NTNU
Norwegian University of
Science and Technology

Predicting monthly bulk shipping freight rates

Jo Kogstad Ringheim
Lord Olav Daniel Stenslet

Industrial Economics and Technology Management
Submission date: June 2017
Supervisor: Sjur Westgaard, IØT

Norwegian University of Science and Technology
Department of Industrial Economics and Technology Management

Predicting monthly bulk shipping freight rates

Abstract

The main purpose of this thesis is to provide best possible one-month forecasts of both dry bulk (Baltic Dry Index) and tanker (Baltic Dirty Tanker Index) freight rates. In order to find superior leading rate predictors, We apply a general-to-specific methodology as outlined by Campos et al. [2005], where a comprehensive, yet well-justified set of variables are collected. In total, 44 dry bulk variables and 37 tanker time series are gathered with monthly resolution from Jan-2000 to Jun-2016, motivated by previous findings in the literature and common perceptions in the financial markets. The result is a total amount of 264 and 222 potential explanatory variables, as a lead-lag relationship up to six months is taken into account. The large amount of variables are reduced to a subset of predictors through a stepwise regression. The forecasts are calibrated by 150 observations from Jul-2000 to Dec-2012, while the remaining 42 observations constitute the out-of-sample window, where they are compared with relevant univariate benchmarks. We find that i) The single most significant dry bulk predictor is the dry bulk equity index[1]. This finding is interesting, as it implies that shipping stocks tend to move prior to freight rates. It is also in correspondence with the recent findings of Westgaard et al. [2017][2]. The single most significant tanker predictor is the oil price, which is consistent with the findings of Poulakidas and Joutz [2009]. ii) The best out-of-sample result in terms of predictive accuracy is achieved by a univariate seasonal model, for both dry bulk and tanker rates. The forecasts are unable to beat this benchmark in terms of predictive accuracy, but two tanker models are better in terms of correlation. iii) Incorporating effects of deterministic seasonality improve the correlation of all forecasts and simultaneously result in equal or better predicting accuracy. Moreover, as the best model in terms accuracy is the univariate seasonal model, the results are in accordance with the conclusions of Kavussanos and Alizadeh-M [2001, 2002], that bulk freight rates exhibit significant deterministic seasonality. Our findings have implications for shipping participants' operational decision making, as acting upon the model with best predictive accuracy could improve their utility[3].

[1] The composition of the equity index is outlined in Appendix A.3.

[2] Westgaard et al. [2017] found that the OSX index serves as a predictor for oil prices. The OSX index (PHLX Oil Service Sector Index) is designed to track the share price performance of a set of companies involved in the oil services sector.

[3] We formulated our OLS estimators such that they minimise MSE, as it is common to maximise utility by minimising a loss function of second order.

Contents

List of Figures

List of Tables

Acknowledgments

We would like to express our sincere gratitude to our advisor and professor, Sjur Westgaard, for his patience, motivation and knowledge. Our sincere thanks also goes to Andreas Wikborg, shipping analyst at Arctic Securities, for his great inputs and for sharing his immense knowledge of shipping economics with us. Finally, we would like to thank Thomas Lorck, Head of Research at Arctic Securities, for providing us with tools and data along the way.

1 Introduction

Forecasting future market conditions is not optional for shipping market participants. Shipowners make investment, scrapping and charter agreement decisions based on their opinions about the future, where a mistake could induce severe financial consequences, and potentially force companies into bankruptcy. Moreover, other shipping market participants such as banks lending money, regulators making laws or shipyards deciding on which ship design to focus, must all have thoughts about what the future shipping environment will bring. Hence, whether it's about an investment decision, a long-term strategic focus, scrapping of ships, the risk of lending money or consequences of regulations, forecasting is vital for anyone taking part in the world of shipping. Considering the importance of predicting future market conditions well, it is not surprising that there have been extensive efforts put into modeling and forecasting freight rates. But yet numerous of attempts, shipping forecasters have poor track records of predicting them accurately [Beenstock and Vergottis, 1993, Randers and Göluke, 2007, Stopford, 2009]. As stated by Hampton [1991], consensus belief is generally a wrong pointer.

Researchers have used several approaches and techniques to model and forecast bulk freight rates. Early research were generally focused on structural models with a manifold of variables (see, for instance, Tinbergen [1934], Koopmans [1939], Zannetos [1966], Strandenes [1984]). However, research provided by Beenstock and Vergottis [1993] together with development in econometrics in the 1990s, shifted researchers' focus from structural models towards advanced time series models [Glen and Martin, 2005, Chen et al., 2014]. Recent studies have devoted more attention to the co-integration relationships of variables, which structural models had neglected earlier. Additionally, recent time series models have studied the time-varying structure and non-linear dynamics of freight rates (see, for instance, Adland and Cullinane [2005, 2006], Alizadeh and Talley [2011]). Recent econometrics time series models include, among others, Vector Autoregressive models (VAR) [Veenstra and Franses, 1997, Kavussanos and Alizadeh-M, 2001], types of Autoregressive Conditional Heteroskedasticity models (ARCH) [Kavussanos, 1996, Kavussanos and Alizadeh-M, 2002], and other stochastic time series models (see, for instance, Benth et al. [2015], Askari and Montazerin [2015]).

The purpose of our thesis is threefold. (i) First, we aim to determine the leading indicators for both dry bulk and tanker freight rates on a horizon of one month. In academia, the consensus view is that cyclical shipping market movements are largely caused by endogenous supply dynamics, and hence that shipowners scrapping and investment decisions could explain shipping freight rates (see, for instance, [Tinbergen, 1934, Koopmans, 1939, Beenstock and Vergottis, 1993, Randers and Göluke, 2007, Stopford, 2009]). On the other hand, shipping market practitioners usually attribute market movements to sudden shifts in demand characteristics and other non-direct shipping factors [Randers and Göluke, 2007, Stopford, 2009]. In attempting to determine the leading bulk market indicators, we apply a general-to-specific methodology on the principles outlined by Campos et al. [2005], where a comprehensive, yet well-justified set of variables are collected. In total, 44 dry bulk and 37 tanker time series are gathered over the time span from Jan-2000 to Jun-2016, motivated by previous findings and suggestions in literature, and by general perceptions in the financial markets. We categorise the variables into fundamental factors, which cover supply and demand characteristics, and non-fundamental factors, which cover prices or measures that are determined by or traded in financial markets. We find that the most significant dry bulk variables are Chinese steel production, U.S. consumer price index, a dollar exchange rate index and two variables of a dry bulk equity index[4]. The fact that a stock index can be used for prediction purposes was recently shown by Westgaard et al. [2017], who concluded that the PHLX Oil Service Sector Index serves as a good predictor for oil prices. For tanker rates, we find that the most significant tanker indicators are Chinese oil imports, oil prices, U.S. consumer price index, second-hand tanker values and fuel prices.

(ii) Second, we aim to investigate whether it is possible to construct forecasting models with predictive power on a one-month horizon. The prediction models are calibrated by 150 observations from Jul-2000 to Dec-2012, while the remaining 42 observations constitute the out-of-sample testing window. In order to evaluate the usefulness of these forecasts we adhere to the guidance proposed by Hyndman [2010], by comparing them with relevant univariate benchmarks, like an in-sample optimised ARIMA model and a random walk. The prediction models and benchmarks are compared based on their predictive accuracy (measured in $MASE$ and scaled $RMSE$) and explained variance (measured in correlation). Furthermore, we apply the Diebold-Mariano test framework in order to evaluate whether the predictive accuracy of two models are significantly different. In the dry bulk market, we find that our best prediction model is unable to beat the benchmarks both in terms of predictive accuracy and correlation, while in the tanker market, we find that the best prediction model beats the benchmarks in terms of correlation, but not in terms of predictive accuracy.

iii) Third, we investigate whether the findings of Kavussanos and Alizadeh-M [2001, 2002] apply in the post-millennial shipping market. During the time span from 1978-1996, they found no evidence of stochastic seasonality in the dry bulk nor the tanker market. On the other hand, they found evidence of deterministic seasonality in both markets, and accordingly proposed that prediction models within both markets would benefit from taking deterministic seasonality into account. Thus, for every non-seasonal[5] model we create, we correspondingly create a "twin" model, where deterministic seasonality is taken into account. In order to specify seasonally adjusted models, we define a set of seasonal components associated with each market. The seasonal components can both be regarded as a seasonal model and a forecasting benchmark, as it is a univariate model. In the dry bulk market, we find that the seasonal components themselves outperform all models and benchmarks in terms of predictive accuracy and correlation. In the tanker market, we find that the seasonal components outperforms all models in terms of predictive accuracy, while they are beaten by two models in terms of correlation - both seasonally adjusted. Thus, our results are consistent with the findings of Kavussanos and Alizadeh-M [2001, 2002].

This paper is organised as follows. In the next chapter, we provide an introduction to shipping market economics. In chapter 3, we provide a study of literature and present freight rate determinants that previously have been discussed in relation to freight rates. We furthermore discuss the inclusion of every variable added to our models in chapter 4, before outlining the methodology in the following chapter. The results are presented in chapter 6 before we finally arrive at the conclusion in chapter 7.

[4]The composition of the equity index is outlined in Appendix A.3.

[5]"Non-seasonal" refers to models that are fitted to the original dependent variable.

2 Shipping market economics

The aim of this section is to provide an understanding of the economic mechanisms of the maritime industry, and how supply and demand dynamics of sea transport are determined. This section is a decisive step in understanding and determining the model indicators presented in section 4. We start by dividing the shipping industry into three segments depending on the cargo that is carried: bulk, liner, and specialised transport. We next discuss the markets in which shipping companies participate, and how these determine the long-run supply characteristics of the industry. We further provide a thorough presentation of the bulk segment, which we have divided into tanker and dry bulk. An overview of the major bulk trades and their geographical distribution is further presented. Finally, we explain various types of freight agreements and describe the freight rates to be analysed throughout this paper.

2.1. The segments of shipping

There are several sectors within the shipping industry. Stopford [2009] splits the industry into three separate segments: bulk transport, liner transport and specialised cargo transport. Bulk is the largest segment with more than two thirds of all seaborne trade in terms of tons transported, and concerns primarily shipment of large parcels of a single commodity or material. The liner service involves transport of small parcels of cargo, which are often shipped in standardised containers. For instance, the cargo can be smaller quantities of commodities and metals, or refrigerated and liquid items. In the liner segment, the cost of shipment is substantially higher per unit than for bulk transport, and shipping companies involved are generally larger and more complex. Specialised shipping can be regarded as somewhere between bulk and liner transport, and ships are designed to carry a specific cargo type such as motor vehicles, refrigerated food, liquid gas or chemicals, to mention some.

Each segment is attached to a fleet that matches the specific needs of the cargoes they carry. The bulk fleet constitutes the largest share with about three quarters of shipping's total deadweight tonnage (dwt)[6], which is a measure of how much a ship is constructed to carry. Dwt is the sum of the cargo, ballast, crew, fuel, and other weight carried by the ship. Total cargo fleet stands at more than two billion dwt. We note that the borderline between bulk and specialised cargo is not well defined, as both segments transport commodities and cargo in large quantities. Nevertheless, the primary distinction between them is that specialised transport utilise ships that are designed for a specific kind of cargo. Hence liquid gas, refrigerated foods, forest products or chemical parcels are examples of cargoes falling into this category.

2.2. Supply dynamics and shipping's four markets

The shipping industry covers four closely related and dynamic markets: the newbuilding market; the freight market; the second-hand market; and the demolition market. These are important as they represent the markets in which shipping companies can participate. In brief, shipowners trade vessel services in the freight market, sell and purchase vessels in the second-hand market, order new ships in the newbuilding market,

[6]Source: Bloomberg [12 February 2016]

and sell vessels to scrapyards in the demolition market. Figure 2.1 shows how the markets are related with each other and how payments and goods the flow between them. Shipowners face a series of strategical decisions. They can charter out their vessels in the freight market in return for payment from the charterer. The various types of freight agreements are explained in subsection 2.4. Shipowners can use funds to purchase vessels in the second-hand or newbuilding markets, respectively purchasing a ship from another shipowner or contracting a new vessel from a shipyard. Shipowners have also the option of trading their ships to demolition yards or in the second-hand market in return for cash.

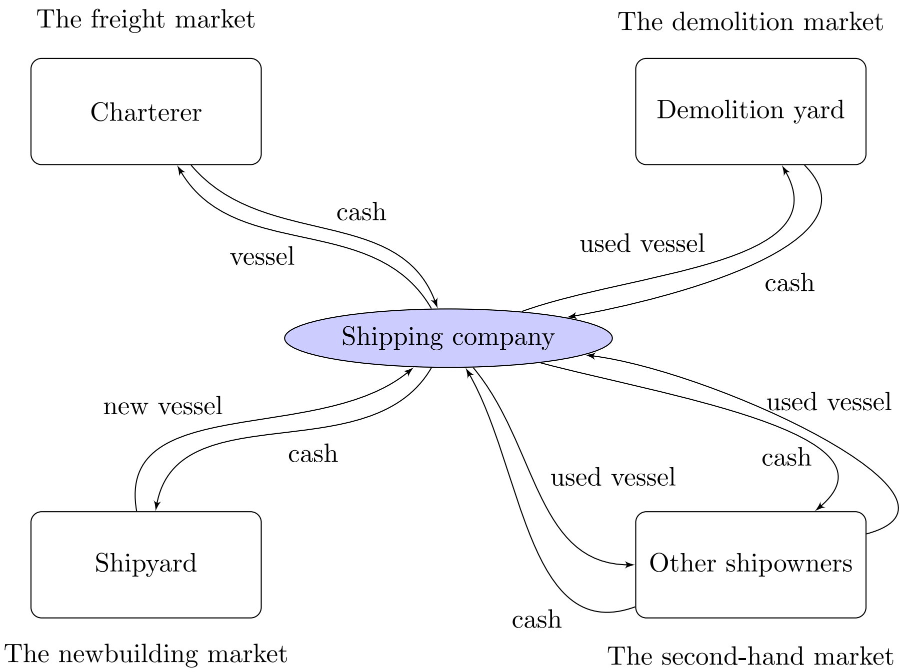

Figure 2.1: The dynamics of the four industry segments of shipping. Inspired by Stopford [2009].

These four interrelated markets form shipping's supply characteristics. Based on Koopmans [1939] findings, it is well recognised in literature that the cyclical nature of shipping is determined by how shipowners trade in these four markets [Randers and Göluke, 2007, Stopford, 2009, Goulielmos, 2010, Anyanwu, 2013]. In periods of excess demand for shipping tonnage, rates tend to spike and shipowners earn profitable returns. Shipowners then tend to invest in new tonnage and few old ones are sold to scrapyards for demolition. Returns are furthermore profitable until new ships have been built, entered the market, and supply has offset demand. Then conditions suddenly shifts from excess demand to excess supply, causing rates to fall. In a distressed environment with poor rates and low returns, shipowners scrap ships or sell them in the second-hand market as they are not profitable or liquidity is needed in order to avoid bankruptcy. Few new ships are ordered and rates stay low until supply falls to the levels of demand or demand surges and catches up with supply. Second-hand ships play a vital economic role in the shipping industry. It gives investors and shipowners the opportunity to buy and sell ships directly, hence allowing direct entry and exit to the freight market. The fact that a ship can change owner relatively swift in the second-hand market, while it takes two-three years to build a new ship, makes it possible to draw some inferences about the market expectations by inspecting the second-hand price relative to the newbuild price. For instance, when freight rates are soaring, second-hand prices may actually be higher than newbuild prices, implying that the market expects freight rates to remain high enough the next couple of years to justify the relatively shorter life expectancy of a second-hand ship. Stopford [2009] found that a shipping cycles averages eight years, but that there are no general rules about the length and timing of the cycle periods. Although shipping cycles are well known and recognised, determining cyclical turning points are a difficult task which shipping market participants devote a lot of time to. This is because each cycle is different, both in nature, circumstances, magnitude and length [Hawdon, 1978, Randers and Göluke, 2007, Stopford, 2009].

Even though shipowners are the ones to decide how to trade with shipyards, demolition yards, charterers and other shipping companies, there are other decision-makers with significant influence on shipowners' behaviour. Banks and investors finance the industry, and their willingness to lend or invest determine shipowners' leeway. For instance, when shipowners have distressed balance sheets, banks can force them to scrap or sell ships in order to to free liquidity and repay debt. Regulators can introduce new legislation and regulatory frameworks, limiting or setting guidelines for shipowners' actions. Charterers could also affect shipowners by becoming shipowners themselves, like some oil companies who ship their crude on owned tankers. Finally, we emphasize that shipping supply is behavioural and dependent on expectations and decisions of a small group of players, and that the lag between investment decisions and market impact intensify the shipping cycles. But the way the four markets interact with each other is not the end of the story. According to Stopford [2009], freight rates are a blend of current and future expectations, and hence we must be precise about which time-frame that is used when explaining rate dynamics. He further presents three time periods to consider: momentary, being the spot market shipowners and charters deal with each day; short-term, when owners and charters have time to respond to rate movements by sending ships in or out of lay-up; and long-term, where the fleet can be adjusted by the scrapping or ordering of ships. Stopford argues that these markets have considerably different dynamics, and that only the long-term dynamics, which concerns a time-frame larger than 2-3 years, can be adequately explained by fundamental variables.

2.2.1. Bulk shipping fleet

Most bulk cargo is transported unpacked in large quantities on long-haul trading distances. A bulk ship generally transports one type of cargo at a time, and the largest vessels typically transport the commodities traded in largest volumes. Bulk cargo can either be liquid or dry. Crude oil is the most important liquid bulk cargo, and is transported in tankers from production facilities to refineries producing gasoline and other oil products. These products are then carried in smaller product, or *clean*, tankers to their destinations. Dry bulk cargo is categorised as either major or minor bulk. Iron ore, coal and grain are referred to as the major bulk commodities.

The tanker and dry bulk fleets consist of ships of different sizes, where many are named after the canals they are able to transit (e.g. Suezmax includes vessels capable of transiting the Suez canal). The largest ships, VLCCs (crude oil) and Capesizes (dry bulk), carry weights of more than 200,000 and 100,000 dwt, respectively. Below is an overview of the bulk fleet, different vessel categories, and the cargo that generally is transported by each type of vessel. Capesizes, the largest dry bulk vessels, primarily transport iron ore and coal. Very large crude carriers (VLCCs) transport crude oil.

Table 2.1: Bulk shipping fleet split by segment and ship type. Source: Clarkson Research Services [15 February 2017], Chen et al. [2014].

Segment	Number of vessels	Vessel size (k.dwt)	Fleet size (mill.dwt)	Main cargo
Tanker fleet				
VLCC	709	>200	218	Crude oil
Suezmax	519	120-200	81	Crude oil
Aframax	976	80-120	106	Crude oil, oil products
Panamax	441	60-80	32	Oil products
Handy	3,776	10-60	125	Oil products
Total tankers*	**6,421**		**562**	
Dry bulk fleet				
Capesize	1,660	>100	317	Iron ore, coal, grain
Panamax	2,472	60-100	198	Iron ore, coal, grain, bauxite, phosphate
Handymax	3,483	40-60	191	Grain, coal, steels, cement, potash, rice, sugar and other minor bulks
Handy	3,335	10-40	95	Gypsum, scrap, sulphur, steels, rice, salt and other minor bulks
Total dry bulk	**10,950**		**801**	
Total bulk fleet	17,371		1,363	

*Does not include tankers of <10k dwt.

2.3. Demand for bulk transport

Shipping companies make a living by transporting goods from one place to another. Bulk cargoes constitutes the largest share of seaborne trading volumes (71% in 2016 measured in weight), and is often shipped in large quantities on long-haul freight routes. The major bulk commodities are iron ore, coal and grain, and together they made up 38% and 27% of bulk and total seaborne trade in 2016, respectively. In the same period, crude oil trade stood at roughly a quarter of all seaborne bulk trade. In order to quantify demand for seaborne transport, we must also examine how far a cargo is transported. This is done by determining the average trading distances, which is referred to as the *average haul* of the trade. Multiplied with the amount traded of a respective cargo furthermore gives us the transport demand, which is determined in terms of *ton-miles*.

Table 2.2 provides a trade overview of the share and amount in terms of volumes and ton-miles, for the most important bulk cargoes.

Table 2.2: 2016 annual seaborne bulk trades in tons and ton-miles. Source: Clarkson Research Services

Cargo	Million tons	Share	Billion ton-miles	Share
Crude oil	1,942	24.6%	9,399	23.7%
Iron ore	1,418	18.0%	8,035	20.3%
Coal	1,130	14.3%	4,903	12.4%
Grain	471	6.0%	3,376	8.5%
Minor bulk	1,860	23.6%	10,819	27.3%
Oil products	1,069	13.5%	3,104	7.8%
Total bulk trade	7,890	100%	39,636	100%
Total seaborne trade	*11,101*		*54,936*	

Comparing the commodities' seaborne trading volumes and their ton-mile contribution reveals interesting trade characteristics. Iron ore is the most important dry bulk cargo. We can observe that its share of ton-miles is higher than its share of total volume, and thus we can conclude that iron ore's average haul is greater than the average haul for total bulk cargo. This fact is largely due to big quantities of iron ore being shipped on long-haul trades between Brazil to China.

We complete the bulk transport presentation by examining the geographical distribution of the most important bulk trades. This is important because the largest importers and exporters of these trades can have significant impact on the amount that is traded. By analysing the regions that are the main trades of these cargoes, we could evaluate potential changes in the trade dynamics. We have chosen ton-miles as the most important trade measure for each commodity, hence we rank importers and exporters after that measure. Figure 2.2 provides an overview of the most important importers and exporters of crude oil, iron ore, coal and grain in terms of ton-miles. We elaborate on these rankings when determining indicators for demand in section 4.

2.4. Freight rates

Shipowners receive payments from charterers in return for vessel services. However, there are several different types of freight agreements. Also, there are different ship sizes, numerous trading routes and plenty of different cargo types. So how can we settle on universal dry bulk and tanker rates?

In order to understand how we can determine a representative freight rate for each of the bulk segments, the types of charter agreements must be evaluated. There are four main types of contractual freight agreements between shipping companies and charterers [Stopford, 2009]. These are different in the way they distribute responsibilities and costs. The first is the *voyage charter* contract, which provides transport for a specific shipload of cargo from one specific port to another. In this case the shipper pays a price for transporting the cargo from A to B, and the shipowner pays all costs related to the trade. The next type of agreement is the *contract of affreightment*. This contract is similar to the voyage charter, but somewhat more complex. The shipping company agrees

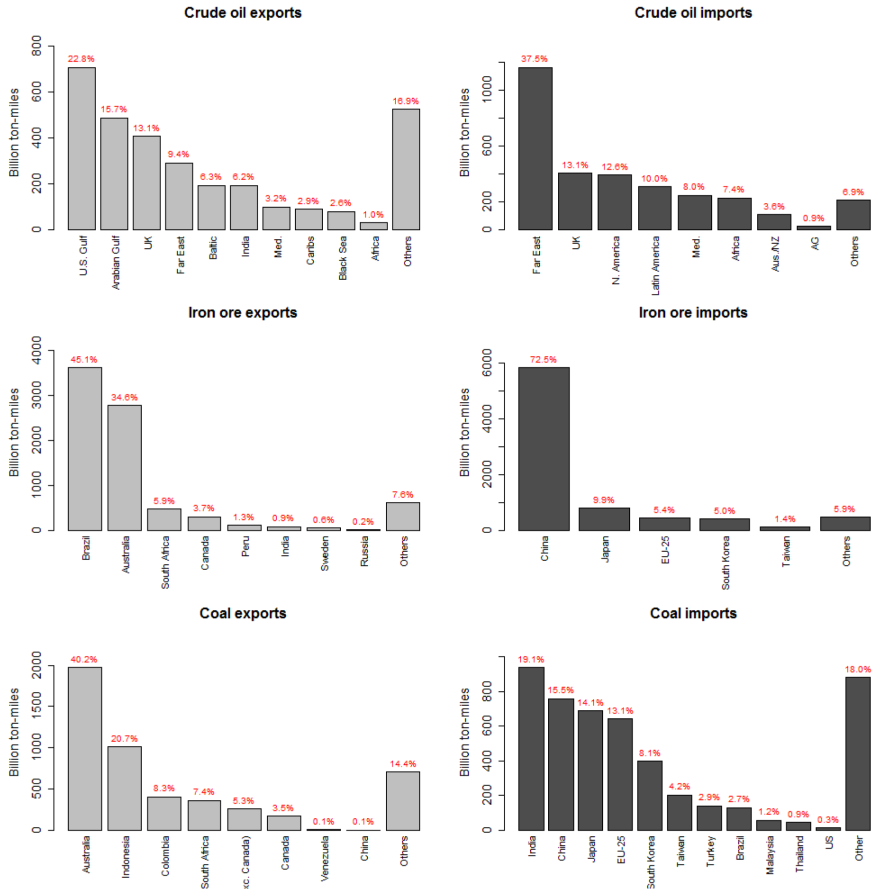

Figure 2.2: Largest importers and exporters of the most important bulk cargoes. Source Clarkson Research Services.

to transport a series of cargo for a fixed price per ton. As with the voyage contract, the shipowner pays all costs. For instance, a contract of affreightment can include delivering ten shiploads of iron ore from Brazil to China, one each month over the next ten months. The third type of contract is the *time charter* contract, where the charterer pays an agreed day-rate over a certain period of time. In this case the charterer gets operational control of the vessel, and must also pay the *voyage costs*, which includes port and bunker costs. The fourth freight agreement is the *bare boat charter*. Here the shipping company provides the charterer with a ship which the charterer gets of full operational control of. Such type of contracts generally stretches over several years.

There are three common ways of measuring freight rates. The *voyage rate* is measured in dollars per ton for a specific route. This measure is generally used for dry bulk trades, for instance for transporting a ton of coal from Australia to Japan. The rate includes voyage costs, and hence the freight rate for a specific route is not directly comparable to rates of other trading routes. The second measure for freight rates is the *time charter rate*, generally measured in dollars per day. This rate is easier to compare for different trading routes because it does not include voyage expenses. However, time charter agreements are fixed for a specific time period, and hence comparing the rates for contracts of different length could yield inadequate and misleading numbers. A third way of measuring freight rates is the *Worldscale index*, commonly used by the tanker industry. The index provides a freight rate that is given in terms of a reference to a standard vessel. Like the voyage rate for dry bulk transport, the Worldscale index includes costs associated with bunker, transit and port costs, and is thus not directly comparable for vessels operating on different routes.

Although voyage rates and the Worldscale index for different trading routes are not directly comparable, there are several indices which are using them in their calculations. For instance, the Clarkson Research Services Limited's[7] Clarksea Index, Clarksons Average Bulker Index and Clarksons Average Tanker Index. The best known freight indices are the the ones provided daily by the London-based Baltic Exchange. Among others, these include the Baltic Dry Index (BDF)[8] and Baltic Dirty Tanker Index (BDTI)[9]. Baltic Exchange constructs the indices from latest fixing prices which are collected from a manifold of brokers. The Baltic indices have become popular for several reasons. They are commonly used to replicate global economic growth as they provide information about global demand and supply for seaborne trade [Bakshi et al., 2011]. Also, the indices are used to settle freight derivatives contracts such as freight futures and options.

We will analyse and forecast the BDI and BDTI as dry bulk and tanker spot rate indices, respectively, throughout this paper. Each include rates for the four largest dry bulk and tanker vessels categories. Hence they capture the dynamics and development of these markets completely. Also, they are calculated exclusively as time charter equivalents from voyage fixings, which reflect the present spot market environment at any point of time. Finally, the indices are credible as they are used for settlement of derivatives. Figure 2.3 shows the development of BDI and BDTI in the period from 2000 to 2017.

In this section we have provided a brief overview of the three segments of shipping. We next presented the four markets in which shipowners can trade and participate, and furthermore how these markets determine the supply dynamics of shipping. Furthermore we analysed the dry bulk and tanker fleet, and dug deeper into demand characteristics of seaborne bulk transport. We completed our shipping market presentation by discussing three different ways of measuring freight rates along with challenges that arise when comparing freight rates for different trade routes. Finally, we discussed the Baltic Dry Index and Baltic Dirty Tanker Index, which will be analysed as freight rates throughout this paper.

[7]Clarkson Research Service Limited publishes daily freight rate indices measured in dollars per day, calculated by subtracting estimated voyage costs from Worldscale and voyage freight rates. Voyage cost estimates are based on trading routes and vessel characteristics. The indices are weighted according to the number of vessels in the fleet.

[8]The BDI is calculated using time charter equivalent rate fixtures (voyage rate minus voyage costs) for Capesize, Panamax, Supramax and Handysize vessels on 22 major trading routes. The index gives each of the categories equal weights (25%).

[9]The BDTI is calculated using time charter equivalent rate fixtures for VLCC, Suezmax, Aframax und Panamax vessels transporting unrefined petroleum products 17 major trading routes for crude oil taken from the Baltic International Tanker Routes.

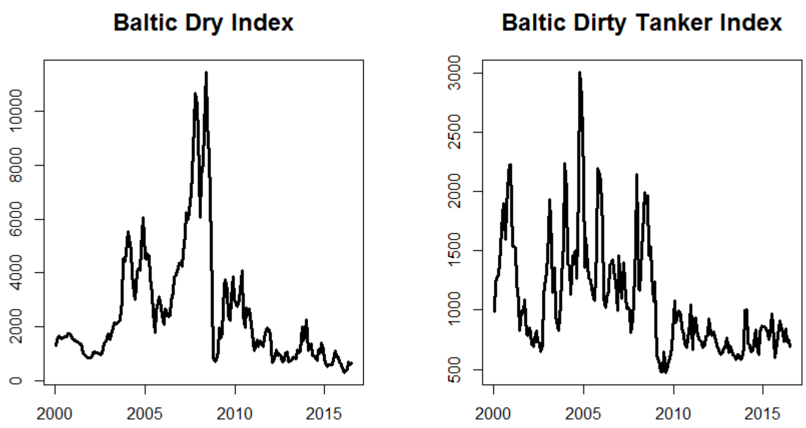

Figure 2.3: The Baltic freight indices to be analysed throughout this paper. BDI (left) is utilised as a proxy for dry bulk freight rates and BDTI (right) corresponds to tanker rates. Source: Clarkson Research Services Limited.

3 Freight rate determinants

This section presents modeling factors that researchers have found to be related to or able to indicate freight rates. Ultimately, our goal is to reveal key determinants that may influence freight rate dynamics and hence can be used in constructing our freight rate model. We first provide a recap of academic discussions and conclusions on the relationship between bulk demand and freight rates. We next swing the pendulum towards the supply side, and discuss concluded relationships and findings. Finally we shed light on financial indicators that could explain freight rate behaviour.

3.1. Demand determinants

Demand for sea transport is considered to have close links to global economic growth. Researchers uniformly agree that global gross domestic product (GDP) and trade demand have a strong and positive relationships [Lun et al., 2006, Lun and Cheng, 2010]. Stopford [2009] points out GDP as the single most important influence on ship demand. He argues that it should be a close relationship between the two since the world economy generates demand for sea transport through commodity imports and manufacturing. Furthermore, he emphasises that global economic growth is fluctuating, usually in periodic movements referred to as business cycles, and hence demand for seaborne trade should follow patterns of similar fashion. Among the causes of business cycles are the interplay between consumption and investment, time-lags between economic decisions and implementation, and build-up of inventories. Stopford also argues that development in seaborne commodity routes and trades are principal indicators for future transport demand. He further presents four other variables that influence shipping demand: transport costs, random shocks, average haul and seaborne commodity trades. On the contrary to former research, Stopford mentions transport costs to influence demand, while other researchers have found demand to be exogenous and unrelated to freight rates (see, for instance, Beenstock and Vergottis [1993], Randers and Göluke [2007]). Moreover, Stopford argues that raw materials only will be transported if the cost levels are acceptable. This view is supported by Coyle et al. [2000], which argue that reduced transport costs stimulate more demand for sea transport through its impact on consumers' purchasing decisions. Randers and Göluke [2007] argue that freight rates never exceed 5% of the cargo value, hence demand should be approximately unaffected by freight rates. Beenstock [1985] used a GDP to model expected second-hand vessel prices and found a significant relationship between them. Lin and Sim [2013] also found a clear relationship between global trade, specifically the BDI, and the gross domestic product of less developed countries. Hyung-geun [2011] analysed the relationship between the BDI and Chinese economic trend, and found that Chinese economic fluctuation does affect the dry bulk market, measured by the BDI. Beenstock and Vergottis [1989b] investigated the tanker market by modeling freight rates, lay-up, shipbuilding activity and new and second-hand tanker prices. It was concluded that an expansion of world trade has a positive impact on freight rates. Anyanwu [2013] examined the tanker segment and, similarly to many of the above-mentioned contributors, found that there is a positive relationship between seaborne trade and freight rate. Stopford [2009] further points out that tanker demand is a derived demand, in the sense that it is derived from the international trade in oil and oil products, which in turn depends on world economic activity through imports and consumption of energy commodities.

3.2. Supply determinants

The much taught dynamic supply theory of shipping was first introduced by Tinbergen [1934] and Koopmans [1939]. This theory describes the cyclical movements of shipping, and involves a dynamic relationship between freight rates and the supply of sea transport. Several models are built upon this theory (see, for instance, Zannetos [1966], Hampton [1991], Beenstock and Vergottis [1993], Engelen et al. [2006], Randers and Göluke [2007]). Beenstock and Vergottis [1993] published a series of studies of the tanker and dry bulk market. Using theory from the model of expected prices for second-hand vessels presented by Beenstock [1985], they presented two separate econometric models for the tanker market [Beenstock and Vergottis, 1989b] and dry bulk market [Beenstock and Vergottis, 1989a]. Demand, denoted in ton-miles, was modelled as exogenous, while supply was modelled as a function of freight rates, fleet size, operational ship- expenses and costs of lay-up. They found that freight rates were determined by the balance between demand and the active fleet size. Later on, Randers and Göluke [2007] argued that it is possible to explain much of the history of shipping since 1950 using an endogenous supply model. Their theory suggests that shipping cycles are determined by the interaction of two loops, one 20-year wave, as a result of shipowners ordering too many new vessels when conditions are good, and one four-year cycle, where shipowners seek to optimise their current fleet. Similar to Beenstock and Vergottis [1993], Engelen et al. [2006], Randers and Göluke treated demand as exogenous, while the supply side included freight rates, ordering of new vessels, average building time, average life of vessels, fleet productivity, change in fleet utilisation and scrap rate. Furthermore, Randers and Göluke explain that supply, denoted by ton-miles, also is flexible and dependent on the way shipowners choose to operate their vessels. When there is not enough capacity to meet demand, shipowners can improve profitability by increasing the speed of their fleet, encouraging fuller cargoes, postponed maintenance, shortened port times, among others. Hence the fleet supply measured in ton-miles is not constant, but rather flexible to market conditions. Fleet productivity is also explained by Stopford [2009]. He argues that the fleet productivity depends upon four main factors, namely speed, port time, fleet utilisation and loaded days at sea. Furthermore, Stopford explains that the supply side of shipping is slow in adjusting to changes in demand, because it generally takes about one to three years from a vessel is ordered until delivery. In the long-run, he explains that supply of sea transport is determined by the size of the fleet, driven by scrapping and deliveries of new vessels. Tsolakis et al. [2003] provided a model of second-hand ship prices for different shipping segments and sizes, and found that time-charter rates and newbuild prices have the greatest effects. They also revealed that each segments react somewhat differently to changes in the indicator variables, and hence they recommended that ship-prices should be analysed at an disaggregated rather than aggregated level. Second-hand prices were also studied by Pruyn et al. [2011], which argued that a model for second-hand prices should consider newbuild prices, orderbook size, vessel earnings and fuel consumption. Alizadeh and Talley [2011] found that the supply of tanker shipping services depends on the size of the tanker fleet, tonnage available for trading, tanker shipbuilding activities, bunker fuel prices, the scrapping rate of the fleet, and the productivity of the tanker fleet at any point in time.

3.3. Financial and other non-fundamental determinants

Our study of literature reveals several relationships between freight rates and variables that are not directly considered as supply or demand factors. We define financial determinants as prices and measures that are determined by or traded in financial markets. Non-fundamental variables are events such as wars taking place, political matters or regulations, which eventually may affect supply or demand. Hence, we note that there may not be a clear line between these and the indicators listed in the two subsections above. Stopford [2009] highlights the importance of behavioural aspects and patterns of other shipping market participants such as banks and regulators. These could act in a way that influences shipowners decisions and affect demand characteristics. Thus financial and non-fundamental shipping variables could impact the shipping market through changes in the behaviour of shipping market participants. Hence factors that do not directly explain supply and demand of sea transport could have behavioural impact, and furthermore have immediate impact on *expectations* about future supply and demand, and furthermore affect freight rates. Also, financial variables could indirectly act as fundamental ones, since fundamental information may be traceable from financial factors [Westgaard et al., 2017]. Bakshi et al. [2011] revealed a link between real and financial markets by showing that the BDI could predict global economic growth, commodity indices and stock markets. The latter relationship was also examined by Alizadeh and Talley [2011], who found that the BDI was a leading indicator for stock market returns. Moreover, they showed that the model could be applied across international stock indices. Built upon the work of Alizadeh and Talley, Apergis and Payne [2013] examined BDI's predictive ability for financial asset markets and industrial production for a panel of G7 countries, using daily data from 1952-2012. They found that the BDI had predictive significance and performed better than the MSCI index and oil prices as predictors for both short and long term developments in industrial production and stock prices. They also found strong unidirectional causality from the BDI to financial asset prices. While shipping freight rates have been well studied and found to have strong predictive significance for financial markets, some newer studies have also revealed that financial markets could lead shipping rates. Shipping derivatives were first introduced in 1985 as futures contracts. In 1992, freight forward agreements (FFA) were introduced and allowed shipping practitioners and speculators to hedge or bet on future levels of the Baltic Freight Index. Since then, the derivatives market has grown exponentially, and become an important market where shipowners can hedge their risk exposure to the highly volatile cyclical movements of the industry. See Kavussanos and Visvikis [2006] for an excellent survey on derivatives contracts in shipping along with their financial applications. In addition, Kavussanos and Visvikis [2006] call for awareness related to currency-risk; usually, the income of shipowners is in US dollars, while payments tend to be in the local currency of the shipyard, such as Japanese yen. Kavussanos and Alizadeh-M [2001] investigated the lead-lag relationship in daily returns and volatilities between spot rates and FFAs, and found a bidirectional lead-lag relationship between them. Moreover, they found that forward contracts discover market information faster than spot rates, which they suggested could be attributed to the higher transaction cost in the spot market. A similar conclusion was reached by Batchelor et al. [2007], applying several econometric forecasting models and found that forward rates do help to forecast spot rates. Furthermore, all models performed better than a random walk out-of-sample. Time-charter rates could be regarded as another form of FFA, because they are related to spot rate expectations over a certain period. Fan et al. [2013] also sought to forecast spot rates, more specifically attempted to predict the BDTI using a wavelet neutral networks technique. The variables included in their model were the oil price, CBOE SPX Volatility Index, and SP Global 1200 Index, and their model outperformed an ARIMA-model out-of-sample. Stopford [2009] also comments on the relationship between tanker rates and the oil price. He argues that demand for oil tankers is not unaffected by the oil price level - rather that an increase in oil price tends to alter the global energy mix, where expensive oil may become substituted by relatively cheap coal, thereby reducing demand in the tanker market. Regarding credit markets' relationship with the shipping industry, limited research have been carried out. Grammenos and Arkoulis [2003] investigated the variability of bond price offerings in the shipping industry, and reached the conclusion that credit rating was the main determinant, while financial leverage and percentage of fleet laid up also had impact on bond spreads.

In this section we have provided a recap of concluded relationships between freight rates and supply, demand, financial and other non-fundamental variables found in literature. As such, we have revealed potential indicators that may influence our tanker and bulk rate models.

4 Data

We apply a general-to-specific approach on the principles presented by Campos et al. [2005], where a comprehensive, yet well-justified set of data should be collected. For the sake of completeness, we have chosen to include all potential and available freight rate determinants reflecting conclusions and discussions in our study of literature presented in the previous section. In total we have collected 44 time series to be analysed in the dry bulk analysis, and 37 time series to be analysed in the tanker analysis. In this section we present the rationale behind the inclusion of each variable. Table 4.1 provides an overview of the variables that are included in dry bulk and tanker models along with their hypothesised directional impact on the respective freight rate. Furthermore, we illustrate and comment on the variables' descriptive statistics.

4.1. Data selection and argumentation

We use the Baltic Dry Index (BDI) and Baltic Dirty Tanker Index (BDTI) as time series to reflect dry bulk and tanker freight rates, respectively, provided by The Baltic Exchange. Despite the frequent argumentation of demand being exogenous with respect to freight rates (see, for instance, Beenstock and Vergottis [1993], Randers and Göluke [2007]), it is argued by Stopford [2009], Coyle et al. [2000] that the freight cost could trigger changes in demand for sea transport, as cheaper freight cost could act as an incentive for increasing shipment of goods. We thus include lagged BDI and BDTI in their respective models. Even though lagged shipping rates could have a negative relationship with themselves because cheaper transport cost could act as an incentive for more transport, we also argue that the relationship could be positive. This is due to the fact that rates tend to have momentum and move in cycles (see section 2). Though Beenstock and Vergottis [1993] found limited spillover effects between shipping segments, Randers and Göluke [2007], Stopford [2009], among others, argue that shipping rates are co-integrated, and that it is meaningful to talk about the shipping market as one entity. We hence include lagged time series of the BDTI and BDI for both the tanker and dry bulk models in order to capture potential lagged cross effects between the markets.

4.1.1. Demand variables

Among demand variables, we first include a geometric mean of global real GDP, provided monthly by EIA, in order to reflect global economic growth. Next, the year-on-year change in OECD industrial production is collected, which refers to the industrial output of manufacturing, mining and public utilities. Motivated by Apergis and Payne [2013], the rationale for including this index is that it is sensitive to changes in consumer demand and interest rates, and hence could provide a leading signal of economic growth. We furthermore use global and Middle East oil production numbers in the tanker models, supposed to reflect explicit demand for crude oil tanker transport. As revealed by Figure 2.2, some countries have particularly strong demand for certain commodities. For instance, China accounts for 73.5% of global iron ore imports in terms of ton-miles. Similarly, India and China combined account for 34.5% of global coal imports. Since iron ore and coal are the main commodities constituting the dry bulk segment, these two countries are the key drivers of the aggregate dry bulk demand. Thus, we argue that it would be reasonable to include Chinese and Indian purchasing managers index (PMI), indicating the economic status of the manufacturing sector, in the dry bulk model. China and

India are also the main importers of seaborne crude oil with 7.5 and 4.6 million barrels per day, respectively [Arctic, 2017]. Alongside the US, these three nations account for 42% of seaborne crude imports [Arctic, 2017], and hence industrial production numbers for these three countries are included in the tanker model. Furthermore, as an attempt to capture indicators for changes in seaborne commodity trades in terms of ton-miles, which is shown in Table 2.2, we have collected import and export data for the most important commodities. For dry bulk, we have included Chinese iron ore and coal imports, Indian coal imports, Brazilian iron ore exports, U.S. grain exports and Australian iron ore and coal exports. Also, we include Chinese steel production, which may reflect the trend in demand for iron ore and coal. Similarly, for tanker rates we have included Indian and Chinese crude imports, and U.S. and OPEC crude exports, as these are the most important crude players. We expect positive relationships for freight rates for both industrial production changes and commodity trades. Stopford [2009] described the interplay between consumption and investments as *multiplier and accelerator*. We include U.S. and Chinese money supplies and consumer price indices (CPI) as such indicators (using the two largest economies in the world as proxies for global multiplier and accelerator effects). The rationale is that CPI development will reflect demand for goods in a given economy - if demand for goods is strong, increased price levels are justified. Also, money supply levels and low interest rates are tied together, and low interest rates will trigger investment activity. Thus, we expect a change in these variables to be a sign of a stimulated economy, which in turn will have a positive impact on freight rates.

