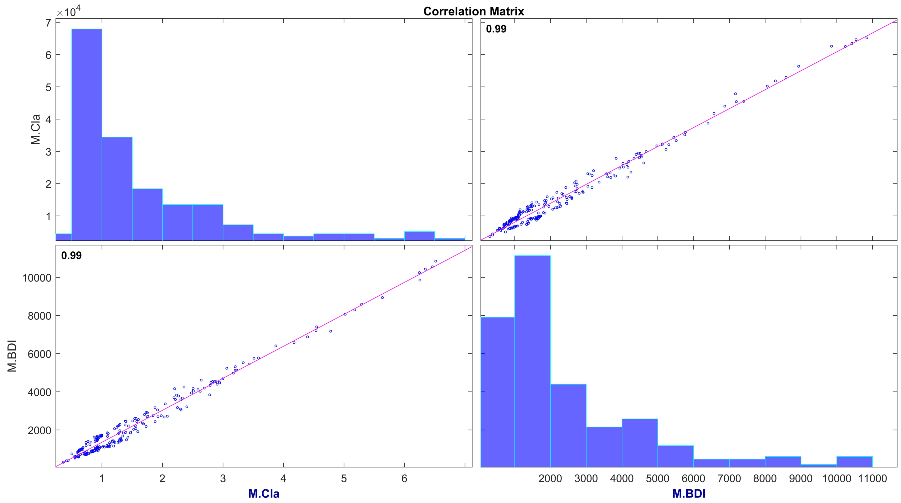

Figure 5.2: Graph that illustrates the correlation between monthly Clarksea Index and monthly BDI and their histograms (Source: Sin.clarksons.net (2018)).

this index affects the BDI.

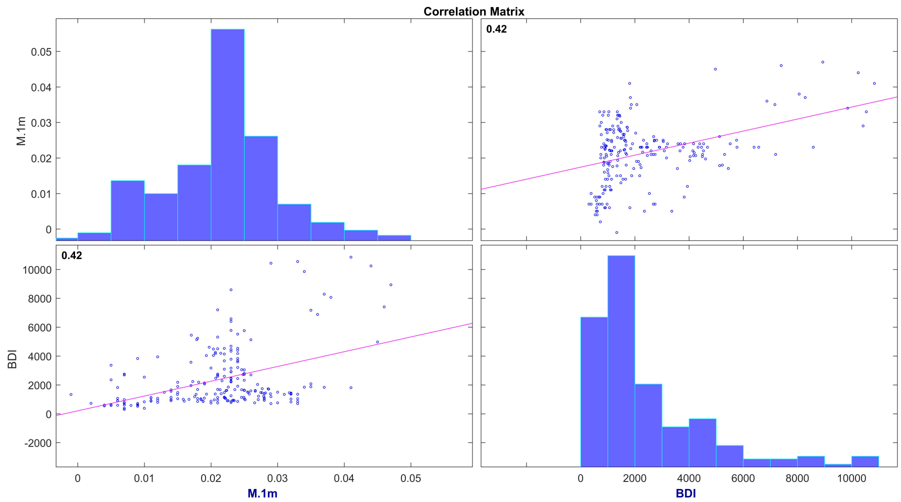

Figure 5.3: Graph that illustrates the correlation between monthly Inflation Indicator OECD Europe (excl. Turkey) and monthly BDI and their histograms (Source: Sin.clarksons.net (2018)).

The correlation between the monthly Trade Weighted Steel Production Index (SPI) and monthly BDI together with their histograms is illustrated in Fig.5.4. One can see a negative correlation, meaning that an increase in one index would mean a potential decrease of the other, with a correlation coefficient of -31%. This coefficient reveals a good relationship between the indexes with the potentiality of the SPI to be used as a potential input variable for the modeling. Their scatter plot is also a sign which illustrates that potentiality.

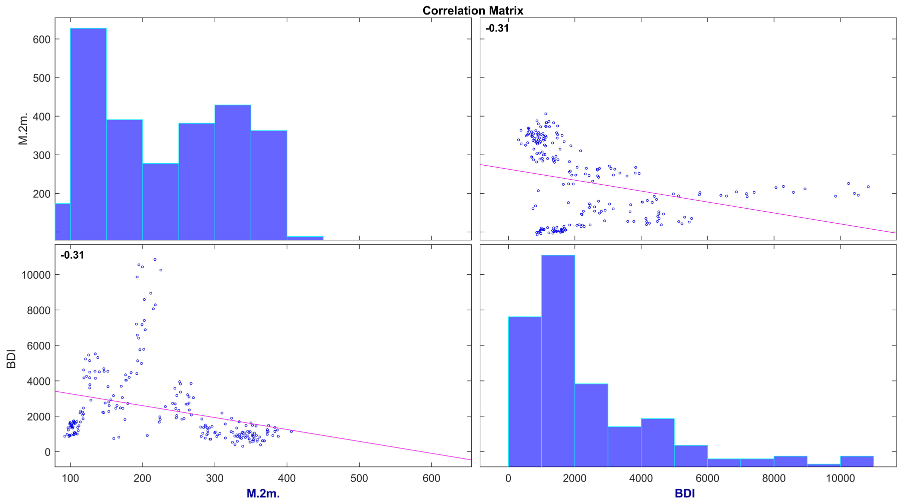

Figure 5.4: Graph that illustrates the correlation between monthly Trade Weighted Steel Production Index (SPI) and monthly BDI and their histograms (Source: Sin.clarksons.net (2018)).

The Industrial Production of OECD and the BDI are correlated as presented in Fig.5.5. A positive, slightly weak, correlation of 14% found. But the examination of their scatter plot revealed the existence of some outliers, which according to the theory may be the reason for the relatively low correlation coefficient, even though the plotted points on the scatter plot seems to be highly concentrated around the line of relationship. Therefore the Industrial Production of OECD could be used as an independent variable of BDI.

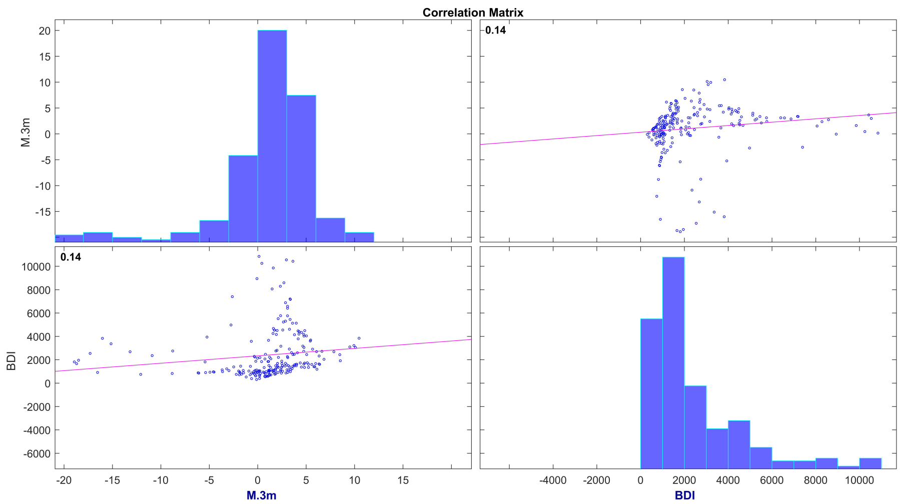

Figure 5.5: Graph that illustrates the correlation between monthly Industrial Production of OECD and monthly BDI and their histograms (Source: Sin.clarksons.net (2018)).

In Fig.5.6 the reader can find the correlation between the Industrial Production of China and the BDI with monthly intervals. A correlation of 59% together with a scatter plot concentrated around the linear relationship illustrates the positive correlation of those two indexes. Their dependence makes the Industrial Production of China a potential input variable of the model of BDI. The same conclusion can be deducted from Fig.5.7, where the correlation of the monthly Averaged Industrial Production of South East Asia and the monthly BDI is presented. A correlation of 45% and the quality of the scatter plot leads to the same assumption regarding this index as previously.

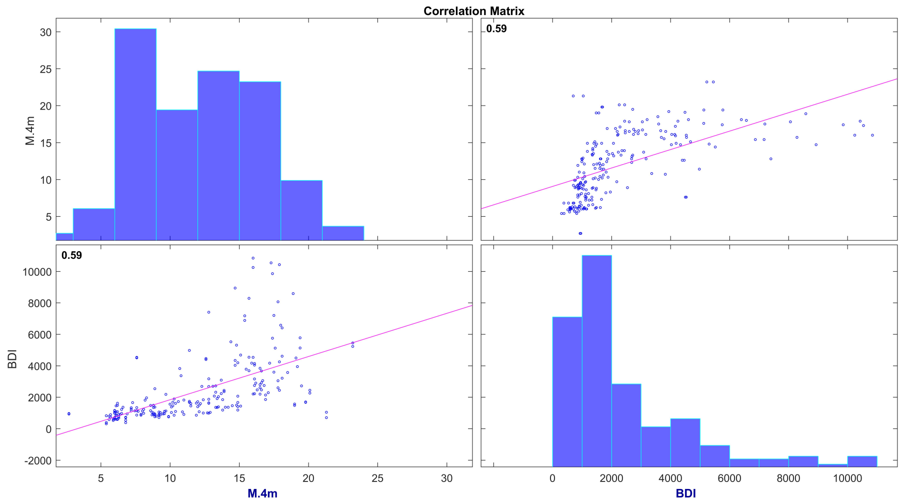

Figure 5.6: Graph that illustrates the correlation between monthly Industrial Production of China and monthly BDI and their histograms (Source: Sin.clarksons.net (2018)).

The results of the correlation analysis between the USA Interest Rates and the BDI are presented in Fig.5.8. The correlation coefficient between those indexes was 34%, revealing a good correlation between them. But as can be seen from both the scatter plot and the Fig.A.9 the USA Interest Rates are, especially after the financial crisis of 2008, been set to a constant number. Therefore, this index, even though it is mathematically correlated to the BDI, is not a valid / pure indicator of the market. Therefore it is not suggested as an input for the model of the BDI.

The same conclusion can be derived from the scatter plots of Fig.5.9 and Fig.5.10, in combination with Fig.A.10 and Fig.A.11, where German / Euro Interest Rates and UK Interest Rates are examined respectively. The resulted correlations where of 33% for the German / Euro Interest Rates and 52% for the UK Interest Rates, but as scatter plots and the figures in the appendix illustrates, those indexes do not represent market balances between supply and demand, but are been set by external factors (like government, central banks etc.). Therefore they are not recommended as inputs for the model of the BDI.

On the other hand the the London Interbank Offered Rate (LIBOR), as presented in Fig.A.12, is an index driven by the "forces" of the market, since it is based on different currencies. In Fig.5.11 we can see that the LIBOR Interest Rates is positively correlated with the BDI with a coefficient of 36%. All the above combined with the

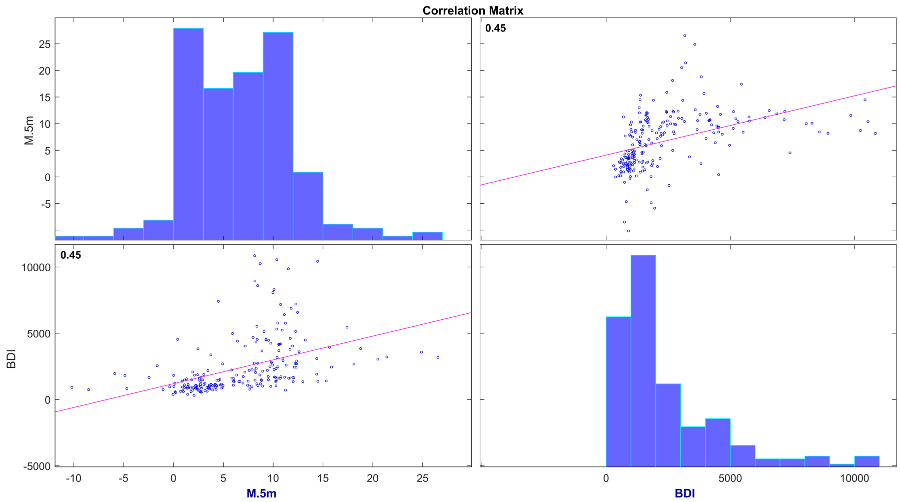

Figure 5.7: Graph that illustrates the correlation between monthly Averaged Industrial Production of South East Asia and monthly BDI and their histograms (Source: Sin.clarksons.net (2018)).

