Norwegian School of Economics
Bergen, Spring 2014

NHH

THE RELATIONSHIP BETWEEN SHIP SPECIFICATIONS AND TIME CHARTER RATES

QUALITY PREMIUMS THROUGH A GENERALIZED ADDITIVE MODEL

Simon Riise & Marius Bjørngaard Rødde

Supervisor: Roar Os Aadland

Master thesis, Business Analysis and Performance Management / Energy, Natural Resources and the Enviroment

NORWEGIAN SCHOOL OF ECONOMICS

ABSTRACT

Nonlinear semi-parametric modeling of freight rates in the dry bulk market has gained increased recognition, although it has not been implemented extensively in previous studies. By allowing the data itself to dictate the functional form of our model we seek to decompose the physical time charter rate according to a combination of macro - and microeconomic variables. Our main goal is to identify quality related aspects of rate determinants and the possible existence of a quality premium. A spline interpolation technique is applied to exclude time–related modifications to the market index along with actual contract fixtures from 2001 to 2014 across the three main dry bulk segments. Supporting previous literature, we identify a clear non-linearity between variables and support for quality segmentation with respect to age and size. Furthermore Japan has increased its position as a provider of quality tonnage and there has been a shift in fuel efficiency, measured as consumption per ton-miles, over time. Contractual specification on cost allocations between shipowner and charterer suggest that the charterer obtains any possible quality related profits, showing evidence of split incentive barriers. In general we find that there has been a shift in market dynamics in time periods before and after 2008.

PREFACE

When searching for a suitable topic for our thesis we both agreed upon maritime economics as an interesting and challenging topic. Ships, oceans and commodities stand initially out as specific, real and measurable in economics terms. After attending Prof. Adlands course in "Commodity Trading and Transport" at NHH in the spring of 2013 we soon realized how complex and dynamic the industry is. However, we found these crossroads between real assets and complexity highly fascinating. After discussions with our supervisor we believe to have found a relevant and forward-looking subject, which could shed some further light on the shipping markets.

As we now are about to finish our degree at NHH, there are some who deserves our attention.

First and foremost we would like to thank our supervisor, Prof. Roar Aadland, for sharing his extensive knowledge and research insight through discussions and previous studies. Furthermore we are grateful to be a part of GREENSHIPRISK, which is a research project in collaboration between NHH, MARINTEK NTNU and several other market participants. Hopefully our work can be of relevance. We would also like to thank DNV-GL for useful discussions and inputs. Finally, we appreciate the financial support provided to us by the Norwegian Shipowners Association fund at NHH.

Bergen, June 2014

Marius Bjørngaard Rødde   Simon Riise

This page intentionally left blank

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

1. INTRODUCTION

The goal of this paper is to investigate the relationship between vessel specifications and time-charter rates with an emphasis on energy efficiency. The nature of the shipping industry is highly volatile and cyclical where freight rates is said to be one of the most volatile commodities traded (Adland, 2013b). Gaining a deeper understanding of how the market prices vessel characteristics could be of great interest, as it might help market participants make sound operational and investments decisions. As with other competitive and homogenous markets, the maritime industry is dominated by cost-efficiency - where the quest for energy efficiency has grown to become one of the major components. Accompanied by economies of scale and increasing commodity prices, market participants seek ways to continuously improve profitability. In addition to this intrinsic market characteristic the Oil Pollution Act (1990, Oil Pollution Act) initiated a quest for regulatory changes considering quality tonnage, which has become prevailing in recent years.

While the industry itself and regulators sees the need for efficiency measures, there is still a cost-benefit issue for both shipowners and charterers. Bunkers cost are the single most important item in voyage cost[1], representing 47% of the total (Stopford, 2009) and as oil prices continue to increase the fuel cost becomes progressively important. Consequently, as bunkers cost are closely related to efficiency measures we would assume the shipping markets to place some price premium on efficiency. Although the efficiency of maritime sea transportation is rarely disputed, there have been debates regarding the quantification of specific energy efficient characteristics of the world fleet (Smith et al., 2013). Skepticism has been directed towards unrepresentative input data in previous studies along with limited real world operational data. The "still water" design parameters often used in analyses of the world fleet has been cited as artificial and insufficient to reflect efficiency parameters. Although this critique is well founded, such real world data are irregular and hard to measure (Smith et al., 2013). Consequently, there is now an increasing interest in a more comprehensive analysis of energy efficiency. Further sophisticated modeling techniques is one way to cope with this prevailing data issue.

---

[1] Accounts for 40% of total cost. Variable voyage costs including fuel, port dues, tugs, pilotage and canal charges (Stopford, 2009)

The general idea behind the methodology applied is to allow the functional form of the relationships between model variables to take on any form. Hence we apply a model that avoids imposing distinctive parameter restrictions, which could lead to model misspecifications. We use a semi-parametric estimation technique within a Generalized Additive Model. By assuming less about how rates and underlying factors are related we believe to discover more accurate results.

Opposed to previous research on the dry bulk market, our dataset includes fixtures for Capesize, Panamax and Handysize from 2001 to 2014. A large dataset allows us to combine time-periods during the extreme market conditions seen in recent years. In addition we investigate if various market indices used as benchmarks are adjusted for changes in the underlying specifications over time. Hence, when searching for efficiency causalities some of these, if not all, could already be incorporated in the index and would consequently be difficult to isolate. Therefore we try to eliminate the effect of changes over time by standardizing the data, making sure the same underlying vessel specifications are used as a comparable. Subsequently we use cubic spline interpolation to estimate a correct index proxy for contracts that falls outside given durations. Our understanding is that this modeling approach would better explain efficiency improvements over time.

The question remains whether the incentives are correctly aligned and if the markets are willing to pay for efficiency practices to be implemented. Without a strong price signal there could be a lack of incentives to justify investments in energy efficiency or operational energy efficiency measures.

Our contributions to existing literature lie in a comparison across all the three main dry bulk segments and a large fixture set of actual cross sectional data where periods before and after the financial crisis are considered. A weakness in comparable studies is the use of spot rates (BDI)[2] as a proxy for period time charter rates. The spot rate

[2] Baltic Dry Index

does not include or differentiate between a market in contango or backwardation[3]. Hence, the length of the contracts is neglected when spot is used as a proxy. Henceforth we use an improved modeling of the market-rate component, using Clarksons time charter indices as proxies.

A brief introduction to dry bulk shipping is presented in section 3 followed by efficiency discussions in section 4. Section 5 provides methodology and a theoretical framework. Section 6 outlines our empirical analysis along with a description of variables. Model specification and selection is discussed in section 7 followed by an analysis of our results in section 8. In section 9 we present a possible model application. Uncertainties regarding the model are presented in section 10 and finally our concluding remarks are given section 11.

[3] Contango refers to a situation where the future price of a commodity is higher than the expected future spot price, hence the forward cure is upward sloping. If a market is in backwardation, the opposite situation occurs and the forward curve is downward sloping.

2. Literature Review

Our research topic suggests a three-folded approach when examining previous literature: namely freight rate determination, efficiency and quality aspects and various modeling techniques.

Following the Oil Pollution Act (1990, Oil Pollution Act), researchers have examined the implications of the increasing importance of quality in dry bulk freight rates, where the focal point has been age as a proxy for quality. Strandenes' (1994) research on the possibility of a two-tier tanker was in many ways a pioneering study setting the course for further research. Thanopoulou and Gardner (2012) argue that although various research methods have been applied, the findings are similar; no distinctive quality segmentation has yet been confirmed or can be deemed sustainable in the long run. Further they argue that the main limitations lie in the methodology used. The same authors also presents a detailed overview of candidate parameters for quality research where they suggest a closer and continuously monitoring of market attitudes along with more realistic and relevant information from charterers regarding how they rank freight rate determinants.

Early studies by Hawdon (1978), Strandenes (1984) and Beenstock and Vergottis (1993) advocate that ocean freight rates are determined mainly through macroeconomic variables. More recent studies by Randers and Göluke (2007) also use such variables in a system dynamic approach to model and forecast freight rates. Furthermore Adland and Koekebakker (2007), Adland and Cullinane (2006) and Kavussanos and Alizadeh (2002) studied the time series dynamics of freight rates through univariate or multivariate models. Adland and Cullinane (2006) investigate the non-linear dynamics of spot freight rates in tanker markets through a non-parametric Markov diffusion model. Their study suggests that a non-linear stochastic model can best describe the oil transportation market. Further they show that spot prices are mean reverting and that the volatility of the rate increases with freight levels. Alizadeh and Talley (2010) proposes that there has been no systematic investigation of microeconomic determinants. Hence, they investigate rate determination through market, vessel and contract specific factors in the dry bulk market along with differences in freight rates across routes, geographical distribution of shipping activities and the duration of the laycan period of shipping contracts.

Using a system of simultaneous equations along with a large sample of individual dry bulk charter contracts from January 2003 to July 2009, findings suggest that microeconomic ship specifications are important determinants of dry bulk rates. A positive relationship is shown between freight rates and dead weight tonnage (dwt) of a vessel and between rates and laycan periods. Findings also suggest that laycan period vary directly with rates and indirectly with rate volatility. A possible existence of a two-tier voyage freight market for Panamax and Capesize vessels of differing age is studied in the works of Tamvakis & Thanopoulou (2000). Using data from four representative years since the end of the 1980s, in which the freight market conditions prevailed, they apply a semi-parametric multiple regression and find no statistically significant difference between rates paid to older and younger tonnage. Any signs of differentiation turn out to be too sporadic to alter current perceptions of the dry bulk carrier market.

Glen (2006) provides an overview of the development of the quantitative modeling techniques so far applied to the analysis of dry bulk shipping markets. He finds that ever since Beenstock and Vergottis published their “Econometric Modelling of World Shipping” in 1993 the development of the quantitative modeling techniques of the shipping markets has shifted attention towards relatively more modern techniques, where specification and estimation of complete structural models has been more or less avoided. The uprising of econometric techniques has also changed the attention towards more specific aspects of the shipping markets, such as demand, the behavior of ship prices, seasonality and assumptions about expectations, which previously had been disregarded. The results also identifies the following modeling trends in recent literature: Reduced form rather than structural modeling, a greater focus on modeling rate volatility rather than rate levels, introduction of models of financial derivatives and their application to shipping markets and lastly segmented models of different vessel types along with higher frequency data.

There are in particular two previous studies we find relevant regarding econometric modeling. For the first time, Adland and Koekebakker (2007) departs from time-series analysis and studies actual cross sectional bulk ship sales data through a non-parametric multivariate density estimation technique. Hence they allow for the presence of non-linearity in addition to microeconomic variables, which is also found to be an important issue in vessel valuation. Although they suggest that the second-hand value of a vessel can be found as a partially non-linear function of size, age and the state of the freight market, the model is not capable of fully explaining observed vessel values. Consequently they suggest other vessel-specific variables to be included in a semi-parametric framework as an extension of the proposed model.

Köhn (2008) revisits the quality-issue and aims to verify the hypothesis of a quality-based segmentation in the Panamax market based on vessel and contract specific determinants. By broadening the range of econometric tools applied in shipping economics through a semi-parametric estimation technique within a Generalized Additive Model, results suggest that freight differentiation has become visible in booming markets with high rates. The works of Köhn and Thanopoluou (2011) is primarily based on Köhn (2008) and results suggest that non-linearity and microeconomic variables are found to explain a great part of the variation in physical time charter rates.

3. DRY BULK SHIPPING

3.1. OVERVIEW

The dry bulk shipping market is by far the largest sector of the worlds shipping markets in terms of cargo weight and by 2013 the total dry bulk trade accounted for approximately 44% of total trade volumes (Clarksons Research, 2014). Iron ore and coking coal are the raw materials in steel production, which again is one of the main building blocks of the modern industrialized society. Growth in gross domestic product is closely linked to energy consumption and as long as the importance of coal continues to increase as a power generation source, it´s importance as a commodity is rarely undisputed. Grain supplies the world with bread, meat and other foods necessary for any society to grow and sustain. Due to their volume and demand, the three major bulk trades mentioned above form the major forces behind the dry bulk carrier market, accounting for almost one quarter of total seaborne cargo showing an annual growth of 4.4% from 1965-2005 (Stopford, 2009). It is worth mentioning that there is no simple pattern and each commodity has its own distinctive characteristic and growth trends.

Dry bulk carriers are designed for low cost and simplicity and the fleet is made up of four main sectors. Out of the three segments listed below, Capesize is the smallest in number, but largest in terms of dwt. There is however a tradeoff between unit cost and cargo flexibility where smaller vessels are flexible but more expensive to operate, while larges vessels offers economies of scale[4], but less flexibility.

Capesize (>100,000 DWT) vessels are ships generally too large to transit the Suez Canal. These ships are used to carry coal, iron ore and other commodity raw materials on long-haul routes to the largest ports around the world.

Panamax (60,000-99,999 DWT) vessels are used to carry the same commodities as Capesize ships in addition to minor bulks. Since these ships are able to pass through the Panama Canal[5] they are more versatile in terms of access to different trade routes, although ships in the larger end of the range are unable to pass until the canal is

[4] Inverse relationship between quantities shipped and per unit fixed-cost.

[5] Maximum beam is 32.3 meters before canal expansion. (Canal de Panama, 2013)

extended. Usually, most of the ships within this category do not carry cargo-handling equipment on board.

Handymax (40,000-59,999 DWT) vessels carry mostly grain and minor bulks and are used in a large variety of global trade routes. Compared to Panamax vessels these ships often carry cargo handling equipment, offering loading and unloading flexibility.

3.2. The Freight Market

Freight rates are the most important factor used by the market to adjust short-term capacity and long-run cost and service improvement. Supply and demand is adjusted by the freight mechanism, where a market balance between supply and demand is found when shipowners and charterers negotiate and establish a freight rate based on available ships and cargo. The conceptualization of the freight mechanism is more or less straightforward and can be found in most microeconomic theory. In a demand driven market, where demand dictates market supply, rates tend to be low. On the contrary, a supply driven market would result in higher rates. Adjustment to this mechanism will eventually bring the market in balance. An important remark considering the supply side is that since the shipowner has an option to put a vessel into lay-up if the rate is too low, it practically sets a minimum rate. The level of freight rates also determines the speed at which an owner will operate[6]. Thus, the market can increase supply capacity by increasing rates. Due to the convexity of the function, supply for shipping services is highly elastic at high freight rates and becomes inelastic at low freight rates, which can be explained by the availability of excess capacity when the market is in recession (Alizadeh and Nomikos, 2009). In good times rates can soar to exceptionally high levels. Mainly since the cost of transport relative to the value of the cargo is low and lack of alternative transport methods (Randers and Göluke, 2007). In response shipowners seek to increase supply of tonnage as long as there is adequate shipyard capacity and available funding.

[6] The supply function has a J-shaped curve.

In reality the supply function is more complex and depends on a variety of other aspects. A relatively long lead-time along with opportunistic market behavior reinforces market cycles, resulting in a highly volatile environment. Further explanation can be found in Stopford (2009). The demand function is almost vertical which reflects charterer's inelasticity due to a lack of substitutes for transportation[7]. This effect could suggest different pricing of vessel specifications when the market peaks compared to a market trough.

Contracts can be classified into various segments differing in terms of duration, freight rate calculation, cost allocation and commercial and operational responsibilities. Single-voyage, trip-charter and time-charter are the most common types of contracts used in the dry-bulk market (Alizadeh and Nomikos, 2009). Freight rates are strongly correlated over time, partly due to the strong degree of substitutability of cargoes among vessels and routes. Another contribution to this property is that speculators and financial institutions are indistinguishable with regard to type of trade and route (Randers and Göluke, 2007).

3.3. The Time Charter Contract

The time charter contract (TC) is an agreement where the charterer hires the vessel, including crew, for a predetermined fee per specified period, usually USD/day. This gives the charterer operational control and flexibility of the vessel as well as cost control, while ownership and management is left to the shipowner. The shipowner incurs capital–and operating costs, while the charterer covers all voyage expenses (bunkers, port charges and canal dues). Contractual arrangements along with legal responsibilities are set out in the charter party. Time-charter contracts are long-term contracts and cover more than one voyage. Under this arrangement freight market risk is redistributed according to owner or charterers risk preferences. By entering such an agreement ship owners could charter out vessels and obtain a steady revenue stream - reducing spot market exposure. On the other hand this would also limit any upside potential if market rates should increase. Although rates are freely negotiable there is a strong interest in recent transactions as they form the starting point for most

[7] Mainly a hypothesis from Stopford (2009)

negotiations. Opposed to other freight contracts the cost allocation between the shipowner and charterer has some market implications as the charterer incurs bunkers cost. For purposes related to the investigation of quality tonnage, the time charter contract is assumed suitable as it, per se, represents “pure” rates.

A particularly interesting feature with the time charter contract is the intrinsic split incentive barrier, sometimes referred to as the principal-agent problem. According to economic theory this represents an economic market failure where efficiency measures are not implemented despite substantial cost savings potential (Rehmatulla, Smith and Wrobel, 2013). The problem arises when the shipowner is faced with the cost of efficiency improvements, while the charterer is receiving the benefits of these improvements through reduced fuel consumption. The shipowner might not be compensated for increased CAPEX through higher rates.

4. Energy Efficiency for Ocean Freight Transportation

4.1. The Emergence of Quality Differences in a Competitive Market

In line with the theoretical idea of perfect market competition the dry bulk market, except liner and some specialized subsectors, is commonly assumed to inherent the same characteristics; Ownership of capacity is well distributed and no independent company can manipulate supply, shipyards control only a small share of total capacity and vessels have great geographical mobility. Consequently, shipping services are traded on the assumption that each vessel of similar size and type are perfect substitutes for each other. Freight rates have until now been undifferentiated in terms of vessel characteristics with individual ship owners being price takers. It has also been shown empirically that traditional bulk shipping has historically been without significant barriers to entry (Thanopoulou and Gardner, 2012) (Beenstock and Vergottis, 1993). Competition has been characterized by cost-efficiency where participants try to survive volatile and cyclical patterns. It is well founded in theory that as long as markets remain fully competitive, the quality of products and services offered is of less importance since homogeneity is a fundamental condition. That is, fully competitive markets are not necessarily deprived of quality characteristics, however competition itself is not based on buyers having diverging perceptions of product characteristics. The Oil Pollution Act (1990) became a turning point for the industry, introducing unlimited liability for environmental damages. Followed by the institution of the International Safety Management Code the dry bulk sector became subject to a quality question. The International Maritime Organization (IMO) followed up with a retroactive legislation for dry bulk carriers, requiring structural modifications of vessels, which was also introduced by classification societies in 1996. These regulatory changes stimulated research into quality segmentation, starting with the hypothesis of a two-tier tanker market (Strandenes, 1994).

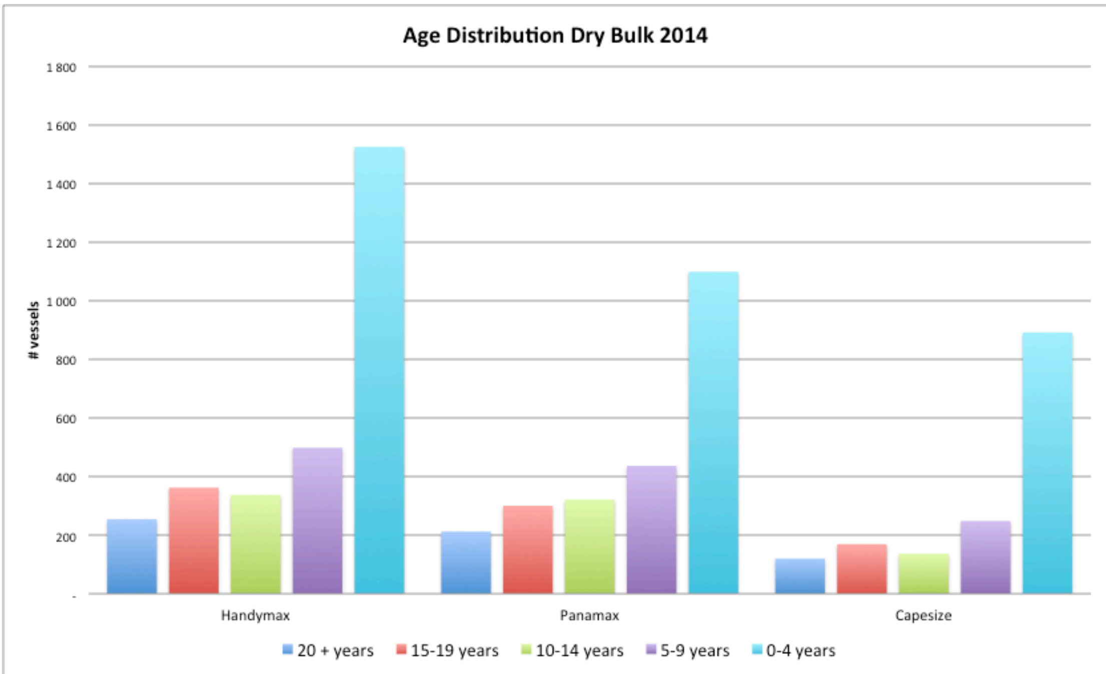


Figure 4-1 – Age distribution in the dry bulk fleet 2014 (Clarksons Intelligence Network, 2014)

As illustrated in figure 4-1 there has been a significant shift towards newer tonnage in the market. Approximately 50% of the current fleet is between 0-4 years old, reflecting the amount of new tonnage entering the market post the financial crisis. Intuitively this indicates that compared to the situation before the financial crisis, the distribution of tonnage is skewed towards newer vessels. In a presentation held by Clarksons for DNV-GL (Clarksons Research Services, 2014), they argue that between 2003-2008 the industry ordered over $800 billions worth of new ships whereas 50% of the orders were placed in the booming years 2007-2008. Furthermore, majorities of the investment were related to standard designs and technology. A somewhat fair assumption would be that shipowners ordering vessels between 2007-2008 was less concerned about efficiency and emphasized on quick deliveries of ships that could enter the highly profitable market. Since the time aspect of delivery was of great importance, shipyards delivered highly standardized ships. In this regard it could be difficult to see any significant efficiency gains across vessels in recent years. ECO-ship is a word prevailing in recent years as a term describing efficient vessels. The definition of an ECO vessel is somewhat relative to the concept of ageing and fuel thirsty vessels. Modern ECO–ships are not necessarily cheap as they features state-of-the art design and equipment including sophisticated engineering solutions, which partly due to an improved focus on environmental issues has become an increasingly important part of the agenda.

