Norwegian School of Economics
Bergen, Autumn 2024

Email circulars as predictive signals in forecasting freight rates

Quantitative supply & demand analysis for freight rate forecasting

Erik Meneilly & Oliver Synnes

Supervisor: Haiying Jia

Master Thesis, Economics and Business Administration, ENE & FIE

NORWEGIAN SCHOOL OF ECONOMICS

Preface

Firstly, we would like to thank our supervisor Haiying Jia for her invaluable guidance, limitless patience, and support throughout our master's thesis. We would also like to thank Zhenming Wu for providing us with important insights into the world of coding.

Secondly, we would like to thank and state our appreciation for the brilliant and active communities on StackExchange and StackOverflow where many questions we too have wondered about have been asked and answered in great detail.

Thirdly, we would like to show our appreciation for and gratefulness to our loved ones whose patience and kindness never wavered throughout our writing of this thesis.

Finally, we would like to extend our gratitude to the Norwegian Shipowners' Association Fund at NHH for their contribution to and support of our thesis.

Due to the fact that the two of us were born and bred in Bergen and Ålesund, two of the most important shipping cities in Norway, it was always a given that we would want to write about shipping.

Abstract

This thesis investigates forecasting freight rates within the Baltic Handysize Index (BHSI), concentrating on the HS7_38 Far East to Southeast Asia route. Leveraging both proprietary data – derived from pre-fixture email circulars containing available tonnage (supply) and demand in deadweight tonnage (DWT) – and publicly accessible data, including the Nominal Broad U.S. Dollar Index and Brent crude oil free on board (FOB) prices, the study use a combination of univariate and multivariate time series models. The research aims to evaluate the forecasting performance of these datasets against a naïve benchmark model to test the Efficient Markets Hypothesis (EMH). Their forecasting accuracy is measured using the mean absolute percentage error (MAPE), mean absolute scaled error (MASE), and root mean squared error (RMSE) metrics.

Using various models – the autoregressive integrated moving average (ARIMA) model, the seasonal ARIMA (SARIMA), ARIMA with eXogenous variables (ARIMAX), and seasonal ARIMAX (SARIMAX) – this thesis finds that incorporating both proprietary data and public data can lead to an improvement in forecasting accuracy. Results demonstrate that the multivariate SARIMAX model with all the variables incorporated, outperform univariate and other multivariate approaches in capturing potential underlying market dynamics, seasonality, and trends. The findings underscore the potential value of incorporating email circulars in improving forecasting accuracy, and how freight rates market can be considered inefficient.

Table of Contents

1. INTRODUCTION **1**

2. LITERATURE REVIEW **4**

3. METHODOLOGY **9**

3.1 ARIMA, ARIMAX; SARIMA, SARIMAX 9
3.2 STATIONARITY 10
3.2.1 Augmented Dickey-Fuller test for stationarity *11*
3.2.2 Kwiatkowski–Phillips–Schmidt–Shin (KPSS) test for stationarity *11*
3.3 MODEL AND METHOD EVALUATION 12
3.3.1 The Akaike's Information Criterion (AIC) *12*
3.3.2 Schwarz's Bayesian Information Criterion (BIC) *12*
3.3.3 The Naïve Method Forecasting *13*
3.4 PERFOMANCE METRICS 14

4. DATA **17**

4.1 VARIABLES 17
4.1.1 Available Tonnage and Demand (DWT) *17*
4.1.2 Freight Rates ($) *20*
4.1.3 Dollar ($) *21*
4.1.4 Oil ($) *22*
4.2 DATA ANALYSIS AND PROCESSING 23
4.2.1 Daily Data *23*
4.2.2 Weekly Data *26*
4.3 DATA PROCESSING 27

5. FORECASTING MODELS **32**

5.1 UNIVARIATE MODELS 32
5.2 MULTIVARIATE MODELS 35

6. FORECASTING **39**

6.1 COMPARISON OF MODELS 39
6.2 IN-SAMPLE FORECAST 43
6.3 FINAL FORECASTS 45

7. DISCUSSION **47**

8. CONCLUSION **50**

8.1 LIMITATIONS AND FURTHER RESEARCH 51

9. AI DECLERATION **53**

10. BIBLIOGRAPHY **54**

11. APPENDIX **58**

11.1 EXHAUSTIVE LIST OF PACKAGES USED IN PYTHON 69

.

1. Introduction

Forecasting is an ancient, enduring, and ubiquitous practice that has been around for millennia (Hyndman & Athanasopoulos, 2021, ch. 1; Stopford, 2009, pp. 698-700)[1]. The reasons for forecasting can be very diverse. On a fundamental level, accurate and timely forecasts can help us plan, prepare, mitigate risks, and even give us a sense of security. Weather forecasts can give estimate of hurricane trajectory or warn of when the monsoon season will fall, giving shipowners a notion to steer away from or exercise caution in the affected area. Inflation forecasts can provide market participants with an idea as to how the market will develop over the next months and years. On a more practical level, the reason for forecasting can be to net financial gains. Freight rates forecasts can be helpful in navigating the complexities of the dry bulk market, enabling shipowners to make better decisions regarding where to allocate their vessels and which charter type to opt for. This can, in turn, potentially lead to returns above the market average if the market is inefficient, which the Efficient Markets Hypothesis (EMH) argues is not the case since markets are efficient.

Freight rates make up the price of transporting goods from port A to port B. They are based on different contracts, such as time charter (leasing a vessel for a specific time (usually on a $/day basis)), voyage charter (hiring a vessel for a specific journey (usually on a $/tonne basis, i.e. the spot market)), or forward freight agreements (where one can hedge against future freight rate fluctuations) (Stopford, 2009, pp. xxii-xxiv, 213). Freight rates, in turn, impact shipping supply and demand as the shipowner views it as revenue and the cargo owner sees it as a cost of transport (Stopford, 2009, pp. 136-138). Revenue from freight, in combination with the size of the world fleet and its productivity, scrapping and losses, and shipbuilding production all influence shipping supply; transport costs, alongside the state of the global economy, average haul, random shocks, and seaborne commodity trades all affect shipping demand. Shipping supply and demand are intertwined variables that play a critical role in shaping the overall equilibrium of the shipping market.

[1] The online book of Hyndman and Athanasopoulos, *Forecasting: Principles and Practice*, is divided into chapters, not pages. To more properly refer to where we have acquired our information and attribute it to the correct author(s), we have detailed in which *chapter* we have found it. The same also applies to Ivan Svetunkov's online book, *Forecasting and Analytics with the Augmented Dynamic Adaptive Model (ADAM)*.

The advancement in labour productivity, driven by specialisation, led to increased productivity (Stopford, 2009, pp. 4-5). To account for the increased productivity, economies had to expand their markets to sustain this growth. Seaborne trade consequently became crucial for the circulation of trade beyond local boundaries, enabling the consumption of excess production and empowering further economic development. Furthermore, seaborne trade plays a crucial role in fostering international specialisation, enabling countries to focus on their comparative advantage, leading to an increase in productivity and overall economic gains (Acharyya, 2023, pp. 8-10).

Seaborne trade accounts for roughly 80-90% of the world trade volume, depending on the source, underscoring its importance in global commerce (International Chamber of Shipping, n.d.; United Nations Conference on Trade and Development, n.d.; Willige, 2024). Moreover, seaborne trade is more cost-effective per unit of goods transported when compared to land-based options (Stopford, 2009, p. 385).

Predictably, given the importance of global commerce through seaborne trade, a considerable amount of time and effort has already been devoted to the research of forecasting freight rates. A relatively novel way of forecasting demand and supply is by way of using email circulars. Email circulars are essentially business letters that are sent to a large number of people, e.g. from shipowners, brokers, and charterers to other shipowners, brokers, and charterers. The information circulated regards available tonnage (i.e. supply) and cargo orders (i.e. demand). Data providers such as Shipfix collects and analyses the information to provide valuable market insights to their subscribed clients.

Our work examines and uses additional exogenous variables – collected from email circulars – that have not been included in past studies to see if they can work as predictive signals in freight rate forecasting in the dry bulk market and thereby improve forecasting accuracy. We also aim to observe whether these signals give indication of efficiency or inefficiency in the freight rate market, in particular for our selected route. The target for our predictions, the Baltic Dry Handysize Index route HS7_38 in the Far East to the Southeast Asia regions, is assumed to already reflect all relevant information. In our thesis, we will make use of email circulars to see if such information will yield better predictive results than a random walk to test for market efficiency. We will elaborate on this in the upcoming chapters: chapter 2 is the literature review; chapter 3 regards our methodology; chapter 4 showcases our data;

chapter 5 compares the different models; chapter 6 examines and compares our models and forecasts; chapter 7 features a discussion of our findings; and, finally, chapter 8 concludes thesis.

