Norwegian School of Economics
Bergen, Fall 2019

Vessel Valuation: An Analysis of the Joint Non-Linear Impact of Vessel Age and Time of Sale in the Market for Handysize Dry Bulkers

By
Gotfred Andenes & Markus Mølkjær Klepelv

Supervisor
Prof. Dr. Roar Os Adland

Master Thesis in Financial Economics

NORWEGIAN SCHOOL OF ECONOMICS

Abstract

This thesis investigates the joint influence of transaction time and vessel age on the valuation of second-hand tonnage. Towards this end, we use individual sales data for Handysize dry bulkers and apply a semiparametric valuation model within the Generalized Additive Model framework. Our empirical results suggest a significant non-linear relationship between the time of the transaction and the asset's age. We find that the volatility of vessels' year-on-year return increases with age, as well as a difference in the depreciation curve depending on the state of the market. In a booming (recovering) market, the average depreciation curve is in major terms concave (convex). These findings may yield valuable insights for market players on the lookout for investments. We argue that asset players, contrary to industrial, should seek older, more volatile, vessels, as these could potentially yield a higher return on invested capital.

Acknowledgments

We would like to express our sincerest gratitude towards those who made this thesis possible. First and foremost, we would like to thank our supervisor, Professor Roar Adland. His guidance and extensive knowledge have been inspiring and instructive throughout the research.

During the development of this paper, we have had the opportunity to use Vaage Ship Management and EOS Corporate's offices, as well as benefit from their industry knowledge, and for that, we would like to express our gratitude. Besides, we would like to thank all of our colleagues at NHH for friendship, cooperation and support throughout our time at the school.

Finally, the financial support provided to us by the Norwegian Shipowners Association is highly appreciated.

Table of Contents

1 Introduction

The first cargo was moved by sea more than 5000 years ago. Since then, the maritime shipping industry has become the single most important market for international transportation. According to Stopford (2009), the shipping industry comprises four different markets: the freight market, the sale and purchase market, the newbuilding market, and the demolition market.

The sale and purchase market facilitates transactions of second-hand vessels for investors. In a market that thrives on price volatility, market players able to predict the market could achieve a competitive advantage. Alizadeh and Nomikos (2006) prove that the introduction of trading strategies can yield superior returns, indicating a breach of the Efficient Market Hypothesis, which suggests that all players are subjected to the market return. Furthermore, Adland and Koekebakker (2007) find evidence of non-linearity in the average depreciation curve of Handysize dry bulkers. These results suggest that the second-hand value of a vessel could be described as a partially non-linear function of size, age, and time.

The objective of this thesis is to develop an econometric model to further analyze and investigate the possibility of non-linearity in the relationship between the age of a vessel and the time of a transaction. In the pursuit of this objective, we build on previous developments in the field of semi-parametric valuation modeling research, notably Generalized Additive Models (Hastie & Tibshirani, 1990), as these provide an appropriate framework for heterogeneous asset valuation (Adland & Koehn, 2019). In this paper, we focus on the world's Handysize bulker fleet, which today consists of 3,738 vessels (Clarksons, 2019). This part of the shipping industry is characterized by its heterogeneous nature and we, therefore, account for individual microeconomic determinants.

Our contributions to the research field are twofold. Firstly, we analyze the joint influence of the age of a vessel and the time of a transaction on vessel valuation. Thus, we are able to follow the price development of a vessel with predetermined microeconomic specifications. A challenge with today's time series analyses is that they present average prices at a constant age, and are subject to brokers' estimates. In other words, they account for the price development of non-existing ships. By allowing vessel age to increase with time, our model reflects the effect of depreciation. Additionally, by basing the valuations on transactional data, brokers' "guesstimation" biases are removed. Secondly, a vessel valuation model that can account for both general market factors and vessel-specific characteristics is valuable for market players. Accordingly, we interpret the implications of our results for long-term and short-term investors.

The remainder of this thesis is structured as follows. In section 2, relevant academic literature is presented, with a focus on the Efficient Market Hypothesis, as well as the development of the econometric model. Subsequently, section 3 contains a description of the data for second-hand transactions on Handysize dry bulkers, retrieved from Clarksons' shipping intelligence network. It also provides insight into the decisive price determinants and their expected sign of the coefficient. Section 4 provides the basis for the methodology and theoretical framework used throughout the research. Furthermore, the empirical analysis and related results are presented and discussed in section 5. Lastly, section 6 presents our concluding remarks, as well as limitations and suggestions for further research.

2 Literature Review

The first studies testing for market efficiency in the second-hand shipping market were conducted by Hale and Vanags (1992). Their studies examined second-hand valuation and cointegration within three different segments of drybulk shipping. Through their research, they found cointegration between two out of three segments, indicating inefficient markets. Some years later, Glen (1997) revisited Hale and Vanags' research by applying Johansen's maximum likelihood approach to it. He also expanded Hale and Vanags' data by including the tanker market and found evidence of cointegration in both the drybulk and tanker segments. However, Glen claimed that the question of whether the second-hand market is efficient remained unsolved, as the cointegration might as well be a result of a random stochastic force.

Further studies on market efficiency investigated the presence of excess return in the market. In 2002, Kavussanos and Alizadeh used four methods of Vector Autoregressive Models to investigate the validity of the Efficient Market Hypothesis in the drybulk sector. In all but a couple of cases, they rejected the hypothesis on the 5% confidence level. On the contrary, Adland and Koekebakker (2004) discovered that "trading rules are generally not capable of producing excess wealth above a buy-and-hold benchmark when accounting for transaction costs and the potential price-slippage in an illiquid market."

In pursuit of excess return, Soedal et al. (2009) applied a theoretical real options model with stochastic freight rates, switching between the drybulk and the tanker markets. In all but one observation, they found the market to be more or less efficient. On the other hand, based on the same dataset, Adland et al. (2018) found evidence of the presence of a "lemon problem." Vessels with previously reported transactions are valued with a premium, as this is a signal of attractiveness and quality, or at least that other investors have found it a suitable candidate.

Pruyn et al. (2011) summarized all research on second-hand vessel value estimation from the last 20 years and claimed that besides varying results, the question of whether the Efficient Market Hypothesis holds is inconclusive. An argument was that the broker bias made it more likely that the extent research had tested brokers' expectations rather than actual market behavior.

The next section of the literature comprehends modeling techniques on second-hand vessel valuation. Charemza and Gronicki (1981) introduced equations with freight rates and activity levels as determinants for ship prices and found a significant correlation between activity levels, freight rates, and ship prices. Strandenes (1984, 1986) came up with similar results stating that vessel prices are a function of short- and long term profits taken depreciation into account. Also, Beenstock (1985) pursue potential interdependency between the freight market and the sale and purchase market. He proposed an econometric model, applied by Beenstock and Vergottis (1989a, 1989b, 1992) to the tanker and drybulk market. Their results indicated that "inter alia, freight rates, lay-up, new and second-hand prices and the size of the fleet are jointly and dynamically determined."

In 2003, Tsalokis et al. presented an econometric approach to second-hand price modeling, suggesting a theoretical error correlation model with a structure based on cyclical businesses. Their results showed that both in the short- and long run, timecharter rates and newbuilding activity are the most decisive price determinants.

To avoid broker bias and measurement error, Adland and Koekebakker (2007) used cross-sectional reported transaction data in their analysis. Thus, they were able to analyze the vessel prices of the Handysize drybulk sector through a multivariate non-parametric approach. Their findings indicated that vessel price is a function of age, DWT and the state of the market. Furthermore, they concluded that a three-factor model is not fully capable of explaining the observed vessel prices. Adland and Koehn (2019) continued Adland and Koekebakker's research by applying a semi-parametric approach, allowing for more price determinants to be included. They suggested the use of a Generalized Additive Model and found that the decisive determinants in the chemical tanker market included age, DWT, market conditions, cargo diversity, and IMO grade.

Our research builds on the semi-parametric valuation model used by Adland and Koehn (2019) for their analysis of individual asset sales in the chemical tanker market. Their study focused on the joint non-linear effect of the age and size of a vessel on second-hand valuation, as well as the impact of microeconomic determinants. We extend the research by investigating the joint effect of the age of a vessel and the time of a transaction on vessel valuation. By including a time parameter as one of our price determinants, we have the possibility to follow the price development of a predetermined ship from its year of build. In addition, we expand the use of semi-parametric modeling to the Handysize dry bulker segment.

3 Data description

3.1 Price Determinants

In this part, a selection of determinants that are expected to influence the valuation of second-hand Handysize drybulk vessels is presented. The variables of choice follow the literature by Adland & Koekebakker (2007), Adland et al. (2018) and Adland & Koehn (2019), as well as economic theory on asset valuation. Table 3.1 displays the variables of choice, arranged in numeric- and dummy variables. Moreover, the expected sign on the coefficient and interpretation of the respective variables is included.

Table 3.1: Price Determinants

Determinants	Expected sign	Interpretation
Numeric Variables		
Saledate	-	Time parameter for the date of the transaction (in years)
Salesage	Negative	The vessels age at the time of the transaction (years).
DWT	Positive	The deadweight carrying capacity of the vessel in tonnes.
FEI	Negative	Fuel efficiency index, effectively measuring fuel consumption on a “grams per tonnemile” basis.
HP	Positive	The number of engine horsepower.
Speedknots	Positive	The vessel's reported speed.
Noholds	-	The number of holds.
Dummy Variables		
Countryindicator.D	-	Dummy for the vessel's country of build other than Japan.
Vesselfueltype.D	-	Dummy for vessel fuel types other than HFO.
Enginebrand.D	-	Dummy for engine manufacturer other than MAN B&W.
Gearindicator Y.D	Positive	Dummy indicating that the vessel is equipped with the appropriate loading gear.

Within the Handysize drybulk sector, there is a lack of recorded macroeconomic data such as timecharter rates, spot earnings, and newbuilding prices. Thus, a natural starting point is to include a time parameter (dummy) for the date of the vessel transactions. Furthermore, Adland et al. (2017) found that including a third-party market index comes with an endogeneity problem, as the index tends to dominate the model. In this report, the data points are based on yearly observations due to the amount of transactional data available. It is expected that the time parameter will account for a substantial amount of the explanatory power, due to its reflection of the market conditions.

Several microeconomic determinants available in the second-hand transaction data are included in the valuation model. The vessel size is important, as more freight capacity is expected to translate into a higher timecharter equivalent. The standard measure of size in the industry is deadweight tonnage (DWT) and is incorporated in favor of the highly correlated grain capacity, referring to Table A1 in the appendix for correlation matrices.

From the perspective of ship owners, the quality of the ship is essential as it affects operating expenses and life expectancy. Thus, a natural parameter is the age of the vessel at the time of the transaction. More, it is a general understanding in the industry, supported by Adland et al. (2018), that there is evidence supporting quality premium and discount in asset prices for ships built in different countries. Taking the central limit theorem into account, a dummy variable, "Countryindicator", for builder countries with more than 20 transactions is derived (Keller, 2009). Based on the research by Adland et al. (2018), ships built in Chinese shipyards, relative to Japanese yards, are expected to trade at a discount. Ships originating from South Korea are believed to differentiate in terms of quality perception, as the shipyards' reputation has changed over the last 20 years (Eckhoff & Sagmo, 2016).

In recent times, focus on fuel efficiency has increased both in terms of environmental concerns, as well as related operational expenses. Initially, consumption was included as a measure of fuel efficiency, however, no significant effect was observed. Furthermore, correlation analysis indicates a high correlation between fuel consumption and speed knots. Thus, an index called fuel efficiency index (FEI) is constructed, effectively measuring fuel consumption on a "grams per tonnemile" basis (Adland et al. 2017). By excluding fuel consumption as a variable, correlation and modeling bias are reduced.

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

As the quality of ship is considered essential, a dummy variable for engine manufacturers is included in order to investigate differences in perceived quality and/or availability and cost of repairs. Further, more horsepower enhances the vessels' ability to keep optimal speeds in different conditions and is believed to have a positive effect on valuation. Additionally, higher speed is expected to have a positive impact on price, as it means a shorter time of transit. Fuel type is included as a dummy variable to investigate its effect on pricing.

Lastly, Handysize vessels carry numerous commodities with different storage attributes and the individual ship's characteristics in terms of the number of holds and related cargo gear are believed to impact price. For cargo gear, a dummy variable indicating if the vessel is equipped with the appropriate loading gear is included. According to Molland (2008), onboard high-capacity cargo handling cranes or derricks assure a speedy loading process, which reflects in a faster turnaround in port and, hence, more trips per year. As costs of cargo handling are excluded from voyage charter costs, appropriate loading gear is expected to yield higher revenue and, therefore, have a positive impact on the vessel price. Further, it is tricky to predict the impact of the number of holds as more holds results in less grain capacity per DWT, meaning it is a trade-off between economies of scale and diversification. Data for the number of hatches are also available but are omitted because of the high correlation with the number of holds (see Appendix A1).

3.2 Second-hand Transaction Data on Handysize Dry Bulkers

In the shipping industry today there is a general lack of standardization. More or less every vessel is unique, with its carrying capacity (DWT), cargo type, age, speed, country of origin and fuel consumption. All of these unique characteristics are expected to impact the price of a vessel in the second-hand market. Further, the second-hand market for ships generally has a low turnover, making it an illiquid market. The absence of standardization in an illiquid market entails the use of shipbrokers estimates instead of actual reported transactional data.