4.2. Efficiency

A major factor associated with the operation of vessels is fuel costs, which fluctuates as a proportion of overall costs between ship types and sizes. Between May 2006 and January 2014 the 380cst bunker price in Rotterdam increased from 324 USD/ton to 570 USD/ton. Notwithstanding the fact that future fuel prices are highly uncertain, the maritime industry will most likely face increasing prices as oil demand from developing countries rises, oil scarcity is likely to increase and the implementation of emission regulations. Also, geopolitical issues in North America and the Middle East will keep bunker prices volatile, at best. Bunker expenses are estimated to represent approximately 60% of total freight costs (Loyd's List, 2012), which gives owners and charterers an incentive to focus on energy efficiency.

According to the 2nd GHG IMO study (Buhaug et. al, 2009), improved energy efficiency means that the same amount of useful work is done, while consuming less energy. Consequently, as a result less fuel is burned and emissions of exhaust gases are reduced. From a regulators point of view the goal is to encourage efficiency and environmental objectives while at the same time sustain the interests of market participants. There are several options for improving energy efficiency, which can be divided into two main categories (Buhaug et. al, 2009):

Technological measures through new-building or retrofitting processes:

- Concept, design speed and capability
- Hull and superstructure
- Power and propulsion systems

Operational measures through vessel operations:

- Fleet management, logistics and incentives
- Voyage optimization
- Energy management

In addition, a less prevailing but increasingly important, market-based mechanisms such as emissions trading and carbon taxes can be applied. Although there are numerous technology-based approaches to improve vessel efficiency, operational changes have been promoted as the most cost-effective change to meet high fuel prices and volatile rates, especially in recent times where rates have been declining and price of bunkers rising, illustrated in figure 4-2.

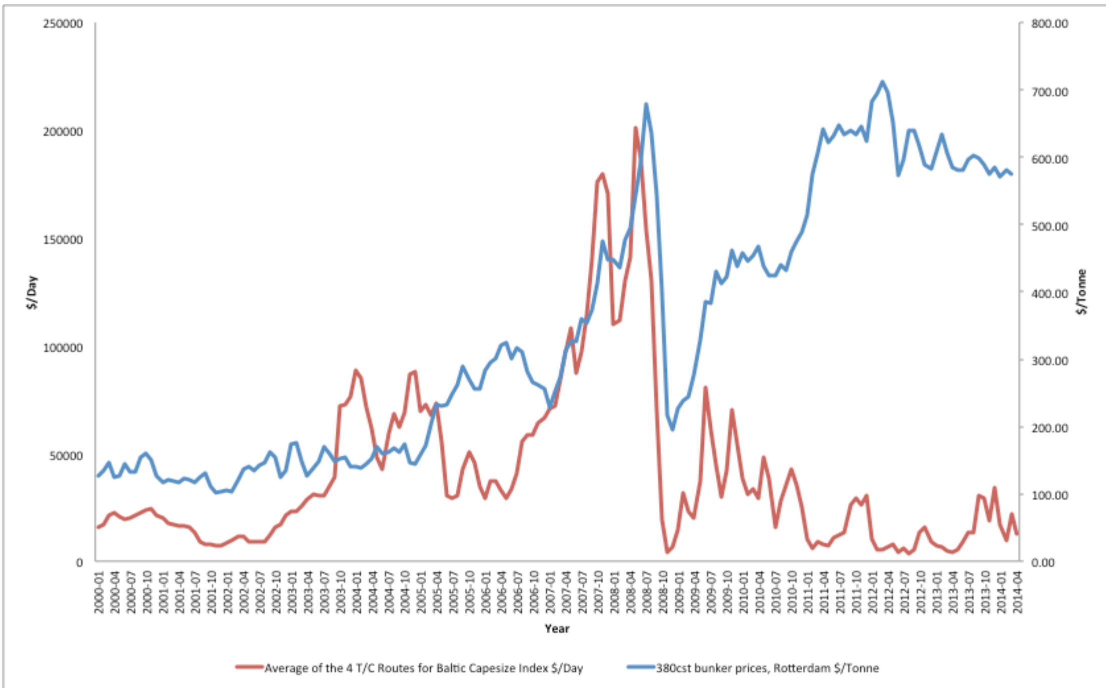


Figure 4-2 - Average BDI 4 TC rates Capesize and 380cst bunker price. (Clarksons Shipping Intelligence Network)

4.3. Efficiency and Time Charter Rate Dynamics

Challenges such as the downturn in the global economy, oversupply, high fuel costs and falling freight rates put a substantial pressure on the maritime shipping value chain. In response the industry has to adapt quickly, which has resulted in a decline in newbuilding orders and more frequent lay-up and idling activity. Faced with this development, the market needs a commercially viable way to reduce costs quickly in order to stay in business. Fuel costs plays an important role in this development and one of the easiest and most effective ways to cut cost is through speed reductions.

Profit maximizing speed is given by:

$$V^* = \left(\frac{24 * R * W}{D * \beta * P_b * \alpha}\right)^{\frac{1}{\beta-1}}$$

R=spot rate, W=Cargo intake, D=Distance, V=Speed, Pb=Bunker price, F=Fuel consumption and β and α are constants.

Clearly the equation shows that as bunker prices increases or rate decreases, optimal speed is reduced. As a result one would expect reduced speed over the past years along with higher actual speed for premium-paying routes along with higher speed for legs with lower consumption, such as ballast legs (Adland, 2013b).

According to recent shipping literature, ships operating at lower than design speed is said to be slow steaming[8]. Reduction in speed means that the engine is continuously running at a low load. To handle possible challenges when deviating from optimized load, operational measures needs to be followed. However, even though two-stroke engines are optimized for a load range of approximately 60% CMCR[9], running at loads down to 10% is still possible without any engine modification (The Motorship, 2012). In addition to engine limitations, factors such as minimum steering (6-7 knots) and contractual clauses sets external limits on speed (Adland, 2013b).

Cost savings in a volatile, competitive and homogenous market is crucial and steaming at reduced speed may be the most effective measure to cut costs and reduce overcapacity. Reduced speed results in a longer voyage time, which again would reduce available transport capacity. Even though owners has the option to put vessels in lay-up, slow steaming would be preferable as it offers greater flexibility to meet market cycles and add capacity when the cycle turns.

---

[8] No clear definition on the term “slow steaming”, however reduced speed relative to “normal” speed is usually applied.

[9] Contract Maximum Continuous Revolution

4.3.1. SPEED-CONSUMPTION RELATIONSHIP

The graph below illustrates the difference between fuel consumption for two ships with different speed/consumption relationships. As knots increase, consumption unsurprisingly increases. However the curve is exponential, indicating a greater advantage for low consumption vessels when steaming at higher speeds.

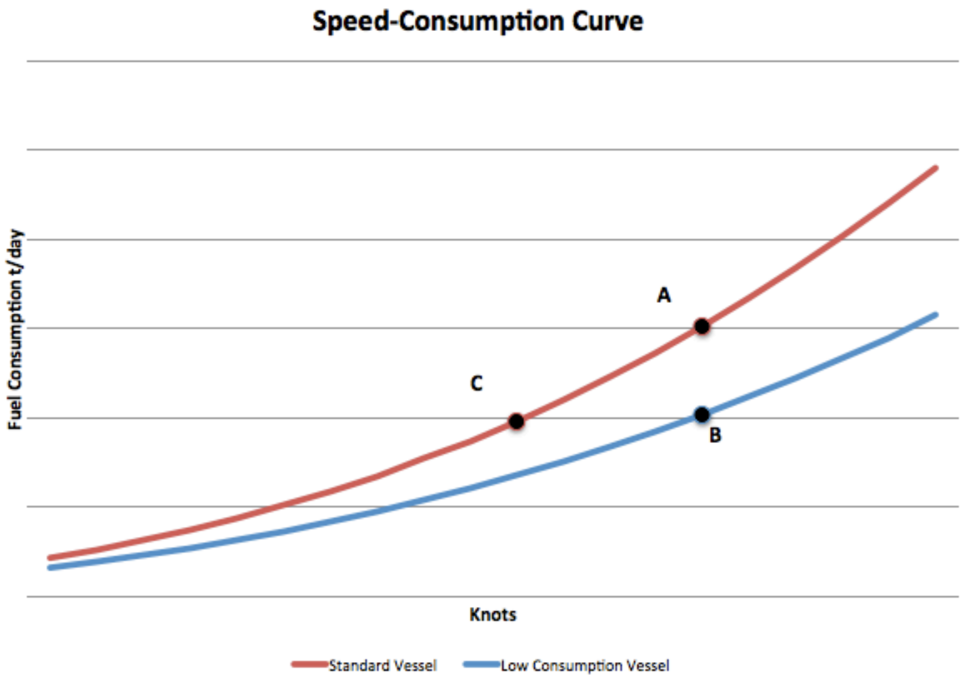


Figure 4-3 - Speed/Consumption-curve (Adland, 2013b)

Ideally a charterer would be able to save fuel cost corresponding to (A-B)*Price of bunkers. From another viewpoint he could operate a low consumption vessel at a higher speed while incurring the same fuel costs where the difference in speed equals B-C. Obviously as knots increases, fuel costs decrease although this advantage is more favorable at speeds above the slow steaming trends we have seen emerging in recent years. Current average speed for bulkers is approximately 2 knots below pre-recession standard speeds (Clarksons Research Services, 2014). Henceforth we would assume the benefits of eco-ships to be greater in good times where slow steaming is less preferable. Note that in this scenario we assume that markets are competitive and shipowners are price-takers, hence we do not incorporate increased capital costs for shipowners due to increased costs for efficient ship measures and designs. If we were to study how owners would react, an analysis of how capital costs increase and at what breakeven point for speed low consumption vessels becomes profitable compared to standardized vessels is needed.

Charles R Weber, an independent US shipbroker firm, estimates that operational retrofitting could lead to 7.5% - 11.5% in fuel savings for a non-eco Capesize vessel (McCarthy, 2013). For a newbuilding the same savings amount to 10.5% - 16% through improved hydrodynamics and a new G-Type engine (MAN D&T, 2012) . Although savings are expected to be greater for newbuildings they also need a higher breakeven rate to compensate for higher CAPEX. Intuitively it might seem more reasonable for eco-ships to operate in the spot market rather than on TC-rates. When fixed on long-term rates the shipowner might not be able to recover increased CAPEX due to the contractual agreements on bunkers cost and the charterer will have greater incentives to focus on fuel consumption as discussed in section 3.2.

4.3.2. Efficient Time Charter Market Dynamics

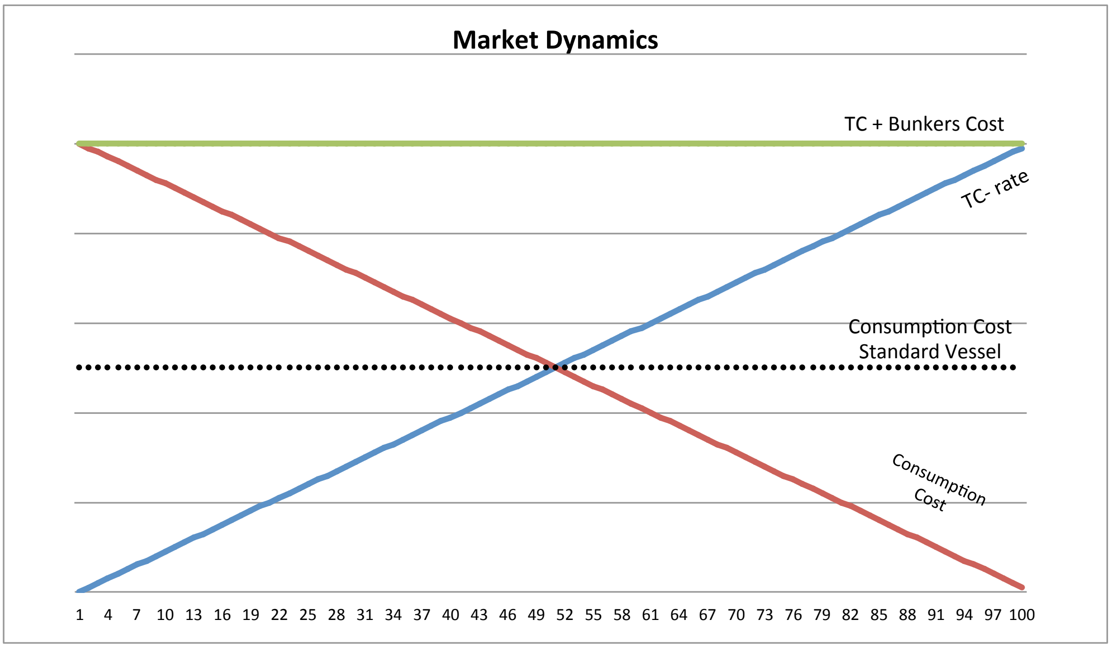


Figure 4-4 - Efficient Time Charter Market Dynamics

Following our argument in the previous section a simplified illustration of efficient market dynamics can be shown in figure 4-4 where bunkers price is constant. The x-axis represents vessels where #1 and #100 is the least and most fuel-efficient, respectively. A charterer would be interested in the TC-rate plus the total bunkers cost for one specific vessel and the rate is observed from indices where the last quoted rate is presumed to be the reference. Assume a charterer choosing to hire vessel #70. In a competitive market he would be willing to pay a higher TC-rate compared to all vessels below #70 due to fuel cost savings. This rate premium equals the difference in consumption and he would be willing to pay maximum TC + Bunkers cost. This situation would describe an efficient market.

However, if the agreed TC rate is deviating from the blue line, signs of market failure occur. Information asymmetries would work in the shipowners favor with respect to rate level. Surveys conducted by DNV-GL suggest that charterers might not be in a position to discover, or at least take advantage of, operational efficiencies. Although the charterer has operational control it is difficult to know whether the vessel has incorporated operational measures such as optimal trim or efficient propeller polishing. Although, more interesting and powerful is the presence of incentive barriers, discussed in section 3.2 and 4.3.1. Noticeably the charterer wants to hire the ship with the best specifications. However, it is questionable whether he is willing to pay a premium since the improved performance is exactly what he wants to keep and the shipowner might not be in a position to influence the rate. This argument proposes that the rate should be somewhere along the dotted line in figure 4-4. Should charterers identify this improved performance, one would assume the demand effect to increase rates in the time charter market. The NPV of this premium should at least equal increased CAPEX due to higher capital investments for more efficient vessels in order to incentivize shipowners. If this is not the case, and the shipowner is not compensated, he would be inclined to offer the vessels in the spot market instead and capture all the performance savings, assuming the whole market to be efficient. Contrary to our previous arguments, this suggests that the rate should be somewhere between the dotted and blue line.

In essence this graph demonstrates that a ship with relatively higher fuel efficiency than the ship underlying the market index should receive a relatively higher rate compared to the index. Hence the rate should, according to our theory, be somewhere on the blue line in figure 4.4.

In the following sections we pursue to verify to what extent efficiency is priced in the dry bulk market. Through a decomposition of the time charter rate we seek to reveal rate determinants and identify quality differences and their potential premiums.

5. METHODOLOGY

The General Linear Model is widely used and desirable because it's simple to fit, results are easy to interpret and there are a wide variety of useful techniques for testing the assumptions involved. However, intrinsic nonlinearities in the data may require semi - parametric modeling.

Since Generalized Linear Models (GLM) are specified in terms of the linear predictor many of the general concepts of linear modeling is applicable, although with some modifications. It allows for incorporation of other types of distributions of the exponential family and includes a link function, which links the estimated fitted values to the linear predictor (Wood, 2006). In essence GLMs are a generalization of the typical linear model where fewer assumptions are made. A basic structure of a GLM would look like

$$g(\mu_i) = X_i\beta_i \quad (1)$$

Where $\mu_i \equiv E(Y_i)$, g is a smooth monotonic link function, X is the i^{th} row of a model matrix, X, and β is a vector of unknown parameters. Y_i are independent and follow some exponential family distribution, $y \sim ExpoFam(\mu, \sigma^2)$.

If a Gaussian distribution is assumed for the response variable along with equal variance of all observations and a direct link of the linear predictor and the expected value, i.e. $X\beta = \mu$, equation (1) would equal a typical linear regression.

5.1. GENERALIZED ADDITIVE MODELS

The Generalized Additive Model (GAM) is a subsequent step in the generalization process and is often seen as an ideal compromise between simple linear models and complex models such as neural nets (Beck, Jackman, 1998). The model uses an algorithm to fit a smooth curve to each variable and combine the results additively. Along with our assumptions as well as findings based on previous research there are several factors influencing the determination of time-charter rates in the dry bulk market that would suggest a non-linear relationship between rates and their determinants. (Adland R., Cullinane K., 2006) (Dick et. al, 1998). GAMs are designed to capitalize on the strengths of GLMs without requiring a priori assessment of response curve shape or specific parametric function estimates. A set of smoothers is employed in order to generalize data into smooth curves by local fitting to subsections of the data. The idea is to plot the value of the dependent variable along a single independent variable and then calculate a smooth curve that passes through the data as good and parsimonious as possible.

An approach to include nonlinearities is to apply a semi-parametric specification within a GAM model. This model offers enough flexibility to take such relations into consideration without making any specific assumptions regarding the functional form of the variables, except additivity of the model terms (Koehn, 2008). According to Hastie and Tibshirani (1990) the idea is to let the data dictate the relationship between the response variable and the explanatory variables. By assuming less, the model will hopefully discover more. When using samples of adequate size with a large number of variables there could be a problem with obtaining reliable results: as the number of predictors increases, the number of points in a local neighborhood for a particular focal point decreases and the regression becomes less local leading to an increase in bias of the estimates. (Andersen, 2007). This issue is referred to as the curse of dimensionality and can be avoided by applying a semi-parametric model. (Koehn, Thanopoulou (2011). However, this flexibility comes at a cost: A smooth function must be chosen and we need to decide how smooth it should be, represented by the basis dimension.

Some theoretical background on GAM models can be found in the following section.

A general GAM model would look like

$$g(\mu_i) = X_i^*\theta + f_1(x_{1i}) + f_2(x_{2i}) + \cdots + f_n(x_{ni}) \quad (2)$$

Where $\mu_i \equiv E(Y_i)$ and Y_i is a response variable following some exponential family distribution. X_i^* is a row of the model matrix for any strictly parametric model components, θ is the corresponding parameter vector and f_i are smooth functions of the covariates.

The main difference between 1 and 2 is that in the latter the linear predictor incorporates some smooth, $f(x)$,function of the covariates. The model allows for flexible specification of the dependence of the response on the covariates, and by using smoothing functions to specify the model we can avoid cumbersome modeling (Wood, 2006).

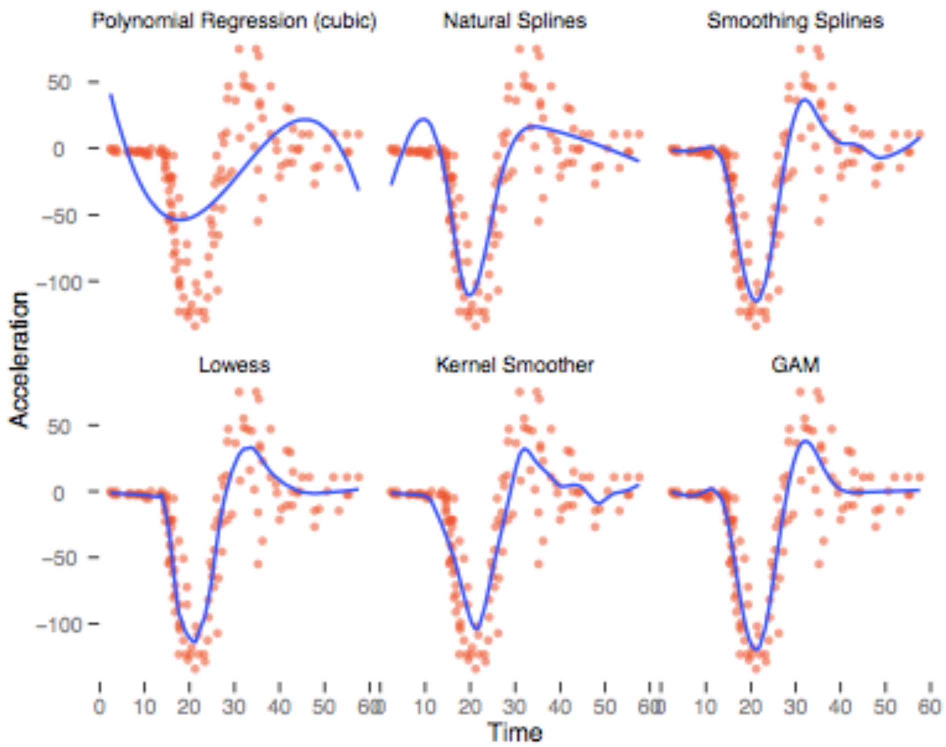


Figure 5-1 - Regression Fits (Clark, 2013)

An Illustration of how GAM modeling works is shown in the figure above. The example shows several different regression approaches. Obviously a simple polynomial regression will fail to capture some of the data relationships with a simple transformation of the predictors. A more flexible model, such as GAM, is needed in order to obtain a best possible fit of the data, as shown in the bottom right illustration.

Consider the following model

$$y_i = f(x_i) + \varepsilon_i \qquad (3)$$

where y_i is the dependent variable, x_i is a covariate, f a smooth function and ε_i a random variable with N(0,σ^2). For simplicity assume x_i to lie between (0,1). If we were to use a standard technique such as OSL, f need to be represented in a way that (3) becomes a linear model. By using a basis, defining the space of functions of which f is an element, this linear approximation is possible. This basis is often called a "spline-basis". Now, consider the following model

$$f(x) = \sum_{j=1}^{q} b_j(x)\beta_j \tag{4}$$

where $b_j(x)$ is the j^{th} basis function with unknown parameters β_j. Substituting (4) into (3) yields a linear model.

