Valuation of the Bet Performer: Price Prediction Report

Compass Maritime Services, LLC

Arunabh Choudhury
Yash Gupta
Bharat Modhwadiya
Purvi Panchal
Priyal Desai
Shashank Kooragayalu

Executive Summary

This report aims to provide an in-depth statistical analysis to predict the price of a capesize bulk carrier called the Bet Performer for a client of ours. The price is a function of various internal variables, which are directly related to the specifications and characteristics of the ship, and external variables, which are macroeconomic or are related to the maritime industry in general.

The report also aims to highlight the recommendations on the predicted price interval that gives a fair idea of the lowest and highest price that the ship can be sold for in an auction. We also point out the limitations and challenges faced in the course of statistical analysis and steps taken to overcome those limitations.

We used the multiple linear regression modeling to compute the predicted price of the Bet Performer for our client. We conclude that the predicted price of the ship is $125.75 million and the prediction interval is between $105.91 million and $145.75 million, with a 95% confidence.

Highlights of the Situation

In the maritime industry, purchase and sale of cargo ships is a common practice. For that, sale and purchase brokers are hired by the clients to facilitate the transactions. A ship can be valued by three methods or approaches: Market Based, Income Based and Cost Based[i]. Market based approach is one of the most prevalent methods in the industry and then is Income based valuation.

A cargo ship's price depends on various factors. These factors can be internal or external. Internal factors are directly related to the specifications of the ship such as age, year built, deadweight ton (DWT), main engine type, and engine hours[ii]. External factors can be a multitude of factors such as macroeconomic factors such as the Baltic Dry Capesize Index, interest rates (if the ship is valued using income method), ocean freight rates, oil prices, volatility in the forex market, dockyard in which it was built or the original ship builders, location[iii] of the ship at the time of the sale, and environmental factors[iv] like air pollution by greenhouse gases and potential threat of oil and chemical spills.

However, the main factors that affect the price of the ship, especially in the case of the Bet Performer are the age of the ship, year in which it was built, DWT of the ship and the capesize index at the time the ship was offered for sale and purchase. The age of the ship and the year built are essentially the same variable. The only difference is that age is negatively correlated with the price and year built is positively. Therefore, we take only age as the variable influencing the price. The economic effect of the variables on the price of the ship are as follows:

- Age: As the ship ages, the valuation tends to become lower. A ship can be considered a depreciating asset, just like a car. With use, its lifespan decreases and the price reduces too.
- Deadweight Ton: The DWT is total weight of the payload that the ship can carry on the waters. As DWT is increased, the size of the ship increases and capacity too goes up. Therefore, the price increases with DWT.
- Capesize Index: The Capesize Index measures the freight rates of the global maritime and shipping industry and is a highly volatile index. It is used to value a ship as one of the factors. The higher the index, the higher the price of the ship but it is not a very strong indicator in the price of the ship.

Statistical Analysis

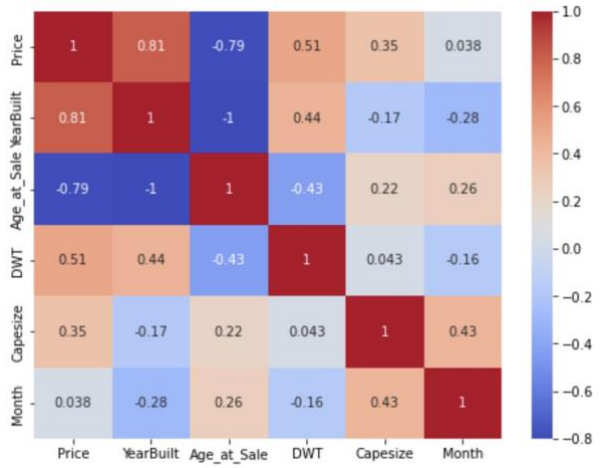
Variables:

The dataset comprises of "Sale Date, Vessel Name, Sale Price, Year Built, Age at Sale, Dead Weight Tons and the Average Monthly Baltic Dry Capesize Index." The Dead Weight Ton (DWT) is the measure of the weight of the ships which factors in the total weight of cargo, fuel, people, and other loads that the ship might be carrying on the waters. "The Baltic Dry Capesize Index (BDI) is an index of average prices paid for the transport of dry bulk materials across more than 20 routes[v]." The variables year built and age are representatives of the same factor at consideration. We decided to select age at sale, as it had the highest correlation with price. Therefore, it is also the best predictor of the price of Bet Performer ship.