2. Literature Review

Today, the methods for forecasting are considerably more sophisticated: with a simple application on your phone, you can discover the probability of precipitation based on complex calculations rather than basing it on, say, which way a spider has spun its web (vertically apparently indicated rain). Forecasting, besides its numerous areas of application such as for the weather, stock returns, capacity, etc., plays an important role in aiding decision making and planning for organisations by for instance quantifying risk (Petropoulos et al., 2022, pp. 709-710).

The choice of using a univariate or multivariate method to forecast with is a fundamental one (Cullinane, 1992, p. 92). The former tests only one variable and the latter tests more (two or more) variables. Equation (2.1) is a simple autoregressive model only considering one variable (Hyndman & Athanasopoulos, 2021, ch. 9.3):

$$y_t = c + \emptyset_1 y_{t-1} + \emptyset_2 y_{t-2} + \cdots + \emptyset_p y_{t-p} + \varepsilon_t \tag{2.1}$$

wherein ε_t is white noise, c is a constant, and the rest constitute the lagged values of y_t at time(s) y_{t-p}.

A multivariate model with an x number of exogenous variables could be a multiple regression where the dependent variable, Y, is presumed to be a function of set k exogenous variables in a population:

$$E(Y_j) = a + \beta_1 X_{1j} + \beta_2 X_{2j} + \cdots + \beta_k X_{kj} \tag{2.2}$$

wherein $E(Y_j)$ is the expected value of Y_j with a change of one unit of β_k, everything else kept the same (Berry, 1985, p. 9).

In our thesis, we will be using the AutoRegressive Integrated Moving Average model, both with and without seasonal components (univariate), as well as with exogenous variables (multivariate) models. Hyndman and Athanasopoulos (2021, ch. 9) state that the univariate ARIMA model and exponential smoothing constitute two of the most popular methods in time series forecasting. Said model incorporates three distinct components, namely *p, d, q*. AR(*p*) is a method (see equation 2.1) to forecast time series exhibiting autocorrelation, i.e. where the time series' values are highly correlated with its own past observations; MA(*q*) describes the relationship between the observed value of the time series at a specific time and its past error terms (Liu et al., 2020, pp. 61-62). The remaining part, I(*d*), came as a result of Box and Jenkins' (1970) seminal work wherein they developed the method on how to render non-stationary data stationary through differencing. SARIMA adds a *Seasonal* component to the model, and the X means 'with eXogenous variables,' whose equations we will show in the methodology chapter.

Where the focus is on forecasting, univariate models are often more practical (Cullinane, 1992, p. 92). New developments within forecasting notwithstanding, older methods such as ARIMA coupled with exponential smoothing remain useful to this day (Petropoulos et al., 2022, p. 710). Other approaches of an equally simplistic nature to exponential smoothing, such as logarithmic transformation, exhibit robustness and is less susceptible to overfitting.

While a multivariate method often yields more accurate forecasts, the data assimilation and analysis required become more complex (Cullinane, 1992, p. 92). Moreover, multivariate models serve an additional purpose in that they attempt to partially address potential, causal relationships. The additional exogenous variables that we include, especially available tonnage and demand, and test on the endogenous freight rates variable are of crucial interest to our research.

The Efficient Market(s) Hypothesis in finance, simply put, asserts that the target for our predictions (e.g. various indices, options, stock market prices, etc.) is wholly unpredictable. The EMH is closely linked to the theory of Random Walk. This hypothesis, and the theory it supports, rests on the idea that if the flow of information runs unimpeded and all information available at any time is reflected in the market, then the change in tomorrow's prices will be due to the news of tomorrow. The change is then random and independent of yesterday's price. In other words, it cannot be forecasted (Malkiel, 2003, p. 3; Timmermann & Granger, 2004, p. 15).

Mainly, there are three forms of market efficiency based on the variables included in an information set, Ω_t, summarised in the table 2.1 (Degutis & Novickytė, 2014, pp. 8-9;

Timmermann & Granger, 2004, p. 17). In the literature on the predictability of stock market returns, but in all likelihood also in other literature set out to testing the EMH, most of the papers aim to test the EMH in either its weak or semi-strong form. Private information, which can be expensive or difficult – or both – to acquire, is usually kept out of the equation. In its strongest form, it is supposed that not even insiders will be able to consistently net abnormal returns.

Type	**What information is included in each Type of EMH?**	**Examples**
Ω_t – weak	Historical prices/returns	Past and current prices; trading volume; dividends
Ω_t – semi-strong	All publicly available information	Dividend payouts; acquisitions; mergers; accounting policy alterations; bankruptcies
Ω_t – strong	All possible information (including private)	Proprietary research; confidential third-party data; insider information

Table 2.1: Types of efficient markets and information contained within the information set

In spite of the emergence and mounting empirical support for this hypothesis, papers are continuously published endeavouring to predict for instance stock returns – more often than not with rather poor results (Timmermann & Granger, 2004, p. 15). Shipping forecasts have a poor track record, too (Stopford, 2009, pp. 697-698). A common argument is that a so-called 'file drawer bias' exists in the published literature due to the difficulty in getting empirical studies with insignificant findings published. Conversely, in the EMH literature there exists a 'reverse file drawer bias.' In the event that researchers were to sincerely believe they had come up with a feasible and effective forecasting method, there is little to induce them into publishing their method rather than bringing it before the highest-paying bidder.

Although the EMH gained considerable traction during the eighties and has generated an extensive body of literature, there seems to be a common consensus that it is not an absolute truth; instead, it can be seen as a truth on relative terms (Degutis & Novickytė, 2014, p. 20). A plethora of studies highlight market inefficiencies, e.g. anomalous volatility, cyclical return patterns, asset bubbles, and instances of investor irrationality, etc., that the EMH fails to fully explain. Another reason for the aforementioned inefficiencies could be attributed to inattentiveness (Ţiţan, 2015, p. 447). Moreover, a perfectly efficient market would negate the incentive for professionals to come up with innovative methods to net extra profits. In actuality, markets require time to completely absorb new information and innovations (Malkiel, 2003, p. 33; Timmermann & Granger, 2004, p. 26).

In his paper, Shostak (1997, pp. 28-33) argues against the EMH, claiming it is destructive to fundamental analysis. He argues that the EMH assumes all participants to have homogenous expectations, which is contradictory towards the reality of a heterogenous market. This diversity, he states, is the basis of trading, because buyers and sellers operate on differing forecasts and valuations. Further he points out that uniform rationality and knowledge among the participants disregard nuanced and subjective interpretations of information that could influence asset prices.

A study conducted by Theissen and Betzer (2009, p. 427) is one of the rare theses that tests the strong market hypothesis (in finance). Their study had a time window of 20 days, where they looked at how insider trading affected stock price, and found that the stock price rose when insider trading happened and declined after insider sale. In addition, they looked at how other factors would contribute to the impact of insider trading. Interestingly, purchases prior to earnings announcement had a more significant effect on price, suggesting that insiders could potentially have a strategic advantage in timing.

For the shipowner, a vital practical question would be to either lease the vessel out on the time charter market or the spot market (Karakitsos et al., 2014, p. 95). If markets are efficient, then the decision will have no bearing on profitability. However, short-term excess profits in shipping may not debunk market efficiency if they are considered as risk premia (Karakitsos et al., 2014, p. 99). Then extra profits would come as a result of taking on extra risk. Therefore, modelling and explaining the risk that affects shipping decisions make up the contributions on this literature, not market efficiency tests specifically. In their book, Karakitsos et al. (2014, p. 126) state that the EMH can be rejected if efficiency is a requirement in every period of time, but not long term, and that this rejection can be the result of time-varying risk premia. Borrowing empirical tests from the financial field and tailoring them to the shipping market by taking into account the finite life of vessels and time charter contracts, Karakitsos et al. (2014, p. 132) further state that the empirical evidence, 'on balance,' suggests that freight rates are inefficient.