In this thesis, the empirical analysis focuses on the Handysize segment, which consists of bulk carriers with carrying capacity in the range 10,000 – 40,000 DWT (Clarksons Research, 2019). This segment, due to the size and onboard cranes, can call at a large number of ports around the world and transport a wide range of commodities such as coal, sugar, grain, alumina and steel products. Compared to larger segments, such as Pana- and Supramax, the operational flexibility and low investment costs make Handysize vessels the preferred vessel size for many small-scale ship owners.

According to Clarksons Research (2019), the global Handysize bulk carrier fleet consists of a total of 3,738 vessels, owned by 1,247 different ship owning companies, and averaging approximately 28,000 DWT. The ownership of the fleet is highly fragmented, with an average fleet size per ship owner of three vessels. Market concentration in terms of ownership by top-ten owners is also low, with 415 vessels, corresponding to 11.1% of the total fleet (Clarksons, 2019). This is an important observation for research purposes, as the impact of sellers and buyers in the second-hand market is expected to be low in comparison to earlier studies in more concentrated markets such as large bulkers and tankers (Adland et al., 2016a).

Our dataset is obtained from Clarksons Research (2019), which collects information from shipbrokers and industry sources. The dataset contains transaction data for second-hand Handysize bulker sales in the period from January 1, 1996, to September 30, 2019 - summing up to a total of 2,371 transactions. The number of transactions per year is highly varying with a low point of 33 transactions in 2014 and a top in 2009 with 145 transactions.

In the dataset, there are essentially two categories of information; first a detailed description of the unique vessel characteristics. This includes variables such as DWT, time of build, builder (yard), number of hatches and holds, gear characteristics, speed, engine manufacturer, main engine rpm, engine power (HP), fuel type and main fuel consumption. The second category of information includes characteristics related to the transaction, such as the date of sale, the price (in million USD) and information regarding the origin of both seller and buyer.

The original dataset is filtered for block sales, newbuilding resales, outliers, transactions from Africa (due to little data basis) and missing values in the decisive determinants. Further, the vessel's age at the sale, FEI and country indicator is derived. The complete input data contains a total number of 1,625 transactions with 11 price determinants presented in 3.1. Descriptive statistics of the complete input data are presented in Table 3.2.

Table 3.2: Descriptive Statistics of Handysize Dry Bulker Data

	No.	Mean	Median	Min	Max	SD
Saledate	1625	2007	2007	1996	2019	
Price	1625	7.03	5.90	0.60	20.75	4.44
Age	1625	18.39	19.19	0.23	40.74	7.11
DWT	1625	28774	28381	10106	42208	6308
Speed	1625	14	14	12	16	0.84
Horsepower	1625	8966	8640	3300	14000	1933
FEI	1625	2.662	2.48	1.406	4.477	0.63
No. Holds	1625	4.847	5	2	9	0.76

Country of build	No.	Percentage
Japan	1223	75.3 %
Europe	164	10.1 %
P.R.C	110	6.8 %
S. Korea	67	4.1 %
Asia other	33	2.0 %
America	28	1.7 %
Engine Manufacturer		
MAN B. & W.	703	43.3 %
Sulzer	455	28.0 %
Mitsubishi	402	24.7 %
Pielstick	47	2.9 %
Other	18	1.1 %
Gear Indicator		
Yes	1596	98.2 %
No	29	1.8 %
Fuel type		
HFO	1377	84.7 %
IFO/MDO	248	15.3 %

4 Methodology

The General Linear Model is a useful framework for comparing how certain variables affect different continuous variables and is widely recognized due to its simple fitting and easy interpretation. Nevertheless, the model has some limitations, as intrinsic non-linearities in the dataset may require semi-parametric modeling.

The Generalized Linear Models (GLMs) is a generalization of the General Linear Model and several concepts of linear modeling are applicable, however with some modifications (McCullagh & Nelder, 1989). In particular, the generalization differs in two major respects: (1) It allows for the distribution of the dependent variable to be (explicitly) non-normal and (2) includes a link function that connects the estimated fitted values to the linear combination of predictors (Wood, 2006). In short, the GLMs are an extension of the traditional linear model where fewer assumptions are necessary and could be structured as

$$g(\mu_i) = X_i \beta_i \tag{1}$$

where $\mu_i \equiv E(Y_i)$, $g(\cdot)$ is a smooth monotonic link function, X_i is the i^{th} row of a model matrix, X, and β is a vector of unknown parameters. The components of Y_i are independent variables and follow a distribution within the exponential family, $Y \sim exp(\mu, \sigma^2)$. Assuming a Gaussian distribution for the response variable, equal variance, σ^2, of all observations and a direct link between the linear predictor and the expected value, i.e. $X\beta = \mu$, the equation in (1) would represent a linear regression.

4.1 Generalized Additive Models

The available methods in Generalized Additive Models (GAMs) are applications of techniques developed by Hastie and Tibshirani (1990) and are a further extension of the GLMs. It is one of the main modeling tools for data analysis, due to the ability to efficiently combine different types of fixed, random and smooth terms in the linear prediction of a regression model (Wood, 2006).

The purpose of GAMs is to combine the GLMs with the notion of Additive Models, i.e. using an algorithm to fit a smooth curve to each variable, determine partial residuals from the fit and refit again (Schimek & Turlach, 2000). The advantage of the GAMs lies in the capitalization of the

GLMs' strength, without having to make assumptions regarding curve shape or a specific parametric function estimate, except the assumption of additive effects in the predictors (Koehn, 2008; Schimek & Turlach, 2000). The idea is to generalize data into smooth curves by local fitting. This is done by plotting the value of the dependent variable along a single independent variable and calculate a smooth curve passing through the data to achieve a parsimonious fit.

An approach for including non-linearity is to implement semi-parametric components for some or all explanatory variables that are expected to have a non-linear relation to the dependent variable. In Hastie and Tibshirani's (1990) original work, the idea of the model is to let the data dictate the relationship between the response variable and the explanatory variables, by making fewer assumptions. With increasing amounts of data available containing a large number of variables, a problem with obtaining reliable results may occur. This is referred to as the curse of dimensionality (Adland & Koekebakker, 2007). With non-parametric regression, estimates are averages of the dependent variable local to the point of which the regression function is to be estimated. With an increasing number of dimensions, i.e. conditioning variables, the number of local to the point averages decrease exponentially and, thus, vast quantities of data are needed. The curse of dimensionality may be overcome by applying a semi-parametric model. However, adding such flexibility comes at the cost of two necessities: (1) The question of how to represent the smooth term needs answering and (2) at what degree should the smoothing be set (Koehn, 2008).

Generally, a basic GAM would have the following construction

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

where $\mu_i \equiv E(Y_i)$ and Y_i is the response variable following a distribution within the exponential family. X_i^* is a row from the model matrix for any strictly parametric model component, θ is the corresponding parameter vector and f_i are smooth functions of the covariates. GAMs allow for a flexible specification of the dependence of the response on covariates. The model is specified in terms of smooth functions rather than detailed parametric relationships (Wood, 2006).

Further, consider the model

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

where y_i is the dependent variable, x_i is a covariate, f a smooth function and ε_ia random variable with $N(0, \sigma^2)$. For a standard technique such as OLS to be applicable, the smooth function needs to be composed in such a way that the model (3) becomes linear. Defining the space of functions of which f is an element, a "spline-basis" is applied and linear approximation is achievable (Wood, 2006). Consider

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

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

4.2 Smoothing

In the context of non-parametric regression techniques, a smoothing algorithm is a summary of the trend of a dependent variable as a function of one or more independent variables. Importantly, a smoother produces a less volatile estimate of the trend in a non-parametric nature, i.e. there is no underlying assumption of a rigid form of the dependency between the dependent and independent variables. Further, it allows for an approximation with a sum of functions and not just one unknown. Importantly, parsimony in the smooth curve is desired in GAM estimation. A univariate function can be represented using a cubic spline. This is a curve, made of subsections of cubic polynomials and joined together so that they are continuous in value as well as first and second derivatives (Wood, 2006; Hastie & Tibshirani, 1990). For conventional splines, the knots appear at a datum. However, in GAM estimation, the locations of the knots must be manually chosen. Typically, knots are either evenly distributed throughout the range of the observed values or places at quantiles of the distribution of unique x values (Wood, 2006).

The basis dimension is crucial for the degree of smoothing for a regression spline. One possibility in order to choose the appropriate degree of smoothing is to make use of hypothesis testing and decide the basis dimension by backward selection. However, this is a problematic approach, due to the fact that a model based on k − 1 evenly spaced knots will not generally be nested within a model based on k evenly spaced knots. Furthermore, one could start with a fine grid of knots and drop knots sequentially, but the resulting uneven knot spacing can itself lead to poor model performance. Also, for such regression spline models, model fit tends to depend strongly on the choice of knot locations (Wood, 2006).

There are alternatives to deciding the smoothness in changing the basis dimension. One is to fix the basis dimension at a size that is slightly larger than believed to be necessary and, further, control the smoothness-degree by adding a penalty for wiggliness (Koehn, 2008). In this approach, the trade-off between model fit and model smoothness is controlled by a smoothing parameter, λ. Whenever $\lambda \rightarrow \infty$, it implies a straight-line estimate for f, while $\lambda = 0$ results in an un-penalized regression spline estimate (Wood, 2006). The estimation of the smoothness' degree of freedom becomes a problem of estimating the smoothing parameter.

Consider,

$$\sum_{i=1}^{n}(y_i - f(x_i))^2 + \lambda \int_a^b [f''(x)]^2 dx \tag{5}$$

where λ is a fixed smoothing parameter concerning the unknown regression function that is found on the bases of the data and $a \leq x_1 \cdots x_n \leq b$. Smooth data is desired instead of interpolating and, thus, a cubic spline smoother is a solution to the optimization problem; among all functions $f(x_1)$ with the second continuous derivatives, find one that minimizes the penalized least square. The OLS method is represented in the first term of the equation (5), while the second term determines the wiggliness of the function as well as penalizing the curvature in the function. In a linear function application, the last term would be equal to 0.

Under the assumption that the basis dimension is sufficiently large in order to represent $f(x)$, the specific basis-choice and precise selection of knot locations do not influence the model fit in a significant manner. Rather, it is the value of the smoothing parameter that determines model flexibility.

The Generalized Cross-Validation (GCV) score is based on a leave-one-out cross-validation estimation process to estimate the mean square prediction error (Clark, 2013). The process is described by Wood (2006) as a model fitting process where a datum, y_i, is omitted, thus making it independent of the model fitted to the remaining data. Hence, the square error in predicting y_i is easily estimated. The process is further repeated by omitting all data to arrive at the best fit. When determining the smoothing parameter's specific nature, it is the GSV score that is minimized:

$$GCV = \frac{n * [scaled\ estimators]}{(n - [effective\ degrees\ of\ freedom] - [number\ of\ parametric\ terms])^2} \tag{6}$$

4.3 The Thin Plate Regression Spline

Several bases could be feasible for modeling purposes, such as p-splines and cubic-splines. However, such bases are open to some criticism. E.g. the necessity of knot location choice introduces an extra degree of subjectivity to the fit of the model. Furthermore, p-spline and cubic-spline bases are only useful for representing smooths of one predictor variable and it is somewhat unclear to what extent some bases are better or worse than others. An approach developed to comprehend with these impediments to some extent; Thin Plate Regression Splines (TPRS). TPRS produces knot-free bases, for smooths of any number, that in a certain limited sense are optimal. Additionally, this spline basis treats the wiggliness in all directions equally (Wood, 2006).

TPRS is constructed by defining exactly what is meant by smoothness, the exact weight to give conflicting goals of matching the data and making $\hat{f}$ smooth, and finding the function that best satisfies the smoothing objective. Further, it estimates the smooth function by finding the function $\hat{f}$ minimizing

$$||y - f||^2 + \lambda J_{md}(f) \tag{7}$$

where y is the vector of y_i data and $f = (f(x_1), f(x_2), \dots, f(x_n))^T$. $J_{md}(f)$ is a penalty function measuring the wiggliness of f, and λ is a smoothing parameter, controlling the trade-off between data fitting and smoothness of f. According to Wood (2006), TPRS are optimal in the sense that no smooth function will better minimize the function (7), thus making it close to an ideal smooth basis.

Worth keeping in mind is that the exact size of the basis dimension is not that critical. The basis dimension only sets an upper bound on the flexibility of a term and the actual effective degree of freedom is controlled by the smoothing parameter. In conclusion, assuming that it is not set to low, the model fit is rather insensitive to the basis dimension.

Lastly, GAMs has its modeling simplicities and it should be noticed that there are some shortcomings: Hypothesis testing is only approximate and a Bayesian approach seems to be required in order to estimate satisfactory interval estimations, thus p-values tend to be rather low due to their conditionality to the uncertain smoothing parameter (Koehn, 2008). Following, results based on significance must be interpreted with caution. As with other non- and semi-parametric techniques, theory and mathematical foundations for GAMs are complex and only selected relevant topics are presented in this paper. Hence, further theoretical background, implementation processes, and discussions can be found in the literature by Hastie and Tibshirani (1990), Wood (2006) and Koehn (2008).