4.1.2. Supply variables

The supply variables we have selected for the tanker and dry bulk models are based on data from their respective fleets. Table 4.1 provides an overview of the variables. We first present four supply variables motivated by the dynamic supply theory presented in section 2.1 and discussed in section 3.2. First, we include the fleet size, measured in dwt, on a global tanker and dry bulk basis, gathered from Clarkson Research Limited. Supply and demand equilibrium mechanisms suggest that a positive change in the fleet (i.e. increased supply) is related to a negative change in freight rates. Orderbook in percentage of total aggregated fleet is the next variable we include. That is, the amount of tonnage that has been ordered, but not yet delivered. The orderbook reflects expectations about the amount of tonnage to be delivered in the future, and hence we expect the orderbook to be an inversely leading indicator for freight rates. We next include scrapping, where we expect increased scrapping to chisel away supply and thus increase freight rates. Consequently, a positive relationship is hypothesised. We further look at the tonnage amount that is delivered, which will have the opposite effect on supply, and hence we expect freight rates to decrease when deliveries increase. The next set of variables we include are ship rates from the second-hand market. The fact that a ship can change owner relatively swift in the second-hand market, while it takes two-three years to build a new ship, makes it possible to draw some inferences about the market expectations by inspecting the second-hand price relative to the newbuild price. For instance, when freight rates are soaring, second-hand prices may actually be higher than newbuild-prices, implying that the market expects freight rates to remain high enough the next couple of years to justify the relatively shorter life expectancy of a second-hand ship. Though second-hand vessels are shown to be tightly correlated with freight rates (see, for instance, Beenstock [1985], Stopford [2009]), we include second-hand prices of five year old[10] VLCC and Aframax tankers and five year old Capesize and Panamax bulk carriers.[11] The rationale for including this variable is that shipowners' expectations about the future may be reflected in the second-hand values in larger extent than in freight rates, and as such second-hand vessels could be leading freight rates. Hence a positive relationship is hypothesised. A variable motivated by Alizadeh and Talley [2011] is vessel fuel prices which, given all else equal, will contribute to cheaper seaborne transportation, but at the same time will follow the fluctuations of the oil price to a great extent. Thus, the variable could possibly have both positive and negative impact on freight rates.

4.1.3. Non-fundamental variables

The first non-fundamental variables we include are one-year time charter rates for dry bulk and tanker ships. These reflect future spot rate expectations, and hence we expect a positive relationship to emerge. Further, we include the oil price for two reasons. First, it may affect the supply side indirectly by being the main determinant of fuel prices. Also, the oil price is correlated positively with global economic activity [Hamilton, 2005, 2008], and hence it could be a proxy for demand. In the case of dry bulk rates, we expect this effect to offset the increase in fuel prices, and accordingly a positive correlation is hypothesised. However, the tanker fleet is differently exposed to the oil price, since oil is the commodity it transports. As mentioned in section 3.3, Stopford [2009] comments specifically on the relationship between tanker rates and the oil price. He argues that an increase in the oil price tends to alter the global energy mix, where coal substitutes oil, thereby reducing demand in the tanker market. Therefore, we argue that tanker rates could be both positively and negatively affected by an oil price increase. Next, a U.S. dollar index is collected. The rationale is that shipowners have their revenues in dollar and could pay costs in other currencies (see Tvedt [2003], Glen and Martin [2005] for discussions on the relationship between shipowners and exchange rates). Hence a strengthening of the dollar could have a positive impact of freight rates as it would improve shipowners financial performance. However, it could also lead to increased newbuild contracting, raising supply expectations, and hence have negative impact on rates. Furthermore, we include the U.S. Dollar to Japanese Yen exchange rate. This is motivated by Tvedt [2003], who argues that the Japanese economy is a major driver of shipping markets, particularly reflecting construction activity. Our hypothesis is that these exchange rates could have both a positive and negative relationship with freight rates. Next we include the contango level for Brent crude oil prices, which is the slope of the oil futures curve. The rationale is that when the curve is steep upwards, traders would buy oil, sell forward derivatives and use oil tankers for storage, hence tanker demand would increase. Consequently, we expect freight rates to increase following an increase in the futures spread level. We have used the slope between the six month futures contract and the one-month front contract, as these are most liquid. We furthermore use the LIBOR interest rate, which we expect will have a negative relationship with freight rates. Lower interest rates could boost investment and economic activity and thus increase demand. Next, we include a high yield bond spread variable in order to

[10] The market for five year old ships is most liquid [Arctic, 2017].

[11] VLCC and Aframax account for the two largest tanker segments measured in fleet size, while Capesize and Panamax account for the two largest bulk segments in fleet size, see 2.1.

reflect investor's willingness to invest money. This index could reflect future economic activity and we expect a tightening of the spread, meaning a less risky investor view, to reflect increased economic activity and seaborne trade demand. Hence we expect a negative relationship between high yield bond spreads and freight rates, which was found by Westgaard et al. [2017] in relation to oil prices. We also include the VIX, a volatility index measuring the implied volatility of S&P500 index options. We expect a negative relationship between freight rates and the VIX, as we expect trading activity to decrease when the "fear" level among investors increases. We furthermore include various commodity prices for the dry bulk model as they are the main cargo of the dry bulk segment, and moreover, it was shown by Bakshi et al. [2011] that commodity returns and the BDI were related to each other. We collect prices for iron ore, coal and grain, which are the three major bulk commodities. As disruptive supply shocks are rare for these commodities (on a global basis), we expect price changes to mostly reflect changes in demand, thus a positive correlation with freight rates is expected. Finally we include share prices in our analysis. First, the S&P500 Index is used in order to capture investors expectations about the future of the largest U.S. listed companies. Similarly, we include the MSCI World Index and MSCI Emerging Markets Index in order to capture the global stock market performance and the development of companies in emerging markets countries. The latter is motivated by Bakshi et al. [2011], who showed a clear relationship between the BDI and economic growth in emerging markets. We hypothesise a positive relationship as rising stock prices reflect optimism and increased economic activity. Furthermore, as we do not find equity indices of pure dry bulk or tanker companies, we construct and introduce two new indices for the largest tanker and dry bulk companies. These indices are constructed using dividend and stock split adjusted share prices for the largest stock exchange listed fleet owners, ranked by fleet tonnage, obtained from Clarkson Research Limited. See Appendix A.3 for an overview of the companies included in the indices and how the indices are calculated. Our intuition is that share prices could serve as strong proxies or price signals as they reflect future shipping market conditions at any point of time. This intuition was confirmed empirically by Westgaard et al. [2017] when examining oil prices, who showed that stocks, specifically the PHLX Oil Service Sector (OSX) index, is a leading indicator for crude oil prices.

4.2. Descriptive statistics

The data set is listed in Table 4.1 and consists of time series of monthly logarithmic changes and absolute values. The sample period is December 2000 to June 2016, corresponding to a sample size of 199 observations. Each data point is obtained at the latest available date of the given month. The arrows in Table 4.1 represent the expected oil price change, as a response to the change in the respective variable as outlined in the previous subsections, based on relationships discussed in section 3. An overview of the descriptive statistics and Jarque-Bera test for each time series is listed in Table 4.2.

Table 4.1: List of variables included in the models

	Variable	Description	Hypothesis	
			Dry bulk	Tanker
	BDI	Baltic Dry Bulk Index	↕	↕
	BDTI	Baltic Dirty Tanker Index	↕	↕
Demand	GDP_W	Weighted geometric mean of world GDP	↑	↑
	IP_OECD	Industrial Production, OECD	↑	↑
	IP_Ch	Industrial Production, China	↑	↑
	IP_I	Industrial Production, India	↑	↑
	IP_US	Industrial Production, U.S.	↑	↑
	Oil_P_G	Oil production, Global	↑	↑
	Oil_P_ME	Oil production, Middle East	↑	↑
	IO_Ch_Imp	Iron ore imports, China	↑	-
	IO_B_Exp	Iron ore exports, Brazil	↑	-
	IO_A_Exp	Iron ore exports, Australia	↑	-
	C_EU_Imp	Coal Imports, EU-25	↑	-
	C_J_Imp	Coal imports, Japan	↑	-
	C_A_Exp	Coal exports, Australia	↑	-
	G_US_Exp	Grain exports, U.S.	↑	-
	S_Ch_Prod	Steel production, China	↑	-
	O_Ch_Imp	Crude oil imports, China	-	↑
	O_US_Exp	Crude oil exports, U.S.	-	↑
	O_AG_Exp	Crude oil exports, Arabian Gulf	-	↑
	S_US_Imp	Steel imports, US	↑	-
	MS_US	Money supply, U.S.	↑	↑
	MS_Ch	Money supply, China	↑	↑
	CPI_US	Consumer price index, U.S.	↑	↑
	CPI_Ch	Consumer price index, China	↑	↑
Supply	Fleet	Global fleet in dwt	↓	↓
	Order	Orderbook dwt in percent of total fleet	↓	↓
	Scrap	Demolition, dwt	↑	↑
	Del	Deliveries, dwt	↓	↓
	New	Newbuild price index	↑	↑
	Sec	5 year old second-hand price index	↑	↑
	Fuel	Vessel bunker fuel price	↕	↕
Financial	TC_Bulk	1 year time-charter rates, Capesize	↑	-
	TC_Tank	1 year time-charter rates, VLCC	-	↑
	Oil	Brent crude oil front month contract	↑	↕
	FX_USD	Dollar exchange rate index	↕	↕
	FX_USD_JPN	Dollar-Yen exchange rate	↕	↕
	Cont_Oil	Oil price contango, Brent 6m relative to 1m	-	↑
	LIBOR	LIBOR USD 3-month	↓	↓
	HY_Spread	BoA Merrill Lynch HY spread index	↓	↓
	VIX	Volatility Index	↓	↓
	P_IO	Iron ore price, Brazil	↑	-
	P_Coal	Coal price, Australia	↑	-
	P_Wheat	Wheat price, U.S.	↑	-
	P_Metals	Metals price index (copper, aluminum, iron Ore, tin, nickel, zinc, Lead, and Uranium)	↑	↑
	P_Gold	Gold price	↓	↓
	DBulk_Index	Dry bulk equity index	↑	-
	Tank_Index	Tanker equity index	-	↑
	SP500	S&P500 Equity Index	↑	↑
	MSCI_W	MSCI World Index Equity Index	↑	↑
	MSCI_EM	MSCI Emerging Markets Index	↑	↑

The table shows the variables that are included in the tanker and dry bulk models. The two rightmost columns explain our hypothesis of the impact a positive change in the respective variable is expected to have on dry bulk and tanker freight rates, respectively. A variable is illustrated with the sign "-" if it is not included in the respective model. See Table A.1 for sources.

Table 4.2: Descriptive statistics

Time series	Mean	Std.err.	Median	Std.dev.	Var	Kurtosis	Skew	Range	Min	Max	Count	J-B	p^{J-B}
ΔBDI	-0.003	0.017	0.013	0.237	0.056	6.104	-1.229	2.001	-1.330	0.671	198	358.9	0.000****
ΔBDTI	-0.002	0.013	-0.005	0.183	0.034	1.330	-0.321	1.174	-0.710	0.464	198	18.3	0.004***
ΔGDP_W	0.003	0.000	0.003	0.002	0.000	10.495	-2.392	0.015	-0.008	0.007	198	1142.8	0.000****
ΔIP_OECD	0.000	0.001	-0.001	0.012	0.000	3.429	0.326	0.098	-0.045	0.053	198	105.9	0.000****
ΔIP_Ch	0.000	0.001	0.000	0.020	0.000	11.329	0.213	0.208	-0.109	0.099	198	1083.8	0.000****
ΔIP_I	-0.001	0.002	0.000	0.033	0.001	1.732	-0.278	0.240	-0.130	0.110	198	27.3	0.003***
ΔIP_US	0.000	0.001	-0.001	0.010	0.000	4.637	0.517	0.091	-0.044	0.047	198	187.4	0.000****
ΔOil_P_G	0.001	0.001	0.001	0.009	0.000	1.656	-0.409	0.058	-0.033	0.025	198	29.0	0.002***
ΔOil_P_ME	0.002	0.001	0.001	0.020	0.000	4.230	-0.126	0.167	-0.082	0.085	198	160.5	0.000***
ΔIO_Ch_Imp	0.013	0.011	0.010	0.160	0.026	-0.092	0.017	0.929	-0.427	0.502	198	0.1	0.936
ΔIO_B_Exp	0.004	0.019	0.012	0.266	0.071	9.914	1.069	2.715	-0.971	1.743	198	856.9	0.000****
ΔIO_A_Exp	0.009	0.007	0.011	0.104	0.011	1.218	-0.075	0.653	-0.325	0.328	198	14.1	0.011**
ΔC_EU_Imp	-0.001	0.007	-0.004	0.093	0.009	0.380	-0.123	0.512	-0.257	0.255	198	1.6	0.404**
ΔC_J_Imp	0.001	0.007	0.003	0.103	0.011	0.942	-0.189	0.656	-0.357	0.299	198	9.5	0.022**
ΔC_A_Exp	0.004	0.006	0.013	0.091	0.008	0.550	-0.384	0.541	-0.300	0.241	198	7.2	0.033**
ΔG_US_Exp	0.001	0.009	0.003	0.133	0.018	0.569	-0.193	0.851	-0.457	0.395	198	4.1	0.099*
ΔS_CH_Prod	0.009	0.004	0.008	0.051	0.003	0.628	0.440	0.290	-0.102	0.188	198	10.5	0.019**
ΔO_Ch_Imp	0.010	0.013	0.012	0.180	0.032	0.107	-0.054	0.944	-0.467	0.477	198	0.2	0.911
ΔO_US_Exp	0.008	0.011	0.014	0.156	0.024	2.666	-0.600	1.078	-0.652	0.425	198	79.2	0.000****
ΔO_AG_Exp	0.002	0.002	0.001	0.032	0.001	4.630	-0.670	0.282	-0.171	0.111	198	207.4	0.000****
ΔMS_US	0.005	0.001	0.005	0.010	0.000	7.260	1.494	0.090	-0.034	0.056	198	512.4	0.000****
ΔMS_Ch	0.000	0.002	0.004	0.035	0.001	0.347	-0.152	0.205	-0.106	0.099	198	2.0	0.328
ΔCPI_US	0.002	0.000	0.002	0.003	0.000	9.130	-1.343	0.032	-0.018	0.014	198	752.7	0.000****
ΔCPI_Ch	0.000	0.001	0.002	0.007	0.000	2.853	-1.248	0.046	-0.033	0.014	198	118.3	0.000****
ΔFleet_Bulk	0.000	0.000	0.000	0.004	0.000	1.530	-0.193	0.024	-0.015	0.009	198	23.1	0.003***
ΔFleet_Tank	0.000	0.000	0.000	0.004	0.000	2.763	-0.324	0.034	-0.021	0.012	198	72.2	0.000****
ΔOrder_Bulk	0.001	0.003	-0.005	0.040	0.002	0.879	0.630	0.241	-0.099	0.142	198	19.7	0.005***
ΔOrder_Tank	0.000	0.004	-0.004	0.056	0.003	11.441	2.117	0.490	-0.119	0.371	198	1260.5	0.000****
ΔScrap_Bulk	0.004	0.041	-0.017	0.570	0.325	3.977	0.255	4.646	-2.376	2.270	198	137.6	0.000****
ΔScrap_Tank	-0.015	0.053	-0.022	0.747	0.558	7.337	0.358	7.057	-2.989	4.069	198	489.8	0.000****
ΔDel_Bulk	0.001	0.265	0.095	3.729	13.908	3.376	-0.511	26.831	-14.856	11.976	198	102.9	0.000****
ΔDel_Tank	0.008	0.093	-0.067	1.311	1.718	0.566	0.151	7.550	-3.627	3.923	198	4.0	0.117
ΔNew_Bulk	0.010	0.009	0.000	0.120	0.014	184.753	13.371	1.738	-0.081	1.657	198	290257.3	0.000****
ΔNew_Tank	0.000	0.001	0.000	0.017	0.000	4.217	-0.105	0.158	-0.080	0.078	198	156.2	0.000****
ΔSec_Bulk	-0.002	0.005	0.003	0.067	0.004	46.228	-4.979	0.826	-0.653	0.173	198	18755.2	0.000****
ΔSec_Tank	0.000	0.003	0.000	0.041	0.002	3.318	-0.562	0.309	-0.199	0.111	198	101.2	0.000****
ΔFuel	0.003	0.008	0.007	0.116	0.013	4.476	-1.053	0.978	-0.595	0.383	198	202.3	0.000****
ΔTC_Tank	0.001	0.008	0.000	0.108	0.012	2.793	0.586	0.767	-0.325	0.442	198	75.6	0.001***
ΔTC_Bulk	-0.002	0.014	0.000	0.194	0.037	7.114	-1.117	1.735	-1.073	0.662	198	461.6	0.000****
ΔOil	0.003	0.007	0.009	0.093	0.009	1.606	-0.650	0.662	-0.407	0.254	198	34.9	0.002***
ΔFX_USD	0.000	0.002	-0.001	0.024	0.001	0.419	0.098	0.139	-0.064	0.075	198	2.1	0.307**
ΔFX_USD_JPN	0.000	0.003	0.000	0.040	0.002	0.522	0.144	0.231	-0.109	0.122	198	3.3	0.162
ΔCont_Oil	0.001	0.002	0.002	0.028	0.001	3.105	0.572	0.204	-0.077	0.127	198	96.6	0.000****
ΔLIBOR	0.000	0.020	0.003	0.282	0.080	27.982	-1.880	3.804	-2.268	1.536	198	6644.0	0.000****
ΔHY_Spread	-0.020	0.052	-0.080	0.739	0.545	13.494	1.539	8.240	-3.190	5.050	198	1602.2	0.000****
ΔVIX	-0.002	0.014	-0.016	0.197	0.039	1.586	0.557	1.339	-0.486	0.853	198	30.9	0.002***
ΔP_IO	0.003	0.006	0.000	0.091	0.008	9.329	0.179	0.994	-0.454	0.539	198	727.8	0.000****
ΔP_Coal	0.004	0.005	0.000	0.068	0.005	6.349	0.160	0.692	-0.329	0.364	198	335.0	0.000****
ΔP_Wheat	0.003	0.005	-0.004	0.067	0.005	3.105	0.630	0.552	-0.230	0.322	198	94.5	0.000****
ΔP_Metals	0.003	0.003	0.004	0.049	0.002	1.804	-0.331	0.363	-0.222	0.142	198	31.3	0.000****
ΔP_Gold	0.008	0.004	0.008	0.050	0.003	0.577	-0.238	0.307	-0.185	0.122	198	4.7	0.088**
ΔDBulk_Index	0.006	0.008	0.008	0.119	0.014	8.729	0.856	1.204	-0.443	0.761	198	656.8	0.000****
ΔTank_Index	0.007	0.007	0.014	0.096	0.009	1.429	-0.582	0.630	-0.367	0.263	198	27.8	0.002***
ΔSP500	0.002	0.003	0.008	0.044	0.002	1.419	-0.701	0.288	-0.186	0.102	198	35.4	0.001***
ΔMSCI_W	0.001	0.003	0.006	0.046	0.002	2.039	-0.836	0.315	-0.211	0.103	198	61.8	0.001****
ΔMSCI_EM	0.003	0.005	0.006	0.067	0.004	2.378	-0.789	0.476	-0.322	0.154	198	67.2	0.000

Significance level: **** 0.001, *** 0.01, ** 0.05, * 0.1.
Descriptive statistics for all time-series in the data set for the period January 2000 - June 2016. Data is obtained monthly. See sources in Table A.1.

5 Methodology

In this section we formulate six dry bulk models and six tanker models - all pure prediction models and hence limited to lags of the variables presented in Table 4.1. The section is organised as follows: First, a stationarity testing method is presented. Next, we discuss how we determine the number of lagged time series for each variable, and how seasonal models are constructed in order to take seasonality patterns into account. Furthermore, we formulate our models, line out the model selection procedure and present the termination criteria. Next, we present a qualitative and quantitative framework for investigating the OLS assumptions, before we finally present the model evaluation framework.

5.1. Testing for stationarity

The regression estimators are not valid if the regression is carried out on time series with non-stationary properties. Qualitatively, a time series will appear stationary if the data generating process seems to be independent of time, meaning that it displays no trend nor seasonality patterns in addition to having an approximately constant variance. More technically, a time series is defined as stationary if the following is true:

$$\mathbf{E}(y_t) = \mu \tag{5.1}$$

$$\mathbf{Var}(y_t) = \mathbf{E}[(y_t - \mu)(y_t - \mu)] = \sigma^2 \tag{5.2}$$

$$\mathbf{Cov}(y_{t_2}, y_{t_1}) = \mathbf{E}[(y_{t_2} - \mu)(y_{t_1} - \mu)] = \Omega_{t2-t1}, \qquad \forall t_2, t_1 \tag{5.3}$$

That is, constant mean, constant variance and constant autocovariance [Nason, 2013]. If a time series is non-stationary, it may heavily influence the behavior and properties of the regression analysis carried out. One serious issue is spurious regression, referring to when an explanatory variables and the response variable are totally unrelated, yet a regression analysis claim there is a significant relation. Typically, if a trend component is present in both the dependent and independent variable, the spurious regressor may display a very high R^2 value. Non-stationarity will also make the error distribution deviate from its asymptotic behavior in the stationary case, causing the t- and F- values to differ from their original distributions. A regression analysis may then misleadingly yield significant variables, i.e. type 1 errors. It is therefore of high importance to test all time series for stationarity.

The data displayed in Figure A.1 is obviously non-stationary, as the time series have strong directional trend components. Ideally, we would like each time series to be stationary and have a distribution close to the normal distribution, as strong departures from normality may lead to bias in the OLS estimators (less of an issue when the sample size is greater than 30 observations). The two transformations that are most commonly employed in order to obtain stationarity, is percentage change or logarithmic change. Most financial time series do not have normally distributed returns, (stock prices tend to increase over time), however returns tend to be log-normally distributed. We therefore initially apply the logarithmic change transformation on each time series. An Augmented Dickey Fuller- test (ADF) is then performed (the test procedure is outlined in Appendix C.2). If the ADF-test in unable to reject non-stationarity at a five percent significance level, the series is transformed by taking the absolute change.

5.2. Determining the number of lags

As summarised in Table 4.1, we have collected 44 dry bulk time series and 37 tanker time series. We aim at constructing pure prediction models, thus the set of potential explanatory variables are simply lagged time series regressed against the non-lagged BDI and BDTI time series. The total number of potential explanatory variables to search from is thus a multiple of the number of lags. Because each variable inevitably exhibit some degree of randomness, the chance of picking false significant variables will increase with the number of potential variables to choose from. With false significance, one refers to variables that happen to be significant by coincidence (Type 1 error). This is one of the reasons why there should always be a clear hypothesis behind every explanatory variable added to the pool of potential variables to choose from. In our case, we have 44 dry bulk and 37 tanker variables, each with v lags, potentially $44v$ and $37v$ variables in each model. Under the assumption of normally distributed residuals, we could expect $\frac{44}{20}v$ and $\frac{37}{20}v$ variables to show up as false significant for the dry bulk and tanker models, respectively, under the significance criterion of 5%. Additionally, most of the time series in our analysis have a higher frequency of extreme observations than the normal distribution function (see Table 4.2, Figure A.5 and Figure A.6), again increasing the risk of finding false significant variables.[12]

Even though one should be cautious about including too many lags, one cannot include too few either. A prediction model needs a reasonable amount of past information in order to capture lagged effects in the market. As a trade-off between these two considerations we have decided to include 6 lags, thereby aiming at capturing lagged effects in the market for the past half year. Most forecasting literature within the field of econometric modeling of financial data use a lag window in the range of 3 to 24 months,[13] and we therefore find our choice towards the conservative side of the spectrum, as it will limit the occurrence of false significant variables, while at the same time capture relevant past market dynamics during the past six months.

5.3. Seasonality

If a time series is measured more frequent than once every year, for instance on a quarterly or monthly basis, it is said to contain seasonal components when systematic patterns occur at specific points during a year. The underlying reason could possibly range from weather and holidays to the timing of decision making in various countries. It is common to sort seasonal effects into three different forms: deterministic seasonality, stochastic seasonality or a combination between the two. A time series with deterministic

[12] It is important to emphasise that the t- distribution is relatively robust to deviations from normality when the sample size is grater than 30. Unless the departure from the normal distribution is very pronounced, the frequency of type 1 errors should not deviate substantially from the selected significance level.

[13] See [Ye et al., 2006] and Baumeister and Kilian [2016] for respective examples.

seasonality will display the same seasonal effect at the same season every year (peaks and troughs every year remain constant), while peaks and trough may be shifted in the case of stochastic seasonality. Kavussanos and Alizadeh-M [2001, 2002] investigated the seasonality patterns in both the dry bulk and tanker shipping markets. They concluded that there was no evidence of stochastic seasonality in neither market, while deterministic seasonality was found in both markets. Accordingly, they proposed that prediction models within both markets would benefit from incorporating deterministic seasonality effects. In order to test their proposition, we construct a seasonally adjusted "twin" model for every non-seasonal model we create. If the seasonally adjusted model deliver significantly better forecasts out-of-sample than their non-seasonal peer, we have strong reason to believe that there are effects of deterministic seasonality. We will carry out the same analysis in both the dry bulk and tanker market segments. The following steps explains how the seasonal models are constructed in order to take deterministic seasonality into account:

Step 1 We decompose the BDI/BDTI- time series into two new series, a trend series and a seasonal series. The seasonal one consist of 12 components (since the data frequency is monthly). The seasonal components are constructed by introducing a matrix of dummy variables $D_{s,t}$, and unique seasonal coefficients ξ_s^{BDI} for the BDI model and ξ_s^{BDTI} for the BDTI model, to account for their respective seasonal effects. Here $s \in S$ and $S = \{Jan, Feb, ..., Dec\}$ since the data frequency is monthly. A common mistake in seasonal regression is to regress η_t on all seasonal dummies *and* the intercept. Either the intercept must be omitted or the number of seasonal components reduced by one, otherwise the variables will become collinear. We choose the latter, and also define January to be the base month where all the dummies are set to zero. The mathematical formulation follows; here the variable η_t account for seasonality at time step t,

$$\eta_t^{BDI} = \sum_{s=Jan}^{Dec} \xi_s^{BDI} D_{s,t} \tag{5.4}$$

$$\eta_t^{BDTI} = \sum_{s=Jan}^{Dec} \xi_s^{BDTI} D_{s,t} \tag{5.5}$$

and where $D_{s,t} = 1 | s = t \cap t \neq 1 + 12n, n \in N$, otherwise $D_{s,t} = 0$. The set N includes the number of years in the data sample, that is, {1, 2, ..., 16}, since our data set includes data from the range 2000 - 2016.

Step 2 The next step is to obtain the trend series. This can be achieved by subtracting the seasonal series from the original BDI/BDTI- series, i.e. subtracting each seasonal component from the original series at each time step. Once we have the trend series, the next step is to formulate a model that predicts the trend series. Here, we carry out a stepwise selection procedure identical to the one outlined in section 5.5.

Step 3 After the trend model is formulated, the last step is to add the seasonal series found in step 1 back to the trend model formulated in step 2 at every time step. The result is a model that takes deterministic seasonality into account.

5.4. Model formulation

Two time series could in some cases have a delayed relationship. If formulated correctly, a model with lagged explanatory variables can successfully capture these effects and hence explain a relationship more accurately [Hyndman and Athanasopoulos, 2013]. We construct models that exclusively incorporate lagged explanatory variables in order to be able to predict the BDI and BDTI. With n variables and v lags of each variable, the one-month forecast for the BDI and BDTI, y_t^{BDI} and y_t^{BDTI}, can be expressed as,

$$\begin{aligned} y_t^{BDI} &= \beta_0 + \beta_{1,1}x_{1,t} + ... + \beta_{1,v}x_{1,t-v} + \beta_{2,1}x_{2,t} + ... \\ &\qquad + \beta_{n,v}x_{n,t-v} + \eta_t^{BDI} + u_t \\ &= \beta_0 + \sum_{k=1}^{n}\sum_{p=1}^{v}\beta_{k,p}x_{k,t-p} + \eta_t^{BDI} + u_t \end{aligned} \tag{5.6}$$

$$\begin{aligned} y_t^{BDTI} &= \beta_0 + \beta_{1,1}x_{1,t} + ... + \beta_{1,v}x_{1,t-v} + \beta_{2,1}x_{2,t} + ... \\ &\qquad + \beta_{n,v}x_{n,t-v} + \eta_t^{BDTI} + u_t \\ &= \beta_0 + \sum_{k=1}^{n}\sum_{p=1}^{v}\beta_{k,p}x_{k,t-p} + \eta_t^{BDTI} + u_t \end{aligned} \tag{5.7}$$

where $x_{k,t-p}$ denotes factor $k \in K$ with lag $p \in P$ and $u_t \sim (0, \sigma)$. The sets of variables and lags are given by $K = \{1, .., n\}$ and $P = \{1, .., v\}$. The seasonality variables η_t^{BDI} and η_t^{BDTI} are explained in section 5.3. In models not accounting for seasonality, the η_t's are set to zero and the y_t's are untouched, while in the seasonal case the seasonal components are subtracted from y_t's, to obtain the trend components, as outlined section 5.3. When the variable $x_{k,t-p}$ is found to be non-significant, the corresponding $\beta_{k,t-p}$ equals zero, as outlined in the F-test description in Appendix C.8.

5.5. Model selection

It is not straight forward to find the optimal model when a regression analysis involves many explanatory variables. First, a model selection strategy for finding the optimal collection of explanatory variables needs to be established. A stepwise selection procedure could be carried out, which involves adding or removing (depending on method) one variable at each step, or by doing a best subset analysis, which involves searching all possible models and selecting the one with the best evaluation criterion. The model is selected by optimising a chosen criterion such as the R^2_{adj}, Akaike information criterion (AIC) [Akaike, 1973] or Bayesian information criterion (BIC) [Schwarz, 1978]. Campos et al. [2005] provide a complete overview of selection criteria. Because the stepwise procedure adds or removes one variable at each time step, it does not evaluate all possible model combinations, thereby making the required model running time relatively short. On the other hand, the best subset analysis searches all possible combinations and will consequently find the best model, but at the cost of exponential processing time. In our case, we aim to construct models from a set of 264 potential dry bulk variables, and 222 tanker variables (44 and 37 variables, respectively, each with six lags). The number of unique model combinations is thus 2^{264} and 2^{222} - astronomical numbers which no supercomputer will have the ability to evaluate in a reasonable amount of time. The best subset analysis of the complete set of data is clearly ruled out as an option. An alternative approach would be to apply the best subset search on a subset of our variables, and then choose model based on, for instance, $Mallow's\ C_p$ (see Appendix B.4). However, most statistical software have an upper limit in best subset search around 20 variables, and it is very likely that some of our models with the least strict significance criteria will yield models with twice as many variables. In order to let our models compete on similar terms, we resist the temptation of improving some of the leaner models, while being unable to do the same with larger models. Thus we restrict our selection procedure to a stepwise search.

Among the class of stepwise procedures there are three main types: Forward, backward and a combination of the two, often referred to as just *stepwise*. In our specific case, backward selection is not an option, because it starts with all possible variables included in the model. In that particular case, the regression would consists of more variables than observations, resulting in "negative degrees of freedom". Hence the backward elimination would need a modified starting point (starting with a subset of variables) in order to function. The forward search starts with an empty model and search for the single variable that will add most explanatory power. Since we have no ability to foresee which variables that will end up as significant, this is a process more suited for our purpose. In the next iteration of the search, the forward procedure search for a new variable that, in combination with the first, will provide most explanatory power. The stepwise procedure differs from the forward procedure by also excluding potential non-significant variables in each iteration.

A mathematical description of the stepwise selection procedure is now outlined. The statistical definitions of R^2, *MSS* (model sum of squares) etc. are defined in (Appendix B.2). The notation $M(b_i)$ refer to the *MSS* where only the variable x_i is included, $M(b_i, b_j)$ refer to the *MSS* where both the variable x_i and x_i are included, and lastly $M(b_i|b_j)$ to the marginal *MSS* added by x_i when x_j is already selected as regressor.

Step 1 Select the variable that gives the largest model sum of squares (MSS) when performing a simple linear regression with y. This is equivalent to that which gives the largest value of R^2. Let us call this variable x_1. If x_1 fails to be significant, the procedure is terminated before x_1 is added.

Step 2 Select the variable that, in combination x_1, gives the largest increase in R^2, minus the R^2 found in step 1. Algebraically, this is the variable x_j that maximise the expression:

$$M(\beta_j|\beta_1) = M(\beta_1, \beta_j) - M(\beta_1) \tag{5.8}$$

We call this variable x_2. The regression model with x_1 and x_2 is then constructed and R^2 observed. Similarly, if x_2 is insignificant, the procedure is terminated before x_2 is added.

The stepwise method differs from forward in the next step, where we test whether x_1 is still significant in the presence of x_2. We perform an F-test by taking the ratio

$$f = \frac{M(\beta_1|\beta_2)}{s^2} \tag{5.9}$$

where s is the standard deviation obtained from the regression model with x_1 and x_2. If $f < f_{alpha}(1, n-3)$ then x_1 is excluded.[14] The sample size is denoted n.

Step 3 Select the variable x_j that maximise the expression:

$$M(\beta_j|\beta_1, \beta_2) = M(\beta_1, \beta_2, \beta_3) - M(\beta_1, \beta 2) \tag{5.10}$$

This will result in the largest increase of R^2 compared to the R^2 in step 2. Calling this variable x_3, we now have a regression model involving x_1, x_2 and x_3, unless x_3 is insignificant. The ratio

$$f = \frac{M(\beta_3|\beta_1, \beta_2)}{s^2} \tag{5.11}$$

evaluates the appropriateness of x_3 in the model. Like in step 2, s represents the standard error obtained from the regression with variables x_1, x_2 and x_3. If x_3 is significant, the next step is to control that x_1 and x_2 is significant under the influence of x_3.

$$f = \frac{M(\beta_1|\beta_2, \beta_3)}{s^2} \qquad f = \frac{M(\beta_2|\beta_1, \beta_3)}{s^2} \tag{5.12}$$

This procedure is repeated until the most recent variable inserted fails to induce a significant increase in the explained regression. For instance in step 3, if the ratio calculated in (5.11) is lower than the critical value, $f < f_{alpha}(1, n-4)$, then x_3 is not included, the process is terminated and the appropriate regression equation contains the variables x_1 and x_2. The ratios in equation 5.12 are compared with the same critical value as 5.11.

An advantage of using the stepwise selection is that it, to a some extent, deals with the problem of multicollinearity (see Appendix B.7). The stepwise selection procedure will not choose a variable if the marginal explanatory power added by that variable is lower than other alternatives. As correlated predictors capture much of the same movements in the response variable, they typically have lower joint explanatory power than a less

[14] Degrees of freedom is n - 3 since three parameters have been estimated in the regression, namely the intercept as well as the slope of x_1 and x_2.

correlated alternative. Thus, it is unnecessary to exclude collinear variables prior to the model selection process. Yet, the problem of multicollinearity may grow as the R^2 of the model increase, since all alternatives may be more or less correlated with variables already present in the model. To monitor the extent of the issue we will display the VIF (variance inflation factor) for every variable in all final models. A presentation of VIF is provided in Appendix B.8.

5.5.1. Termination criteria

When there is a large number of potential explanatory variables, a too relaxed termination criterion could overestimate the amount of explained variance in the data, i.e. over-fitting of the model [Harrell, 2001, Stephen Olejnik, 2000, Moutinho and Huarng, 2013, Campos et al., 2005]. Consequently, the model will have a much better fit in-sample than out-of-sample. Techniques that have been discussed in the literature to handle this issue include imposing a stricter termination criterion, evaluating the variance of the explanatory variables and placing an upper limit on R^2 in the complete model. Blanchet et al. [2008] suggested the latter by implementing an additional step in the stepwise selection algorithm. L. S. Freedman [1992] and J. B. Copas [1991] suggested a implementing a stopping criterion where the variance of the explanatory variables are compared relative to a Bayesian criterion. In any case, a stricter termination criterion will likely reduce the number of variables in the final model, and thus reduce the risk of over-fitting the model. Since this has proven to be a simple, yet efficient way of handling the issue of over-fitting, we limit ourselves to this technique. On the other hand, placing a too strict significance criteria could possibly result in an under-specified model. As it is impossible to foresee which significance criteria that will yield the best trade-off between under-specifying and over-fitting the model, we define three unique models by utilising three different significance criteria.

In the $Alpha_1$ model, we formulate a rather conservative termination criterion of a maximum significance level of *one percent* for all variables selected. Meaning that at each iteration, the stepwise procedure is terminated if the entering explanatory variable has a significance level above the one percent mark. For $Alpha_5$ and $Alpha_{10}$, the criteria are relaxed somewhat to *five* and *ten percent*, respectively. The null hypothesis is rejected when the t-statistic is above its critical value. As all hypotheses are two sided, the critical t value contains the index $\frac{\alpha}{2}$.[15] The criteria is displayed below for model $Alpha_i$ by inserting α=i%, where $i = \{1,5,10\}$.

$$|t_{Alpha_i,v}| > t_{\frac{\alpha}{2},v} \tag{5.13}$$

Since the degrees of freedom, v, equals the sample size minus the number of parameters estimated (including the intercept), and our in-sample window is $n = 150$, the distribu-

[15] Even though the column in Table 4.1 containing arrows is titled "hypothesis", it does not refer to hypothesis in the statistical sense, but it refers to the anticipated impact on freight rates based on our study of literature. All statistical hypotheses are two-sided, despite the fact that some arrows in the table are unidirectional.

tion of the t-statistic approach the distribution of z-statistic, unless approximately 120+ variables are found to be simultaneously significant.

5.5.2. Model overview

By utilising three different significance criteria, we end up with a total amount of 12 prediction models: We create three non-seasonally adjusted dry bulk models and three seasonally adjusted dry bulk models, and in a similar manner three plus three tanker models. Hereafter, we will refer to the non-seasonally adjusted dry bulk models as $BDI-Alpha_i$, where $i = \{1,5,10\}$ represents the termination criteria, and their seasonally adjusted twins[16] are denoted $SBDI - Alpha_i$. Correspondingly, we will refer to the non-seasonal tanker models as $BDTI-Alpha_i$ and their seasonal peers as $SBDTI-Alpha_i$.[17]

5.6. Investigating the OLS assumptions

The OLS assumptions can be investigated both quantitatively, by conducting misspecification tests, and graphically, by studying residual plots. A quantitative approach is most commonly documented in forecasting literature, at least within the field of econometrics. While a quantitative approach is usually sufficient, a combination with a graphical inspection will always give supplementary information that is useful when evaluating the degree of model validity. We will therefore combine the two approaches. In the next section we will present the OLS- assumptions, then explain how each assumption can be checked from a graphical/qualitative viewpoint, before introducing the quantitative misspecification tests and how they should be interpreted. We will also address how violations of the model assumptions will affect the OLS estimators, and discuss potential remedies.

5.6.1. Brief presentation OLS assumptions

The *ordinary least squares* (OLS) theory is outlined in Appendix B.1, along with a mathematical presentation of the four principal assumptions underlying OLS. In words, the four assumptions are:

i) Linear relationship between response- and explanatory variables.
ii) No autocorrelation in errors.
iii) Homoscedastic (constant) errors.
iv) Normally distributed errors with zero mean.

5.6.2. Qualitative approach - link between residual plots and OLS assumptions

Since the "true" error term (u_t) cannot be observed, we have to rely on the obtained residuals when investigating whether the OLS assumptions are (approximately) valid.[18] A residual at timestep t is denoted

$$\hat{u}_t = y_t - \hat{y}_t \tag{5.14}$$

[16]When we use the terms "twin" or "peer" we refer to a model with identical termination criteria and market segment, i.e. BDI_1 and $SBDI_1$ are twin models.

[17]For convenience, we will sometimes skip the Alpha notation, e.g. $BDTI_{10}$ refers to $BDTI-Alpha_{10}$ and $SBDI_1$ refers to $SBDI - Alpha_1$.

[18]Note that all real life models will have slight departures from the model assumptions. There is no perfect model. As George Box famously pointed out: "All models are wrong, but some are useful."

where $\hat{y}_t$ is the fitted regression value. Residual plots for all models along with a discussion are listed in Appendix G. Each discussion will be based upon the "checklist" being laid out in the section below.

i) The assumption is most effectively checked by studying a plot of the residuals versus fitted values. If the linearity assumption holds, the points should be close to symmetric around the horizontal line, the points should be randomly distributed - but the point density should decrease as the distance from the horizontal line increases - and there should be no specific pattern in the "cloud" of data-points. For instance, a missing positive second order term would typically reveal itself by displaying a pattern with positive residuals for small fitted values, negative for medium ranged fitted values and again positive for large fitted values. In a similar fashion, residuals versus each explanatory variables could be checked, to identify potential non-linear relations. (The latter approach is more practical when working with fewer explanatory variables or if one have reason to suspect non-linearity in a subset of the variables.)

ii) The assumption is examined by studying a plot of the residuals versus observation order. If there is no significant autocorrelation, the data points should be randomly distributed around the horizontal line without any visible patterns/ trends. However, it may be difficult to discern whether there is significant autocorrelation from the residual versus order chart (especially if its negative, as it might be hard to differentiate a series of alternating signs relative to a series of random signs), hence one should also investigate ACF plots of the residuals (see Appendix G.1). If there is significant autocorrelation, it means the information about a residual automatically gives you information about its neighbouring residuals.[19] In the case of positive autocorrelation, it means that neighbouring residuals most likely will have the same sign, while negative autocorrelation means that neighbouring residuals tend to have alternating signs.

iii) The assumption is most effectively checked by studying a plot of the residuals versus fitted values. Heteroscedasticity reveal itself by displaying a particular shape in the diagram. For instance if the distance between the residuals increase with fitted values it means that the variance of the residuals increase with larger values of the explanatory variables, i.e. non-constant variance. Any sign of a non-uniform residual distribution along the horizontal axis indicate heteroscedasticity, meaning that an increase in variance with smaller/ negative explanatory variable equally well points to heteroscedasticity. If the distance from the residuals to the x-axis seem independent of the x-value, there is no reason to suspect non-constant variance. One may also look at plots displaying residuals versus specific covariates. Variance could indeed vary as a function of just one covariate without showing apparent signs heteroscedasticity on the plot of residuals versus fitted values. However, as pointed out earlier, this task is tedious when working with larger models.

iv) The degree of validity of the normality-assumption could be interpreted from both the Q-Q plot and the histogram. Though, it must be pointed out, that even

[19]Since we are performing a time-series regression, "neighbouring residuals" refer to residuals associated with consecutive months.

if the diagrams should seem to indicate a non-normal distribution, the apparent non-normality could originate from a non-linear relation between the dependent variable and the covariates. That is, non-normality could be a symptom of function misspecification. If, on the other hand, the residual plot versus fitted values indicate that the linear model specification is adequate, we have reason to interpret the two normality diagrams in a straight forward fashion.