5.2. SMOOTHER

A smoother is a tool for summarizing the trend of a dependent variable as a function of one or more independent variables. Its called a smoother since it produces an estimate of the trend that is less volatile than the variable itself. The most important property of a smoother is its non-parametric nature. It does not assume a rigid form of the dependency between dependent and independent variables. Further it allows for an approximation with a sum of functions and not just one unknown. When estimating GAM models, the trick is the parsimony of the smooth curve. The curve might wiggle considerably and fail to represent a parsimonious fit. By dividing the data into subsections, joined together by knots, a low order polynomial or spline function is used to fit the curve where the condition is that the second derivative for each section joined together by a knot must be equal. This will result in a smooth and continuous curve (Wood, 2006)(Hastie, Tibshirani, 1990).

In order to illustrate, imagine a univariate function, which can be represented using a cubic spline in the figure below. This is a curve made up of sections of cubic polynomials, joined together so that they are continuous in value as well as first and second derivatives. The points where the sections join are known as the knots of the spline. For a conventional spline, the knots occur wherever there is a datum. However, for other regressions splines the knots must be chosen. Knots may be evenly spaced through the range of observed x values or placed at quartiles of the distribution of unique x values.

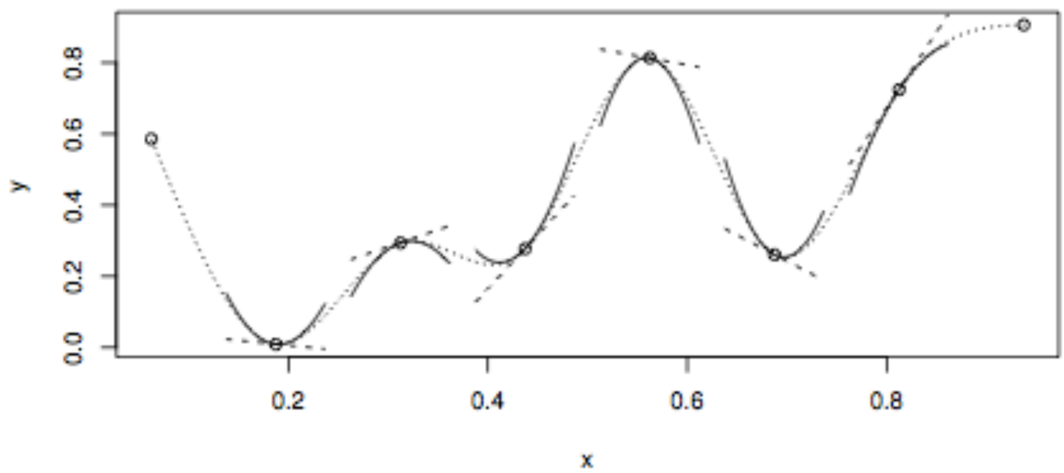


Figure 5-2 - Cubic Spline Interpolation (Wood, 2006)

5.3. Degree of Smoothing

For any regression spline the basis dimension is crucial for the degree of smoothing. The basis dimension could be determined through backward-selection and hypothesis testing. However, this is cumbersome and problematic since a model based on k-1 evenly spaced knots is not necessarily nested within a model based on k evenly spaced knots. Backwards selection of knots is a possibility, but uneven knot spacing can lead to poor model performance. Also, for regression spline models, model fit tends to depend strongly on the locations chosen for the knots. (Wood, 2006).

One alternative to controlling the smoothness by changing the basis dimension is to fix the basis dimension at a size slightly larger than believed to be necessary and control the smoothness by adding a penalty for wiggliness (Koehn, 2008). Following this procedure the trade off between goodness of fit to the data and the wiggliness of the function can be controlled by a smoothing parameter. A smoothing parameter reaching infinity leads to a straight line estimate for $f(x)$, while a parameter equal to 0 results in an un-penalized regression spline estimate. In essence, the problem of estimating the degree of smoothness for the model now becomes a problem of estimating the smoothing parameter.

Consider the following equation

$$\sum_{i=1}^{n}(y_i - f(x_i))^2 + \lambda \int_a^b [f''(x)]^2 dx$$

Where λ is a fixed smoothing parameter with respect to the unknown regression function that is found on the basis of the data and $a \leq x_1 \leq \dots \leq x_n \leq b$.

Since we want to smooth the data instead of interpolating, a (cubic spline) smoother is a solution to the optimization problem above: among all functions $f(x_i)$, with second continuous derivatives, find one that minimizes the penalized least square.

The first term of the equation represents the OLS method. Only considering this part yields an interpolated curve, which would not be particularly smooth since it only measures closeness to the data without any curvature penalization. The second term measures the wiggliness of the function and penalizes curvature in the function. By applying a linear function the last term would equal 0.

5.4. Choosing an Appropriate Value for the Smoothing Parameter

Assuming the basis dimension is large enough so that it is more flexible than we expect in order to represent $f(x)$, the exact choice of basis and the precise selection of knot locations do not have significant influence on the model fit. It is rather the value of the smoothing parameter that determines model flexibility. Generalized Cross Validation (GVC) is an estimate of the mean square prediction error based on a leave-one-out cross validation estimation process (Clark, 2013). It resamples the original sample and is the most commonly used method to choose the smoothing parameter. It uses n subsets of the data where each subset removes one observation from the dataset. In simple terms, GCV compares the fit of all models based on all possible values of the smoothing parameter and chooses the one with best fit. When determining the specific nature of the smoothing parameter, it is the GCV score that is minimized.

$$GCV = \frac{n * scaled\ estimators}{(n - edf - [number\ of\ parametric\ terms])^2}$$

5.5. THE THIN PLATE REGRESSION SPLINE

Although there are several bases that could be used for modeling purposes, some impediments could pose particular concerns. The choice of knot locations introduces subjectivity to the model fit, certain bases are only useful for smoothing of one predictor variable and it is unclear whether some bases are better or worse than others. All these issues could to a degree be addressed by using thin plate regression splines (TPRS), producing knot free bases, for any number of predictors, that are in a limited sense optimal. In addition this spline treats the wiggliness in all directions equally (Wood, 2006).

TPRS introduces a sophisticated and general solution to the smoothing estimation problem. According to Wood (2006), TPRS might be the closest to an ideal smoother we can find. It has been constructed by defining exactly what is meant by smoothness, exactly how much weight to give to the conflicting goals of matching the data and finding the function that best satisfies the smoothing objective. In other words we avoid the problem with knot placement and it is relatively cheap to compute.

TPRS estimates the smooth function by finding the function $\hat{f}$ minimizing:

$$\|y - f\|^2 + \lambda J_{md}(f)$$

Where y is the vector of y_i data and $f = (f(x_1), f(x_2), \ldots, f(x_n))^T$. $J_{md}(f)$ is a penalty function measuring the wiggliness of f and λ is a smoothing parameter, controlling the tradeoff between goodness of fit of the data and smoothness of f.

It is worth to keep in mind that the exact size of the basis dimension is not that critical. It sets only an upper bound on the flexibility of a term and it is the smoothing parameter that controls the actual effective degrees of freedom. Consequently, the model fit is rather insensitive to the basis dimension, assumed that it is not set restrictively too low to capture the true effects in the best way possible. Although GAM has its modeling simplicities, it should be noticed that there are some drawbacks: Hypothesis testing is merely approximate and by using a Bayesian posterior covariance matrix to estimate satisfactory interval estimates, p-values tend to be rather low since they are conditional on the smoothing parameter, which is uncertain (Koehn, 2008). Hence results based on significance have to be interpreted with caution. As with other nonparametric techniques, theory and mathematical foundations for GAMs is a complex. Henceforth we outline only selected and relevant topics and refer to literature by Hastie and Tibshirani (1990) and Wood (2006) for further discussions.

6. Empirical Analysis

We have chosen the three main segments in the dry bulk market for three main reasons. Out of the current markets, dry bulk is the one with largest activity within the time charter market. Furthermore the market is desirable from a modeling point of view as it features merely comparable characteristics. In addition the market has been subject to limited research using modern econometric techniques.

6.1. Data Description

We use cross-sectional data provided by Clarksons Research Services for TC contract fixtures in the dry bulk market. The dataset contains 10738 reported contracts for Capesize, Panamax and Handymax from 1st of January 2001 until 16th of May 2014. For modeling purposes the dataset is divided in two periods; 2001-2007 and 2008-2014. The first period was an extremely good period with historically high rates, while the latter is best characterized as a market trough. Some fixtures have been excluded to create a complete and consistent dataset. Reported and non-positive values for rate, age, dwt, speed, consumption and contract length are assumed to be necessary for our study. Therefore any fixture not fulfilling these criteria is removed. Since our dataset is large and stems from an external source, potential outliers could pose modeling interference. Trimming of data to remove low-density observations etc. is found in comparable studies and might result in narrower confidence band. However, we chose otherwise to keep the data as authentic as possible. Nevertheless, obvious misspecifications are excluded or altered after thoroughly investigating data entries. The issue was found most prominent for option and contract length, number of days forward, dwt and age. Furthermore, fixtures with obviously misreported data are also omitted[10]. After filtering out irrelevant, missing and misreported values our dataset consists of 7763 observations, or approximately 75% of the initial data.

6.2. Variable Description

Since our goal is to investigate the relationship between ship specifications and TC-rate we have included both macroeconomic and microeconomic variables assumed to be of relevance for rate determination. Microeconomic aspects are captured through ship and contract specific variables.

Ship specific:

- Age
- Dwt
- Fuel efficiency index (FEI)
- Fuel
- Flag
- Builder country (BUILD)
- Engine type (ENGT)
- Gear
- Speed
- Consumption

Contract specific:

- Option length (OPTL)
- Contract length (CONTL)
- Number of days forward (NDF)
- Place of delivery (POD)

Macroeconomic elements are captured through market specific variables:

- Index
- Bunkers cost

[10] Some ships were reported with fixture values considered to be highly unlikable.

Information considering TC-rate, speed, consumption and dwt is fully available in reported fixtures. Note that since we use FEI (consumption per ton/mile), we do not explicitly include speed and consumption as they are highly correlated. Age is a explanatory factor for rate determination and for our purpose it is applicable since we use three size segments across time periods.

6.2.1. *Smoothed Variables*

As a representation of the market weekly TC indices, figures for all vessel segments are gathered from Clarksons Shipping Intelligence Network. These indexes are matched by closest date (before and after actual date), and by ship type. Clarksons define "standard" ships as underlying for the indices, divided into several different ship types and specifications. This would adjust the technical specifications of a standard ship in order to reflect changes in fleet development over time. As size, speed and consumption changes, the underlying specifications of the standard ship used as a proxy for the market is altered accordingly. Since adjustment of the underlying index is changed over time, we suspect efficiency improvements to be incorporated in the index; hence it would be difficult to detect how each ship specification is reflected in TC-rates. This is an issue not properly addressed in previous research. A standardization of data to ensure that the same underlying ship specification is used as a comparable for the index during the whole time period would be preferable. After studying how Clarksons adjust their indexes we found that no significant adjustments were made and reported indexes are suitable.

Three different indices for each ship size group is used; 1-month, 6 months, 1 year and 3 year contracts. Since there are no reported figures for the 1-mont index, average spot earnings are used as a proxy[11]. The various indexes would be piecewise linear between each contract length, but obviously not linear in time. As a result a contract that does not match any index with respect to length would obtain a value that does not reflect the actual contract length. To solve this non-linearity issue we use a smooth cubic normal spline function that passes through each observation non-

[11] Spot earnings exclude contractual TC costs and are therefore assumed to be an appropriate proxy.

linearly to determine a correct index proxy. We believe this proxy is a better representation of the market since its possible for us to eliminate the time component and only differentiate on ship specifications. In this case, use of BDI does not reflect differences with respect to contract lengths. The effects of option length are not addressed. (See Adland & Koekebakker (2006))

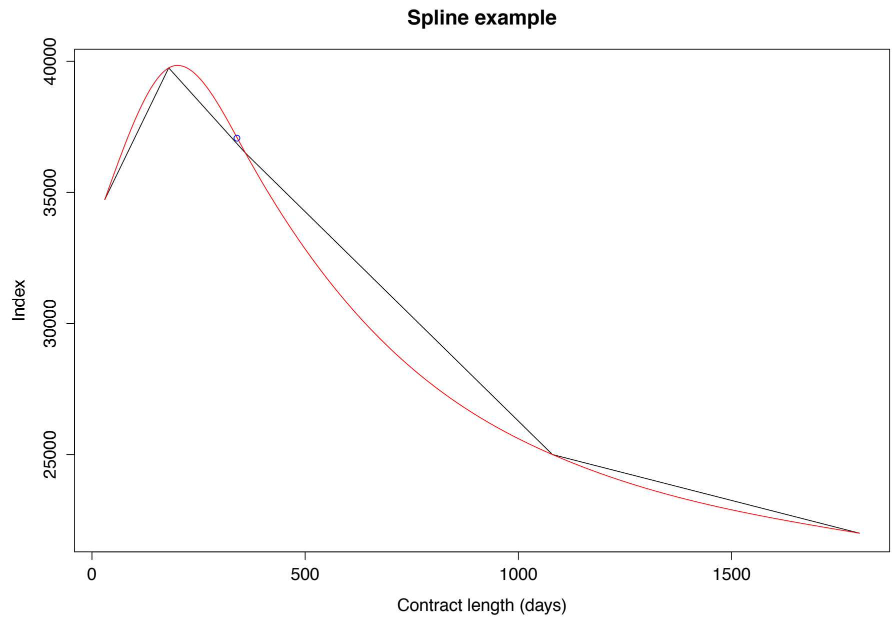


Figure 6-1 - Index Spline Interpolation

Figure 6-1 illustrates an example of the spline interpolation based on given index contract lengths. The dark line represents the piecewise linearity between actual contract lengths, while the red line is the spline-interpolated curve. Clearly contracts between 500 and 1000 days would obtain a higher index value based on linear assumptions than by assuming a non-linear relationship.

Prior to 2008 there exists no index values for 5-year contracts. Therefore the closest approximation was to use the 3-year index as a proxy. This might overestimate the index proxy values since the spline now assigns 3-year values to 5-year contracts. However, only 2% of the contracts in our dataset, evenly distributed among segments, are affected.

A comparable bunkers price is obtained though the same source as for index data. Data from all stated suppliers are gathered for HFO and IFO and further averaged to create two bunkers indices. The data are matched by ship fuel type and closest date to assign one bunkers price per fixture. Fixtures containing HFO fuel type is matched to the HFO index. Since our dataset only contains fixtures with HFO and IFO fuel type, those not containing any bunkers information are assigned IFO.

By constructing a parameter measuring consumption in tons per million ton-mile we are able to compare consumption between vessels of various size. Initially we included consumption as a measure of efficiency, though no significant effect was observed. Another reason for using FEI is the correlation between speed and consumption. By including FEI and excluding consumption and speed we assume to remove some modeling bias.

$$FEI = \left[\frac{consumption\ tons/day}{dwt * speed * 24}\right] * 10^6$$

As earlier studies have seen signs of a two-tier market (Koehn, 2008)(Strandenes, 1994), we correspondingly assume age to be a significant indicator of quality differences in the market. The age we use in our model is computed as the difference from the contract year and the year of build.

Since there is a speculative element of forward and time charter contracts the timing of the contracts is an important element. Laycan period and contract date are both reported in our fixture data. Although it is uncertain exactly at what date the vessel is supposed to be delivered we use the latest date as an assumption along with the stated contract date.

6.2.2. Factor Variables

Factor variables take on a limited number of different values, which are often referred to as categorical variables. These variables enter into statistical models differently than continuous variables and storing data as factors insure that the modeling function treats such data correctly. In our model we use factor variables to quantify textual variables. Usually they are also known as dummy variables.

The only information regarding place of delivery in our dataset is delivery area. Timmermann & McConville (1996) and Rowlinson (1996) discuss the distribution of quality tonnage between the Pacific and Atlantic Ocean where the authors argue that there has been an increasing concentration of high quality tonnage in the Pacific basin at the expense of available capacity in the Atlantic. As a consequence, market rates between these two oceans are found to differ significantly (Koehn, 2008). We have therefore searched and matched each fixture to either the Pacific or Atlantic Ocean based on proximity.

According to fixtures data, contract length is given as a time period, e.g. 5-8 months. We have interpreted the first value as contract length and set the option period equal to the difference, which would be a fair assumption. In this example the contract length and option period would be 5 and 3 months, respectively.

According to representatives from DVN-GL, these variables could represent quality differences in tonnage. For both variables the dataset has been sorted descending and by accumulated percentage for comparable reasons. All observations within the 90% accumulated range are used as individual countries or flags, while those outside this interval are categorized as "other" within these variables.

6.3. Summary Statistics Smoothed Variables

	CAPESIZE							
	2001-2007		n=749		2008-2014		n=494	
	mean	sd	min	max	mean	sd	min	max
RATE	47019	32526	7000	178000	45194	45376	6000	210000
NDF	31	48	0	453	20	40	0	380
AGE	9	6	0	28	8	6	-1	27
DWT	163568	14118	111695	211485	171062	11375	100336	239999
SPEED	14.1	0.8	11.5	17.0	14.5	0.6	12.0	16.2
CONS	54	7	32	90	57	7	32	75
FEI	0.9709	0.1205	0.5410	1.8400	0.9625	0.0870	0.5410	1.3000
CONTL	407	466	30	4320	283	320	60	3600
OPTL	40	34	0	360	60	35	0	270
BUNKERS	261	95	108	520	538	126	228	765
INDEX	52881	35570	9563	187552	50519	49795	3544	200096

PANAMAX

	2001-2007		n=2803		2008-2014		n=2186	
	mean	sd	min	max	mean	sd	min	max
RATE	31489	19386	3750	105000	29067	22732	4000	100500
NDF	18	37	0	558	14	30	0	702
AGE	7	5	0	30	8	5	0	39
DWT	73020	3469	60158	98362	75337	3852	61455	95712
SPEED	14.1	0.6	12.0	17.8	14.2	0.5	11.5	16.3
CONS	33	4	22	82	34	3	22	55
FEI	1.3417	0.1521	0.9910	2.6800	1.3183	0.1052	0.9910	2.2900
CONTL	232	223	5	1800	217	219	4	3600
OPTL	49	35	0	720	58	36	0	720
BUNKERS	264	98	107.04	527.61	536	124	228.04	778.1
INDEX	31623	19200	6181.4823	93002.016	29287	22691	3476.4321	87592.178

HANDYMAX

	2001-2007		n=687		2008-2014		n=843	
	mean	sd	min	max	mean	sd	min	max
RATE	28750	14576	4200	77000	22774	16615	3000	89000
NDF	18	32	0	411	8	22	0	424
AGE	7	6	0	27	6	5	0	29
DWT	49240	4296	40048	59076	52725	4131	40045	59888
SPEED	14.1	0.6	10.5	17.0	14.3	0.5	10.5	16.9
CONS	29	4	10	48	30	3	14	61
FEI	1.7475	0.2237	0.7400	3.8100	1.6753	0.1679	0.7400	3.2600
CONTL	215	197	10	1800	169	141	40	1080
OPTL	47	24	0	90	47	26	0	210
BUNKERS	288	94	107.89	515.92	544	125	228.04	778.1
INDEX	30025	15414	7100	73500	23886	17158	7750	70500

Table 6-1 - Summary Statistics Smoothed Variables

Across all segments and both time periods there is a great variability in observed rates, illustrated by standard deviations. However there seems to be consistently higher rates for Capesize vessels. Although the volatility increases in all periods the mean rate is lower in the last period. Average numbers of days forward, contract length and option length is higher in the first period along with higher volatility. In addition option length is longer in the last period. Contract length has increased with size and decreased between the two periods. This could illustrate that charterers prefers more flexibility post 2007 and are reluctant to enter relatively long-term contracts. Average age is fairly unchanged, which could seem contrary to our suggestion earlier where 50% of the current fleet is between 0-4 years; however, note that the vessel age in our dataset is reported from the contracting day. The average Fuel Efficiency Index (FEI) for all segments has declined; thus it could be reasons to expect some efficiency improvements. As anticipated there is no significant differences in speed due to the fact that reported speed in contract fixtures are design speed, which haven't changed considerably. The same argument is also valid for consumption.

6.4. Additional Variables

Fuel type, engine, gear and ship discharge are also considered as important factor variables influencing the rate. However, the distribution within our dataset might suggest that they are not suited for modeling purposes. Table 6-2 includes a description of the factor variables considered.

China has increased its share of total vessel supply in our reported dataset and the distribution is considered to be suitable for model inclusion. Between fuel types there seem to be an even distribution between HFO and IFO, although HFO has reduced its share for both Panamax and Handymax. On the engine side MAN B. & W. is clearly the leading provider. The delivery pattern between Pacific and Atlantic seems to be consistent for Handysize, while Capesize and Panamax delivery in the Pacific seems to be dominating. The distribution of loading and unloading gear on vessels is obvious. Capesize vessels do not carry such gear, only one percent of Panamax and 98% of Handymax. Between flag states there seems to be no obvious changing pattern, however Panama is the leading flag state and there seems to be a greater distribution within the Handymax segment.