5 Empirical Analysis

In this section, the empirical analysis and related results are presented. For empirical analysis, the smooth terms of the empirical model are based on the Thin Plate Regression Spline basis, with the selected default of dimension k = 30. This specification sets the upper limit in order to handle the identifiability constraint equal to k - 1. Through inspection of the optimal number of knots (see Appendix A3), this seems reasonable considering the bias-variance trade-off. Further, a gamma distribution on the dependent variable is applied. Since gamma distribution is assumed (see Appendix A2), there is no need for transformation of the dependent variable.

Model selection is initially based on the literature by Adland & Koekebakker (2007), Adland et al. (2018) and Adland & Koehn (2019). The correlation matrix (see Appendix A1) is used to investigate linear relationships and a data mining algorithm is applied to find the optimized GCV-score. Lastly, plots of residual deviation vs. theoretical quantiles, residuals vs. linear predictors, response vs. fitted values and residual frequency are studied in order to substantiate the choice of variables (see Appendix A3). The model optimization process yields the following model:

$$\begin{aligned} g(E(PRICE_i \mid .)) = \gamma_0 + s(SALEDATE_i, SALESAGE_i) + FEI_i + DWT_i + HP_i + SPEEDKNOTS_i \\ + NOHOLDS_i + I_i^{COUNTRYINDICATOR} + I_i^{ENGINEBRAND} + I_i^{VESSELFUELTYPE} + I_i^{GEARINDICATOR} \end{aligned}$$

To analyze the impact of the linear microeconomic determinants and how they differ depending on the state of the market (Stopford, 2009), the dataset is divided into three different periods: (1) 1996-2003, (2) 2004-2008, and (3) 2009-2019. The periods consist of a total of 521, 390, and 714 individual transactions, respectively. Furthermore, a standardized vessel is determined for further analysis of the non-linear term. The standard vessel is based on the entire dataset and has the attributes of the median - meaning a 28,000 DWT Japanese-built vessel with FEI of 2.5, five holds, and 8,640 horsepower. The dummy variables are set to their base.

5.1 Results

The model presented in this thesis is semi-parametric in nature; it includes both parametric and non-parametric variables. The model output is, therefore, represented in two panels: one in which the parametric regression is represented by point estimates along with the degree of significance, recognizable from traditional linear modeling. As for the non-parametric terms, there is an absence of point estimates. This is an important characteristic of GAMs; there are no coefficients for smoothed variables. The results of the second panel are, therefore, represented in terms of effective degrees of freedom (EDF), reflecting the degree of non-linearity. This is accompanied by the significance of the variables. The implication is that fitted smooth curves require plotting for interpretation purposes. Table 5.1 presents the model results throughout the periods.

In general, the developed model yields a satisfactory coefficient of determination, especially when considering major macroeconomic factors such as rates, newbuild price, and scrap price are not accounted for. Noticeably, the achieved R squared is considerably lower in the years between 2004 and 2008. This is expected due to the increased price volatility in this period, particularly as our time parameter is yearly based. Thus, the predicted price volatility of the model is smaller. This is reinforced by a lack of transactions for vessels under 5-years-old.

Before looking at the non-linear relationship between the time of sale and the vessel's age at the sale, we provide a brief insight into the linear microeconomic determinants of the model. As per expectations, the size of the vessels (DWT) is one of the most decisive determinants, with a positive sign of the point estimate. The determinant is significant at the 0.1% level in all periods. Other noticeable effects are the significant discount in value for Chinese-built vessels. Relative to a Japanese-built ship, our results indicate a discount of 0.9 million dollars with 0.1% significance over the entire period. When considering the different periods, the discount is only significant after 2008. However, this could be related to a bias in the data due to a lack of transactions for vessels built in China prior to 2008. Some interesting insight is to be found in the coefficients of the dummy variable for the engine manufacturer. At the 5% significance level, it may seem that Mitsubishi has made progress to receive a price premium for its engines relative to MAN B. & W. in the period between 2009 and 2019. Further, both Pielstick and Sulzer come at a discount between 2004 and 2008, this is significant at the 1% level.

Prior to the analysis, we expected that certain ship-specific factors would have increased significance post-2008. Inter alia, we expected FEI to have a negative significant impact. However, our results indicate no such effect. This may be due to several factors. Adland et al. (2017), for instance, state that, in the Capesize and Panamax segments, only 14-27% of the fuel cost savings are shared with ship owners through higher timecharter rates. It could also be that FEI has a non-linear effect on vessel value and that including a smoother may yield greater significance.

Table 5.1: Regression results from GAM for periods 1996-2003, 2004-2008 and 2009-2019.

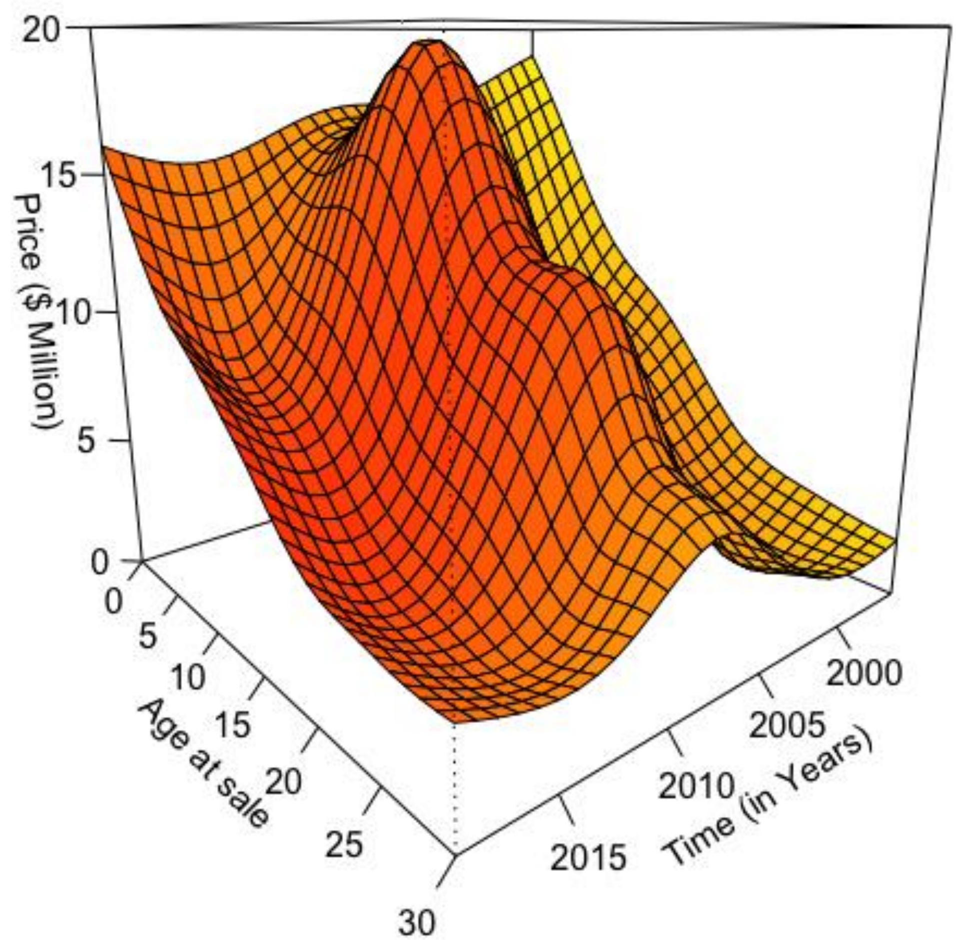
	1996-2003		2004-2008		2009-2019		1996-2019	
Parametric Coeff.	**Estimate**	**Sig.**	**Estimate**	**Sig.**	**Estimate**	**Sig.**	**Estimate**	**Sig.**
(Intercept)	2.995	***	2.680		1.669		3.578	***
DWT (in Thousands)	0.039	***	0.201	***	0.112	***	0.082	***
Country of build other than Japan								
America	-0.687	**	-1.732	**	0.249		-0.562	*
Asia Other	-0.378		-0.684		0.003		-0.253	
Europe	-0.194		-0.026		-0.416	-	-0.254	-
P.R.C	-0.072	-	0.255		-0.699	***	-0.930	***
S. Korea	-0.895	-	-0.250		-0.256		-2.557	
Enginebrand other than MAN B&W								
Mitsubishi	-0.071	-	0.236		0.259	*	0.048	
Pielstick	-0.052		-1.235	**	0.123		-0.135	
Sulzer	-0.069	-	-0.087	**	-0.053		-0.102	
Other	-0.169		2.921		0.550		0.213	
Speedknots	0.079		0.039		0.104	-	0.012	
FEI	-0.142		0.219		0.077		-0.045	
HP (in Thousands)	0.042		-0.050		-0.024		-0.023	
Noholds	-0.003		0.002		0.225	*	0.064	
Vesselfueltype IFO / MDO	-0.114		0.817		0.031		0.206	
Gearindicator Y	-0.036		0.779		-0.530		0.103	
Smooth Coeff.	**EDF.**	Sig.	**EDF.**	Sig.	**EDF.**	Sig.	**EDF.**	Sig.
s(Saledate, Saleage)	24.15	***	12.77	***	27.30	***	28.64	***
N	521		390		714		1625	
R-sq (adj)	0.826		0.657		0.836		0.708	
GCV	0.072		0.107		0.066		0.119	

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

5.2 Effects of Time and Vessel Age at Sale

Further, we present the results from the smoothed interaction term of our model and the valuable insights gained by the non-parametric feature of GAMs. We look at the results estimated across the entire dataset in order to understand the long-term effects and to capture the major market cycles. Figure 5.1 illustrates the joint non-linear effect of time and vessel age on the second-hand value. Our results indicate a clear non-linear relationship in the interaction term, indicated by the estimated EDF of 28.64, as well as in the contour plot. This is significant at the 0.1% level. The model and associated visualization provide the opportunity to analyze the price development for second-hand vessels at a predetermined age or time. Furthermore, the value development for a ship built in a specific year can be studied, which has not been subject to previous research.

Figure 5.1: Visualization of Model Output for the Interaction Between Salesdate and Salesage.



The model output can further be examined from different perspectives. First, we look at a model where vessel age is kept constant throughout, thus, examining the isolated impact of time on vessel value. As seen in Figure 5.2, the major shipping cycles are recognized with a clear peak in 2008 before the onset of the global financial crisis. In 1996, at the start of the analyzed dataset, the shipping sector appears to be in the recovery phase, before moving into the trough in the years before the peak. Further, we see the collapse post-2008 and the sector moving into a new cycle (Stopford, 2008).

From the model output, it is clear that the standard vessel's price volatility differs based on age. The 5- and 25-year-old vessels are less volatile in price compared to the 10-, 15- and 20-year-old vessels. This can be explained by the fact that the newbuild- and scrap prices act as an upper- and lower limit to the valuation, respectively. According to Adland and Koekebakker (2007), an important determinant for the valuation of a vessel is its age, hence the 25-year-old vessel's remaining life expectancy is taken into account when estimating its value. Note that there is no reason for a 10-year-old vessel to be valued higher than a 5-year-old. After studying the transaction data, we find that this is a result of bias in the sample as only 43 transactions for vessels under 5 years were recorded since 1996. Also, none of these transactions happened between 2004 and 2008. Thus, it is likely that the valuation of vessels younger than 10 years are biased downwards. In addition, to highlight the price development in percentage, the log-transformed basis of the model is added in Figure 5.3. Noticeably, the 20- and 25-year-old vessels have a larger percentage increase in price in the period 2000-2008. Following the global financial crisis, the younger vessels have a lower percentage decrease in value, but a larger capital loss.

Figure 5.2: Model Output for the Vessel Value Across Time, Constant Vessel Age.