In summary, Random Walk moves at random and cannot be predicted solely based on past values. In the same vein, the EMH argues that stocks and indices, like the BHSI, exhibit the same arbitrariness, and that, at least according to Malkiel (2003, p. 5), markets do not allow investors to secure 'above-average risk-adjusted returns.' This supports the notion that markets are efficient and that it is difficult to continuously outperform the market through forecasting (Hyndman & Athanasopoulos, 2021, ch. 9.1.). However, by applying models such as the univariate and multivariate models, this study hopes to uncover and capitalise on any underlying patterns in the targeted BHSI HS7_38 route to increase forecast accuracy.

3. Methodology

We will in this section delve into the methodology behind our research wherein we will discuss model choice, different transformation methods, stationarity tests, and our performance metrics.

3.1 ARIMA, ARIMAX; SARIMA, SARIMAX

The ARIMA model equation with the application of the backshift notation, *B*, which effectively shifts the data back a certain number of periods, give us a shorter equation than the original (Hyndman & Athanasopoulos, 2021, ch. 9.2, 9.5):

$$\underbrace{\left(1-\emptyset_1 B-\cdots-\ \emptyset_p B^p\right)}_{\text{AR}(p)}\underbrace{(1-B)^d y_t}_{d\text{ differences}}=c+\underbrace{\left(1+\theta_1 B+\cdots+\theta_q B^q\right)\varepsilon_t}_{\text{MA}(q)} \tag{3.1}$$

An ARIMAX model is a further extension of the ARIMA model. It applies the autoregressive integrated moving average to the dependent variable while also benefitting from the additional explanatory power of one or more independent variables (Chen et al., 2012, p. 501). Consequently, the model does not exclusively base itself on its own endogenous variable's historical data, but also considers the exogenous variables. Equation 3.2 is given for an ARIMAX model:

$$\left(1-\emptyset_1 \mathrm{B}-\cdots-\ \emptyset_\mathrm{p} \mathrm{B}^\mathrm{p}\right)(1-\mathrm{B})^\mathrm{d} \mathrm{y}_\mathrm{t}=\mathrm{c}+\left(1+\theta_1 \mathrm{B}+\cdots+\theta_\mathrm{q} \mathrm{B}^\mathrm{q}\right)\varepsilon_\mathrm{t}+\beta x_t \tag{3.2}$$

wherein x_t is a covariate at time *t* and β is its associated coefficient (Hyndman, 2010).

The ARIMA and ARIMAX models are both non-seasonal models. By including seasonality, we arrive at a SARIMA(X) model. Below is an overview of the two portions that make up a SARIMA model (Hyndman & Athanasopoulos, 2021, ch. 9.9):

$$\text{ARIMA}: (p, d, q)(P, D, Q)m \tag{3.3}$$

where the m equals the seasonal period, or, in other words, the number of observations per annum (in our case that is 52 for 52 weeks/year). The seasonal terms are similar to the non-seasonal terms, the difference being that the former contains backshifts of the seasonal period. These terms are then multiplied with each other. Equation 3.4 is for a SARIMA model (backshift notated):

$$\underbrace{\Phi(B)}_{P}\underbrace{\phi(B)}_{p}\underbrace{(1-B)^{d}}_{d}\underbrace{(1-B)^{D}}_{D}y_t = c + \underbrace{\Theta(B)}_{Q}\underbrace{\theta(B)}_{q}\varepsilon_t \tag{3.4}$$

A data-related prerequisite to these models is stationarity – for the exogenous variable(s) where that applies as well – which we will explain in the upcoming section.

3.2 Stationarity

A stationary time series exhibits stable statistical properties over time (Liu et al., 2020, p. 63). The mean and variance of the series are both constant, meaning that it displays no trend and heteroskedasticity. Shortly put, trends signify a sustained directional shift in data, whether upward or downward; this movement need not be linear and may even reverse course. Heteroskedasticity refers to the variance of the residuals possibly not being constant, which would warrant applying for instance logarithm or square root on the forecasted variable (Hyndman & Athanasopoulos, 2021, ch. 2.3, 7.3). Periodic variations should also have been removed. Stationary series are considerably easier to model because their fluctuations are kept within certain bounds, and the previous and future mean and variance are not all too dissimilar. With that said, we cannot, in most cases, come to the conclusion that a time series is stationary based on observations alone. Instead, we need to use well-established statistical approaches to conclude that, which brings us over to our next part, namely the Augmented Dickey-Fuller test followed by the Kwiatkowski–Phillips–Schmidt–Shin test.

3.2.1 Augmented Dickey-Fuller test for stationarity

The Augmented Dickey-Fuller (ADF) test is used to check whether a time series is stationary or not, and it helps to control if a time series has a unit root, which would suggest non-stationarity. As stated by Dickey and Fuller (1979, pp. 427-431), in the context of the ADF tests, the hypotheses are formally stated as follows:

1. Null hypothesis (H_0): The time series has a unit root (i.e. it is non-stationary)

2. Alternative hypothesis (H_1): The time series does not have a unit root (i.e. it is stationary)

To evaluate if the null hypothesis can be rejected, we must interpret the results. Typically, a p-value less than 0,05, and an ADF test-statistics smaller than the critical values (e.g. 1%, 5%, or 10%), suggests that we can reject the null hypothesis, indicating that the time series is stationary. Models where the underlying variables are based on time, such as the ARIMA models, differencing can be applied to achieve stationarity (Mushtaq, 2011, p. 7).

3.2.2 Kwiatkowski–Phillips–Schmidt–Shin (KPSS) test for stationarity

The Kwiatkowski–Phillips–Schmidt–Shin (KPSS) test is a test used to assess the stationarity of a time series (Kwiatkowski et al., 1992, pp. 159-167). Whereas traditional tests like the ADF test assumes a unit root as the null hypothesis (i.e. non-stationary), the KPSS test assumes stationarity as the null hypothesis. The KPSS test compares the null hypothesis, that a time series is stationary around the deterministic trend (i.e. a predictable pattern), to the alternative hypothesis, that a unit has a root. The test statistic is calculated by summing the partial sums of the residual obtained from regressing the time series on a deterministic trend. The Lagrange Multiplier (LM) is used to assess the significance of the residuals, and the test is then used to determine if the residuals resemble a random walk. Generally speaking, a higher LM statistic indicates that the residuals are most likely a random walk, suggesting that the time series is non-stationary.

The critical values are obtained from the distribution that the statistic would follow if we had an infinitely large sample size (i.e. asymptotic distribution) under the null hypothesis. A p-value, generally lower than 0,05, indicates that the null hypothesis can be rejected, suggesting that the time series is stationary. Combining the KPSS with other stationary tests provides us with a more robust and comprehensive assessment of the stationary properties of a time series (Kwiatkowski et al., 1992, pp. 159-167).

3.3 Model and Method Evaluation

In this part of our thesis we will introduce, in brief, the different metrics used to gauge the fitness of model as well as the quality of the forecast.

3.3.1 The Akaike's Information Criterion (AIC)

One such metric is the Akaike's Information Criterion (AIC). This is a well-established metric within the modelling framework of ARIMA (Petropoulos et al., 2022, p. 749). The AIC balances the model's goodness of fit with model complexity, penalising models with more parameters (Hyndman & Athanasopoulos, 2021), striving for parsimoniousness. Parsimoniousness means that the model is kept simple with the purpose of precluding overfitting, a state in which the model fails to capture the patterns in the observed data, but instead starts capturing noise. The AIC formula is taken from (Hyndman & Athanasopoulos, 2021):

$$\mathrm{AIC} = T \log\left(\frac{\mathrm{SSE}}{T}\right) + 2(k+2) \tag{3.5}$$

In the formula, SSE stands for the Sum of Squared Errors (i.e. measuring the discrepancy between observed and predicted values), T denotes the sample size used for estimation, and k refers to the number of predictors present in the model. The reason behind the contents of the last parenthesis is that there are k coefficients in the predictor, as well as the intercept and the residual variance. Usually, the model with the lowest AIC value is the one best suited for forecasting (Hyndman & Athanasopoulos, 2021, ch. 7.5). In their book, Hyndman and Athanasopoulos recommend using, among others, the AIC. Nevertheless, we also choose to include BIC for reference.