Correlation Analysis:

The price of the ship is dependent on the independent variables in the dataset. The correlation coefficient denotes that if one dependent variable moves in a certain direction, how much the independent variable will change with respect to the dependent variable. One thing to note in the dataset is that both "Age at Sale" and "Year Built" are very strongly correlated amongst themselves. Thus, giving rise to multicollinearity. In order to avoid multicollinearity in the dataset, do not consider the Year Built and take only age into account. The following Exhibit A explains the correlation of each of the variables with each other.

Exhibit A



The data suggests that the price is strongly correlated with age with a correlation coefficient of -0.78. Therefore, the age of the ship can be considered as the single best predictor for the price of the ship.

The initial analysis was focused on finding a comparable ship by using the market valuation approach. To approach that, we used the Euclidean Distance method to find out the closest ship which was comparable to the Bet Performer. The Euclidean Distance is calculated by taking the square root of the squares of difference of the coordinate points with the points for the target ship, i.e., Bet Performer. The variables used to calculate the Euclidean Distance are first normalized by scaling the values between 0 and 1. This operation is done by subtracting the minimum of each variable from every data point and dividing it by the range of the variable. The normalized values of the variables are used to calculate the Euclidean distance to get to the nearest matching ship of the Bet Performer. The

Bet Performer has an age of 11, DWT of 172,000 tons and as of May 2008, the Capesize index was 12,479. The parameters of the Bet Performer were normalized and calculated the Euclidean Distance with each ship in the dataset. We use the K-Nearest Neighbors method with K=5. We select the 5 nearest neighbors of the Bet Performer ship and get the table below in Exhibit B:

Exhibit B

Vessel Name	Sale Price	Year	Age	DWT	Capesize Index	Euclidean Distance
Cape Sun	135	1999	9	171.7	11193	0.186
Cape Falcon	87.2	1993	15	161.5	12479	0.199
Nightflight	158	2004	4	170	11614	0.324
Sumihou	106	1996	11	171.1	9663	0.360
Sinfonia	83.7	1991	17	184.4	10526	0.379

As the data in Exhibit B suggests, vessel Cape Sun has the shortest Euclidean distance as it can be considered as the closest comparable ship. Cape Sun's age is 9 years, 2 years younger than Bet Performer and the DWT is 171,700. The Cape Sun was sold at a price of $135 million. Hence, the estimate that basis the market-based valuation approach, the Bet Performer should be worth

around $135 million. For further concise prediction of price, multiple linear regression can be used to estimate the price using the various parameters given.

Multiple Regression Analysis with Ordinary Least Square Method

In this case, for the prediction of the price of the Bet Performer, given more than one independent variable, it is prudent to use a Multiple Linear Regression (MLR) model with Ordinary Least Squares (OLS). Simple linear regression cannot work, simply because there are more than one independent variable and each regression on an independent variable will ignore the other two variables.

Multiple Linear Regression Analysis

The dataset provided had data for 48 ships. The process was to split the data into training and test data sets to perform the MLR, which was done in the ratio of 80/20 respectively. The results of the regression analysis are shown Exhibit C below:

Exhibit C

OLS Regression Results

Dep. Variable:	Price	**R-squared:**	0.920
Model:	OLS	**Adj. R-squared:**	0.915
Method:	Least Squares	**F-statistic:**	169.7
Date:	Mon, 07 Nov 2022	**Prob (F-statistic):**	3.39e-24
Time:	18:45:23	**Log-Likelihood:**	-175.97
No. Observations:	48	**AIC:**	359.9
Df Residuals:	44	**BIC:**	367.4
Df Model:	3		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
Intercept	44.2255	16.383	2.699	0.010	11.207	77.244
DWT	0.2422	0.092	2.643	0.011	0.058	0.427
Age_at_Sale	-4.5438	0.261	-17.378	0.000	-5.071	-4.017
Capesize	0.0072	0.001	12.051	0.000	0.006	0.008