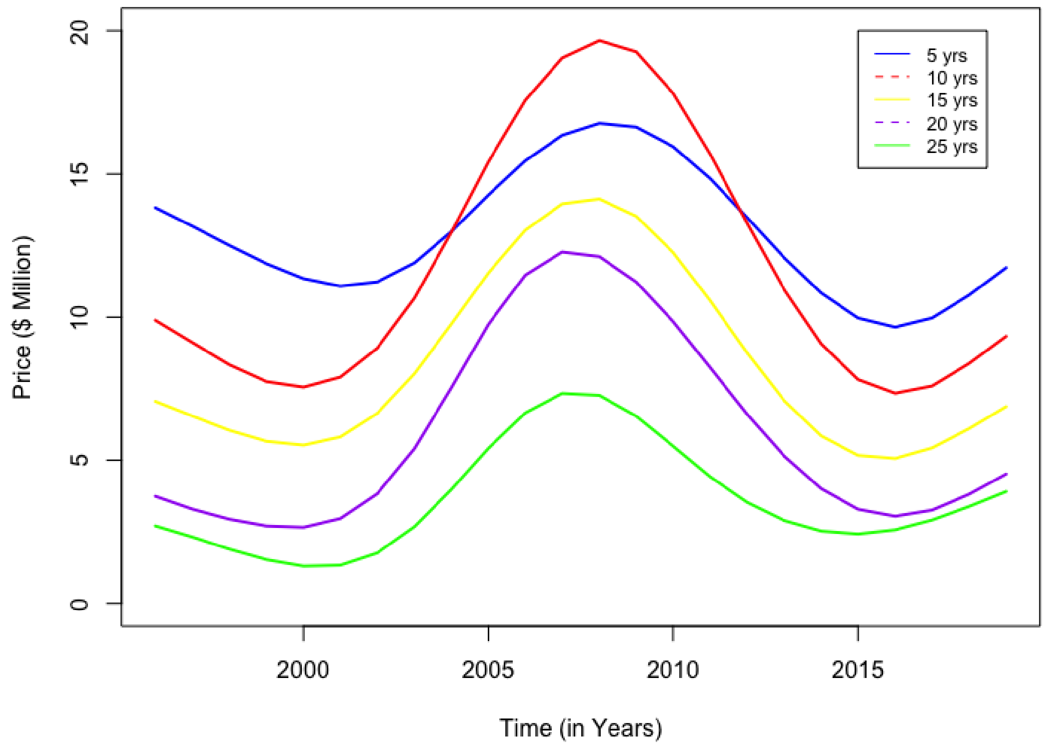
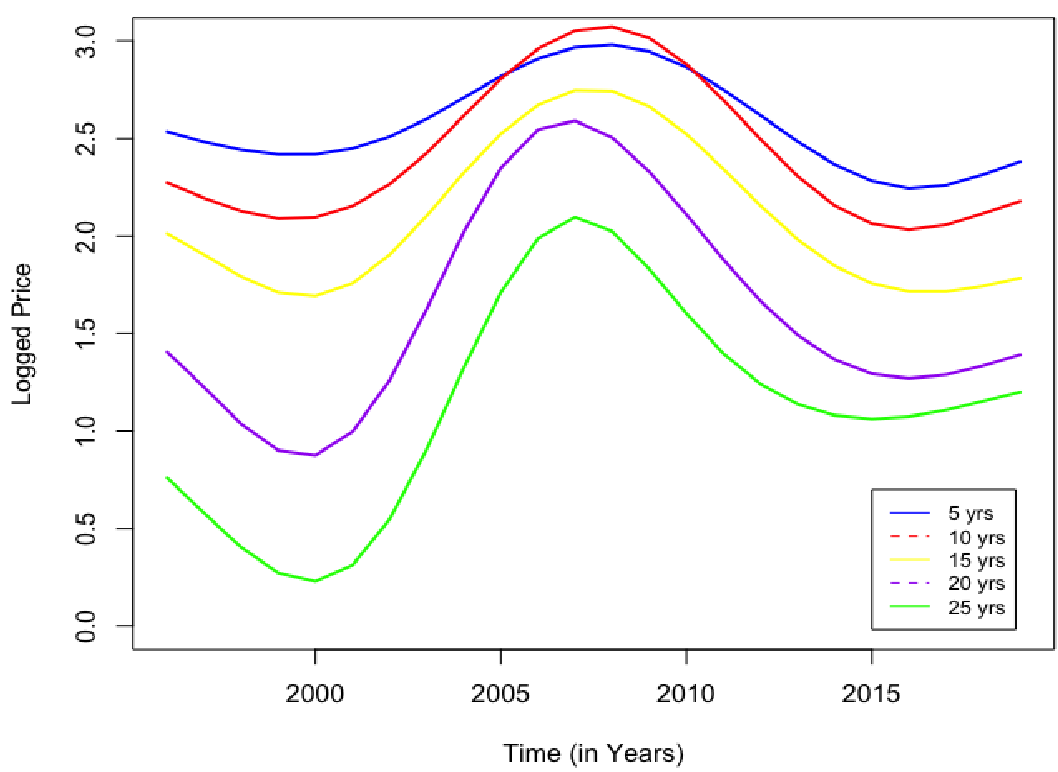


Figure 5.3: Logged Model Output the Vessel Value Across Time, Constant Vessel Age.



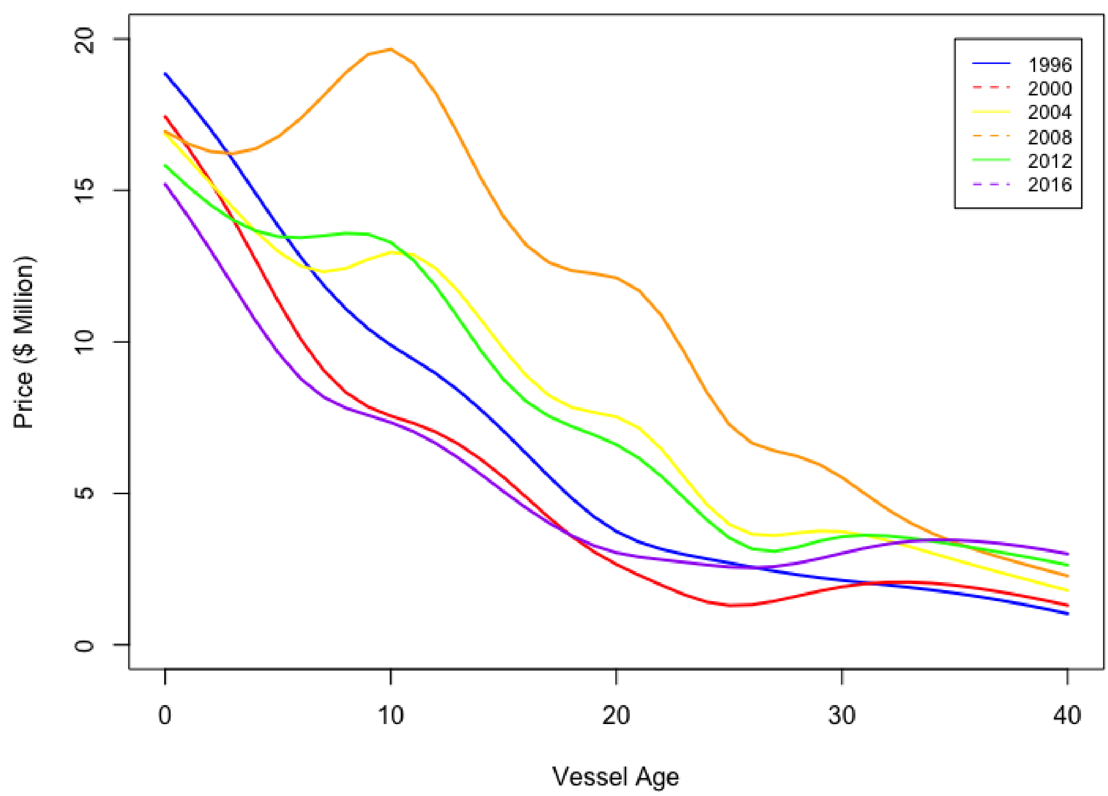
Next, we look at a model where the time parameter is kept constant, thus examining the isolated effect of vessel age on its value. In other words, the depreciation curve is analyzed, as presented in Figure 5.4. Previous studies by Adland and Koekebakker (2007) and Adland and Koehn (2019) on vessel valuation have found a decisive non-linear application to the depreciation curve. In addition to considering the average depreciation for the Handysize drybulk segment, it is here further studied in accordance with Stopford's four stages of the shipping cycle (2009).

While the studies (Adland & Koekebakker, 2007; Adland & Koehn, 2019) finds a slightly concave depreciation curve, particularly for younger vessels. Our results indicate a difference in this relationship in a booming market relative to a market in recovery. In a recovering market, such as in 2000, the average depreciation curve is in major terms convex. As for a booming market, such as in 2008, second-hand vessel value depreciates in a slightly concave manner. The differences in depreciation can be explained by the term structure of freight rates related to the different states of the market. The timecharter rates are in general less volatile than spot rates and, therefore, do not reflect the highs and lows of the spot market (Kavussanos, 2003). Thus, in a booming market, the freight market is generally in backwardation (spot rates higher than forward rates), while in recovery the market is generally in contango (spot rates lower than forward rates) (Ko, 2013). According to Alizadeh and Nomikos (2011), the timecharter rates are a form of forward freight rates and a change in future earnings has a bigger impact on the net present value for an older ship. Thus, when earnings are expected to decrease (increase), the effect on depreciation is concave (convex). In accordance with Adland and Koekebakker's (2007) findings, we see evidence that second-hand tonnage value eventually will converge towards scrap price. It is also observable that a booming market results in prolonged life expectancy, as higher earnings cause higher net present values and, thus, prolong the time until vessel value converges towards the scrap price.

Another observation is the price for 10-year-old vessels at the peak in 2008, in contrast to younger vessels. As previously mentioned, there is no reason that a 10-year-old vessel should be valued more highly than a younger one and this is a result of bias in the sample. On the other hand, it is well-known in the literature that second-hand vessel prices can surpass newbuilding prices in a booming market. This can be explained by the stickiness of newbuilding prices (Beenstock, 1985). As a result of the time-varying delivering lag, the volatility of newbuilding prices is relatively low compared to second-hand prices. This is correlated with the alternative cost of operating in the freight market (Adland & Jia, 2015). In booming markets, ship owners want their share of the increased earnings, and the time of delivery is distinctively shorter for second-hand vessels.

Additionally, it is worth noticing an individual vessel's depreciation curve's tendency to fluctuate between concavity and convexity. This trend can be somewhat explained by the required hull and machinery intermediate- and special surveys. According to IACS (2019), an intermediate survey is to be carried out within the window of three months before the second to three months after the third anniversary. The special survey is to be carried out at five-year intervals. Apostolidis et al. (2013) find that these dry-dockings are a significant determinant of profitability and that performance could be improved by proper maintenance cost monitoring. This can explain the development of the depreciation curve prior to surveys, as the cost of maintenance is a substantial part of the net present value and post-survey this cost is accounted for. Knowing that a vessel is approaching scrap at the later special surveys, it is reasonable to believe that the surveys become more expensive both in general and relative to remaining earnings, making the effect more distinct.

Figure 5.4: Model Output for the Vessel Value Given Age in Different Parts of the Shipping Cycle.



Our further contribution to the research on vessel valuation is the model's ability to follow the price development of a predetermined ship from its year of build. This is a valuable contribution, as today's time series present the average price for certain ships at a constant age, as well as being subject to brokers' estimates. The model applies the attributes of the standard vessel, with both year and age increasing relatively throughout the years. The valuation will, therefore, be influenced both by the major shipping cycles and the vessels' depreciation. Figure 5.5 and Table 5.2 presents the model output for vessels built between 1988 and 2012, at a four-year interval.

Results from the vessel value development, suggests some distinctive differences for older ships, i.e. 1988-, 1992- and 1996 built, compared to newer 2000- and 2004 built ships. During the booming market prior to 2008, older ships have experienced a larger percentage return on asset value, relative to newer. Investing in an older built ship would, thus, yield a significantly higher return on investment, not accounting for earnings from operations. In the years following 2008, the entire fleet experienced major capital losses, on average around 80 percent of asset value. Furthermore, we see that older vessels, in general, have higher volatility in their year-on-year (YoY) return through the period of analyzation. Consequently, we argue that investing in an older vessel has greater upside potential, but comes with more capital risk. Note that the results are subject to bias due to the mentioned lack of transactions for vessels under 5-years-old.

Lastly, the results presented in this paper may be of interest to market players, such as short term asset- and long term industrial players. Asset players seek a return in buying and selling second-hand tonnage and tend to enter and exit the market within one shipping cycle. On the other hand, industrial players gain their income in the freight market. Short term investors are in general accepting higher risk, as they seek higher capital returns. Long term investors have a lower yield requirement and would not be interested in taking unnecessary risk. Furthermore, industrial investors are in general more concerned with the condition of the freight market and seek revenue through operations. Also, industrial players are more likely to operate their vessels until they reach scrap value.

Based on these findings, we argue that asset players should seek older, more volatile, vessels, as these potentially yield a higher return on invested capital. On the contrary, industrial players should pursue the acquisition of newer tonnage. By running a linear regression on the raw data analyzing the relationship between FEI and year of build, a significant negative relationship is found, i.e. newer vessels are more fuel-efficient. Adland et al. (2017) further state that fuel cost savings results in higher timecharter rates, even though the savings are shared with the charterers. To summarize, newer vessels have characteristics that are expected to increase revenue in the freight market and should be of interest to industrial players.