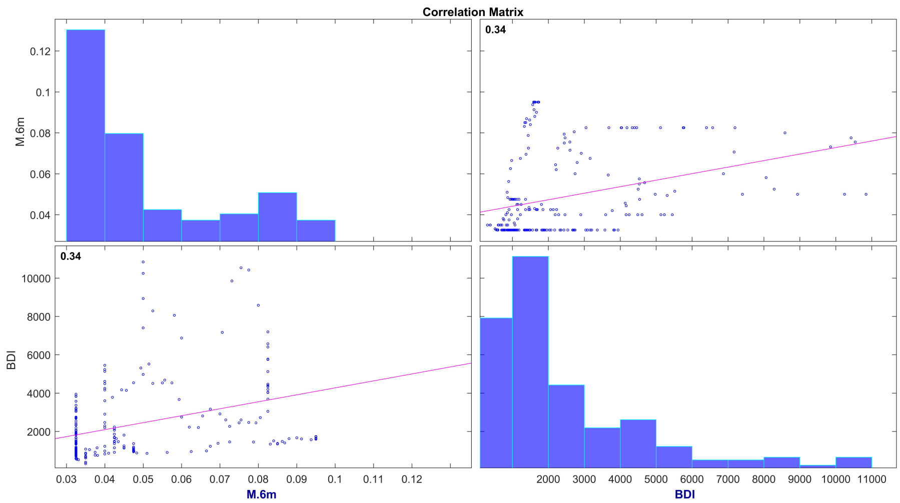

Figure 5.8: Graph that illustrates the correlation between monthly USA Interest Rates and monthly BDI and their histograms (Source: Sin.clarksons.net (2018)).

conclusions obtained by the study of their scatter plot reveals the potentiality of the LIBOR Interest Rates to be used as input parameter for modeling the BDI.

As suggested by the bibliography, oil production and prices indicators are correlated with the BDI. Figs.5.12, 5.16 and 5.17 are a proof of this perception. In those figures the correlation of Brent Crude Oil Price, Global Oil Production and OPEC Crude Oil Production with the BDI is graphically illustrated. The results are correlation coefficients of 17%, -20% and 7% respectively. Indicating that those are potentials influencers of the BDI. One may contradict this idea, especially for the OPEC Crude

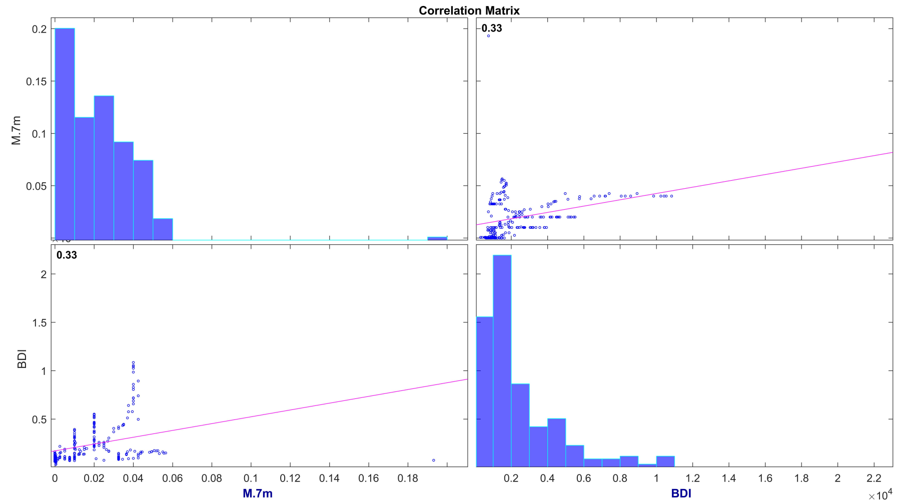

Figure 5.9: Graph that illustrates the correlation between monthly German / Euro Interest Rates and monthly BDI and their histograms (Source: Sin.clarksons.net (2018)).

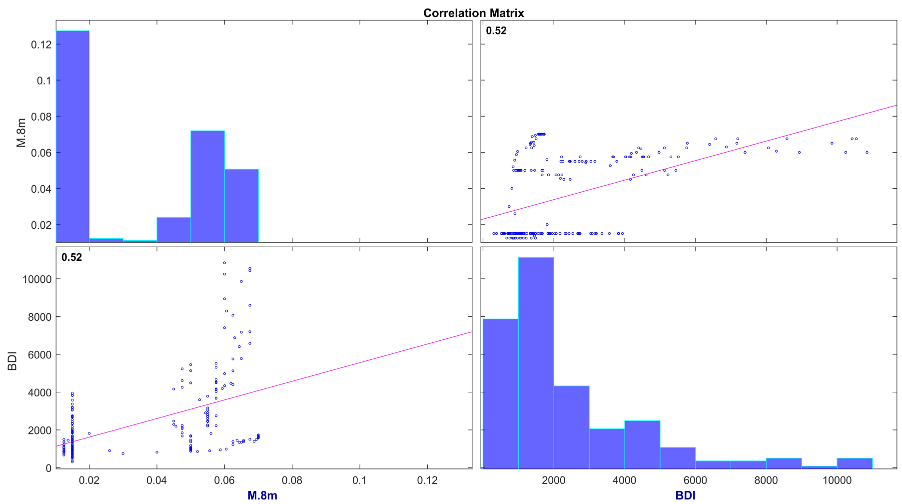

Figure 5.10: Graph that illustrates the correlation between monthly UK Interest Rates and monthly BDI and their histograms (Source: Sin.clarksons.net (2018)).

Oil Production index with only 7% correlation, a suggestion though that is rejected by the observation of the scatter plot in Fig.5.17.

Moreover, the literature review (Ch.3) revealed that also the market indexes of the transported by the bulk carriers commodities may potentially affect the BDI. Therefore the further analysis of such indexes, like the US Gulf Wheat Price, the US Gulf Corn Price and the World Steel Production, has been conducted. Two of these indexes are related with the market rice of two main major bulk commodities, i.e. wheat and corn, whereas the world steel production represents the amount of steel produced monthly worldwide. The results of each statistical analysis are illustrated in Figs.5.13,

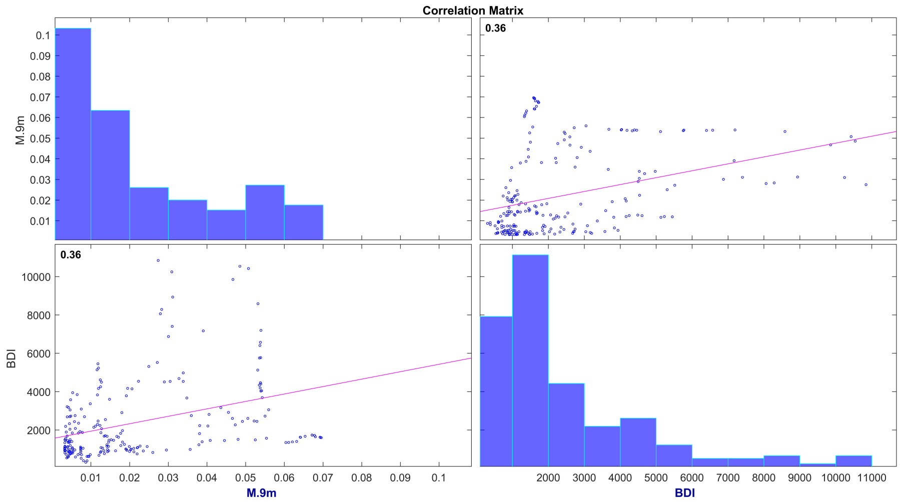

Figure 5.11: Graph that illustrates the correlation between monthly LIBOR Interest Rates and monthly BDI and their histograms (Source: Sin.clarksons.net (2018)).

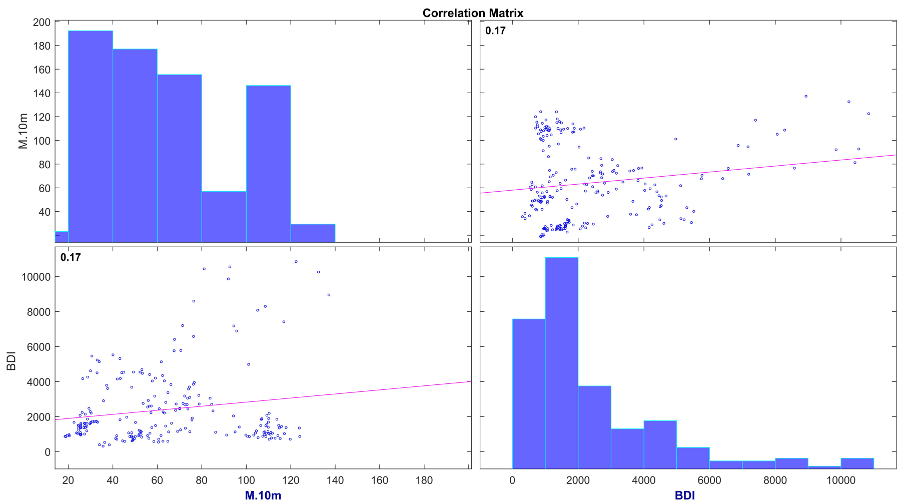

Figure 5.12: Graph that illustrates the correlation between monthly Brent Crude Oil Price and monthly BDI and their histograms (Source: Sin.clarksons.net (2018)).

5.14 and 5.18 respectively. In conclusion we can see that US Gulf Wheat Price is by 20% correlated with the BDI, where on the contrary US Gulf Corn Price revealed no significant correlation. World Steel Production is also related with the BDI by -13%. Therefore both US Gulf Wheat Price and World Steel Production could be potential independent variables of the BDI.

Similarly with LIBOR Interest Rates, the Exchange Rates of Euro against the US Dollar (see Fig.A.16) is an index derived from the supply and demand equilibrium of these two currencies. Therefore represents the tendencies of the market at a given period of time. This index is positively correlated with the BDI by 51%, resulting the

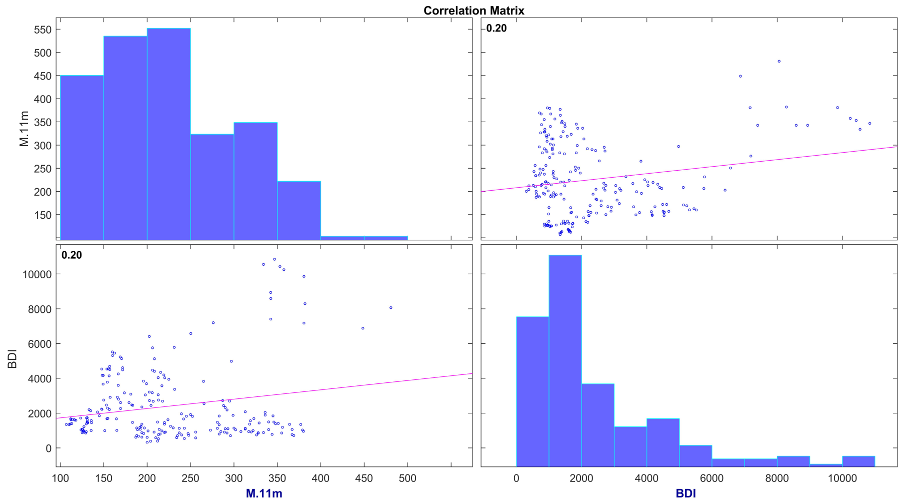

Figure 5.13: Graph that illustrates the correlation between monthly US Gulf Wheat Price and monthly BDI and their histograms (Source: Sin.clarksons.net (2018)).

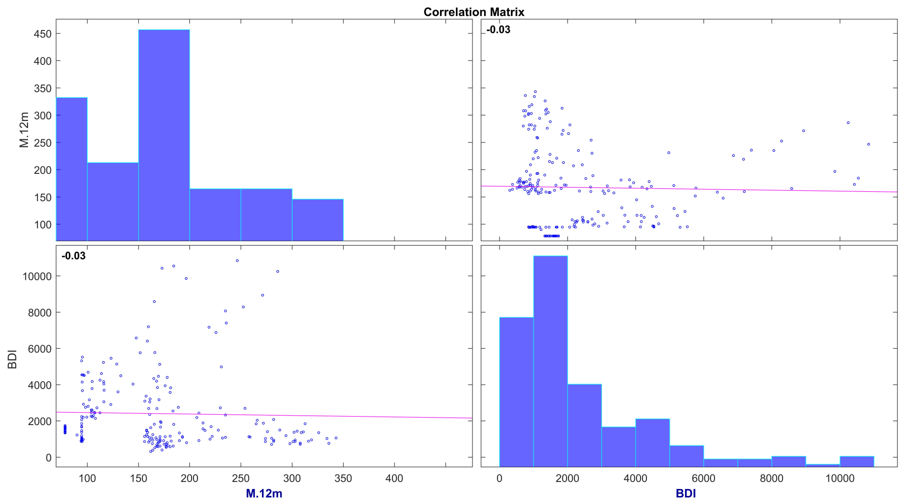

Figure 5.14: Graph that illustrates the correlation between monthly US Gulf Corn Price and monthly BDI and their histograms (Source: Sin.clarksons.net (2018)).