	CAPESIZE		PANAMAX		HANDYMAX	
	2001-2007	2008-2014	2001-2007	2008-2014	2001-2007	2008-2014
BUILD						
China P.R.	9 %	26 %	9 %	19 %	12 %	28 %
Japan	33 %	31 %	63 %	60 %	66 %	53 %
South Korea	37 %	33 %	21 %	16 %	7 %	4 %
Other	20 %	11 %	7 %	5 %	15 %	15 %
FUEL						
HFO	77 %	81 %	73 %	67 %	69 %	56 %
IFO	23 %	19 %	27 %	33 %	31 %	44 %
ENGT						
MAN B. & W.	88 %	92 %	73 %	86 %	78 %	89 %
Sulzer	11 %	5 %	26 %	14 %	18 %	7 %
Other	0 %	3 %	1 %	1 %	4 %	3 %
Unknown	0 %	0 %	0 %	0 %	0 %	0 %
POD						
Atlantic	29 %	9 %	33 %	22 %	28 %	33 %
Pacific	62 %	87 %	61 %	76 %	68 %	66 %
Unknown	9 %	4 %	5 %	2 %	4 %	1 %
GEAR						
Yes	0 %	0 %	1 %	1 %	99 %	100 %
No	97 %	99 %	96 %	97 %	0 %	0 %
Unknown	3 %	1 %	3 %	2 %	0 %	0 %
DIS						
World Wide	95 %	99 %	95 %	96 %	92 %	88 %
Other	1 %	1 %	4 %	4 %	8 %	12 %
Unknown	4 %	0 %	2 %	0 %	0 %	0 %
FLAG						
Bahamas	3 %	1 %	3 %	4 %	3 %	3 %
China P.R.	2 %	1 %	9 %	3 %	3 %	2 %
Cyprus	2 %	2 %	5 %	5 %	1 %	1 %
Greece	9 %	14 %	10 %	11 %	6 %	5 %
Hong Kong	9 %	13 %	10 %	12 %	13 %	14 %
Liberia	11 %	7 %	8 %	10 %	6 %	6 %
Malta	9 %	10 %	7 %	6 %	10 %	11 %
Marshall Is.	6 %	13 %	10 %	13 %	14 %	19 %
Panama	24 %	19 %	22 %	23 %	21 %	21 %
Singapore	7 %	5 %	1 %	1 %	3 %	5 %
South Korea	5 %	6 %	2 %	2 %	2 %	1 %
Other	14 %	9 %	11 %	9 %	18 %	12 %

Table 6-2 - Factor Variable Data Description

6.5. CORRELATION

A correlation matrix can be used to gain some prior understanding of linear relationships between variables (Appendix 1). There is however worth to notice that these correlation coefficients reflects isolated relations between two variables and does not capture all non-linear or multivariate relations between two or more variables. Correlation only applies to linear relationships and even though there is a strong non-linear relationship, the correlation coefficient may be misleading due to curved relationships. Therefore one could expect deviations from the linear correlation matrix and the results from our model. Furthermore we refer to these relationships whenever found relevant in the following sections.

7. MODEL DESCRIPTION

7.1. MODEL SPECIFICATION

We apply thin plate regression splines as a basis to represent the smooth terms in our model, which selects the default dimension k=10. This sets an upper limit equal to k-1 in order to handle the identifiability constraint. For all models we apply a Gamma distribution on the response variable and the natural logarithm has been used as a link function. A log transformation of the response variable is often used to make regression modeling more applicable. Nevertheless, since we assume a Gamma distribution we see no need for a response variable transformation. Distribution of the response variable can be found in appendix 2.

7.2. MODEL SELECTION

Model selection is primarily based on a combination of data mining, maritime economic theory and rationale assessments where the first was applied to study patterns in our dataset whilethe two latter was used to assess each variables predicted influence on the response variable. Initially we listed all factors assumed to be of relevance, including ship-, vessel and contract specific variables. Furthermore we used an algorithm to find an optimized model GCV-score. This method raised some concerns since it always chose the model with larges amount of variables. A correlation matrix was then used to investigate linear relationships. Due to the nature of vessel characteristics, most of ship specific variables are highly correlated. Hence, dwt was chosen as measurement of size, FEI represents speed and consumption. Correlation between dwt and FEI was insignificant, thus we assume no model interference. Bunkers price and index showed a high correlation in the first time period and close to no relationship in the last. After model testing, excluding the variable did not change model fit or power and was therefore excluded in the first period for all models. A procedure similar to Koehn (2008), where we added the chosen variables stepwise proved additivity of the model. Explanatory power increased along with reduced GCV score. As a model check, plots of residual deviance vs. theoretical quantiles and residuals vs. linear predictors were studied[12].

---

[12] gam.check() in R was used to study model fit

We emphasize that the final model is not a result of data mining, but rater a reflection of the relationships we set out to analyze.

7.2.1. Capesize

$$g(E(RATE_i|.)) = \gamma_0 + s(INDEX_i) + s(BUNKERS_i) + s(CONTL_i) + s(AGE_i) + s(DWT_i) + s(FEI_i) + s(NDF_i) + \sum_{DIS} \gamma FUEL\, I_i^{FUEL} + \sum_{POD} \gamma POD\, I_i^{POD} + \sum_{FLAG} \gamma FLAG\, I_i^{FLAG} + \sum_{BUILD} \gamma BUILD\, I_i^{BUILD} + \sum_{ENGT} \gamma ENGT\, I_i^{ENGT}$$

7.2.2. Panamax & Handymax

$$g(E(RATE_i|.)) = \gamma_0 + s(INDEX_i) + s(BUNKERS_i) + s(OPTL_i) + s(CONTL_i) + s(AGE_i) + s(DWT_i) + s(FEI_i) + s(NDF_i) + \sum_{DIS} \gamma FUEL\, I_i^{FUEL} + \sum_{POD} \gamma POD\, I_i^{POD} + \sum_{FLAG} \gamma FLAG\, I_i^{FLAG} + \sum_{BUILD} \gamma BUILD\, I_i^{BUILD} + \sum_{ENGT} \gamma ENGT\, I_i^{ENGT}$$

As discussed, bunkers were removed from the first period due to problems of multicollinearity. Model testing indicated that there was no significant difference when the variable was excluded. Low data density on option length along with insignificant results suggested that the variable was removed from the Capesize model.

8. RESULTS

Since our models include both parametric and non-parametric variables the model output consists of two panels. Parametric regression is represented by point estimates along with significance and relationships can be interpreted directly. The absence of regression parameters for non-parametric variables reflects an important characteristic of GAMs; there are no coefficients for smoothed parameters. The fitted smooth curve needs to be plotted in order to reveal any estimated nonlinearities in the relationship between the dependent and smoothed parameters since the effects of the components differs with scale. Confidence bands are plotted as guidance to the degree of estimation error. The results are represented in terms of effective degrees of freedom (EDF), which reflects the degree of non-linearity, along with significance of the variables. Table 8-1 presents the results for our three models for both time periods.

	Capesize				Panamax				Handymax			
	2001-2007		2008-2014		2001-2007		2008-2014		2001-2007		2008-2014	
Parametric coeff	**Estimate**	**Sig.**	**Estimate**	**Sig.**	**Estimate**	**Sig.**	**Estimate**	**Sig.**	**Estimate**	**Sig.**	**Estimate**	**Sig.**
(Intercept)	10.6446	***	10.3770	***	10.1946	***	10.1387	***	9.8928	***	9.7800	***
BUILD Japan	0.0176		0.0710	*	-0.0241	**	0.0200	**	0.0199	.	0.0273	*
BUILD Other	-0.0126		-0.0017		-0.0333	**	-0.0121		0.0330	*	0.0321	*
BUILD South Korea	0.0134		0.0213		-0.0244	**	0.0016		0.0130		-0.0246	
FUEL IFO	0.0117		0.0139		0.0044		-0.0143	**	-0.0086		-0.0022	
FLAG China P.R	-0.1708	***	0.0888		0.0120		-0.0147		0.0544	.	-0.0282	
FLAG Cyprus	-0.0861	.	0.0531		0.0214		-0.0006		0.0319		0.0121	
FLAG Greece	-0.0765	*	0.0964		0.0110		-0.0191		0.0410	.	0.0213	
FLAG Hong Kong	-0.0912	**	0.0875		0.0172		-0.0066		0.0133		0.0085	
FLAG Liberia	-0.0912	**	0.0739		0.0213	.	-0.0077		0.0353		-0.0222	
FLAG Malta	-0.0825	*	0.0926		0.0600	***	0.0274	.	0.0439	*	0.0137	
FLAG Marshall Is.	-0.0595		0.0972		0.0302	*	-0.0059		0.0447	*	0.0251	
FLAG Other	-0.0887	**	0.1284		0.0271	*	0.0053		0.0326		0.0102	
FLAG Panama	-0.0859	**	0.0866		0.0238	*	-0.0039		0.0303		0.0085	
FLAG Singapore	-0.0684	.	0.1096		0.0342		-0.0324		0.0066		-0.0278	
FLAG South Korea	-0.0909	*	0.0258		0.0284	.	-0.0089		0.0268		0.0011	
ENGT Other	0.0605		-0.0728		0.0103		-0.0155		0.0197		-0.0405	
ENGT Sulzer	-0.0306	.	-0.0122		0.0014		-0.0050		-0.0042		-0.0449	*
POD Pacific	-0.0499	***	-0.1789	***	-0.0298	***	-0.1566	***	-0.0054		-0.1461	***
POD Unknown	-0.0283		-0.1611	**	-0.0389	**	-0.0869	***	-0.0433	.	-0.2219	***
Smooth coeff	**edf**	**Sig.**	**edf**	**Sig.**	**edf**	**Sig.**	**edf**	**Sig.**	**edf**	**Sig.**	**edf**	**Sig.**
s(NDF)	1.0010	.	4.1790	**	6.1720	***	5.8460	***	2.4390	*	2.8710	*
s(AGE)	3.5170	***	8.9920	***	7.8470	***	8.1350	***	1.6440	***	1.0000	***
s(DWT)	2.8550	***	8.4740	***	6.2550	***	4.1300	***	5.2160	***	2.0130	
s(FEI)	3.6060	*	2.3020		5.5510	**	6.1770	**	7.8840	***	1.3620	
s(CONTL)	6.8920	***	3.7730	**	8.1180	***	9.0000	***	6.8840	***	2.0830	.
s(OPTL)			-		8.3190	***	5.2880	***	1.8690	*	1.0010	**
s(INDEX)	5.7020	***	7.4730	***	8.8220	***	8.1980	***	7.0680	***	5.3880	***
s(BUNKERS)	-		4.0320	***	-		5.6990	***	-		2.2910	**
n	750		494		2803		2186		687		843	
R-sq(adj)	0.8990		0.9630		0.9560		0.9790		0.9760		0.9620	
GCV	0.0163		0.0282		0.0108		0.0119		0.0075		0.0197	

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Table 8-1 - Model Output Summary

Parametric terms are constructed such that reported results in figure 8.1 is benchmarked to the first alphabetic variable, which is normalized to 0. This is to avoid the dummy variable trap[13]. The parametric model outputs should henceforth be compared to the one not listed in the table, but stated in table 6.2. E.g. all BUILD variables are benchmarked to the performance of China P.R.

The explanatory power of all models is remarkably high, indicating they are able to predict the rate with a high degree of accuracy. Although the index itself, fairly regardless of choice within this market, would probably represent a significantly high share, we believe our choice of index proxy contributes to further increase the overall power of the model. Referring to our discussion in section one, this is in line with our initial objective. Since our models are comparable to Koehn (2008) we find an overall higher explanatory power for all models, which indicate that our choice of index is more accurate.

Capesize vessels built in Japan seem to receive a 7% higher rate compared to China in the second period, while the other countries are insignificant. For the Panamax segment South Korea obtained the best rate in the first period, although they were lower than China. In the second period the relationship changes and rates for vessels built in Japan are 2% higher than China compared to negative 1.7% in the first period. South Korea and other countries appear to have no significant impact on rates. Handymax vessels built in South Korea received a higher rate in both periods and the change is fairly minor for both Japan and other countries. Japan looks to have no significant impact. According market participants, such as DNV-GL and Wikborg Rein, Japanese shipyards are known for quality and timely delivery. This could support our findings and suggest that quality has become more pronounced in recent years (Willumsen, 2014).

Regarding flag state the most interesting result is that during the first period several countries received a lower rate compared to Bahamas. China P.R is the most noticeable significant variable with a discount of 17%, although a fairly small amount of the observations has a Chinese flag. Panama, Malta, Hong Kong and Liberia have a

[13] Perfect multicollinearity between the dummy variables and the intercept. (Brooks, 2008)

relatively large amount of total observations and the discount is quite similar in the first period. The second period for all segments reveals no clear significant results. One possible hypothesis could be that Chinese built ships with Chinese flag appears more risky due to political and regulatory insecurities. Data from Clarksons World Fleet Monitor shows that there is an almost perfect correlation between Chinese flag and ownership.

Out of the parametric variables reported, place of delivery has the most interesting result. POD Pacific is highly significant, except Handymax in the first period. Rates are consistently lower for Pacific compared to Atlantic and there is a substantial increase in discount in the last period. Since we have emphasized on Pacific and Atlantic delivery we place no further comment on unknown delivery, although we note that there seems to be an increase in discount as well for the last period. This result is in line with findings in Koehn & Thanopoulou (2010). Apparently it seems that Atlantic delivery is more preferable where the assumption is that less tonnage capacity in Atlantic would put an upward pressure on rates. Furthermore a theory could be that as dry bulk trade follows a pattern from west to east, delivery in the Pacific seems less attractive as it would require a ballast leg. Figure 8-1 presents a graphical illustration of the POD results.

Vessel gear is only relevant for Handymax, however table 6-2 shows that the distribution within this segment is fairly undistributed. It would therefore not be relevant to include this parameter. Model testing with and without the parameter also shows that there is no change in results or model fit when excluded.

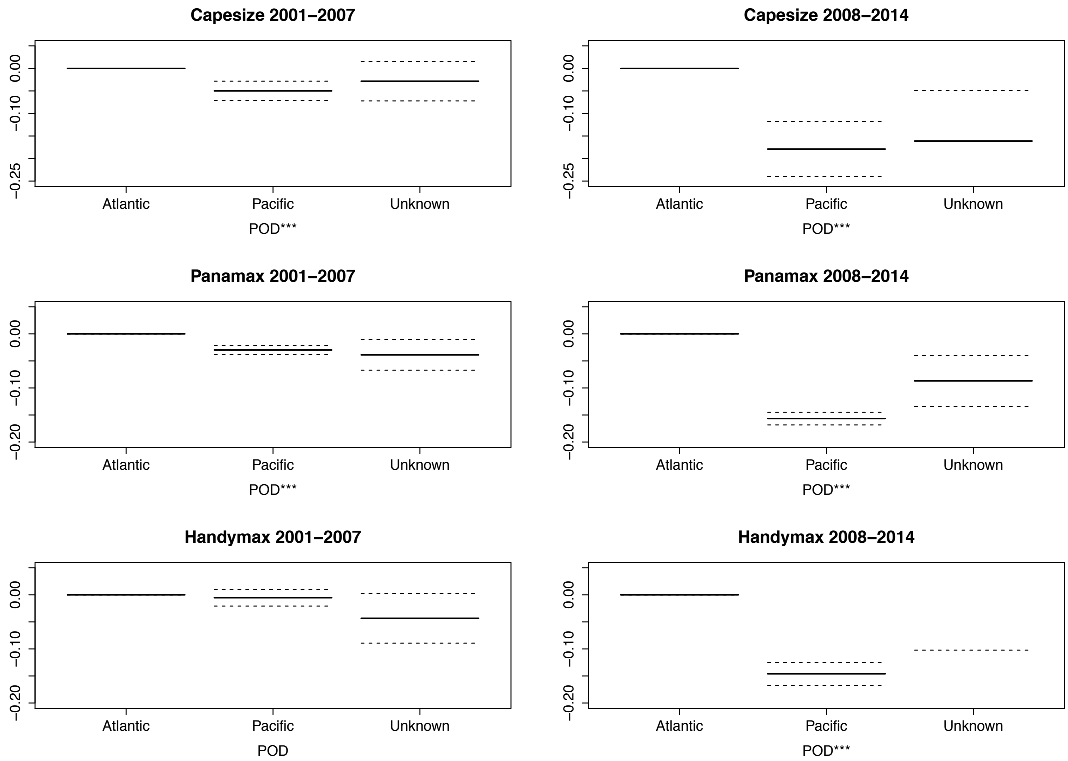


Figure 8-1 - Plot of Place of Delivery

8.1.1. *AGE*

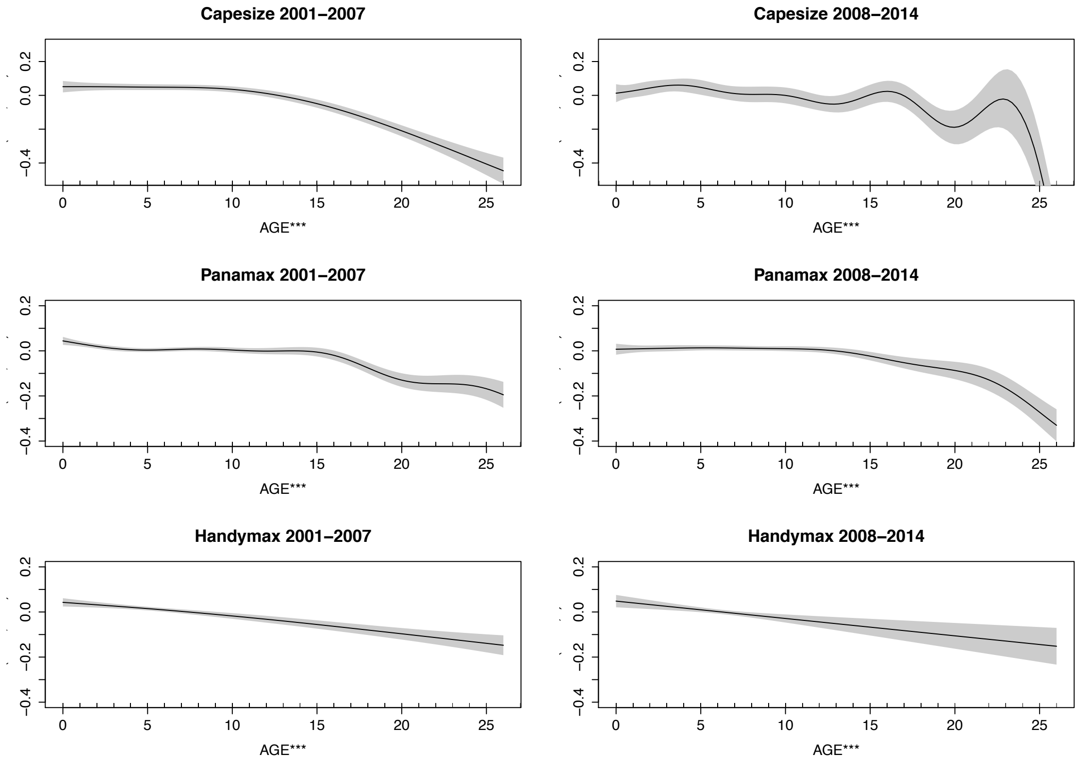


Figure 8-2 - Plot Age

As expected, age has a significant negative impact on charter rates across all segments. Previous studies have found evidence for a two-tier market, which is further supported by our results. Compared to Koehn (2008) we find a different dynamic for Panamax in the first period. Apparently the market shows minor differentiation between 20-25 years, compared to a strong differentiation the last period. An explanation could be a shift in the quality of vessels built in the early 1980s compared to 1990s and beyond. We see that the market places a clear discount for Capesize vessels older than 10 years in the first period and there are no structural shifts. With the assumption that new vessels are preferred over older ones, our assessment would be that pre 2008 the market situation dictated that high demand for vessels resulted in high degree of fleet utilization. This would eventually lead to a smooth downward trending rate as age increases. On the contrary this trend does not continue post 2008 as the market was in a trough, yielding high overcapacity. Rates altered from being driven by supply to demand and older vessels was either put in lay up or demolished. Available contracts were assigned to relatively newer vessels at low rates.

Possible due to the same reasons, increasing confidence band from 17 years for Capesize in the upper right panel indicates fewer observations and higher uncertainty. There seems to be no obvious explanation for the sudden increase in rates from between year 20-23, hence we assume it to be caused by data driven inference. For Handymax vessels the relationship is fairly linear, indicated by ~1 effective degrees of freedom, with a negative trend. This could suggest a possible n-tier market. However, the uncertainty increases with age, especially in the last period.

8.1.2. DWT

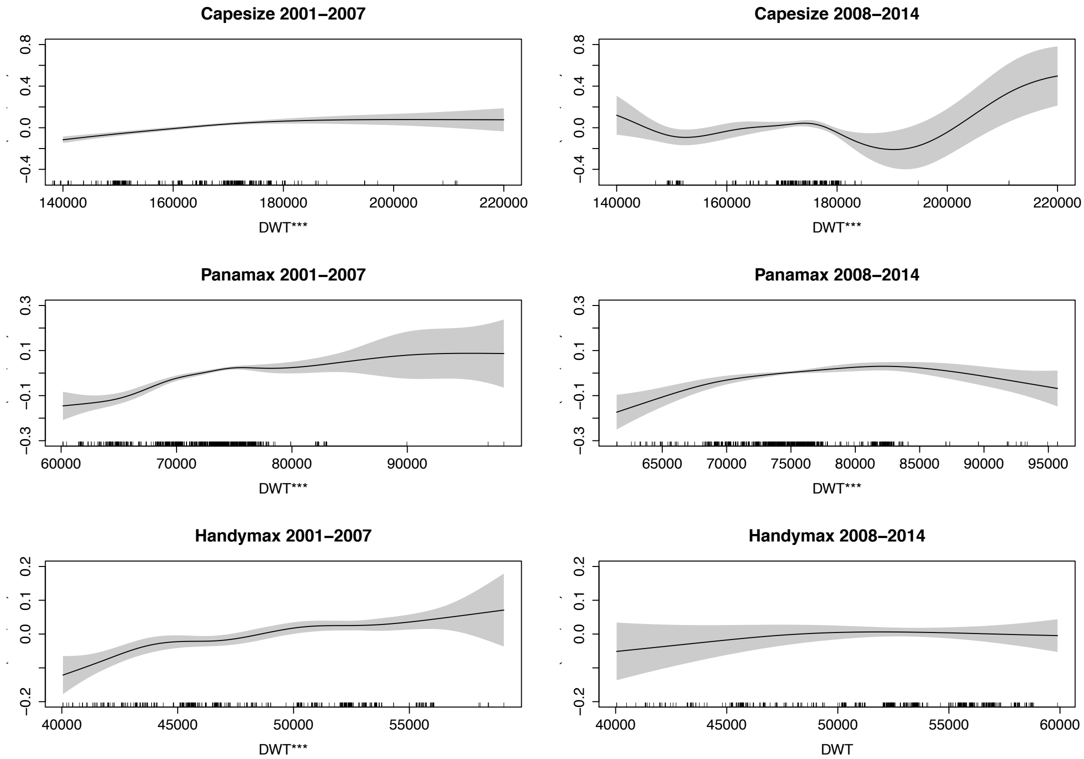


Figure 8-3 - Plot Size (DWT)

It seems apparent that, across all segments, rate increases with size in the first period and the relationships are non-linear. For Panamax the confidence interval beyond 85000 reveals some uncertainty. This relationship changes in the last period. For mid size vessels rate increases with size. Some large Capesize vessels seem to have obtained large rates, though the number of observations is limited. There is an increasing discount for Panamax vessels above 85000 dwt. Observations in our dataset reveals that the beam of these vessels is larger than the Panama Canals limitations or other limitations on logistical infrastructure. One possible explanation could be that the reduction in rate is due to reduced flexibility compared to their smaller peers. Charterers who do not need the flexibility to pass through the canal might as well choose a small Capesize. Lower rates for the largest vessels are obviously demand driven and one explanation to why the shipowner accepts lower rates, compared to smaller vessels, is the relatively higher lay-up costs.