From the Exhibit C, we can observe that age has the highest absolute magnitude of the regression coefficient. That validates our statement earlier that age is the best predictor of the ship price. We also can see that the R-squared is 0.92 which indicates that 92% of the variance is jointly explained by the variables DWT, Age at Sale, and Capesize Index together. R-squared also indicates how tightly the data fits the model. A higher value is considered better because the unexplained variance is lower. The regression equation is defined as:

$$price = age\ at\ sale(-4.54) + DWT(0.242) + capesize(0.00720) + 44.22$$

To check the statistical significance of the variables in the model, we can create a hypothesis to test the significance. The null hypothesis states that the mean of the variables equates to 0 and the alternative hypothesis states that the mean of the variables does not equate to 0. After testing the hypothesis, we found that all the variables are statistically significant as their p-values are less than alpha (assuming alpha is 0.05) for a 95% Confidence Interval.

Checking Assumptions of MLR

For any linear regression, we need to check the 4 basic assumptions:

1. Linearity- This states that the relationship between the dependent variable and the independent variables should be linear in nature. The following plots show that the relationship between the variables is linear and hence it satisfies the criterion. The plots in Exhibit D are for Price vs Age at Sale, Exhibit E is for Price vs DWT and Exhibit F shows Price vs Capesize Index.

Exhibit D

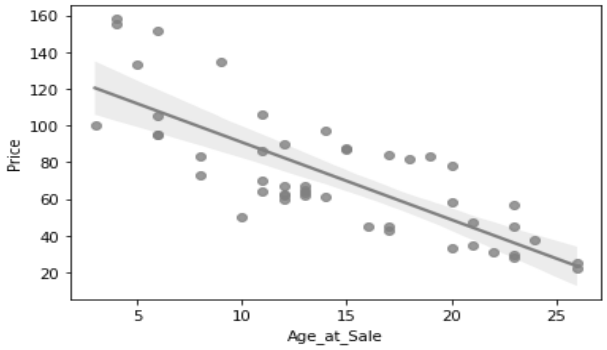


Exhibit E

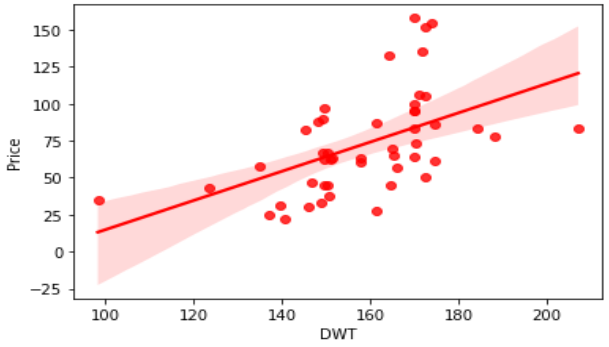
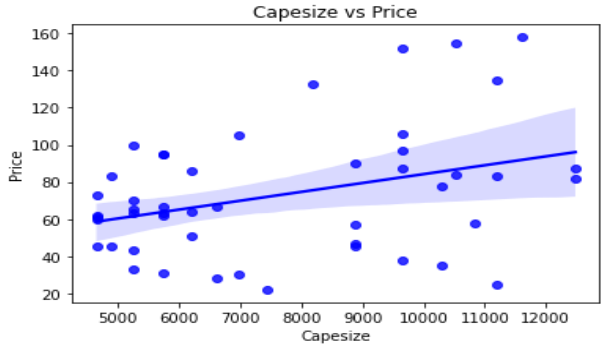
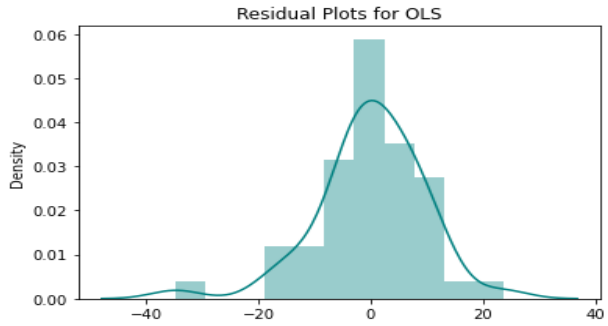


Exhibit F



2. Residuals are normally distributed- From the plot below in Exhibit G, it can be observed that the residuals have a nearly normal distribution.