"good" scatter plot in Fig.5.15. Hence, it can be used for the purposes of this thesis as an independent variable.

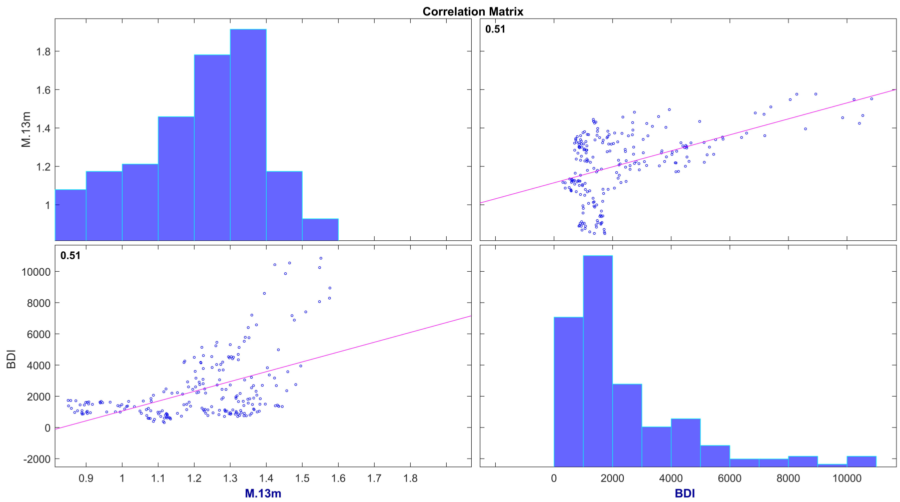

Figure 5.15: Graph that illustrates the correlation between monthly Exchange Rates of Euro against the US Dollar and monthly BDI and their histograms (Source: Sin.clarksons.net (2018)).

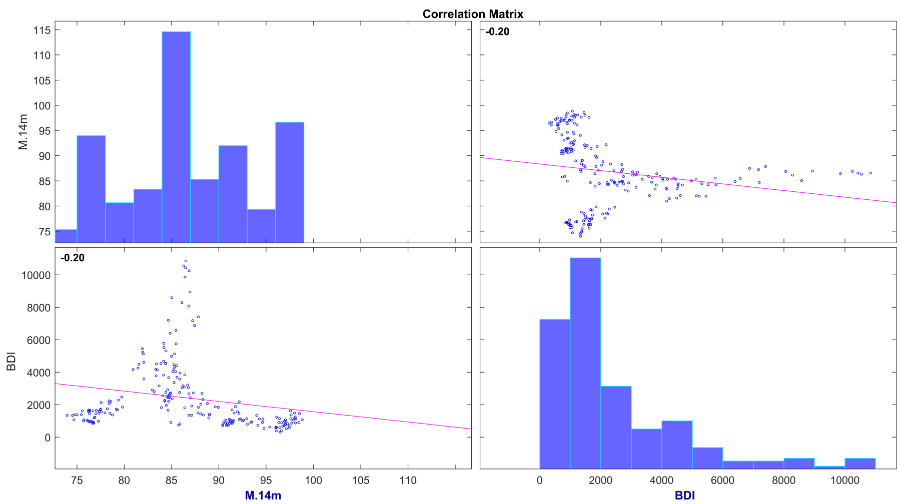

Figure 5.16: Graph that illustrates the correlation between monthly Global Oil Production and monthly BDI and their histograms (Source: Sin.clarksons.net (2018)).

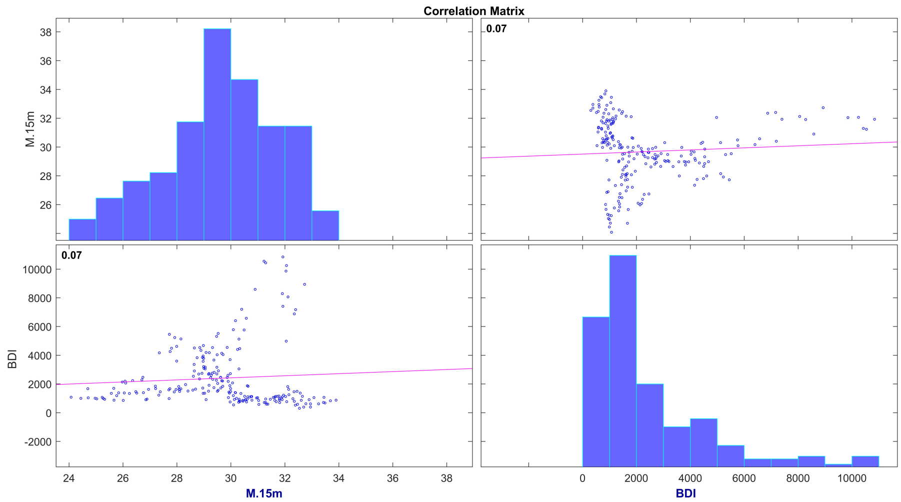

Figure 5.17: Graph that illustrates the correlation between monthly OPEC Crude Oil Production and monthly BDI and their histograms (Source: Sin.clarksons.net (2018)).

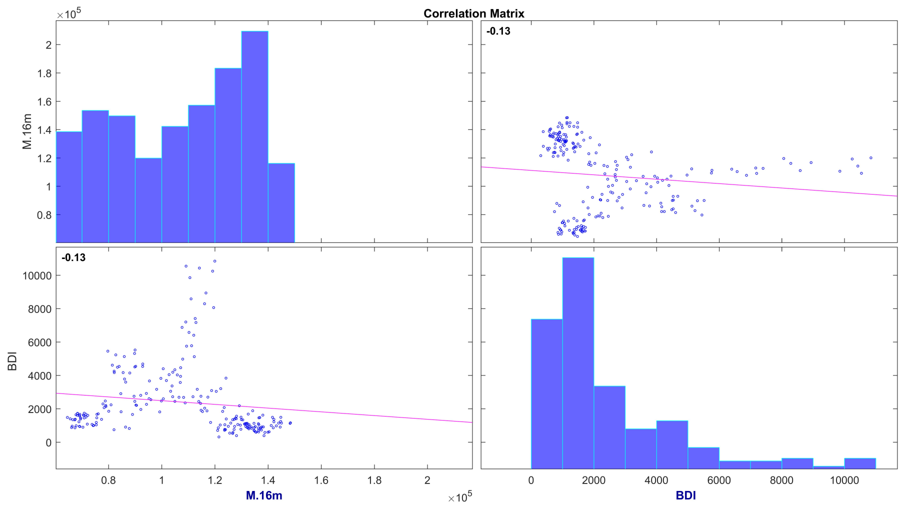

Figure 5.18: Graph that illustrates the correlation between monthly World Steel Production and monthly BDI and their histograms (Source: Sin.clarksons.net (2018)).

5.4 Modeling

The basic idea of each model is to utilize the findings of the correlation analysis as presented in Ch.5.3.2 and extract models based on the most favourable independent variables of the BDI. The models were constructed by the N4SID method as presented in Ch.4.4.1 and from which the linear state space model of the BDI was identified from the input - output datasets.

The main concept of the modeling procedure was that BDI was the dependent variable, which was affected by the independent variable ($Index_i$, where $i = 1, 2, \ldots$), in the following general function representation:

$$BDI = f(Index_1, Index_2, \ldots) \tag{5.1}$$

All the available data, as can be derived by the time frame of Table 5.1, were of a total of 222 months, i.e. from the 1/11/1999 till the 1/4/2018 with monthly intervals.

5.4.1 First Model

As presented in Eq.5.1 this model was based on the following concept:

$$BDI = f(Monthly\_1m,\ Monthly\_2m,\ Monthly\_4m,\ Monthly\_8m,\ Monthly\_6m,\\ Monthly\_10m,\ Monthly\_13m) \tag{5.2}$$

or according to Table 5.1

$$BDI = f(Inflation\ Indicator\ OECD\ Europe\ (excl.\ Turkey),\ Trade\ Weighted\\ Steel\ Production\ Index\ (SPI),\ Industrial\ Production\ of\ China,\\ UK\ Interest\ Rates,\ USA\ Interest\ Rates,\ Brent\ Crude\ Oil\\ Price,\ Exchange\ Rates\ of\ Euro\ against\ US\ Dollar) \tag{5.3}$$

So, as can be derived by Eqs.5.2 and 5.3 for the identification of this model both data suggested by the statistical analysis and not have been used as input of the model. The dependent variable is the BDI and the function f of the equations represent the resulted by the identification model.

In Fig.5.19 the reader can see all the used available data plotted against time. For the time intervals months have been used, as also stated on the figure.

In Fig.5.20 the data used for the training of the model are illustrated. For the training purposes the last 10 month series of the data have been neglected, approximately one (1) year of results, as to have a percentage of unbiased available data for further evaluation reasons.

In Fig.5.21 the model response to the training data can be seen. The model seems to responds well to the given data, with a fitting percentage of around 70%. A good fitting with which we can continue our comparison of the model against all the available data.

Fig.5.22 illustrates the fitting of the model against all the available of the analysis data. One can see that the model is responding surprisingly well, even though that the fitting percentage has fallen by around 2% to a rounded down 68% compared to the Fig.5.21, where the training data have only been used. The usage of the whole

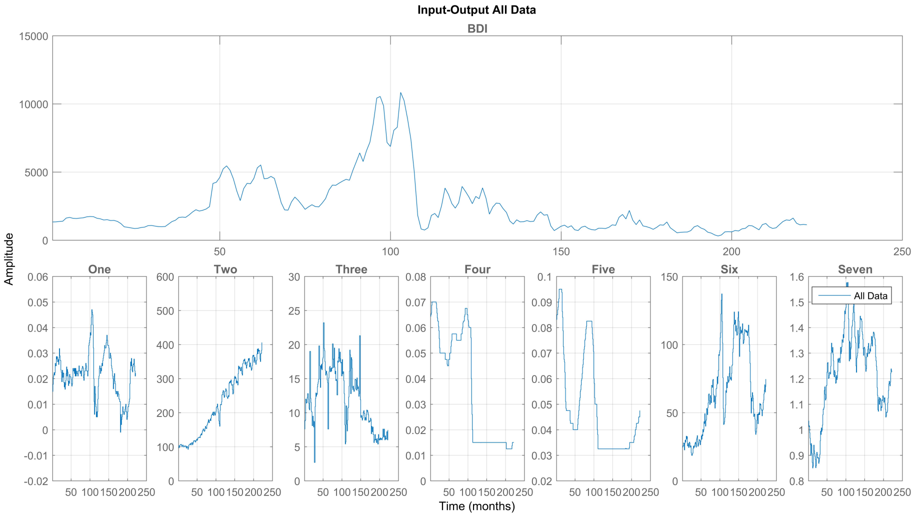

Figure 5.19: All available data.