Figure 5.5: Illustration of Value Development for Vessels Built Between 1988 and 2012

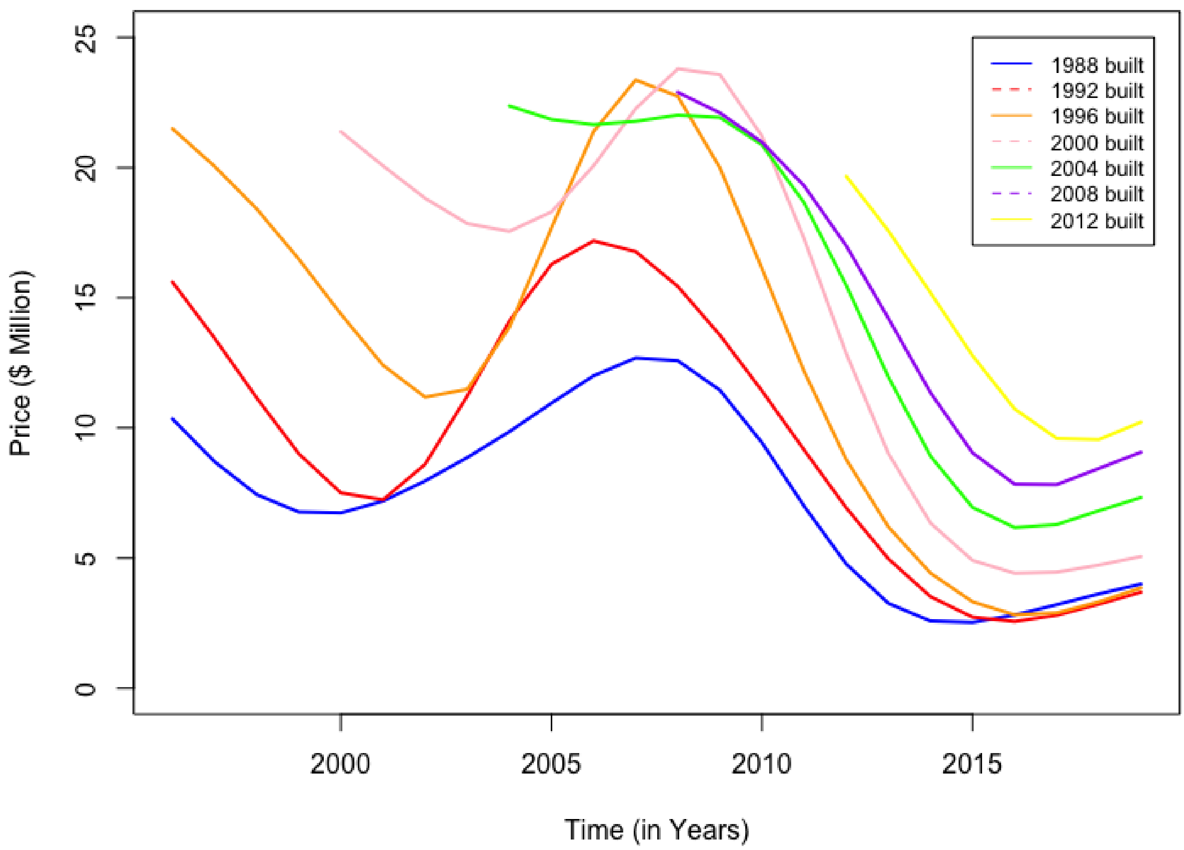


Table 5.2: Model Output for the Value Development for Vessels Built Between 1988 and 2012.

Year	1988 built	YoY Return	1992 built	YoY Return	1996 built	YoY Return	2000 built	YoY Return	2004 built	YoY Return	2008 built	YoY Return	2012 built	YoY Return
1996	10.35		15.60		21.49									
1997	8.70	-15.89 %	13.45	-13.83 %	20.05	-6.72 %								
1998	7.44	-14.50 %	11.15	-17.05 %	18.41	-8.16 %								
1999	6.77	-9.02 %	9.01	-19.25 %	16.48	-10.48 %								
2000	6.73	-0.52 %	7.50	-16.69 %	14.37	-12.82 %	21.38							
2001	7.19	6.76 %	7.24	-3.51 %	12.41	-13.66 %	20.07	-6.12 %						
2002	7.96	10.77 %	8.60	18.83 %	11.18	-9.89 %	18.82	-6.22 %						
2003	8.86	11.27 %	11.23	30.49 %	11.48	2.70 %	17.85	-5.15 %						
2004	9.85	11.21 %	14.10	25.58 %	13.86	20.69 %	17.55	-1.67 %	22.37					
2005	10.95	11.15 %	16.28	15.51 %	17.68	27.57 %	18.30	4.24 %	21.85	-2.32 %				
2006	12.00	9.58 %	17.18	5.52 %	21.38	20.95 %	20.07	9.68 %	21.65	-0.90 %				
2007	12.68	5.69 %	16.77	-2.37 %	23.36	9.25 %	22.26	10.94 %	21.78	0.57 %				
2008	12.58	-0.81 %	15.44	-7.98 %	22.74	-2.67 %	23.79	6.87 %	22.02	1.10 %	22.89			
2009	11.45	-9.00 %	13.56	-12.12 %	19.99	-12.09 %	23.57	-0.92 %	21.92	-0.42 %	22.10	-3.45 %		
2010	9.43	-17.62 %	11.42	-15.80 %	16.12	-19.37 %	21.21	-10.04 %	20.87	-4.80 %	20.97	-5.10 %		
2011	6.99	-25.84 %	9.15	-19.87 %	12.17	-24.47 %	17.27	-18.58 %	18.65	-10.66 %	19.30	-7.98 %		
2012	4.77	-31.76 %	6.93	-24.27 %	8.79	-27.83 %	12.86	-25.55 %	15.48	-17.00 %	17.00	-11.94 %	19.67	
2013	3.26	-31.62 %	4.97	-28.23 %	6.20	-29.41 %	9.02	-29.84 %	11.96	-22.71 %	14.21	-16.37 %	17.57	-10.64 %
2014	2.58	-20.91 %	3.52	-29.25 %	4.42	-28.67 %	6.35	-29.62 %	8.92	-25.43 %	11.36	-20.10 %	15.19	-13.54 %
2015	2.53	-1.97 %	2.73	-22.37 %	3.32	-24.95 %	4.91	-22.67 %	6.94	-22.13 %	9.04	-20.40 %	12.77	-15.93 %
2016	2.81	11.16 %	2.56	-6.21 %	2.82	-15.02 %	4.42	-10.03 %	6.17	-11.11 %	7.84	-13.30 %	10.73	-16.02 %
2017	3.21	14.15 %	2.80	9.40 %	2.89	2.56 %	4.46	0.91 %	6.29	1.85 %	7.82	-0.20 %	9.60	-10.55 %
2018	3.62	12.84 %	3.23	15.22 %	3.32	14.70 %	4.73	6.13 %	6.82	8.44 %	8.44	7.90 %	9.55	-0.44 %
2019	4.00	10.39 %	3.69	14.20 %	3.85	16.02 %	5.06	7.00 %	7.33	7.45 %	9.07	7.40 %	10.22	6.98 %
St. Dev	3.39	15.28 %	5.01	17.90 %	7.04	17.26 %	7.33	13.31 %	6.96	11.02 %	5.98	9.96 %	3.90	8.66 %

.

5.3 Model Uncertainty

There are uncertainties in our research that should be highlighted. Firstly, the developed model operates with yearly data in the time parameter. Due to this, one might suspect that the model will be unable to capture some of the price volatility in the second-hand market. Furthermore, this paper does not account for operational earnings and is merely concerned with asset valuation. If the intention of the valuation is to maximize returns, one should consider both gains on assets and market earnings. The focus of this paper is essentially the non-linear interaction between the age of a vessel and the time of a transaction. Therefore, the relationships of the other microeconomic determinants are assumed linear. It is reasonable to believe that non-linearity in some of these price determinants could increase explanatory power.

Since the second-hand transaction data is built on reported fixture data, parameters such as main fuel consumption and speed may differ for actual operational numbers. There is also a general lack of recorded data regarding timecharter- and spot rates for the Handysize drybulk sector, a major price determinant as this reflects the ships' earnings. Additionally, transaction-, finance- and opportunity costs are not considered. Lastly, one might suspect that some ship owners are more experienced in the sales and purchase market, resulting in premiums and discounts for certain transactions. Besides, there is a lack of information regarding technical conditions for ships at the transaction time, causing investor badwill.

6 Concluding Remarks

Throughout this report, we have developed a comprehensive General Additive Model for the estimation of second-hand prices in the Handysize drybulk sector. Previous research by Adland and Koehn (2019) has shown that semi-parametric models provide an appropriate framework for the valuation of highly heterogeneous assets, as is the case in the bulker sector. This paper contributes to the existing research, by analyzing the joint non-linear effect of time of the age of a vessel and the time of a transaction, hence, making it feasible to follow the price development of a predetermined ship from its year of build. Accordingly, the vessel valuation model can account for the influence of both the major shipping cycles and the vessel depreciation, as well as their isolated effect.

From the analysis of the microeconomic determinants, we find a significant positive effect of size. More, a ship built in China trades at a discount relative to a Japanese built ship. These results are in accordance with Adland et al. (2018). Contrary to our expectations and previous research findings (Adland et al., 2018), we could not prove that FEI had a negative impact post-2008.

Further, results from the smoothed interaction term of our model yield a non-linear relationship between the time of sale and age of the vessels at the sale. When analyzing the isolated effect of time, i.e. keeping age constants, we find that the volatility of vessel value increases with age in terms of percentage. Additionally, findings suggest that the depreciation curve differs on the state of the market conditions. In a booming (recovering) market, the average depreciation curve is in major terms concave (convex). Based on the combined effect of sales date and depreciation, we argue that asset players, in contrary to industrial, should seek older, more volatile, vessels. These vessels could potentially yield a higher return on invested capital. On the other hand, newer vessels could have characteristics that are expected to increase revenue in the freight market and should be of interest to industrial players.

Finally, for further research, we suggest two topics that would continue to development of the research on second-hand valuation modeling. Firstly, as we have developed a model examining structural shifts and changes across time, it could be interesting to implement a time parameter based on monthly- or quarterly observations instead of yearly. By doing this the model should be able to obtain more of the observed volatility in the dataset, but due to the average sales rate, other uncertainties could occur. As new environmental regulations are being implemented in 2020, it could be interesting to include new microeconomic determinants. An example is scrubber installation. Secondly, it could be interesting to develop a trading strategy based on the valuation model to investigate its ability to yield excess returns. Examples of trading strategies could be moving average or filter-rules.

7 References

Adland, R., Alger, H., Banyte, J., & Jia, H. (2017). Does fuel efficiency pay? Empirical evidence from the drybulk timecharter market revisited. *Transportation Research Part A: Policy and Practice,* 95. 10.1016/j.tra.2016.11.007.

Adland, R., Cariou, P., & Wolff, F. C. (2016). The influence of charterers and owners on bulk shipping freight rates. *Transportation Research Part E: Logistics and Transportation Review*. 86. 69-82.

Adland, R., Cariou, P., & Wolff, F. C. (2017). What makes a freight market index? An empirical analysis of vessel fixtures in the offshore market. *Transportation Research Part E: Logistics and Transportation Review*, 104, 150-1.

Adland, R., Cariou, P., & Wolff, F. C. (2018). Does energy efficiency affect ship values in the second-hand market? *Transportation Research Part A: Policy and Practice*, 111, 347-359.

Adland, R., & Jia, H. (2015). Shipping market integration – the case of sticky newbuilding prices, *Maritime Economics and Logistics*, 17(4), 389 – 398.

Adland R., & Koehn S. (2019). Semiparametric Valuation of Heterogeneous Assets. In: Mathew J., Lim C., Ma L., Sands D., Cholette M., Borghesani P. (eds) Asset Intelligence through Integration and Interoperability and Contemporary Vibration Engineering Technologies. Lecture Notes in Mechanical Engineering. Springer, Cham.

Adland, R., & Koekebakker, S. (2004). Market Efficiency in The Second-hand Market for Bulk Ships. *Martime Economics & Logistics 6,* 1-15.

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, N. (2006). Trading strategies in the market for tankers. *Maritime Policy & Management. 33.* 119-140.

Alizadeh, A., & Nomikos, N. (2011). Dynamics of the Term Structure and Volatility of Shipping Freight Rates. *Journal of Transport Economics and Policy (JTEP).* 45. 105-128.

Apostolidis, A., Merikas, A. & Merika, A. (2013). The Modelling of Maintenance Cost: The Case of Container-Ships in Dry-Dock. *The Journal of Computational Optimization in Economics and Finance. 5.*

Beenstock, M. (1985). A theory of Ship Prices. *Maritime Policy & Management 12*, 215-225.

Beenstock, M., & Vergottis, A. (1989a). An econometric Model of the World Shipping Market for Dry Cargo, Freight and Shipping. *Applied Economics, 339-356.*

Beenstock, M., & Vergottis, A. (1989b). An econometric Model of the World Tanker Market. *Journal of Transport Economics and Policy 23*, 263-280.

Beenstock, M., & Vergottis, A. (1992). The interdependence between the Dry Cargo and Tanker Markets. *Logistics and Transportation Review 29(1)*, 3-38.

Byoung Wook, Ko. (2013). Analysis of Term Structure in Dry Bulk Freight Market. The Asian *Journal of Shipping and Logistics*. 29. 1–22.

Campbell, J. Y., & Schiller, R. F. (1988). Stock prices, earnings and expected dividends. *Journal of Finance 43*, 661-676.

Charemza, W., & Gronicki, M. (1981, 01 01). An econometric model of world shipping and shipbuilding. *Maritime policy and management 8, 21-30.* Available at: http://www.ics-shipping.org/shipping-facts/shipping-and-world-trade.

Clark, M. (2013). Generalized Additive Models. Chicago: University of Notre Dame. Available at: https://m-clark.github.io/docs/GAMS.pdf.

Clarkson Research, (2019). Shipping Intelligence Network database. Available at: www.clarksons.net.

Eckhoff J., & Sagmo K. (2016). Microeconomic Determinants of Secondhand Vessel Prices. Norwegian School of Economics.

Glen, D. R. (1997). The market for second-hand ships: Further results on efficiency using cointegration analysis. *Maritime Policy & Management*, *245-260.*

Hale, C., & Vanags, A. (1992). The market for second hand ships: Some results of efficiency using cointegration analysis. *Maritime Policy and Management 19(1), 31-39.*

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

International Association of Classification Societies. (2019). Classification Societies - what, why and how? Available at: http://www.iacs.org.uk/media/3785/iacs-class-what-why-how.pdf.