8.1.3. FEI

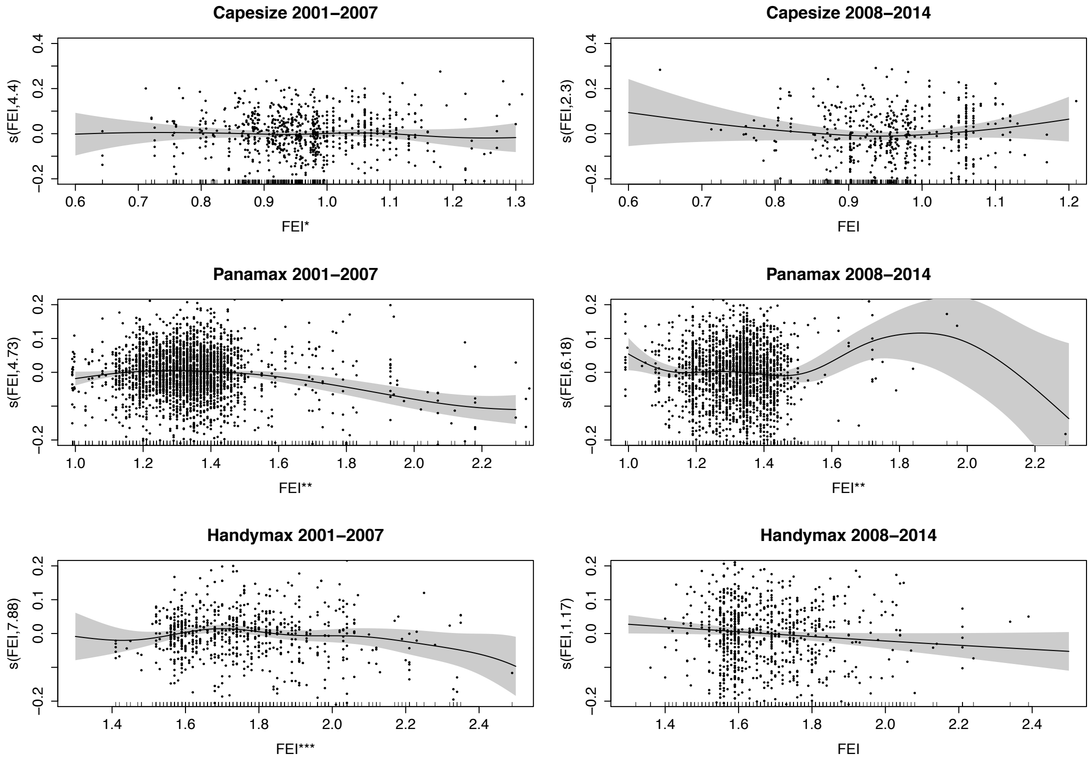


Figure 8-4 - Plot Fuel Efficiency (FEI)

Residuals are included for the FEI variable. Contrary to the interpretation of other results we draw our attention towards the distribution of observations. Although the significance is fairly low or non-existing for most of the variables, we discover some interesting results. The Capesize segment shows limited or no change in observations, which also is illustrated by significance levels. However, for Panamax vessels there seems to be a clear shift towards lower consumption per ton-mile in the last period. The distribution is altered further to the left on the x-axis, suggesting a more efficient fleet. These results are also supported by the data description where average FEI is reduced by 2% between periods. Also, there is some evidence suggesting that increased efficiency yields higher rates, seen by the downward trending graph. Due to increased confidence bands for the last period we find it difficult to make any reasonable interpretations. We find the same results for Handymax vessels. The variable has a significant effect on rates and there is a downward sloping trend. Average FEI is reduced by 4% between periods, however results for the last period is not significant. Following our argument for Panamax vessels, there could be evidence of efficiency premiums in the Handymax market as well. Although it might be reasonable to assume that these effects could be explained by reduced design speed between time periods along with increased consumption and size. We find that speed has increased slightly over time, which could indicate the economies of scale of larger vessels. Drawing our attention towards the rate influence of FEI we see that the graph is fairly flattened around observations of high density. Considering our discussion in section 4.3.2 this could indicate that rate levels lie along the dotted line, where the charterer obtains any existing efficiency premiums. Following the arguments we should be able to detect a shift from the time charter market to the spot market. However, no clear indication of market shifts is identified (see appendix 3).

8.1.4. NDF

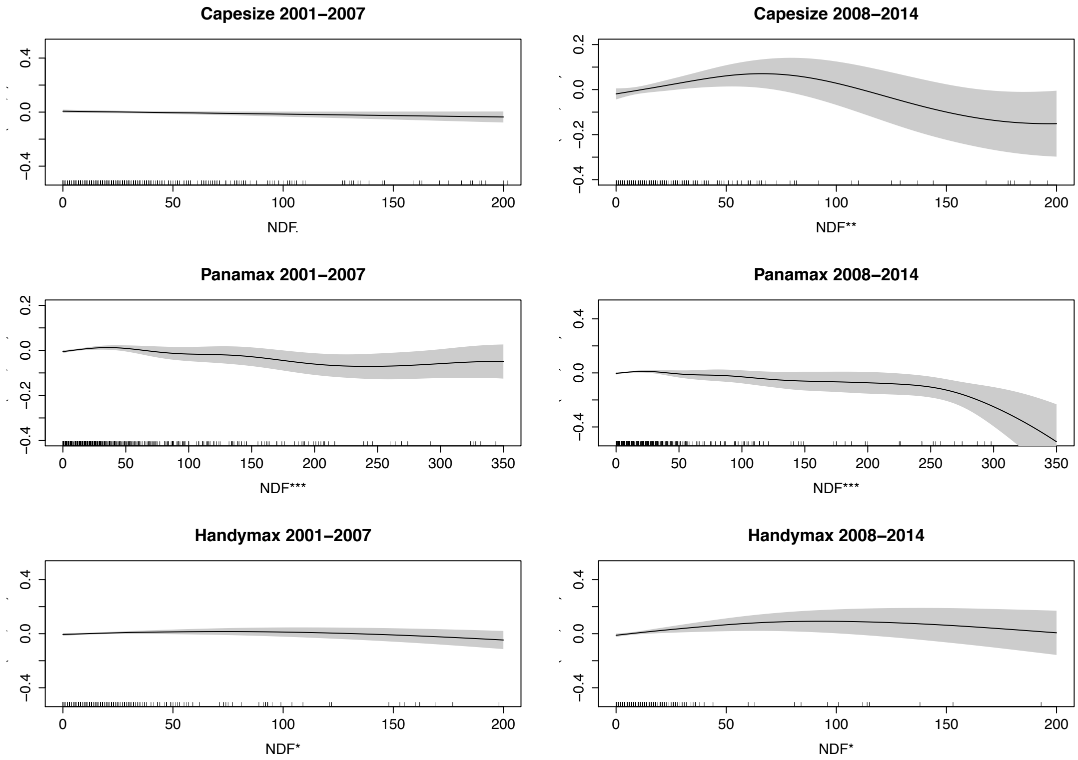


Figure 8-5 - Plot Number of Days Forward (NDF)

There seems to be no significant change in rates for contracts between 0 and 50 days forward in the first period. Beyond 50 days rates receive a small discount, however the results are relatively more non-linear for Panamax. In the second period we see an opposite effect for Capesize and Handymax between 0 and roughly 75 days, where the rates have a significant increase. Beyond roughly 30 days Panamax contracts receive a slight increasing discount until 260 days. The most distant contracts are heavily discounted. I line with financial theory we see a trend where contracts entered more than 2-3 months in advance are discounted. Also, there has been a reduction in contracts entered more than 50 days in advance after 2008.

8.1.5. CONTL

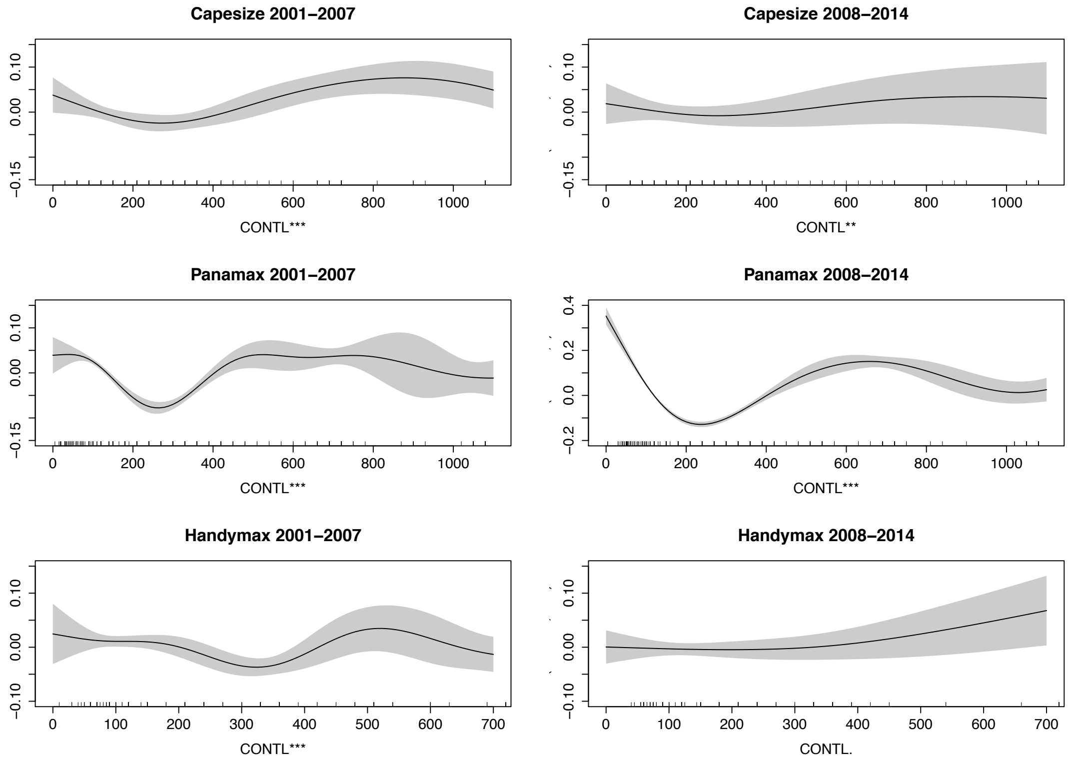


Figure 8-6 - Plot Contract Length (CONTL)

Compared to findings in Koehn (2008) we find the same clusters around 3, 12 and 24 months in the first period, although the dynamics are different. All significant variables across segments and time periods have a negative trend for contracts up to approximately 6 months, which seems to be a global trough for rates after 2008, except Capesize and Handy in the first period. Rates for Capesize vessels appear to be less influenced by the length of the contract. Since these vessels tend to be hired on longer contracts it could explain this result. Panamax vessels seem to be more volatile to longer contracts in the latest period. The rate increases up to roughly 24 months and decreases beyond this point. Compared to the first period this shifting trend is not as evident. It appears that the market is rather indifferent between approximately 15 and 27 months. In the Handymax segment there seems to be a preference for shorter contracts along with contracts between 15 and 24 months, though the reliability of the estimates decreases. In general it appears that a premium is attached to contracts deviating from the mean.

8.1.6. OPTL

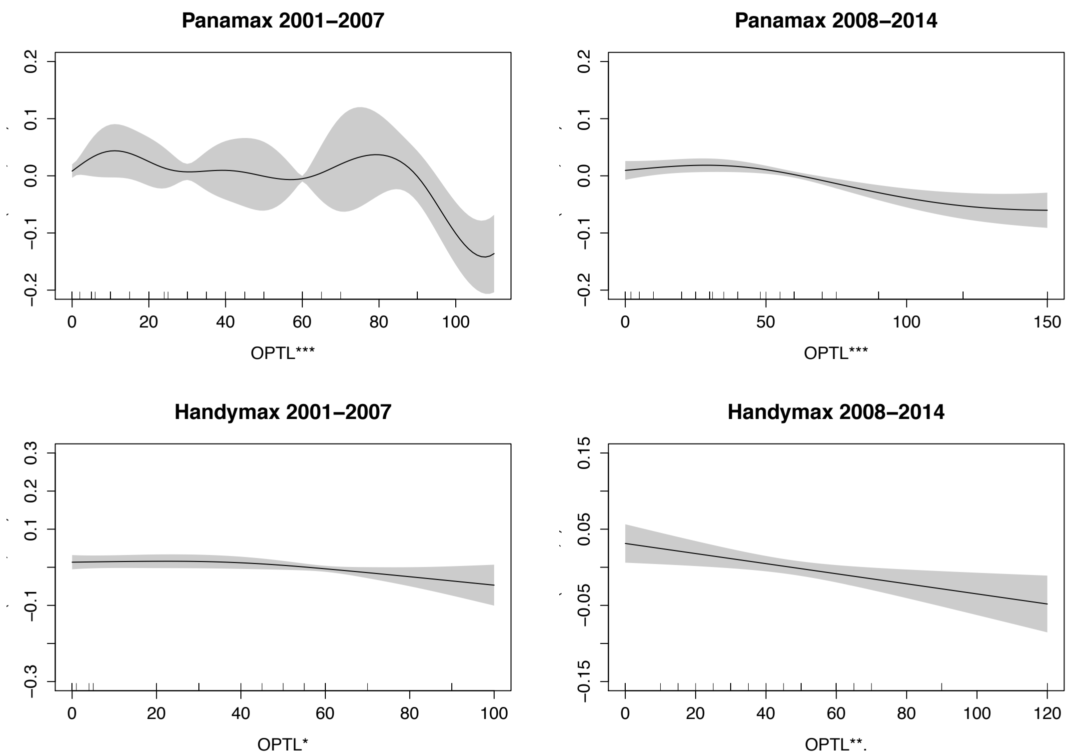


Figure 8-7 - Plot Option Length (OPTL)

A general observation is that longer options results in a rate discount. This effect is however more consistent in the last period. For both Panamax and Handymax there is a cluster around 30 and 60 days in the first period. Data density is low for other periods. Option length on contracts in the last period seems to be more linear related with a negative impact. Due to the fact that there is not enough observation different from zero in the Capesize segment, option length has been excluded from the model.

8.1.7. INDEX

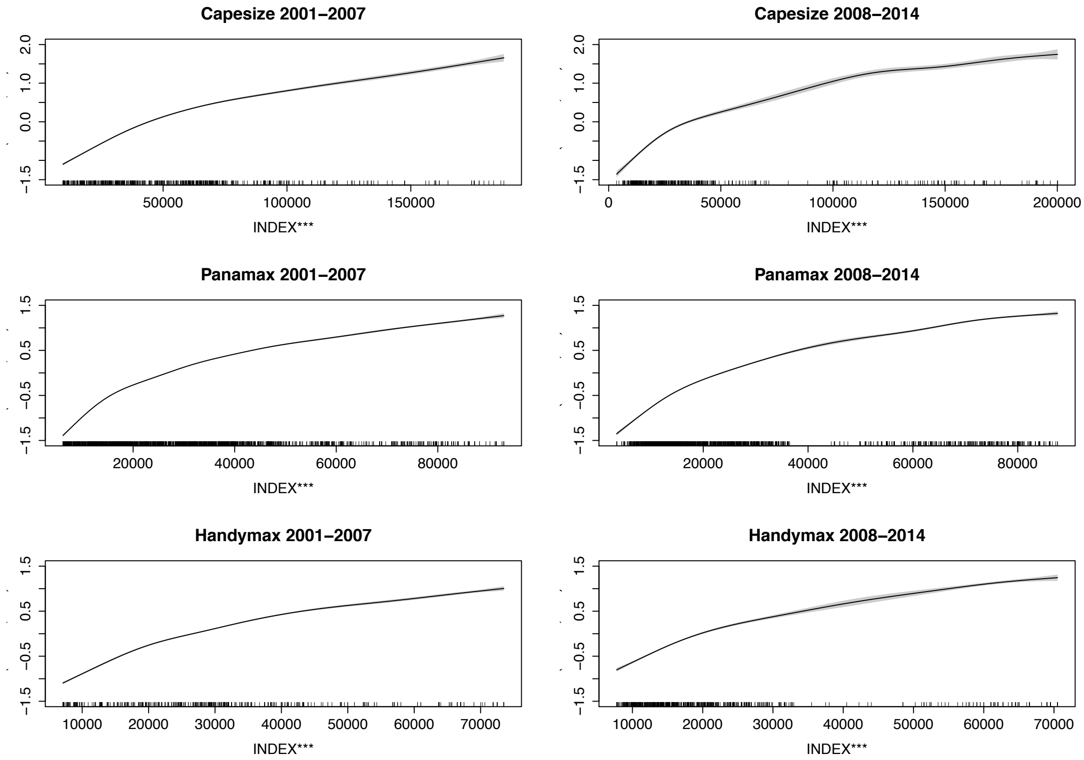


Figure 8-8 - Plot Index

Observations from the correlation matrix suggest an almost perfect positive correlation between index and rate. However, plots in figure 8-8 reveal a technical non-linear relationship with parsimonious confidence bands. Taking the logs of index data would normalize the data around the mean[14] and result in a more linear relationship. Variable transformation due to possible heteroskedasticity was considered and we refer to section 9.1 for further discussions.

[14] ”Tame” outliers or deal with skewness

8.1.8. BUNKERS

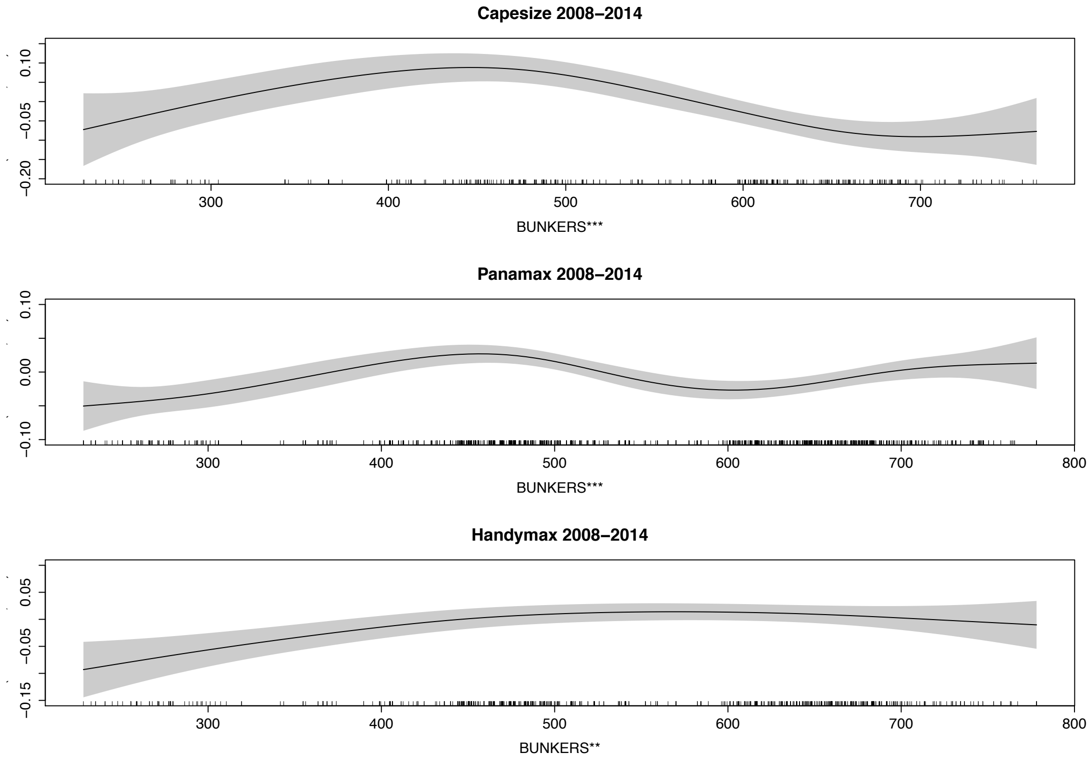


Figure 8-9 - Plot Bunkers

During the first period there was a substantial demand pressure in the market where we assume that the bunkers price had a relatively small impact on the rate, as it represented only a marginal part. However, in the last period when the market was tougher there is a nonlinear relationship. At a bunkers price of roughly below $500, Capesize and Panamax rates receive a discount. For Handymax vessels this trend is evident between $500-600. After a dramatic increase from 2010 the price of bunkers has been relatively stable from mid 2011 while rates has slightly recovered. This could indicate a market shift where the price of bunkers became relatively less important, explaining the inverse relationship between Capesize and Panamax at high bunkers levels.

9. MODEL APPLICATION

As an extension to our research we create two scenarios where we use our models to predict rates for two individual Panamax vessels measured over time. One is specified as a standard ship while the other is a hypothetical ECO-ship. In scenario one we compare two vessels with different age, size and consumption. Scenario two only considers different consumption, ruling out the effects of size and age. Input parameters for both scenarios are shown in table 9-1. Please note that for each scenario there are two models shown in one continuous graph, clearly visible before and after 2008.