Exhibit G



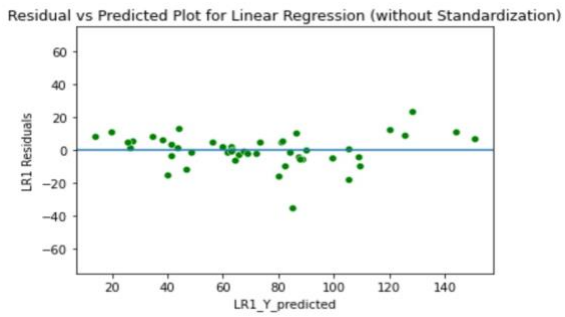
3. Multicollinearity- We used the Variance Inflation Factor (VIF) to test for no multicollinearity. There is no multicollinearity if the VIF for each variable is between 1 and 10. From the below Exhibit H, we observe that it satisfies the above-mentioned criteria of having no multicollinearity.

Exhibit H

	VIF	variable
0	131.944976	Intercept
1	1.258727	DWT
2	1.318729	Age_at_Sale
3	1.075427	Capesize

4. Homoscedasticity- It means that the error terms or the noise have a constant variance. The below scatterplot (Exhibit I) of the error term shows that the spread is constant.

Exhibit I



Results

Using the OLS Regression for multiple independent variables, we find that the model predicts the price of the Bet Performer as $125.75 million. The mean absolute error (MAE) for the residuals of the regression is 6.86. The price can of the ship considering the MAE ranges from ($118.9 to $132.7) million. The prediction interval for regression comes out to be ($105.91, $145.75) million, with 95% confidence. The prediction interval is calculated by multiplying the t_cutoff with the standard error of the regression. t_cutoff for a 95% CI with degrees of freedom for residuals of 44 is 2.02 and standard error is 9.88. We estimate the prediction interval because it is an auction for the ship, the highest bidder buys the ship. Therefore, the lowest the ship can be bought for is $105.91 million and the maximum estimated bid can go up to $145.75.

Limitations

However, due to limitations on the number of data points, the problem of overfitting of the model arises. The model will run precisely on the available data but the prediction result on an external data point may have a higher-than-expected margin of error or deviation from the actual result. Moreover, there are other external and macroeconomic factors and risks which are not controlled by any individual stakeholder. The model cannot factor those volatility and factors in computation of the price prediction.

Multiple Regression using Synthetic Data

To avoid this, we tried to try a new method to make the model more robust and avoid overfitting. We decided to create synthetic data points to have more datapoints which mimics the actual data set. We created 5000 rows of synthetic data using Gretel.ai[vi] which uses Generative Adversarial Networks[vii] (GANs). We selected 250 rows from the synthetic generated data and analyzed the regression results on those data sets. After creation of synthetic data points, the total dataset is split into training and test and then we run the regression on the complete dataset. The R-squared reduced from 0.92 to 0.90

and thus we can deduce that overfitting of the model is avoided.