3.3.2 Schwarz's Bayesian Information Criterion (BIC)

The BIC formula is thusly denoted (Hyndman & Athanasopoulos, 2021, ch. 7.5):

$$\text{BIC} = T\log\left(\frac{\text{SSE}}{T}\right) + (k+2)\log(T) \tag{3.6}$$

Following in the same vein of the AIC, the aim is minimisation of the BIC, ultimately leading to the selection of the 'best' model. Due to its stricter penalty on model complexity (in other words, its greater emphasis on model parsimony), the BIC will either select the same model as the AIC or a simpler one with fewer parameters. According to simulation studies with small to moderate samples, the BIC is likely to select the correct model structure than the AIC or other criteria (Neath & Cavanaugh, 2012, pp. 199, 201).

3.3.3 The Naïve Method Forecasting

The Naïve Method is a simple forecasting method where the forecasts are the value of the last observation, hence its adjective: naïve. The Naïve Method is optimal when the data follow a random walk (for h time periods), and is called Random Walk forecasts, and can be expressed mathematically as (Hyndman & Athanasopoulos, 2021, ch. 5.2):

$$\check{y}_{T+h|T} = y_T \tag{3.7}$$

Random Walk is one of the simplest theories in forecasting. The theory suggests that the future value of a variable is unpredictable and follows a path of random steps. The theory assumes that all information is reflected in the current value, making it equally likely to go up or down. As a result, the current value provides the most accurate prediction of future values (Hyndman & Athanasopoulos, 2021, ch. 9.1). The mathematical representation for a random walk is:

$$Y_t = Y_{t-1} + \epsilon_t \tag{3.8}$$

where Y_t is the value of the time series at time t, y_{t-1} is the value of the previous time step, and ε_t which is an error term, assumed to be white noise, which is characterised with being independently and identically distributed (i.i.d.).

3.4 Perfomance metrics

Performance metrics are used to evaluate the effectiveness of different models. There are several ways in which you can measure the performance between different models. Among the various metrics available, mean absolute error (MAE) and root mean squared error (RMSE) are the most widely used (Hyndman & Athanasopoulos, 2021, ch. 5.8). These are both scale-dependent evaluations that are based on absolute errors or squared errors. In addition to the aforementioned pair, we will also use the mean absolute percentage error (MAPE) and the mean absolute scaled error (MASE). Finally, it should be noted that error, or forecast error, specifically, does not mean mistake; instead, it represents the unpredictable component of an observation, calculated as the difference between the actual value and the predicted value.

MAE assesses the magnitude of prediction errors by averaging the absolute deviations between actual and forecasted values. The formula looks like this:

$$MAE = mean(|e_t|) \tag{3.9}$$

MAPE determines the average absolute percentage difference between the actual values and the predicted values. This gives us a percentage that shows how far, on average, our prediction differs from the actual values. The mathematical formula for MAPE can be expressed as:

$$MAPE = \left(\frac{1}{n}\sum_{i=1}^{n}\frac{|A_t - F_t|}{A_t}\right) \times 100\% \tag{3.10}$$

where A_t is the actual values at time *t*, F_t is the forecasted value at time *t* and *n* is the number of observations.

Where a series is of different scales, such as ours is ($ and DWT), MASE may be preferrable than to for instance MAE (Hyndman & Koehler, 2006, p. 687). Since it is based on absolute errors rather than squared errors it is robust towards outliers (extreme values). Additionally, it has the quality of possessing scale independence, because it scales the forecast error using the average absolute error of a naïve benchmark model. The mathematical formula for MASE (Svetunkov, 2023, ch. 2.1) can be expressed as:

$$\text{MASE} = \frac{1}{h}\sum_{j=1}^{h}\frac{|y_{t+j} - \hat{y}_{t+j}|}{\bar{\Delta}_y}, \tag{3.11}$$

of which

$$\bar{\Delta}_y = \frac{1}{t-1}\sum_{j=2}^{t}|\Delta y_j| \tag{3.12}$$

is the mean absolute value of the first differences, thus denoted:

$$\Delta y_j = y_j - y_{j-1} \tag{3.13}$$

of the in-sample data.

MASE, arguably, counters the weaknesses of MAPE at the cost of interpretability (Svetunkov, 2023, ch. 2.1). Hyndman & Koehler (2006, p. 687), on the other hand, argue that it *is* easily interpretable:

Value	**Result**
MASE = 1	In-sample mean absolute error acquired from the one-step forecasts of the naïve method (benchmark).
MASE > 1	The forecasts are, on average, worse than the MASE = 1 results.
MASE < 1	The forecasts are, on average, better than the MASE = 1 results.

RMSE, the extension of mean square error (MSE), is used to measure the root of the squared errors. First, we calculate the sum of the squared differences between the actual value and the predicted value. To get the average of the squared errors, we divide the total square error by the number of observations *n*, which gives us the MSE. Finally, we take the root of the MSE; this step is crucial as it converts the average squared errors back to its original form. This process measures the average magnitude of errors (Willmott & Matsuura, 2005, pp. 79-82).

$$RMSE = \sqrt{\frac{1}{n}\sum_{I=1}^{n}\frac{(A_t - F_t)^2}{n}} \tag{3.14}$$

4. Data

In the following section, we aim to explain the different data sources and expound on the chosen variables. We will also give descriptive statistics of each variable as well. These will be of the weekly-aggregated data which we will, in the following chapter, compare to the daily data to clarify the choice of weekly data for our analysis. A table containing the full descriptive statistics of the variables can be found in table 11.1 in the appendix.

Table 4.1 features a tabulated summary of the variables.

Variables Description			
Data	Specification(s)	Period	Acquired from
Available Tonnage	Handysize vessel, e-mail circulars for tonnage opening.	06.01.2019 to 02.06.2024	Shipfix
Demand	Handysize vessel, e-mail circulars for cargo orders.	06.01.2019 to 02.06.2024	Shipfix
Freight Rate	Far East Asia to Southeast Asia region route: HS7_38.	06.01.2019 to 02.06.2024	Clarksons Research
Dollar	Nominal broad U.S. Dollar Index, a weighted average of foreign exchange values compared to the U.S. Dollar.	06.01.2019 to 02.06.2024	FRED
Oil	Brent Crude FOB Europe.	06.01.2019 to 02.06.2024	EIA

Table 4.1: Table featuring a short description of the variables

4.1 Variables

4.1.1 Available Tonnage and Demand (DWT)

Shipfix – from where our data derive – has an AI-driven, collaborative data platform that provides maritime market information, intended to improve pre- to post-fixture workflows and provide continuity (Veson Nautical, n.d.). Shipfix's business model revolves around a subscription-based platform, which utilises advanced technology to automate content extraction (i.e. email circulars) and provide granular shipping filters in order to help their users make informed decisions and optimize their operations. Their revenue is generated through subscription fees and premium features. More users on their platform would, in turn, offer a higher collaborative value – more information available overall – indicating potential network effects.

At the time we received the data, Shipfix had a 79% coverage of the Handysize segment. This means that they were counting unique vessels that appeared in those circulars and comparing them to all the vessels that make up the global fleet, ultimately scoring a 79-percentage coverage ratio.

The product the shipping companies supply is sea transport; the product the customers demand is not the ship itself, but transport (Stopford, 2009, pp. 53, 567). The customers vary from energy companies to steel mills to sugar refiners and more. Both parties work closely together as one transports the cargo and the other generates and uses cargo in a form of seeming symbiosis. Shipping companies provide different types of oceanic and short-sea conveyance dependent on the needs of the particular customer, occasioning the emergence of major segments in the market, e.g. specialised, bulk, and liner shipping (Stopford, 2009, pp. 53-54). Whereas counting available tonnage can be a relatively straightforward process, tallying demand, however, is not. This is, in part, supported by Stopford (2009, p. 56): the rich variety of cargoes make trade flow analysis between industries intricate; primary materials go from areas of surplus to areas of shortage, but specialist cargoes are bartered for competitive reasons, adding another layer of complexity to the analysis.