Scenario 1		
Input parameters	**Standard**	**ECO**
NDF	16	16
Age	**10**	**4**
DWT	**74000**	**80000**
***FEI**	**1.3272**	**1.0435**
CONTL	360	360
OPTL	50	50
BUILD	Japan	Japan
FUEL	HFO	HFO
FLAG	Panama	Panama
ENGT	MAN B. & W.	MAN B. & W.
POD	Atlantic	Atlantic

Scenario 2		
Input parameters	**Standard**	**ECO**
NDF	16	16
Age	**7**	**7**
DWT	**75000**	**75000**
***FEI**	**1.3095**	**1.0747**
CONTL	360	360
OPTL	50	50
BUILD	Japan	Japan
FUEL	HFO	HFO
FLAG	Panama	Panama
ENGT	MAN B. & W.	MAN B. & W.
POD	Atlantic	Atlantic

Table 9-1 - Vessel Specifications

The prediction technique takes a fitted GAM object produced from our model in section 9 and produces predictions given a new set of values for the model covariates. Hence, we keep the original coefficients and apply new input data. This way we can calculate a rate for any vessel, given its specifications, using the relationships found in the original models. Figure 9-1 shows the difference between estimated rates for our ECO-ship and standard ship for both scenarios. Pre 2008 the ECO-vessel in scenario 1 receives a constantly lower rate than the standard vessel, although the difference is relatively low. In scenario two the ECO-vessel obtains a moderately increasing rate premium. The fact that the ECO-vessel is both larger and younger in scenario 1 does not seem to make a difference in the market. From 2008 and beyond the results become more interesting. First of all rate and size seems to make a difference as scenario one is constantly above the second, illustrating the results for age and size found in section 8. Note that from the results in section 8 the effect from size is probably most prominent as the age coefficients are fairly different between 4 and 10 years. In addition our ECO-vessel in scenario 2 has a lower consumption and hence a lower FEI. Our prediction indicates that an ECO-vessel should be able to obtain a higher rate, given the current market dynamics in the last period. Apparently the market dynamics changed before and after 2008. One remark regarding this simulation is that as smoothed terms result in point estimates opposed to a linear regression. Hence, it only considers threshold values given out input variables. Choosing a greater difference for age and size between the vessels would change the relationship depending on the size of the estimated coefficient at that particular point. The question remains whether the positive difference in the last period justifies investments in more capital-intensive new buildings. Retrofitting is less costly, it is however not considered any further. From our point if view, an average difference of approximately \$1000/day over a lifetime of 20-25 years might not be sufficient. The same technique can also be applied to model effects of other variables as well, such as the presence of options and its effect on rates.

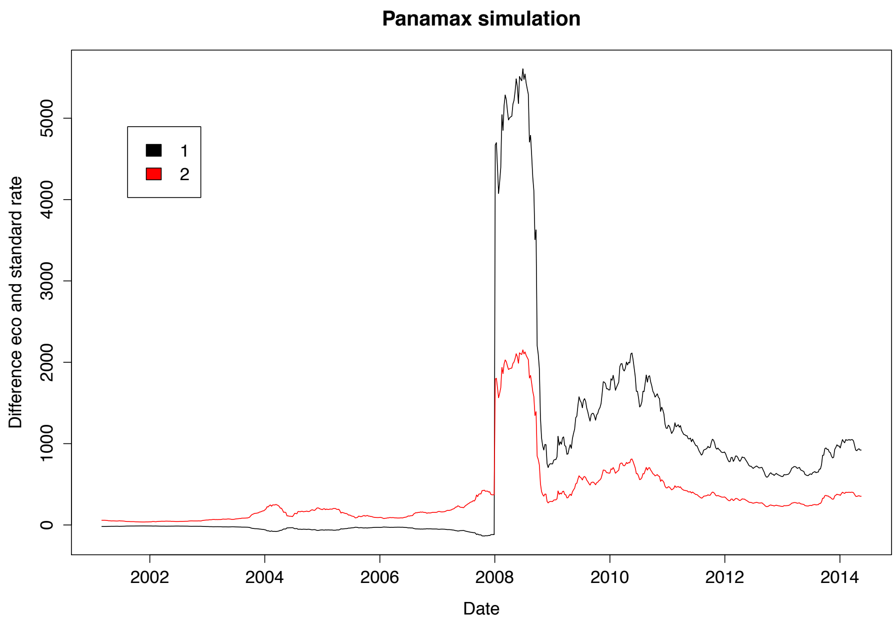


Figure 9-1 - Panamax Simulation

10. MODEL UNCERTAINTY

10.1. GENERAL

Conventional tests for heteroscedasticity are not applicable for GAM models and according to Wood (2006), the most intuitive way for detecting issues with non-constant variance around the error term is to examine data plots of variables. Common practice to deal with heteroscedasticity is to manually transform variables assumed to inherit such characteristics to make them more normally distributed. It would also be possible to apply models such as General Additive Mixed Models or Generalized Additive Models for Location, Scale and Shape to investigate random and fixed effects, distributions along with structural behavior of variables. This is, however, considered to be beyond our scope.

The relationship between rate and index seen in figure 10.1 is clearly linear and there seems to be no consistent indication of a significant increase in variance around the mean. A log transformation of the index variable was tested in the model, without yielding any obvious differences with regard to standard errors. Hence, we assume no suspect inferences with regard to rate and index.

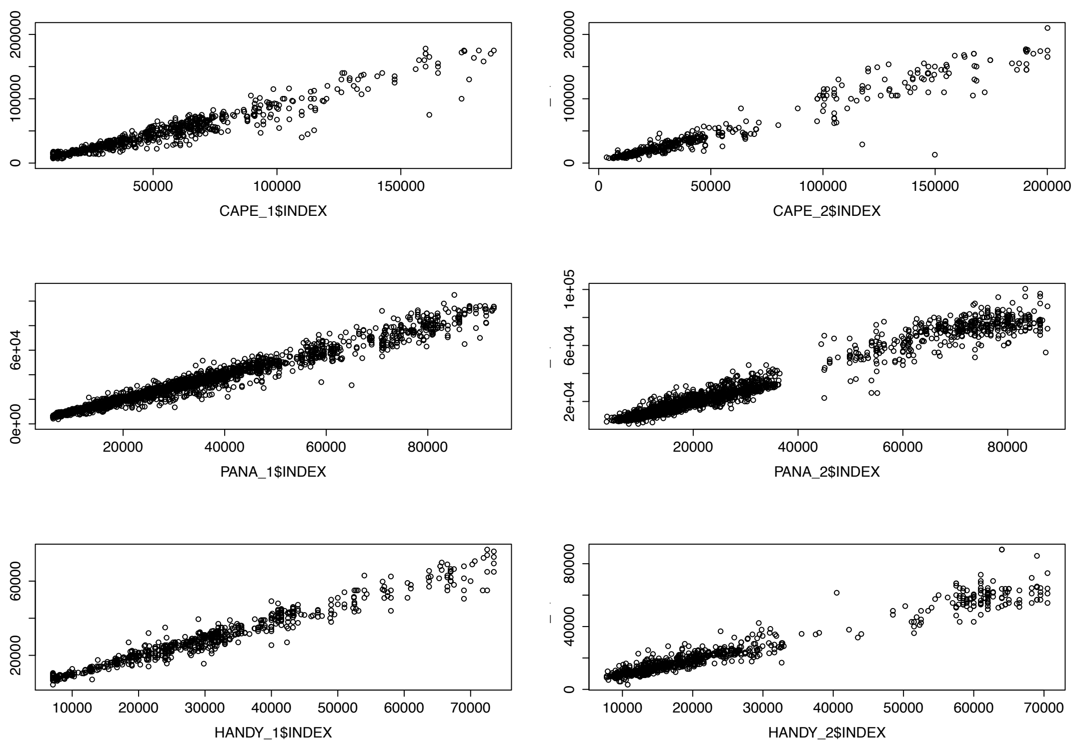


Figure 10-1 - Plot Index/Rate

10.2. KNOT LOCATIONS

As mentioned there are other splines that could have been applied, given that knot locations was chosen. Manually deciding number of knots to variables such as index could be applied, however we found this approach to be more uncertain and computational costly due to the amount of model coefficients, in addition to be somewhat out of our scope.

10.3. SELECTION BIAS

One concern regarding the intentions of our study is the possibility of a skewed distribution between vessels operating within the spot and time charter market. The idea is that if efficient vessels were consistently operating in the spot market, which would be a reasonable assumption if the time charter market does not compensate efficiency, the results from our model would be inaccurate. This is referred to as selection bias. Possible selection bias is investigated by plotting the FEI for our dataset against the total fleet from Clarksons World Fleet Register. Vessels in our dataset are removed from the total fleet such that vessels remaining in the fleet have never been in the time charter market since 2001. From the graphs in section 13.3 the distributions seems to be fairly equal and no obvious bias is assumed.

10.4. DATA UNCERTAINTY

Since we use reported fixture data we are aware that design parameters such as speed and consumption, might be different from actual operational data. However, as pointed out earlier real world data is hard to obtain and measure. Misreported values are obviously a concern since they stem from an external source, though we see no realistic alternative other than being aware of the issue. Another concern is that there probably exists a large amount of contracts not reported in the data from Clarksons, which could indicate results that are not representative. Yet, we see this as a data limitation problem we cannot influence.

11. CONCLUDING REMARKS

In this thesis our aim has been to test whether quality premiums are evident in the dry bulk market. By identifying relevant determinants we have decomposed the rate according to market, vessel and contractual specific variables through a semi-parametric generalized additive model. Some of the questions asked in this thesis are studied in previous literature, which gave us some prior expectations of the relationship between rate and its determinants. Hence, we further support previous results along with identifying other relevant variables that could be related to efficiency. In addition to what has been done in existing literature we employ a much larger dataset across the three main dry bulk segments. Furthermore we have applied a more comprehensive composition of the index variable, which accounts for a great part of rate determination, along with a combination of speed, consumption and size in an attempt to incorporate the impact of those variables in a more realistic manner.

Beginning with market based variables we find a clear non-linear relationship between time charter rates and the index. By interpolating the index variable and excluding time effects we found a significant increase of the explanatory power in the model. By eliminating time-driven effects of the underlying index specifications, we assume to have increased possible significance of other variables expected to be of relevance. Based on our knowledge this would not have been accounted for in previous research, resulting in possible model interference. Supporting previous literature we find a rate discount for vessels delivered in the Pacific. In addition we discover that the rate discount has increased over the two time periods considered. Generally we see one possible explanation: Along with increased production activity in the Far East, trade flow from west to east increased considerably leading to greater supply of tonnage in the Pacific basin. Another explaining factor is the need for backhaul sailing.

Through analysis of vessel specific variables related to efficiency we see a change in the impact of flag state. A shift can be found between time periods where as an example China has gone from being discounted in rates to insignificant with regard to flag in the second period. One possible rationalization could be that regulation differences between flag states has been reduced over time. Furthermore Japan has apparently gained increased recognition over time as a quality provider as it yields a substantial discount compared to China and South Korea. Age has a non-linear negative effect on rates across segments and time periods, showing possible existence of a two-tier market or even an n-tier market for Panamax and Handysize vessels. Following results from Koehn (2008) this relationship seems to have been consistent over time. Also, we find evidence for increasing rate with size, confirming economies of scale. Between time periods there has been a change for Capesize and Panamax where the largest Panamax vessels receive a discount while there is a premium for Capesize. A possible reason could be that Panamax vessels in the largest end of the scale are unable to pass through the Panama Canal, leading to reduced flexibility, and that large Capesizes are built for special purposes or through non-standard contracts. Results in both ends of the size range are clouded due to low-density observations. According to consumption per ton mile there has been a shift towards more efficient vessels in the last period. However, one drawback is that a decomposition of the influence of each variable within the FEI is limited. Within high density observation ranges the relationships seems to be leveled, indicating that the Charterer obtains any vessel specific efficiency gains. This contradicts the assumptions of efficient markets we discussed in section 4 and results in shipowners not being sufficiently compensated for efficiency investments, confirming the notion of split incentive barriers.

Most of the limitations within out thesis lie in the chosen methodology. Our model is able to capture non-linear dynamics and hence serves its purpose well. However, further research on the topic could take advantage of more sophisticated models such as Generalized Additive Mixed Models. In addition one could deviate from default choices in the model, such as splines and knot selections to explore relationships even further. Further a more detailed study of each segment separately could result in more thorough insight on how ship specific variables are related.

12. REFERENCES

1990, Environmental Protection Agency. (Oil Pollution Act) [pdf] Washington D.C: United Printing Office. Available at http://Http://www.epw.senate.gov/opa90.pdf [Accessed 12 March 2014].

Adland, R. (2013a). Can Shipowners Economise on Fuel Consumption. [Online] Norwegian School of Economics Available at: https%3A%2F%2Fwww.nhh.no%2F...%2FDWSDownload.aspx%3FFile&h=wAQEMGCj [Accessed 20 April 2014].

Adland, R. (2013b) Freight Market Economics, ENE430 - Commodity Trading and Transport. [Print] Norwegian School of Economics, Unpublished

Adland, R., & Cullinane, K. (2006) 'The non-linear dynamics of spot freight rates in tanker markets'. Transportation Research 42, (3) 211-224.

Adland, R., Koekebakker, S., (2007). Ship valuation using cross-sectional sales data: a multivariate non-parametric approach. Maritime Economics and Logistics 9 (2), 105–118.

Alizadeh, A & Nomikos K.N. (2009) Shipping derivatives and risk management. 1st ed. New York: Palgrave Macmillan.

Alizadeh, A., & Talley, W. (2011) 'The non-linear dynamics of spot freight rates in tanker markets'. Transportation, Economics and Policy 38, (3).

Andersen, R. (2007) Nonlinearity and Generalized Additive Models, [Print] McMaster University

Beck, N., & Jackman, S. (1998) 'Beyond Linearity by Default: Generalized Additive Models'. *American Journal of Political Science* 42, (2) 596-627.

Beenstock, M & Vergottis, A. (1993) Econometric modelling of world shipping. 1st ed. London: Chapman & Hall.

Brooks, C (2008) Introductory Econometrics for Finance. 2nd ed. London: Cambridge University Press.

Canal de Panama (2013) This is the Canal [Online] Available at http://Http://www.acp.gob.pa/eng/acp/asi-es-el-canal.html [Accessed 12 April 2014]

Clark, M. (2013) Generalized Additive Models. Chicago: University of Notre Dame. Available at: http://www3.nd.edu/~mclark19/learn/gams.pdf. [Accessed 12 April 2014].

Clarkson Research Services, (2014) Shipping Market Overview, Presentation to DNV GL. Oslo, Unpublished

Clarksons Research, 4. ed. (2014) 'World Fleet Monitor'.

Dick, van D., Haralambides, H.E., Veenstra, A.W., 1998. Modelling of Ocean Charter rates; an investigation into non-linearity and parameter. In: 8th World Conference on Transport Research (WCTR), Antwerp, Belgium, July 1998.

Glen, D.R. (2006) 'The modelling of dry bulk and tanker markets: a survey'. Maritime Policy and Management 33, (5) 431-445.

Hastie, T & Tibshirani, R.J,(1999) Generalized additive models. Boca Raton, Fla: Chapman & Hall/CRC.

Hawdon, D. (1978) 'Tanker freight rates in the short and long run'. Applied Economics 10, 203-2017.

Kavussanos, M.G., Alizadeh, A., (2002). The expectation hypothesis of the term structure and premiums in dry bulk shipping freight markets. Journal of Transport Economics and Policy 36 (2), 267–304.

Koehn, S., 2008. Generalized Additive Models in the Context of Shipping Economics, PhD thesis, University of Leicester.

Kohn, S. and Thanopoulou, H.A. (2011). A GAM assessment of quality premia in the dry bulk time-charter market. Transportation Research Part E: Logistics and Transportation Review 47, (5) 709-721.

MAN Diesel & Turbo SE. 2012. G-Type Order Book Reflects Growing Interest. [Press release]. [Accessed 12 April 2014]. Available from: http://www.mandieselturbo.com/files/news/filesof17998/6510-0299%20PR%20G70.pdf

McCarthy, L. (2013) '*CMA 2013: Are eco-ships necessary when slow steaming still rules the waves?*'. Loyds List [Online] 20 March. Available at http://www.lloydslist.com/ll/sector/ship-operations/article419237.ece [Accessed 12 March 2014].

Randers, J., & Göluke, U. (2007) 'Forecasting turning points in shipping freight rates: lessons from 30 years of practical effort'. Journal of Transport Economics and Policy 36, (2) 267-304.

Rehmatulla, N., Smith, T., & Wrobel, P. (2013) 'Implementation Barriers to Low Carbon Shipping'. *Low Carbon Shipping 2013*, (London). Available at http://www.lowcarbonshipping.co.uk/files/ucl_admin/LCS%202013/Rehmatulla_et_al.pdf [Accessed 21 March 2014]

Rowlinson, M. (1996) 'Comment: An analysis of the quality and redistribution of dry capesize tonnage'. Maritime Policy and Management 23, 407-409.

Smith, T., O'Keeffe, E., Aldous , L., & Agnolucci, P. (2013) 'Assessment of Shipping's Efficiency Using Satellite AIS data'. LowCarbonShipping.com.

Stopford, M (2009) Maritime economics. 3rd ed. Abingdon, Oxon: Routledge.

Strandenes, S.P. (1984), Price determination in the time charter and second hand markets. Discussion paper 0584, Norwegian School of Economics and Business Administration, Bergen, Norway.

Strandenes, S.P. (1994). Is there a Potential for a two-tier Tanker Market?. Discussion Paper, Economics series 37, Wirtschaftswissenschaftliche Facultat, Humboldt Universitat zu Berlin.

Tamvakis, M.N., Thanopoulou, H.A., 2000. Does quality pay? The case of the dry bulk market. Transportation Research: Part E 36 (4), 297–307.

Thanopoulou, H.A., & Gardner, G.M. (2012) 'Defining Quality Bulk Tonnage: A Task for Researchers and Policy-Makers'. Journal of Economics and Business 62, (2-3).

The Motorship, (2012). Continuous slow steaming with two-stroke engines . [ONLINE] Available at: http://www.motorship.com/features101/fuels-and-oils/continuous-slow-steaming-with-two-stroke-engines. [Accessed 12 March 2014].

Timmermann, K., & McConville, J. (1996) 'An analysis of the quality and resdistri-bution of dry capesize tonnage'. Maritime Policy and Management 23, 45-53.

Willumsen, T. (2014) 'Why the sun has yet to set on Japanese shipbuilding'. Seatrade Global [Online] 27 February. Available at http://Http://www.seatrade-global.com/news/asia/why-the-sun-has-yet-to-set-on-japanese-shipbuilding.html [Accessed 15 March 2014].

Wood Simon, 2006. Generalized Additive Models: An Introduction with R (Chapman & Hall/CRC Texts in Statistical Science). 1 Edition. Chapman and Hall/CRC.

13. APPENDICES

APPENDIX 1 - CORRELATION

Capesize 2	2008-2014								
	RATE	NDF	AGE	DWT	FEI	CONTL	OPTL	BUNKERS	INDEX
RATE	1								
NDF	0.1134	1							
AGE	0.0318	-0.0798	1						
DWT	0.0124	0.0156	-0.5939	1					
FEI	-0.0986	0.0561	-0.2361	-0.0482	1				
CONTL	0.0894	0.4709	-0.0630	0.0132	0.0639	1			
OPTL	-0.2194	-0.1478	-0.0565	0.0682	-0.0040	-0.2170	1		
BUNKERS	-0.0135	-0.0873	-0.1889	0.2188	0.0037	-0.0171	0.1795	1	
INDEX	0.9680	0.1356	0.0995	-0.0750	-0.0948	0.1162	-0.2407	-0.0142	1

Capesize 1	2001-2007								
	RATE	NDF	AGE	DWT	FEI	CONTL	OPTL	BUNKERS	INDEX
RATE	1								
NDF	0.0263	1							
AGE	-0.1649	-0.0518	1						
DWT	0.2357	0.0543	-0.5649	1					
FEI	-0.0381	-0.0799	0.1564	-0.3441	1				
CONTL	-0.0783	0.3816	-0.1089	0.1426	-0.0300	1			
OPTL	-0.0341	-0.2415	0.1118	-0.1191	0.0877	-0.3092	1		
BUNKERS	0.6613	0.1023	-0.0525	0.1883	-0.0068	0.1207	-0.0753	1	
INDEX	0.9129	0.0802	-0.0057	0.1157	0.0186	-0.0619	-0.0395	0.7318	1

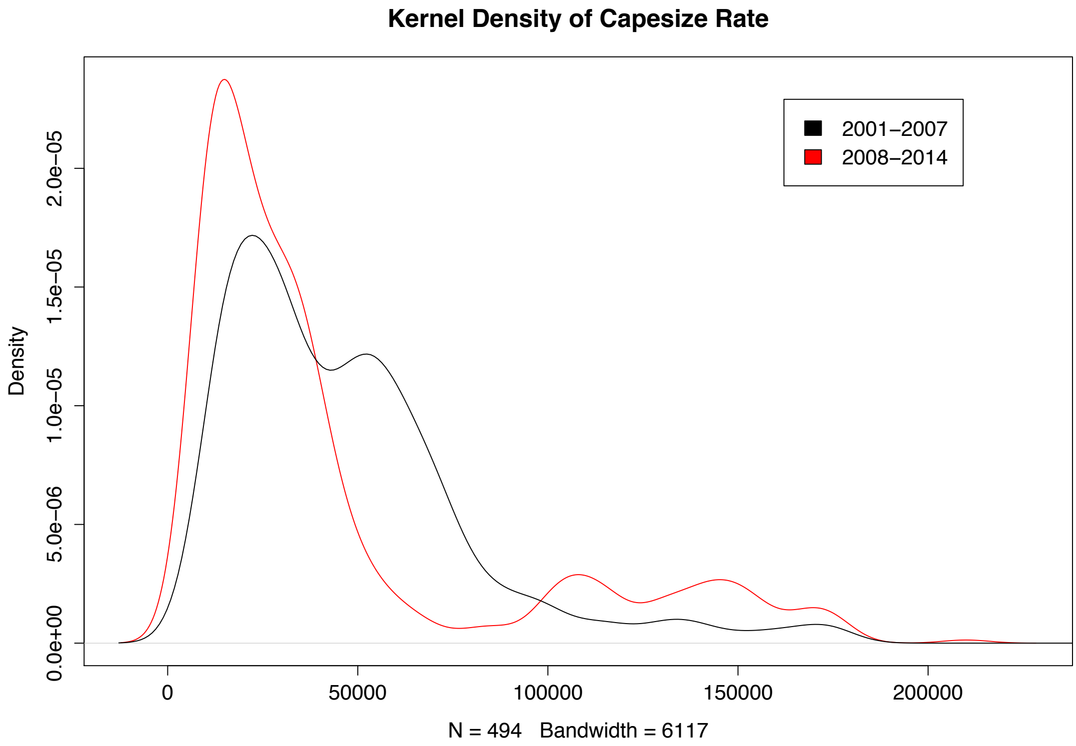
Panamax 2	2008-2014								
	RATE	NDF	AGE	DWT	FEI	CONTL	OPTL	BUNKERS	INDEX
RATE	1								
NDF	0.0757	1							
AGE	-0.0279	-0.0404	1						
DWT	-0.1464	-0.0224	-0.6813	1					
FEI	-0.0254	0.0054	0.0554	-0.0472	1				
CONTL	0.0232	0.5235	-0.0783	0.0537	0.0512	1			
OPTL	-0.2430	0.0061	-0.0253	0.1030	0.0490	0.0866	1		
BUNKERS	0.0695	0.0053	-0.0268	0.1505	-0.0281	0.0243	0.1374	1	
INDEX	0.9810	0.1078	0.0131	-0.1777	-0.0096	0.0445	-0.2328	0.0694	1