Kavussanos, M. G. (2004). Time Varying Risks Among Segments of the Tanker Freight Markets. *Maritime Economics & Logistics 5(3): 227-250.*

Kavussanos, M. G., & Alizadeh, A. H. (2002). Efficient pricing of ships in the dry bulk sector of the shipping industry. *Maritime Policy & Management 29:3, 303-330.*

Keller, G. (2009). Statistics for Management and Economics, 8th ed, Cengage Learning.

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

Løvås, G. (2013). Statistikk for universiteter og høgskoler. 3rd edition. Universitetsforlaget.

McCullagh P., & Nelder J.A. (1989). Generalized Linear Models (Chapman & Hall/Monographs on Statistics and Applied Probability 37). 2nd edition. Chapman and Hall/CRC.

Molland, A. F. (2008). The maritime engineering reference book: A guide to ship design, construction and operation. Amsterdam: Butterworth-Heinemann.

Pruyn, J. F., Voorde, E. V., & Meersman, H. (2011). Second hand vessel value estimation in maritime economics: A review of the past 20 years and the proposal of an elementary method. *Maritime Economics & Logistics*, vol 55, No. 3, 213-236.

Riise R., & Rødde M. (2014). The Relationship Between Ship Specifications and Time Charter Rates. Norwegian School of Economics.

Schimek, M., & Turlach, B. (2000). Additive and Generalized Additive Models: a Survey. Austria: Karl Franzens University of Graz & Australia: University of Adelaide. Available at: https://edoc.hu-berlin.de/bitstream/handle/18452/4423/97.pdf?sequence=1&isAllowed=y.

Soedal, S., Koekebakker, S., & Adland, R. (2009). Value based trading of real assets in shipping under stochastic freight rates. *Applied Economics*, 2793-2807.

Stopford, Martin. (2009). Maritime Economics 3e. Abingdon, Oxon: Routledge.

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

Strandenes, S. (1986). Norship: a simulation model of markets in bulk shipping. Bergen, Norway: Norwegian School of Economics.

Tsalokis, S. D., Cridland, C., & Haralambides, H. E. (2003). Econometric Modelling of Second-hand Ship Prices. *Maritime Economics and Logistics 5,* 347-377.

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

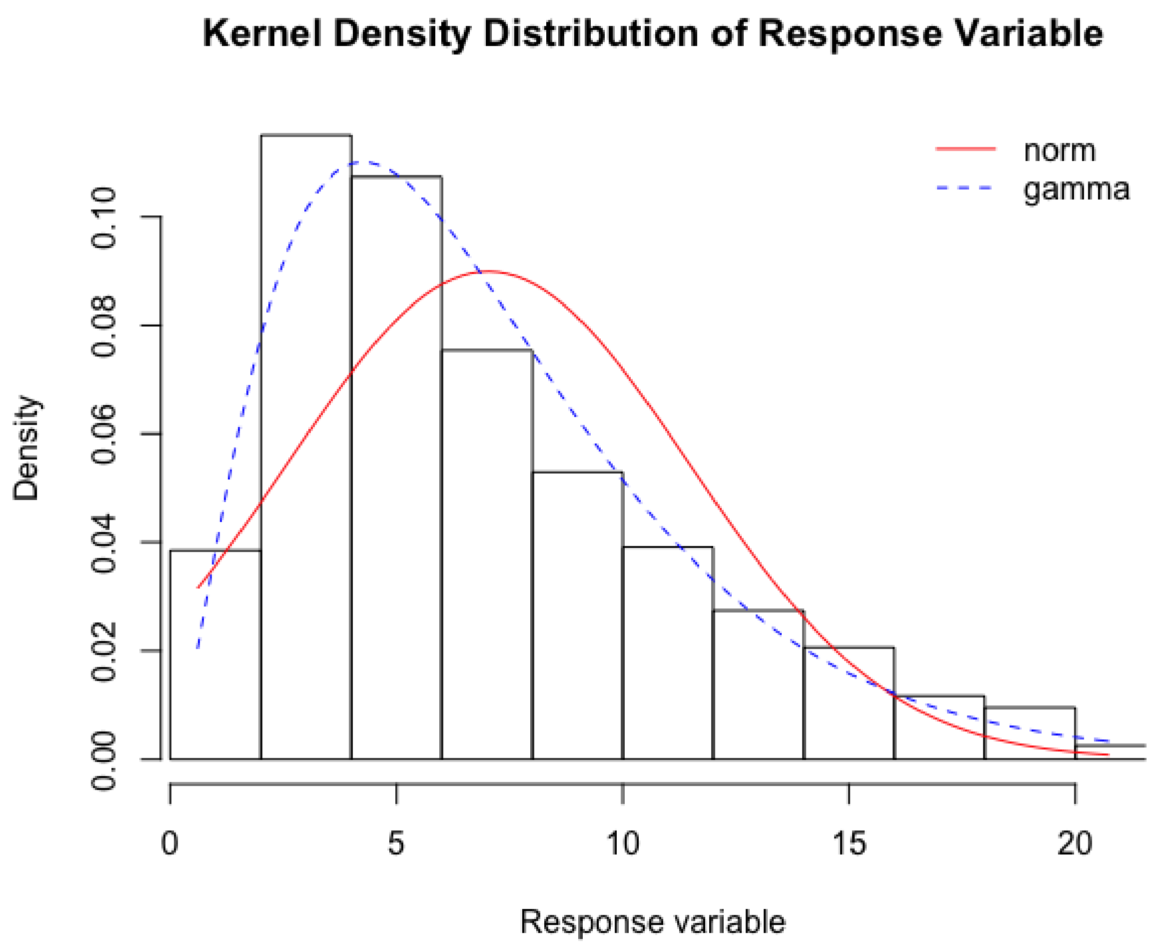
8 Appendices

A1 Correlation matrix

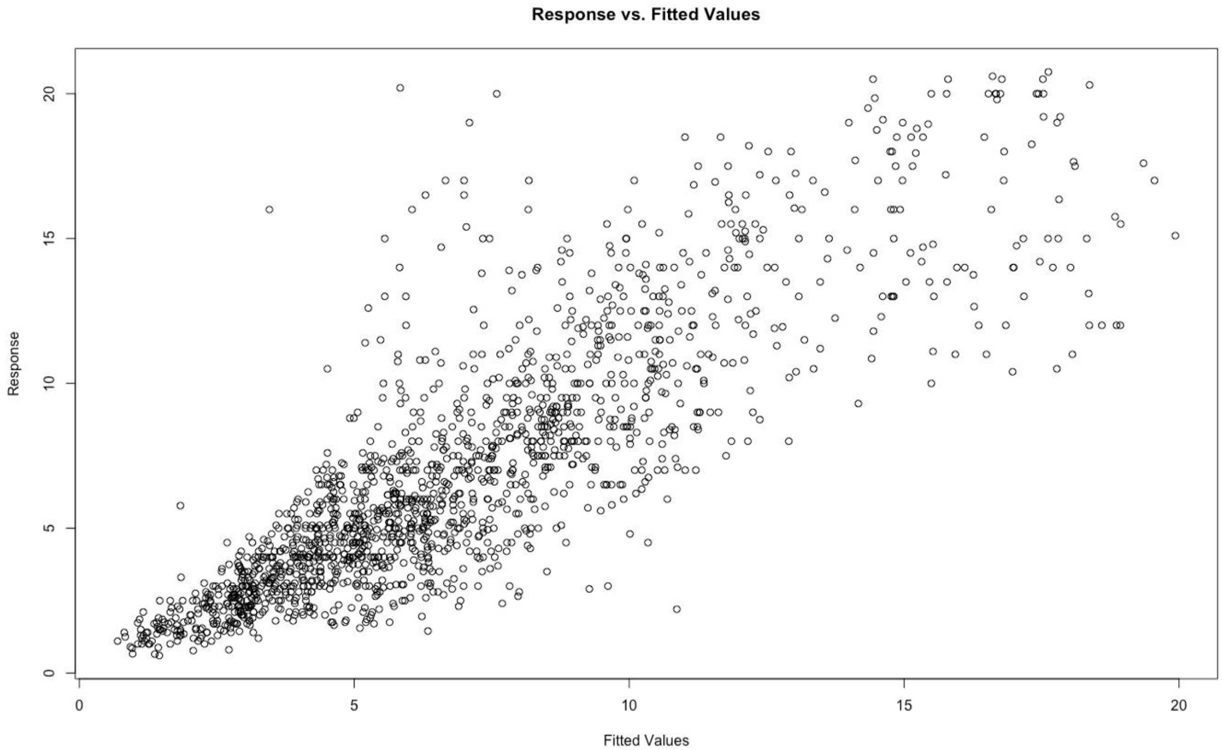
	Saledate	Salesage	Dwt	Countryindicator	Price	Speedknots	Mainfuelconsumptiontpd	HP	Vesselfueltype	Enginebrand	Noholds	Nohatches	Gearindicator	FEI
Saledate														
Salesage	-0.125***													
Dwt	0.078***	-0.022												
Countryindicator	0.092***	-0.079***	0.059*											
Price	0.106***	-0.492***	0.098***	-0.003										
Speedknots	-0.128***	0.125***	0.129***	0.079***	-0.073**									
Mainfuelconsumptiontpd	-0.234***	0.291***	0.333***	0.149***	-0.218***	0.521***								
HP	-0.149***	0.218***	0.468***	0.252***	-0.139***	0.389***	0.727***							
Vesselfueltype	0.228***	-0.211***	0.032	0.020	0.146***	-0.132***	-0.121***	-0.083***						
Enginebrand	-0.287***	0.320***	0.219***	-0.054*	-0.175***	0.147***	0.271***	0.224***	-0.107***					
Noholds	-0.028	0.090***	0.596***	0.249***	-0.033	0.222***	0.438***	0.540***	-0.026	0.216***				
Nohatches	-0.081***	0.128***	0.426***	0.244***	-0.080***	0.218***	0.421***	0.496***	-0.034	0.174***	0.800***			
Gearindicator	0.083***	-0.089***	0.055*	-0.125***	0.081***	0.039	-0.031	-0.032	0.049*	-0.038	-0.116***	-0.099***		
FEI	-0.151***	0.171***	-0.383***	0.043	-0.186***	0.043	0.522***	0.086***	-0.040	0.027	-0.143***	-0.056*	-0.059*	

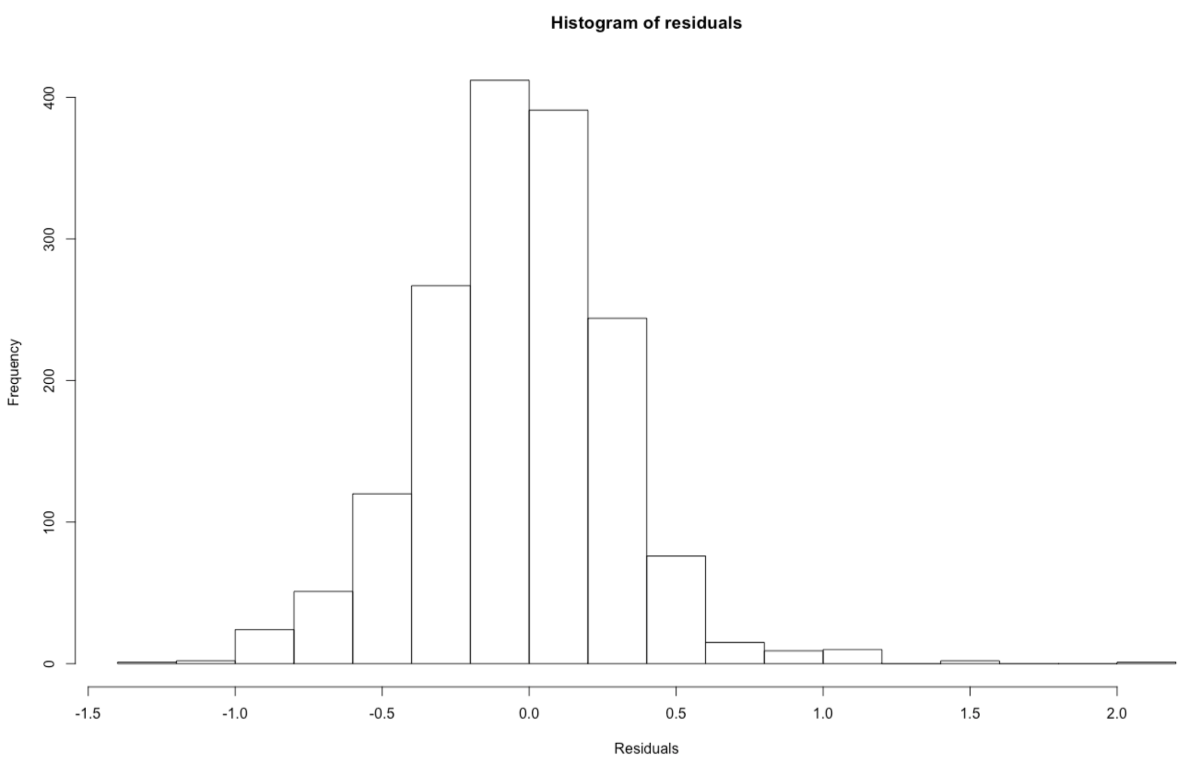
Computed correlation used pearson-method with listwise-deletion.