The available tonnage and demand data are private data shared with us. From a combination of methods including, among others, machine learning, the pre-fixture email circulars of leading shipowners, commodity traders, and brokers from across the market have been deciphered, analysed, and contextualised into extensive datasets. These have been aggregated and anonymised into global datasets, designed to offer unique insights into regional disparities in available tonnage and demand. The circulars encompass available tonnage openings and cargo orders (i.e. demand).

	count	mean	std	min	max	Skewness	Kurtosis	J-B
AVAILABLE_TON_R_7_raw	283	381511,3	179198,3	60900	1043624	0,8812	0,773659	43,68668
AVAILABLE_TON_G_7_raw	283	1215821	421819,6	294400	2763464	0,7937	0,281417	30,64925
AVAILABLE_TON_R_9_raw	283	647026,8	249722,8	218000	1737149	0,8692	0,959464	46,48652

AVAILABLE_TON_G_9_raw	283	1977780	608823,4	720199	4180539	0,9483	0,897271	51,90933
AVAILABLE_TON_R_12_raw	283	1072923	326460,8	425600	2422976	0,7677	0,707416	33,70226
AVAILABLE_TON_G_12_raw	283	3197691	824594,1	1669599	6343489	1,0518	1,179602	68,58825
AVAILABLE_TON_R_15_raw	283	1478546	395765	447700	2958541	0,7041	0,72509	29,58149
AVAILABLE_TON_G_15_raw	283	11316085	3128877	3235245	22768600	0,8305	0,939134	42,93385
DEMAND_R_7_raw	283	983371,5	598438,8	105998	3377561	1,5952	2,835506	214,8243
DEMAND_G_7_raw	283	3395150	1180329	845107	7259172	0,7764	0,602681	32,71558
DEMAND_R_9_raw	283	1566357	859719,5	300254	5137336	1,5746	2,923043	217,7003
DEMAND_G_9_raw	283	5376384	1683493	1240144	12124776	0,7349	0,848334	33,9568
DEMAND_R_12_raw	283	2552971	1284216	691080	7897845	1,5725	2,627144	198,019
DEMAND_G_12_raw	283	8422705	2423810	2331568	18337213	0,7384	1,025248	38,11042
DEMAND_R_15_raw	283	3508372	1661010	898479	10435164	1,6242	2,917141	224,7657
DEMAND_G_15_raw	283	11316085	3128877	3235245	22768600	0,8305	0,939134	42,93385

Table 4.2*:* Descriptive statistics of available tonnage and demand data

The AVAILABLE_TON (available tonnage) and DEMAND variables from table 4.2 are sums equal to the sum of the individual fixtures' vessel specifications' normalised deadweight tonnage. In both datasets that we eventually combined, we have extracted the pertinent data based on several factors:

First and foremost, the vessel class made available or sought after must be Handysize.

Second, for *cargo orders* (DEMAND), the discharge area is *in* the specified region surrounding our route, i.e. Far East Asia and Southeast Asia; for *available tonnage* (AVAILABLE_TON (supply)), the loading area is *in* the specified region surrounding our route. It is not of vital importance where the vessels in demand load but discharge; they could be loaded in Asia, or somewhere else entirely, but they discharge within the regions we have specified, bringing in cargo and their presence. Regarding *cargo orders,* the important feature is not where they discharge their cargo but where they load their cargo. They could discharge the cargo in Asia, or in another continent, but they load the cargo within the region we have specified. The rationale for this is that these fixtures represent demand-supply activity as well as a (high or low) presence of vessels within the determined area, which would subsequently either drive prices up or down. Per the EMH, these changes should be reflected in the FREIGHT_RATE either instantaneously or with a slight time lag. Chances are that our filtering will result in us missing out on some observations; we endeavour to check just that later on by removing the geographical parameter.

AVAILABLE_TON and DEMAND are both separated into regional (R) and global (G): AVAILABLE_TON _R or AVAILABLE_TON _G; DEMAND_R or DEMAND_G.

Third, we use only spot data, which is measured by the time from the initial reception date (when the e-mail circular was first circulated) to the earliest given loading date (start of laycan). The number set for this condition is very subjective. We will experiment and investigate if by changing the number of days, we end up getting widely disparate results. Fifteen days could give a mid-delivery vessel a decent amount of time to discharge the cargo at point X and go to the loading port at point Y, but this can be much more time than needed for shorter distances. We have selected four different time windows (i.e. a limitation of maximum number of days) to check the difference: 7, 9, 12, and 15 days. These are applied both to regional and global available tonnage and demand with variables defined by _7, _9, _12, or _15 to indicate the number of days used. AVAILABLE_TON_R_12 would indicate the available tonnage is summed under the conditions set by the regional and 12 days limitations.

Fourth, we avoid any overlap or repeated fixtures by ensuring that the id number of each fixture is unique.

4.1.2 Freight Rates ($)

The time-charter rates variable has been sourced from Clarksons Research Shipping Intelligence Platform (Clarksons Research, n.d.-a) for the HS7_38 route, including North China-South Korea-Japan trip to South-East Asia for the period 06.01.2019 to 02.06.2024. The freight rate used for this thesis is for the Handysize segment.

Clarksons Research provides data and insights on shipping, offshore energy, and global trade. They help businesses make decisions with real-time data and expert analysis (Clarksons Research, n.d.-c). This data was, in turn, sourced from the Baltic Exchange, which we do not have access to. It is the sole independent source of maritime market information for trading and settling physical and derivative contracts worldwide (The Baltic Exchange, n.d.).

The Baltic Handysize Index constitutes our dependent – endogenous – variable, i.e. the variable that we would like to forecast the future fluctuations of. We have narrowed down the scope, both geographically and in granularity, by focusing on a single route located in the Far East – Southeast Asia regions instead of the entire index itself. More specifically, it is the HS7_38 route, North China – South Korea – Japan trip to Southeast Asia. The route is weighted at 10% out of the seven routes that comprise the index. The data were collected from the Clarksons Research Database. To read more on the specificities of the route, please refer to the images 11.1 and 11.2 provided in the *Appendix*.

The Handysize segment, a critical subsector within the bulk carrier market, is one that merits further exploration. These vessels, typically ranging from 20,000 to 40,000 DWT (the thresholds vary slightly depending on the given source and literature), offer a unique dynamic within the shipping industry. These bulk carriers were thusly termed because they were 'handy' (or flexible) (Stopford, 2009, p. 224). In contrast to their larger competitor, the Handymax vessels, that mimics the trading patterns of new Panamax bulkers but with a more extensive list of commodities it can carry, the Handysize can be found all over the world, engaged in the transport of any number of commodities, including but not limited to bulk coal, iron ore, grain, bulk phosphate, cassava, bauxite, steels, scrap, forestry products, etc. (Institute of Chartered Shipbrokers, 2023a, pp. 57, 60). These vessels are particularly sought-after in regions with spatial restrictions, such as the Great Lakes. They are usually geared, and some even have stanchions for the purpose of transporting logs (Institute of Chartered Shipbrokers, 2023b, p. 101).

Handysize vessels' virtual omnipresence make them very difficult to pin down; by instead setting a geographical limitation focusing solely on a single route of the Dry Bulk Handysize Index, we still retain a relatively large area to investigate. The index provides assessments on a $/day time charter rate that are published Monday to Friday, excluding certain holidays.

	count	mean	std	min	max	Skewness	Kurtosis	J-B
Freight_Rate_raw	283	13325,76	8631,041	3526	34966,4	1,102642	-0,1190	57,51302

Table 4.3: Descriptive statistics of the freight rate variable

4.1.3 Dollar ($)

FRED is an online database managed by the Federal Reserve Bank of St. Louis, offering of economic time series from national and international sources (Federal Reserve Economic

Data, n.d.-b). The data on the Nominal Broad U.S. Dollar Index was collected from their platform (Federal Reserve Economic Data, n.d.-a).

The index is essentially a weighted average of foreign exchange values compared to the U.S. Dollar (USD), based on a large group of U.S. trading partners (FRB, 2024). The main reason behind including the index is that the USD is widely accepted and stable, making it an ideal choice for global trade transactions. According to the Federal Reserve Board (Bertaut et al., 2023), the USD is the more frequently used currency in international transactions and financial markets. Similarly, shipping costs and revenues are predominantly expressed in USD; however, the expenses a shipping company incurs are rarely all denominated in USD but in a domestic currency instead, requiring conversion, thus occasioning an exchange rate risk (Akatsuka & Leggate, 2001, p. 236). Fluctuations in exchange rates can have a serious impact on the shipping industry; therefore, by incorporating the Nominal Broad U.S. Dollar Index, we hope to be better able to capture said impact on the shipping industry's performance.

	count	mean	std	min	max	Skewness	Kurtosis	J-B
DOLLAR_raw	283	117,846	3,910295	110,76	128,33	0,2832	-0,6595	8,911211

Table 4.4: Descriptive statistics of the dollar variable

4.1.4 Oil ($)

The U.S Energy Information Administration (EIA) provides independent and impartial energy data, analyses, and insights to foster sound policy, efficient markets, and public awareness of energy's role in the economy and the environment (U.S. Energy Information Administration, n.d.-a). The Europe Brent Spot Price FOB data have been sourced from their website (U.S. Energy Information Administration, n.d.-b).