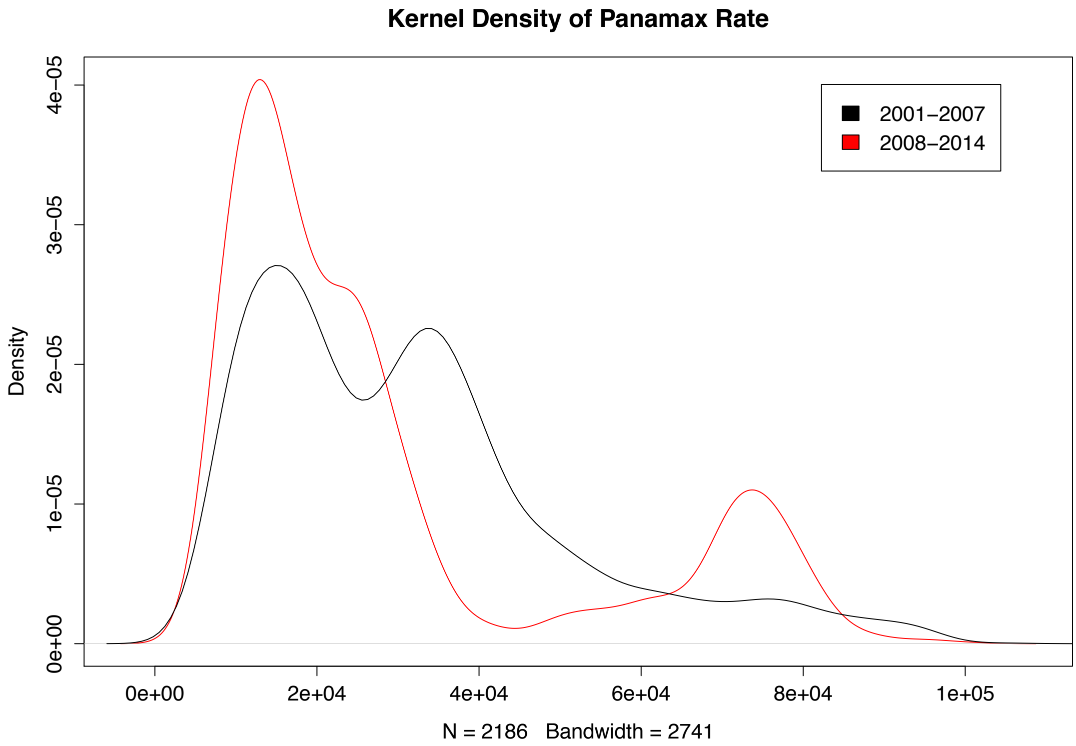
Panamax 1	2001-2007								
	RATE	NDF	AGE	DWT	FEI	CONTL	OPTL	BUNKERS	INDEX
RATE	1								
NDF	0.1005	1							
AGE	-0.0183	-0.0230	1						
DWT	0.2179	0.0855	-0.7478	1					
FEI	-0.0793	-0.0425	0.3543	-0.2949	1				
CONTL	-0.0161	0.4485	-0.0276	0.1008	-0.0155	1			
OPTL	-0.0871	-0.1242	-0.0241	-0.0324	-0.0272	-0.1454	1		
BUNKERS	0.6755	0.1579	0.0433	0.2225	-0.0775	0.0552	-0.0976	1	
INDEX	0.9257	0.1450	0.0660	0.1546	-0.0365	-0.0022	-0.1147	0.7143	1

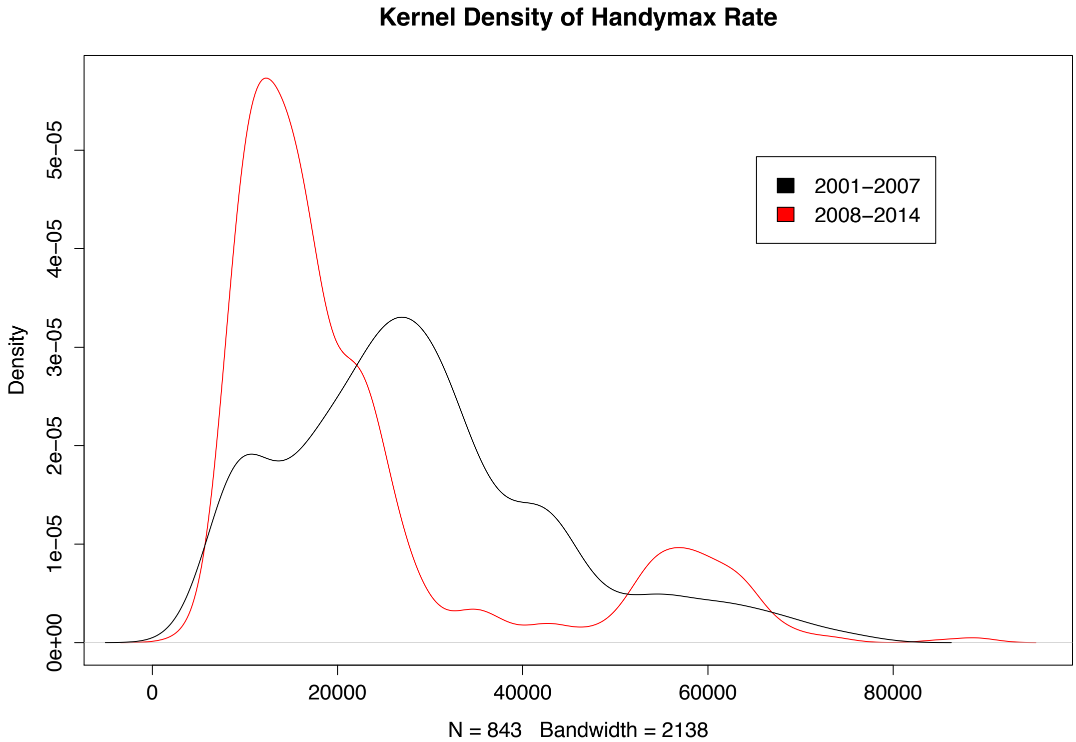
Handymax 2	**2008-2014**								
	RATE	NDF	AGE	DWT	FEI	CONTL	OPTL	BUNKERS	INDEX
RATE	1								
NDF	0.1358	1							
AGE	0.0775	-0.0053	1						
DWT	-0.2620	-0.0308	-0.8394	1					
FEI	0.0192	0.0348	0.3095	-0.3375	1				
CONTL	0.0516	0.3456	-0.0350	0.0041	0.0005	1			
OPTL	-0.0257	-0.1272	0.0071	0.0333	-0.0187	0.0236	1		
BUNKERS	-0.0149	-0.0683	-0.0601	0.2255	-0.0381	-0.1336	0.1065	1	
INDEX	0.9727	0.1412	0.1442	-0.3240	0.0668	0.0282	-0.0526	-0.0177	1

Handymax 1	**2001-2007**								
	RATE	NDF	AGE	DWT	FEI	CONTL	OPTL	BUNKERS	INDEX
RATE	1								
NDF	0.0841	1							
AGE	-0.1614	-0.0457	1						
DWT	0.3027	0.1268	-0.6679	1					
FEI	-0.1439	-0.0778	0.3446	-0.3578	1				
CONTL	0.0434	0.4937	-0.1161	0.2131	-0.1221	1			
OPTL	-0.1650	-0.2539	0.1035	-0.1608	0.0766	-0.5319	1		
BUNKERS	0.7179	0.1834	-0.0996	0.3074	-0.1349	0.1593	-0.1557	1	
INDEX	0.9762	0.1130	-0.0479	0.2062	-0.0900	0.0580	-0.1549	0.7256	1

APPENDIX 2 - RESPONSE VARIABLE DISTRIBUTION

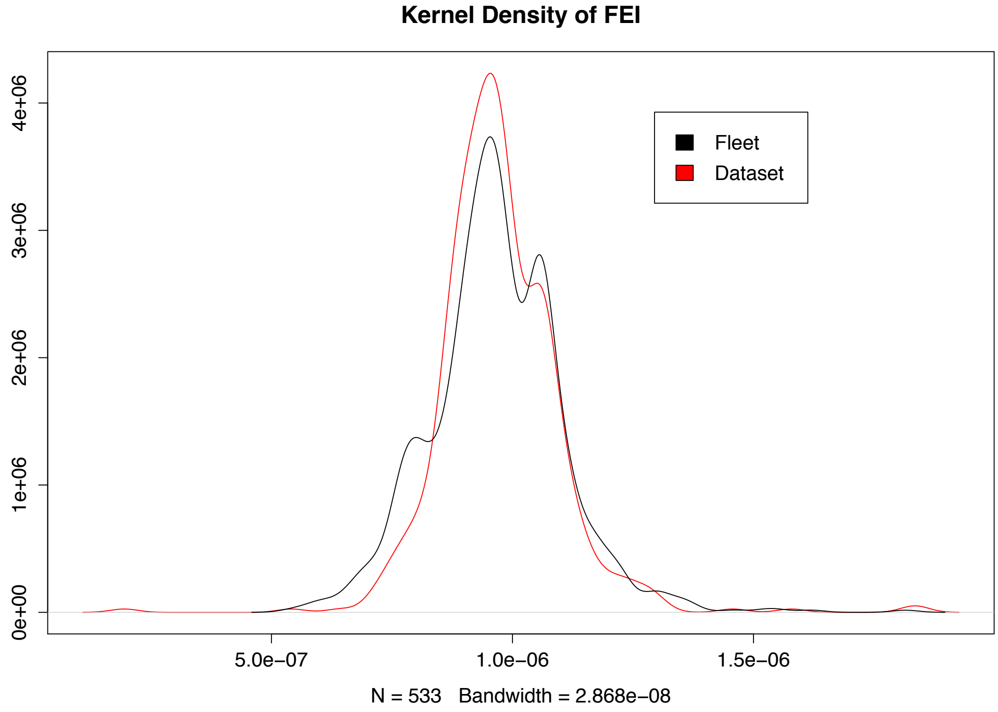




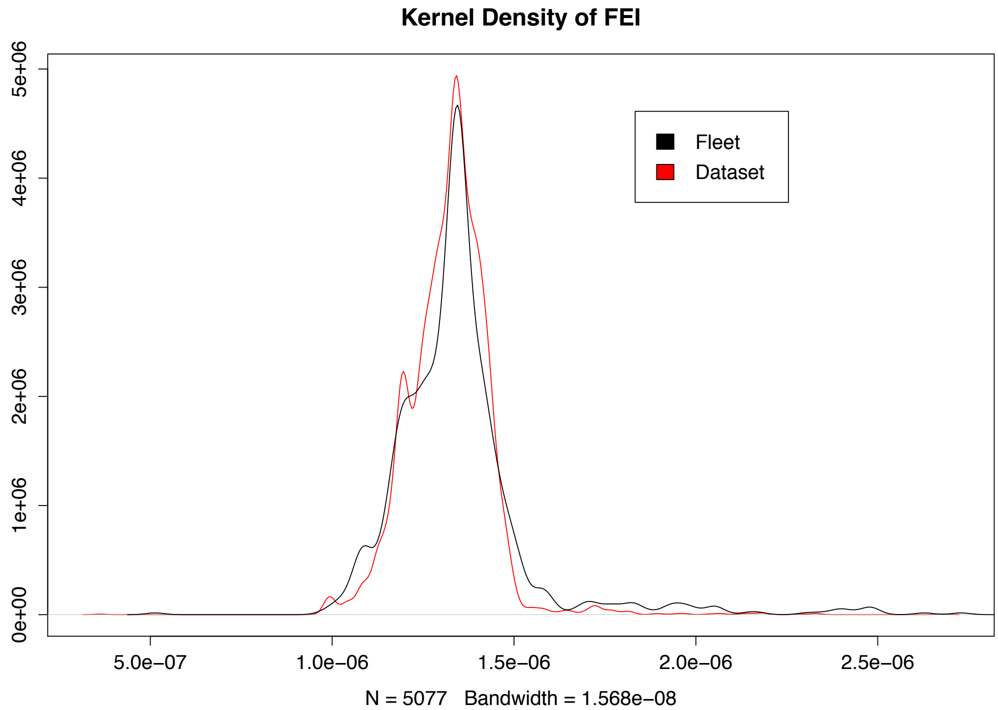

Kernel Density of Handymax Rate
Density
0e+00
1e−05
2e−05
3e−05
4e−05
5e−05
0
20000
40000
60000
80000
N = 843   Bandwidth = 2138
2001−2007
2008−2014

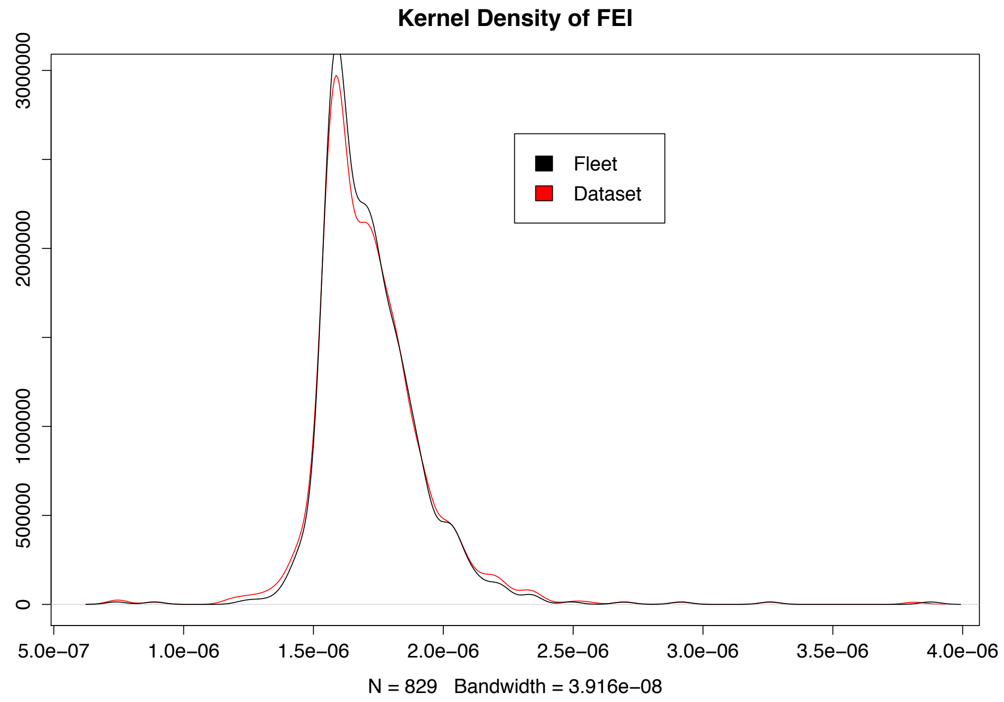
APPENDIX 3 - SELECTION BIAS

CAPESIZE



PANAMAX




Kernel Density of FEI
Fleet
Dataset
0
500000
1000000
2000000
3000000
5.0e−07
1.0e−06
1.5e−06
2.0e−06
2.5e−06
3.0e−06
3.5e−06
4.0e−06
N = 829   Bandwidth = 3.916e−08

Appendix 4 – R Code

Throughout our thesis, we have extensively used R as our modeling tool. R is a software package for statistical computing and graphics and provides a wide variety of statistical and graphical techniques. In the following section we have described a subset of code used in our modeling.

```
####################
###  FORMATTING  ###
####################

##INDEX##

####CAPESIZE####
#Remove empty rows.
capei <- subset(CapesizeIndexes, !is.na(WEEK))

#Make date-row a factor
factor(capei$DATE)

#Transform date from text to date object and store in data.frame
capei$DATE <- as.Date(capei$DATE,"%d.%m.%y")

####PANAMAX####
#Remove empty rows.
panai <- subset(PanamaxIndexes, !is.na(WEEK))

#Make date-row a factor
factor(panai$DATE)

#Transform date from text to date object and store in data.frame
panai$DATE <- as.Date(panai$DATE,"%d.%m.%y")

####HANDYMAX####
#Remove empty rows.
handyi <- subset(HandymaxIndexes, !is.na(WEEK))

#Make date-row a factor
factor(handyi$DATE)

#Transform date from text to date object and store in data.frame
handyi$DATE <- as.Date(handyi$DATE,"%d.%m.%y")


##BUNKERS##
#Remove empty rows
bunkers <- subset(Bunkers, !is.na(X380))
#Factor date column
bunkers$DATE <- factor(Bunkers$DATE)
#Transform date from text to date object
bunkers$DATE <- as.Date(bunkers$DATE,"%d.%m.%y")
#Remove text column
bunkers$Date <- NULL

##DATA##
#Transform date column from text to date object
fullData$DATE <- as.Date(fullData$DATE, "%d.%m.%y")

#SUBSETTING DATA#
```

```
##CAPESIZE
dataCapeAll <- subset(fullData, DWT >= 100000)

##PANAMAX
dataPanaAll <- subset(fullData, DWT < 100000)
dataPanaAll <- subset(dataPanaAll, DWT >= 60000)

##HANDYMAX
dataHandyAll <- subset(fullData, DWT >= 40000)
dataHandyAll <- subset(fullData, DWT < 60000)

#Split data into 2 periods for each segment
CAPE_1 <- subset(dataCapeAll, DATE < as.Date("2008-01-01") )
CAPE_2 <- subset(dataCapeAll, DATE > as.Date("2008-01-01") )
HANDY_1 <- subset(dataHandyAll, DATE < as.Date("2008-01-01") )
HANDY_2 <- subset(dataHandyAll, DATE > as.Date("2008-01-01") )
PANA_1 <- subset(dataPanaAll, DATE < as.Date("2008-01-01") )
PANA_2 <- subset(dataPanaAll, DATE > as.Date("2008-01-01") )

#FILTER DATA#

#Remove entries with missing values
CAPE_1 <- subset(CAPE_1, RATE > 0) #1459
CAPE_1 <- subset(CAPE_1, DWT > 0) #1444
CAPE_1 <- subset(CAPE_1, DRAUGHT > 0) #1325
CAPE_1 <- subset(CAPE_1, SPEED > 0) ##1308
CAPE_1 <- subset(CAPE_1, FEI > 0) #1029
CAPE_1 <- subset(CAPE_1, CONTL > 0) #1012
CAPE_1 <- subset(CAPE_1, OPTL >= 0) #713


#EXAMPLE OF RE-GROUPING OF FACTOR VARIABLES
fullDataCopy <- fullData
levels(fullDataCopy$BUILD) <- c(levels(fullDataCopy$BULD), "Other")
fullDataCopy$BUILD[fullDataCopy$BUILD == ""] <- 'Other'
fullDataCopy$BUILD <- factor(fullDataCopy$BUILD)
fullDataCopy$BUILD[fullDataCopy$BUILD == 'Philippines'] <- 'Other'
fullDataCopy$BUILD[fullDataCopy$BUILD == 'Denmark'] <- 'Other'
fullDataCopy$BUILD[fullDataCopy$BUILD == 'Italy'] <- 'Other'
fullDataCopy$BUILD[fullDataCopy$BUILD == 'South Korea'] <- 'Other'
fullDataCopy$BUILD[fullDataCopy$BUILD == 'Romania'] <- 'Other'
fullDataCopy$BUILD[fullDataCopy$BUILD == 'Spain'] <- 'Other'
fullDataCopy$BUILD[fullDataCopy$BUILD == 'Taiwan'] <- 'Other'
fullDataCopy$BUILD[fullDataCopy$BUILD == 'Vietnam'] <- 'Other'


##########################################
#### INDEX/BUNKERS MATCHING & SPLINE  ####
##########################################

#Function to match two dates. Return value is the closest date or 01-01-1970
closestDt <- function(searchDate, dateList, roundDown=FALSE)
  as.Date( sapply( searchDate , function (x) if( roundDown ){
    max( dateList[ dateList <= x ] ) } else {
      min( dateList[ dateList >= x])  }
  ), "1970-01-01")



#Iterate through every row in given data.frame
for (i in 1:nrow(HANDY_2))
{

 #Copy the current fixture row
  currentFixture <- HANDY_2[i,]

 #Find the closest date in a data.frame containing indices
  date <- closestDt(currentFixture$DATE,handyi$DATE)

 #Get the index entry for given date and store it in a variable
```

```
  indexData <- handyi[which(handyi$DATE == date),]

 #Store each index value in a y variable
  y <-
c(indexData$X1MONTH,indexData$X6MONTH,indexData$X1YEAR,indexData$X3YEAR,inde
xData$X5YEAR)

 #Store length as days in a x variable
  x <- c(0.083*360,0.5*360,1*360,3*360,5*360)

 #Spline x and y using a spline function with method "natural", and store in
variable
  indexSpline <- spline(x,y,n=1770,method="natural")

 #Use contract length in fixture to find correct index value
  fixtureIndex <- indexSpline$y[which(abs(indexSpline$x-
currentFixture$CONTL)==min(abs(indexSpline$x-currentFixture$CONTL)))]

 #Store the index value in data.frame
  HANDY_2[i,]$INDEX <- fixtureIndex[1]
}

#Iterate though all data and match with a bunkers value
for (i in 1:nrow(fullData))
{
  currentFixture <- fullData[i,]
  date <- closestDt(currentFixture$DATE,bunkers$DATE)
  bunkersData <- bunkers[which(bunkers$DATE == date),]
  if(currentFixture$FUEL == "HFO"){
    fullData[i,]$BUNKERS <- bunkersData$X380
  }else{
    fullData[i,]$BUNKERS <- bunkersData$X180
  }
}


#Correlation and description
write.table(cor(PANA_1), file = "Correlation_PANA_1.csv", sep = ",",
col.names = NA, qmethod = "double")
write.table(describe(PANA_1), file = "Description_PANA_1.csv", sep = ",",
col.names = NA, qmethod = "double")




#############################
#### MODEL OPTIMIZATION  ####
#############################

##Method is derived from page 130 in wood

##INPUT PARAMETERS
parameters <-
c("s(INDEX)","s(BUNKERS)","s(OPTL)","s(CONTL)","s(BEAM)","s(AGE)","s(SPEED)"
,"s(CONS)","s(DWT)","s(FEI)","s(LOA)","s(NDF)","POD","DIS","FLAG","BUILD","E
NG","FUEL")


##DECIDE DATA INPUT
dataInput <- dataPana

##CREATE AN EMPTY ARRAY FOR PARAMETER ORDER COMBINATIONS
parameterArray <- array(1:100,c(length(parameters),length(parameters)))

##CREATE AN EMPTY ARRAY FOR ALL COMBINATIONS OF PARAMETERS
allParametersArray <-
array(1:100,c(length(parameters)*length(parameters),length(parameters)))

##FILL INN ARRAY FOR ORDER COMBINATIONS
```

```
for(x in 1:length(parameters))
{
  parameterArray[x,] <- parameters
  object <- parameters[1]
  parameters <- c(parameters,object)
  parameters <- parameters[-1]
}

##MAKE ALL THE COMBINATIONS OF THE PARAMETERS
zeros = 1
parameterSelection = 1
for (r in 1:(length(parameters)*length(parameters)))
{
  para <- parameterArray[parameterSelection,]
  allParametersArray[r,] =
c(para[1:zeros],rep(c(0),times=length(parameters)-zeros))
  zeros <- zeros + 1
  if(zeros > length(parameters))
  {
    zeros =  1
    parameterSelection = parameterSelection + 1
  }
}


##DELETE DUPLICATE ROWS
deleteIndex = 0
for(p in 1:length(parameters))
{
  if(p!=1)
  {
    allParametersArray <- allParametersArray[-(p*length(parameters)-
deleteIndex),]
    deleteIndex = deleteIndex+1
  }
}

#CREATE EMPTY LISTS FOR RESULTS
models = list()
gcv = c()
radj = c()
nHandy = c()
##ITERATE THROUGH ALL COMBINATIONS AND DO THE REGRESSION. STORE THE RESULTS
for(row in 1:nrow(allParametersArray))
{
  clean <- allParametersArray[row,]
  clean <- clean[clean!=0]
  model <- gam(
as.formula(paste("RATE~",paste(clean,collapse="+"))),family=Gamma(link="log"
),data=dataInput)
  models <- list(models,allParametersArray[row,])
  gcv[row] <- model$gcv.ubre
  radj[row] <- summary(model)$r.sq
  nHandy[row] <- summary(model)$n
}


#REMOVE USED VARIABLES
remove(clean)
remove(deleteIndex)
remove(object)
remove(p)
remove(para)
remove(parameters)
remove(parameterSelection)
remove(r)
remove(row)
remove(x)
remove(zeros)
```

```
##PLOT RESULTS
plot(gcv, type="l", col="blue")




#CHECK AND PRINT ONE SPECIFIC MODEL
checkModel <- function(modelNumber){
  clean <- allParametersArray[modelNumber,]
  clean <- clean[clean!=0]
  model <- gam(
as.formula(paste("RATE~",paste(clean,collapse="+"))),family=Gamma(link="log"
),data=dataInput)
  return(summary(model))
}

#PRINT THE BEST MODEL BASED ON GCV
printBestModel <- function(){

  clean <- allParametersArray[which(gcv==min(gcv)),]
  clean <- clean[clean!=0]
  model <- gam(
as.formula(paste("RATE~",paste(clean,collapse="+"))),family=Gamma(link="log"
),data=dataInput)
  return(summary(model))
}

#GET THE BEST MODEL AS A VARIABLE
bestModel <- function(){
  clean <- allParametersArray[which(gcv==min(gcv)),]
  clean <- clean[clean!=0]
  model <- gam(
as.formula(paste("RATE~",paste(clean,collapse="+"))),family=Gamma(link="log"
),data=dataInput)
  return(model)
}

#PLOT
par(mfrow=c(1,1))
plot(pana1, residuals=T, pch=10, cex=0.25, scheme=1, col='#FF8000',
shade=T,shade.col='gray90')



############################################
#################  MODELS  #################
############################################

#MODEL TESTING EXAMPLE#

#MODEL TEST 1
cape1 <- gam(RATE ~
             s(INDEX)
             ,family = Gamma(link="log")
             ,data=CAPE_1)
summary(cape1)

#MODEL TEST 2
cape1 <- gam(RATE ~
             s(INDEX)
          +  s(AGE)
             ,family = Gamma(link="log")
             ,data=CAPE_1)
summary(cape1)

#MODEL TEST 3
cape1 <- gam(RATE ~
             s(INDEX)
          +  s(AGE)
          +  s(DWT)
```

```
             ,family = Gamma(link="log")
             ,data=CAPE_1)
summary(cape1)
```