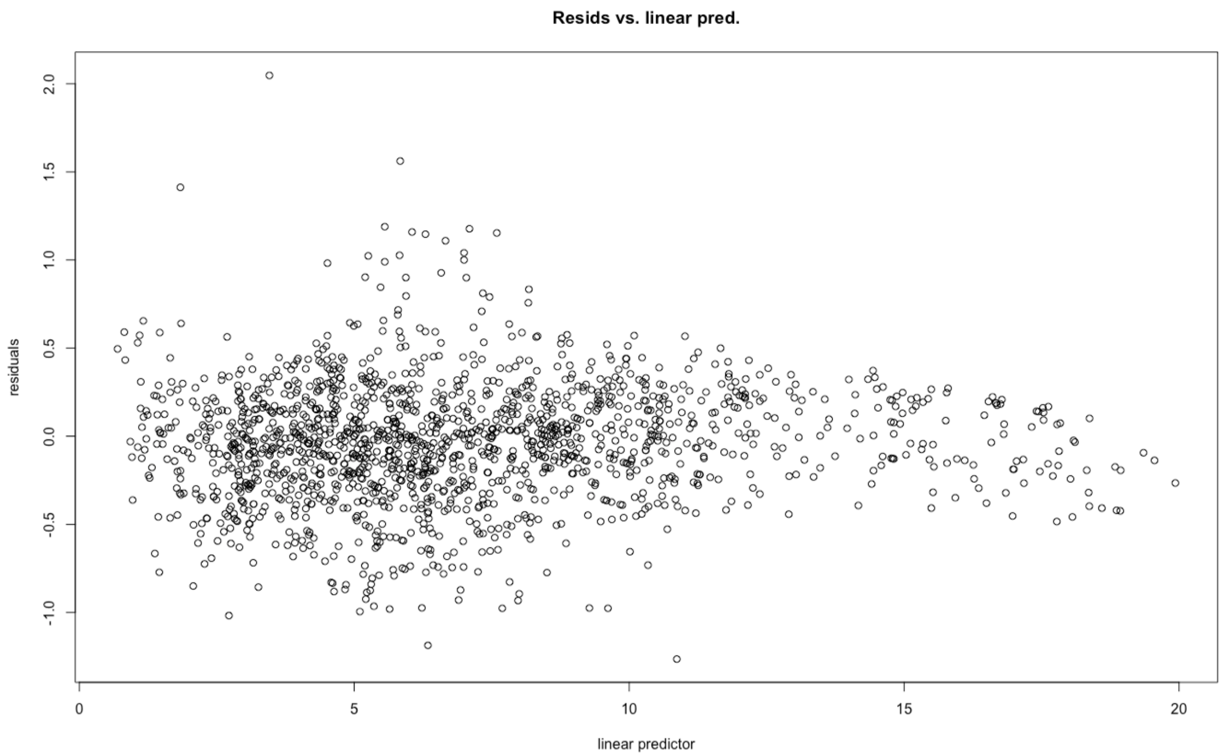
A2 Response Variable Distribution

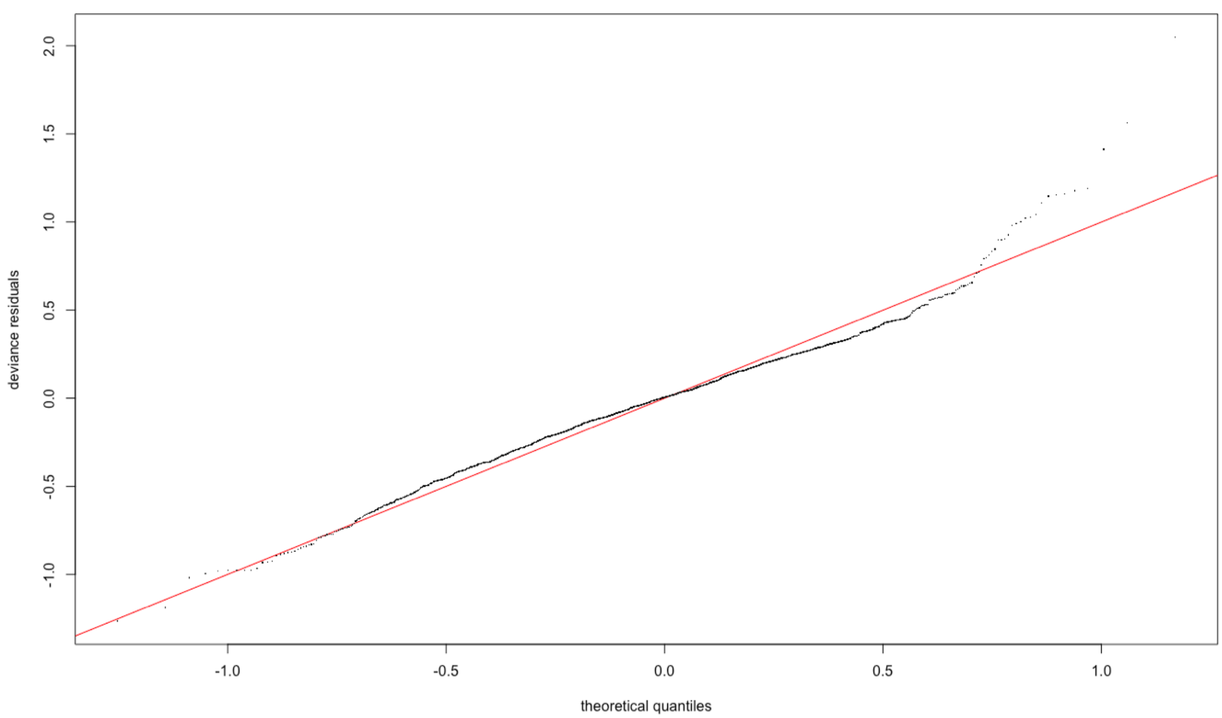


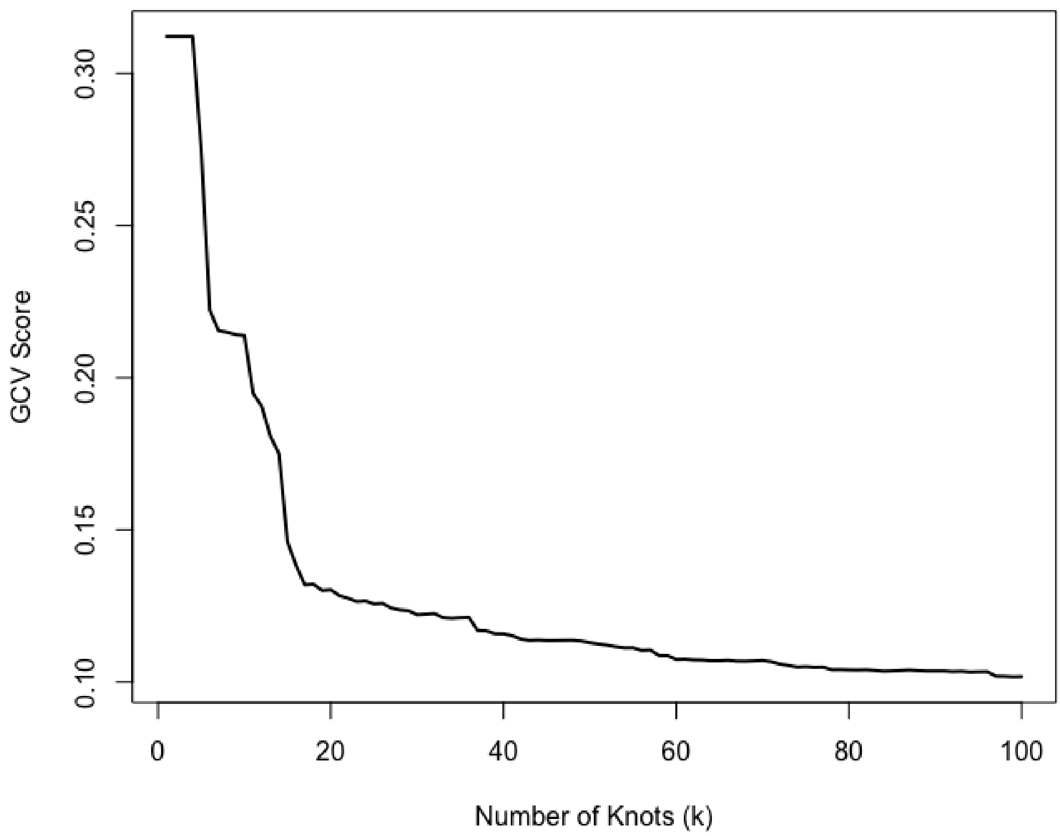
A3 Diagnostic Plots






Resids vs. linear pred.
residuals
2.0
1.5
1.0
0.5
0.0
-0.5
-1.0
0
5
10
15
20
linear predictor


deviance residuals
2.0
1.5
1.0
0.5
0.0
-0.5
-1.0
-1.0
-0.5
0.0
0.5
1.0
theoretical quantiles


GCV Score
0.10
0.15
0.20
0.25
0.30
0
20
40
60
80
100
Number of Knots (k)

A4 R Code

In this thesis, the preferred modelling tool is R. In the following appendix, a subset of our R script is presented. The main package for general additive modelling is the “mgcv”-package.

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

# transform date from POSIXct to date object
handysize$Saledate <- as.Date(handysize$Saledate, "%Y/%m/%d", tz = "GMT")

# remove block sales (2371 obs -> 2114 obs)
handysize <- handysize %>% filter(!grepl('#', handysize$Enblocindicator))

# remove resale hulls (2114 obs -> 2105 obs)
handysize <- handysize %>% filter(!grepl('N/B RESALE HULL', handysize$Vesselname))

# Remove NAs in model variables (2105 obs -> 1850 obs)
handysize <- handysize %>% filter_at(vars(Dwt, Buildercountry, Region, Countryindicator, Price,
Speedknots, Mainfuelconsumptiontpd, HP, Vesselfueltype, Noholds, Nohatches,
Gearindicator),all_vars(!is.na(.)))

# remove africa-transavtions due to data basis (1850 obs -> 1848 obs)
handysize <- handysize %>% filter(!grepl('Africa', handysize$Countryindicator))

# construction of FEI (31 -> 32 vars)
handysize$FEI <- (handysize$Mainfuelconsumptiontpd/(handysize$Dwt*handysize$Speedknots*24))*10^


#####################################################################################################
                                         FILTER INPUT DATA
#####################################################################################################

# Subset raw data for selected variables (32 vars -> 14 vars)
Input <- handysize %>% select (-c(Vesselname, IMOnumber, Vesseltype, Monthbuilt, Deliverydate,
Mainenginerpm, Builder, Enblocindicator, Seller, Sellercountry, Buyer, Buyercountry, Enginesummary,
Graincapacitycum, Gearsummary, Yearbuilt, Buildercountry, Region))

# rename for modelpurposes
colnames(Input)[which(names(Input) == "Age at sale")] <- "Salesage"

# factor cols due to GAM attributes
cols <- c( "Vesselfueltype", "Gearindicator", "Countryindicator", "Enginebrand")
Input[cols] <- lapply(Input[cols], factor)

# extract year from data col
Input$Saledate <- format(as.Date(Input$Saledate, format="%Y/%m/%d"),"%Y")

# numeric for correlation purposes
Input$Saledate <- as.numeric(Input$Saledate)

# Data Correlation (integer cols for correlation purposes)
cor.data <- Input
integ <- c("Countryindicator", "Vesselfueltype", "Gearindicator", "Enginebrand")
cor.data[integ] <- lapply(cor.data[integ], as.integer)

# correlation plot
sjp.corr(cor.data)

# correlation table
sjt.corr(cor.data, triangle = "lower”)

# remove highly correlated variables (14 vars -> 12 vars)
Input <- Input %>% select (-c(Mainfuelconsumptiontpd, Nohatches))

# boxplot dataset to identify outliers
boxplot(Input)

# store FEI-outliers
FEI_outliers <- boxplot(Input$FEI, plot = FALSE)$out
```

```
# store Price-outliers
Price_outliers <- boxplot(Input$Price, plot = FALSE)$out

# store HP-outliers
HP_outliers <- boxplot(Input$HP, plot = FALSE)$out

# store Speed-outliers
Speed_outliers <- boxplot(Input$Speedknots, plot = FALSE)$out

# remove outliers (1848 -> 1625 vars)
Input <- Input[-which(Input$FEI %in% FEI_outliers),]
Input <- Input[-which(Input$Price %in% Price_outliers),]
Input <- Input[-which(Input$HP %in% HP_outliers),]
Input <- Input[-which(Input$Speedknots %in% Speed_outliers),]


 ####################################################################################################
                                              PERIODS
 ####################################################################################################

# subset for data pre 2003: 521 obs.
Pre03 <- subset(Input, Saledate <= "2003")

# subset for data between 2003-2008: 390 obs.
Pre08 <- subset(Input, Saledate > "2003" & Saledate <= "2008")

# subset for data post 2008: 714 obs.
Post08 <- subset(Input, Saledate > "2008")


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

# Method derived from Wood (2006) p. 130->

## Distibution fitting

# fit normal/gaussian distribution to response variable
fit.gaussian <- fitdist(Input$Price, "norm")

# fit gamma distribution to response variable
fit.gamma <- fitdist(Input$Price, "gamma", lower = c(0, 0))

# kernel density of response variable
denscomp(list(fit.gaussian, fit.gamma), fitcol = c("red", "blue"), main = "Kernel Density Distribution
of Response Variable", xlab = "Response variable")