The main rationale behind including crude oil prices is because bunker prices, representing a substantial proportion of the shipping rates, is related to crude oil prices (Geman & Smith, 2012, pp. 9-10). In fact, considering that oil remains the main energy source for the shipping sector, it is suggested that nearly half of operating expenses is apportioned to the purchase of the commodity (Yilmazkuday, 2024, p. 3).

	count	mean	std	min	max	Skewness	Kurtosis	J-B
OIL_raw	283	72,91618	21,12886	14,24	127,4	-0,1237	0,271296	1,589473

Table 4.5: Descriptive statistics of the oil variable

4.2 Data Analysis and Processing

In the following section, we will go through the data analysis, data pre-processing and processing as well as manipulation parts. For these parts, we have used Python with its extensive arsenal of packages. For instance, similar to R's *fable* package, PMDARIMA's library in Python comes decked with the auto_arima() function. It will help us compute, search, and select the model based on information criteria, such as choosing the one with the lowest AIC score (Nkongolo, 2023, pp. 114-115). Additionally, we employ essential packages such as Pandas, SciPy, Statsmodels, and more for our research. For an exhaustive list, please refer to appendix 10.1.

For the sake of brevity and simplicity, we will from here on out refer to our targeted route (HS7_38) for the Baltic Handysize Index, nominal broad U.S. Dollar Index, Europe Brent Spot Price FOB, and our available tonnage and demand datasets by their original variable name, i.e. FREIGHT_RATE, DOLLAR, OIL, AVAILABLE_TON, and DEMAND, respectively, unless otherwise specified.

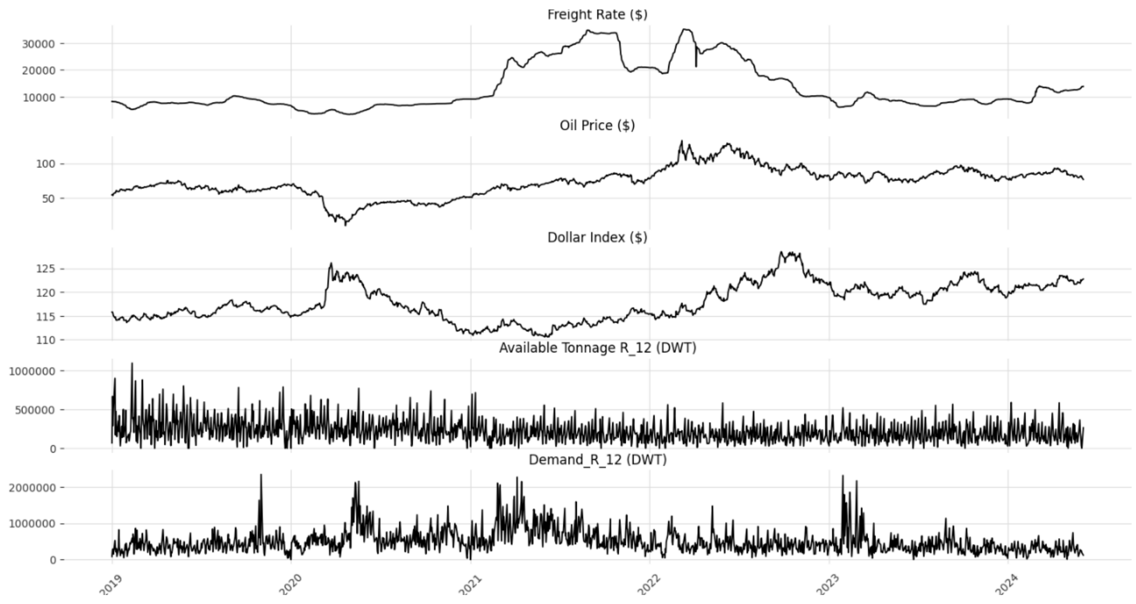
4.2.1 Daily Data

The FREIGHT_RATE, OIL, and DOLLAR variables are daily estimates. Due to a discrepancy in publication days of the prices and indices, there are periods, especially around Christmas and New Year holidays or American banking holidays for DOLLAR, where certain patches appear. It should also be specified that there are no updates during the weekend either. For the AVAILABLE_TON and DEMAND variables, the data flow is much more frequent, even on holidays.

	Count	Mean	Min	Max	SD	Sk.	Kurt.	J-B	p-value
FREIGHT_RATE	1415	13312,2	3482	35125	8644,588	1,105173	-0,12207	288,9269	1,82E-63
OIL	1415	73,05104	9,12	133,18	21,30035	-0,0962	0,232843	5,37887	0,067919
DOLLAR	1415	117,8594	110,5179	128,4544	3,916302	0,280153	-0,66815	44,82985	1,84E-10
AVAILABLE_TON_R_12	1415	212044,7	0	1090225	140251,1	1,384553	3,072531	1008,683	9,3E-220
DEMAND_R_12	1415	502606,5	0	2339858	331497,3	1,891325	5,273518	2483,233	0

Table 4.6: Table containing the descriptive statistics of the FREIGHT_RATE, OIL, DOLLAR, AVAILABLE_TON_R_12, and DEMAND_R_12. Statistics' names vertically placed; variables' names horizontally placed

Please refer to graph 4.1 for a graphic presentation of the data. Observing table 4.6 above, we can see volatility in all variables, but the greatest volatility is in DEMAND_R_12 with 0 (DWT) as the minimum value and 2 339 858 as its maximum value, and the least in DOLLAR. This is also apparent by observing the standard deviation. The Jarque-Bera p-value of less than 0,05 rejects the hypothesis of normality (skew = 0 and kurtosis = 3) (Jing et al., 2008, p. 243).

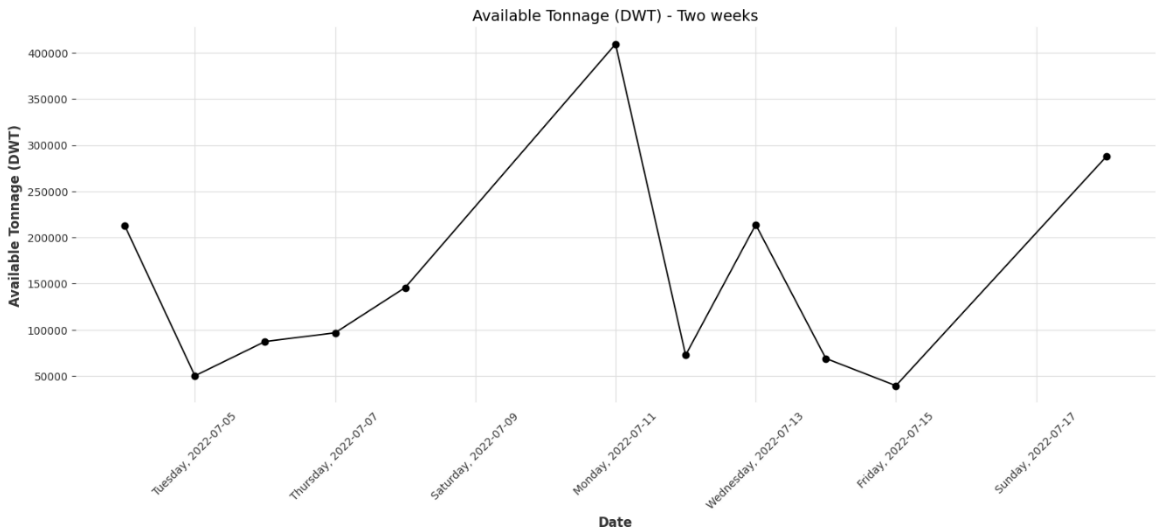


Graph 4.1: Original daily dataset including five variables (from top to bottom): freight rate, oil price, dollar index, available tonnage regional 12 days limit, and demand regional 12 days limit

As depicted in graph 4.1, FREIGHT_RATE remained relatively stable until 2021. Between the first quarter of 2021 and the same quarter the following year, there were several sharp rises and falls, before a decreasing trend ensued lasting till 2023 where it eventually stabilised. OIL has obvious trends. After its plunge in 2020, it steadily increases until circa mid-2022, where it decreases and subsequently stabilises. DOLLAR's trends show some resemblance to OIL. Where OIL initially faces a drop, DOLLAR increases, after which it declines until mid-2021 where it continuously increases till end of 2023, then subsides and more or less stabilises. AVAILABLE_TON_R_12 and DEMAND_R_12 are accentuated by high-frequency spikes, making it hard to spot any trends.