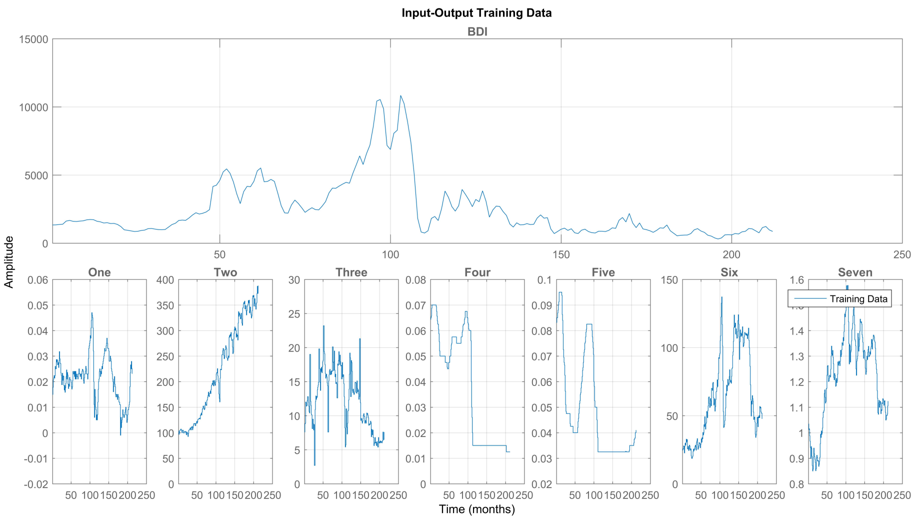

Figure 5.20: Used Training data for the identification of the model.

available data means that a decrease could be expected. Rather than that, the model responds well to every shock and every sudden change of the BDI. Especially in the time frame between 80th month and 120. i.e. the booming and the collapse phase of the dry bulk shipping industry, the model responded surprisingly well. Overall an oscillation around the real value of the BDI can be observed, an expected phenomenon though due to the dependency of the model from other indexes. Regarding the last 10 months, where the model was plotted against data not used during the training phase, the model also responded well following the real trajectory of the BDI. A small rise was expected, especially due to factors like the simultaneous increase of the independent variables during those last 10 months, as can be seen in Fig.5.19. The steady rise of the factor five, i.e. the USA Interest Rates during those months after a steady phase

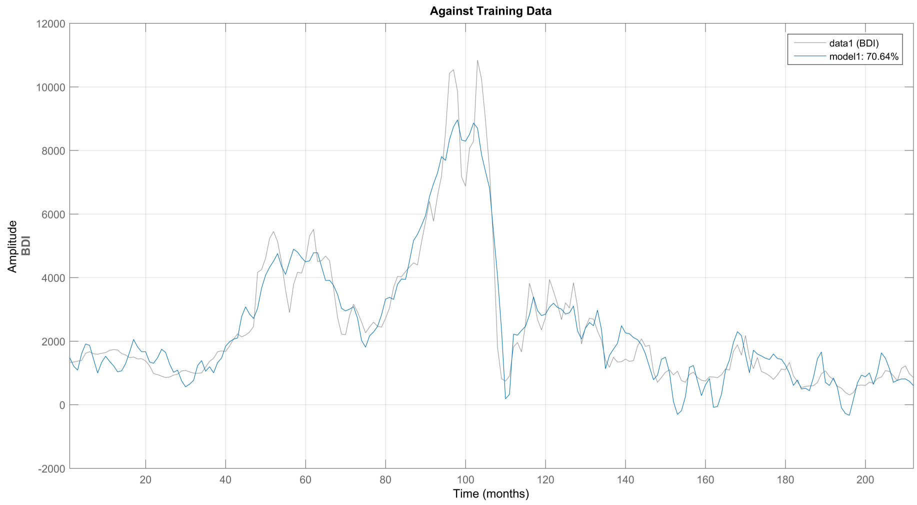

Figure 5.21: Obtained model against training data.

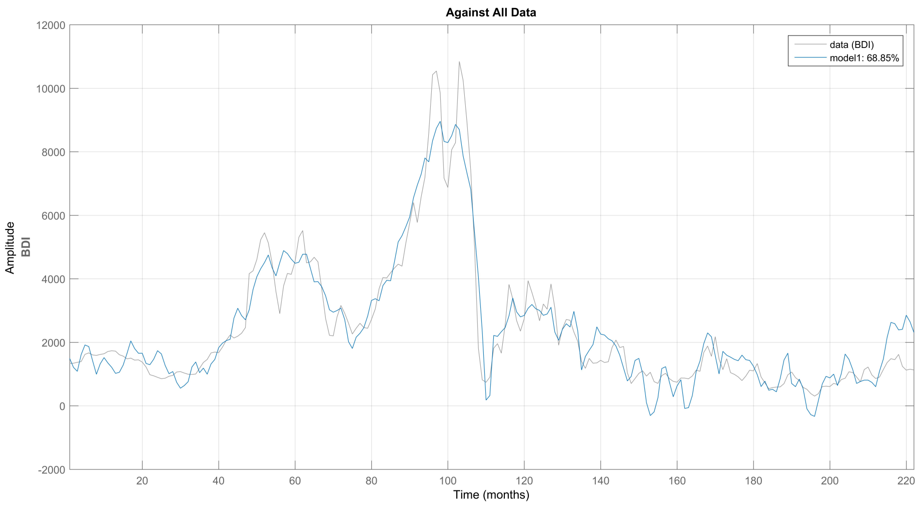

Figure 5.22: Obtained model against all available data.

of about 80 months also contribute to that. A fact that also justified the findings of the statistical analysis.

5.4.2 Second Model

As presented in Eq.5.1 this model was based on the following concept:

$$BDI = f(Monthly\_2m,\ Monthly\_4m,\ Monthly\_9m,\ Monthly\_10m,\ Monthly\_11m) \tag{5.4}$$

or according to Table 5.1

$$BDI = f(\,Trade\ Weighted\ Steel\ Production\ Index\ (SPI),\ Industrial\ Production\ of\ China,\ LIBOR\ Interest\ Rates,\ Brent\ Crude\ Oil\ Price,\ US\ Gulf\ Wheat\ Price) \tag{5.5}$$

So, as can be derived by Eqs.5.4 and 5.5 for the identification of this model only data suggested by the statistical analysis have been used as input of the model. The dependent variable is the BDI and the function f of the equations represent the resulted by the identification model.

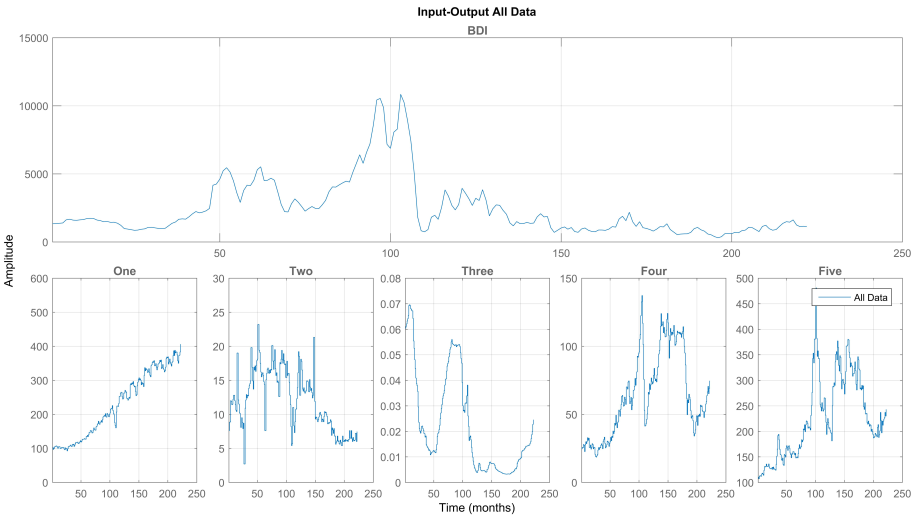

Figure 5.23: All available data.

In Fig.5.23 the reader can see all the used available data plotted against time. For the time intervals months have been used, as also stated on the figure.

In Fig.5.24 the data used for the training of the model are illustrated. For the training purposes the last 25 month series of the data have been neglected, approximately two (2) years of results, as to have a percentage of unbiased available data for further evaluation reasons.

In Fig.5.25 the model response to the training data can be seen. The model seems to responds well to the given data, with a fitting percentage of around 50%. A good fitting with which we can continue our comparison of the model against all the available data.

Fig.5.26 illustrates the fitting of the model against all the available of the analysis data. One can see that the model is responding surprisingly well, compared to the Fig.5.25, where the training data have only been used, no significant difference can be observed in terms of fitting percentages, since a fitting of around 50% was achieved again. The model again responds well to every shock and every sudden change of the BDI. Especially in the time frame between 80th month and 120. i.e. the booming and the collapse phase of the dry bulk shipping industry, the model responded surprisingly well. Also one can see a predictive ability of the model, since it started declining and / or rising prior the real BDI. Overall an oscillation around the real value of the BDI can be observed again, an expected fact due to the dependency of the calculated BDI from

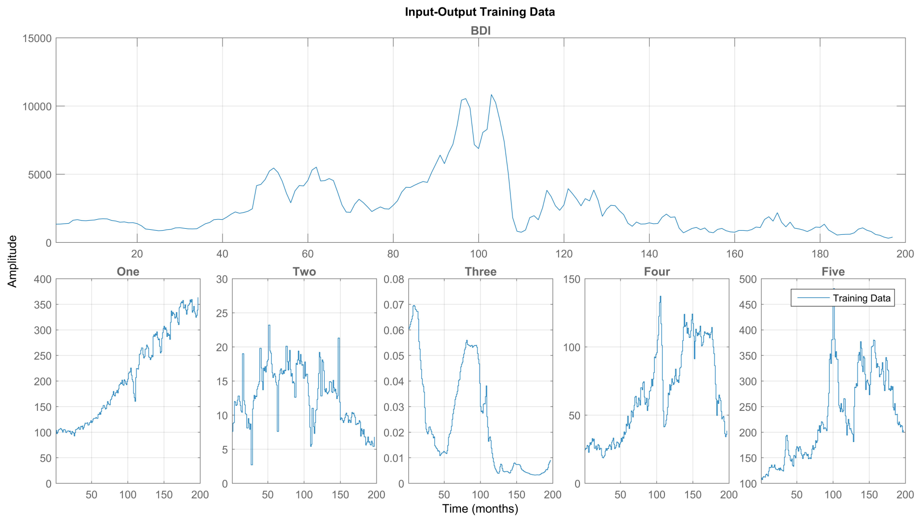

Figure 5.24: Used Training data for the identification of the model.

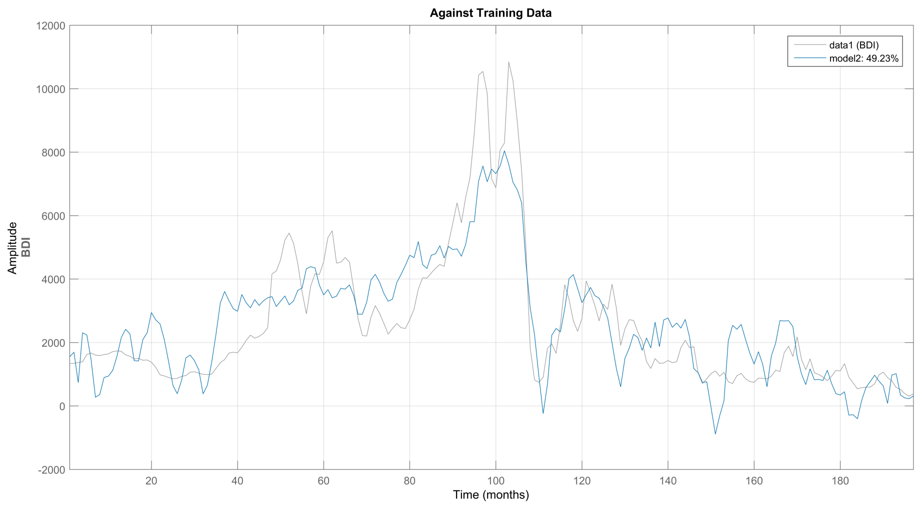

Figure 5.25: Obtained model against training data.

other indexes. Regarding the last 25 months, where the model was plotted against data not used during the training phase, the model responded surprisingly well following the real trajectory of the BDI really closely.

5.4.3 Third Model

As presented in Eq.5.1 this model was based on the following concept:

$$BDI = f(Monthly\_1m,\ Monthly\_2m,\ Monthly\_4m,\ Monthly\_8m,\ Monthly\_6m,\ Monthly\_10m,\ Monthly\_13m) \tag{5.6}$$

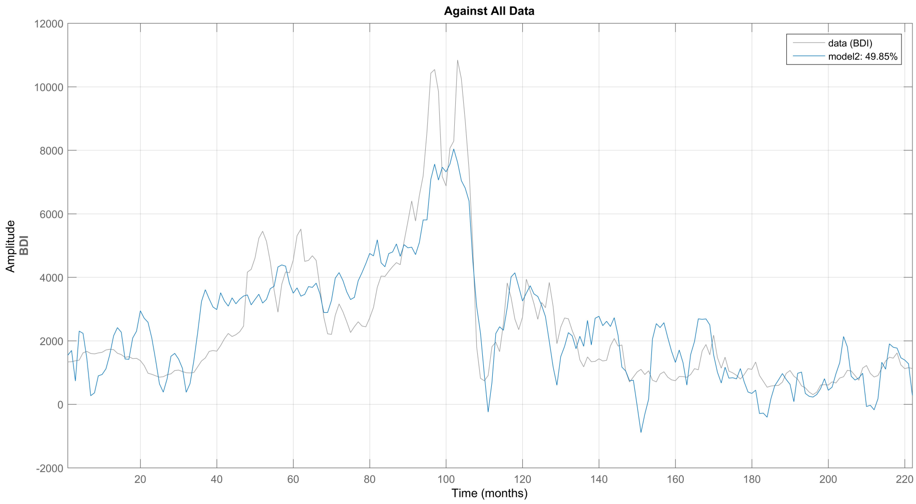

Figure 5.26: Obtained model against all available data.

or according to Table 5.1

$$BDI = f(Inflation\ Indicator\ OECD\ Europe\ (excl.\ Turkey),\ Trade\ Weighted\ Steel\ Production\ Index\ (SPI),\ Industrial\ Production\ of\ China,\ LIBOR\ Interest\ Rates,\ Exchange\ Rates\ of\ Euro\ against\ US\ Dollar) \tag{5.7}$$

So, as can be derived by Eqs.5.6 and 5.7 for the identification of this model only data suggested by the statistical analysis have been used as input of the model. The dependent variable is the BDI and the function f of the equations represent the resulted by the identification model.