5.6.3. Quantitative approach

i) The test most frequently utilised to test the linearity assumption is the Reset-test (see Appendix C.6). Usually, one have no idea which variable(s) that may poorly adhere to the linearity assumption. In that case, the Reset-test is advantageous, as it test for possible misspecification in the model as an entirety, and not on the level of individual variables. The Reset-test is an F-test, and accordingly there is a t-statistic associated with each term constituting the test. In our case, we will search for possible model terms of second and third order.

ii) There exist various quantitative tests for autocorrelation. One of the most frequently used is the Durbin-Watson (DW) test, which tests for autocorrelation in the first lag. It is is obviously useful to evaluate whether the first lag is autocorrelated, and the DW statistic is frequently reported in the literature, therefore we will perform the DW test on all our models. However, the Breusch-Godfrey (See Appendix C.3) test is more general, and tests for autocorrelation in a certain number of pre-determined lags. Since our data-resolution is monthly, we will look for autocorrelation in every lag from 1-12. The B-G test is a joint F- test, providing a t-statistic for every lag tested. Instead of printing the entire F-table with 12 individual t-statistics, we will print the F- statistic along with ACF-plots for all models. Each individual bar in an ACF-plot provides the same amount of information as the individual t-tests making up the F-test, but the graphical representation makes it easier to see general autocorrelation-pattern across lags.

iii) Heteroscedasticity refers to non-constant variance in the residuals. Among several statistical tests to identify heteroscedasticity, the GoldfeldâQuandt test [Goldfeld and Quandt, 1965], the BreuschâPagan test [Breusch and Pagan, 1979]) and the White's test [White, 1980] are most frequently used. The latter is particularly advantageous as it makes no assumptions about the form of the heteroscedasticity. Similarly to the BG- test, the White test is an F-test. In the test, every variable (and possibly every cross-term, if one have a sufficient degrees of freedom) is inspected for their ability to explain the residual variance. Because the F-test provides individual t-values for every term tested, it is straight forward to locate the potential sources of non-homogeneity.

iv) To test for normality, we apply the Anderson-Darling test statistic as Stephens [1974] found it to be one of the EDF's that provide best power, when it comes to detect departures from normality in our sample range. The Anderson Darling statistic is presented in Appendix C.7. A second option would have been to employ the famous test statistic presented by Bera and Jarque [1981]. While we do not use Jarque-Bera to test for normality in the residuals obtained from our prediction models, the test statistic is shown in Table 4.2 to test for normality in every individual time-series.

5.6.4. Consequence of model violations and possible remedies

i) If the linearity assumption is violated, then the OLS estimators will no longer be unbiased, and we essentially have the wrong equation specified. If the Reset-test rejects the null hypothesis of linearity, one option is to add terms second-order or third order. We could potentially also add interaction terms (i.e. cross product terms), however, if significant interaction are found, we cannot draw conclusions about the main effects, but would first need to investigate interaction plots for every significant combination. Such analysis would be outside the scope of this thesis. Moreover, the available degrees of freedom would limit the number of interactions that could be simultaneously tested. The issue of misspecification may also be handled by introducing piece wise linear variables that are well specified in their domain. Another option is to introduce variable transformations, for instance taking the logarithm if there seem to be exponential relations between variables, or if the model is multiplicative (which in itself is an implicit violation of the linearity assumption), a log transformation will turn it additive. Another issue that may cause the Reset-test to reject the null hypothesis is the "omitted variable bias" - leaving out important variables from the model, causing the present model variables to compensate for the ones left out. According to Brooks [2008], this may in fact be the case regardless of the Reset-test outcome. The remedy in that case is obviously to add more variables to the model. Lastly, extreme outliers, especially in the dependent variable, may lead the Reset- test to reject the null hypothesis. However, the occurrence of outliers is not per se related to the linearity assumption, but belongs to the normality domain, and will be further discussed in that section.

ii) If residuals are autocorrelated, then the OLS estimators are no longer be most efficient, but they will remain unbiased (See Appendix B.1 for a summary of the Gauss-Markov theorem). That is, some other estimator would more precisely estimate the regression parameters with the same number of data points. The intuition behind this is that some information about the residuals is not captured by the OLS-estimators. If the B-G test rejects the hypothesis of no autocorrelation, there could be several possible explanations. One possible source could be an underlying seasonality pattern in the dependent variable that is not accounted for. For instance, if there is significant autocorrelation in all our models which are not seasonally adjusted, but not in the adjusted models, there is reason to believe that a repeated seasonal pattern at least may partially explain the autocorrelation. The source of the autocorrelation could also be the "omitted variable bias" discussed above. In that case, the model is underspecified, and slacking the significance criteria such that more variables are found to be significant could resolve the issue.

iii) The consequence of heteroscedasticity bears resemblance to the one discussed for autocorrelation - the OLS estimators become inefficient, yet they remain unbiased. Since the standard error in the coefficient estimates may be incorrectly estimated, inferences made from regression may be misleading. For instance the standard errors may be inflated in certain covariate regions relative to others, so the true standard errors should be a function of the covariates being responsible for the non-homogeneity. Heteroscedasticity is commonly dealt with by variable transformation or by using a software that adjusts the standard errors of the coefficients to account for the non-constant variance.

iv) Violation of the normality assumption could lead the OLS estimators to become biased. For instance, a very large, positive outlier may cause the the coefficients to over-adjust to reduce that particular error- distance. The good news is that this is mainly a problem when the sample size is small, as the t-distribution is relatively robust to small to medium departures from normality, when the sample size is greater than 30. In our case we have an in-sample window size of 150 observations, thus the effective number of degrees of freedom is 150 minus the number of parameters estimated (including the intercept). As we have less than 50 variables in all our models, we always have more than 100 degrees of freedom in every model. Having this in mind, we know that unless the deviations from normality are severe, we can rely on the coefficient estimates provided by the t-distribution. If, on the other hand, the residuals are very non-normal, the best remedy is to amke use of a variable transformation, where the best suited transformation will depend on the specific deviation from normality.

5.7. Model evaluation

We evaluate two aspects of our selected models: The degree of over-fitting and the predictive power.

5.7.1. Over-fitting

Empirical models will inevitably capture idiosyncratic noise. Increasing the complexity of a model will increase its ability to fine-tune itself to fit the idiosyncratic noise. Since the noise is unlikely to repeat itself in the future, more complex models are rarely superior to simpler ones when evaluated out-of-sample. A model that captures too much noise is said to be over-fitted.

Since the purpose of our research is to provide insight into freight rate changes, we intend to construct models that are able to capture the real underlying dynamics impacting freight rates, and hence seek to limit the degree of over-fitting. In order to monitor the degree of over-fitting, we compare the residuals obtained in-sample versus out-of-sample. Specifically, we compare the mean absolute error (MAE), root mean square error ($RMSE$), percentage of correct direction predicted ($\%CD$), mean absolute scaled error ($MASE$) and scaled root mean square error ($SRMSE$) in-sample and out-of-sample. Definitions of MAE, $RMSE$, $MASE$ and $SRMSE$ follows,

$$MAE(y, \hat{y}) = \frac{\sum_{t=1}^{N} |y_t - \hat{y}|}{N} \tag{5.15}$$

$$RMSE(y, \hat{y}) = \sqrt{\frac{\sum_{t=1}^{N} (y_t - \hat{y})^2}{N}} \tag{5.16}$$

$$MASE(y, \hat{y}, s(y)) = \frac{1}{s(y)} \frac{\sum_{t=1}^{N} |y_t - \hat{y}|}{N} \tag{5.17}$$

$$SRMSE(y, \hat{y}, s(y)) = \frac{1}{s(y)} \sqrt{\frac{\sum_{t=1}^{N} (y_t - \hat{y})^2}{N}} \tag{5.18}$$

Where $\hat{y}_t$ denotes the fitted model value at time step t, $\bar{y}$ is the out-of-sample mean and $\{1, .., N\}$ is the out-of-sample space. The MAE and $RMSE$ metrics are included mainly for completeness. As the standard error of the dependent variable may vary considerably in-sample and out-of-sample, the best suited test metrics are those measuring the errors relative to the standard error. For instance, the MAE-metric may indicate that the out-of-sample accuracy outperforms the in-sample accuracy if the standard error is considerably lower out-of-sample. As it is easier to score a low MAE when the standard error is lower, the $MASE$- score is more appropriate as it will readjust the imbalance caused by different standard errors. Similarly, $SRMSE$ is a more appropriate metric than $RMSE$. The $MASE$ and $SRMSE$ metrics are expected to increases slightly out-of-sample versus in-sample. An excessive increase, however, points towards an over-fitted model.

5.7.2. Predictive power

Due to the effect of over-fitting, the predictive power of a model will not be properly reflected by the in-sample R^2. The out-of-sample prediction errors or the out-of-sample R^2 give a substantially clearer indication. However, to identify the real usefulness offered by any prediction model, it must be compared with relevant benchmarks. Hyndman [2010] recommends two classes of benchmarks for prediction; one "naive" and one standardised. A random walk (which is an ARIMA model with only one autoregressive component) and a well-specified ARIMA(p, d, q) model[20] serves the purpose. The random walk, ARIMA$(0, 1, 0)$, is given by

$$y_t = y_{t-1} + u_t \tag{5.19}$$

where y_t is the lagged dependent variable. Note that both benchmarks are univariate models, i.e. models composed of only the dependent variable. One might say that the benchmarks represent our "best guesses", if no explanatory variables are involved. The predictive performance of our models is evaluated by comparing the following two aspects with the benchmarks: i) Predictive accuracy and ii) Correlation with the dependent variable. The first point (i) refers to the amount of error, either in absolute or squared terms, the forecasts accumulate during the out-of-sample period. The second point (ii) refers to the amount of variation in the dependent variable the models are able to explain.

[20]Optimised by minimising the AIC in-sample. See Appendix B.5 for a presentation of AIC, and Appendix B.6 for a definition of an ARIMA model.

If our generated prediction models are unable to beat the benchmarks, they offer limited value. The contrary - if our models beat the benchmarks, how confidently can we say the models generally provide superior predictions? When two models "compete" out-of-sample, one will draw the longest straw. But if the out-of-sample window is extended or changed, how sure can we be that the one-time winner would turn out to be a consistent winner? The Diebold-Mariano (DM) test, presented in Appendix C.9, aims to answer that question. However, the DM-test only measure aspect (i) of the forecast, and accordingly, only the conclusion regarding predictive accuracy is drawn based on the outcome of the DM-test.

5.7.3. Determining the out-of-sample window

Choosing a proper out-of-sample window is a trade-off between model calibration and testing. A reasonable share of the total degrees of freedom must be spent on calibrating the model - the more degrees of freedom spent, the more likely will the model be able to separate idiosyncratic noise from the true underlying market dynamics. At the same time, a portion of the degrees of freedom must be saved up to the out-of-sample window, to test the models ability to differentiate noise from actual freight rate signals. We have decided to use an in-sample period from Jul-2000 to Dec-2012, corresponding to 150 observations.[21] Consequently, the out-of-sample period will consist of 42 observations from Jan-2013 to Jun-2016. Even though there exist robustness tests for window size, it is common in forecasting literature to report results based on one window only [Rossi and Inoue, 2012]. Moreover, the uncertainty associated with reporting results from only one out-of-sample window is less of an issue when a DM-test is conducted on the test results obtained out-of-sample.

[21]The first six observations in 2000 are not useful as there are six lags included in the model specification.

6 Results

This section begins with a brief summary of the of the stationarity test results, followed by a presentation of the seasonal components. We continue by presenting the results obtained by the stepwise model selection procedure, before analysing the misspecification tests results. Then, we provide a thorough review of significant variables and interpret their economical meaning. Furthermore, we outline the out-of-sample results and discuss the degree of over-fitting. Lastly, we present the Diebold-Mariano test results, and evaluate the predictive power of all models.

6.1. Stationarity

The Augmented Dickey-Fuller (ADF) test was performed on all time series. Table D.1 provides an overview of the test statistics and significance levels. At a five percent significance level, signs of non-stationarity was detected in nine time series, and consequently, they were transformed. After the transformation they were all found to be stationary at a significance level of either 1% or 5%.

6.2. Seasonality

The monthly seasonal components of dry bulk and tanker freight rates in the in-sample period are illustrated in Figure 6.1. We observe that the BDI and BDTI seasonal patterns have some commonalities, but also some apparent differences. They both have a vast drop in January followed by a relative rise in February. BDI reaches its seasonal peak at this point, while the BDTI component still has a negative derivative. In March, BDTI continues to climb, unlike the BDI component that drops moderately. Both models follow a decline-rise-decline pattern in April, May and June, and reach a local minimum in July and August, respectively. In August, BDI rises before gradually diminishing as the year passes by, while the BDTI picks up in September and remains on the positive side until the end of the year. Some of the patterns here are expected and worth commenting on. The seasonal components of the BDTI rates during the end of the year might reflect larger oil consumption and inventory build-up on the northern hemisphere, right before the winter is coming. The drop in January in both markets could be a consequence of the Chinese new year celebration, which put a damper on demand for several days each January. Also, the dry bulk peak in February is expected, as Japan tend to increase their demand of dry bulk commodities before they enter a new financial year every March.

Kavussanos and Alizadeh-M [2001, 2002] also investigated seasonality patterns in both dry bulk and tanker freight markets. Although their analysis were performed on data in the period 1978-1996, and hence had no overlap with our data set from 2000-2013, their deterministic seasonality observations appear to be consistent with our findings. Similarly to us, Kavussanos and Alizadeh-M [2002] found that tanker rates usually increase in November and December, while rates tend to decrease from January to April. Also, we both find that rates increase in May, before dropping during mid-summer. When it comes to the dry bulk segment, we agree on the derivative sign of every month except for April, where Kavussanos and Alizadeh-M observed an increase in rates, while we found a decline.

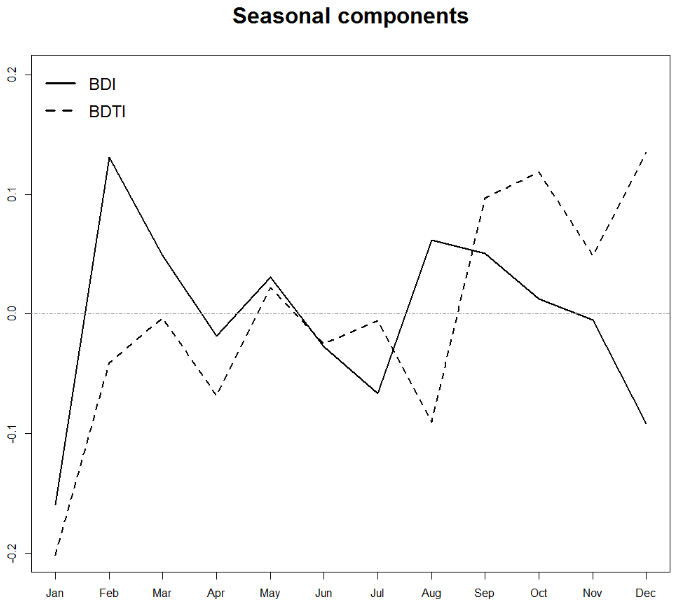

Figure 6.1: Deterministic seasonality components calculated with the in-sample monthly mean change in the period of January 2000 to December 2012.

6.3. Model selection

This subsection displays the results obtained from the stepwise regression procedure. In Table 6.1 through 6.6, there are a total of 12 models presented, representing a set of models matching unique combinations of market segment (dry bulk/tanker), adjustment (seasonally adjusted/ non-adjusted) and significance criteria (1%, 5% or 10%). Each table shows the set of variables found to be significant in the respective regression, along with their coefficient, standard deviation, significance and variance inflation factor. Before we begin interpreting the explanatory variables in each model, we provide a thorough discussion of model properties and misspecifications.

Table 6.1: Dry Bulk $Alpha_1$ freight rate regression output

Baltic Dry Index - $Alpha_1$											
Not adjusted for seasonality						**Adjusted for seasonality**					
Variable, $\mathbf{x_{k,t-p}}$	$\beta_{\mathbf{k,p}}$	$SD[\beta_{\mathbf{k,p}}]$	**t**	$\mathbf{P > \|t\|}$	*VIF*	**Variable, $\mathbf{x_{k,t-p}}$**	$\beta_{\mathbf{k,p}}$	$SD[\beta_{\mathbf{k,p}}]$	**t**	$\mathbf{P > \|t\|}$	*VIF*
Intercept	0.00130	0.0178	0.17	0.86	-	Intercept	0.0226	0.0183	1.23	0.219	-
ΔS_Ch_Prod_2	1.061	0.330	3.21	0.002***	1.07	ΔCPI_US_3	-14.81	4.75	-3.12	0.002***	1.03
ΔCPI_US_3	-14.07	4.18	-3.37	0.001***	1.04	ΔFX_USD_2	-1.772	0.628	-2.82	0.005***	1.03
ΔFX_USD_2	-1.810	0.643	-2.81	0.006***	1.04	ΔBulk_Index_1	0.721	0.149	4.84	0.000***	1.02
ΔBulk_Index_1	0.702	0.153	4.59	0.000***	1.03	ΔBulk_Index_3	-0.484	0.149	-3.24	0.000***	1.02
ΔBulk_Index_3	-0.570	0.156	-3.65	0.001***	1.08						

Significance level: *** 0.01, ** 0.05, * 0.1.

Table 6.2: Dry Bulk $Alpha_5$ freight rate regression output

Baltic Dry Index - $Alpha_5$											
Not adjusted for seasonality						**Adjusted for seasonality**					
Variable, $\mathbf{x_{k,t-p}}$	$\beta_{\mathbf{k,p}}$	$SD[\beta_{\mathbf{k,p}}]$	**t**	$\mathbf{P > \|t\|}$	*VIF*	**Variable, $\mathbf{x_{k,t-p}}$**	$\beta_{\mathbf{k,p}}$	$SD[\beta_{\mathbf{k,p}}]$	**t**	$\mathbf{P > \|t\|}$	*VIF*
Intercept	-0.00640	0.0127	-0.51	0.614	-	Intercept	-0.0326	0.0137	-2,37	0,019**	-
ΔIP_OECD_3	2.526	0.817	3.09	0.002***	1.34	ΔSBDI_5	-0,3026	0,0959	-3,15	0.002***	3,07
ΔIP_Ch_2	1.122	0.452	-2.48	0.014**	1.21	ΔIP_OECD_2	-2,91	1,20	-2,42	0,017**	1,80
ΔIP_I_6	1.507	0.339	4.45	0.000***	1.23	ΔIP_I_6	2,360	0,453	5,21	0.000***	1,36
ΔOil_P_G_3	-2.75	1.15	-2.38	0.019**	1.23	ΔIP_US_1	7,39	1,37	5,39	0.000***	1,56
ΔIO_Ch_Imp_4	-0.1457	0.0630	-2.31	0.022***	1.26	ΔG_US_Exp_4	0,247	0,106	2,34	0.021**	1,17
ΔIO_A_Exp_1	-0.3232	0.0988	-3.27	0.001***	1.27	ΔS_CH_Prod_2	1,087	0,273	3,99	0.000***	1,22
ΔC_EU_Imp_2	0.475	0.118	4.04	0.000***	1.18	ΔCPI_Ch_5	-6,73	1,72	-3,91	0.000***	1,14
ΔG_US_Exp_2	0.1928	0.0830	2.32	0.022***	1.25	ΔOrder_Bulk_2	0,755	0,305	2,47	0.015**	1,21
ΔS_CH_Prod_2	0.893	0.232	3.85	0.000***	1.42	ΔScrap_Bulk_6	0,1835	0,0340	5,39	0.000***	1,18
ΔCPI_US_3	-10.62	3.36	-3.16	0.002***	1.40	ΔDel_Bulk_4	-0.0169	0.0042	-4.03	0.000***	1.19
ΔCPI_Ch_2	6.81	1.53	4.45	0.000***	1.46	ΔSec_Bulk_2	-0.951	0.0231	-4.11	0.000***	2.01
ΔCPI_Ch_3	4.98	1.47	3.38	0.001***	1.35	ΔTC_Bulk_5	0,214	0,106	2,02	0,046**	3,17
ΔCPI_Ch_5	-7.27	1.41	-5.18	0.000***	1.23	ΔFX_USD_1	-2,322	0,569	-4,08	0.000***	1,42
ΔOrder_Bulk_4	0.653	0.238	2.74	0.007***	1.16	ΔLIBOR_4	-0.177	0.0427	-4.15	0.000***	1.32
ΔScrap_Bulk_6	0.0997	0.0275	3.63	0.000***	1.25	ΔP_IO_2	-0,523	0,176	-2,97	0.004***	1,82
ΔDel_Bulk_5	0.1504	0.0034	4.39	0.000***	1.27	ΔP_Gold_2	0.970	0.264	3.68	0.000***	1.19
ΔSec_Bulk_3	-0.688	0.163	-4.22	0.000***	1.61	ΔP_Gold_3	-0,550	0,273	-2,02	0.046**	1,28
ΔFX_USD_1	-1.829	0.452	-4.05	0.000***	1.45	ΔBulk_Index_1	0,789	0,139	5,66	0.000***	1,50
ΔLIBOR_4	-0.0871	0.0352	-2.48	0.015**	1.45	ΔBulk_Index_2	0.390	0.139	2.81	0.006***	1.47
ΔP_IO_2	-0.444	0.123	-3.60	0.000***	1.44						
ΔP_Coal_1	-0.486	0.150	-3.23	0.002***	1.45						
ΔP_Gold_5	-0.988	0.214	4.61	0.000***	1.27						
ΔBulk_Index_1	0.488	0.110	4.45	0.000***	1.49						
ΔBulk_Index_2	0.339	0.107	3.16	0.002***	1.43						

Significance level: *** 0.01, ** 0.05, * 0.1.

Table 6.3: Dry Bulk $Alpha_{10}$ freight rate regression output

Baltic Dry Index - $Alpha_{10}$

Not adjusted for seasonality						Adjusted for seasonality					
Variable, $\mathbf{x_{k,t-p}}$	$\beta_{\mathbf{k,p}}$	$SD[\beta_{\mathbf{k,p}}]$	**t**	$\mathbf{P} > \lvert\mathbf{t}\rvert$	*VIF*	**Variable, $\mathbf{x_{k,t-p}}$**	$\beta_{\mathbf{k,p}}$	$SD[\beta_{\mathbf{k,p}}]$	**t**	$\mathbf{P} > \lvert\mathbf{t}\rvert$	*VIF*
Intercept	-0.0119	0.0125	-0.95	0.344	-	Intercept	-0.0101	0.0132	-0.86	0.381	-
ΔBDI_2	-0.1479	0.0503	-2.94	0.004***	1.64	ΔSBDI_5	-0,1525	0,0544	-2.80	0.006***	2.25
ΔBDI_5	-0.1191	0.0491	-2.43	0.017**	1.54	ΔSBDTI_3	-0,1234	0,0556	-2.22	0.028**	1.41
ΔIP_OECD_3	3.083	0.812	3.80	0.000***	1.43	ΔIP_OECD_3	4.375	0.765	5.72	0.000***	1.66
ΔIP_Ch_2	1.229	0.452	2.72	0.008***	1.30	ΔIP_Ch_2	1.55	0.408	3.81	0.000**	1.39
ΔIP_I_6	1.315	0.329	3.99	0.000***	1.25	ΔIP_I_6	0.905	0.337	2.68	0.008***	1.71
ΔIO_B_Exp_2	0.0711	0.0365	1.95	0.054*	1.32	ΔIO_A_Exp_1	-0.2164	0.0901	-2.40	0.018**	1.48
ΔIO_A_Exp_1	-0.2713	0.0988	-3.75	0.007***	1.36	ΔC_EU_Imp_2	0.664	0.109	6.10	0.000***	1.42
ΔC_EU_Imp_2	0.291	0.118	2.47	0.015**	1.27	ΔC_J_Imp_6	-0.1720	0,0909	-1.89	0.061*	1.31
ΔG_US_Exp_6	0.178	0.0810	2.14	0.034**	1.24	ΔC_A_Exp_6	0.267	0.106	2.52	0.013**	1.52
ΔS_CH_Prod_2	0.836	0.224	3.72	0.000***	1.44	ΔG_US_Exp_2	0,1430	0,0722	1.98	0.050*	1.33
ΔS_CH_Prod_4	0.605	0.231	2.62	0.010**	1.52	ΔG_US_Exp_6	0.219	0.0772	2.85	0.005***	1.40
ΔCPI_US_3	-8.69	3.29	-2.64	0.009***	1.44	ΔS_CH_Prod_2	0.727	0.0219	3.32	0.001***	1.79
ΔCPI_Ch_2	6.53	1.52	4.30	0.000***	1.55	ΔS_CH_Prod_3	0.660	0.0231	2.86	0.005***	1.99
ΔCPI_Ch_3	4.07	1.46	2.78	0.006***	1.44	ΔS_CH_Prod_4	1.23	0.218	5.64	0.000***	1.77
ΔCPI_Ch_5	-7.63	1.39	-5.49	0.000***	1.30	ΔMS_Ch_2	-1.336	0.315	-4.24	0.000***	1.60
ΔOrder_Bulk_4	0.854	0.231	3.70	0.000***	1.18	ΔMS_Ch_6	-1.006	0.321	-3.14	0.002***	1.71
ΔScrap_Bulk_6	0.105	0.0266	3.95	0.000***	1.26	ΔCPI_US_3	-11.92	3.13	-3.80	0.000***	1.71
ΔDel_Bulk_5	0.0194	0.0033	5.92	0.000***	1.24	ΔCPI_Ch_5	-7.29	1.40	-5.20	0.000***	1.73
ΔSec_Bulk_3	-0.689	0.164	-4.21	0.000***	1.74	ΔFleet_Bulk_1	-12.58	3.01	-4.18	0.000**	1,73
ΔFX_USD_1	-1.761	0.443	-3.97	0.000***	1.50	ΔOrder_Bulk_2	0,691	0.224	3.09	0.003***	1.48
ΔLIBOR_4	-0.108	0.0341	-3.17	0.002***	1.46	ΔScrap_Bulk_3	0.0649	0.0283	2.29	0.024***	1.93
ΔP_IO_2	-0.339	0.127	-2.66	0.009***	1.65	ΔScrap_Bulk_4	-0.0644	0.0285	-2.26	0.026**	1.85
ΔP_Coal_1	-0.472	0.152	-3.10	0.002***	1.60	ΔScrap_Bulk_6	0.0911	0.0250	3.65	0.000***	1.45
ΔP_Coal_2	-0.309	0.147	-2.10	0.037**	1.49	ΔDel_Bulk_4	-0.0242	0.0034	-7.16	0.000***	1.77
ΔP_Wheat_5	0.254	0.140	1.81	0.073*	1.21	ΔSec_Bulk_2	-0.387	0.183	-2.12	0.037**	2.85
ΔP_Gold_5	0.815	0.207	3.93	0.000***	1.29	ΔSec_Bulk_3	-0.408	0.169	-2.42	0.017**	2.43
ΔBulk_Index_1	0.527	0.106	4.96	0.000***	1.51	ΔTC_Bulk_6	-0.153	0.0522	-2.93	0.004***	1.76
ΔBulk_Index_2	0.452	0.108	4.20	0.000***	1.55	ΔFX_USD_1	-2.516	0.408	-6.16	0.000***	1.66
						ΔFX_USD_6	-0.883	0.421	-2.10	0.038**	1.82
						ΔLIBOR_4	-0.120	0,031	-3.88	0.000***	1.58
						ΔVIX_2	0.128	0,0592	2.16	0.033**	1.81
						ΔVIX_3	0.161	0,0547	2.94	0.004***	1.54
						ΔP_IO_2	-0.667	0.129	-5.19	0.000***	2.21
						ΔP_Coal_2	-0.492	0.137	-3.59	0.000***	1.69
						ΔP_Coal_6	0.307	0.134	2.30	0.023**	1.59
						ΔP_Wheat_5	0.417	0.127	3.29	0.001***	1.28
						ΔP_Metals_5	-0.975	0,221	-4.40	0.000***	2.08
						ΔP_Gold_5	1.091	0.195	5.59	0.000***	1.49
						ΔBulk_Index_1	0.498	0.0967	5.15	0.000***	1.64
						ΔBulk_Index_2	0.622	0,111	5.61	0.000***	2.15
						ΔMSCI_W_6	0.577	0.230	2.51	0.014**	2.06

Significance level: *** 0.01, ** 0.05, * 0.1.

Table 6.4: Tanker $Alpha_1$ freight rate regression output

Baltic Dirty Tanker Index - $Alpha_1$

Not adjusted for seasonality						Adjusted for seasonality					
Variable, $\mathbf{x_{k,t-p}}$	$\beta_{\mathbf{k,p}}$	$SD[\beta_{\mathbf{k,p}}]$	**t**	$\mathbf{P > \|t\|}$	*VIF*	**Variable,** $\mathbf{x_{k,t-p}}$	$\beta_{\mathbf{k,p}}$	$SD[\beta_{\mathbf{k,p}}]$	**t**	$\mathbf{P > \|t\|}$	*VIF*
Intercept	-0.00441	0.0150	-0.49	0.622	-	Intercept	0.0198	0.0119	1.67	0.097*	-
ΔO_Ch_Imp_2	-0.397	0.0794	-3.44	0.001***	1.00	ΔCPI_US_5	-11.45	3.98	-2.91	0.004***	1.23
ΔOil_1	0.575	0.161	3.56	0.001***	1.00	ΔSec_Tank_1	1.054	0.287	3.67	0.000***	1.01
						ΔFuel_5	0.464	0.115	4.03	0.000***	1.22
						ΔOil_1	0.619	0.129	4.80	0.000***	1.01

Significance level: *** 0.01, ** 0.05, * 0.1.

Table 6.5: Tanker $Alpha_5$ freight rate regression output

Baltic Dirty Tanker Index - $Alpha_5$

Not adjusted for seasonality						Adjusted for seasonality					
Variable, $\mathbf{x_{k,t-p}}$	$\beta_{\mathbf{k,p}}$	$SD[\beta_{\mathbf{k,p}}]$	**t**	$\mathbf{P > \|t\|}$	*VIF*	**Variable,** $\mathbf{x_{k,t-p}}$	$\beta_{\mathbf{k,p}}$	$SD[\beta_{\mathbf{k,p}}]$	**t**	$\mathbf{P > \|t\|}$	*VIF*
Intercept	-0.00981	0.0140	-0.743	0.771	-	Intercept	0.0249	0.0145	1.77	0.096*	-
ΔBDI_5	-0.176	0.0616	-2.85	0.005***	1.18	ΔSBDI_5	-0.142	0.0578	-2.46	0.015**	1.38
ΔO_Ch_Imp_2	-0.274	0.0716	-3.83	0.000***	1.07	ΔIP_OECD_2	2.783	0.870	3.20	0.002**	1.17
ΔO_AG_Exp_5	-0.777	0.386	-2.01	0.046**	1.04	ΔOil_P_ME_4	1.204	0.526	2.29	0.024**	1.11
ΔCPI_US_3	10.94	4.14	2.64	0.009***	1.12	ΔMS_US_1	-2.37	1.03	-2.29	0.023**	1.09
ΔSec_Tank_1	0.791	0.328	2.41	0.017**	1.10	ΔSec_Tank_1	0.804	0.264	3.04	0.003***	1.04
ΔFuel_5	0.477	0.126	3.78	0.000***	1.23	ΔFuel_5	0.464	0.108	4.31	0.000***	1.30
ΔOil_1	0,619	0.147	4.21	0.000***	1.10	ΔOil_1	0,631	0.122	5.18	0.000***	1.10
ΔCont_Oil_6	0.991	0.431	2.30	0.023**	1.04	ΔFX_US_JP_1	0.852	0.301	2.84	0.005***	1.14
ΔLIBOR_1	0.1271	0.0418	3.04	0.003***	1.07	ΔCont_Oil_6	1.204	0.371	3.25	0.001***	1.12
ΔLIBOR_3	0.1178	0.0423	2.78	0.006***	1.10	ΔTank_Index_6	0.261	0.126	-2.07	0.040**	1.33

Significance level: *** 0.01, ** 0.05, * 0.1.

Table 6.6: Tanker Alpha_{10} freight rate regression output

Baltic Dirty Tanker Index - $Alpha_{10}$

Not adjusted for seasonality						**Adjusted for seasonality**					
Variable, $\mathbf{x_{k,t-p}}$	$\beta_{\mathbf{k,p}}$	$SD[\beta_{\mathbf{k,p}}]$	**t**	$\mathbf{P > \|t\|}$	*VIF*	**Variable,** $\mathbf{x_{k,t-p}}$	$\beta_{\mathbf{k,p}}$	$SD[\beta_{\mathbf{k,p}}]$	**t**	$\mathbf{P > \|t\|}$	*VIF*
Intercept	-0.0689	0.0214	-3.68	0.002***	-	Intercept	0.0269	0.0122	1.77	0.066*	-
ΔBDI_5	-0.1946	0.0496	-3.93	0.000***	1.51	ΔSBDI_3	0.1368	0.0488	2.81	0.006***	1.32
ΔBDTI_2	-0.2412	0.0496	-3.93	0.000***	1.75	ΔSBDI_5	-0.0963	0.0500	-1.93	0.056*	1.38
ΔBDTI_4	-0.2370	0.0573	-4.14	0.000***	1.48	ΔSBDTI_2	-0.2131	0.0631	-3.38	0.001***	1.32
ΔGDP_W_1	19.57	7.45	2.63	0.010**	2.82	ΔIP_OECD_2	2.812	0.808	3.48	0.001***	1.32
ΔIP_OECD_4	2.506	0.962	2.61	0.010**	1.94	ΔIP_I_4	-0.759	0.336	-2.26	0.026**	1.24
ΔIP_I_1	1.325	0.408	3.25	0.002***	1.32	ΔO_P_ME_4	-1.798	0.467	3.85	0.000***	1.16
ΔIP_I_2	1.484	0.388	3.82	0.000***	1.78	ΔO_P_ME_6	0.888	0.462	1.92	0.057*	1.20
ΔIP_US_1	3.85	1.13	3.41	0.001***	1.77	ΔO_AG_Exp_2	0.869	0.314	2.77	0.007***	1.30
ΔOil_P_G_5	3.43	1.29	2.67	0.009***	1.59	ΔO_AG_Exp_5	-0.495	0.287	-1.72	0.087*	1.11
ΔOil_P_ME_2	1.183	0.549	2.16	0.033**	1.58	ΔCPI_US_5	-15.25	4.60	-3.31	0.001***	2.67
ΔOil_P_ME_4	3.763	0.539	6.99	0.000***	1.57	ΔFleet_Tank_1	11.14	2.58	4.32	0.000***	1.59
ΔO_Ch_Imp_3	0.2167	0.0623	3.68	0.000***	1.26	ΔFleet_Tank_2	8.62	2.55	3.38	0.001***	1.55
ΔO_US_Exp_4	0.2167	0.0623	3.48	0.001***	1.26	ΔDel_Tank_5	0.021	0.0072	2.88	0.001***	1.07
ΔO_AG_Exp_5	-0.881	0.326	-2.70	0.008***	1.46	ΔSec_Tank_1	1.148	0.239	4.81	0.000***	1.13
ΔO_AG_Exp_6	0.734	0.333	2.20	0.030**	1.54	ΔFuel_5	0.893	0.141	6.32	0.000***	2.99
ΔMS_Ch_3	-1.049	0.372	-2.82	0.006***	1.64	ΔOil_1	0.641	0.107	6.00	0.000***	1.12
ΔCPI_US_3	9.28	3.45	2.69	0.008***	1.54	ΔOil_5	-0.615	0.171	-3.60	0.000***	2.96
ΔCPI_Ch_1	7.91	1.63	4.87	0.000***	1.71	ΔOil_6	0.528	0.189	2.80	0.006***	3.60
ΔCPI_Ch_2	7.27	1.89	3.83	0.000***	2.32	ΔFX_US_JP_1	1.092	0.268	4.07	0.000***	1.21
ΔCPI_Ch_3	7.74	1.72	4.49	0.000***	1.93	ΔFX_US_JP_4	-0.917	0.260	-3.53	0.001***	1.27
ΔCPI_Ch_4	5.51	1.60	3.45	0.001***	1.66	ΔCont_Oil_6	1.896	0.460	4.12	0.000***	2.30
ΔOrder_Tank_6	1.176	0.235	5.00	0.000***	1.44						
ΔNew_Tank_2	1.345	0.551	2.44	0.016**	1.25						
ΔSec_Tank_1	0.709	0.265	2.68	0.009***	1.41						
ΔFuel_5	0.747	0.109	6.83	0.000**	1.82						
ΔTC_Tank_2	-0.238	0.111	-2.16	0.033**	1.69						
ΔTC_Tank_6	0.209	0.106	1.96	0.052*	1.56						
ΔOil_1	0.979	0.140	7.00	0.000***	1.95						
ΔFX_US_JP_1'	0.873	0.289	3.02	0.003***	1.43						
ΔFX_US_JP_4'	-0.816	0.275	-2.97	0.004**	1.45						
ΔCont_Oil_3	-1.109	0.370	-3.00	0.003***	1.32						
ΔCont_Oil_6	1.738	0.355	4.90	0.000***	1.39						
ΔP_Metals_1	-0.733	0.229	-3.20	0.002***	1.65						
ΔP_Metals_4	-1.197	0.238	-5.02	0.000***	1.80						
ΔP_Metals_6	-0.434	0.234	-1.86	0.066*	1.71						
ΔTank_Index_6	-0.364	0.122	-2.98	0.004***	1.70						
ΔMSCI_EM_1	-0.554	0.191	-2.90	0.004***	2.18						

Significance level: *** 0.01, ** 0.05, * 0.1.
'ΔFX_USD_JPN_1

6.3.1. Model misspecification test results

Misspecification test results are presented in accordance with the quantitative approach outlined in section 5.6.3.

i) Ramsey's RESET test did not indicate functional misspecifications in any of the dry bulk nor tanker models. Anyhow, we would like to highlight an interesting observation. Both the non-seasonal and seasonal dry bulk models seem to increase their degree of misspecification when more variables are added to the models. One plausible explanation is that the most significant variables are in fact well approximated by the linear relation, whereas among the less significant variables, there are at least some that substantially deviate from the linear form. If the $Alpha_{10}$ models contain a higher share of inherently non-linear variables than the $Alpha_5$'s, which again contain a higher share than the $Alpha_1$'s, the $Alpha_{10}$'s will appear more misspecified overall compared to the leaner $Alpha_5$'s, which again appears more misspecified than the $Alpha_1$'s.

ii) Two general observations from the BG-test results are worth commenting on. First, the results indicate that both BDI_1 and $BDTI_1$ are autocorrelated at a 10% significance level, with respective p-values of 8.53% and 9.97%. P-values in that range are not definite signals of a wrong hypothesis, but should in any case be taken seriously. We inspect the issue by taking a closer look at the ACF-plots (see Appendix G.1). It is evident that the BDI_1- model have significant negative autocorrelation in lag eight. Even more interestingly, the autocorrelation pattern across lags shows a wave-like pattern with a period approximately equal to six months. By looking at the seasonal components obtained from the dry bulk series (see figure 6.1), we observe that the two local minima are six months apart, and similarly that the two local maxima are six months apart as well. Thus, a wave-like pattern with a period of six months is consistent with what we would expect when seasonality is ignored (if the seasonality hypothesis is correct). By examining the ACF belonging to $BDTI_1$, we observe that there is significant negative autocorrelation in lag 4. We might see tendencies to a wave-like shape like in the dry bulk model, though not as apparent. The reason behind this might be the overall shape of the tanker seasonal components, which are less sinusoidal than the dry bulk components. In summary, the fact that two non-seasonal models are autocorrelated, while no seasonally adjusted models are found to be autocorrelated, combined with the fact that we observe a wave-like lag shape in BDI_1 and $BDTI_1$, provide some minor support to the theory of of deterministic seasonality.

Second, all $Alpha_1$ models, regardless of seasonality adjustments and market segment, are found to have a higher degree of autocorrelation than the $Alpha_5$ and $Alpha_{10}$ models. This finding is consistent with the remedies we presented in section 5.6.4, namely that models with fewer variables usually are more prone to be autocorrelated, as they have less ability to be capture possible complexities in the dependent variable.

As the autocorrelation is relatively weak in BDI_1 and $BDTI_1$ (not found to be significant at a 5% level), combined with the fact that ignoring present autocorrelation in a model do not cause severe consequences[22], we will accept both BDI_1 and $BDTI_1$ in

[22] OLS estimators are no longer efficient - it would be possible to narrow the confidence intervals of the coefficients with the same amount of data points. However, with 100+ degrees of freedom, we have reason to believe the test power is already acceptable.

their current state.

iii) White's test reveals no clear evidence of heteroscedasticity in any of the tanker models. Turning to the dry bulk segment, $SBDI_5$ is found to be significant at 10% significance level, with a p-value of 6.18%. Thus, we look at the residual plot (see Figure G.7) to investigate the apparent non-uniform variance. It seems to be larger spread in the residuals generated by negative fitted values relative to positive fitted values. (Sometimes referred to as "funneling shape" in the residuals.). This implies that the coefficients in $SBDI_5$ generally will be more uncertain (have larger standard errors) as covariates take more negative values. With this piece of information in mind, we are aware of the model's weakness. As other misspecification tests seem to validate the model (except normality in residuals, which five out of six dry bulk models rejects at the 5% level), and since the p-value is above 5%, we will not experiment with variable transformations, but rather move forward with the current model.

iv) The Anderson- Darling (AD) statistics claim that the residuals generated by the majority of our dry bulk models are unlikely to originate from a normal distribution. By looking at the normal probability plots in Figure A.5 and A.6, this result should not come as a surprise. Since the majority of our collected variables, especially the dry bulk varaible, clearly are non-normal, mostly due to leptokurtosis, skew, or both, the residuals are unlikely to be normally distributed. This is expected as a linear combination of non-normal distributions in general not will be normal. When we initially transformed each time series to obtain stationary series, the log transformation was the best suited candidate in order to attain normality.[23] Despite the transformation, most of time series are inherently non-normal. It is worth pointing out, that many models seem to have approximately normal residuals within the range of $\pm$ two standard deviations (see Appendix G.2). Yet, the distribution as a whole is often rejected, as negative extreme observations tend to occur too frequent to be compatible with the normal distribution. Since the in-sample size is 150, and the number of parameters estimated is less than 50 in all models, we have 100+ degrees of freedom, which makes the t-distribution relatively robust. Thus, there is no clear reason to mistrust the estimated variable coefficients.

[23]It would have been possible to "tailor" a transformation to every time series, i.e. test several possible transformation for every series and each time select the one that would lead to a distribution most closely resembling the normal. However, the main priority was to reach stationary at at 5% significance level, which we accomplished.

Table 6.7: Dry Bulk regression output

Dry bulk models (Baltic Dry Index)							
Not adjusted for seasonality				**Adjusted for seasonality**			
	$\boldsymbol{Alpha_1}$	$\boldsymbol{Alpha_5}$	$\boldsymbol{Alpha_{10}}$		$\boldsymbol{Alpha_1}$	$\boldsymbol{Alpha_5}$	$\boldsymbol{Alpha_{10}}$
S	0.195	0.115	0.111	S	0.190	0.147	0.0972
R^2	0.330	0.691	0.722	R^2	0.298	0.624	0.822
R^2_{adj}	0.307	0.632	0.658	R^2_{adj}	0.279	0.569	0.755
R^2_{pred}	0.257	0.549	0.542	R^2_{pred}	0.216	0.466	0.637
D-W stat	1.747	2.021	1.943	D-W	1.814	1.735	1.987
RESET (p-stat)	0.857	0.267	0.128	RESET (p-stat)	0.563	0.273	0.112
B-G (p-stat)	0.0853*	0.549	0.763	B-G (p-stat)	0.111	0.722	0.243
White (p-stat)	0.239	0.755	0.258	White (p-stat)	0.721	0.0618*	0.1430
AD (p-stat)	0.008***	0.022**	0.006***	AD (p-stat)	0.007***	0.029**	0.610

Significance level: *** 0.01, ** 0.05, * 0.1.