There are three quite interesting takeaways from the graph in relation to the three first variables. Firstly, they all start off quite stable, and that stability lasts the longest for FREIGHT_RATE. In about the first quarter into 2023 they stabilise again. Secondly, OIL and DOLLAR both end up stabilising at an elevated level from their initial value. Thirdly, all three variables undergo marked fluctuations and trends. This period starts from about the first quarter of 2021 for FREIGHT_RATE and the year before that for OIL and DOLLAR, lasting until the same quarter 2023. Throughout 2020 up until 2023, several major geopolitical events occurred, but it suffices to mention the COVID-19 pandemic and the Russian invasion of Ukraine. We believe that these events have had an effect on FREIGHT_RATE, OIL, and DOLLAR.

However, issues with the different daily datasets were consecutive missing dates due to different publication days, especially around the different holidays, and a frequency of zeroes as illustrated in graph 4.1 for AVAILABLE_TON_R_12 and DEMAND_R_12. The former was resolved by linear interpolation, at the cost of inducing a slight biasness; the latter was only prevalent in the available tonnage and demand datasets as for particular dates there were no recorded fixtures. Given its simplicity and suitability where a gradual linear change between known data points is reasonable, we opted for the linear interpolation method to fill in the gaps. Another cause for concern was the fact that the earlier days of the week, in particular Monday, had a predominantly larger portion of fixtures recorded than the succeeding days of the week, illustrated in the graph 4.2:

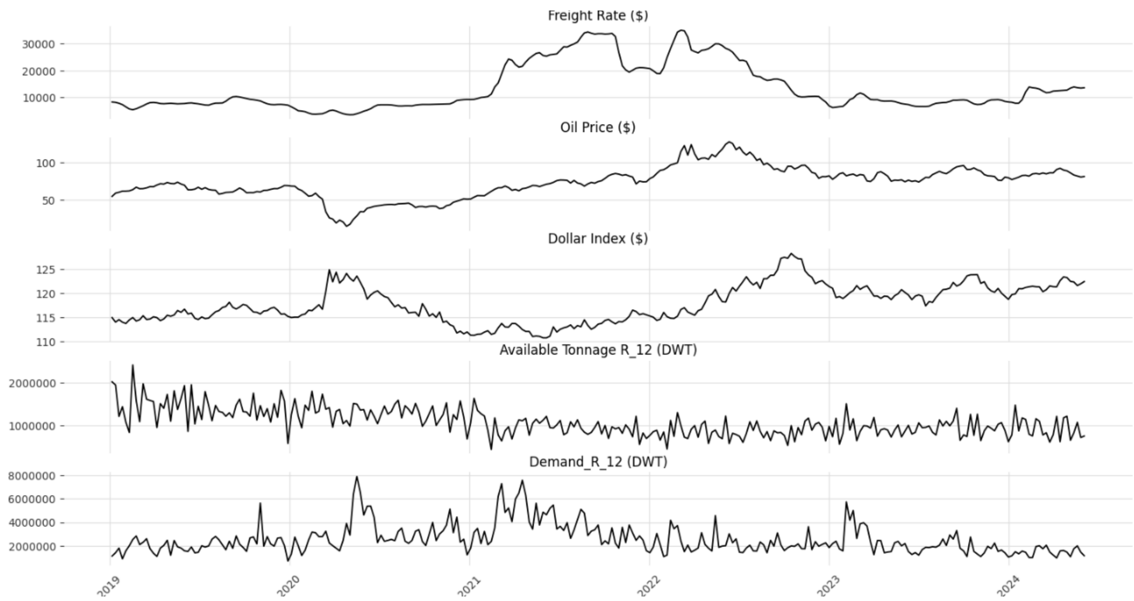


Graph 4.2: Graph illustrating how Mondays show considerably higher sum of fixtures (in DWT) than the following days. Here: AVAILABLE_TON_R_12

Graph 4.2 shows, from the period Monday July 5th, 2022, to Monday July 18th the same year, how Mondays are accentuated with very high values compared to the subsequent days of the week.

To address the issues with patches that cropped up surrounding the holidays in the datasets and the number of zeroes in the private datasets as well as our concern, we have aggregated the daily data into weekly data. This gives us 283 observations. Although the lower number of observations could make the forecasts more inaccurate and biased, we are of the conviction that this is, procedurally, a reasonable trade-off between data quantity and quality.

4.2.2 Weekly Data



Graph 4.3: Original *weekly* dataset including five variables (from top to bottom): freight rate, oil price, dollar index, available tonnage regional 12 days limit, and demand regional 12 days limit

Observing graph 4.3, we see the data has become smoother, especially for AVAILABLE_TON_R_12 and DEMAND_R_12. Their trends have become more distinguishable. AVAILABLE_TON_R_12 decreases to below 1 million DWT in 2021, but then moves back up again, and subsequently stabilises around the 1 million mark. DEMAND_R_12 has more noticeable peaks and troughs, but generally increases and eventually tapers off after first quarter of 2021, and stabilises at the 2 million DWT mark. A final remark is that DEMAND_R_12 has a sharp spike at about the same time as OIL and DOLLAR do, early-to-mid 2020.

The long downward trend for AVAILABLE_TON_R_12 could be due to a number of different reasons: 1) as the oil prices increase, breakeven is harder to achieve, forcing companies to lay up, scrap, or sell their vessels, decreasing available tonnage; 2) in the tumultuous times surrounding COVID-19, more vessels might have opted for time charters, thus becoming unavailable in the market for a long period of time; 3) a non-market-related reason might simply be that Shipfix's customer base went down, limiting coverage; or 4) that a continual increase in the number of Handysize vessels worldwide over the years, by increasing the total, diminishes the overall, percentual coverage of Shipfix in the Handysize segment. From 2013-2023, the average annual increase to the Handysize orderbook based on the overall fleet percentage was 11,9 % (Clarksons Research, n.d.-b).

Reasons 2) and 3) can apply for DEMAND_R_12, too. If a cargo owner has arranged a time charter contract with a shipping company, then that would result in fewer cargo orders circulated. Should both parties belong to Shipfix's customer base, this would have a double impact on email circulation. For reason 3), Shipfix's customer base might have become smaller, leading to the decreasing trend.

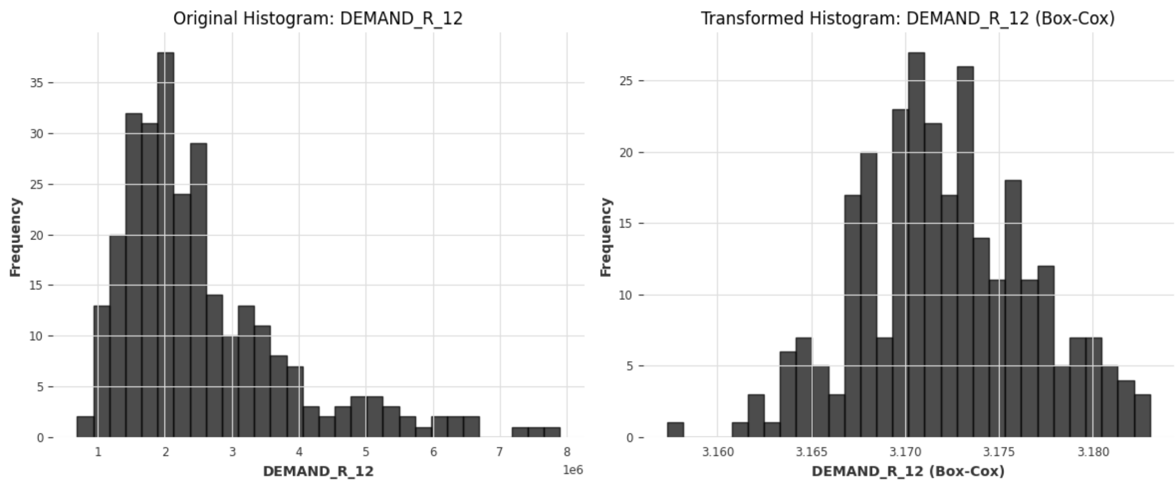
4.3 Data Processing

We have applied the Box-Cox and differencing transformations to our data. This is due to (1) reduce the skewness and kurtosis that can render forecasting more difficult and induce biasness; (2) meet assumptions of homoskedasticity of variance, normality in the data, as well as stationarity; and (3) make the data's statistical properties (e.g. mean and variance) more consistent over time, so that modelling and analysis can become more accurate (Feng et al., 2014, pp. 105-108; Hyndman, 2010; Hyndman & Athanasopoulos, 2021, ch. 9.1;

Osborne, 2010, p. 2). The Box-Cox transformation is shown in equation 4.1 (Hyndman & Athanasopoulos, 2018, ch. 3.2; Osborne, 2010, p. 3):

$$w_t = \begin{cases} \log(y_t) & if\ \lambda = 0; \\ \dfrac{y_t^\lambda - 1}{\lambda} & otherwise. \end{cases} \tag{4.1}$$

If λ = 0, then the natural logarithm is used; if λ ≠ 0, then a power transformation is applied instead coupled with some scaling.