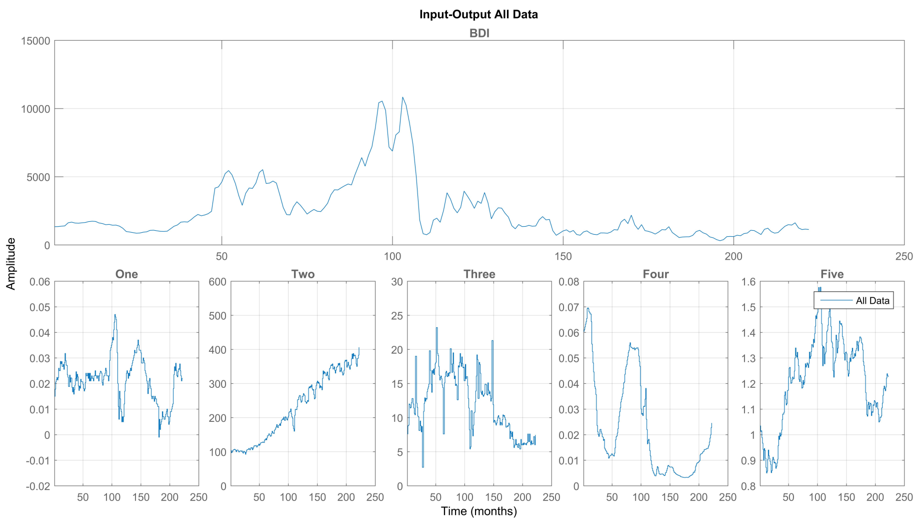

Figure 5.27: All available data.

In Fig.5.27 the reader can see all the used available data plotted against time. For the time intervals months have been used, as also stated on the figure.

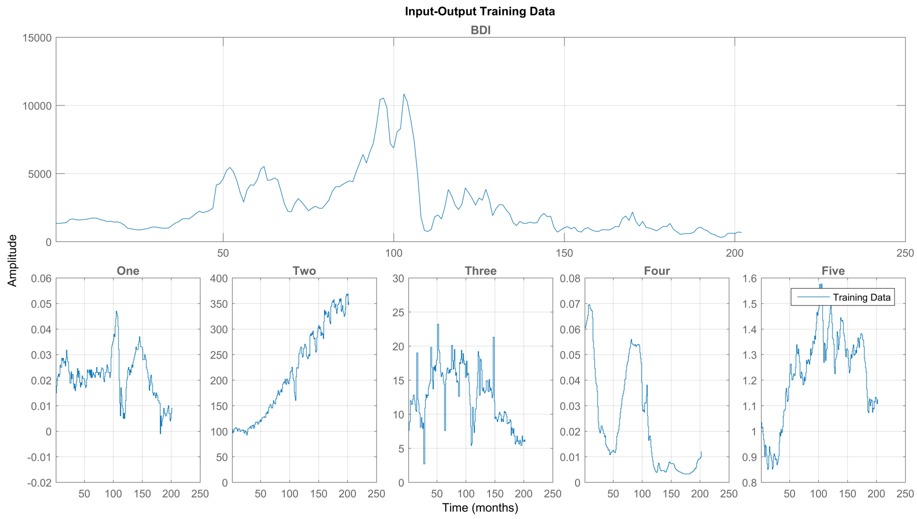

Figure 5.28: Used Training data for the identification of the model.

In Fig.5.28 the data used for the training of the model are illustrated. For the training purposes the last 20 month series of the data have been neglected, approximately one and a half (1.5) years of results, as to have a percentage of unbiased available data for further evaluation reasons.

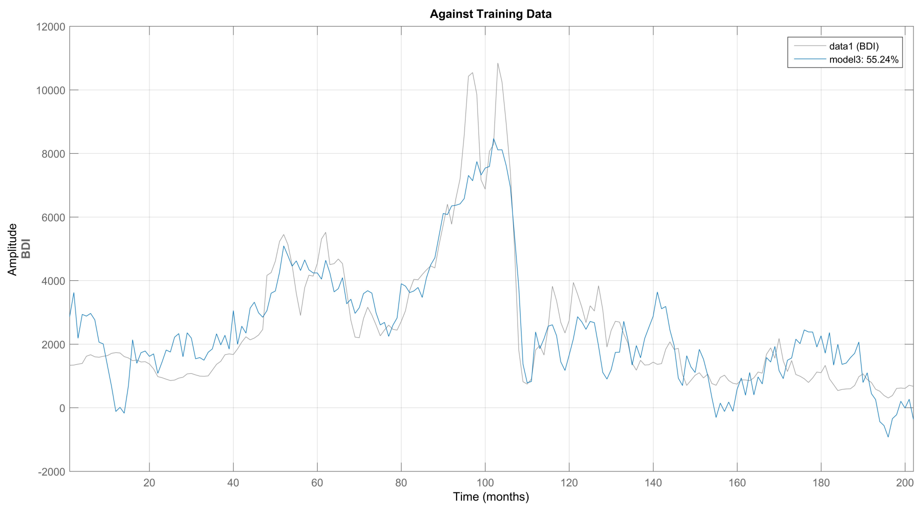

Figure 5.29: Obtained model against training data.

In Fig.5.29 the model response to the training data can be seen. The model seems to responds well to the given data, with a fitting percentage of around 55%. A good fitting with which we can continue our comparison of the model against all the available data.

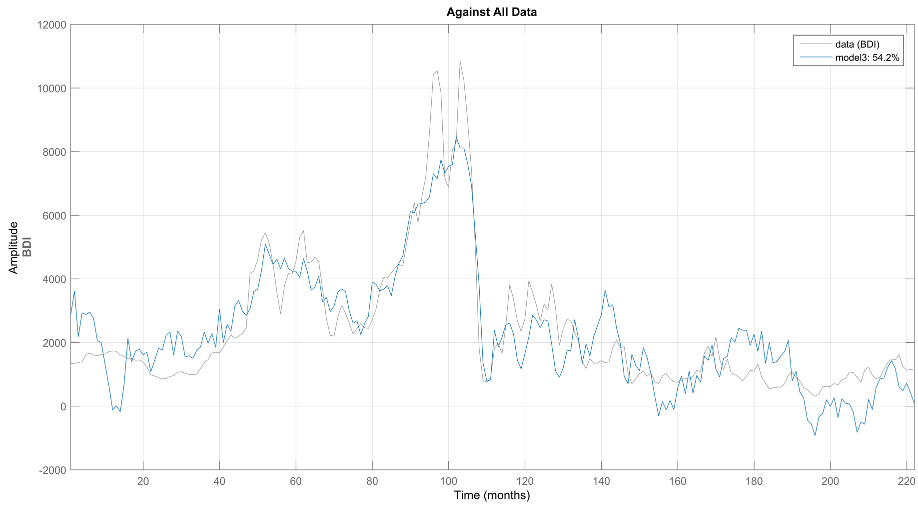

Figure 5.30: Obtained model against all available data.

Fig.5.30 illustrates the fitting of the model against all the available of the analysis data. One can see that the model is responding surprisingly well, with a fitting of 54%. The usage of the whole available data means that a decrease of 1% could be expected. Rather than that, the model responds well to every shock and every sudden change of the BDI. Especially in the time frame between 80th month and 120. i.e. the booming and the collapse phase of the dry bulk shipping industry, the model responded surprisingly well. Overall an oscillation around the real value of the BDI can be observed, an expected phenomenon though due to the dependency of the model from other indexes. Regarding the last 20 months, i.e. the excluded months from the training data, the model also responded well following the real trajectory of the BDI. A small depression of the value of the BDI can only be spotted in the beginning of those months, with a corrected trajectory towards the end of the available periods.

5.5 Discussion / Final choice of Model

Given all the results both from the statistical analysis been done in this thesis, as well as the resulted modelings of the BDI, whose results can be briefly illustrated in Table 5.3, we can conclude on which one of the suggested models is the more optional to derive any conclusions regarding the BDI. Even though all the models behaved in a good manner; following the required trajectory and responding well to the sudden changes of the BDI over the examined period of time; the first model was dependent on variables which were rejected by the statistical analysis for usage. A rejection which was also based on factors like the nature of the indexes, since indexes which where set to a price by external of the market driven forces where also been analysed. Therefore, the first model can be rejected for the further analysis and estimation of the BDI trajectory over time, due to the dependency on such factors. On top of that the increase of one of those factors at the end of the examined period revealed an increase of the projected BDI.

On the other hand the other two obtained models, even the fact that they behaved

Table 5.3: Synopsis of the resulted fittings of the obtained model against both the training and all the available data.

	Fitting against Training Data [%]	Fitting against All Data [%]	Abs. Difference [%]
First Model	70.64	68.85	1.79
Second Model	49.23	49.85	0.62
Third Model	55.24	54.2	1.04

worse in terms of fitting percentages, mainly due to the more freedoms that they had, since only five (5) indexes were used instead of seven (7) in the first model, they were mainly dependent on variables accepted and / or suggested by the statistical analysis. In other words the independent variables were well related with the BDI and are set by the supply / demand equilibrium of the market.

Therefore models two and three are the suggested ones by the author of this thesis for the modeling of the BDI, with the second model revealing a better forecasting ability of the BDI, since as can be illustrated in Fig.5.26, the model responded to the sudden changes usually prior they have occurred to the real BDI trajectory. Therefore it can be a potential good indicator for the following' s months trajectory of the BDI, regardless the fact that it revealed the least percentage of fitting among the examined models.

Chapter 6

Conclusions and Future Work

6.1 Conclusions

In Ch.1 the reader can find the posed main research question with four more sub-questions, which have been tried to analysed and finally conclude through the previous chapters. The main findings of the research that has been done, starting from analysing the sub-questions, were that there are indeed suggested, by the bibliography, economical indexes that are related to the BDI with acceptable correlation coefficients of linear relationship. This relationship has been examined again and the analysis has further provide evidence over the correlation of those indexes between them and worked as an indicator for a modeling attempt of the BDI. Hence, with the usage of both suggested and rejected (for the sake of completeness of the thesis) by the statistical analysis indexes, modeling attempts of the BDI by the usage of data in form of input - output data has been performed. As inputs were the identified as independent variables and as output of the model was always the BDI. The findings of this procedure has further revealed the potentiality of connecting the suggested by the bibliography indexes with the BDI. Eventually the obtained models revealed a good fitting percentage with the real BDI, both against the training data and against all the available data. The modelled BDI responded well to almost any given change over time, including the economic shocks etc..

Therefore, as to answer the main research question, i.e. of whether *"the Baltic Dry Index (BDI) relates with major macroeconomic and/or microeconomic indexes and whether there is a modeling feasibility of the BDI by them"*, we can conclude from the findings of this thesis that there is indeed a relation between the BDI and several economical indicators, both related with the dry bulk shipping industry and not, as well as there is a feasibility of constructing a model based on those indexes, as to predict and examine the rationality of the monthly BDI; i.e. whether the resulted BDI is according to the trajectory that other economical factors indicate.

6.2 Limitations and Future Work

The main limitations while conducting this thesis were regarding the data been obtained from databases. Firstly, as to access the used databases a subscription was essential. Furthermore, not all obtained data were of the same length, meaning that a filtering of the datasets based on the selected data range was essential. In addition, the sample size was not sufficient for an annual investigation, hence an investigation of monthly and/or quarterly intervals was only applicable, since modeling the BDI on a daily basis would not be inside the scope of this work, which was to provide the reader an indicator of the tendency of this index. Regarding the modeling method, even though there was not an extensive usage in prior research examining the BDI, several other applications of N4SID could be found on other non-BDI-relevant topics.