Table 6.8: Tanker regression output

Tanker models (Baltic Dirty Tanker Index)							
Not adjusted for seasonality				**Adjusted for seasonality**			
	$\boldsymbol{Alpha}_1$	$\boldsymbol{Alpha}_5$	$\boldsymbol{Alpha}_{10}$		$\boldsymbol{Alpha}_1$	$\boldsymbol{Alpha}_5$	$\boldsymbol{Alpha}_{10}$
S	0.183	0.159	0.114	S	0.146	0.132	0.114
R^2	0.147	0.390	0.752	R^2	0.283	0.436	0.609
R^2_{adj}	0.136	0.346	0.668	R^2_{adj}	0.263	0.395	0.545
R^2_{pred}	0.109	0.258	0.552	R^2_{pred}	0.231	0.346	0.468
D-W stat	1.910	1.902	2.283	D-W stat	2.176	2.091	2.079
RESET (p-stat)	0.159	0.382	0.578	RESET (p-stat)	0.573	0.355	0.577
B-G (p-stat)	0.0997*	0.344	0.435	B-G (p-stat)	0.243	0.467	0.835
White (p-stat)	0.364	0.599	0.404	White (p-stat)	0.883	0.879	0.926
AD (p-stat)	0.363	0.816	0.400	AD (p-stat)	0.130	0.163	0.246

Significance level: *** 0.01, ** 0.05, * 0.1.

6.3.2. Model properties

We give a brief summary of some important properties of each model displayed in Table 6.7 and 6.8. Not surprisingly, stricter significance criteria yielded models with fewer explanatory variables. We observe that models with fewer explanatory variables typically have larger standard errors and lower in-sample R^2. Among the dry bulk models, the best fit, denoted by the highest R^2 value of 0.822, belongs to $SBDI_{10}$, closely followed by BDI_{10} with an R^2 of 0.722. These two models also yield highest R^2_{Adj} values, in that same order, suggesting that the $Alpha_{10}$ models tend to explain more variation than the $Alpha_5$'s and $Alpha_1$'s, even taking into account the number of parameters the model. In the tanker segment $BDTI_{10}$ shows the best fit in-sample, with an R^2 of 0.752, followed by $SBDTI_{10}$ with an R^2 of 0.609. The same two models have the highest R^2_{adj}, in that same order. It is also worth pointing out that the R^2_{pred} is lower than R^2 for all models, and the gap between them increases with a slacker significance criteria, hinting to more serious over-fit in larger models.

6.3.3. Explanatory variables

The dry bulk models are displayed in Table 6.1, 6.2 and 6.3. Let's begin with examining the models that are not adjusted for seasonality, which are displayed in the left part of the tables. The BDI_1, BDI_5 and BDI_{10} models include 5, 24 and 28 variables, respectively (excluding the intercept). The most significant variables are Chinese steel production (lagged by two months), the U.S. consumer price index (lagged by three months), the US dollar exchange index (lagged by two months) and two variables (lagged by one and three months) of the composed dry bulk stock index[24]. The dry bulk equity index is significant with a positive sign as anticipated (See Table 4.1 for hypotheses), which is particularly interesting, as it implies that correct future freight rate information is traceable in shipping stock fluctuations. The leading ability of the variable is intuitively meaningful, as shipping company earnings, and hence valuations, are heavily dependent on freight rates. Moreover, as shipping investors have strong incentives, leading knowledge and access to superior sources of information, they are well-positioned to predict future freight rates meticulously. This result is consistent with the research of Westgaard et al. [2017], who found that oil company valuations, specifically changes in the OSX index[25], serves as a leading predictor for oil price changes. The fact that the variable lagged with three months have the opposite sign, could be explained by stock market behaviour; stock prices tend to over-react to both good and bad news and later on correct themselves. The positive coefficient of Chinese steel production is consistent with our hypothesis, and implies that an increase in the variable is associated with positive development of the BDI two months later. Steel is produced from iron ore and coal, two major seaborne dry bulk commodities, of which China is the largest importer (see Figure 2.2 for the largest importers and exporters of the most important bulk cargoes). An increase in the steel production in China (i.e. a positive change in the variable) could reflect increased use and demand of these bulks. Moreover, iron ore and coal are sourced mainly from Australia and Brazil on Capesize vessels, which use around

[24]see Appendix A.3 for a thorough description of the index.

[25]The PHLX Oil Service Sector (OSX) index is designed to track the performance of a set of companies involved in the oil services sector.

one and two months on a round trip, respectively, meaning that a voyage charter would keep a Capesize occupied for a period of up to two months. When increased steel production stimulate higher demand for iron ore and coal, while at the same time relevant vessels are fixed on a round trip taking up to two months, utilisation will increase and prompt an upsurge in freight rates. The next variable is the US-dollar lagged with two months, which, in contrast to our hypothesis, has a negative sign on the β. Even though shipowners typically have their revenues in dollars, there could be at least one adverse effect caused by an uplift in dollar rates. When the dollar is high, the oil-price, which is quoted in dollars, becomes relatively more expensive in other currencies, impairing the demand side in other countries than the US. Though oil is not a dry bulk commodity, the BDI and BDTI are known to be co-integrated [Tvedt, 2003, Veenstra and Franses, 1997], and the oil price is found to be one of the leading indicators of BDTI (discussed in the next section). Moreover, commodity prices are often correlated; for instance the Brent oil price and the steel price (e.g. represented by US HRC) have been strongly correlated the past ten years. Thus, a strong dollar may be a signal of a weak commodity market, and this effect might outweigh the favourable currency/revenue-effect. The next variable is the US CPI-variable, and we observe that the sign is not consistent with our hypothesis. The negative sign implies that a rise in the consumer price level three months ago tend be related to a decrease in dry bulk rates. One possible explanation could be that cheaper goods, proxied by a sliding CPI, could induce higher purchasing power for U.S. consumers, furthermore pouring water to consumers' "spending-mill". In other words, a weaker CPI reflects declining prices, which would inevitably raise U.S. consumers' purchasing power, and hence lead to higher demand for seaborne cargo transport.

Turning to the seasonal adjusted dry bulk models found in the rightmost half of Table 6.1, 6.2 and 6.3, the $SBDI_1$, $SBDI_5$ and $SBDI_{10}$ models comprise 4, 19 and 41 variables (excluding intercept), respectively. The leanest model with a 1% stopping criteria, $SDBI_1$, includes one less variable than its non-seasonal twin. A very interesting result is that the Chinese steel production with two lags is no longer significant. The implication is that there must be commonalities in the seasonal components and in the Chinese steel production variable. This seems very plausible, as China is the largest importer of dry bulks, and, as mentioned above, China use iron ore and coal, two main dry bulks, to produce steel. Thus, we have reason to believe demand fluctuations in China throughout the year is at least partly responsible for the seasonal components in the dry bulk market. Apart from the lack of the Chinese steel-variable, the other $SBDI_1$-variables are found significant in BDI_1 with the same sign.

Moving on to the tanker models shown in Table 6.4, 6.5 and 6.6, we first examine the original BDTI models found in the leftmost part of the tables. The output shows that $BDTI_1$, $BDTI_5$ and $BDTI_{10}$ consist of 2, 10 and 37 variables (excluding intercept), respectively. The significant variables in the $BDTI_1$ model are Chinese oil imports (two months lag) and the Brent crude oil price (one month lag). The latter is consistent with the research of Poulakidas and Joutz [2009], who concluded that there is a lead-lag relationship between tanker rates and the oil price, meaning that when oil prices surge, tanker demand tend to rise, making an opportunity for shipowners to raise rates. It is also consistent with our hypothesised sign, which was based upon the presumption that the oil price is correlated positively with global economic activity (See Hamilton [2005, 2008]), and hence could be a proxy for demand. Chinese oil imports, on the other hand, has the opposite sign. As China is among the largest importers of oil, and oil is the most prominent commodity in the tanker market, this appears to be rather counter-intuitive. How could it be that a current increase in Chinese oil demand induces a reduction in rates on a two month horizon? As this result seems confusing, we have included a specific correlation matrix (See Figure E.8 for correlations between oil prices and Chinese oil imports, including all lags)[26], to inspect the issue further. We observe a fairly strong positive correlation in the first lag, and even stronger positive correlation in the third lag, while the second lag has the strongest correlation albeit associated with a negative sign. Furthermore, all Chinese oil imports are very negatively correlated across consecutive lags, while every second lag have positive correlation. Apparently, the Chinese oil imports time series has a vast negative autocorrelation in the first lag, and consequently positive autocorrelation in the second. Based on these findings, it seems that the fundamental relationship between demand and rates still applies, but is "disturbed" by a second effect, which is the lack of regularity in China's oil importing routines. Unfortunately, we do not have extensive knowledge regarding China's oil import strategy, but it appears as if they import large quantities of oil one month, then build inventories, thus requiring less oil the following month, before the cycle repeats. If that is indeed the case, the variable serves as a proxy of Chinese demand despite the counter-intuitive sign.

For the models fitted to the seasonal component of the BDTI, shown in the right part of Table 6.4, 6.5 and 6.6, the three models consist of 4, 10 and 21 variables ($SBDTI_1$, $SBDTI_5$ and $SBDTI_{10}$, respectively). The most significant variables are the the U.S. CPI (lagged five months), second-hand tanker values (lagged one month), the fuel price (lagged five months) and the oil price (lagged one month). The oil price has the same sign as in the previously discussed model, and is thus covered. The fuel price is interesting, as we initially argued it could possibly have both a positive and negative impact on freight rates. The positive sign could be explained by the strong correlation with the oil price (see Figure E.7 - the correlation between $dFuel\_5$ and d_Oil_5 is 0.75). Nevertheless, some other effects must also be inherent in the fuel variable, otherwise the oil price lagged with five months could equally well have been found significant. Alternatively, the idiosyncrasies in the fuel variable may just have a slightly better fit in-sample than idiosyncrasies of the oil variable. The impact of second-hand tanker values is positive, which is in accordance with our presented hypothesis, i.e. that short term future market expectations are traceable from the second-hand market. The last variable, U.S. CPI, lagged by five months, has the opposite sign of what we anticipated, but the variable is similar to the significance found in the seasonally adjusted dry bulk models. A rise in the CPI, reflecting increasing prices, could inevitably weaken U.S. consumers' purchasing power, and hence cause their demand for oil to fall, thus supporting the negative relationship. Additionally, it must be pointed out, that in a manner similar to the dry bulk models, yet another fundamental China variable (Chinese oil imports) was found significant in the non-seasonal model, while the significance subsides after the seasonal adjustment. When presenting the rationale for the inclusion of seasonal components in Chapter 5, we argued that, for instance, the negative January components could stem from lower Chinese demand, in both the dry bulk and tanker market, as a result of their large-scale New Years's connection. The fact that the Chinese steel production variable in the bulk model and the Chinese oil import variable in the tanker model, are non-

[26] Matrices illustrating correlations between the dependent variable and lags of significant covariates in a respective model are provided in Appendix E. Figure E.5 and E.6 show the BDTI correlation matrix.

significant in their respective seasonally adjusted peer model, supports the hypothesis we carried out.

6.4. Out-of-sample evaluation

This subsection presents out-of-sample model evaluation metrics. At first, we will analyse the degree of over-fit in each model. Next, we present the results obtained from the out-of-sample tests measuring predictive power, and attempt to mutually rank all models and benchmarks based on various metrics. Then, we present the Diebold-Mariano test results and, based on these findings, discuss whether some models deliver significantly better forecasts accuracy than others. Lastly, we present charts illustrating out-of-sample forecasts and squared errors, for all models and relevant benchmarks.

6.4.1. Over-fitting

Test results are displayed in Table 6.9. As mentioned in chapter 5, MSE and MAE reflect, respectively, the mean square error and absolute deviation between the dependent variable and the fitted model. As it is more convenient to compare the two metrics when they are in the same scale, we also provide $RMSE$. Furthermore, as the standard error of the dependent variable may vary considerably in-sample and out-of-sample, $MASE$ and $SRMSE$ are more appropriate metrics, as imbalances caused by different standard errors are taken into account (i.e. allowing us to compare apples with apples).

By looking at the $MASE$ and $SRMSE$ results obtained in-sample, we observe that all models improve as the significance criteria is slacked. This tendency is expected, as a more complex model, i.e. a model with more variables, is more apt to fine-tune itself to in-sample idiosyncratic noise. The big question is how the more complex models will perform relative to the simpler ones out-of-sample. A leaner model is always preferred relative to a complex one if it offers equal or better predictive power, as the leaner model obviously is less likely to contain false predictors, i.e. be over-fitted. In other words, unless the $MASE$ and $SRMSE$ results improve as the significance criteria is relaxed, the $Alpha_5$'s and $Alpha_{10}$'s offer limited value. Interestingly, we observe both $MASE$ and $SRMSE$ worsen in all models as the significance criteria is relaxed. By comparing the ratio of the out-of-sample metrics and the in-sample metrics, it is evident that the $Alpha_{10}$'s are more over-fitted than the $Alpha_5$'s, which again are more over-fitted than the $Alpha_1$'s. It is not unexpected that the contrast between the in-sample fit and out-of-sample fit increase as models grow more complex, but the outcome was not given. If for instance the out-of-sample performance of an $Alpha_5$ model had been comparable to its in-sample performance, it would indicate under-specification in the associated $Alpha_1$ model.

By investigating the frequency of correct direction predicted for the various model, we observe a similar pattern. In-sample, the more complex model the better, while out-of-sample, the simpler is (usually) better or equally good. The tendency is especially clear-cut in the seasonally adjusted tanker models. The only exception is in $BDTI_5$ and $BDTI_{10}$, where the latter has a higher frequency of correct direction predicted than the former. In any case, the spread between the in-sample and out-of-sample correct direction increases with larger complexity for all four model types. Thus, this test metric provides data in accordance with what we found for $MASE$ and $SRMSE$. Also, as mentioned in section 6.3.2, the spread between R^2_{pred} and R^2 increase with model complexity (see Table 6.7 and 6.8). This observation is consistent with the findings from this section.

Table 6.9: Model validation

Dry bulk model validation

	Not adjusted for seasonality							Adjusted for seasonality					
	Alpha$_1$		***Alpha***$_5$		***Alpha***$_{10}$			***Alpha***$_1$		***Alpha***$_5$		***Alpha***$_{10}$	
Metric	In-sample	Out-of-sample	In-sample	Out-of-sample	In-sample	Out-of-sample	**Metric**	In-sample	Out-of-sample	In-sample	Out-of-sample	In-sample	Out-of-sample
MAE	0.137	0.244	0.0880	0.252	0.0822	0.288	MAE	0.141	0.238	0.113	0.285	0.0650	0.284
RMSE	0.178	0.306	0.117	0.323	0.107	0.349	RMSE	0.180	0.301	0.142	0.335	0.0825	0.340
MASE	0.655	0.925	0.422	0.957	0.394	1.09	MASE	0.674	0.903	0.573	1.08	0.331	1.08
SRMSE	0.851	1.16	0.559	1.22	0.512	1.32	SRMSE	0.861	1.14	0.679	1.27	0.394	1.29
Correct Direction	0.640	0.500	0.773	0.500	0.813	0.500	Correct Direction	0.607	0.524	0.713	0.524	0.813	0.452

Tanker model validation

	Not adjusted for seasonality							Adjusted for seasonality					
	Alpha$_1$		***Alpha***$_5$		***Alpha***$_{10}$			***Alpha***$_1$		***Alpha***$_5$		***Alpha***$_{10}$	
Metric	In-sample	Out-of-sample	In-sample	Out-of-sample	In-sample	Out-of-sample	**Metric**	In-sample	Out-of-sample	In-sample	Out-of-sample	In-sample	Out-of-sample
MAE	0.133	0.1085	0.120	0.1262	0.0765	0.2082	MAE	0.108	0.1066	0.0980	0.1204	0.0843	0.1723
RMSE	0.1808	0.1405	0.1530	0.1581	0.09798	0.2534	RMSE	0.1432	0.1395	0.1269	0.1571	0.1054	0.2238
MASE	0.676	0.7889	0.608	0.9171	0.389	1.513	MASE	0.639	0.7750	0.578	0.8749	0.498	1.252
SRMSE	0.920	1.02	0.778	1.15	0.498	1.84	SRMSE	0.915	1.01	0.727	1.14	0.419	1.63
Correct Direction	0.620	0.619	0.700	0.500	0.833	0.571	Correct Direction	0.673	0.667	0.687	0.619	0.760	0.5476

The in-sanple and out-of sample periods correspond to July 2000 to December 2012 and January 2013 to June 2016, respectively.
Test values are calculated out-of-sample with models that are fitted to the respective dependent variable using a rolling window of 150 observations (corresponding to the in-sample window size), where the coefficients are held fixed.

6.4.2. *Predictive power*

Table 6.10 shows the out-of-sample test results for all models and their relevant benchmarks. Here, the models' predictive power is evaluated, both in terms of accuracy (MAE etc.) and correlation[27]. As correlation is the square root of the R^2, it similarly measures how much variation in the dependent variable the model is able to explain. The two ARIMA models are respectively a random walk model (hereafter referred to as RW^k, where k=$\{BDI, BDTI\}$) and an in-sample optimised ARIMA model (hereafter ISO^k). Additionally, we include the seasonal components as benchmarks (hereafter SC^k), as they, similarly to the other benchmarks, are univariate models. However, SC^k should not only be regarded as benchmarks, but also as two of our seasonal models - the simplest possible seasonal models. It is also worth pointing out that rank of models are similar in terms of MAE and $MASE$, as they are all adjusted for the same standard deviation. The same is true for $RMSE$ and $SRMSE$. The reason why $MASE$ and $SRMSE$ are included in this section, is that they may be used to compare errors across markets. For instance, the errors between BDI_1 and $BDTI_1$ are more correctly compared based on these metrics.

By studying dry bulk models, we observe that the all models are ranked identically in terms of MAE and $RMSE$, except $SBDI_5$ and $SBDI_{10}$. In terms of $RMSE$, the $Alpha_1$'s are superior to the 5's, which again are superior to the 10's, while $SBDTI_{10}$ outperforms $SBDTI_5$ in terms of MAE. Further, we see that $SBDI_1$ beats BDI_1 and $SBDI_{10}$ beats BDI_{10}, while BDI_5 beats $BDTI_5$, so two out of three seasonally adjusted models outperform their non-seasonal twins. The SC^{BDI} delivers best predictive accuracy, followed by ISO^{BDI}. The RW^{BDI} delivers poor results, which is not surprising, taking the vast fluctuations in ΔBDI into account. Regarding correlation, SC^{BDI} again takes the lead and RW^{BDI} is ranked last, while the rank of the other models differ from their accuracy rank. Among the non-seasonal models, the more complex models are able to explain somewhat more of the variation than the simpler models. Among the seasonally adjusted models, the $SBDI_5$ explains more variation than $SBDI_{10}$, which again explain more than $SBDI_1$. If the marginal variables found significant in the more complex were purely over-fitted, we would probably have observed a decrease in the correlation. The fact that the correlation increases somewhat signals that at least some of the added variables are not false predictors. It is also worth pointing out that all three seasonally adjusted models have higher correlation than the non-seasonal models. Regarding the $\%CD$- metric, the results are very uniform; eight out of nine models predicts correct direction either 22 or 21 times out of 42.

Among the tanker models, we see that best predictive accuracy is delivered by SC^{BDTI}. In terms of MAE is $SBDTI_1$ the second best, closely followed by ISO^{BDTI}. However, they appear in the reversed order in terms of $RMSE$. We observe that the $Alpha_1$'s are superior to the 5's, which again are superior to the 10's, both in terms of MAE and $RMSE$. We also see that three out of three seasonal models beat their non-seasonal twins, both in terms of MAE and $RMSE$. When it comes to the correlation test, the best result is attained by $SBDTI_1$, closely followed by $SBDTI_5$ and SC^{BDTI}. Interestingly, all three seasonal models outperform their non-seasonal peers in terms of

[27] The $\%CD$ metric is neither a pure accuracy nor correlation test metric in a strict sense, but does in any case provide complementary information regarding predictive power.

correlation as well. When it comes to the $\%CD$- metric, the $SBDTI_1$-model again comes out ahead, while $SBDTI_5$, SC^{BDTI} and $BDTI_1$ share the second place.

Table 6.10: Model predictability evalution

Dry bulk model evaluation

	MAE	*Rank*	RMSE	*Rank*	%CD	*Rank*	MASE	*Rank*	SRMSE	*Rank*	Correlation	*Rank*
Not adjusted for seasonality												
Alpha1	0.244	4	0.3056	4	0.500	4	0.9248	4	1.16	4	0.0394	8
Alpha5	0.252	5	0.3328	5	0.500	4	0.9566	5	1.22	5	0.0430	7
Alpha10	0.2877	8	0.3490	8	0.500	4	1.091	8	1.32	8	0.0521	6
Adjusted for seasonality												
Alpha1	0.238	3	0.3015	3	0.524	1	0.9029	3	1.14	3	0.0746	5
Alpha5	0.285	7	0.3351	6	0.524	1	1.082	7	1.27	6	0.225	2
Alpha10	0.284	6	0.3406	7	0.452	9	1.077	6	1.29	7	0.104	4
Benchmarks												
ARIMA(0,1,0)	0.305	9	0.3886	9	0.500	4	1.16	9	1.47	9	-0.0780	9
ARIMA(3,0,1)	0.214	2	0.2608	2	0.524	1	0.814	2	0.989	2	0.117	3
SeasonalComponents	0.203	1	0.2532	1	0.500	4	0.769	1	0.960	1	0.239	1

Tanker model evaluation

	MAE	*Rank*	RMSE	*Rank*	%CD	*Rank*	MASE	*Rank*	SRMSE	*Rank*	Correlation	*Rank*
Not adjusted for seasonality												
Alpha1	0.1085	4	0.1405	4	0.619	2	0.789	4	1.02	4	0.255	4
Alpha5	0.1262	6	0.1581	6	0.500	7	0.9171	6	1.14	5	0.233	5
Alpha10	0.2082	9	0.2533	9	0.571	5	1.513	9	1.84	9	0.0980	7
Adjusted for seasonality												
Alpha1	0.1066	2	0.1395	3	0.667	1	0.7730	2	1.01	3	0.4284	1
Alpha5	0.1204	5	0.1571	5	0.619	2	0.8749	5	1.14	5	0.4058	2
Alpha10	0.1723	8	0.2238	8	0.548	6	1.252	8	1.63	7	0.1774	6
Benchmarks												
ARIMA(0,1,0)	0.145	7	0.190	7	0.500	7	1.056	7	1.83	8	0.0183	8
ARIMA(3,0,2)	0.1070	3	0.1382	2	0.452	9	0.7750	3	1.00	2	-0.0699	9
SeasonalComponents	0.1060	1	0.1304	1	0.619	2	0.7680	1	0.948	1	0.399	3

The rank provides an overview of the test scores in comparison to other models' score. A rank value of 1 corresponds to superiority in the respective test.

We formulated our OLS estimators such that they minimise MSE, and our loss function is consequently sum of squared residuals, i.e. of second order. In order to have consistency in the loss function, we will use squared residuals in the Diebold-Mariano test (see Appendix C.9 for the DM-test formulation).

The matrix containing all DM-test outcomes is displayed in Table 6.11. First, we inspect the dry bulk results. We observe that the model ranks are consistent with their $RMSE$ ranks, which is expected, as second order residuals were used in the test. Only positive terms in the rightmost column indicates that SC^{BDI} indeed beats all other models in terms of accuracy. It is found to be significantly superior to all models with a significance level of 10% or lower, except ISO^{BDI}, which ranked as number two on the $RMSE$ metric. Similarly, RW^{BDI} is found to be the weakest model, reflected by entirely negative signs in its respective column. It is found to be significantly weaker than the top four models at a 5% significance level. Even though $SBDI_1$ is the most accurate regression model, it is not found to be significantly different from any other regression model. In fact, no dry bulk regression models are found to be significantly superior or inferior to each other.

Moving to the tanker models, we similarly observe that the models' rank are in accordance with their $RMSE$ ranks. The best model, SC^{BDTI}, is significantly superior to both $Alpha_{10}$'s and RW^{BDTI}. The weakest model, $BDTI_{10}$, is significantly inferior to six out of eight models at a 1% significance level. This is not surprising, taking into account that the marginal $RMSE$ gap from the top six models to $BDTI_{10}$ is roughly 60%. The most accurate regression model, $SBDTI_1$, is not significantly inferior to any benchmark, while it is significantly superior to three regression models with higher complexity in addition to RW^{BDTI}, at a 10% significance level or lower. All significant superiorities among the regression models involve a weak $Alpha_{10}$, either seasonally adjusted or not, except that $SBDTI_1$ is superior to $SBDTI_5$ at a 10% level.

Some interesting findings from this section are important to shed light upon. To measure predictive accuracy, we earlier presented two main types of loss functions, namely the MAE and $RMSE$. Most models obtain fairly equal rank on both metrics. The best model in terms of accuracy in both the dry bulk and tanker market is SC^k, closely followed by ISO^k and $SBDI_1/SBDTI_1$. More complex models are clearly inferior in terms of accuracy. The models that yield high accuracy have one commonality - they have small "amplitudes"[28]. In fact, both ISO^{BDI} and ISO^{BDTI} very closely resemble a straight line, i.e. their "amplitudes" are way smaller than our prediction models. If we had included a straight line as a benchmark, it would probably yield accuracy results not differing too much from the ISO's. However, its associated correlation score would be 0, as it explains no variance in the dependent variable at all. This is why the correlation test is a very useful complimentary test - it measures a completely different quality of the forecast. Therefore, it is particularly interesting that the SC^k, despite being univariate models, are able to provide some of the best correlations *and* best predictive accuracy - in both markets! As mentioned in chapter 5, predictive power is based on a total evaluation of accuracy and correlation. Thus, SC^{BDI} is indeed the dry bulk model with best predictive power. Among the tanker models, SC^{BDTI} has a slightly better accuracy than $SBDTI_1$ and IOS^{BDTI}, while $SBDTI_1$ explains slightly

[28]We do not refer to amplitude in a strict mathematical sense, but rather as a convenient way to describe the average height on the monthly forecast graph, see Figure 6.2

more variation than SC^{BDTI}, and way more than IOS^{BDTI} (in fact, since IOS^{BDTI} has negative correlation, it would have been beaten by a straight line). Thus, based on a total evaluation, we cannot unambiguously conclude whether $SBDTI_1$ or SC^{BDTI} has best predictive power.

Table 6.11: Diebold-Mariano test results

Dry bulk Diebold-Mariano test results

	Not adjusted for seasonality			**Adjusted for seasonality**					
	Alpha $_1$	***Alpha*** $_5$	***Alpha*** $_{10}$	***Alpha*** $_1$	***Alpha*** $_5$	***Alpha*** $_{10}$	***ARIMA(0,1,0)***	***ARIMA(3,0,1)***	***Seasonal Comps.***
Not adjusted for seasonality									
Alpha $_1$	-	-0.938	-1.472	0.325	-1.036	-1.302	-1.983**	1.554	1.874*
Alpha $_5$	0.938	-	-1.226	0.971	-0.108	-0.406	-1.094	1.793*	2.275**
Alpha $_{10}$	1.472	1.226	-	1.441	0.612	0.414	-0.766	2.294**	2.800***
Adjusted for seasonality									
Alpha $_1$	-0.325	-0.971	-1.441	-	-1.076	-1.324	-2.143**	1.471	1.833*
Alpha $_5$	1.036	0.108	-0.612	1.076	-	-0.201	-1.111	2.136**	2.656***
Alpha $_{10}$	1.302	0.406	-0.414	1.324	0.201	-	-0.985	2.250	2.634*
Benchmarks									
ARIMA(0,1,0)	1.983*	1.094	0.766	2.143***	1.111	0.985	-	3.232***	2.962***
ARIMA(3,0,1)	-1.554	-1.793	-2.294**	-1.471	-2.136**	-2.250**	-3.232***	-	0.879
Seasonal Comps.	-1.874*	-2.275**	-2.800***	-1.833*	-2.656***	-2.634***	-2.962***	-0.879	-

Tanker Diebold-Mariano test results

	Not adjusted for seasonality			**Adjusted for seasonality**					
	Alpha $_1$	***Alpha*** $_5$	***Alpha*** $_{10}$	***Alpha*** $_1$	***Alpha*** $_5$	***Alpha*** $_{10}$	***ARIMA(0,1,0)***	***ARIMA(3,0,2)***	***Seasonal Comps***
Not adjusted for seasonality									
Alpha $_1$	-	-1.380	-3.250***	0.068	-0.833	-2.170**	-2.620***	0.204	0.672
Alpha $_5$	1.380	-	-3.350***	1.080	0.053	-1.820*	-1.470	0.985	1.270
Alpha $_{10}$	3.250***	3.350***	-	3.300***	3.090***	0.805	1.780*	2.930***	3.140***
Adjusted for seasonality									
Alpha $_1$	-0.068	-1.080	-3.300***	-	-1.820*	-2.730***	-2.220**	0.094	0.720
Alpha $_5$	0.833	-0.053	-3.09***	1.820*	-	-2.330**	-1.330	0.796	1.470
Alpha $_{10}$	2.170**	1.820*	-0.805	2.730***	2.330**	-	0.900	2.200**	2.470**
Benchmarks									
ARIMA(0,1,0)	2.620***	1.470	-1.780*	2.220**	1.330	-0.900	-	2.610***	2.770***
ARIMA(3,0,2)	-0.204	-0.985	-2.930***	-0.0944	-0.796	-2.200**	-2.610***	-	0.565
Seasonal Comps.	-0.672	-1.270	-3.140***	-0.720	-1.470	-2.470**	-2.770***	-0.565	-

Significance level: *** 0.01, ** 0.05, * 0.1.

In the Diebold-Mariano test, the forecast accuracy of model j is compared with the accuracy provided by model i, where i is the row index and j the column index of the matrix. The test statistic associated with entry (i, j) represents the test outcome. If (i, j) is positive, then the model associated with column j is superior, while a negative sign imply the opposite. Note that the diagonal serve as a symmetry-line, except for the change of sign.

6.4.3. Final discussion of results

Graphs illustrating various characteristics of all 12 models and their corresponding benchmarks are displayed in Figure 6.2 through 6.5. Figure 6.2 displays monthly out-of-sample forecasts, Figure 6.3 accumulated monthly out-of-sample forecasts, Figure 6.4 squared forecasting errors, and Figure 6.5 accumulated squared forecasting errors. In this section, we will first address whether the seasonally adjusted models outperforms their non-seasonal twins or not. As the above-mentioned graphs are quite dense, we have provided a section in Appendix F, designated to clearly compare all six twin-forecasts.[29] Furthermore, we will discuss why our tanker models consistently underestimate the BDTI-index. Lastly, we discuss whether we indeed have found some good predictors for the BDI and BDTI index.

In section 5.7.2, we briefly discussed why a forecast should be measured in terms of predictive accuracy and correlation. The two measures captures completely different qualities, like the flat line example provided in section 6.4.2. It is possible to have good accuracy and poor correlation, like the flat line, but it is also possible to have a correlation of 1 accompanied with poor accuracy, for instance if the shape of the forecast perfectly mimics the shape of the dependent variable, but its "amplitude" is much larger. In order to graphically evaluate the accuracy of two models, the most convenient way is by looking at the accumulated squared forecasting graphs, as presented in Figure 6.5 and Appendix F. Here, we can track the cumulative loss function for every time step. If the model is significantly superior, it should reside below its peer model for most or all time steps. For instance, the fact that SC^{BDI} outperforms BDI_{10} at a 1% significance level according to the DM-test, is very consistent with the upper-left graphs in Figure 6.5. The preferable way to graphically evaluate the correlation of a model, is by comparing the shape of the forecast and the dependent variable either in terms of monthly out-of-sample forecasts or accumulated out-of-sample forecasts. In Appendix F we have done the latter, as the development in absolute terms is more straight forward to interpret.

First, let's investigate the impact of seasonal adjustments in the dry bulk models. By looking at the accumulated squared forecasting errors in F.1, F.2 and F.3, we observe that it is fairly close race between all three pair of twins. Among the $Alpha_1$-models, the seasonally adjusted model comes out ahead, but with very little margin. Also, if a different sample window had been selected, it is not obvious to guess the winning model, correspondingly reflected by the lack of significance in the DM-test. If, for instance, the test period had ended early 2016, BDI_1 would have come out ahead. A similar inspection of F.2 and F.3 support the DM-test results. Thus, we cannot necessarily draw the conclusion that a seasonal adjustment improve the forecast accuracy of the dry bulk models. When it comes to correlation, a characteristic feature of all our dry bulk models, is that they have low correlation in general. For instance, by studying the upper half of Figure 6.2, we observe that none of the regression models are able to predict the peak in late 2014. The best correlation (0.225) is attained by $SBDI_5$, and the positive correlation can be observed by studying Figure F.2, where the development in $SBDI_5$ apparently tracks the dry bulk index relatively well, especially from mid-2014 to mid-2016. The correlation of its non-seasonal twin is clearly lower, as its fluctuations seem to be slightly more "out of sync". Among the $Alpha_1$ models, we similarly observe a slight

[29] For the sake of clearness, a twin model refers to a model with identical termination criteria and market segment, i.e. BDI_1 and $SBDI_1$ are twin models.

improvement in correlation after the seasonal adjustment. The same tendency is apparent in the $Alpha_{10}$'s, where the seasonal model also have a slightly better correlation. Despite having a somewhat better correlation than its peer, $SBDTI_{10}$ also seem to be over-fitted, for instance illustrated by the erroneously predicted increase during spring 2015. In summary, the seasonal adjustment of dry bulk models seem to induce an increase the forecast correlation, but there seem to be no notably increase in accuracy.

Next, we investigate the seasonal adjustments of the tanker models. The accuracy of the $Alpha_1$ models are compared in Figure F.4, and it is apparent that the seasonally adjusted model is somewhat stronger than its peer. Also, the victory appears to be somewhat more robust than what we observed among the dry bulk models, as any test window extending beyond autumn of 2014 would yield the same winner. Among the $Alpha_5$'s (see Figure F.5), the seasonally adjusted model also draw the longest straw, though with a tiny margin. Likewise, the victory is not as robust as in the aforementioned case. Out of the $Alpha_{10}$'s (see Figure F.6), the accuracy of the seasonally adjusted model is notably better, but not sufficiently to make the DM-test significant. Overall, the spread in model accuracy between the pair of twins seem to increase somewhat relative to what we observed among the dry bulk twins, but not sufficiently to bring significant spreads. In terms of correlation, the results are interesting. By comparing the $Alpha_1$-models, we observe that the seasonally adjusted twin is able to explain the BDTI-variations to a much larger extent than its twin. While $BDTI_1$ predicts almost a flat BDTI during the first 20 months, $SBDTI_1$ tracks the BDTI development much closer. Also, the variations in the latter half of the out-of-sample window is better explained by $SBDTI_1$. Among the $Alpha_5$'s, the same tendency is apparent; $BDTI_5$ is more or less a straight line until the second half of 2014, then gradually diminishes, while $SBDTI_5$ tracks the small trends in BDTI relatively good. Moving to the $Alpha_{10}$'s, we see that both models are highly over-fitted, for example illustrated by the wrong predictions during the first half of 2013. But regardless of over-fit, $SBDTI_{10}$ tracks the correct BDTI fluctuations much more successfully than its non-seasonal peer. In summary, the seasonal adjustment of the tanker models seem to cause small, but not significant improvements in accuracy, while the correlations seem to increase substantially.

We now investigate the three $BDTI$ models and specifically their apparent bear-ish predictions. The reason why $BDTI_1$ provides a bear-ish forecast is obvious; it consists of only two variables, and one of them is the oil price with a positive sign. The oil price was stable in the range 80-120 dollars per barrel between early 2010 to the summer of 2014, but as a response to the supply boost from U.S. shale-oil, the price dropped considerably during the second half of 2014 and slid further into 2015, before experiencing a slight recovery during 2016 [Arctic, 2017]. When it comes to $BDTI_5$, the model consists of 10 variables. As the oil price appears to be the dominant factor in this model as well, the other marginal explanatory variables either take small values or cancel each other out. The sole oil price sensitivity of $BDTI_5$ remain approximately equal, because its marginal variables relative to $BDTI_1$ as a whole neither are particularly positively or negatively correlated with the oil price (See Appendix E for correlations). The sensitivity to the oil price appears to be a lot more diluted in $BDTI_{10}$, as the model does not predict a constant BDTI during first 18-20 months. This is not surprising, taking into account that the number of variables abruptly increase from 10 to 37. Also the seasonally adjusted tanker models appears to be bear-ish. The fact the all $SBDTI$'s contain the same oil price variable, the bear-is prediction explanation applies to all of them.

In order to evaluate whether we have found some good predictors for the BDI and BDTI, we wrap up the discussion by examining all relevant elements we have discussed so far. There are at least two different ways of evaluating whether the models are good predictors or not. One way is through the path of false significance. The fact that our complex models are subject to a greater degree of over-fitting (e.g. $Alpha_{10}$'s exhibit more over-fit than $Alpha_1$'s), combined with our empirical results showing that our complex models do not display superior predictability compared to the leaner ones, is indeed suggesting that we should favour the leaner $Alpha_1$'s over the 5's and 10's. The other way is through further analysis of the out-of-sample results. We have seen that all $Alpha_1$ models outperform the more complex ones in terms of accuracy in both markets. Among the tanker models, the $Alpha_1$'s also achieves best correlation, and as we have already established, the seasonally adjusted twin is preferred among the $Alpha_1$'s. Thus, in the tanker market, the $SBDTI_1$ is superior among our constructed models. In the dry bulk market, $SBDI_5$ has a substantially higher correlation than $SBDI_1$, so there might be some good predictors entering $SBDI$ after the significance criteria is slacked to 5%. Among the BDI models, we also see that the correlation increases marginally with complexity. However, if we had calculated R^2_{adj} or any other metric measuring explained variance, but at the same time taking the number of variables into account, we would have seen that the marginal increase in correlation would not justify the large number of variables entering. Therefore, we disregard BDI_5 and BDI_{10}. Furthermore, as $SBDI_1$ marginally beats BDI_1 in terms of both accuracy and correlation, we disregard all models except $SBDI_1$ and $SBDI_5$ in the dry bulk market. In summary, we can say that among our initial 264 dry bulk explanatory variables, we have reason to believe that the subset of best predictors is significant in either $SBDI_1$ (4 variables) or $SBDI_5$ (19 variables). Similarly, among the initial 222 tanker explanatory variables, we have reason to believe the best subset of predictors should be the ones found significant in $SBDTI_1$ (4 variables).

The next step is to establish what a *good* predictor is. One way to rank a set of potential predictors is by comparing them to a set of other potential predictors, as we just did. Another way is to classify a predictor is based on its performance relative to a benchmark. The relevant benchmark is the best benchmark, which in our analysis is the univariate SC^k, applicable to both dry bulk and tanker models. None of our models beats it in terms of accuracy, though it should be pointed out that $SBDTI_1$ has marginally better correlation. Hypothetically, if one of our regression models had delivered forecast accuracy that was significantly better than it (ideally combined with high correlation), we could have claimed, beyond reasonable doubt, that we had found a set of good predictors. However, as this is not the case, we cannot claim that we have found a "superior" set of predictors. Despite this fact, our best forecast models are not useless, as they provide complementary information about the expected future movement of BDI/ BDTI, as they have fairly good correlation. Thus, they can be used in conjunction with other forecasts/methods. In summary, the SC^{BDI} is the superior dry bulk model. It would give the best utility, because it is the most accurate, and give the best guidance in terms of future price movement, because it has highest correlation. The best tanker model in terms of utility is SC^{BDTI}, while $SBDI_1$ provides best guidance of future price movement.

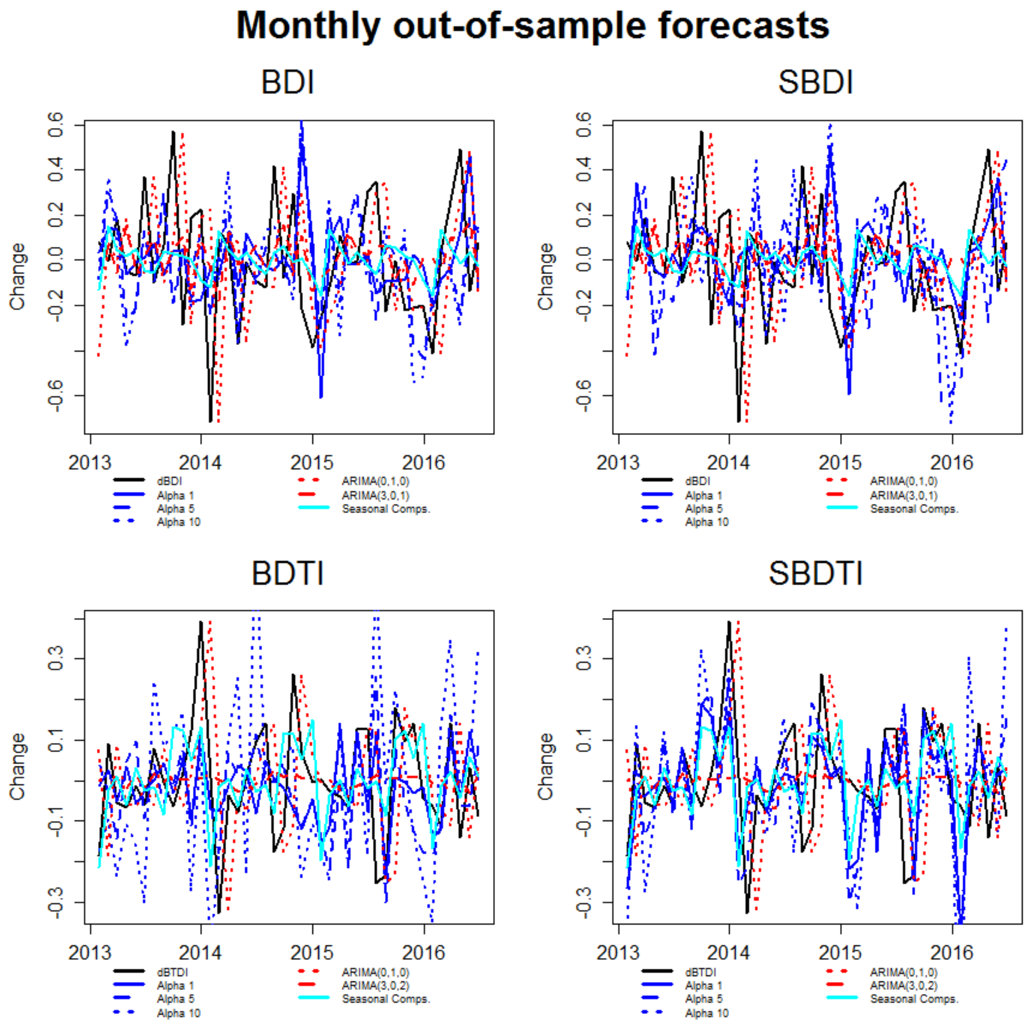

Figure 6.2: Monthly out-of-sample forecasts. The plots shows monthly point forecasts produced by the Alpha1, Alpha5, Alpha10 and benchmark ARIMA models. We note that the forecasts are calculated with models that are fitted to the respective dependent variable using a rolling window of 150 observations (corresponding to the in-sample window size), where the coefficients are held fixed.