Exhibit J

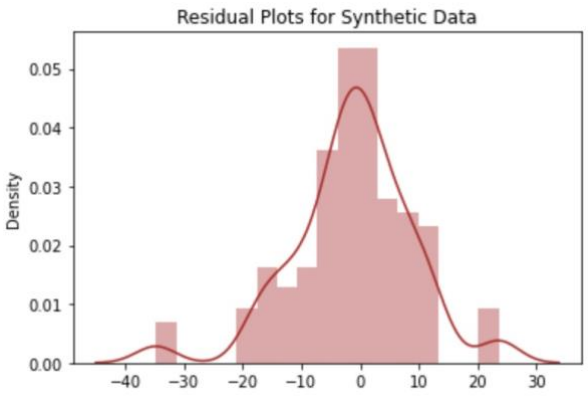
	Price	YearBuilt	Age_at_Sale	DWT	Capesize	Month
count	250.000000	250.000000	250.000000	250.000000	250.000000	250.000000
mean	73.448800	1993.216000	13.984000	157.834000	7737.304000	5.316000
std	33.571131	6.109554	6.141549	19.416699	2511.903797	3.625011
min	22.000000	1981.000000	3.000000	98.400000	4647.000000	1.000000
25%	50.500000	1988.000000	10.000000	149.300000	5245.000000	3.000000
50%	65.000000	1994.000000	13.000000	161.500000	6980.000000	4.000000
75%	90.000000	1997.000000	20.000000	170.200000	10299.000000	9.000000
max	158.000000	2004.000000	26.000000	207.100000	12479.000000	12.000000

Exhibit K

	Price	YearBuilt	Age_at_Sale	DWT	Capesize	Month
count	48.00000	48.000000	48.000000	48.000000	48.000000	48.000000
mean	72.95625	1992.916667	14.270833	158.935417	7643.708333	5.312500
std	33.89537	6.330720	6.330405	17.650984	2499.309368	3.543987
min	22.00000	1981.000000	3.000000	98.400000	4647.000000	1.000000
25%	46.50000	1987.750000	10.750000	149.275000	5245.000000	3.000000
50%	66.00000	1994.000000	13.000000	161.450000	6799.000000	4.000000
75%	88.12500	1996.250000	20.000000	170.125000	9663.000000	8.250000
max	158.00000	2004.000000	26.000000	207.100000	12479.000000	12.000000

Exhibit J and Exhibit K are the summary statistics of the synthetic data and real dataset respectively. The comparison of the means and standard deviations in the tables above shows the real data set and generated data sets are similar in characteristics. However, the coefficients of the regression do not change and the price of the Bet Performer remained same.

Exhibit L



The Exhibit L shows the residual plot for regression done using the synthetic data and it is nearly normally distributed. The regression also satisfies all the other assumptions of linear regression, as mentioned before in another section.

Conclusions

Findings

The Bet Performer ship's valuation has been done using two different techniques. The market valuation method to find the most comparable ship which has been sold and purchased recently. Cape Sun becomes the best reference transaction and is valued around $135 million. Therefore, using market valuation, the Bet Performer should be worth around $135 million. By building a multiple linear regression model using OLS and also by creating synthetic data points to overcome the problem of overfitting, the results of both the regression were very similar. The price of the ship predicted by both the methods is $125.75 million. As per our analysis, we get the MAE of the regression as $6.86 million. If we take the upper limit, the price comes to be $132.69 million.

Recommendations

Basis our analysis, we recommend the following to our client along with 3 different scenarios into consideration:

- Taking into consideration our analysis using the market value approach and the upper limit after adding MAE in the predicted price using MLR, the recommended price of the ship is $132.69 million.
- **Scenario 1:** If the client wants to buy a ship younger than the Bet Performer, say 5 years younger, the predicted price comes to be $148.45 million dollars
- **Scenario 2:** If the client wishes to buy a different ship which is 20,000 DWT lighter, the price comes to be $120.9 million.
- **Scenario 3:** If the charter rates reduce by 30% before the sale of the Bet Performer is completed, the new price recommended to the client is $98.79 million.

References:

[i] HBS Case Study: Compass Maritime Services, LLC: Valuing Ships.
[ii] https://boaterrific.com/how-engine-hours-affect-a-boats-value/
[iii] https://link.springer.com/article/10.1007/s13437-022-00272-4
[iv] https://tethys.pnnl.gov/publications/environmental-effects-marine-transportation
[v] https://www.investopedia.com/terms/b/baltic_dry_index.asp
[vi] https://gretel.ai
[vii] https://en.wikipedia.org/wiki/Generative_adversarial_network

Appendix:

1. Link to the code file used to analyze and predict the price of the Bet Performer: Outliers STAT Ship Group Project Final.pdf