Based on the finding of this thesis, future readers are advised to re-examined the final results with up-to-date data. Moreover, a manipulation of the examined models is advised, such as the introduction of a closed loop state space model based on the suggested ones. Furthermore, the examination of more indexes as potential independent variables is suggested, as well as the testing of more models.

Appendices

Appendix A

Appendix

A.1 Graphical Representation of Used Data

In this section of the appendix the reader can find the graphical representations of all the used data / indexes in this thesis. All the data have been plotted with data within the time frame of 01/11/1999 till the 01/04/2018 with monthly intervals.

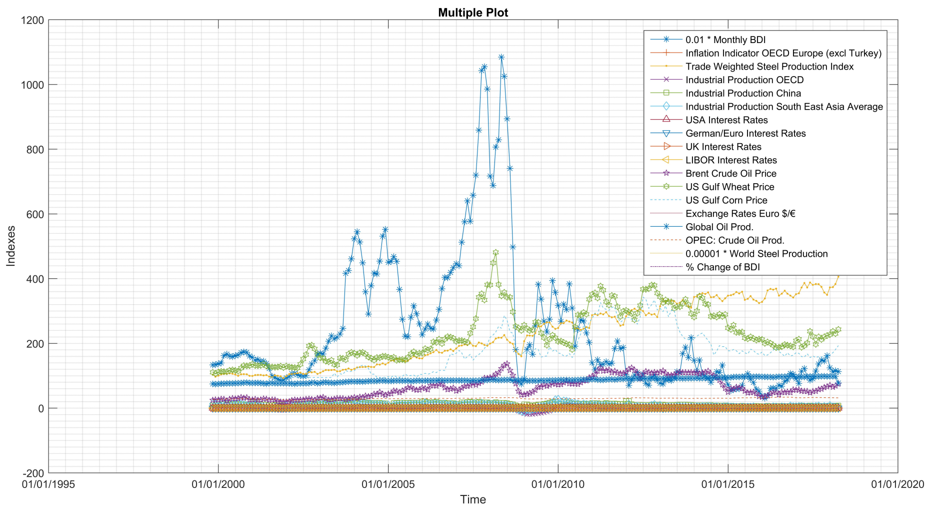

Figure A.1: Graphical Representation of all used data combined (Source: Sin.clarksons.net (2018)).

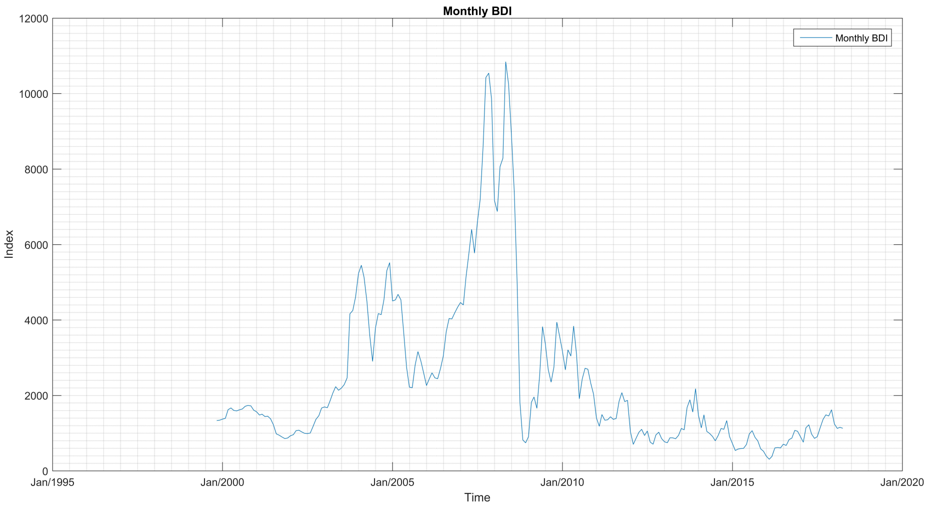

Figure A.2: Graphical Representation of the monthly BDI (Source: Sin.clarksons.net (2018)).

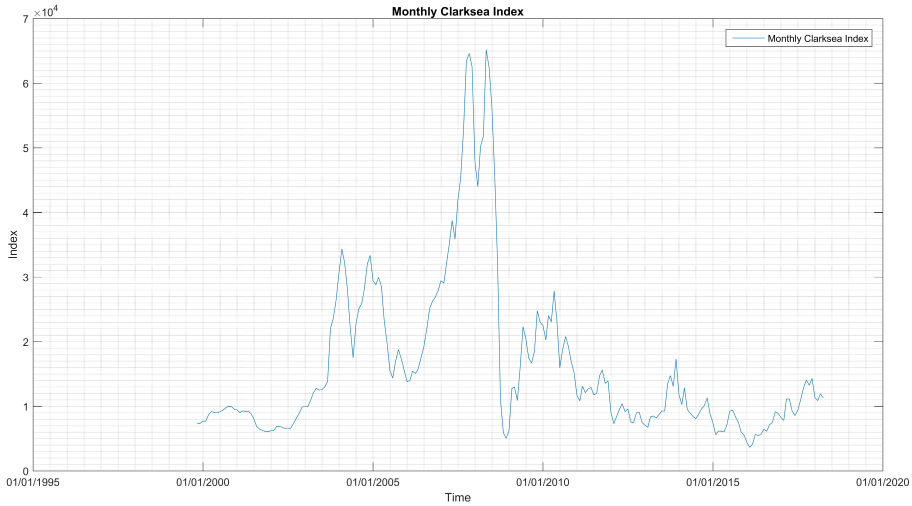

Figure A.3: Graphical Representation of the monthly Clarksea Index (Source: Sin.clarksons.net (2018)).

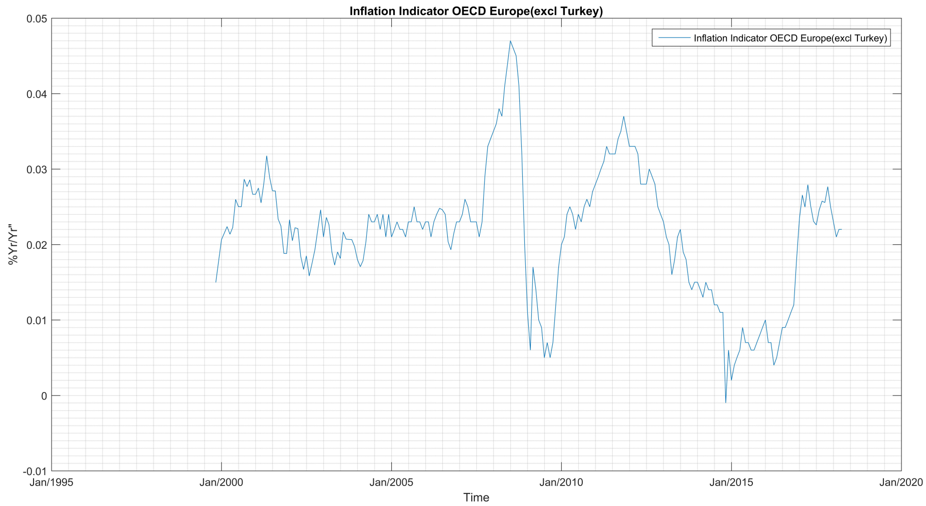

Figure A.4: Graphical Representation of the Inflation Indicator OECD Europe (excl. Turkey) (Source: Sin.clarksons.net (2018)).

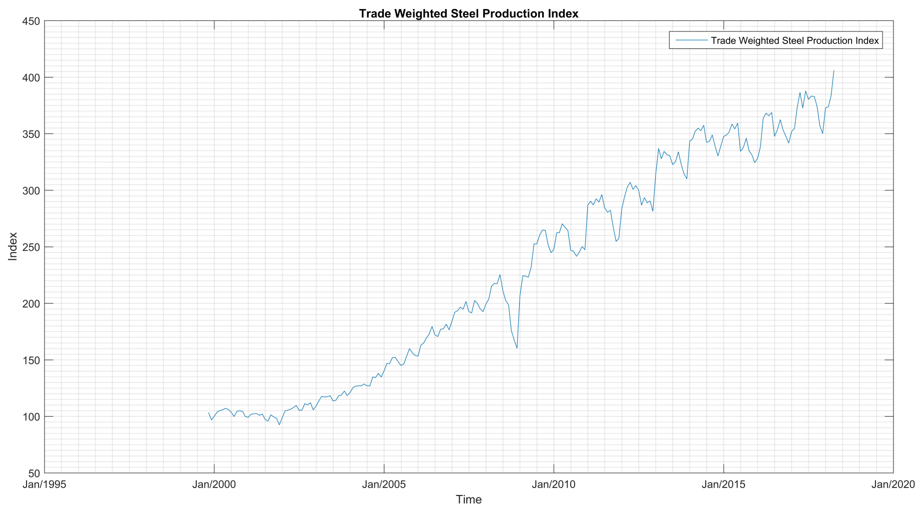

Figure A.5: Graphical Representation of the Trade Weighted Steel Production Index (SPI) (Source: Sin.clarksons.net (2018)).

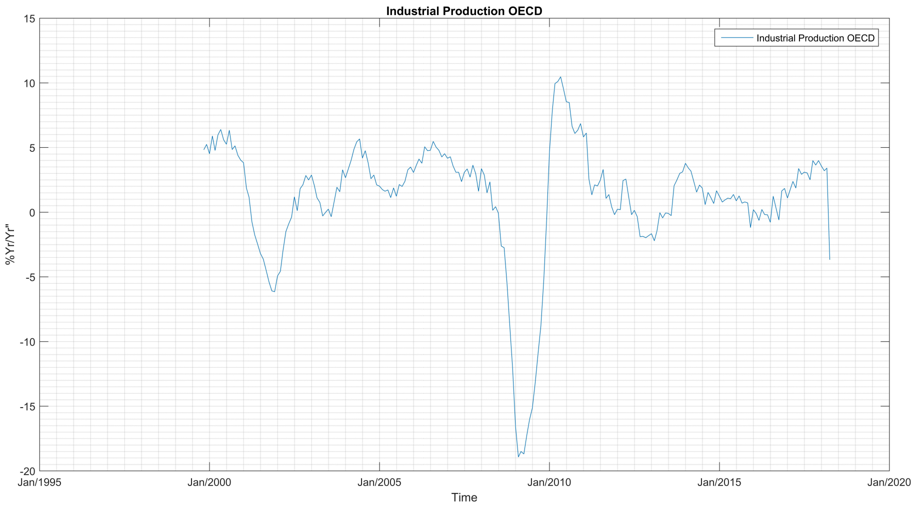

Figure A.6: Graphical Representation of the Industrial Production of OECD (Source: Sin.clarksons.net (2018)).

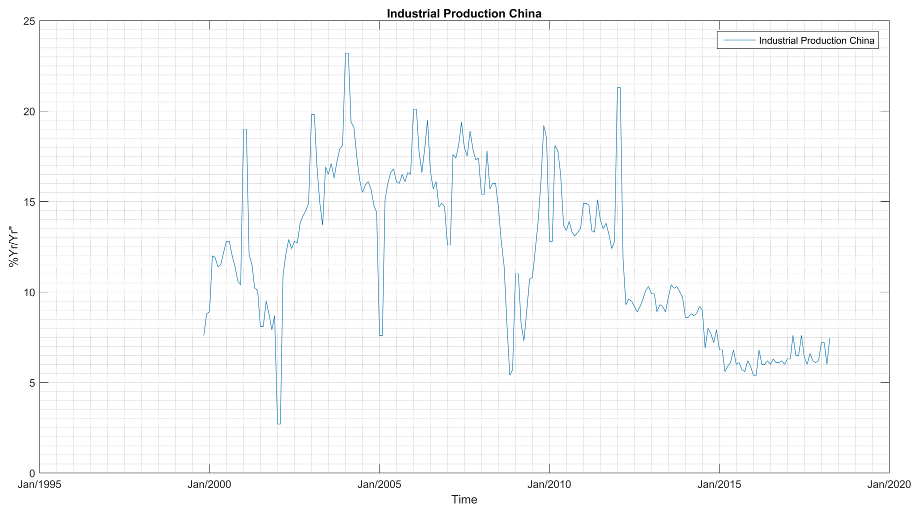

Figure A.7: Graphical Representation of the Industrial Production of China (Source: Sin.clarksons.net (2018)).