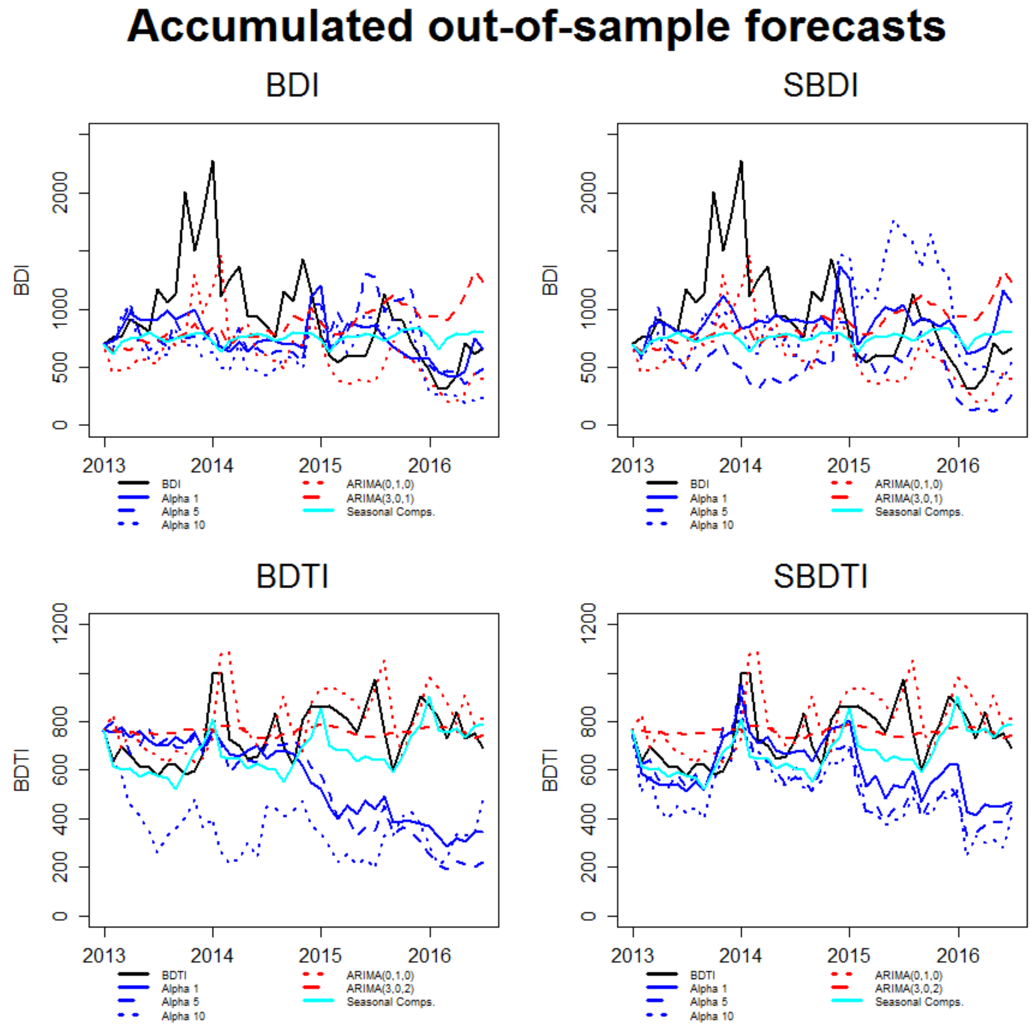

Figure 6.3: Accumulated out-of-sample monthly point forecasts. The plots show freight rate development for Alpha1, Alpha5, Alpha10 and benchmark ARIMA models, generated using the monthly point forecasts in Figure 6.2.

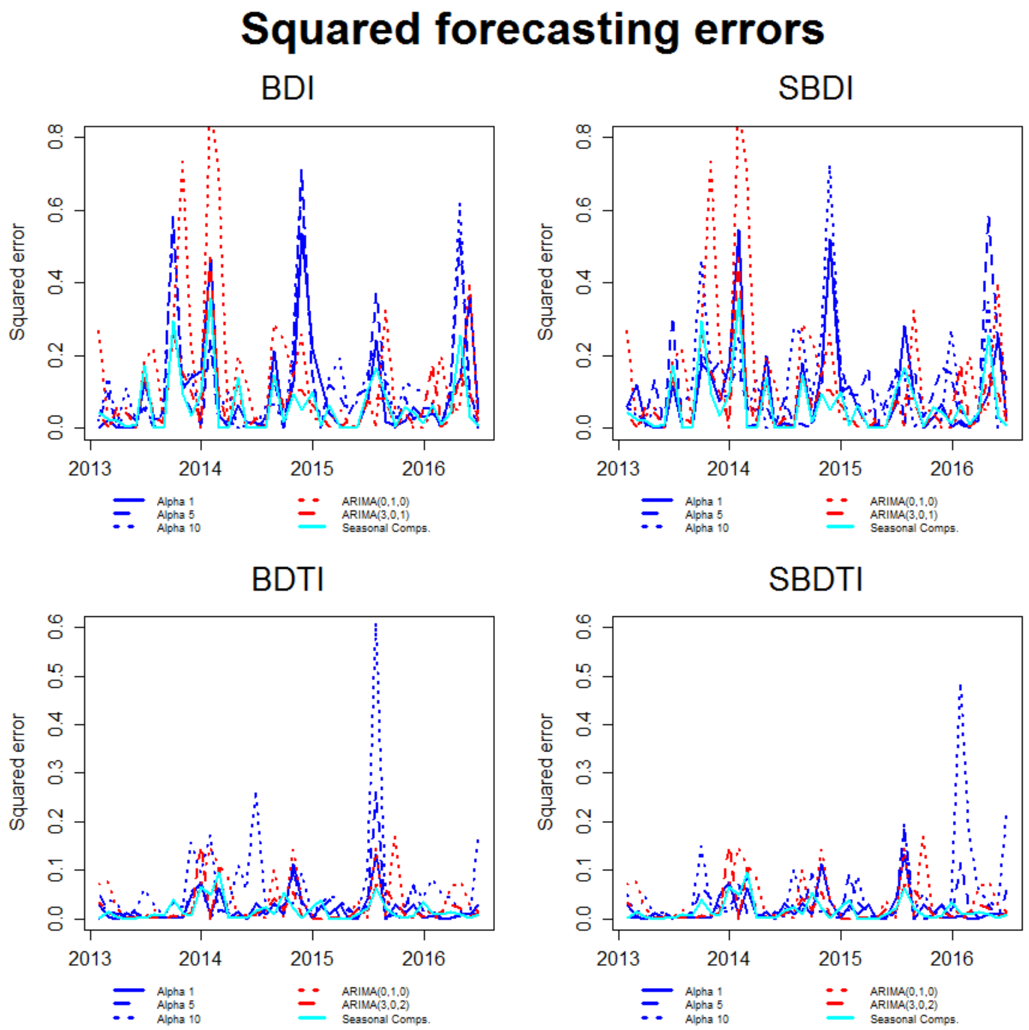

Figure 6.4: Squared out-of-sample forecasting errors. The plots show forecasting error generated for the Alpha1, Alpha5 and Alpha10 models along with the ARIMA benchmark forecasts.

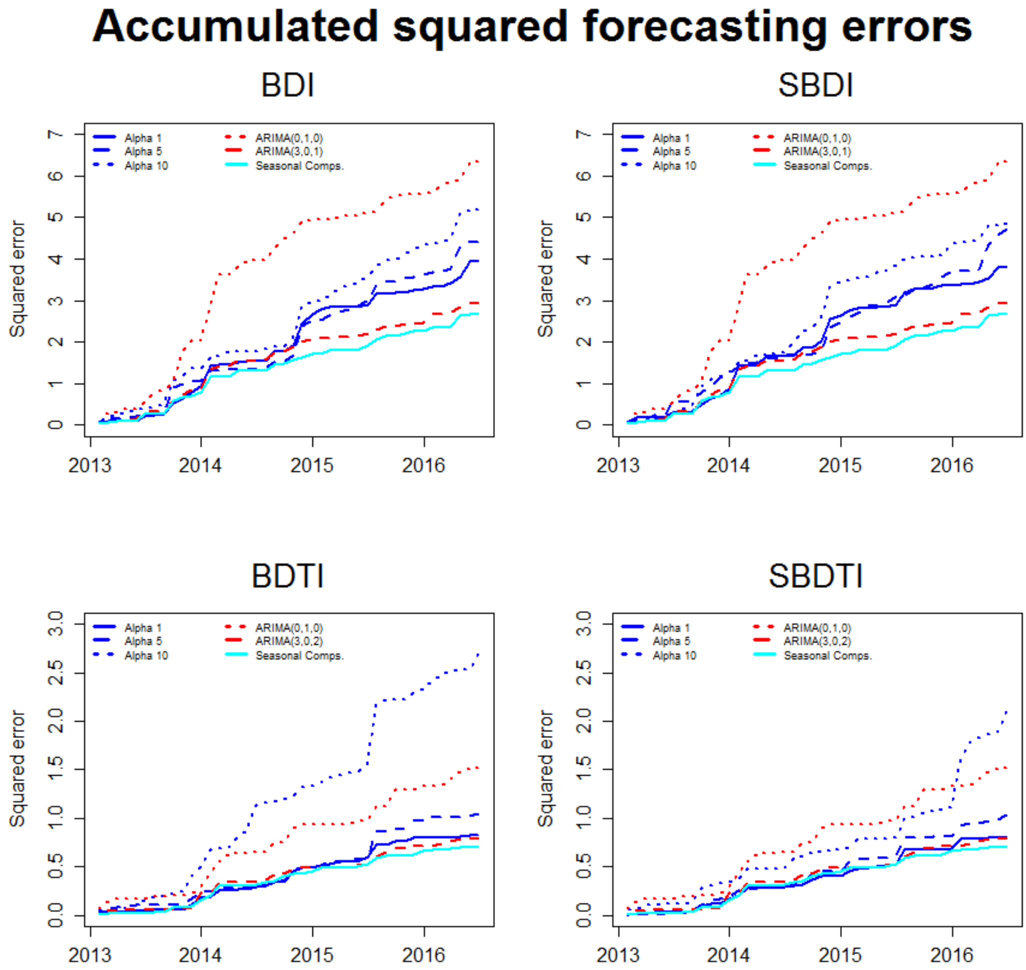

Figure 6.5: Accumulated out-of-sample forecasting errors. The plots show squared forecasting errors for Alpha1, Alpha5, Alpha10 and benchmark ARIMA models, generated from squared forecasting errors in Figure 6.4.

7 Conclusion

In accordance with the general-to-specific methodology outlined by Campos et al. [2005], we present an analysis of a large set of fundamental and non-fundamental indicators for bulk shipping freight rates. Our data set is motivated by a comprehensive study of literature and by general perceptions in the financial markets. We construct 12 pure forecasting models, six dry bulk and six tanker models, characterised by unique combinations of three different model selection criteria and whether the dependent variable is seasonally adjusted or not.

We find that i) The most significant dry bulk freight rate predictors in the period Jul-2000 to Dec-2012 are Chinese steel production, U.S. consumer price index, a dollar exchange rate index and two variables of a dry bulk equity index, which we formulate and introduce in this thesis. The latter is particularly interesting, as no research, to our knowledge, has analysed the stock market's ability to predict shipping rates prior to our analysis. The relationship implies that correct future freight rate information is traceable from shipping stock fluctuations. As shipping stocks are sensitive to freight rates and priced by the collective view of investors exposed to the shipping market - investors who possess leading market expertise and have a strong incentive to collect correct market information, the relationship is is intuitively meaningful. Also interestingly, the finding is consistent with the research of Westgaard et al. [2017], who recently found that the OSX index serves as significant leading predictor for oil prices. Turning to tanker rates, we find that Chinese oil imports, oil prices, U.S. consumer price index, second-hand tanker values and fuel prices are the most significant leading indicators. The positive relationship with lagged oil prices is consistent with the research of Poulakidas and Joutz [2009], who found that an increase in oil prices could reflect increased demand for the commodity, furthermore lifting demand for transportation and thus increase rates. It also supports the conclusion of Hamilton [2005, 2008], who found a positive correlation between oil prices and global economic growth.

ii) The seasonal components, which initially were defined in order to specify the seasonal adjustment, give superior predictive accuracy in both the dry bulk and tanker market. For dry bulk rates, the Diebold-Mariano test concludes that the seasonal components are significantly superior in terms of accuracy, to all regression models on a 10% significance level or lower. In the tanker market, the seasonal components are found to be significantly superior to only the two most complex models, on a 5% significance level or lower. Also, the seasonal components deliver best correlation in the dry bulk market, while a seasonally adjusted regression model delivers best correlation in the tanker market.

iii) The conclusions of Kavussanos and Alizadeh-M [2001, 2002] still apply in the post-millennial shipping market. In particular, we find three aspects that support their conclusion. First, the univariate seasonal component model provides best predictive accuracy in both markets, as mentioned in (ii). Second, all "twin" models that are adjusted for seasonality beat their original peer in terms of correlation, while they deliver equal or better predictive accuracy. Third, the non-seasonal models are slightly more autocorrelated than their respective seasonal "twin", which could indicate that there is a simple repeating pattern in the dependent variable they do not capture (i.e. that the freight rates exhibit seasonality). It is worth mentioning that the seasonal effects appear to be slightly more pronounced in the tanker market, as the tanker models (including the tanker seasonal components) achieve higher correlation and superior accuracy (in terms of $MASE$ and scaled $RMSE$, when compared across markets) relative to the dry bulk models (including the dry bulk seasonal components).

Our findings are useful for shipping market participants' operational (i.e. short term) decision making. By acting upon the model with best predictive accuracy in their respective market, they may improve their utility[30]. Additionally, the model providing best correlation in their respective market, may be used in conjunction, as it gives an indication of the future price movement.

[30]We formulated our OLS estimators such that they minimise MSE, as it is common to maximise utility by minimising a loss function of second order.

Appendix A Data

Appendix A.1. Data description and sources

Table A.1: The table shows an overview of the model's input data along with respective sources.

Variable	**Description**	**Source**
BDI	Baltic Dry Bulk Index	Baltic Exchange/Bloomberg
BDTI	Baltic Dirty Tanker Index	Baltic Exchange/Bloomberg
GDP_W	Weighted geometric mean of world GDP	EIA/Quandl
IP_OECD	Industrial Production, OECD	OECD
IP_Ch	Industrial Production, China	National Bureau of Statistics of China
IP_I	Industrial Production, India	World Bank Group
IP_US	Industrial Production, U.S.	Federal Reserve
Oil_P_G	Oil production, Global	Clarkson Research
Oil_P_ME	Oil production, Middle East	Clarkson Research
IO_Ch_Imp	Iron ore imports, China	Clarkson Research
IO_B_Exp	Iron ore exports, Brazil	Clarkson Research
IO_A_Exp	Iron ore exports, Australia	Clarkson Research
C_EU_Imp	Coal Imports, EU-25	Clarkson Research
C_J_Imp	Coal imports, Japan	Clarkson Research
C_A_Exp	Coal exports, Australia	Clarkson Research
G_US_Exp	Grain exports, U.S.	Clarkson Research
S_Ch_Prod	Steel production, China	Clarkson Research
O_Ch_Imp	Crude oil imports, China	Clarkson Research
O_US_Exp	Crude oil exports, U.S.	Clarkson Research
O_AG_Exp	Crude oil exports, Arabian Gulf	Clarkson Research
S_US_Imp	Steel imports, U.S.	Clarkson Research
MS_US	Money supply, U.S.	Federal Reserve
MS_Ch	Money supply, China	The People's Bank of China
CPI_US	Consumer price index, U.S.	Bureau of Labor Statistics
CPI_Ch	Consumer price index, China	National Bureau of Statistics of China
Fleet	Global fleet in dwt	Clarkson Research
Order	Orderbook dwt in percent of total fleet	Clarkson Research
Scrap	Demolition, dwt	Clarkson Research
Del	Deliveries, dwt	Clarkson Research
New	Newbuild price index	Clarkson Research
Sec	5 year old second-hand price index	Clarkson Research
Fuel	Vessel bunker fuel price	Clarkson Research
TC_Bulk	1 year time-charter rates, Capesize	Clarkson Research
TC_Tank	1 year time-charter rates, VLCC	Clarkson Research
Oil	Brent crude oil front month contract	Bloomberg
FX_USD	Dollar exchange rate index	Bloomberg
FX_USD_JPN	Dollar-Yen exchange rate	Bloomberg
Cont_Oil	Oil price contango, Brent 6m relative to 1m	Quandl
LIBOR	LIBOR USD 3-month	Bloomberg
HY_Spread	BoA Merrill Lynch HY spread index	Bank of America Merill Lynch
VIX	Volatility Index	Bloomberg
P_IO	Iron ore price, Brazil	WIKI Commodity Prices/Quandl
P_Coal	Coal price, Australia	Clarkson Research
P_Wheat	Wheat price, U.S.	Clarkson Research
P_Metals	Metals price index (copper, aluminum, iron Ore, tin, nickel, zinc, Lead, and Uranium)	IMF Cross Country Macroeconomic Statistics
P_Gold	Gold price	Bloomberg
DBulk_Index	Dry bulk equity index	Yahoo Finance
Tank_Index	Tanker equity index	Yahoo Finance
SP500	S&P500 Equity Index	Bloomberg
MSCI_W	MSCI World Index Equity Index	Bloomberg
MSCI_EM	MSCI Emerging Markets Index	Bloomberg

Appendix A.2. Time series charts

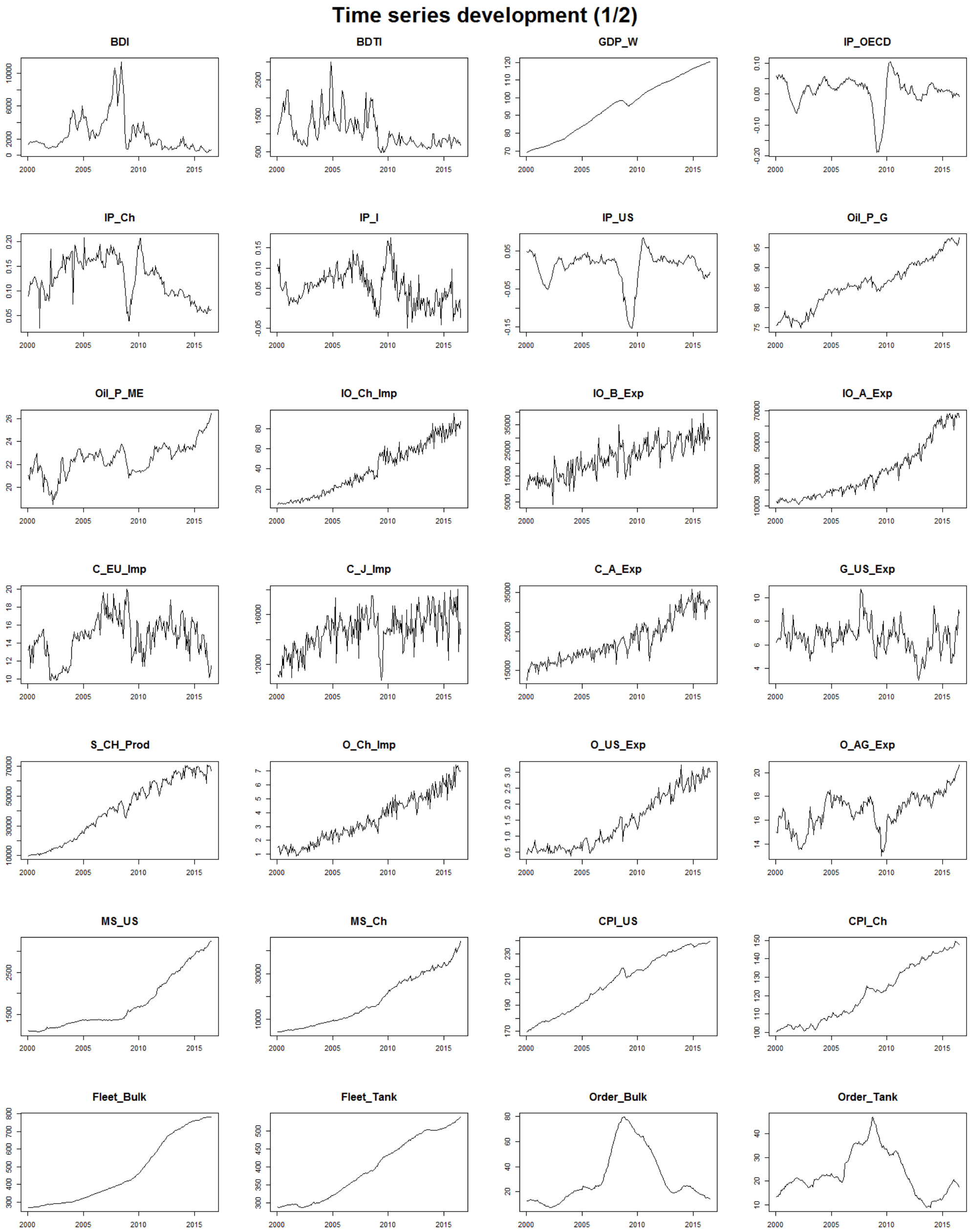

Figure A.1: Time series development Jan-2000 - Jun-2016

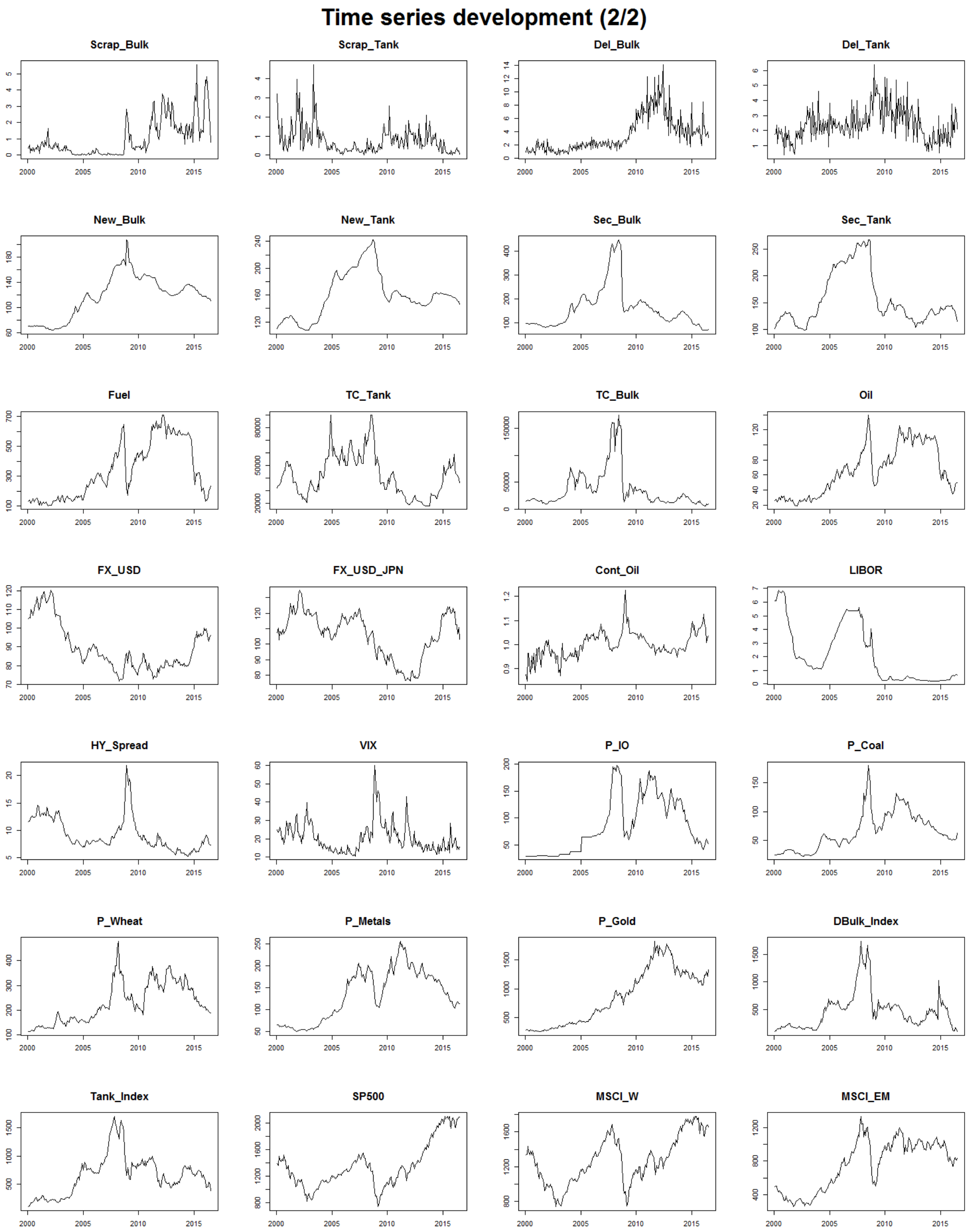

Figure A.2: Time series development Jan-2000 - Jun-2016

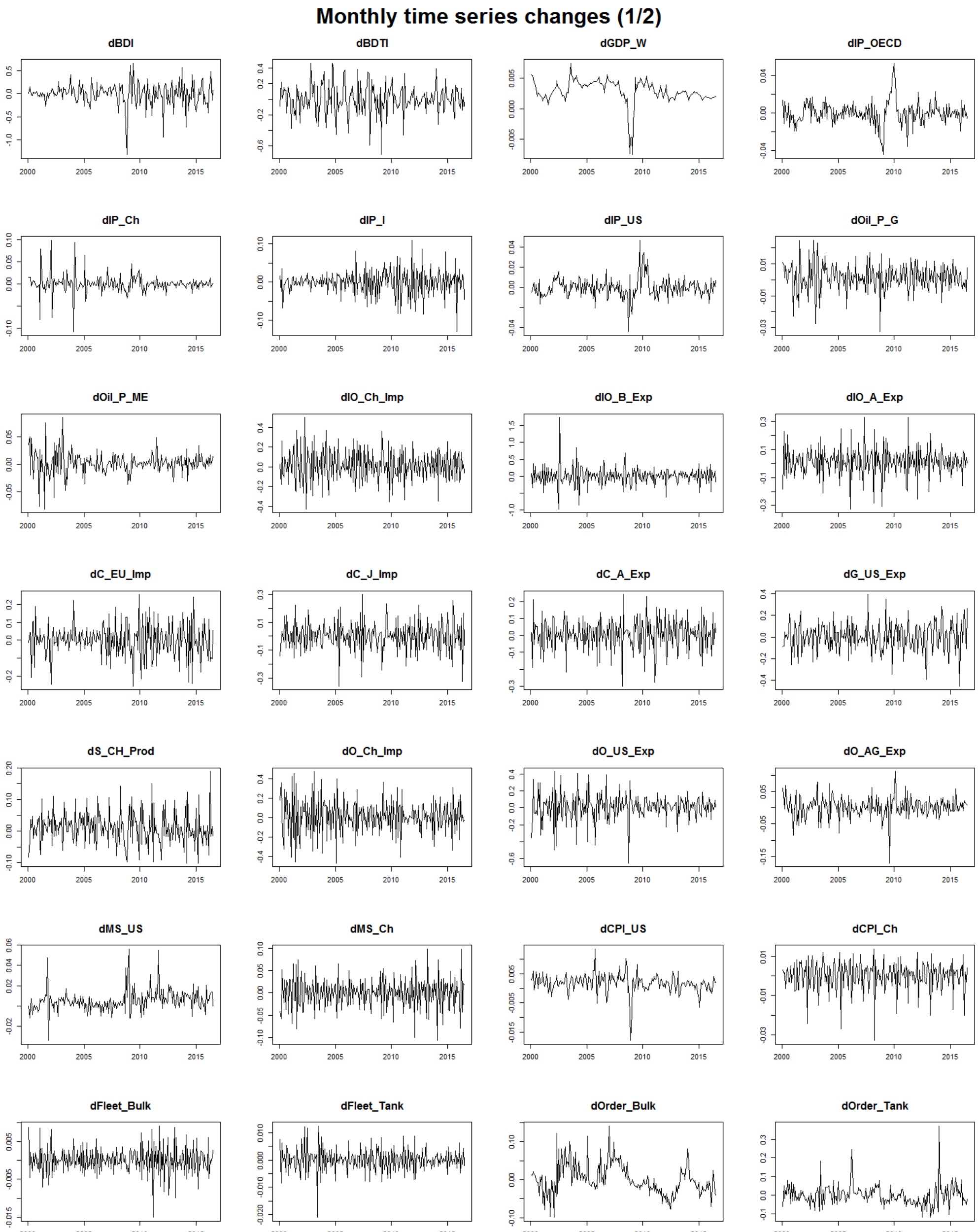

Figure A.3: Monthly logarithmic change in time series from Jan-2000 to Jun-2016

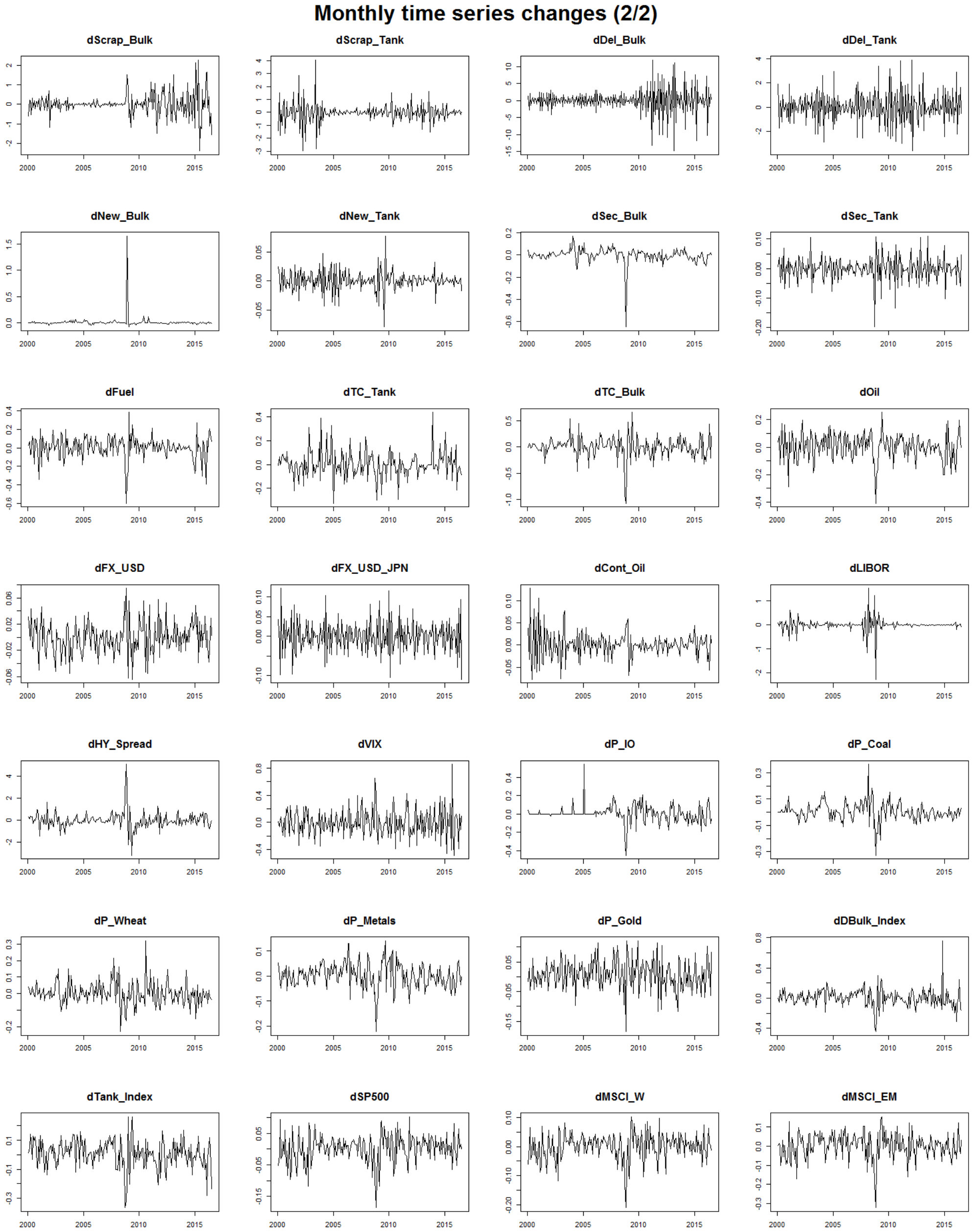

Figure A.4: Monthly logarithmic change in time series from Jan-2000 to Jun-2016

Figure A.5: Distribution plots (part 1)

Figure A.6: Distribution plots (part 2)

Appendix A.3. Dry Bulk and Tanker stock market indices

We introduce two stock market indices constructed using major listed shipping companies operating in the two respective segments. The 30 largest companies ranked by fleet size (dwt) is extracted from Clarkson Research Services Limited [14 February 2017]. Furthermore, stock market data for the publicly listed companies is sourced from Yahoo Finance, where adjusted close[31] price is utilised to compute the index development. The Dry Bulk Index (DBI) and Tanker Index (TI) consist of Y an Z companies, respectively. Moreover, companies operating in both dry bulk and tanker markets are included in the index corresponding to their most important segment measured in dwt. However, if its respective fleet is among the largest three players in a segment, we view the company's industry representation as extensive and thus the company is included in any such case. The DBI and TI are listed in Table A.2 along with their Yahoo Finance stock tickers and owned deadweight tonnage.

The indices are fixed to 100 at December 1999. Each index element is calculated using the average of the accumulated return of companies in the index in the time period of analysis. Let $d \in \{1, ..D\}$ and $x \in \{1, ..X\}$ denote dry bulk and tanker companies, respectively, represented in the index. Let $\gamma_{d,j}$ and $\gamma_{x,j} \in \{0, 1\}$ correspond to whether company d or x is listed ($\gamma = 1$) or not at time $j \in \{1, ..T\}$. The index values is equal to

$$\begin{aligned} DBI_t = 100 \cdot & \left[\left(1 + \frac{\sum_{d=1}^{D} \gamma_{d,j} \Delta_{d,1}}{D \cdot \sum_{d=1}^{D} \gamma_{d,j}} \right) \cdot .. \cdot \left(1 + \frac{\sum_{d=1}^{D} \gamma_{d,j} \Delta_{d,t}}{D \cdot \sum_{d=1}^{D} \gamma_{d,j}} \right) \right] \\ &= 1 + 100 \cdot \prod_{j=1}^{t} \left(1 + \frac{\sum_{d=1}^{D} \gamma_{d,j} \Delta_{d,j}}{D \cdot \sum_{d=1}^{D} \gamma_{d,j}} \right) \end{aligned} \tag{A.1}$$

$$\begin{aligned} TI_t = 100 \cdot & \left[\left(1 + \frac{\sum_{x=1}^{X} \gamma_{x,j} \Delta_{x,1}}{X \cdot \sum_{x=1}^{X} \gamma_{x,j}} \right) \cdot .. \cdot \left(1 + \frac{\sum_{x=1}^{X} \gamma_{x,j} \Delta_{x,t}}{X \cdot \sum_{x=1}^{X} \gamma_{x,j}} \right) \right] \\ &= 100 \cdot \prod_{j=1}^{t} \left(1 + \frac{\sum_{x=1}^{X} \gamma_{x,j} \Delta_{x,j}}{X \cdot \sum_{x=1}^{X} \gamma_{x,j}} \right) \end{aligned} \tag{A.2}$$

where $\Delta_{d,j}$ is the return of company d in time step j. January 2000 corresponds to $t = 1$.

[31] Adjusted Close is adjusted for all applicable share splits and dividends according to Center for Research in Security Prices (CRSP) standards. See http://www.crsp.com/products/documentation/crsp-calculations for description and formulas.

Table A.2: Dry Bulk and Tanker stock market indices. Source: Clarkson Research Services Limited, Yahoo Finance [14 February 2017].

Tanker Index			Dry Bulk Index		
Company	Ticker	Fleet, m.dwt	Company	Ticker	Fleet, m.dwt
China Merchants Grp	0144.HK	17.015	China COSCO Shipping	1919.HK	35.989
China COSCO Shipping	1919.HK	15.849	Nippon Yusen Kaisha	NPNYY	18.238
Euronav NV	EURN	13.294	China Merchants Grp	0144.HK	13.661
Teekay Corporation	TK	12.122	Mitsui O.S.K. Lines	MSLOY	13.473
Frontline	FRO	10.148	Golden Ocean	GOGL	11.767
Gener8 Maritime	GNRT	9.519	Pan Ocean	AZY.SI	8.803
DHT Holdings	DHT	6.723	Star Bulk Carriers	SBLK	7.885
Tsakos Group	TNP	5.895	Navios Group	NM	7.337
Nordic American Tankers	NAT	4.946	Diana Shipping	DSX	5.659
Navios Group	NM	4.161	Wisdom Marine Group	2637.TW	5.610
			ICBC	1398.HK	5.602
			ICBC	1398.HK	5.602
			Genco Shpg & Trading	GNK	4.858
			Scorpio Bulkers	SALT	3.614
			Eagle Bulk Shipping	EGLE	2.271
			Dryships	DRYS	1.422

Appendix B Statistical models and concepts

Appendix B.1. OLS regression

The method of ordinary least squares (OLS) estimates the parameter values $(\hat{\beta}_0, \hat{\beta}_1, ..., \hat{\beta}_k)$ in a linear regression by minimizing the sum of squared residuals, $\sum_{t=1}^{T} \hat{u_t}^2$, hereafter denoted *residual sum of squares* (RSS). With k regressors, the model becomes,

$$y_t = \beta_0 + \beta_1 x_{1t} + \beta_2 x_{2t} + ... + \beta_k x_{kt} + u_t, \quad t \in \{1, .., T\} \tag{B.1}$$

Where β_0 represents the intercept and $\beta_1, \beta_2, ..., \beta_k$ represents the regressors' effect on y and u_t denotes the the random error at time t. Since the error term u_t cannot be directly observed, it is estimated as the deviation between the fitted model value $\hat{y_t}$ and the observation y_t, such that $\hat{u_t} = y_t - \hat{y_t}$. Here, $\hat{y_t}$ is defined as

$$\hat{y_t} = \hat{\beta}_0 + \hat{\beta}_1 x_{1t} + \hat{\beta}_2 x_{2t} + ... + \hat{\beta}_k x_{kt} + \hat{u_t}, \quad t \in \{1, .., T\} \tag{B.2}$$

In order to minimize RSS, one can differentiate RSS with respect to every x_i, resulting in a set of i equations with i unknowns. In vector form, this may equivalently be expressed as

$$\hat{\beta} = \begin{bmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \\ \vdots \\ \hat{\beta}_k \end{bmatrix} = (X'X)^{-1}X'y \tag{B.3}$$

Appendix B.1.1. Gauss-Markov theorem

The Gauss-Markov theorem states that when the four OLS- assumptions are fulfilled (see B.1.2), then the OLS- estimator is in fact the *best linear unbiased estimator* (BLUE). An estimator is said to unbiased if and only if

$$\mathbf{E}(\hat{\beta}_j) = \beta_j \tag{B.4}$$

An estimator is said to be best if the uncertainty associated with estimation is not larger than any other estimator. (Here, among the class of linear unbiased estimators). More precisely, the mean square error of estimation is sought to be minimized,

$$\mathbf{E}\left[(\sum_{i=j}^{k}\lambda_j(\hat{\beta}_j - \beta_j))^2\right] \tag{B.5}$$

where λ_j is the relative weight of each coefficient.

Appendix B.1.2. OLS assumptions

There are four principal assumptions underlying the OLS method:

i) Linear relationship between response- and explanatory variables,

ii) Errors should be independent (no autocorrelation),

$$cov(u_i, u_j) = 0 \tag{B.6}$$

iii) Errors should have constant variance (homoscedasticity), both versus time and versus covariates/fitted model value,

$$var(u_t) = \sigma^2 < \infty \quad \cap \quad cov(u_t, x_j) = 0 \tag{B.7}$$

iv) Normally distributed errors,

$$u_i \sim \mathcal{N}(0, \sigma^2) \tag{B.8}$$

Appendix B.2. Correlation R^2 and R^2_{adj}

The correlation, i.e. linear association between variable x and y, often denoted ρ, may be estimated by the sample correlation coefficient

$$R = \sqrt{\frac{MSS}{TSS}} \tag{B.9}$$

under the assumption that the joint distribution $f(x, y)$ is a bivariate normal distribution.

Here, *MSS* stands for the model sum of squares, defined as

$$MSS = \sum_{i=1}^{T} (\hat{y}_i - \bar{y})^2 \tag{B.10}$$

while *TSS* stands for total sum of squares and is defined as

$$TSS = \sum_{i=1}^{T} (y_i - \bar{y})^2 \tag{B.11}$$

The residual sum of squares, encountered in section B.1, is related to *MSS* and *RSS* through the following identity.

$$TSS = MSS + RSS \tag{B.12}$$

Again, y_i is the ith observation, $\hat{y}_i$ the ith fitted model value and $\bar{y}$ the of y. The variation in the response variable accounted for by the regression model is described by R^2, which is simply the square of the correlation, or in terms of *RSS*, *MSS* and *TSS*:

$$R^2 = 1 - \frac{RSS}{TSS} = \frac{MSS}{TSS} \tag{B.13}$$

R^2 will always increase as more explanatory variables are added. Therefore, *adjusted* R^2 is often preferred, which will decrease if variables with little or no explanatory power are added. R^2_{adj} is defined

$$R^2_{adj} = 1 - \frac{(1 - R^2)(n-1)}{n-p-1} \tag{B.14}$$

where p is the number of explanatory variables (constant term not included) and n the sample size.

Appendix B.3. PRESS and R^2_{pred}

PRESS (predicted sum of squares) was introduced in order to obtain more realistic R^2 values for prediction purposes. *PRESS* is calculated in this manner: First an observation is excluded from the sample and a model is fitted (without that observation). Then, the error between the excluded data-point and the predicted value by the model is calculated. The process is repeated by excluding each and every data-point, one-by-one, and calculating the error. The errors are often called *PRESS* residuals and the *PRESS* value is the sum of squares of these errors. Formally, *PRESS* residuals are denoted

$$\delta_i = y_i - \hat{y}_{i,-i}, \quad i = 1, 2, ..., n, \tag{B.15}$$

where $\hat{y}_{i,-i}$ represents the fitted model where variable i is excluded. Thus, we have

$$PRESS = \sum_{i=1}^{N} \delta_i^2 \tag{B.16}$$

A low $PRESS$ value indicates a good predictive performance. However, the value can take the value of any positive number, thus a more intuitive interpretation of the predictive performance is provided by

$$R^2_{pred} = 1 - \frac{PRESS}{TSS} \tag{B.17}$$

In form it is similar to R^2, the difference is that $PRESS$ is replaced by RSS. Like R^2, R^2_{pred} will take values from 0 to 1. If R^2_{pred} deviates substantially from R^2, it signals that the model might be over-fitted.

Appendix B.4. Mallow's C_p

The idea behind Mallow's C_p is that choosing too few variables in a model will lead to excessive bias, while choosing too many variables will cause unnecessary prediction variance. Finding the compromise between these two considerations is the purpose of Mallow's C_p. The C_p for a given model is estimated by the following:

$$\Gamma_p = \frac{1}{\sigma^2}\sum_{i=1}^{N} Var(\hat{y}_i) + \frac{1}{\sigma^2}\sum_{i=1}^{N}[Bias(\hat{y}_i)]^2 \tag{B.18}$$

Under the standard OLS- assumptions and further assuming that all explanatory variables included indeed constitute the "true" model. (see [Mallows, 1973] for a complete treatment) that under

Appendix B.5. AIC

AIC is a (relative) measure indicating the goodness of fit of a model. The definition follows.

$$AIC = 2p - 2ln(L) \tag{B.19}$$

Here, p is the number of regressors and L the likelihood function[32] of the model. Lowest AIC-value is equivalent with the best goodness of fit. Hence, AIC will penalise a model based on the number of variables it contains, reflected by the positive $2p$-term. In the special of case of OLS, minimising AIC and Mallow's C_p is shown to be approximately equivalent Boisbunon et al. [2014].

[32] Further discussion on the method of maximum likelihood is provided by Alexander [2008].