# goodness of fit for distributions
gofstat(list(fit.gamma, fit.gaussian))

## Model development

# set input parameters
parameters <- c("s(Saledate, Salesage)", "Dwt", "Speedknots", "Countryindicator", "FEI",
"HP","Enginebrand", "Noholds", "Vesselfueltype", "Gearindicator")

# create an emptu array for parameters order combinations
ParamArray <- array(1:100, c(length(parameters), length(parameters)))

# create an empty array for all combinations of parameters
allParamArray <- array(1:100, c(length(parameters)*length(parameters), length(parameters)))

# Fill in array for order combinations
for(x in 1:length(parameters))
{
  ParamArray[x,] <- parameters
  object <- parameters[1]
  parameters <- c(parameters,object)
  parameters <- parameters[-1]
}

# Make all the combinations of the parameters
zeros = 1
ParamSelection = 1
```

```
for (r in 1:(length(parameters)*length(parameters)))
{
  para <- ParamArray[ParamSelection,]
  allParamArray[r,] =
c(para[1:zeros], rep(c(0), times=length(parameters)-zeros))
  zeros <- zeros + 1
  if(zeros > length(parameters))
  {
    zeros = 1
    ParamSelection = ParamSelection + 1
  }
}

# Delete duplicate rows
deleteIndex = 0
for(p in 1:length(parameters))
{
  if(p!=1)
  {
    allParamArray <- allParamArray[-(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(allParamArray))
{
  clean <- allParamArray[row,]
  clean <- clean[clean!=0]
  model <- gam(as.formula(paste("Price ~",paste(clean, collapse = "+"))), family =
Gamma(link="identity"), data = Input)
  models <- list(models, allParamArray[row,])
  gcv[row] <- model$gcv.ubre
  radj[row] <- summary(model)$r.sq
  nHandy [row] <- summary(model)$n
}

# plot gcv valus for all interations
plot(gcv, col="blue", type="l")

# Check and print one specific model
checkModel <- function(modelNumber){
  clean <- allParamArray[modelNumber,]
  clean <- clean[clean!=0]
  model <- gam(as.formula(paste("Price~", paste(clean, collapse = "+"))), family =
Gamma(link="identity"), data = Input)
  return(summary(model))
}

# Print the best model based on gcv
printBestModel <- function(){
  clean <- allParamArray[which(gcv==min(gcv)),]
  clean <- clean[clean!=0]
  model <- gam(as.formula(paste("Price~", paste(clean, collapse = "+"))), family =
Gamma(link="identity"), data = Input)
  return(summary(model))
}

# Get the best model as a variable
bestModel <- function(){
  clean <- allParamArray[which(gcv==min(gcv)),]
  clean <- clean[clean!=0]
  model <- gam(as.formula(paste("Price~", paste(clean, collapse = "+"))), family =
Gamma(link="identity"), data = Input)
  return(model)
}
```

##

```
                                           KNOTS OPTIMIZATION
#######################################################################################################

# create empty list for gcv values
testgcv <- 0

# iterate with increasing k
for (i in 1:100){
  gamTEST <- gam(Price ~ s(Saledate, Salesage, k=i) + Dwt + Countryindicator + FEI + HP + Noholds +
Vesselfueltype + Gearindicator + Speedknots + Enginebrand, family = Gamma(link="identity"), data =
Input)
  testgcv[i] <- gamTEST$gcv.ubre
}

# plot gcv-values given k
plot(testgcv, type = "l", xlab="Number of knots (k)", ylab="GCV score")


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

# GAM running entire period
gam_handy1 <- gam(Price~ s(Saledate, Salesage) + Dwt + Countryindicator + FEI + HP + Noholds +
Vesselfueltype + Gearindicator + Speedknots + Enginebrand, family = Gamma(link="identity"), data =
Input)

# GAM running pre 2003
gam_handy2 <- gam(Price~ s(Saledate, Salesage) + Dwt + Countryindicator + FEI + HP + Noholds +
Vesselfueltype + Gearindicator + Speedknots + Enginebrand, family = Gamma(link="identity"), data =
Pre03

# GAM running period between 2004 and 2008
gam_handy3 <- gam(Price~ s(Saledate, Salesage) + Dwt + Countryindicator + FEI + HP + Noholds +
Vesselfueltype + Gearindicator + Speedknots + Enginebrand, family = Gamma(link="identity"), data =
Pre08)

# GAM running post 2008
gam_handy4 <- gam(Price~ s(Saledate, Salesage) + Dwt + Countryindicator + FEI + HP + Noholds +
Vesselfueltype + Gearindicator + Speedknots + Enginebrand, family = Gamma(link="identity"), data =
Post08)

# GAM running entire period (logged)
gam_handy5 <- gam(Price~ s(Saledate, Salesage) + Dwt + Countryindicator + FEI + HP + Noholds +
Vesselfueltype + Gearindicator + Speedknots + Enginebrand, family = Gamma(link="log"), data = Input)


#######################################################################################################
                                                 PLOTS
#######################################################################################################

# visualization 1st model
fvisgam(gam_handy1, ticktype="detailed", view=c("Saledate","Salesage"), theta=140, ylim=c(0,30),
color="heat", zlim=c(0,20), plot.type="persp", zlab="Price ($ Million)", ylab="Age at sale", xlab="Time
(in Years)")

plot(gam_handy1, all.terms = T)

# visualization 2nd model
fvisgam(gam_handy2, ticktype="detailed", view=c("Saledate","Salesage"), theta=140, ylim=c(0,30),
color="heat", zlim=c(0,20), plot.type="persp", zlab="Price ($ Million)", ylab="Age at sale", xlab="Time
(year)")

plot(gam_handy2, all.terms = T)

# visualization 3rd model
fvisgam(gam_handy3, ticktype="detailed", view=c("Saledate","Salesage"), theta=140, ylim=c(0,30),
color="heat", zlim=c(0,20), plot.type="persp", zlab="Price ($ Million)", ylab="Age at sale", xlab="Time
(year)")

plot(gam_handy3, all.terms = T)

# visualization 4th model
fvisgam(gam_handy4, ticktype="detailed", view=c("Saledate","Salesage"), theta=140, ylim=c(0,30),
```

```
color="heat", zlim = c(0,20), plot.type="persp", zlab="Price ($ Million)", ylab="Age at sale",
xlab="Time (year)")

plot(gam_handy4, all.terms = T)


#####################################################################################################
                                              MODEL CHECK
#####################################################################################################

# basic GAM-check (diagnostic plots)
gam.check(gam_handy1)

# plot of price vs. residuals
plot(Input$Price,residuals(gam_handy1))

# anova-analysis
anova(gam_handy1)


#####################################################################################################
                                           MODEL APPLICATION
#####################################################################################################


## Model output for vessel with constant age

# create empy data frame to store values
stdage <- data.frame(matrix(, nrow = 24, ncol = 6))

# rename columns in data frame
colnames(stdage) <- c("Year", "5 yrs", "10 yrs", "15 yrs", "20 yrs", "25 yrs")

# fill in years in column one
stdage$Year <- c(1996:2019)

# fill in prediction from GAM for 5 year old vessel
stdage$`5 yrs` <- predict.gam(gam_handy1, Std5)

# fill in prediction from GAM for 10 year old vessel
stdage$`10 yrs` <- predict.gam(gam_handy1, Std10)

# fill in prediction from GAM for 15 year old vessel
stdage$`15 yrs` <- predict.gam(gam_handy1, Std15)

# fill in prediction from GAM for 20 year old vessel
stdage$`20 yrs` <- predict.gam(gam_handy1, Std20)

# fill in prediction from GAM for 25 year old vessel
stdage$`25 yrs` <- predict.gam(gam_handy1, Std25)

# plot model output for vessels with constant age
plot(stdage$Year, stdage$`5 yrs`, type = "l", col = "blue", lwd = 2, ylab = "Price ($ Million)", xlab =
"Time (in Years)", ylim = c(0,20))
lines(stdage$Year, stdage$`10 yr`, type = "l", col = "red", lwd = 2)
lines(stdage$Year, stdage$`15 yrs`, type = "l", col = "yellow", lwd = 2
lines(stdage$Year, stdage$`20 yrs`, type = "l", col = "purple", lwd = 2)
lines(stdage$Year, stdage$`25 yrs`, type = "l", col = "green", lwd = 2)
legend(2015, 20, legend = c("5 yrs", "10 yrs", "15 yrs","20 yrs", "25 yrs"), col = c("blue", "red",
"yellow", "purple", "green"), lty=1:2, cex=0.8)

## Model output for vessel depreciation

# create empy data frame to store values
stdyear <- data.frame(matrix(, nrow = 41, ncol = 7))

# rename columns in data frame
colnames(stdyear) <- c("Age", "1996", "2000", "2004", "2008", "2012", "2016")

# fill in vessel age in column one
stdyear$Age <- c(0:40)
```

```
# fill in prediction from GAM in 1996 market
stdyear$`1996` <- predict.gam(gam_handy1, Std96)

# fill in prediction from GAM in 2000 market
stdyear$`2000` <- predict.gam(gam_handy1, Std00)

# fill in prediction from GAM in 2004 market
stdyear$`2004` <- predict.gam(gam_handy1, Std04)

# fill in prediction from GAM in 2008 market
stdyear$`2008` <- predict.gam(gam_handy1, Std08)

# fill in prediction from GAM in 2012 market
stdyear$`2012` <- predict.gam(gam_handy1, Std12)

# fill in prediction from GAM in 2016 market
stdyear$`2016` <- predict.gam(gam_handy1, Std16)

# plot prediction for ship depreciation at predetermined time
plot(stdyear$Age, stdyear$`1996`, type = "l", col = "blue", ylab = "Price ($ Million)", xlab = "Vessel
Age", ylim = c(0,20), lwd=2)
lines(stdyear$Age, stdyear$`2000`, type = "l", col = "red", lwd=2)
lines(stdyear$Age, stdyear$`2004`, type = "l", col = "yellow", lwd=2
lines(stdyear$Age, stdyear$`2008`, type = "l", col = "orange", lwd=2)
lines(stdyear$Age, stdyear$`2012`, type = "l", col = "green", lwd=2)
lines(stdyear$Age, stdyear$`2016`, type = "l", col = "purple", lwd=2)
legend(35, 20, legend = c("1996", "2000", "2004", "2008", "2012", "2016"), col=c("blue", "red",
"yellow", "orange", "green", "purple"), lty=1:2, cex=0.8)

## Following vessel value development for ship built in..

# create empy data frame to store values
stdship <- data.frame(matrix(, nrow = 24, ncol = 8))

# rename columns in data frame
colnames(stdship) <- c("Year", "GAM prediction 1988-build","GAM prediction 1992-build" ,"GAM prediction
1996-build", "GAM prediction 2000-build", "GAM prediction 2004-build", "GAM prediction 2008-build",
"GAM prediction 2012-build")

# fill in years in column one
stdship$Year <- c(1996:2019)

# fill in prediction from GAM for vessel built in 1988
stdship$`GAM prediction 1988-build` <- predict.gam(gam_handy1, Std1988)

# fill in prediction from GAM for vessel built in 1992
stdship$`GAM prediction 1992-build` <- predict.gam(gam_handy1, Std1992)

# fill in prediction from GAM for vessel built in 1996
stdship$`GAM prediction 1996-build` <- predict.gam(gam_handy1, Std1996)

# fill in prediction from GAM for vessel built in 2000
stdship[5:24,5] <- predict.gam(gam_handy1, Std2000)

# fill in prediction from GAM for vessel built in 2004
stdship[9:24,6] <- predict.gam(gam_handy1, Std2004)

# fill in prediction from GAM for vessel built in 2008
stdship[13:24,7] <- predict.gam(gam_handy1, Std2008)

# fill in prediction from GAM for vessel built in 2012
stdship[17:24,8] <- predict.gam(gam_handy1, Std2012)

# plot predictions for vessel value development
plot(stdship$Year, stdship$`GAM prediction 1988`, ylim = c(0,25), type = "l", lwd=2, col = "blue", ylab
= "Price ($ Million)", xlab = "Time (in Years)")
lines(stdship$Year, stdship$`GAM prediction 1992`, type = "l", col = "red", lwd=2)
lines(stdship$Year, stdship$`GAM prediction 1996`, type = "l", col = "orange", lwd=2)
lines(stdship$Year, stdship$`GAM prediction 2000`, type = "l", col = "pink", lwd=2)
lines(stdship$Year, stdship$`GAM prediction 2004`, type = "l", col = "green", lwd=2)
lines(stdship$Year, stdship$`GAM prediction 2008`, type = "l", col = "purple", lwd=2)
lines(stdship$Year, stdship$`GAM prediction 2012`, type = "l", col = "yellow", lwd=2)
legend(2015, 25, legend=c("1988 built", "1992 built", "1996 built", "2000 built", "2004 built", "2008
built", "2012 built"), col=c("blue", "red", "orange", "pink", "green", "purple", "yellow"), lty=1:2,
cex=0.8)
```