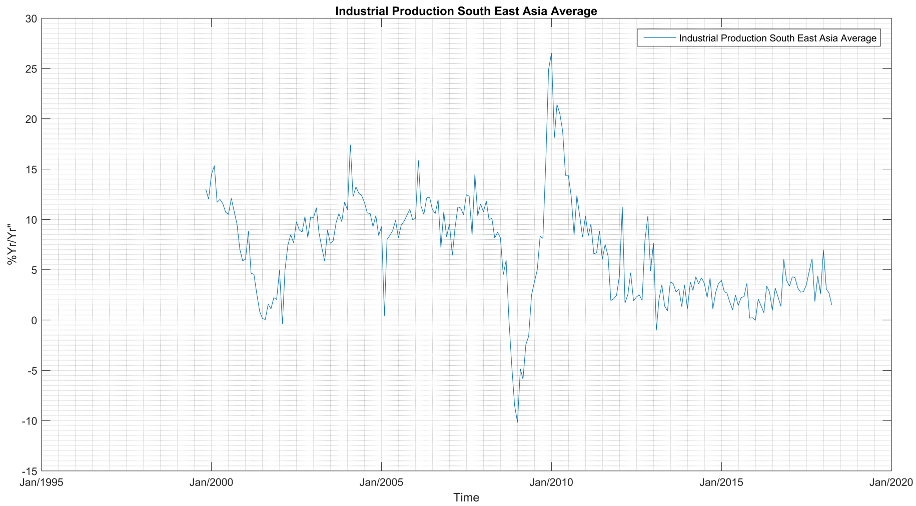

Figure A.8: Graphical Representation of the Averaged Industrial Production of South East Asia (Source: Sin.clarksons.net (2018)).

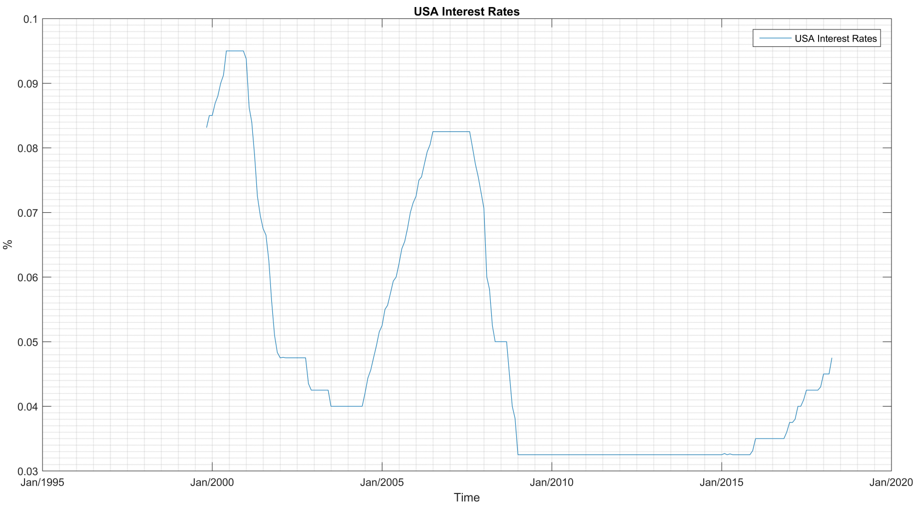

Figure A.9: Graphical Representation of the USA Interest Rates (Source: Sin.clarksons.net (2018)).

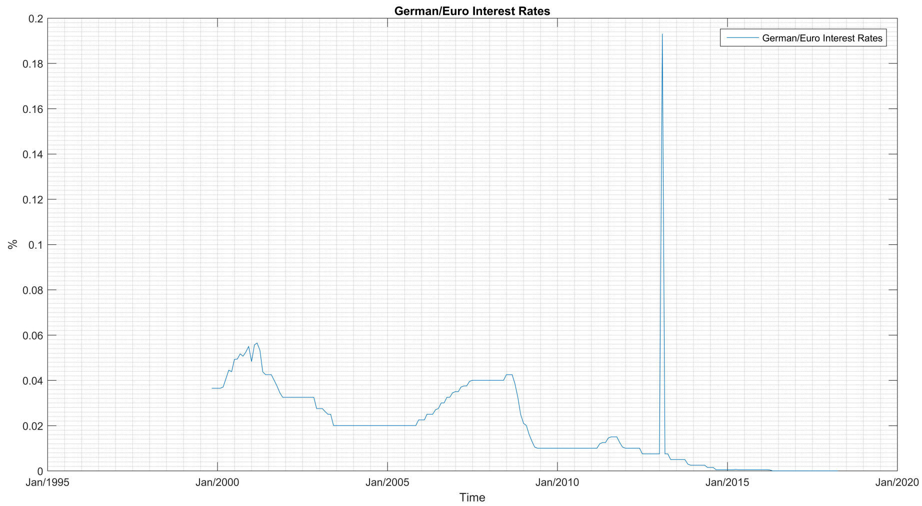

Figure A.10: Graphical Representation of the German / Euro Interest Rates (Source: Sin.clarksons.net (2018)).

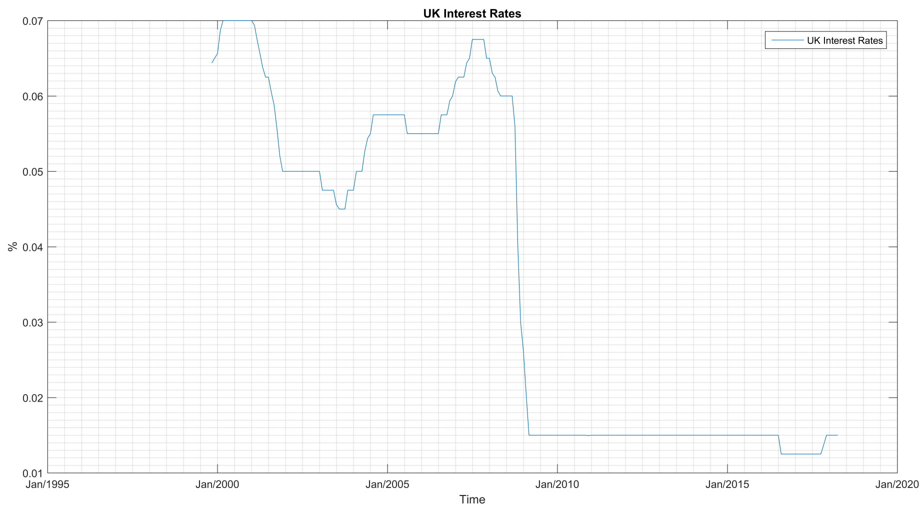

Figure A.11: Graphical Representation of the UK Interest Rates (Source: Sin.clarksons.net (2018)).

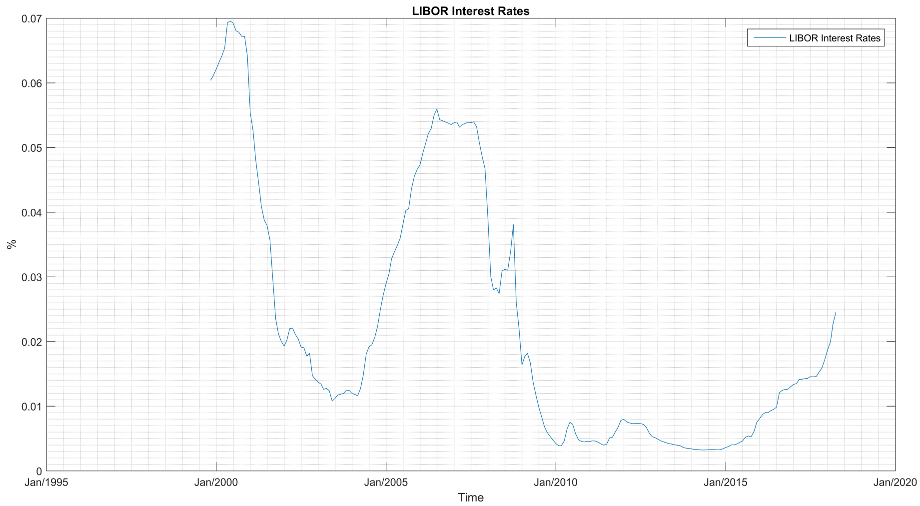

Figure A.12: Graphical Representation of the LIBOR Interest Rates (Source: Sin.clarksons.net (2018)).

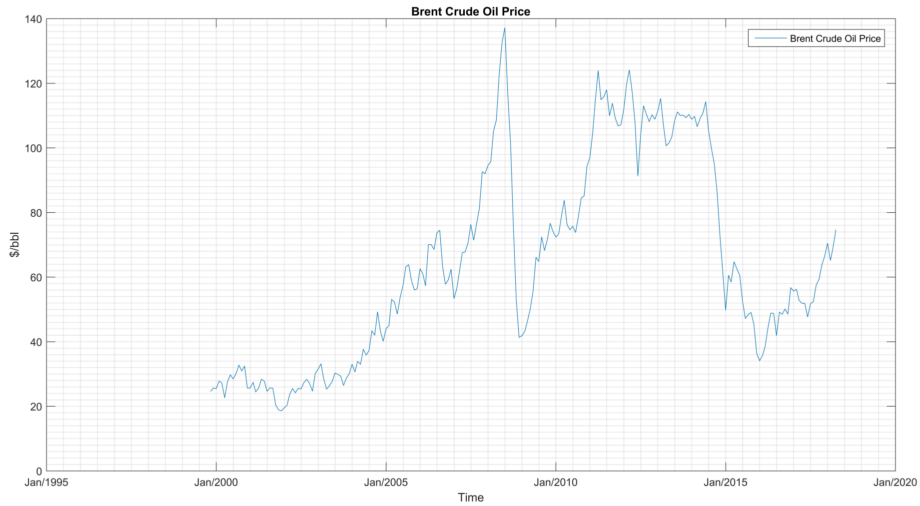

Figure A.13: Graphical Representation of the Brent Crude Oil Price (Source: Sin.clarksons.net (2018)).

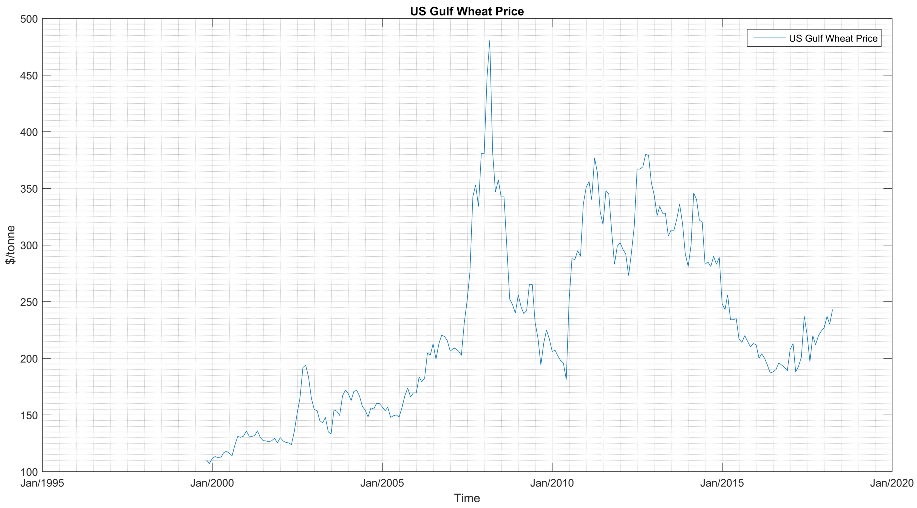

Figure A.14: Graphical Representation of the US Gulf Wheat Price (Source: Sin.clarksons.net (2018)).

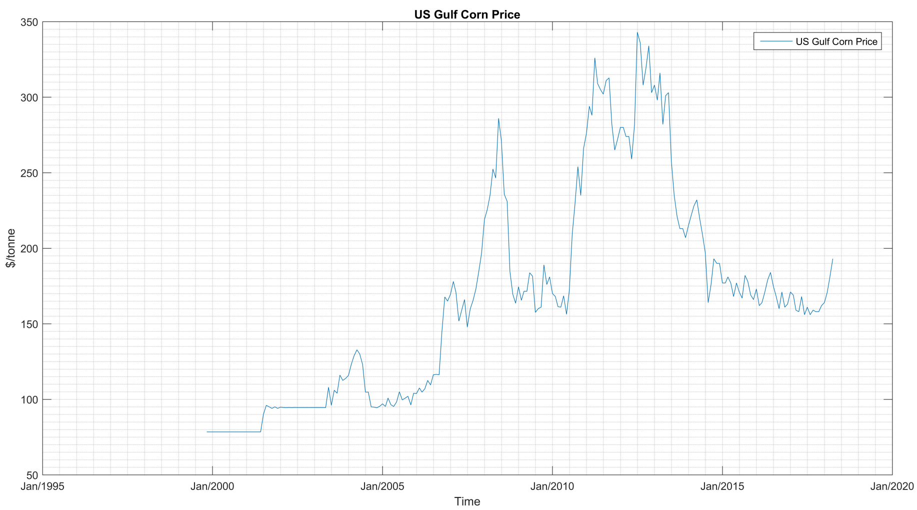

Figure A.15: Graphical Representation of the US Gulf Corn Price (Source: Sin.clarksons.net (2018)).