Appendix B.6. ARIMA(p, d, q)

An ARIMA(p, d, q) (Autoregressive integrated moving average) model is a given as

$$(1 - \sum_{k=1}^{p} \alpha_k L^k)(1 - d)X_t = (1 + \sum_{k=1}^{q} \beta_k L^k)\epsilon_t \tag{B.20}$$

where the parameters $\alpha_1, .., \alpha_p$ and $\beta_1, .., \beta_k$ concern the autoregressive and moving average parts, respectively. $\epsilon_1, .., \epsilon_k$ denotes the error terms. L is the lag operator and X_t denotes the time series. The p, d, and q properties represents the number of autoregressive, seasonal and moving average components, respectively.

Appendix B.7. Multicollinearity

The phenomenon of multicollinearity arise in multiple regression when two or more explanatory variables are correlated to such extent that one may be predicted with a considerable degree of accuracy by the other(s). (In the case of only two correlated variables they are said to be collinear, otherwise multicollinear.) More formally, the vector containing coefficients for all variables in the model is given $\hat{\beta} = (X'X)^{-1}X'y$ and the covariance matrix of the regression parameters is $Cov(\hat{\beta}) = \sigma^2(X'X)^{-1}$. In the case of multicollinearity, $(X'X)^{-1}$ may have large diagonal elements, thus the diagonal elements of $Cov(\hat{\beta})$ becomes large, which is the variance of the regression parameters.

If explanatory variables are correlated, it may not be straight forward to determine which variables to include in the model. As they provide information with similarities, their distinct significance, coefficient estimates and standard errors will depend on whether other collinear variables are already included in the model. For instance, the estimate of β_1 in a model with only x_1 will change if x_2 is collinear with x_1 and added to the model. Also x_1 will be less significant if x_2 is added to the model, since x_2 provides information that was initially individually provided by x_1. In our case, the stepwise selection procedure handles the issue of multicollinearity relatively well, since a variable that is highly collinear with any present variables in the model is unlikely to add most marginal explanatory power, and is thus unlikely to be selected as the incoming variable.

Unless an experiment is designed to orthogonal, independent variables tend to be correlated to a certain degree, meaning that there will be some level of multicollinearity in the data. Hence the problem is not an binary issue, but needs to be evaluated in each individual case. There are some ways to monitor the level of multicollinearity. One is to calculate the VIF, and compare it versus a pre-determined upper value (see section VIF).Alexander [2008] presented a different approach to detect multicollinearity: If the square of correlation between two independent variables is greater than the R^2_{adj} of the regression output, multicollinearity is an issue. There are a few ways to handle the issue, depending on its severity. In our case, we are unlikely to find a high degree of multicollinear in models with few variables, but the extent of the problem may increase with the number of variables. If many variables are significant, but only a few have high VIF-values, the problem can be solved by simply removing those variables, and then perform a new regression analysis with the remaining variables. This is legitimate, as a variable with a very high VIF-value will not bring much new information to the model, however, the regression should be repeated, so to re-calculate the significance, coefficients and standard errors of the remaining variables.

Appendix B.8. VIF-Variance inflation factor

The VIF measures the degree of multicollinearity associated with each variable in an OLS- regression. It represents a scale factor of how many times the standard deviation of a particular parameter is inflated as the result of not being an orthogonal variable in the regression. We will demonstrate how the VIF associated with a particular variable is obtained. Assume we wish to fit a regression model consisting of p variables. Next, we assume the VIF associated with variable i is of interest, where $1 \leq i \leq p$. For simplicity lets assume we $i = 1$.

Step 1 Perform a regression on the form

$$X_1 = \beta_2 X_2 + \beta_3 X_3 + ... + \beta_p X_p + \beta_0 + \epsilon \tag{B.21}$$

and calculate the R^2 of the model. Note that only X_i is on the LHS and the remaining set of variables are on the RHS. We denote the R^2 associated with this model R_i^2.

Step 2 Calculate the VIF_i associated with x_i by the following formula

$$VIF_i = \frac{1}{1 - R_i^2} \tag{B.22}$$

If VIF equals 1, it means that the variable is orthogonal to the remaining set of variables, implying that the estimated standard deviation of the coefficient is unaffected by whether other variables are added to/removed from the model (in fact, the coefficient of an orthogonal variable is also unaffected by other variables in the model). If $VIF > 1$, then the variable is not orthogonal, implying that the variable to some extent/ in some regions display overlapping tendencies relative the remaining set of variables in the model. If one variable is a pure linear combination of the rest of the set, it cannot be estimated as its VIF would approach infinity. Most statistical software would exclude the variable, before carrying out the analysis. Note that VIF never can be less than 1 since R^2 is always positive.

Various levels of multicollinearity have been deemed acceptable by the literature. An established rule of thumb is to say that the degree of multicollinearity is high if $VIF > 10$ Neter et al. [1989]. However, more recently, stricter VIF have been recommended, for [Jackson, 2008] suggested an upper VIF of 4.

Appendix B.9. ACF-plots and their interpretation

If a time series consists of a random component and the sample size is large, the lagged correlation coefficient is approximately normally distributed with a mean of zero and variance $1/N$ [Chatfield, 2004]. We may construct a H_0 stating that the correlation between each lag is zero (no autocorrelation). The critical value where H_0 is rejected at a certain significance level α [33], is given by

$$r_\alpha = 0 \pm \frac{z_{\frac{\alpha}{2}}}{\sqrt{N}} \tag{B.23}$$

Most statistical software provides autocorrelation function (ACF)- plots of the residuals obtained from a regression. Normally, the default significance is set to 5%, thus $z_{\frac{\alpha}{2}}$ = 1.96. Then r_α appears as the critical lines in the ACF-plots. ACF- plots are useful as they explicitly point which lags that appears to be autocorrelated, as well as their direction.

[33] $\frac{\alpha}{2}$ appears in the formula because the alternative hypothesis is two-sided.

Appendix C Statistical tests

Appendix C.1. Jarque-Bera test for normality

Bera and Jarque [1981] developed one of the most common tests for normality of the residuals. Let u and σ^2 denote the residuals and the residual variance, respectively. The Jarque-Bera test examines if the coefficients of skewness and kurtosis, respectively denoted by b_1 and b_2, are jointly zero. These coefficients are given by,

$$b_1 = \frac{E[u^3]}{(\sigma^2)^{\frac{3}{2}}} \tag{C.1}$$

$$b_2 = \frac{E[u^4]}{(\sigma^2)^2} \tag{C.2}$$

Let T and and denote the sample size. The test statistic, JB, follows a chi-squared distribution with two degrees of freedom and is given by,

$$JB = \frac{T}{6} \cdot [b_1 + \frac{(b_2 - 3)^2}{4}] \sim \chi^2(2) \tag{C.3}$$

The null hypothesis H_0, states that residuals are normally distributed. H_0 is rejected if the test statistic is above its critical value.

Appendix C.2. Augmented Dickey-Fuller test

The objective of the Augmented Dickey-Fuller (ADF) test is to examine whether a time series contains a unit root. A time series is stationary if it has a constant mean, variance and autovariance. The null hypothesis, H_0, states that the series contains a unit root and hence the series being non-stationary. The model is given by,

$$\Delta y_t = \psi y_{t-1} + \sum_{n=1}^{p} \alpha_n y_{t-n} + u_t \tag{C.4}$$

where ϕ denotes the root and $\psi = 1 - \phi$. u_t denotes the residuals and p is the number of lags of the dependent variable. The test statistic is defined by,

$$Test\ statistic = \frac{\Psi}{SE(\Psi)} \tag{C.5}$$

Where $SE(\Psi)$ is the standard error. H_0 of non-stationarity is rejected in if the test statistic is more negative than the critical value.

Appendix C.3. Breusch-Godfrey test for autocorrelation

For the OLS-estimator to be BLUE, the residuals obtained from a regression cannot be autocorrelated, i.e correlated across time. (However, in the presence of autocorrelation, OLS will remain unbiased.) Breusch-Godfrey tests for autocorrelation among lags up to order p. First, the standard regression is performed,

$$y_t = \beta_1 + \beta_2 x_{2t} + ... + \beta_n x_{nt} + u_t \tag{C.6}$$

Next, the residuals are saved and a new regression is performed on the lagged residuals,

$$u_t = \rho_0 + \rho_1 u_{t-1} + \rho_2 u_{t-2} + ... + \rho_p u_{t-p} + v_t, \quad v_t \sim N(0, \sigma_v^2) \tag{C.7}$$

The null hypothesis, H_0, states that the series is not autocorrelated for any lag. H_0 is rejected if there is significant autocorrelation between any lag.

$$\begin{aligned} H_0 &: \rho_0 + \rho_1 + \rho_2 + ... + \rho_p = 0 \\ H_1 &: \rho_0 \cup \rho_1 \cup \rho_2 \cup ... \cup \rho_p \neq 0 \end{aligned} \tag{C.8}$$

Appendix C.4. White's test for heterocedasticity

First, a standard linear regression is estimated,

$$y_t = \beta_0 + \beta_1 x_{1t} + ... + \beta_n x_{nt} + u_t \tag{C.9}$$

where $\hat{u_t}$ denotes the residual at time t when the residual sum of squares is minimised. The following auxiliary regression is then carried out,

$$\hat{u_t}^2 = \delta_0 + \delta_1 \hat{Y_n} + \delta_2 \hat{Y}_n^2 \tag{C.10}$$

where $\hat{Y_n}$ represents the predicted values from the regression

$$\hat{Y}_n = \hat{\beta_0} + \hat{\beta_1} x_{1t} + ... + \hat{\beta_n} x_{nt} \tag{C.11}$$

where $\hat{\beta_1}, .., \hat{\beta_n}$ are the optimal independent variable parameters. Furthermore, it is possible to use two test frameworks. First, the Lagrange Multiplier (LM) approach is applied, which uses the value of R^2 obtained from the auxiliary regression. The value of R^2 is then multiplied by the number of observations, T, and it can be shown that,

$$TR^2 \sim \chi^2(m) \tag{C.12}$$

where m is obtained from the auxiliary regression, being the the number of regressors less the constant term. Second, the F-test framework is used. The test statistic is calculated with two and three degrees of freedom in the numerator and denominator, respectively. The null hypothesis of homoscedasticity, H_0, is rejected if either of the F-test or chi-squared test statistics are significant. If not, H_0 is not rejected.

Appendix C.5. Durbin-Watson test for autocorrelation

The Durbin-Watson (DW) is a test that evaluates whether there is first order autocorrelation, i.e. it checks whether there is a relationship between an error and the previous error. The test statistic can be approximately written as

$$DW \approx 2(1 - \hat{\beta}) \tag{C.13}$$

where $\hat{\beta} = corr(\hat{u_t}, \hat{u_{t-1}})$. Since $-1 \leq \hat{\beta} \leq 1$, we have that $0 \leq DW \leq 4$, where $DW = 0$ corresponds to perfectly positive autocorrelation, while $DW = 4$ corresponds to perfectly negative autocorrelation. The DW test statistic do not follow a standard statistical distribution, hence the critical values are found in tables. There are two critical values associated with DW, namely d_u and d_l: Between d_u and 4 - d_u is there no evidence of autocorrealtion, while outside d_l and 4 - d_l is either positive or negative autocorrelation evident.

Appendix C.6. Ramsey's RESET test for misspecification

The *Ramsey regression equation specification error test* (RESET) test [Ramsey, 1969] examines whether non-linear combinations of the predicted values explain the dependent variable, y_t. The test uses an auxiliary regression with higher order terms of the predicted values together with the explanatory variables. The auxiliary regression is formulated,

$$y_t = \alpha_1 + \alpha_2 \hat{y}_t^2 + ... + \alpha_p \hat{y}_t^p + \sum \beta_i x_{it} + v_t \tag{C.14}$$

where v_t is the error term and p is the highest order term of the predicted values in the auxiliary regression. Let TR^2 denote the test statistic, which is chi-squared distributed with p-1 degrees of freedom,

$$TR^2 \sim \chi^2(p-1) \tag{C.15}$$

The null hypothesis, H_0, that the model's functional form is correct, is rejected if the test statistic is greater than critical value given by the χ^2 distribution.

Appendix C.7. Anderson-Darling test for distributional adequacy

The Anderson-Darling test is used to test whether it is reasonable to believe that a sample of data is drawn from population with a specific distribution.

$$\begin{aligned} H_0 &: \text{The data follow the specified distribution} \\ H_1 &: \text{The data do not follow the specified distribution} \end{aligned} \tag{C.16}$$

The test statistic is defined as

$$A^2 = -N - S \tag{C.17}$$

where N is the number of observations and S is defined as

$$S = \sum_{i=1}^{N} \frac{2i-1}{N} [lnF(Y_i) + ln(1 - F(Y_{N+1-i}))] \tag{C.18}$$

In the latter expression, F is the CDF of the specified distribution and $Y_1, Y_2, ..., Y_N$ are ordered data. Critical values for the AD- test depends on which distribution function that is tested. Stephens [1974] has published critical values for a range of distribution functions, including the normal distribution.

Appendix C.8. F-test for joint significance

Consider a multiple linear regression model,

$$y_t = \beta_1 + \beta_2 x_{2t} + ... + \beta_n x_{nt} + u_t \tag{C.19}$$

where the subject of interested is whether the variables have any predictive power, both individually and jointly. The hypotheses may be formulated in the following way

$$\begin{aligned} H_0 &: \beta_0 + \beta_1 + \beta_2 + ... + \beta_p = 0 \\ H_1 &: \beta_0 \cup \beta_1 \cup \beta_2 \cup ... \cup \beta_p \neq 0 \end{aligned} \tag{C.20}$$

The test statistic following the F-distribution may now be written

$$F_0 = \frac{(SSR_r - SSR_{ur})/q}{SSR_{ur}/(n-(k+1))} \tag{C.21}$$

By imposing that the null is true and removing variables from the model, we are essentially "restricting" the model. Thus, the difference $SSRr - SSR_{ur}$ is telling us how much larger the residuals are in the model where the null hypothesis is (assumed) true. If the residuals are sufficiently larger in the restricted model, then F_0 will also be big. Since large residuals implies worse regression fit, we have reason to question the null hypothesis when F_0 is big. If these variables really had no predictive power, then removing them should not affect the residuals.

Appendix C.9. Diebold-Mariano test

The Diebold-Mariano framework evaluates the significance of apparent predictive superiority. For instance, if model 1 outperforms model 2 in an out-of-sample test, the predictive strength of model 1 relative to model 2 seems better. However, if the models were to be compared in a different out-of-sample window, how confidently could we claim that 1 would remain ahead? The Diebold-Mariano test will answer that question - it assesses whether model 1 generally provides a better predictive accuracy than model 2.

We define the loss differential between model 1 and 2 for each time step. $d_{12t} = e_{1t}^2$ - e_{2t}^2, where e_{it}^2 denotes the squared residual of model i at time step t, where $i = \{1, 2\}$. Furthermore, let $\bar{d_{12}}$ denotes the time average of the loss differential between the models. If the following assumptions hold,

$$\mathbf{E}(d_{12t}) = \mu \tag{C.22}$$

$$\mathbf{Var}(d_{12t}) = \sigma^2 < \infty \tag{C.23}$$

$$\mathbf{Cov}(d_{12t}, d_{12t-\tau}) = \lambda(\tau), \qquad \forall t \tag{C.24}$$

then, under the null-hypothesis of $E(\bar{d_{12}}) = 0$ (i.e. equal predictive power), $\mathrm{DM}_{12} = \frac{\bar{d_{12}} - E(\bar{d_{12}})}{St.Dev.(\bar{d_{12}})} \rightarrow \mathcal{N}(0,1)$.

Appendix D Test results

Appendix D.1. Stationarity test results

Table D.1: Augmented Dickey-Fuller test results

	Logarithmic change		**Transformed series**	
Variable	**ADF τ-stat**	**$P > \lvert t \rvert$**	**ADF τ-stat**	**$P > \lvert t \rvert$**
ΔBDI	-3.909	0.0149**		
ΔBDTI	-4.459	<0.01***		
ΔGDP_W	-3.347	0.048**		
ΔIP_OECD	-5.430	<0.01***		
ΔIP_Ch	-5.578	<0.01***		
ΔIP_I	-5.345	<0.01***		
ΔIP_US	-5.976	<0.01***		
ΔOil_P_G	-3.865	0.017**		
ΔOil_P_ME	-3.761	0.022**		
ΔIO_Ch_Imp	-5.668	<0.01***		
ΔIO_B_Exp	-8.463	<0.01***		
ΔIO_A_Exp	-5.961	<0.01***		
ΔC_EU_Imp	-3.609	0.034**		
ΔC_J_Imp	-4.854	<0.01***		
ΔC_A_Exp	-5.488	<0.01***		
ΔG_US_Exp	-3.723	0.0241**		
ΔS_Ch_Prod	-3.789	0.021**		
ΔO_Ch_Imp	-5.274	<0.01***		
ΔO_US_Exp	-6.496	<0.01***		
ΔO_AG_Exp	-3.874	0.017**		
ΔMS_US	-3.245	0.049*		
ΔMS_Ch	-2.917	0.192	-6.502	<0.01***
ΔCPI_US	-4.602	<0.01***		
ΔCPI_Ch	-2.436	0.392	-6.591	<0.01***
ΔFleet_Bulk	-1.968	0.589	-3.809	0.020**
ΔFleet_Tank	-2.683	0.290	-4.312	<0.01***
ΔOrder_Bulk	-3.521	0.042**		
ΔOrder_Tank	-3.235	0.045**		
ΔScrap_Bulk	-4.619	<0.01***		
ΔScrap_Tank	-3.844	0.018**		
ΔDel_Bulk	-2.608	0.322	-8.128	<0.01***
ΔDel_Tank	-4.557	<0.01***		
ΔNew_Bulk	-3.922	0.014**		
ΔNew_Tank	-2.929	0.187	-4.979	<0.01***
ΔSec_Bulk	-3.727	0.024**		
ΔSec_Tank	-2.723	0.274	-4.783	<0.01***
ΔFuel	-4.116	<0.01***		
ΔTC_Bulk	-4.195	<0.01***		
ΔTC_Tank	-3.633	0.032**		
ΔOil	-4.370	<0.01***		
ΔFX_USD	-4.728	<0.01***		
ΔFX_USD_JPN	-3.067	0.089*	-6.982	<0.01***
ΔCont_Oil	-3.807	0.020**		
ΔLIBOR	-2.422	0.399	-5.220	<0.01***
ΔHY_Spread	-4.406	<0.01***		
ΔVIX	-4.644	<0.01***		
ΔP_IO	-4.612	<0.01***		
ΔP_Coal	-4.491	<0.01***		
ΔP_Wheat	-4.775	<0.01***		
ΔP_Metals	-3.566	0.038**		
ΔP_Gold	-3.183	0.049**		
ΔDBulk_Index	-4.309	<0.01***		
ΔTank_Index	-3.691	0.026**		
ΔSP500	-3.737	0.023**		
ΔMSCI_W	-3.694	0.026**		
ΔMSCI_EM	-4.363	<0.01***		

Significance level: *** 0.01, ** 0.05, * 0.1.

Appendix E Model correlation matrices

	dBDI	dBDI_2	dBDI_5	dIP_OECD_3	dIP_Ch_2	dIP_I_6	dOil_P_G_3	dIO_Ch_Imp_4	dIO_B_Exp_2	dIO_A_Exp_1	dC_EU_Imp_2	dG_US_Exp_6	dS_CH_Prod_2	dS_CH_Prod_4	dCPI_US_3	dCPI_Ch_2	dCPI_Ch_3	dCPI_Ch_5	dOrder_Bulk_4	dScrap_Bulk_6
dBDI	1.00																			
dBDI_2	-0.11	1.00																		
dBDI_5	-0.05	-0.08	1.00																	
dIP_OECD_3	0.04	0.06	0.27	1.00																
dIP_Ch_2	0.10	0.10	0.07	0.16	1.00															
dIP_I_6	0.15	-0.01	0.00	0.10	0.01	1.00														
dOil_P_G_3	-0.05	0.17	0.10	0.07	0.08	-0.02	1.00													
dIO_Ch_Imp_4	-0.13	-0.01	0.11	0.03	0.07	-0.07	-0.02	1.00												
dIO_B_Exp_2	0.00	0.11	-0.05	0.09	-0.22	0.01	-0.17	-0.27	1.00											
dIO_A_Exp_1	-0.14	0.23	-0.03	-0.02	-0.05	-0.25	0.16	0.04	-0.05	1.00										
dC_EU_Imp_2	0.04	-0.05	-0.16	0.02	-0.23	-0.07	-0.01	-0.01	0.23	0.02	1.00									
dG_US_Exp_6	0.12	-0.01	-0.02	0.02	-0.16	0.13	0.08	-0.24	-0.02	-0.12	0.12	1.00								
dS_CH_Prod_2	0.19	0.12	-0.06	0.04	0.17	-0.04	-0.01	-0.03	0.08	-0.12	-0.02	0.06	1.00							
dS_CH_Prod_4	0.07	0.22	0.25	0.03	0.02	0.05	-0.14	0.01	0.18	-0.04	-0.04	-0.01	0.11	1.00						
dCPI_US_3	-0.29	0.09	0.18	0.19	0.07	-0.02	-0.06	0.03	0.09	0.02	0.04	0.04	-0.04	0.15	1.00					
dCPI_Ch_2	0.02	-0.12	0.02	0.03	-0.17	-0.07	-0.03	-0.06	0.00	0.08	-0.14	0.05	-0.31	-0.09	-0.13	1.00				
dCPI_Ch_3	0.11	0.09	0.05	-0.03	0.09	-0.01	0.00	0.22	-0.13	-0.08	-0.04	-0.07	0.01	0.28	0.09	-0.30	1.00			
dCPI_Ch_5	-0.12	0.09	-0.11	0.02	-0.09	0.12	0.03	-0.12	0.19	-0.04	0.13	0.20	0.13	0.01	-0.01	-0.01	-0.17	1.00		
dOrder_Bulk_4	0.12	0.09	0.11	0.00	0.05	0.04	0.16	-0.01	-0.13	0.02	0.02	0.00	-0.01	-0.05	0.15	-0.07	0.05	-0.05	1.00	
dScrap_Bulk_6	0.20	-0.08	-0.03	-0.10	-0.06	-0.19	-0.03	0.04	-0.04	0.13	-0.03	-0.06	-0.02	0.11	-0.08	0.07	0.06	0.08	-0.08	1.00
dDel_Bulk_5	0.14	0.09	-0.12	-0.09	-0.06	0.00	0.06	-0.28	-0.01	-0.05	0.00	0.05	-0.16	-0.19	0.01	0.02	-0.10	-0.11	0.08	-0.18
dSec_Bulk_3	-0.19	0.10	0.26	0.28	0.15	0.11	-0.07	0.09	0.08	-0.11	-0.02	-0.08	0.04	0.24	0.37	-0.03	0.08	-0.03	0.18	-0.05
dFX_USD_1	-0.26	-0.05	-0.22	0.03	-0.01	0.15	-0.12	-0.14	-0.01	-0.14	0.06	-0.02	-0.16	-0.08	0.01	0.09	-0.08	0.01	-0.02	-0.10
dLIBOR_4	-0.10	-0.08	0.10	0.13	-0.09	-0.15	0.05	0.09	0.08	-0.01	-0.03	-0.12	-0.10	0.02	0.18	0.12	-0.18	-0.04	-0.02	0.17
dP_IO_2	-0.14	0.37	0.09	0.26	0.15	0.09	0.10	0.09	-0.03	0.01	-0.05	0.04	0.15	-0.01	0.06	0.07	0.01	-0.01	0.08	-0.14
dP_Coal_1	0.00	0.12	0.21	0.29	0.11	0.02	0.04	0.02	0.04	-0.11	0.01	0.05	0.13	-0.01	-0.02	0.02	-0.06	0.02	0.16	-0.07
dP_Coal_2	-0.16	0.17	0.13	0.24	0.04	0.00	0.19	0.04	-0.03	0.09	-0.04	0.13	0.01	0.00	0.19	0.18	0.02	-0.05	0.20	-0.09
dP_Wheat_5	0.03	0.00	0.22	0.05	0.08	0.06	0.07	0.04	-0.01	-0.06	0.06	0.08	0.09	0.02	0.09	-0.07	0.01	0.10	0.07	-0.06
dP_Gold_5	0.11	-0.05	0.20	0.13	0.04	0.08	0.04	0.11	-0.01	0.03	-0.08	-0.01	-0.06	0.03	0.04	-0.03	-0.08	0.19	0.08	0.13
dDBulk_Index_1	0.33	0.14	0.19	0.11	0.06	0.01	0.07	0.13	-0.04	0.09	-0.05	-0.04	0.07	-0.01	-0.02	-0.08	0.03	0.03	0.11	0.09
dDBulk_Index_2	0.16	0.35	0.10	0.15	0.10	-0.01	0.04	-0.08	0.04	0.10	-0.02	0.01	0.18	0.09	-0.01	0.06	-0.08	0.13	0.14	-0.04
dDBulk_Index_3	-0.27	0.35	0.05	0.17	0.05	0.05	0.15	0.08	0.06	0.07	-0.08	-0.13	0.20	0.08	0.14	-0.05	0.06	0.03	0.17	-0.13

Figure E.1: Correlation matrix (part 1) of significant variables in BDI-models.

	dCPI_Ch_3	dCPI_Ch_5	dOrder_Bulk_4	dScrap_Bulk_6	dDel_Bulk_5	dSec_Bulk_3	dFX_USD_1	dLIBOR_4	dP_IO_2	dP_Coal_1	dP_Coal_2	dP_Wheat_5	dP_Gold_5	dDBulk_Index_1	dDBulk_Index_2	dDBulk_Index_3
dBDI																
dBDI_2																
dBDI_5																
dIP_OECD_3																
dIP_Ch_2																
dIP_I_6																
dOil_P_G_3																
dIO_Ch_Imp_4																
dIO_B_Exp_2																
dIO_A_Exp_1																
dC_EU_Imp_2																
dG_US_Exp_6																
dS_CH_Prod_2																
dS_CH_Prod_4																
dCPI_US_3																
dCPI_Ch_2																
dCPI_Ch_3	1.00															
dCPI_Ch_5	-0.17	1.00														
dOrder_Bulk_4	0.05	-0.05	1.00													
dScrap_Bulk_6	0.06	0.08	-0.08	1.00												
dDel_Bulk_5	-0.10	-0.11	0.08	-0.18	1.00											
dSec_Bulk_3	0.08	-0.03	0.18	-0.05	0.02	1.00										
dFX_USD_1	-0.08	0.01	-0.02	-0.10	0.12	0.05	1.00									
dLIBOR_4	-0.18	-0.04	-0.02	0.17	0.04	-0.07	-0.19	1.00								
dP_IO_2	0.01	-0.01	0.08	-0.14	-0.06	0.31	0.02	-0.17	1.00							
dP_Coal_1	-0.06	0.02	0.16	-0.07	0.02	0.26	-0.12	0.04	0.29	1.00						
dP_Coal_2	0.02	-0.05	0.20	-0.09	0.04	0.27	-0.03	0.01	0.35	0.36	1.00					
dP_Wheat_5	0.01	0.10	0.07	-0.06	-0.09	0.15	-0.09	0.04	0.05	0.22	0.09	1.00				
dP_Gold_5	-0.08	0.19	0.08	0.13	-0.12	0.23	-0.09	-0.01	0.03	0.07	-0.04	0.14	1.00			
dDBulk_Index_1	0.03	0.03	0.11	0.09	-0.04	-0.03	-0.43	0.18	0.17	0.24	0.10	0.07	0.00	1.00		
dDBulk_Index_2	-0.08	0.13	0.14	-0.04	0.03	0.20	0.00	-0.13	0.32	0.38	0.25	-0.01	0.07	0.21	1.00	
dDBulk_Index_3	0.06	0.03	0.17	-0.13	-0.03	0.45	0.02	-0.08	0.44	0.28	0.38	0.17	-0.02	0.02	0.21	1.00

Figure E.2: Correlation matrix (part 2) of significant variables in BDI-models.

	dBDI	dSBDI_5	dSBDTI_3	dIP_OECD_3	dIP_Ch_2	dIP_L6	dIO_A_Exp_1	dC_EU_Imp_2	dC_J_Imp_6	dC_A_Exp_6	dG_US_Exp_2	dG_US_Exp_6	dS_CH_Prod_2	dS_CH_Prod_3	dS_CH_Prod_4	dMS_Ch_2	dMS_Ch_6	dCPI_US_3	dCPI_Ch_3	dCPI_Ch_5	dFleet_Bulk_1	dOrder_Bulk_2	dScrap_Bulk_3
dBDI	1.00																						
dSBDI_5	-0.06	1.00																					
dSBDTI_3	-0.10	0.09	1.00																				
dIP_OECD_3	0.04	0.28	0.14	1.00																			
dIP_Ch_2	0.10	0.06	-0.04	0.16	1.00																		
dIP_L6	0.15	0.03	-0.03	0.10	0.01	1.00																	
dIO_A_Exp_1	-0.14	-0.02	-0.07	-0.02	-0.05	-0.25	1.00																
dC_EU_Imp_2	0.04	-0.12	0.07	0.02	-0.23	-0.07	0.02	1.00															
dC_J_Imp_6	-0.09	0.08	0.09	-0.07	-0.02	0.01	0.05	-0.07	1.00														
dC_A_Exp_6	-0.02	-0.08	-0.03	0.02	-0.12	0.21	-0.07	-0.03	0.15	1.00													
dG_US_Exp_2	0.10	-0.03	0.05	0.05	0.05	-0.03	-0.07	-0.05	-0.07	0.00	1.00												
dG_US_Exp_6	0.12	-0.02	0.12	0.02	-0.16	0.13	-0.12	0.12	0.10	0.15	0.12	1.00											
dS_CH_Prod_2	0.19	-0.05	0.04	0.04	0.17	-0.04	-0.12	-0.02	-0.19	-0.09	0.20	0.06	1.00										
dS_CH_Prod_3	-0.06	-0.06	0.02	0.14	-0.04	0.12	0.20	-0.20	-0.02	0.02	-0.02	-0.13	-0.21	1.00									
dS_CH_Prod_4	0.07	0.12	0.06	0.03	0.02	0.05	-0.04	-0.04	0.09	0.00	0.12	-0.01	0.11	-0.21	1.00								
dMS_Ch_2	0.03	-0.01	-0.05	0.04	0.11	-0.02	-0.14	0.11	-0.16	-0.18	0.22	-0.03	0.30	-0.15	0.01	1.00							
dMS_Ch_6	0.00	0.04	-0.02	-0.01	-0.13	0.11	-0.04	0.04	0.12	0.30	-0.03	0.24	0.02	0.02	-0.06	-0.17	1.00						
dCPI_US_3	-0.29	0.22	0.37	0.19	0.07	-0.02	0.02	0.04	0.02	0.01	0.12	0.04	-0.04	0.00	0.15	-0.04	0.09	1.00					
dCPI_Ch_3	0.11	-0.05	0.04	-0.03	0.09	-0.01	-0.08	-0.04	-0.06	-0.01	-0.01	-0.07	0.01	-0.31	0.28	0.04	-0.10	0.09	1.00				
dCPI_Ch_5	-0.12	-0.03	-0.02	0.02	-0.09	0.12	-0.04	0.13	0.12	0.37	0.11	0.20	0.13	-0.02	0.01	-0.04	0.42	-0.01	-0.17	1.00			
dFleet_Bulk_1	0.02	-0.05	0.02	0.12	-0.05	-0.08	-0.08	0.09	-0.04	0.11	0.06	-0.07	-0.11	0.04	0.17	0.21	-0.01	0.00	0.05	-0.01	1.00		
dOrder_Bulk_2	0.11	0.17	-0.05	0.10	0.02	0.08	-0.04	0.04	0.00	0.01	0.06	0.09	-0.05	0.17	-0.03	0.02	0.12	0.00	-0.06	0.00	0.20	1.00	
dScrap_Bulk_3	0.09	-0.33	-0.06	-0.11	-0.15	-0.08	0.05	0.00	0.10	0.00	-0.13	-0.02	-0.01	0.16	-0.20	0.13	-0.04	-0.20	-0.04	-0.03	0.10	-0.01	1.00
dScrap_Bulk_4	-0.09	-0.42	-0.01	-0.15	0.07	0.00	0.05	0.10	-0.05	-0.01	-0.01	0.07	0.11	-0.01	0.18	-0.01	0.06	-0.10	0.08	-0.03	0.06	-0.08	0.00
dScrap_Bulk_6	0.20	-0.02	-0.05	-0.10	-0.06	-0.19	0.13	-0.03	0.06	-0.15	0.04	-0.06	-0.02	0.03	0.11	-0.10	-0.04	-0.08	0.06	0.08	-0.08	-0.16	-0.08
dDel_Bulk_4	-0.13	0.07	-0.03	0.09	0.07	-0.05	0.08	0.08	0.02	-0.12	0.09	-0.02	0.16	-0.19	0.24	0.03	-0.25	-0.07	0.14	0.03	-0.29	-0.07	-0.15
dNew_Bulk_6	0.26	0.03	-0.01	-0.01	0.15	-0.15	0.00	-0.02	0.02	-0.04	0.24	-0.07	0.14	-0.04	0.12	0.13	-0.03	0.04	-0.09	0.06	-0.02	-0.03	-0.04
dSec_Bulk_2	-0.11	0.13	0.10	0.24	0.12	-0.07	0.18	-0.05	-0.12	-0.03	0.14	0.02	0.20	0.24	0.13	-0.08	-0.01	0.18	0.00	-0.01	0.02	0.23	-0.04
dSec_Bulk_3	-0.19	0.29	0.14	0.28	0.15	0.11	-0.11	-0.02	0.02	-0.03	0.10	-0.08	0.04	0.20	0.24	0.00	0.02	0.37	0.08	-0.03	0.09	0.20	-0.26
dTC_Bulk_6	0.06	0.29	0.13	0.24	-0.01	0.07	0.01	0.04	0.00	0.00	-0.03	0.05	-0.16	-0.09	-0.07	-0.03	0.01	0.00	0.01	-0.07	-0.04	0.13	0.04
dFX_USD_1	-0.26	-0.21	-0.04	0.03	-0.01	0.15	-0.14	0.06	-0.05	0.12	-0.05	-0.02	-0.16	0.01	-0.08	0.02	0.03	0.01	-0.08	0.01	0.10	-0.07	0.11
dFX_USD_2	-0.26	-0.08	-0.02	-0.06	0.02	-0.14	-0.10	0.02	0.01	-0.08	-0.08	-0.16	-0.10	-0.16	0.01	-0.11	-0.02	0.11	0.09	-0.07	-0.05	-0.16	-0.13
dFX_USD_6	0.01	-0.30	0.01	-0.21	0.10	-0.22	0.10	-0.05	0.08	-0.05	-0.03	-0.07	0.09	-0.12	0.09	-0.02	-0.11	-0.13	-0.02	-0.05	-0.05	-0.18	0.08
dLIBOR_4	-0.10	0.12	0.16	0.13	-0.09	-0.15	-0.01	-0.03	-0.03	0.02	0.00	-0.12	-0.10	0.13	0.02	-0.14	0.02	0.18	-0.18	-0.04	0.02	0.02	-0.16
dVIX_2	-0.06	0.01	0.03	-0.01	-0.02	0.12	-0.10	-0.04	0.04	0.07	-0.01	-0.03	-0.11	-0.04	-0.01	-0.11	-0.01	0.12	0.07	-0.15	0.00	0.11	-0.10
dVIX_3	0.08	0.02	-0.01	-0.07	-0.06	-0.16	0.06	0.08	-0.02	0.03	-0.09	0.04	-0.20	-0.11	-0.04	0.07	0.01	0.07	0.09	0.01	-0.09	-0.01	-0.12
dP_IO_2	-0.14	0.09	-0.05	0.26	0.15	0.09	0.01	-0.05	-0.11	0.03	-0.02	0.04	0.15	0.24	-0.01	-0.06	0.04	0.06	0.01	-0.01	0.00	0.23	0.08
dP_Coal_2	-0.16	0.12	0.03	0.24	0.04	0.00	0.09	-0.04	-0.07	-0.01	-0.02	0.13	0.01	0.13	0.00	-0.19	-0.01	0.19	0.02	-0.05	0.02	0.24	-0.15
dP_Coal_6	-0.11	0.01	0.17	-0.04	-0.15	-0.17	0.03	0.06	0.05	-0.10	-0.07	-0.05	-0.05	-0.08	0.03	-0.03	-0.18	0.07	-0.06	-0.15	-0.03	0.03	0.11
dP_Wheat_5	0.03	0.23	0.07	0.05	0.08	0.06	-0.06	0.06	-0.05	-0.04	0.00	0.08	0.09	-0.10	0.02	0.07	-0.05	0.09	0.01	0.10	0.01	0.06	-0.14
dP_Metals_5	-0.04	0.27	0.12	0.34	-0.01	0.18	-0.03	0.12	0.04	-0.04	0.03	0.08	-0.06	0.13	0.10	0.00	-0.01	0.17	-0.21	0.05	-0.09	0.13	-0.06
dP_Gold_5	0.11	0.21	0.07	0.13	0.04	0.08	0.03	-0.08	0.13	0.04	-0.07	-0.01	-0.06	0.05	0.03	-0.01	0.06	0.04	-0.08	0.19	-0.02	0.02	-0.01
dDBulk_Index_1	0.33	0.22	0.01	0.11	0.06	0.01	0.09	-0.05	-0.01	-0.03	0.03	-0.04	0.07	0.10	-0.01	-0.04	0.07	-0.02	0.03	0.03	0.01	0.17	-0.19
dDBulk_Index_2	0.16	0.10	0.05	0.15	0.10	-0.01	0.10	-0.02	0.02	0.03	0.08	0.01	0.18	0.08	0.09	-0.01	0.08	-0.01	-0.08	0.13	0.02	0.11	0.07
dDBulk_Index_3	-0.27	0.06	0.14	0.17	0.05	0.05	0.07	-0.08	-0.10	-0.09	0.08	-0.13	0.20	0.18	0.08	0.00	-0.08	0.14	0.06	0.03	0.14	0.18	0.01
dMSCI_W_6	0.02	0.26	0.11	0.28	0.01	0.19	0.00	0.01	0.02	0.06	0.07	0.09	-0.06	0.04	-0.05	0.03	0.05	0.14	0.02	0.01	0.03	0.09	0.08

Figure E.3: Correlation matrix (part 1) of significant variables in SBDI-models.

	dScrap_Bulk_4	dScrap_Bulk_6	dDel_Bulk_4	dNew_Bulk_6	dSec_Bulk_2	dSec_Bulk_3	dTC_Bulk_6	dFX_USD_1	dFX_USD_2	dFX_USD_6	dLIBOR_4	dVIX_2	dVIX_3	dP_IO_2	dP_Coal_2	dP_Coal_6	dP_Wheat_5	dP_Metals_5	dP_Gold_5	dDBulk_Index_1	dDBulk_Index_2	dDBulk_Index_3	dMSCI_W_6
dBDI																							
dSBDI_5																							
dSBDTI_3																							
dIP_OECD_3																							
dIP_Ch_2																							
dIP_I_6																							
dIO_A_Exp_1																							
dC_EU_Imp_2																							
dC_J_Imp_6																							
dC_A_Exp_6																							
dG_US_Exp_2																							
dG_US_Exp_6																							
dS_CH_Prod_2																							
dS_CH_Prod_3																							
dS_CH_Prod_4																							
dMS_Ch_2																							
dMS_Ch_6																							
dCPI_US_3																							
dCPI_Ch_3																							
dCPI_Ch_5																							
dFleet_Bulk_1																							
dOrder_Bulk_2																							
dScrap_Bulk_3																							
dScrap_Bulk_4	1.00																						
dScrap_Bulk_6	-0.06	1.00																					
dDel_Bulk_4	0.22	0.12	1.00																				
dNew_Bulk_6	-0.12	0.33	0.01	1.00																			
dSec_Bulk_2	-0.06	0.03	0.02	0.00	1.00																		
dSec_Bulk_3	-0.03	-0.05	0.00	0.02	0.49	1.00																	
dTC_Bulk_6	-0.25	-0.11	-0.02	-0.17	0.06	0.11	1.00																
dFX_USD_1	0.09	-0.10	-0.20	-0.03	-0.04	0.05	-0.09	1.00															
dFX_USD_2	0.11	0.11	0.22	-0.10	-0.19	-0.04	-0.20	0.04	1.00														
dFX_USD_6	0.12	0.01	0.02	0.04	-0.04	-0.14	-0.19	-0.06	-0.05	1.00													
dLIBOR_4	-0.05	0.17	-0.05	0.13	-0.07	-0.07	0.04	-0.19	0.14	-0.12	1.00												
dVIX_2	0.13	0.06	0.07	-0.02	-0.17	0.00	-0.08	0.09	0.24	-0.09	0.07	1.00											
dVIX_3	-0.12	0.08	0.00	0.04	-0.25	-0.17	0.05	-0.09	0.08	-0.03	0.03	-0.04	1.00										
dP_IO_2	0.01	-0.14	0.00	-0.16	0.60	0.31	-0.08	0.02	-0.07	-0.05	-0.17	-0.10	-0.26	1.00									
dP_Coal_2	-0.06	-0.09	-0.06	-0.22	0.44	0.27	0.15	-0.03	-0.13	-0.15	0.01	0.05	-0.16	0.35	1.00								
dP_Coal_6	-0.04	-0.03	0.03	-0.13	-0.11	-0.06	0.27	0.09	0.06	-0.13	0.15	0.17	0.08	-0.16	0.08	1.00							
dP_Wheat_5	-0.05	-0.06	0.06	-0.05	0.06	0.15	0.10	-0.09	-0.09	-0.15	0.04	0.00	-0.14	0.05	0.09	0.05	1.00						
dP_Metals_5	-0.18	-0.01	0.03	-0.14	0.09	0.21	0.27	-0.02	-0.13	-0.36	0.14	-0.01	0.11	0.09	0.21	0.19	0.18	1.00					
dP_Gold_5	0.00	0.13	0.10	0.10	0.06	0.23	0.05	-0.09	-0.10	0.10	-0.01	-0.03	-0.08	0.03	-0.04	-0.13	0.14	0.18	1.00				
dDBulk_Index_1	-0.14	0.09	0.02	0.20	0.20	-0.03	0.10	-0.43	-0.14	-0.07	0.18	-0.17	0.05	0.17	0.10	-0.10	0.07	0.02	0.00	1.00			
dDBulk_Index_2	-0.18	-0.04	-0.03	0.16	0.45	0.20	0.21	0.00	-0.43	0.01	-0.13	-0.52	-0.16	0.32	0.25	-0.15	-0.01	0.14	0.07	0.21	1.00		
dDBulk_Index_3	0.07	-0.13	0.05	-0.21	0.53	0.45	0.06	0.02	0.00	-0.05	-0.08	0.04	-0.52	0.44	0.38	0.07	0.17	-0.03	-0.02	0.02	0.21	1.00	
dMSCI_W_6	-0.14	0.00	0.03	-0.11	0.10	0.10	0.27	0.06	-0.08	-0.50	0.09	0.09	0.02	0.14	0.14	0.15	0.04	0.51	0.04	-0.01	0.11	0.01	1.00

Figure E.4: Correlation matrix (part 2) of significant variables in SBDI-models.