#MODEL TEST 3

```
cape1 <- gam(RATE ~
             s(INDEX)
          +  s(AGE)
          +  s(DWT)
          +  s(FEI)
             ,family = Gamma(link="log")
             ,data=CAPE_1)
summary(cape1)
```

#############################
FINISHED MODELS
#############################

##CAPESIZE##

#CAPE1 - 2001-2007

```
cape1 <- gam(RATE ~ s(NDF) + s(AGE) + s(DWT) + s(FEI) + s(CONTL) + s(INDEX)
+ BUILD + FUEL + FLAG + ENGT + POD ,family = Gamma(link="log") ,data=CAPE_1)
summary(cape1)
```

#CAPE2 - 2008-2014

```
cape2 <- gam(RATE ~ s(NDF) + s(AGE) + s(DWT) + s(FEI) + s(CONTL) + s(INDEX)
+ s(BUNKERS) + BUILD + FUEL + FLAG + ENGT + POD
,family = Gamma(link="log")    ,data=CAPE_2)
summary(cape2)
```

##PANAMAX##

#PANA1 2001-2007

```
pana1 <- gam(RATE ~
s(NDF) + s(AGE) + s(DWT) + s(FEI) + s(CONTL) + s(OPTL) + s(INDEX) + BUILD
+ FUEL + FLAG + ENGT + POD ,family = Gamma(link="log") ,data=PANA_1)
summary(pana1)
```

#PANA2 2008-2014

```
pana2 <- gam(RATE ~
s(NDF) + s(AGE) + s(DWT) + s(FEI) + s(CONTL) + s(OPTL) + s(INDEX)
+ s(BUNKERS) + BUILD + FUEL + FLAG + ENGT + POD
,family = Gamma(link="log")   ,data=PANA_2)
summary(pana2)
```

##HANDYMAX##

#HANDY1 - 2001-2007

```
handy1 <- gam(RATE ~  s(NDF)+ s(AGE)+ s(DWT)+ s(FEI)+ s(CONTL) + s(OPTL) +
s(INDEX) + BUILD  + FUEL + FLAG + ENGT + POD + GEAR
,family = Gamma(link="log")   ,data=HANDY_1)
summary(handy1)
```

#HANDY2 - 2008-2014

```
handy2 <- gam(RATE ~ s(NDF) + s(AGE) + s(DWT) + s(FEI) + s(CONTL) + s(OPTL)
+ s(INDEX) + s(BUNKERS) + BUILD + FUEL + FLAG + ENGT + POD + GEAR
,family = Gamma(link="log")   ,data=HANDY_2)
summary(handy2)
```

#############################
KNOTS OPTIMIZATION
#############################

```
testgcv1<-0
for (i in 1:60)
{
```

```
  gamTEST <- gam(RATE ~ s(INDEX,k=i)
                 ,family = Gamma(link="log")
                 ,data = dataPana,gamma=1.4)
  testgcv1[i] <- gamTEST$gcv.ubre

}

gam_pana5 <- gam(RATE ~ s(LOA) + s(NDF) + POD + DIS + FLAG + BUILD + ENG +
FUEL +
                 s(INDEX) + s(BUNKERS) + s(OPTL) + s(CONTL) +
                 s(BEAM) + s(AGE) + s(SPEED) + s(CONS) + s(DWT)
               ,family = Gamma(link="log")
               ,data = dataPana)
indexOfModel <- c(1:length(gcv))

for(i in 1:nrow(comparisonArray))
{
  comparisonArray[i,2] <- gcv[i]
  comparisonArray[i,3] <- radj[i]
}

#########################################
#### MODEL APPLICATION / SIMULATION  ####
#########################################

#Create standard ship
standardShip <- NULL
standardShip$NDF <- 16
standardShip$AGE <- 10
standardShip$DWT <- 74000
standardShip$FEI <- 1.32722e-06
standardShip$CONTL <- 360
standardShip$OPTL <- 50
standardShip$INDEX <- 0
standardShip$BUILD <- "Japan"
standardShip$FUEL <- "HFO"
standardShip$FLAG <- "Panama"
standardShip$ENGT <- "MAN B. & W."
standardShip$POD <- "Atlantic"
standardShip$BUNKERS <- 500

#Modify standard ship to create eco ship
ecoShip <- standardShip
ecoShip$DWT <- 80000
ecoShip$FEI <- 1.04353e-06
ecoShip$AGE <- 4

#Divide index into two time periods
panai_1 <- subset(panai, DATE < as.Date("2008-01-01") )
panai_2 <- subset(panai, DATE >= as.Date("2008-01-01") )

#Divide bunkers into two time periods
bunkers_1 <- subset(bunkers, DATE < as.Date("2008-01-01") )
bunkers_2 <- subset(bunkers, DATE >= as.Date("2008-01-01") )

#Iterate through the index
for (i in 1:nrow(panai_2))
{
  #Set index for each ship
  ecoShip$INDEX <- panai_2[i,]$X1YEAR
  standardShip$INDEX <- panai_2[i,]$X1YEAR
  #Set bunkers for each ship
  ecoShip$BUNKERS <- standardShip$BUNKERS <- bunkers_2[i,]$X380
  #Use model to predict new rate with eco/standard ship and store values
       panai_2[i,]$ECO <-
       Gamma(link="log")$linkinv(predict.gam(pana22,ecoShip))
       panai_2[i,]$STD <-
       Gamma(link="log")$linkinv(predict.gam(pana22,standardShip))
}
```

```
#Plot the results
par(new=TRUE)
plot(panai_2$ECO-
panai_2$STD,type="l",col="blue",xaxt="n",yaxt="n",xlab="",ylab="")
axis(4)
mtext("y2",side=4,line=3)
legend("topleft",col=c("red","blue"),lty=1,legend=c("y1","y2"))



###############################
#### DISTRIBUTION FITTING  ####
###############################

plot(ecdf(PANA_1$RATE),main="Empirical cumulative distribution function")

z.norm<-(PANA_1$RATE-mean(PANA_1$RATE))/sd(PANA_1$RATE) ## standardized data
qqnorm(z.norm) ## drawing the QQplot
abline(0,1) ## drawing a 45-degree reference line

curve(dgamma(PANA_1$RATE, scale=1.5, shape=2),from=0, to=15, main="Gamma
distribution")

alpha <- mean(PANA_1$RATE)/var(PANA_1$RATE)
beta <- (mean(PANA_1$RATE))*2/var(PANA_1$RATE)

x.gamma = rgamma(n=1000,scale=0.83,shape=10.59)
x.weibull = rweibull(n=1000,scale=3.5,shape=14.1)

hist(x.gamma)
qqplot(PANA_1$RATE,x.gamma)
plot(density(CAPE_2$RATE))
plot(d)
plot(density(x.weibull))
hist(x.weibull)
qqplot(PANA_1$RATE,x.gamma)

d <- density(PANA_1$RATE)
d1 <- density(PANA_2$AGE)
p <- density(CAPE_1$RATE)
p1 <- density(CAPE_2$RATE)

c <- density(FDF_HANDY$FEI)
c1 <- density(FNF_HANDY$FEI)
plot(c1, main="Kernel Density of RATE",col="red")
points(c,t="l")

polygon(c1, col="black", border="black")
polygon(c1, col="black", border="red")

colfill<-c(2:(2+length(levels(cyl.f))))
legend(locator(1), c("1","2"), fill=c("black","red"))

######################
#### MODEL PLOTS  ####
######################

par(mfrow=c(3,2))

#AGE
plot(cape1,select=2,scheme=1,ylim=c(-
0.5,0.3),xlim=c(0,26),xlab="AGE***",main="Capesize 2001-2007",seWithMean=T)
plot(cape22,select=2,scheme=1,ylim=c(-
0.5,0.3),xlim=c(0,26),xlab="AGE***",main="Capesize 2008-2014",seWithMean=T)
plot(pana1,select=2,scheme=1,ylim=c(-
0.4,0.2),xlim=c(0,26),xlab="AGE***",main="Panamax 2001-2007",seWithMean=T)
plot(pana22,select=2,scheme=1,ylim=c(-
0.4,0.2),xlim=c(0,26),xlab="AGE***",main="Panamax 2008-2014",seWithMean=T)
plot(handy1,select=2,scheme=1,ylim=c(-
0.4,0.2),xlim=c(0,26),xlab="AGE***",main="Handymax 2001-2007",seWithMean=T)
```

```
plot(handy22,select=2,scheme=1,ylim=c(-
0.4,0.2),xlim=c(0,26),xlab="AGE***",main="Handymax 2008-2014",seWithMean=T)
```

#DWT

```
plot(cape1,select=3,scheme=1,ylim=c(-
0.5,0.8),xlim=c(140000,220000),xlab="DWT***",main="Capesize 2001-
2007",seWithMean=T)
plot(cape22,select=3,scheme=1,ylim=c(-
0.5,0.8),xlim=c(140000,220000),xlab="DWT***",main="Capesize 2008-
2014",seWithMean=T)
plot(pana1,select=3,scheme=1,ylim=c(-0.3,0.3),xlab="DWT***",main="Panamax
2001-2007",seWithMean=T)
plot(pana22,select=3,scheme=1,ylim=c(-0.3,0.3),xlab="DWT***",main="Panamax
2008-2014",seWithMean=T)
plot(handy1,select=3,scheme=1,ylim=c(-0.2,0.2),xlab="DWT***",main="Handymax
2001-2007",seWithMean=T)
plot(handy22,select=3,scheme=1,ylim=c(-0.2,0.2),xlab="DWT",main="Handymax
2008-2014",seWithMean=T)
```

#FEI

```
plot(cape1,select=4,scheme=1,ylim=c(-0.2,0.4),xlim=c(6.0e-07,1.2e-
06),xlab="FEI*",main="Capesize 2001-
2007",seWithMean=T,residuals=T,pch=1,cex=0.3,)
plot(cape22,select=4,scheme=1,ylim=c(-0.2,0.4),xlim=c(6.0e-07,1.2e-
06),xlab="FEI",main="Capesize 2008-
2014",seWithMean=T,residuals=T,pch=1,cex=0.3)
plot(pana1,select=4,scheme=1,ylim=c(-0.2,0.2),xlim=c(1.0e-06,2.3e-
06),xlab="FEI**",main="Panamax 2001-
2007",seWithMean=T,residuals=T,pch=1,cex=0.3)
plot(pana22,select=4,scheme=1,ylim=c(-0.2,0.2),xlim=c(1.0e-06,2.3e-
06),xlab="FEI**",main="Panamax 2008-
2014",seWithMean=T,residuals=T,pch=1,cex=0.3)
plot(handy1,select=4,scheme=1,ylim=c(-0.2,0.2),xlim=c(1.3e-06,2.5e-
06),xlab="FEI***",main="Handymax 2001-
2007",seWithMean=T,residuals=T,pch=1,cex=0.3)
plot(handy22,select=4,scheme=1,ylim=c(-0.2,0.2),xlim=c(1.3e-06,2.5e-
06),xlab="FEI",main="Handymax 2008-
2014",seWithMean=T,residuals=T,pch=1,cex=0.3)
```

#CONTL

```
plot(cape1,select=5,scheme=1,ylim=c(-
0.15,0.15),xlim=c(0,1100),xlab="CONTL***",main="Capesize 2001-
2007",seWithMean=T)
plot(cape22,select=5,scheme=1,ylim=c(-
0.15,0.15),xlim=c(0,1100),xlab="CONTL**",main="Capesize 2008-
2014",seWithMean=T)
plot(pana1,select=5,scheme=1,ylim=c(-
0.15,0.15),xlim=c(0,1100),xlab="CONTL***",main="Panamax 2001-
2007",seWithMean=T)
plot(pana22,select=5,scheme=1,ylim=c(-
0.2,0.4),xlim=c(0,1100),xlab="CONTL***",main="Panamax 2008-
2014",seWithMean=T)
plot(handy1,select=5,scheme=1,ylim=c(-
0.1,0.15),xlim=c(0,700),xlab="CONTL***",main="Handymax 2001-
2007",seWithMean=T)
plot(handy22,select=5,scheme=1,ylim=c(-
0.1,0.15),xlim=c(0,700),xlab="CONTL.",main="Handymax 2008-
2014",seWithMean=T)
```

#OPTL

```
plot(pana1,select=6,scheme=1,ylim=c(-
0.2,0.2),xlim=c(0,110),xlab="OPTL***",main="Panamax 2001-2007",seWithMean=T)
plot(pana22,select=6,scheme=1,ylim=c(-
0.2,0.2),xlim=c(0,150),xlab="OPTL***",main="Panamax 2008-2014",seWithMean=T)
plot(handy1,select=6,scheme=1,ylim=c(-
0.3,0.3),xlim=c(0,100),xlab="OPTL*",main="Handymax 2001-2007",seWithMean=T)
plot(handy22,select=6,scheme=1,ylim=c(-
0.15,0.15),xlim=c(0,120),xlab="OPTL**.",main="Handymax 2008-
2014",seWithMean=T)
```

```
#INDEX
plot(cape1,select=6,scheme=1,ylim=c(-1.5,2),xlab="INDEX***",main="Capesize
2001-2007",seWithMean=T)
plot(cape22,select=6,scheme=1,ylim=c(-1.5,2),xlab="INDEX***",main="Capesize
2008-2014",seWithMean=T)
plot(pana1,select=7,scheme=1,ylim=c(-1.5,1.5),xlab="INDEX***",main="Panamax
2001-2007",seWithMean=T)
plot(pana22,select=7,scheme=1,ylim=c(-1.5,1.5),xlab="INDEX***",main="Panamax
2008-2014",seWithMean=T)
plot(handy1,select=7,scheme=1,ylim=c(-
1.5,1.5),xlab="INDEX***",main="Handymax 2001-2007",seWithMean=T)
plot(handy22,select=7,scheme=1,ylim=c(-
1.5,1.5),xlab="INDEX***",main="Handymax 2008-2014",seWithMean=T)



#NDF
plot(cape1,select=1,scheme=1,ylim=c(-
0.5,0.5),xlim=c(0,200),xlab="NDF.",main="Capesize 2001-2007",seWithMean=T)
plot(cape22,select=1,scheme=1,ylim=c(-
0.4,0.2),xlim=c(0,200),xlab="NDF**",main="Capesize 2008-2014",seWithMean=T)
plot(pana1,select=1,scheme=1,ylim=c(-
0.4,0.2),xlim=c(0,350),xlab="NDF***",main="Panamax 2001-2007",seWithMean=T)
plot(pana22,select=1,scheme=1,ylim=c(-
0.1,0.1),xlim=c(0,50),xlab="NDF***",main="Panamax 2008-2014",seWithMean=T)
plot(handy1,select=1,scheme=1,ylim=c(-
0.5,0.5),xlim=c(0,200),xlab="NDF*",main="Handymax 2001-2007",seWithMean=T)
plot(handy22,select=1,scheme=1,ylim=c(-
0.5,0.5),xlim=c(0,200),xlab="NDF*",main="Handymax 2008-2014",seWithMean=T)


#BUNKERS
#plot(cape1,select=7,scheme=1,ylim=c(-
0.2,0.15),xlab="BUNKERS",main="Capesize 2001-2007",seWithMean=T)
plot(cape22,select=7,scheme=1,ylim=c(-
0.2,0.15),xlab="BUNKERS***",main="Capesize 2008-2014",seWithMean=T)
#plot(pana1,select=8,scheme=1,ylim=c(-
0.1,0.1),xlab="BUNKERS***",main="Panamax 2001-2007",seWithMean=T)
plot(pana22,select=8,scheme=1,ylim=c(-
0.1,0.1),xlab="BUNKERS***",main="Panamax 2008-2014",seWithMean=T)
#plot(handy1,select=8,scheme=1,ylim=c(-
0.15,0.1),xlab="BUNKERS",main="Handymax 2001-2007",seWithMean=T)
plot(handy22,select=8,scheme=1,ylim=c(-
0.15,0.1),xlab="BUNKERS**",main="Handymax 2008-2014",seWithMean=T)

#POD
termplot(cape1,se=T,ylim=c(-0.25,0.05),term="POD",main="Capesize 2001-
2007",col.term="black",col.se="black",lty.se=2,xlab="POD***")
termplot(cape22,se=T,ylim=c(-0.25,0.05),term="POD",main="Capesize 2008-
2014",col.term="black",col.se="black",lty.se=2,xlab="POD***")
termplot(pana1,se=T,ylim=c(-0.2,0.05),term="POD",main="Panamax 2001-
2007",col.term="black",col.se="black",lty.se=2,xlab="POD***")
termplot(pana22,se=T,ylim=c(-0.2,0.05),term="POD",main="Panamax 2008-
2014",col.term="black",col.se="black",lty.se=2,xlab="POD***")
termplot(handy1,se=T,ylim=c(-0.2,0.05),term="POD",main="Handymax 2001-
2007",col.term="black",col.se="black",lty.se=2,xlab="POD")
termplot(handy22,se=T,ylim=c(-0.2,0.05),term="POD",main="Handymax 2008-
2014",col.term="black",col.se="black",lty.se=2,xlab="POD***")

#BUILD
termplot(pana1,se=T,ylim=c(-0.05,0.05),term="BUILD",main="Panamax 2001-
2007",col.term="black",col.se="black",lty.se=2,xlab="BUILD")
termplot(pana2,se=T,ylim=c(-0.05,0.05),term="BUILD",main="Panamax 2008-
2014",col.term="black",col.se="black",lty.se=2,xlab="BUILD")
```