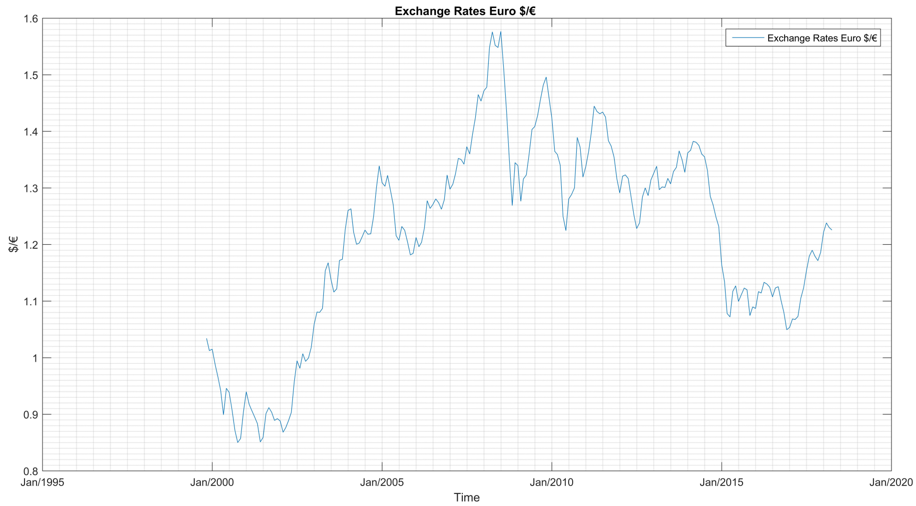

Figure A.16: Graphical Representation of the Exchange Rates of Euro against the US Dollar (Source: Sin.clarksons.net (2018)).

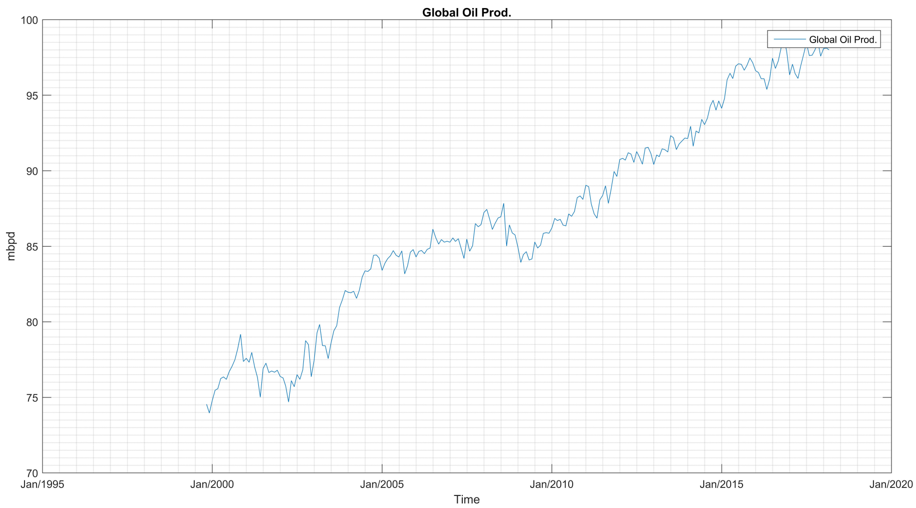

Figure A.17: Graphical Representation of the Global Oil Production (Source: Sin.clarksons.net (2018)).

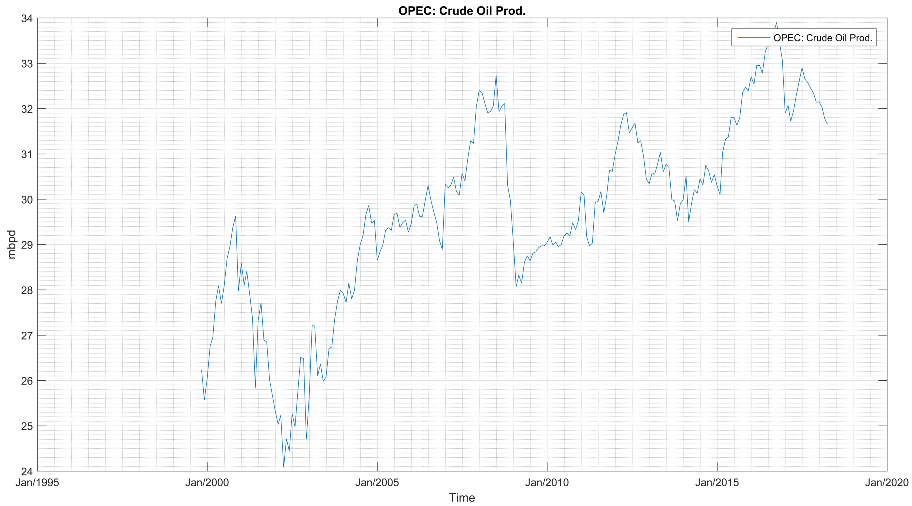

Figure A.18: Graphical Representation of the OPEC Crude Oil Production (Source: Sin.clarksons.net (2018)).

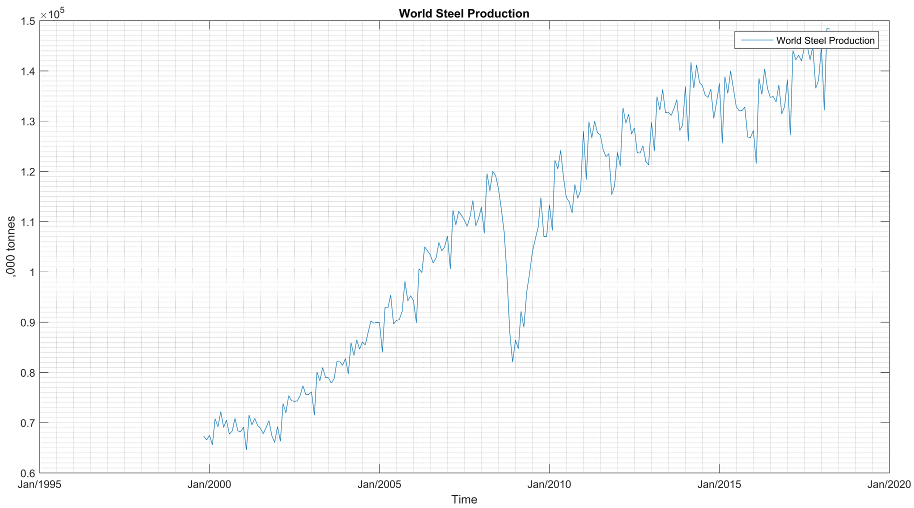

Figure A.19: Graphical Representation of the World Steel Production (Source: Sin.clarksons.net (2018)).

Bibliography

Abdelghani, M., Verhaegen, M., Overschee, P. V. & Moor, B. D. (1998), 'Comparison study of Subspace Identification Methods applied to flexible structures', *Mechanical Systems and Signal Processing* **12**(5), 679 – 692.
URL: *http://www.sciencedirect.com/science/article/pii/S0888327098901698*

Balticexchange.com (2018), 'The Baltic Exchange', [online] Available at: `https://www.balticexchange.com/`. [Accessed 5 Jul. 2018].

Bildirici, M. E., Kayikci, F. & Onat, I. S. O. (2015), 'Baltic Dry Index as a Major Economic Policy Indicator: The relationship with Economic Growth', *Procedia-Social and Behavioral Sciences* **210**, 416–424.

Chen, S. (2011), *Modelling and forecasting in the dry bulk shipping market*, Shun Chen.

Chen, S., Meersman, H. & Van de Voorde, E. (2012), 'Forecasting spot rates at main routes in the dry bulk market', *Maritime Economics & Logistics* **14**(4), 498–537.

Chou, C.-C., Tsai, C.-L. & Lin, K.-S. (2017), 'A regression model for analyzing the relationship between Baltic Dry Index and economic indexes', *Proceedings of 12th International Conference on Eastern Asia Society for Transportation Studies (EASTS 2017)* .

Chou, M. T. & Lin, S. C. (2010), 'An analysis of the relationships between Baltic Dry Index and Steel Price Index - An application of the vector AR model', *Journal of the Chinese Institute of Transportation* **22**(2), 211–232.

Chou, M.-T., Su, Y.-L., Chou, T.-Y. & Liang, H.-U. (2015), 'An analysis of the relationship between Asian Steel Index and the Baltic Capsize Index', *Modern Economy* **6**(02), 207.

Favoreel, W., De Moor, B. & Van Overschee, P. (2000), 'Subspace state space system identification for industrial processes', *Journal of process control* **10**(2-3), 149–155.

G. Divya, C. V. (2014), 'An analysis on Baltic Dry Index', *IRJA-Indian Research Journal* **1**(3).

Goulas, L. (2012), Forecasting freight rates: evidence from the baltic exchange indices and the IMAREX freight futures, Master's thesis, University of Pireaus.

Grammenos, C. (2013), *The handbook of maritime economics and business*, The Grammenos Library, Taylor & Francis.
URL: *https://books.google.gr/books?id=k9Z7Kk2y7yUC*

Iacs.org.uk (2018), 'Safer and Cleaner Shipping - IACS', [online] Available at: http://www.iacs.org.uk/. [Accessed 5 Jul. 2018].

Ics-shipping.org (2018), 'Ics-shipping.org', [online] Available at: http://www.ics-shipping.org/shipping-facts/shipping-and-world-trade. [Accessed 5 Jul. 2018].

Isotalo, J. (2001), 'Basics of statistics', *Finland: University of Tampere* .

Kavussanos, M. G. & Alizadeh-M, A. H. (2001), 'Seasonality patterns in dry bulk shipping spot and time charter freight rates', *Transportation Research Part E: Logistics and Transportation Review* **37**(6), 443–467.

Kavussanos, M. G. & Visvikis, I. D. (2016), *The International Handbook of Shipping Finance*, Palgrave Macmillan UK.

Keller, G. W. L. U. (2013), *Statistics for management and economics*, Cengage Learning, Inc.

Ljung, L. (1999), 'System Identification-Theory for the user', *2nd ed. Prentice Hall, Upper Saddle River, NJ.* .

Lyzell, C. (2009), Initialization methods for system identification, PhD thesis, Linköping University Electronic Press.

Papailias, F., Thomakos, D. D. & Liu, J. (2017), 'The Baltic Dry Index: cyclicalities, forecasting and hedging strategies', *Empirical Economics* **52**(1), 255–282.

Sin.clarksons.net (2018), 'Shipping intelligence network', [online] Available at: https://sin.clarksons.net/. [Accessed 5 Jul. 2018].

Stopford, M. (2013), *Maritime economics*, Routledge.

Tsioumas, V., Papadimitriou, S., Smirlis, Y. & Zahran, S. Z. (2017), 'A novel approach to forecasting the Bulk Freight Market', *The Asian Journal of Shipping and Logistics* **33**(1), 33–41.

Tsoukalas, V. D. & Fragiadakis, N. G. (2016), 'Prediction of occupational risk in the shipbuilding industry using multivariable linear regression and genetic algorithm analysis', *Safety Science* **83**, 12–22.

Van Dokkum, K. (2003), *Ship knowledge*, DOKMAR.

Van Overschee, P. & De Moor, B. (1994), 'N4SID: Subspace algorithms for the identification of combined deterministic-stochastic systems', *Automatica* **30**(1), 75–93.

Van Overschee, P. & De Moor, B. L. (2012), *Subspace identification for linear systems: Theory-Implementation-Applications*, Springer Science & Business Media.

Veenstra, A. W. (1999), Quantitative analysis of shipping markets, Master's thesis, the Netherlands: Delft University Press.

Weiss, N. A. & Hassett, M. J. (1999), *Introductory statistics*, Addison-Wesley Boston.

Xin Yan, X. G. S. (2009), *Linear regression analysis*, World Scientific Publishing Company.