	dBDTI	dBDI_5	dBDTI_2	dBDTI_4	dGDP_W_1	dIP_OECD_4	dIP_I_1	dIP_I_2	dIP_US_1	dOil_P_G_5	dOil_P_ME_2	dOil_P_ME_4	dO_Ch_Imp_2	dO_Ch_Imp_3	dO_US_Exp_4	dO_AG_Exp_5	dO_AG_Exp_6	dMS_Ch_3	dCPI_US_3	dCPI_Ch_1
dBDTI	1.00																			
dBDI_5	-0.10	1.00																		
dBDTI_2	-0.05	0.12	1.00																	
dBDTI_4	-0.10	0.15	-0.06	1.00																
dGDP_W_1	0.15	0.31	0.09	0.00	1.00															
dIP_OECD_4	0.14	0.23	0.21	0.14	0.39	1.00														
dIP_I_1	-0.04	0.09	-0.04	0.07	0.01	0.04	1.00													
dIP_I_2	0.14	-0.02	0.06	-0.05	0.09	0.01	-0.52	1.00												
dIP_US_1	0.10	0.10	0.15	0.04	0.37	0.46	0.05	-0.03	1.00											
dOil_P_G_5	0.06	-0.09	-0.15	0.05	0.06	0.07	0.04	-0.15	0.03	1.00										
dOil_P_ME_2	0.19	0.08	0.24	0.02	0.13	0.16	-0.03	-0.03	0.10	-0.02	1.00									
dOil_P_ME_4	0.09	0.15	0.19	0.24	-0.02	0.06	-0.15	0.01	-0.10	-0.23	0.04	1.00								
dO_Ch_Imp_2	-0.27	0.06	-0.04	-0.04	0.06	0.09	-0.03	0.14	0.07	-0.06	-0.13	-0.13	1.00							
dO_Ch_Imp_3	0.22	-0.01	0.13	0.10	0.06	0.01	-0.03	-0.03	0.01	-0.10	0.19	0.19	-0.59	1.00						
dO_US_Exp_4	0.07	-0.03	0.01	-0.10	-0.03	0.01	-0.10	-0.05	-0.01	0.00	-0.03	-0.06	0.14	-0.19	1.00					
dO_AG_Exp_5	-0.03	0.05	0.00	0.02	-0.02	0.08	-0.10	-0.07	0.02	0.32	-0.03	0.01	-0.07	0.01	0.07	1.00				
dO_AG_Exp_6	-0.09	0.01	0.02	0.15	-0.01	-0.04	0.09	-0.10	-0.06	0.14	0.06	0.03	-0.08	-0.08	0.05	-0.11	1.00			
dMS_Ch_3	-0.03	-0.05	-0.01	0.00	-0.02	0.05	-0.03	0.00	0.09	0.04	0.00	0.07	-0.01	0.06	0.11	-0.03	0.06	1.00		
dCPI_US_3	0.14	0.18	0.08	0.07	0.26	0.13	0.01	-0.05	0.07	0.10	0.18	0.14	0.08	0.08	0.01	0.03	-0.09	-0.06	1.00	
dCPI_Ch_1	0.09	0.02	-0.08	-0.03	0.03	-0.13	-0.12	0.12	-0.09	-0.14	-0.05	-0.11	-0.08	0.00	-0.08	0.01	0.01	-0.03	-0.13	1.00
dCPI_Ch_2	0.02	0.02	-0.06	0.17	-0.03	0.04	0.01	-0.12	0.17	-0.06	0.02	0.04	0.00	-0.08	0.15	-0.18	0.02	0.43	-0.13	-0.30
dCPI_Ch_3	0.11	0.05	0.08	-0.08	0.01	0.03	0.02	0.01	-0.07	0.12	0.12	-0.05	-0.07	0.01	-0.08	0.10	-0.19	-0.42	0.09	-0.18
dCPI_Ch_4	0.02	-0.05	0.02	-0.06	0.03	-0.03	-0.01	0.02	-0.05	-0.03	-0.02	0.01	0.17	-0.07	0.03	-0.02	0.10	0.04	0.07	-0.01
dOrder_Tank_6	-0.04	-0.11	0.02	0.00	-0.10	-0.08	0.00	0.01	-0.11	0.01	-0.16	-0.11	0.15	-0.10	0.06	-0.04	-0.06	0.05	0.03	0.01
dNew_Tank_2	0.19	0.14	0.21	0.00	0.09	0.08	0.03	-0.01	0.15	-0.09	0.21	0.02	-0.16	0.22	-0.07	0.05	-0.07	-0.04	0.02	0.01
dSec_Tank_1	0.18	-0.07	0.15	-0.01	-0.06	0.01	0.11	-0.06	0.12	-0.05	0.06	-0.07	0.01	-0.06	-0.14	0.01	-0.08	-0.11	-0.13	-0.02
dFuel_5	0.22	0.28	0.08	0.22	0.09	0.25	-0.02	-0.02	0.07	-0.01	0.12	0.10	-0.16	0.16	-0.03	0.13	-0.01	0.07	0.15	0.12
dTC_Tank_2	-0.12	-0.02	0.42	0.09	0.09	0.16	-0.04	-0.02	0.18	-0.01	0.30	0.15	0.10	-0.11	0.15	-0.01	0.09	-0.02	0.11	-0.13
dTC_Tank_6	0.00	0.02	0.05	-0.12	-0.05	0.02	-0.12	0.01	-0.16	0.05	-0.02	0.10	-0.03	0.05	-0.08	0.19	0.18	0.02	-0.07	-0.16
dOil_1	0.28	0.16	-0.04	0.02	0.35	0.13	0.11	0.16	0.06	0.07	0.07	-0.08	-0.03	0.03	0.00	0.08	-0.05	0.10	-0.10	0.04
dFX_USD_JPN_	0.12	-0.11	-0.08	0.01	-0.03	-0.01	-0.15	0.11	-0.09	0.03	-0.09	-0.04	-0.02	-0.07	0.05	0.08	-0.23	-0.04	-0.12	-0.11
dFX_USD_JPN_	-0.11	0.01	-0.11	-0.03	0.03	0.04	0.17	-0.06	0.12	-0.02	0.00	-0.07	0.13	-0.11	-0.12	0.01	0.02	-0.02	-0.04	-0.21
dLIBOR_1	0.17	0.01	-0.25	-0.04	0.08	-0.01	-0.20	0.19	-0.21	0.06	0.07	0.06	0.06	-0.11	0.11	0.05	-0.08	0.02	-0.01	0.05
dLIBOR_3	0.17	0.00	0.17	-0.25	0.05	0.03	-0.10	0.21	0.07	-0.20	0.06	0.08	0.12	-0.03	0.19	0.05	-0.07	-0.08	0.08	0.12
dCont_Oil_3	0.00	-0.10	-0.22	0.08	-0.14	-0.07	-0.07	0.02	-0.07	0.01	0.04	-0.03	-0.20	0.12	0.04	0.04	0.13	0.09	-0.15	0.00
dCont_Oil_6	0.16	-0.11	-0.12	-0.05	-0.07	-0.07	0.08	-0.05	-0.06	0.08	0.09	-0.11	-0.02	0.03	-0.03	0.04	-0.07	-0.03	0.08	-0.09
dP_Metals_1	0.06	0.08	0.04	0.06	0.47	0.26	-0.03	0.18	0.16	0.00	0.12	0.02	0.16	-0.03	0.01	-0.09	-0.11	0.00	0.04	0.03
dP_Metals_4	0.00	0.17	0.10	0.05	0.40	0.36	0.11	0.05	0.20	0.16	0.19	0.03	0.06	0.09	-0.09	0.01	0.07	0.01	0.31	0.00
dP_Metals_6	-0.09	0.20	0.01	0.10	0.27	0.35	0.10	-0.10	0.21	-0.07	0.08	0.18	0.04	-0.04	0.00	0.02	0.02	0.01	0.10	-0.08
dTank_Index_6	-0.07	0.36	0.04	0.09	0.18	0.19	0.02	-0.06	0.02	-0.05	-0.06	0.11	0.03	0.05	0.00	0.13	0.04	-0.02	0.10	0.03

Figure E.5: Correlation matrix (part 1) of significant variables in BDTI-models.

	dCPI_Ch_2	dCPI_Ch_3	dCPI_Ch_4	dOrder_Tank_6	dNew_Tank_2	dSec_Tank_1	dFuel_5	dTC_Tank_2	dTC_Tank_6	dOil_1
dBDTI										
dBDI_5										
dBDTI_2										
dBDTI_4										
dGDP_W_1										
dIP_OECD_4										
dIP_I_1										
dIP_I_2										
dIP_US_1										
dOil_P_G_5										
dOil_P_ME_2										
dOil_P_ME_4										
dO_Ch_Imp_2										
dO_Ch_Imp_3										
dO_US_Exp_4										
dO_AG_Exp_5										
dO_AG_Exp_6										
dMS_Ch_3										
dCPI_US_3										
dCPI_Ch_1										
dCPI_Ch_2	1.00									
dCPI_Ch_3	-0.30	1.00								
dCPI_Ch_4	-0.17	-0.31	1.00							
dOrder_Tank_6	0.00	-0.03	0.00	1.00						
dNew_Tank_2	-0.01	0.06	-0.04	-0.10	1.00					
dSec_Tank_1	-0.06	0.02	-0.04	0.00	-0.05	1.00				
dFuel_5	0.00	-0.09	-0.07	-0.08	0.05	0.06	1.00			
dTC_Tank_2	0.00	0.08	0.02	0.09	0.13	-0.06	-0.08	1.00		
dTC_Tank_6	-0.06	0.05	0.09	-0.03	-0.07	0.15	-0.10	-0.07	1.00	
dOil_1	-0.01	-0.03	-0.09	-0.10	0.12	0.04	-0.04	-0.03	-0.03	1.00
dFX_USD_JPN_	0.11	0.01	-0.02	0.08	-0.07	0.12	0.02	-0.07	0.07	-0.05
dFX_USD_JPN_	0.21	-0.04	-0.12	0.03	-0.02	0.07	0.13	-0.03	0.05	-0.04
dLIBOR_1	-0.04	0.06	-0.01	0.03	0.09	-0.18	-0.02	-0.11	-0.02	0.14
dLIBOR_3	-0.18	0.05	-0.04	0.01	0.02	0.15	-0.15	0.10	0.06	0.10
dCont_Oil_3	0.09	-0.09	-0.11	0.09	-0.08	0.04	0.07	-0.12	0.08	0.10
dCont_Oil_6	-0.03	0.05	0.02	-0.11	0.07	0.01	-0.11	-0.11	0.00	-0.03
dP_Metals_1	0.02	-0.01	0.10	-0.01	-0.05	0.00	-0.05	0.05	-0.04	0.35
dP_Metals_4	-0.20	0.01	0.04	0.14	0.02	-0.06	0.16	0.05	0.02	0.16
dP_Metals_6	0.06	-0.01	-0.21	0.12	-0.01	0.00	0.15	0.13	0.06	-0.09
dTank_Index_6	-0.03	0.03	-0.04	0.11	0.03	-0.09	0.39	0.01	0.18	0.00

	dFX_USD_JPN_	dFX_USD_JPN_	dLIBOR_1	dLIBOR_3	dCont_Oil_3	dCont_Oil_6	dP_Metals_1	dP_Metals_4	dP_Metals_6	dTank_Index_6
dFX_USD_JPN_	1.00									
dFX_USD_JPN_	0.04	1.00								
dLIBOR_1	0.18	-0.13	1.00							
dLIBOR_3	0.08	-0.07	-0.08	1.00						
dCont_Oil_3	0.00	-0.05	0.09	-0.03	1.00					
dCont_Oil_6	-0.26	-0.08	0.02	0.02	0.09	1.00				
dP_Metals_1	-0.01	-0.02	0.15	-0.01	-0.02	0.01	1.00			
dP_Metals_4	-0.12	-0.02	-0.12	0.14	-0.10	-0.01	0.11	1.00		
dP_Metals_6	0.02	-0.01	0.04	-0.12	0.00	-0.12	0.07	0.16	1.00	
dTank_Index_6	-0.12	0.00	-0.06	-0.15	0.07	0.00	0.00	0.09	0.30	1.00

Figure E.6: Correlation matrix (part 2) of significant variables in BDTI-models.

	dSBDTI	dSBDI_3	dSBDI_5	dSBDTI_2	dIP_OECD_2	dIP_I_4	dOil_P_ME_4	dOil_P_ME_6	dO_AG_Exp_2	dO_AG_Exp_5	dMS_US_1	dCPI_US_5	dFleet_Tank_1	dFleet_Tank_2	dDel_Tank_5	dSec_Tank_1	dFuel_5	dOil_1	dOil_5	dOil_6	dFX_USD_JPN_1	dFX_USD_JPN_4	dCont_Oil_6	dTank_Index_6
dSBDTI	1.00																							
dSBDI_3	0.14	1.00																						
dSBDI_5	-0.08	-0.04	1.00																					
dSBDTI_2	-0.08	0.11	0.14	1.00																				
dIP_OECD_2	0.26	0.23	0.29	0.14	1.00																			
dIP_I_4	-0.01	0.02	0.16	0.15	0.18	1.00																		
dOil_P_ME_4	0.08	0.02	0.18	0.10	-0.03	-0.06	1.00																	
dOil_P_ME_6	0.01	-0.07	0.02	0.10	-0.11	-0.02	0.06	1.00																
dO_AG_Exp_2	0.14	0.00	0.09	0.38	0.10	-0.01	-0.04	0.05	1.00															
dO_AG_Exp_5	-0.01	0.05	0.01	0.00	-0.06	0.01	0.01	0.08	0.05	1.00														
dMS_US_1	-0.21	-0.31	0.06	-0.06	-0.05	0.09	-0.14	0.05	0.05	0.00	1.00													
dCPI_US_5	-0.09	-0.23	0.23	0.07	-0.01	-0.07	0.19	0.16	0.03	0.14	0.08	1.00												
dFleet_Tank_1	0.15	0.02	-0.04	0.13	-0.05	0.15	-0.02	0.17	0.03	0.07	-0.04	-0.03	1.00											
dFleet_Tank_2	-0.01	-0.02	-0.05	-0.08	0.03	-0.12	-0.12	-0.12	0.01	-0.13	0.01	0.03	-0.53	1.00										
dDel_Tank_5	0.07	-0.04	-0.08	-0.03	-0.03	-0.06	-0.08	-0.06	-0.06	-0.02	0.01	0.03	-0.13	0.07	1.00									
dSec_Tank_1	0.27	-0.12	-0.06	0.09	0.00	0.05	-0.07	0.02	0.19	0.01	-0.04	0.08	-0.02	-0.02	-0.06	1.00								
dFuel_5	0.23	-0.11	0.28	0.05	0.21	-0.01	0.10	0.05	0.10	0.13	0.10	0.43	-0.01	-0.07	-0.04	0.06	1.00							
dOil_1	0.35	0.15	0.15	0.04	0.19	0.13	-0.08	-0.04	0.06	0.08	-0.13	-0.05	-0.01	-0.01	0.01	0.04	-0.04	1.00						
dOil_5	0.12	0.02	0.27	0.09	0.22	-0.01	0.11	0.08	0.17	0.11	0.06	0.36	0.06	-0.02	-0.08	0.00	0.75	-0.04	1.00					
dOil_6	-0.08	-0.16	0.17	0.05	0.03	0.06	0.18	0.21	-0.04	0.11	0.06	0.70	-0.12	0.06	-0.03	-0.02	0.27	-0.03	0.16	1.00				
dFX_USD_JPN_1	0.15	-0.02	-0.15	-0.05	-0.08	0.07	-0.04	0.06	0.05	0.08	0.09	0.08	0.00	0.00	-0.04	0.12	0.02	-0.05	0.05	0.11	1.00			
dFX_USD_JPN_4	-0.07	0.07	0.01	-0.08	0.02	-0.19	-0.07	0.19	-0.01	0.01	0.05	0.09	-0.06	0.04	0.02	0.07	0.13	-0.04	-0.07	0.10	0.04	1.00		
dCont_Oil_6	0.14	-0.01	-0.10	-0.09	-0.01	-0.03	-0.11	-0.15	-0.06	0.04	-0.08	-0.33	0.22	-0.20	0.00	0.01	-0.11	-0.03	-0.02	-0.65	-0.26	-0.08	1.00	
dTank_Index_6	-0.09	-0.22	0.37	0.04	0.12	0.11	0.11	0.03	-0.15	0.13	-0.03	0.46	-0.02	-0.03	0.01	-0.09	0.39	0.00	0.34	0.38	-0.12	0.00	0.00	1.00

Figure E.7: Correlation matrix of significant variables in SBDTI-models.

	dBDTI	dOil_1	dOil_2	dOil_3	dOil_4	dOil_5	dOil_6	dO_Ch_Imp_1	dO_Ch_Imp_2	dO_Ch_Imp_3	dO_Ch_Imp_4	dO_Ch_Imp_5	dO_Ch_Imp_6
dBDTI	1.00												
dOil_1	0.28	1.00											
dOil_2	0.06	0.18	1.00										
dOil_3	0.11	-0.02	0.17	1.00									
dOil_4	0.10	0.03	-0.03	0.17	1.00								
dOil_5	0.11	-0.04	0.04	-0.03	0.17	1.00							
dOil_6	-0.06	-0.03	-0.04	0.04	-0.03	0.16	1.00						
dO_Ch_Imp_1	0.14	0.06	0.15	-0.01	-0.02	-0.04	-0.04	1.00					
dO_Ch_Imp_2	-0.27	-0.03	0.04	0.15	0.01	-0.03	-0.04	-0.61	1.00				
dO_Ch_Imp_3	0.22	0.03	-0.05	0.05	0.16	0.00	-0.03	0.22	-0.59	1.00			
dO_Ch_Imp_4	-0.04	0.08	0.04	-0.04	0.05	0.16	0.00	-0.05	0.22	-0.60	1.00		
dO_Ch_Imp_5	-0.08	-0.03	0.10	0.03	-0.06	0.06	0.17	-0.13	-0.07	0.21	-0.58	1.00	
dO_Ch_Imp_6	0.13	-0.04	-0.02	0.10	0.02	-0.06	0.06	0.20	-0.14	-0.07	0.21	-0.57	1.00

Figure E.8: Correlation matrix of oil prices and Chinese oil imports including lagged variables.

Appendix F Comparison of seasonal and non-seasonal models

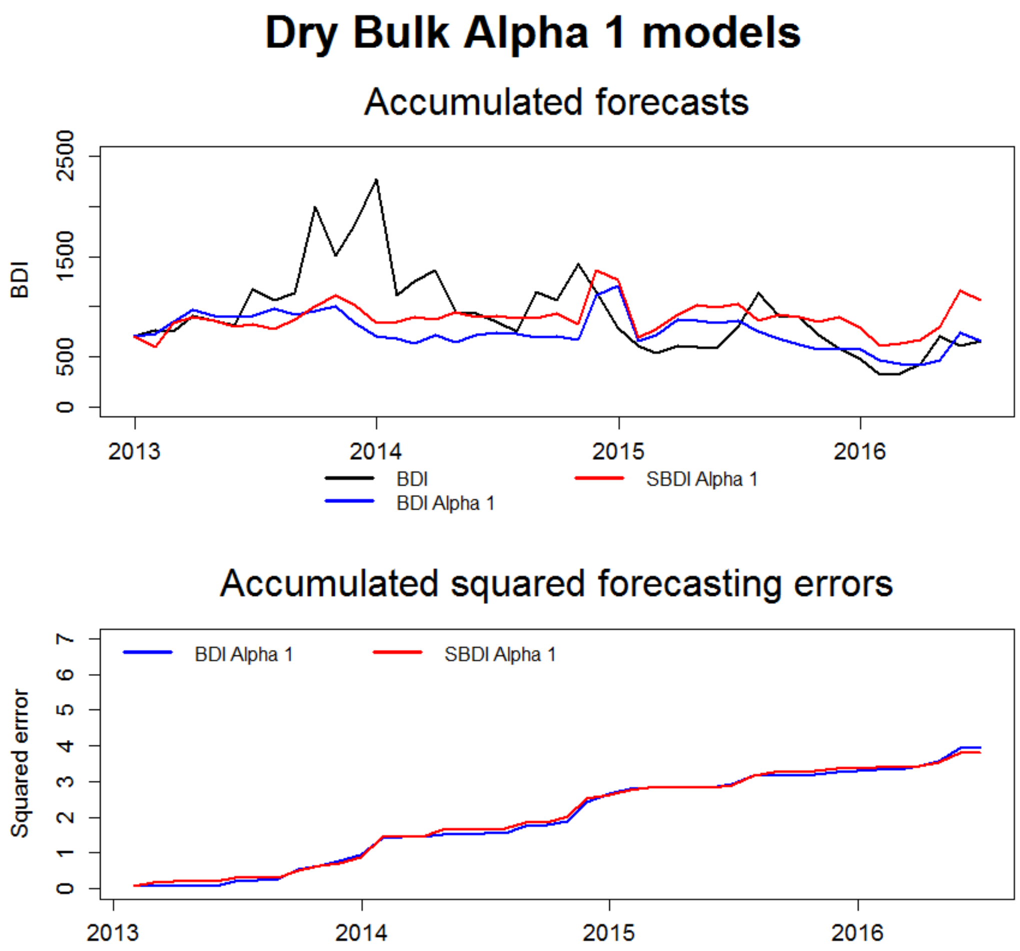

Figure F.1: Accumulated forecasts (top) and accumulated squared forecasting errors for Dry Bulk Alpha$_1$ models.

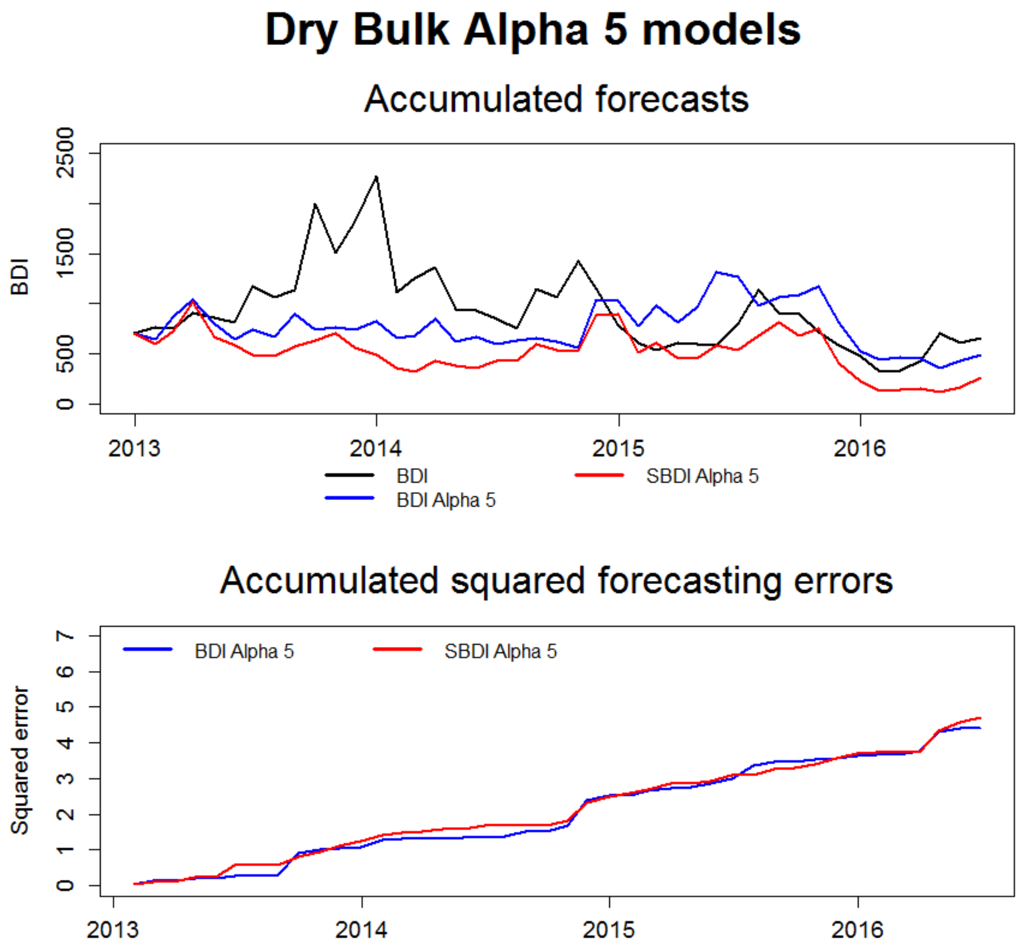

Figure F.2: Accumulated forecasts (top) and accumulated squared forecasting errors for Dry Bulk Alpha$_5$ models.

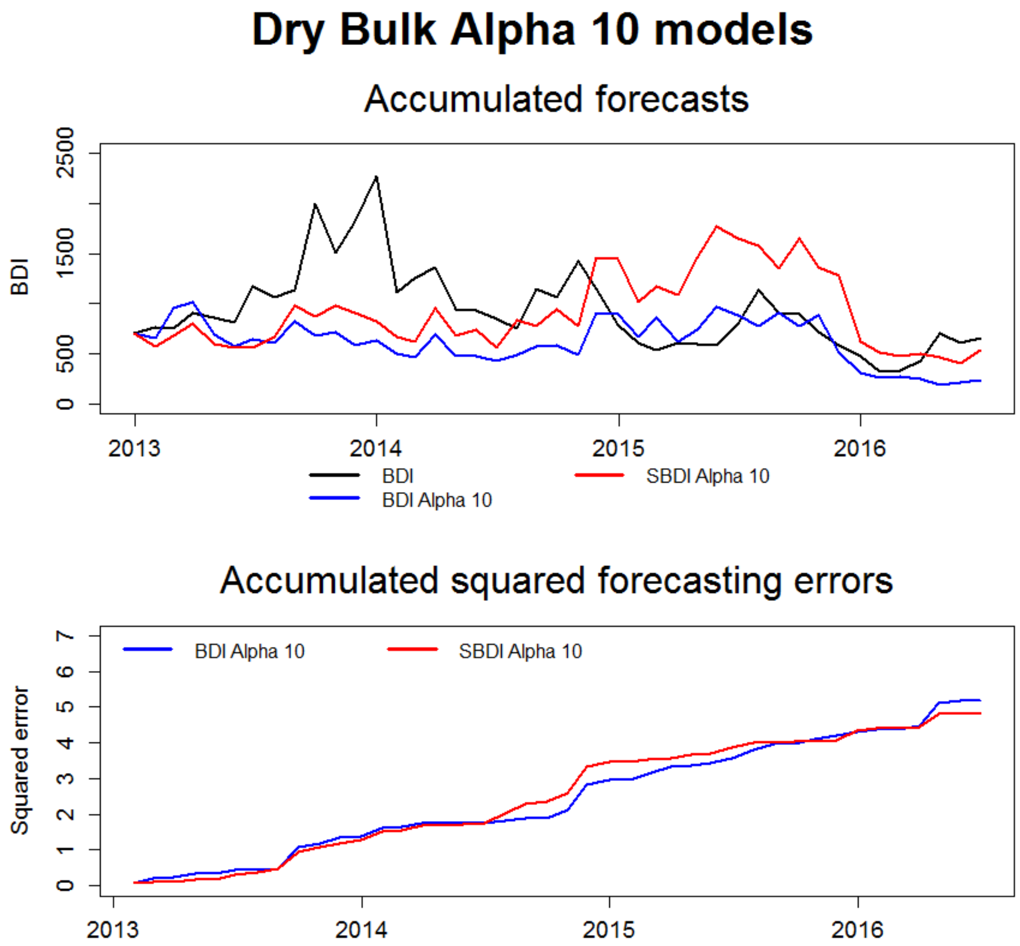

Figure F.3: Accumulated forecasts (top) and accumulated squared forecasting errors for Dry Bulk Alpha_{10} models.

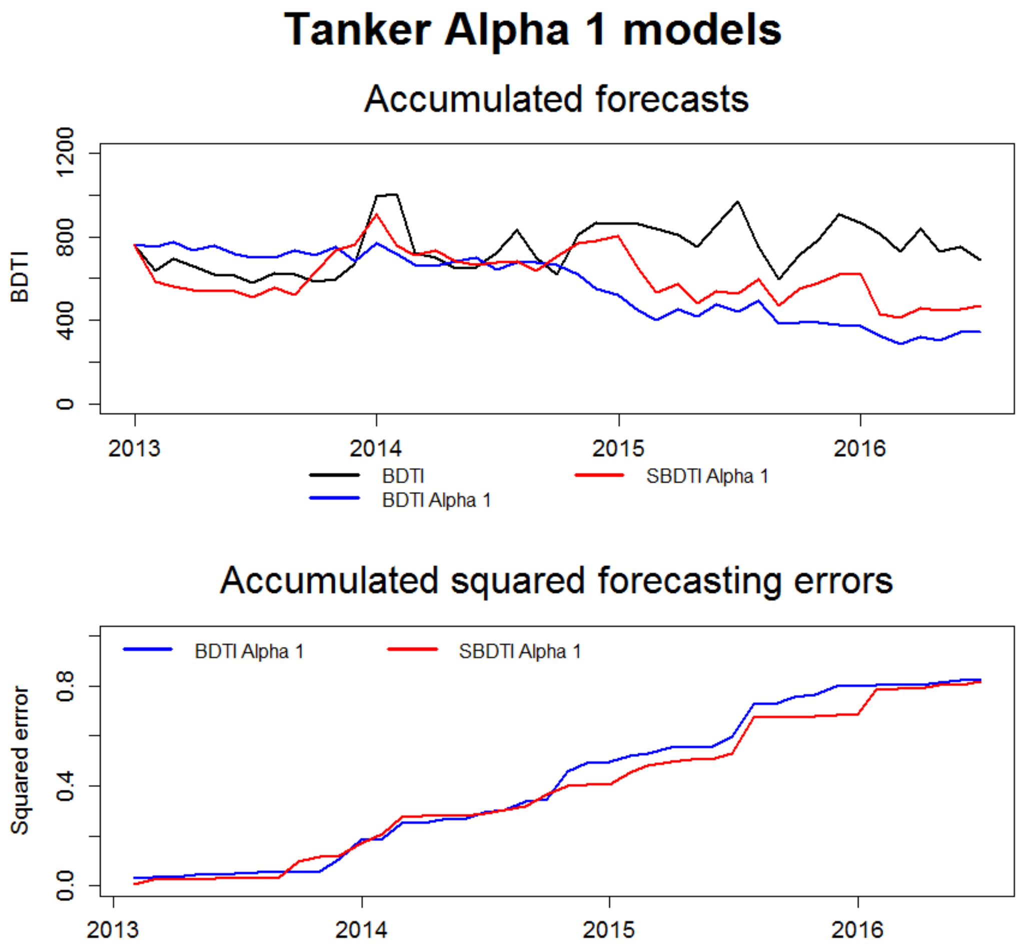

Figure F.4: Accumulated forecasts (top) and accumulated squared forecasting errors for Dry Bulk Alpha$_1$ models.

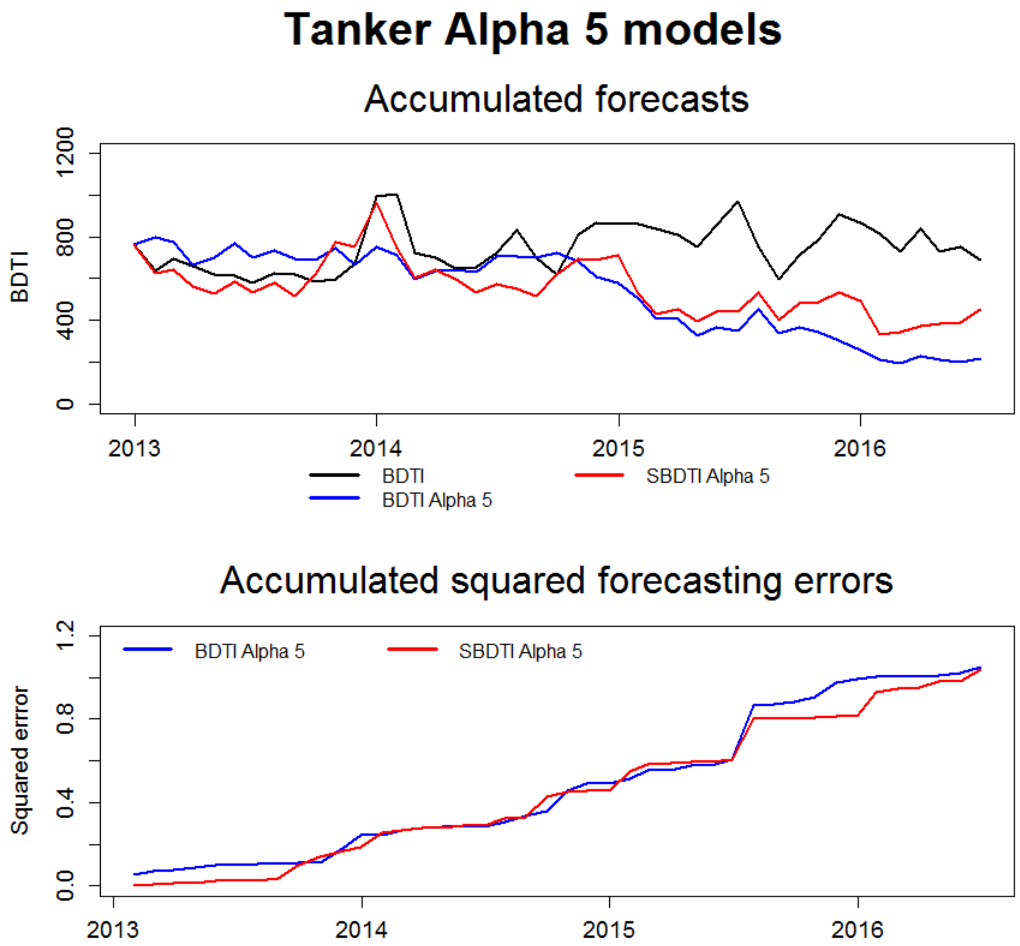

Figure F.5: Accumulated forecasts (top) and accumulated squared forecasting errors for Dry Bulk Alpha$_5$ models.

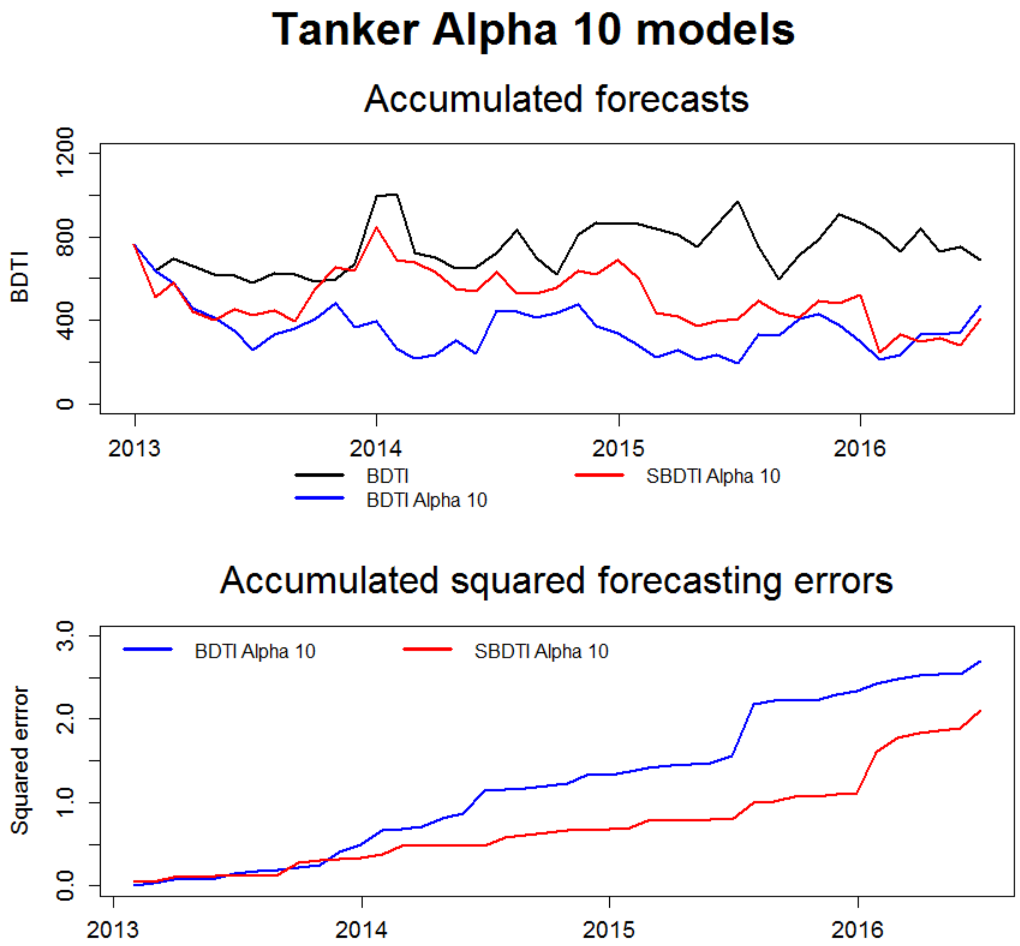

Figure F.6: Accumulated forecasts (top) and accumulated squared forecasting errors for Dry Bulk Alpha_{10} models.

Appendix G Autocorrelation functions and residual plots

This section provides an overview of the model's autocorrelation functions and residual plots. The latter subsection provides an illustration of each models' normality probability plot, fit, historgram and order, along with a discussion on how these observations relates to the qualitative approach (link between residual plots and OLS assumptions) "checklist" laid out in section 5.6.2.

Appendix G.1. ACF-plots

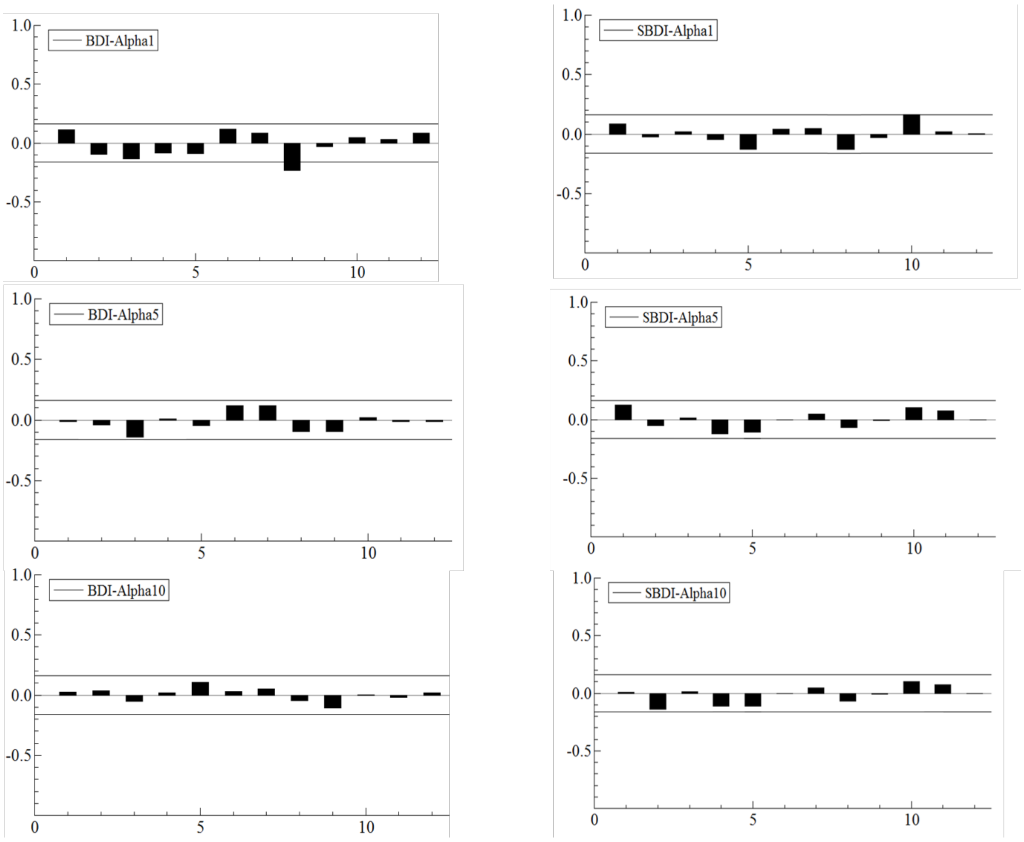

Figure G.1: ACF-plots for all six dry bulk models, based on the in-sample observations from from Jul-2000 to Dec-2012.

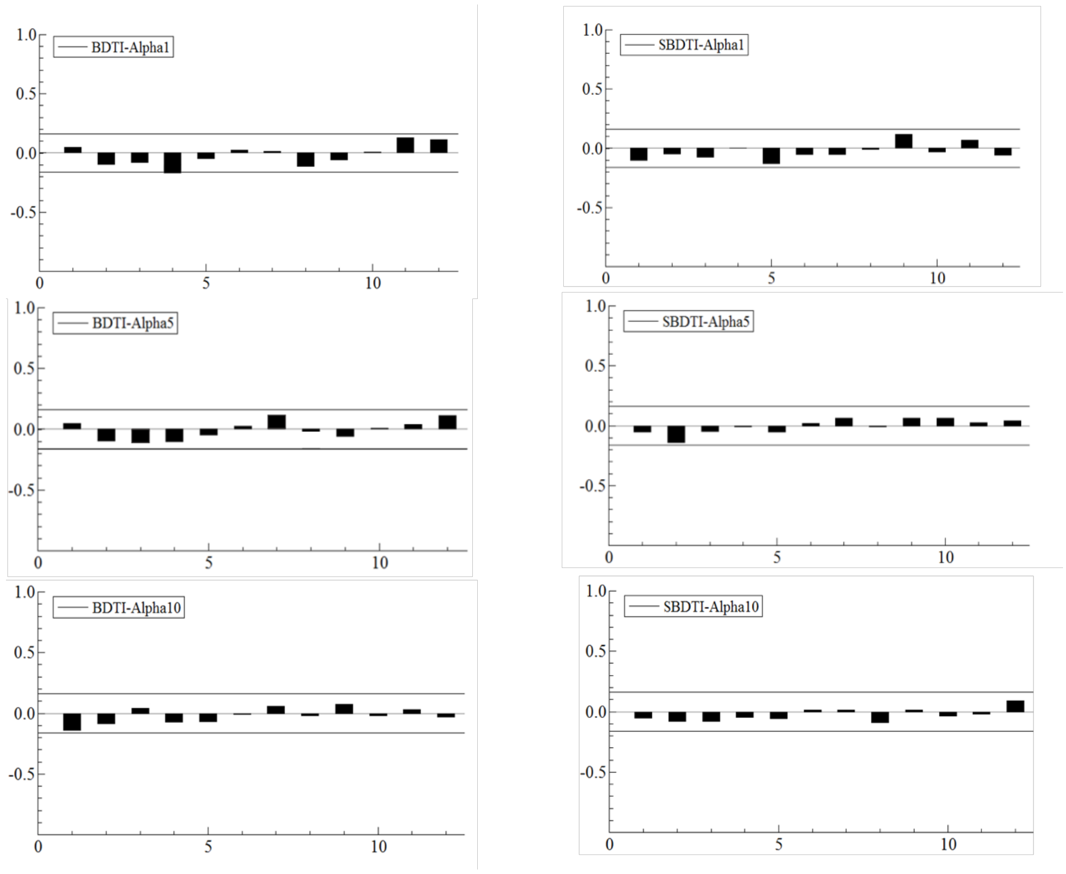

Figure G.2: ACF-plots for all six tanker models, based on the in-sample observations from from Jul-2000 to Dec-2012.

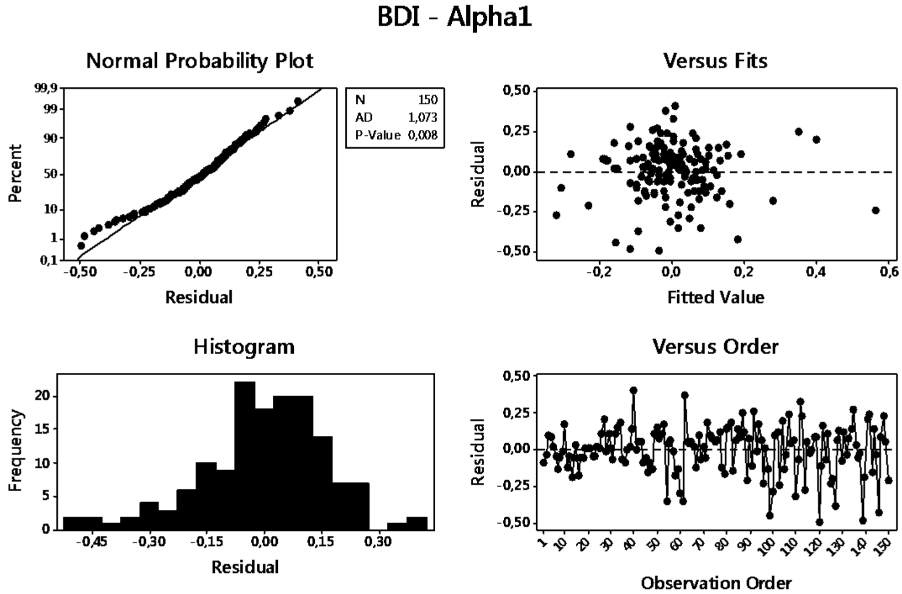

Figure G.3: Residual plot for BDI Alpha$_1$ model

i) There is no clear pattern/trend in the residuals versus fit- diagram, and consequently there is little reason to question the linearity assumption.

ii) The ACF- plot shows that there is significant negative autocorrelation in lag 8. Furthermore, there is a wave-like autocorrealtion pattern across lags, that may hint to seasonality effects that are not accounted for.

iii) From the residuals versus fit diagram, it is apparent that the residuals seem to be approximately evenly distributed along the horizontal axis, thus there is no clear evidence of non-constant variance.

iv) From both normality plots, tendencies to a negative skew is apparent, i.e. a "fatter" negative tail causing the mean to be less than the median. This outcome is not surprising, as the distribution of the dry bulk time series also is characterised by large negative outliers. Even though the residual distribution deviates from the normal, these deviation are not severe, as the asymmetry is not vast and we recognise the bell shape. In the relevant sample range, the t-distribution will accordingly be robust.

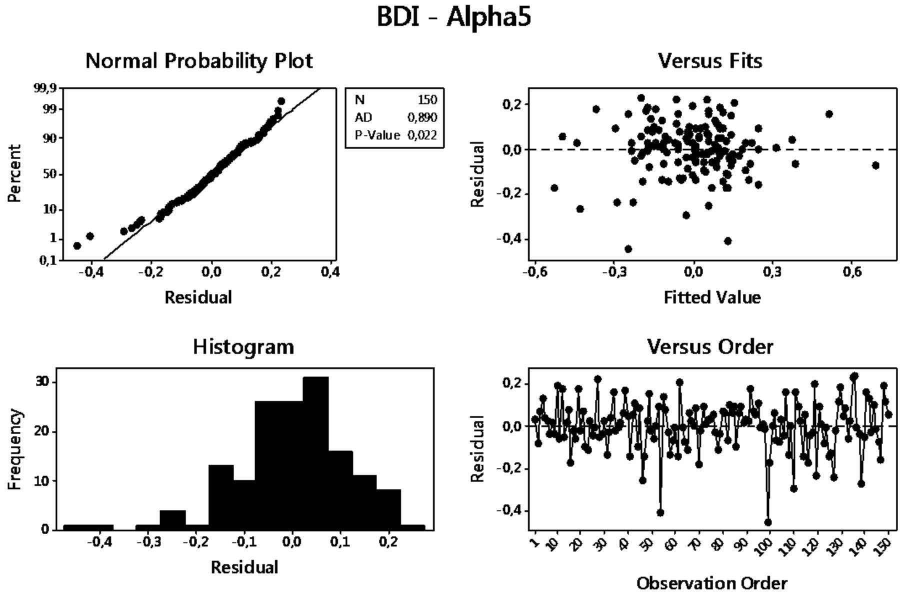

Figure G.4: Residual plot for BDI Alpha$_5$ model

i) There is no clear pattern/trend in the residuals versus fit- diagram, and consequently there is little reason to question the linearity assumption.

ii) Though no significant autocorrelation is detected, the ACF- plot shows small tendencies to a wave-like autocorrealtion across lags, that may hint to seasonality effects that are not accounted for.

iii) From the residuals versus fit diagram, it is apparent that the residuals seem to be approximately evenly distributed along the horizontal axis, thus there is no clear evidence of non-constant variance.

iv) From both normality plots, tendencies to a negative skew is apparent, i.e. a "fatter" negative tail causing the mean to be less than the median. This outcome is not surprising, as the distribution of the dry bulk time series also is characterised by large negative outliers. Even though the residual distribution deviates from the normal, these deviation are not severe, as the asymmetry is not vast and we recognise the bell shape. In the relevant sample range, the t-distribution will accordingly be robust.

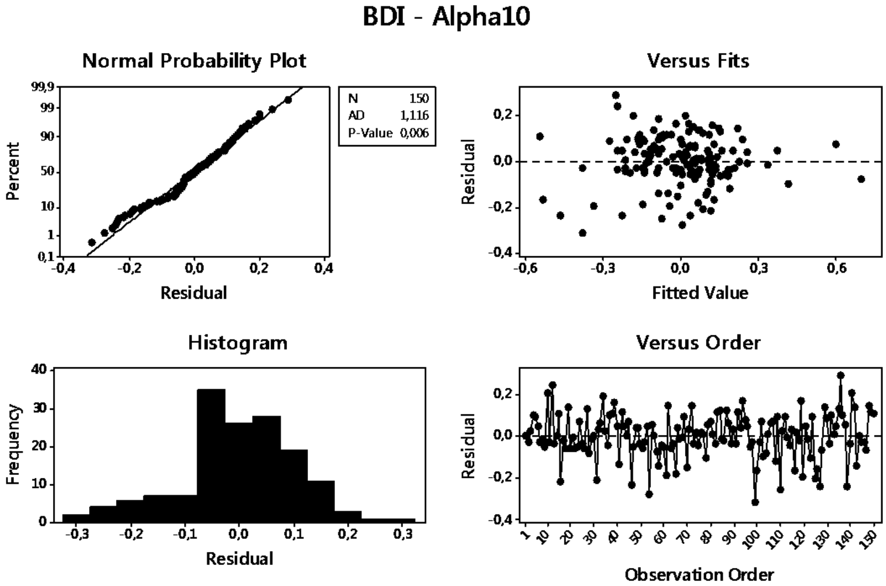

Figure G.5: Residual plot for BDI Alpha_{10} model

i) There is no clear pattern/trend in the residuals versus fit- diagram, and consequently there is little reason to question the linearity assumption.

ii) Though no significant autocorrelation is detected, the ACF- plot shows small tendencies to a wave-like autocorrealtion across lags, that may hint to seasonality effects that are not accounted for.

iii) From the residuals versus fit diagram, we observe a slightly larger spread in the residuals associated with negative fitted values relative to the positive fitted values. The pattern is more pronounced for larger negative and positive values. However, there are also more points to the far left than to the far right, which makes it harder to draw inferences. Thus, there is not sufficient evidence to claim there is non-constant variance.

iv) From the histogram, we observe asymmetries in the central region of the normal distribution, along with a "fat" negative tail. This outcome is not surprising, as the distribution of the dry bulk time series also is characterised by large negative outliers. Even though the residual distribution deviates from the normal, these deviation are not severe, as the asymmetry is not extreme and we recognise the bell shape. In the relevant sample range, the t-distribution will accordingly be robust.

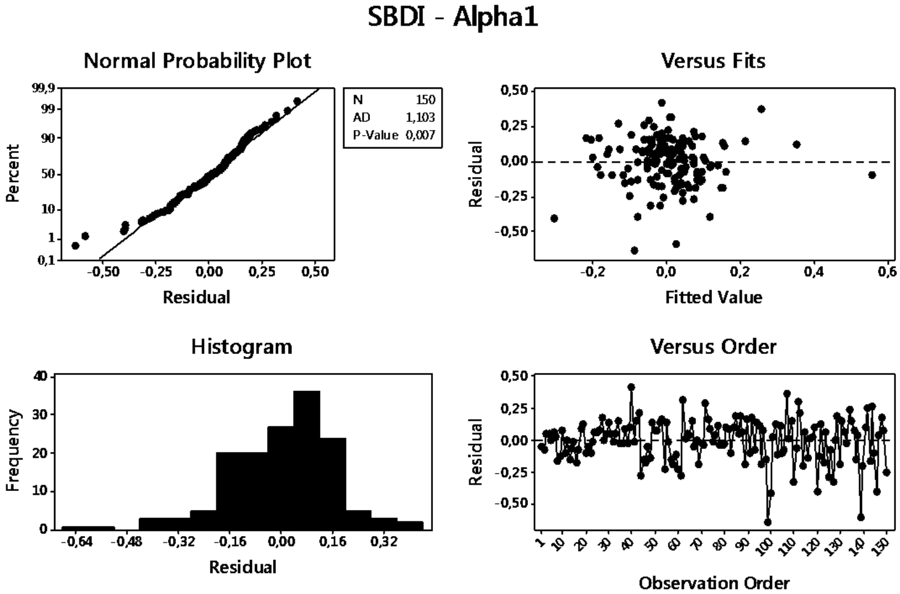

Figure G.6: Residual plot for SBDI Alpha_1 model

i) There is no clear pattern/trend in the residuals versus fit- diagram, and consequently there is little reason to question the linearity assumption.

ii) No lags are found to be significant, though the ACF- plot reveals that lag five, eight and ten are close.

iii) From the residuals versus fit diagram, it is apparent that the residuals seem to be approximately evenly distributed along the horizontal axis, thus there is no clear evidence of non-constant variance.

iv) From both normality plots, we observe an apparent negative skew, i.e. a "fatter" negative tail causing the mean to be less than the median. This outcome is not surprising, as the distribution of the seasonal dry bulk time series, similarly to the non-seasonalised dry bulk series, is characterised by large negative outliers. Even though the residual distribution deviates from the normal, these deviation are not severe, as the asymmetry is not extreme and we recognise the bell shape. In the relevant sample range, the t-distribution will accordingly be robust.

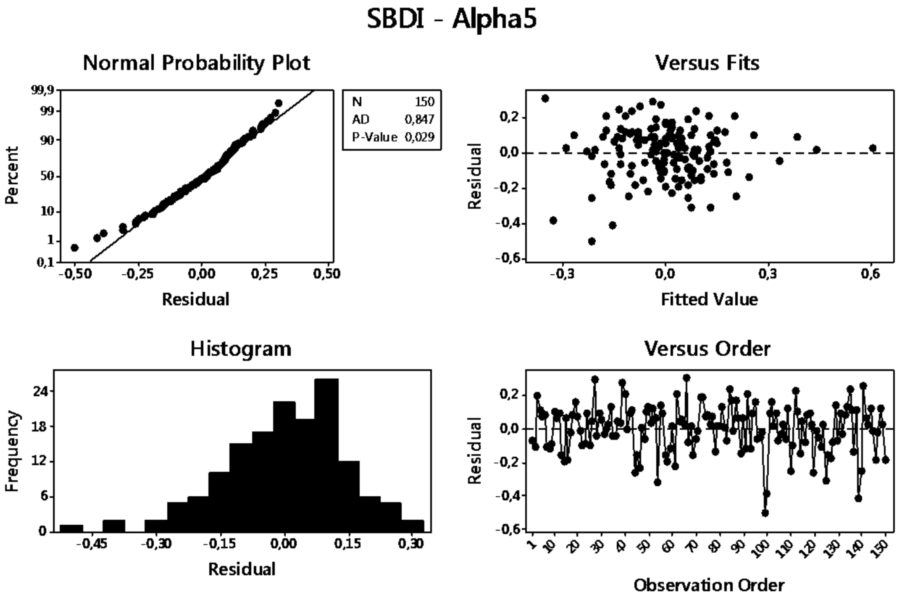

Figure G.7: Residual plot for SBDI Alpha$_5$ model

i) There is no clear pattern/trend in the residuals versus fit- diagram, and consequently there is little reason to question the linearity assumption.

ii) The ACF-plot reveals that no significant autocorrelation is detected in any lags.

iii) From the residuals versus fit diagram, it is apparent that the spread in the residuals seem to increase when fitted value decrease. We have reason to believe there is non-constant variance in this model.

iv) From both normality plots, we observe a relatively symmetric distribution, except for a couple of negative outliers and some unusually frequent positive residuals occurring a little less than one standard deviation from the mean. Even though the residual distribution deviates slightly from the normal, these deviation are not severe, as the asymmetry is not extreme and we recognise the bell shape. In the relevant sample range, the t-distribution will accordingly be robust.

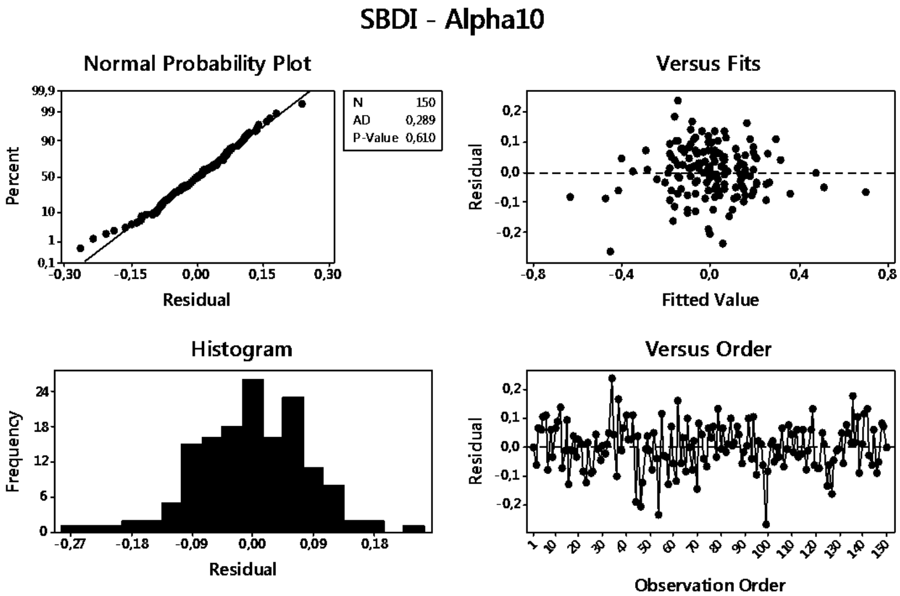

Figure G.8: Residual plot for SBDI Alpha$_{10}$ model

i) There is no clear pattern/trend in the residuals versus fit- diagram, and consequently there is little reason to question the linearity assumption.

ii) The ACF-plot reveals that no significant autocorrelation is detected in any lags.

iii) From the residuals versus fit diagram, it is apparent that the spread in the residuals seem to decrease slightly when fitted value increase. However, there are too few observations to the far left and far right to conclude.

iv) From both normality plots, we observe a relatively symmetric distribution, except for a couple of negative outliers. Even though the residual distribution deviates slightly from the normal, these deviation are not severe, as the asymmetry is not extreme and we recognise the bell shape. In the relevant sample range, the t-distribution will accordingly be robust.

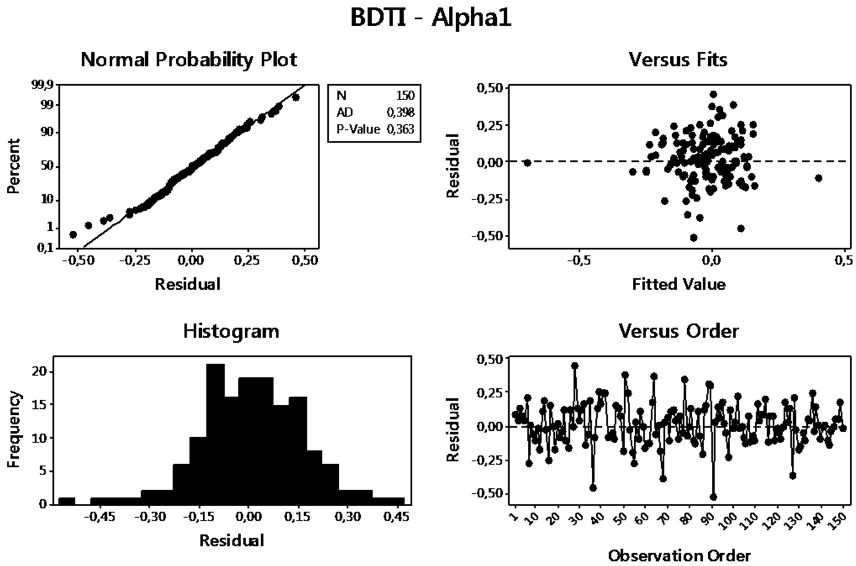

Figure G.9: Residual plot for BDTI Alpha$_1$ model

i) There is no clear pattern/trend in the residuals versus fit- diagram, and consequently there is little reason to question the linearity assumption.

ii) The ACF-plot reveals significant negative autocorrelation in lag four.

iii) From the residuals versus fit diagram, it is apparent that the residuals seem to be approximately evenly distributed along the horizontal axis, thus there is no clear evidence of non-constant variance.

iv) The normal probability plot reveals the occurrence of some negative outliers, while the histogram indicates that the symmetry is relatively good. Overall, the residual distribution seems to be fairly well approximated by the normal distribution.

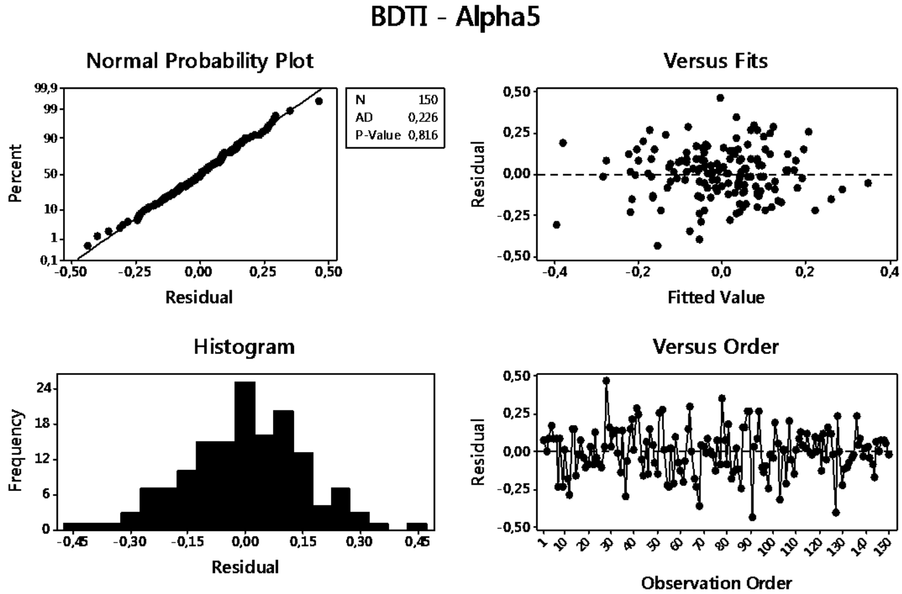

Figure G.10: Residual plot for BDTI $Alpha_5$ model

i) There is no clear pattern/trend in the residuals versus fit- diagram, and consequently there is little reason to question the linearity assumption.

ii) The ACF-plot reveals no significant autocorrelation in any lag.

iii) From the residuals versus fit diagram, it is apparent that the residuals seem to be approximately evenly distributed along the horizontal axis, thus there is no clear evidence of non-constant variance.

iv) The normal probability plot reveals no extensive occurrence of outliers in either direction, while the histogram indicates that the overall symmetry is relatively good. Thus, the residual distribution seems to be fairly well approximated by the normal distribution.

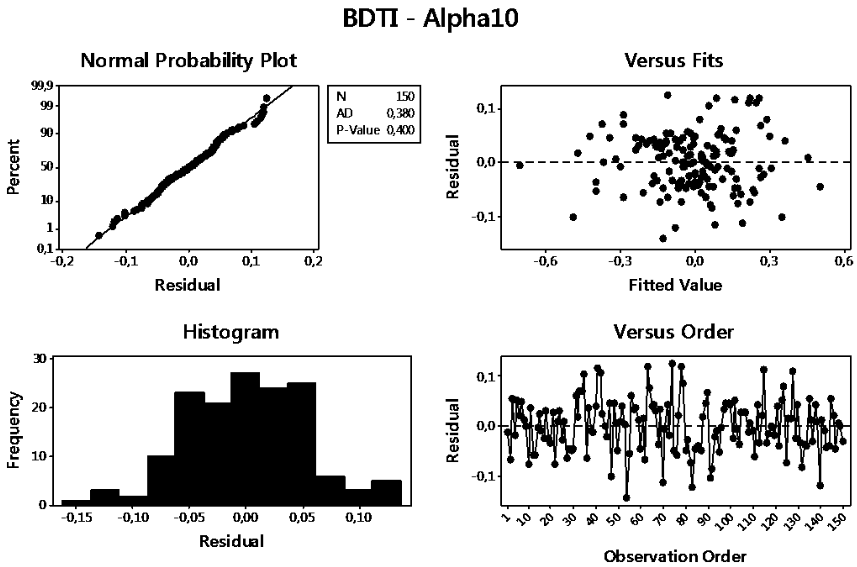

Figure G.11: Residual plot for BDTI Alpha_{10} model

i) There is no clear pattern/trend in the residuals versus fit- diagram, and consequently there is little reason to question the linearity assumption.

ii) The ACF-plot reveals no significant autocorrelation in any lag.

iii) From the residuals versus fit diagram, it is apparent that the residuals seem to be approximately evenly distributed along the horizontal axis, thus there is no clear evidence of non-constant variance.

iv) The normal probability plot reveals no extensive occurrence of outliers in either direction, while the histogram indicates that the overall symmetry is relatively good. Thus, the residual distribution seems to be fairly well approximated by the normal distribution.

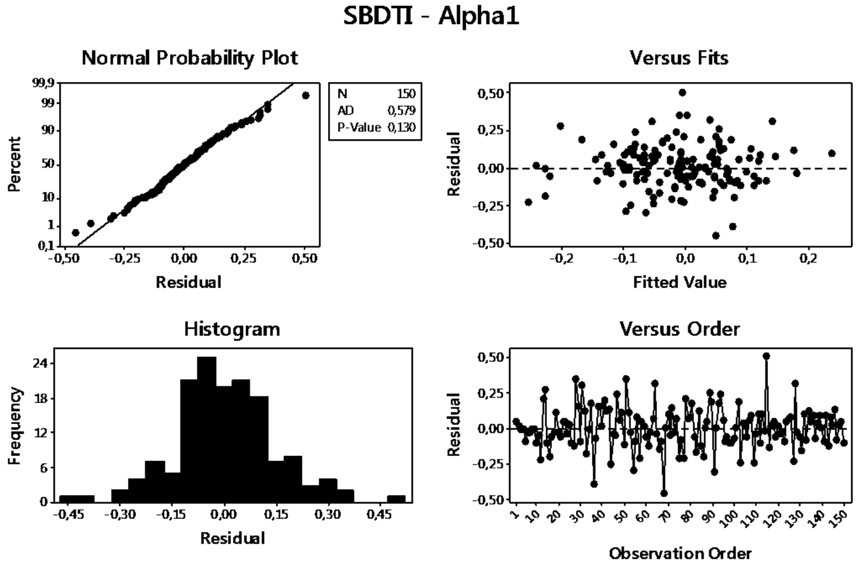

Figure G.12: Residual plot for SBDTI Alpha$_1$ model

i) There is no clear pattern/trend in the residuals versus fit- diagram, and consequently there is little reason to question the linearity assumption.

ii) The ACF-plot reveals no significant autocorrelation in any lag.

iii) From the residuals versus fit diagram, it is apparent that the residuals seem to be approximately evenly distributed along the horizontal axis, thus there is no clear evidence of non-constant variance.

iv) The normal probability plot reveals no extensive occurrence of outliers in either direction, while the histogram indicates that the overall symmetry is relatively good. Thus, the residual distribution seems to be fairly well approximated by the normal distribution.

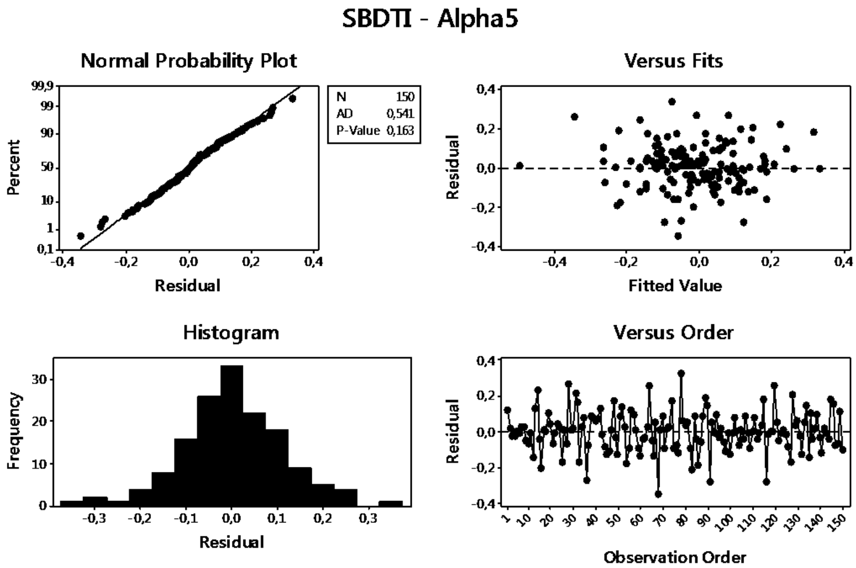

Figure G.13: Residual plot for SBDTI $Alpha_5$ model

i) There is no clear pattern/trend in the residuals versus fit- diagram, and consequently there is little reason to question the linearity assumption.

ii) The ACF-plot reveals no significant autocorrelation in any lag.

iii) From the residuals versus fit diagram, it is apparent that the residuals seem to be approximately evenly distributed along the horizontal axis, thus there is no clear evidence of non-constant variance.

iv) The normal probability plot reveals no extensive occurrence of outliers in either direction, while the histogram indicates that the overall symmetry is relatively good. Thus, the residual distribution seems to be fairly well approximated by the normal distribution.

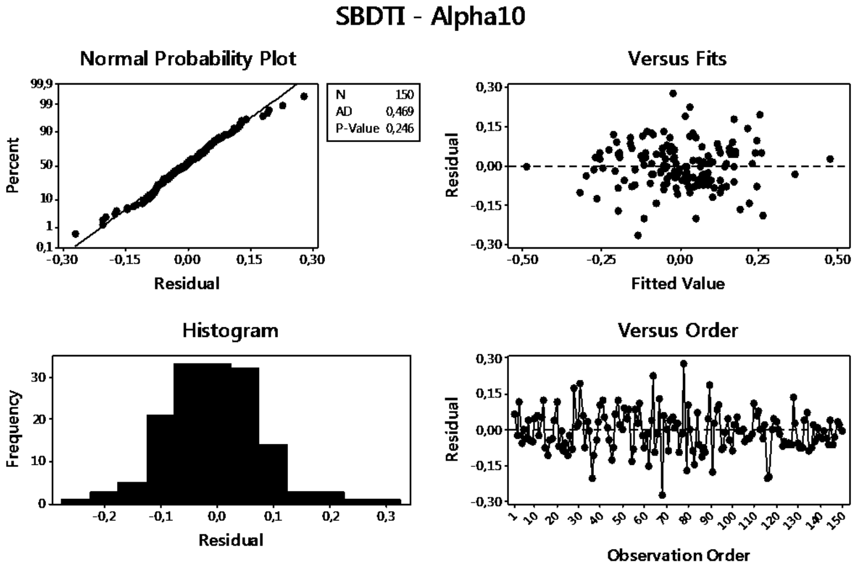

Figure G.14: Residual plot for SBDTI $Alpha_{10}$ model

i) There is no clear pattern/trend in the residuals versus fit- diagram, and consequently there is little reason to question the linearity assumption.

ii) The ACF-plot reveals no significant autocorrelation in any lag.

iii) From the residuals versus fit diagram, it is apparent that the residuals seem to be approximately evenly distributed along the horizontal axis, thus there is no clear evidence of non-constant variance.

iv) The normal probability plot reveals no extensive occurrence of outliers in either direction, while the histogram indicates that the overall symmetry is relatively good. Thus, the residual distribution seems to be fairly well approximated by the normal distribution.

References

Roar Adland and Kevin Cullinane. A time-varying risk premium in the term structure of bulk shipping freight rates. *Journal of Transport Economics and Policy (JTEP)*, 39(2):191–208, 2005.

Roar Adland and Kevin Cullinane. The non-linear dynamics of spot freight rates in tanker markets. *Transportation Research Part E: Logistics and Transportation Review*, 42(3):211–224, 2006.

H Akaike. Information theory and an extension of the maximum likelihood principle. *2nd International Symposium on Information Theory (Petrov and, B. N. and Csaki, F., eds.)*, (2):267–281, 03 1973.

Carol Alexander. *Market Risk Analysis, Practical Financial Econometrics (The Wiley Finance Series) (Volume I)*. John Wiley Sons, volume i edition, 2008. ISBN 978-0470998007.

Amir H. Alizadeh and Wayne K. Talley. Microeconomic determinants of dry bulk shipping freight rates and contract times. *Transportation*, 38(3):561–579, 2011. ISSN 1572-9435. doi: 10.1007/s11116-010-9308-7. URL http://dx.doi.org/10.1007/s11116-010-9308-7.

Julius Okechukwu Anyanwu. Analysis of the interrelationships between the various shipping markets. *Industrial Engineering Letters*, 3(10), 2013. ISSN 2225-0581. URL http://www.iiste.org/Journals/index.php/IEL/article/view/8008.

Nicholas Apergis and James E Payne. New evidence on the information and predictive content of the baltic dry index. *International journal of financial studies*, 1(3):62–80, 2013.

Arctic. Data provided by Arctic Securities, 17 June 2017, 6 2017.

S Askari and N Montazerin. A high-order multi-variable fuzzy time series forecasting algorithm based on fuzzy clustering. *Expert Systems with Applications*, 42(4):2121–2135, 2015.

Gurdip Bakshi, George Panayotov, and Georgios Skoulakis. The baltic dry index as a predictor of global stock returns, commodity returns, and global economic activity. 2011.

Roy Batchelor, Amir Alizadeh, and Ilias Visvikis. Forecasting spot and forward prices in the international freight market. *International Journal of Forecasting*, 23(1):101–114, 2007.

Christiane Baumeister and Lutz Kilian. Understanding the decline in the price of oil since june 2014. *Journal of the Association of Environmental and Resource Economists*, 3(1):131–158, 2016. doi: 10.1086/684160.

Michael Beenstock. A theory of ship prices. *Maritime Policy & Management*, 12(3):215–225, 1985. doi: 10.1080/03088838500000028. URL http://dx.doi.org/10.1080/03088838500000028.

Michael Beenstock and Andreas Vergottis. An econometric model of the world market for dry cargo freight and shipping. *Applied Economics*, 21(3):339–356, 1989a. doi: 10.1080/758522551. URL http://dx.doi.org/10.1080/758522551.

Michael Beenstock and Andreas Vergottis. An econometric model of the world tanker market. *Journal of Transport Economics and Policy*, pages 263–280, 1989b.

Michael Beenstock and Andreas Vergottis. The interdependence between the dry cargo and tanker markets. *Logistics and Transportation Review*, 29(1):3, 03 1993. URL https://search.proquest.com/docview/197442403?accountid=12870.

Fred Espen Benth, Steen Koekebakker, and Che Mohd Imran Che Taib. Stochastic dynamical modelling of spot freight rates. *IMA Journal of Management Mathematics*, 26(3):273–297, 2015.

Anil K. Bera and Carlos M. Jarque. Efficient tests for normality, homoscedasticity and serial independence of regression residuals. *Economics Letters*, 7(4):313 – 318, 1981. ISSN 0165-1765. doi: http://dx.doi.org/10.1016/0165-1765(81)90035-5. URL http://www.sciencedirect.com/science/article/pii/0165176581900355.

F. Guillaume Blanchet, Pierre Legendre, and Daniel Borcard. Forward selection of explanatory variables. *Ecology*, 89(9):2623–2632, 2008. ISSN 1939-9170. doi: 10.1890/07-0986.1. URL http://dx.doi.org/10.1890/07-0986.1.

A. Boisbunon, S. Canu, D. Fourdrinier, W.E. Strawderman, and M.T. Wells. Aic, cp and estimators of loss for elliptically symmetric distributions. *International Statistical Review*, 2014. in press.

T. S. Breusch and A. R. Pagan. A simple test for heteroscedasticity and random coefficient variation. *Econometrica*, 47(5):1287–1294, 1979. ISSN 00129682, 14680262. URL http://www.jstor.org/stable/1911963.

Chris Brooks. *Introductory Econometrics for Finance*. Cambridge University Press, 3 edition, 6 2008.

Julia Campos, Neil R. Ericsson, and David F. Hendry. General-to-specific modeling: An overview and selected bibliography. *FRB International Finance Discussion Paper*, (838), 2005. doi: http://dx.doi.org/10.2139/ssrn.791684. URL https://ssrn.com/abstract=791684.

Chris Chatfield. Book reviews. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 167(3):567–568, 2004. ISSN 1467-985X. doi: 10.1111/j.1467-985X.2004.02057_4.x. URL http://dx.doi.org/10.1111/j.1467-985X.2004.02057_4.x.

Shun Chen, Hilde Meersman, Eddy Van de Voorde, and Koos Frouws. *Modelling and Forecasting in Dry Bulk Shipping*. CRC Press, 2014.

J.J. Coyle, E.J. Bardi, and R.A. Novack. Transportation. 2000.

Steve Engelen, Hilda Meersman, and Eddy Van De Voorde. Using system dynamics in maritime economics: an endogenous decision model for shipowners in the dry bulk sector. *Maritime Policy & Management*, 33(2):141–158, 2006.

Shuangrui Fan, Tingyun Ji, Wilmsmeier Gordon, and Bergqvist Rickard. Forecasting baltic dirty tanker index by applying wavelet neural networks. 2013.

DR Glen and BT Martin. A survey of the modelling of dry bulk and tanker markets. *Research in Transportation Economics*, 12(1):19–64, 2005.

Stephen M. Goldfeld and Richard E. Quandt. Some tests for homoscedasticity. *Journal of the American Statistical Association*, 60(310):539–547, 1965. doi: 10.1080/01621459.1965.10480811. URL http://www.tandfonline.com/doi/abs/10.1080/01621459.1965.10480811.

Alexander M. Goulielmos. What can we learn from 259 years of shipping cycles? *International Journal of Shipping and Transport Logistics*, 2(2):125–150, 2010. doi: 10.1504/IJSTL.2010.030863. URL http://www.inderscienceonline.com/doi/abs/10.1504/IJSTL.2010.030863.

Costas Th. Grammenos and Angelos G. Arkoulis. Determinants of spreads on new high yield bonds of shipping companies. *Transportation Research Part E: Logistics and Transportation Review*, 39(6): 459 – 471, 2003. ISSN 1366-5545. doi: http://dx.doi.org/10.1016/S1366-5545(03)00054-1. URL //www.sciencedirect.com/science/article/pii/S1366554503000541.

James D. Hamilton. Oil and the macroeconomy. *The New Palgrave Dictionary of Economics*, pages 201–228, 2005. URL http://www.dictionaryofeconomics.com/dictionar.

James D. Hamilton. Understanding crude oil prices. (14492), November 2008. doi: 10.3386/w14492. URL http://www.nber.org/papers/w14492.

Michael J Hampton. Long and short shipping cycles. *Cambridge Academy of Transport, Cambridge*, 1991.

Frank Harrell. *Regression Modeling Strategies*. Springer-Verlag New York, 1 edition, 2001.

D. Hawdon. Tanker freight rates in the short and long run. *Applied Economics*, 10(3):203–218, 1978. doi: 10.1080/758527274. URL http://dx.doi.org/10.1080/758527274.

R.J. Hyndman and G Athanasopoulos. *Forecasting: principles and practice*. 2013. URL http://otexts.org/fpp/.

Robert J. Hyndman. Benchmarks for forecasting. 2010. URL http://robjhyndman.com/hyndsight/benchmarks/#disqus_thread.

Kim Hyung-geun. Study about how the chinese economic status affects to the baltic dry index. *International Journal of Business and Management*, 6(3), 2011.

Tianyong Long J. B. Copas. Estimating the residual variance in orthogonal regression with variable selection. *Journal of the Royal Statistical Society. Series D (The Statistician)*, 40(1):51–59, 1991.

R T Jackson. *Ethnic difference in the relationship between acute inflammation and serum ferritin in US adult males*. 2008.

Manolis G Kavussanos. Comparisons of volatility in the dry-cargo ship sector: Spot versus time charters, and smaller versus larger vessels. *Journal of Transport economics and Policy*, pages 67–82, 1996.

Manolis G. Kavussanos and Amir H. Alizadeh-M. Seasonality patterns in dry bulk shipping spot and time charter freight rates. *Transportation Research Part E: Logistics and Transportation Review*, 37 (6):443 – 467, 2001. ISSN 1366-5545. doi: http://dx.doi.org/10.1016/S1366-5545(01)00004-7. URL //www.sciencedirect.com/science/article/pii/S1366554501000047.

Manolis G. Kavussanos and Amir H. Alizadeh-M. Seasonality patterns in tanker spot freight rate markets. *Economic Modelling*, 19(5):747 – 782, 2002. ISSN 0264-9993. doi: http://dx.doi.org/10.1016/S0264-9993(01)00078-5. URL http://www.sciencedirect.com/science/article/pii/S0264999301000785.

Manolis G Kavussanos and Ilias D Visvikis. Shipping freight derivatives: a survey of recent evidence. *Maritime Policy & Management*, 33(3):233–255, 2006.

Tjalling Charles Koopmans. *Tanker Freight Rates and Tankship Building: An Analysis of Cyclical Fluctuations, by Dr. T. Koopmans*. Number 27. De erven F. Bohn nv, 1939.

D. N. Midthune L. S. Freedman, D. Pee. The problem of underestimating the residual error variance in forward stepwise regression. *Journal of the Royal Statistical Society. Series D (The Statistician)*, 41 (4):405–412, 1992.

Faqin Lin and Nicholas C.S. Sim. Trade, income and the baltic dry index. *European Economic Review*, 59:1 – 18, 2013. ISSN 0014-2921. doi: http://dx.doi.org/10.1016/j.euroecorev.2012.12.004. URL //www.sciencedirect.com/science/article/pii/S0014292112001596.

Y.H.V. Lun and T.C.E. Cheng. Shipping and logistics management. 2010.

Y.H.V. Lun, K.H. Lai, and T.C.E. Cheng. Shipping and transport logistics. 2006.

C. L. Mallows. Sequential sampling of finite populations with and without replacement. 24(2):164–168, March 1973. ISSN 0036-1399 (print), 1095-712X (electronic).

Luiz Moutinho and Kun-Huang Huarng. *Quantitative modelling in marketing and management.* Singapore : World Scientific, 2013.

G. P. Nason. A test for second-order stationarity and approximate confidence intervals for localized autocovariances for locally stationary time series. *Journal of the Royal Statistical Society, Series B*, 75:879–904, 2013. URL http://www.stats.bris.ac.uk/~guy.

John Neter, William Wasserman, and Michael H. Kutner. *Applied linear regression models.* Irwin, Burr Ridge, Ill. [u.a.], 2. ed. edition, 1989. ISBN 0-256-07068-7.

Angela Poulakidas and Fred Joutz. Exploring the link between oil prices and tanker rates. *Maritime Policy & Management*, 36(3):215–233, 2009.

Jeroen Frederik Josef Pruyn, Eddy van de Voordeb, and Hilde Meersman. Second hand vessel value estimation in maritime economics: A review of the past 20 years and the proposal of an elementary method. *Maritime Economics Logistics*, 13(2):213–236, 2011.

J. B. Ramsey. Tests for specification errors in classical linear least-squares regression analysis. *Journal of the Royal Statistical Society. Series B (Methodological)*, 31(2):350–371, 1969. ISSN 00359246. URL http://www.jstor.org/stable/2984219.

Jørgen Randers and Ulrich Göluke. Forecasting turning points in shipping freight rates: lessons from 30 years of practical effort. *System Dynamics Review*, 23(2-3):253–284, 2007. ISSN 1099-1727. doi: 10.1002/sdr.376. URL http://dx.doi.org/10.1002/sdr.376.

Barbara Rossi and Atsushi Inoue. Out-of-sample forecast tests robust to the choice of window size. *Journal of Business & Economic Statistics*, 30(3):432–453, 2012. doi: 10.1080/07350015.2012.693850. URL http://dx.doi.org/10.1080/07350015.2012.693850.

Gideon Schwarz. Estimating the dimension of a model. *Ann. Statist.*, 6(2):461–464, 03 1978. doi: 10.1214/aos/1176344136. URL http://dx.doi.org/10.1214/aos/1176344136.

Harvey Keselman Stephen Olejnik, Jamie Mills. Using wherry's adjusted rÂ$^2 andmallow's < tex - math >c_p$¡/tex-math¿ for model selection from all possible regressions. *The Journal of Experimental Education*, 68(4):365–380, 2000.

M. A. Stephens. Edf statistics for goodness of fit and some comparisons. *Journal of the American Statistical Association*, 69:730–737, 1974.

Martin Stopford. *Maritime Economics.* Routledge, 3 edition, 2009.

SP Strandenes. Price determination in the time charter and second hand markets. *Center for Applied Research, Norwegian School of Economics and Business Administration, Working Paper MU*, 6, 1984.

Jan Tinbergen. Annual survey of significant developments in general economic theory. *Econometrica: Journal of the Econometric Society*, pages 13–36, 1934.

S D Tsolakis, C. Cridland, and H E Haralambides. Econometric modelling of second-hand ship prices. *Maritime Economics & Logistics*, 5(4):347–377, 2003. ISSN 1476-0592. doi: 10.1057/palgrave.mel.9100086. URL http://dx.doi.org/10.1057/palgrave.mel.9100086.

Jostein Tvedt. A new perspective on price dynamics of the dry bulk market. *Maritime Policy & Management*, 30(3):221–230, 2003. doi: 10.1080/0308883032000133413. URL http://dx.doi.org/10.1080/0308883032000133413.

Albert Willem Veenstra and Philip Hans Franses. A co-integration approach to forecasting freight rates in the dry bulk shipping sector. *Transportation Research Part A: Policy and Practice*, 31(6): 447 – 458, 1997. ISSN 0965-8564. doi: http://dx.doi.org/10.1016/S0965-8564(97)00002-5. URL //www.sciencedirect.com/science/article/pii/S0965856497000025.

Sjur Westgaard, Daniel Stenslet, Jo Ringheim, and Petter Osmundsen. Modeling superior predictors for crude oil prices. *Journal of Energy Markets*, 10(2):11–22, 2017. ISSN 1756-3615. doi: https://doi.org/10.21314/JEM.2017.162. URL https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2983716.

Halbert White. A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. *Econometrica*, 48(4):817–838, 1980. ISSN 00129682, 14680262. URL http://www.jstor.org/stable/1912934.

Michael Ye, John Zyren, and Joanne Shore. Forecasting short-run crude oil price using high- and low-inventory variables. *Energy Policy*, 34(17):2736 – 2743, 2006. ISSN 0301-4215. doi: https://doi.org/10.1016/j.enpol.2005.03.017. URL http://www.sciencedirect.com/science/article/pii/S0301421505001023.

Zenon S Zannetos. *The theory of oil tankship rates*. Massachusetts Institute of Technology, 1966.