Histogram *4.1:* *Histogram of DEMAND_R_12 following Box-Cox transformation*

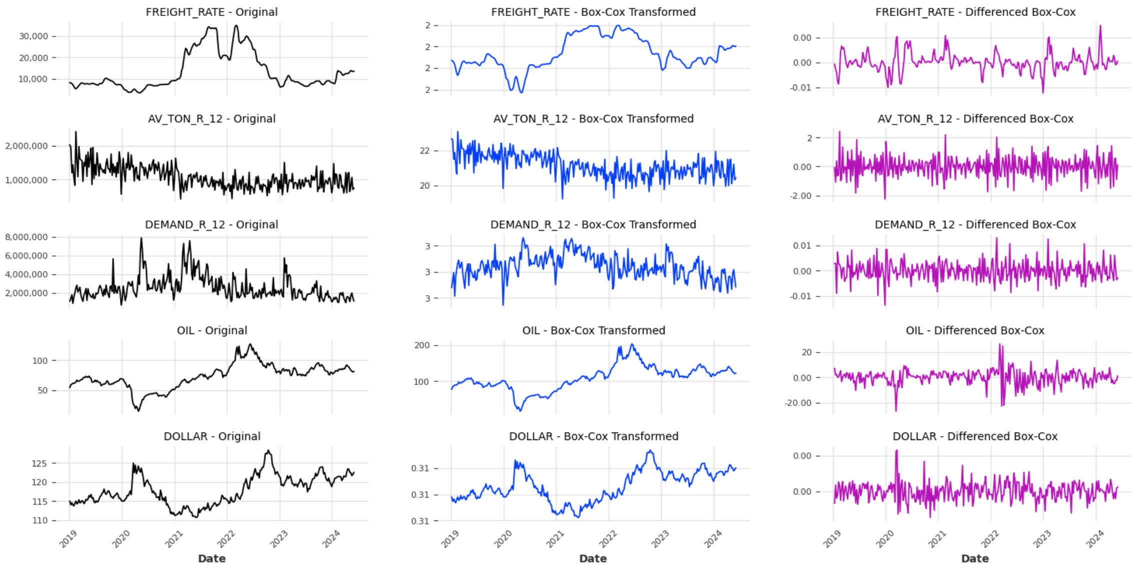
Histogram 4.1 illustrates the improvement in normalisation of the DEMAND_R_12 variable following the Box-Cox transformation. Please see histogram 11.1-11.4 in the appendix for the results of the other transformed variables. An overall improvement was achieved with every variable.

First order differencing is achieved through

$$z_i = y_i - y_{i-1} \tag{4.2}$$

where y_i is the value of the time series at time i, and y_{i-1} is the value of the time series at the previously observed time period $(i-1)$ (Hyndman & Athanasopoulos, 2021, ch. 9.1).

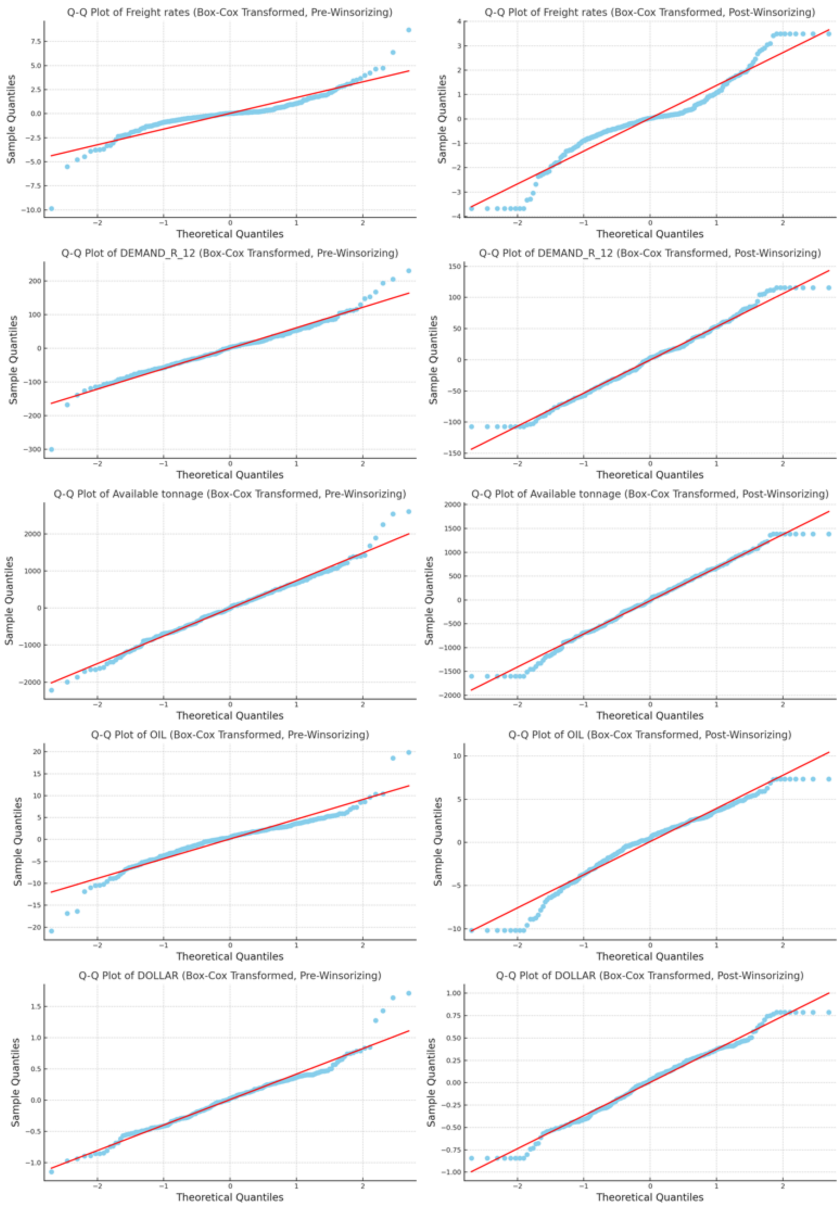
The application of differencing introduces non-stationarity across all variables. Therefore, unit root tests have been conducted for the Box-Cox with and without differencing to ensure stationarity. Following Box-Cox, differencing and winsorization – a technique that caps extreme values to a certain percentage to mitigate the impact of outliers – both the KPSS and ADF showed stationarity at the 1% significance level for all variables (see table 11.2 in the appendix).



Graph 4.4: The graph illustrates the original dataset, weekly, then with Box-Cox transformation applied, and finally with both transformations applied simultaneously

The Box-Cox transformation coupled with differencing have reduced or removed trends. An interesting detail to point out for FREIGHT_RATE in the upper-right corner in graph 4.4 is that, every year, roughly around January/February, there appears to be an increase followed by a dip. Given the specific region we are focusing on and China's overall prominence in the dry bulk market, this could be due to the Chinese New Year celebrations. Our supposition is that before the holidays, the warehouses are stocked up and demand diminishes; then, at the conclusion of the holidays, demand resurges alongside FREIGHT_RATE.

Winsorization was applied at the 95th percentile; this involves truncating the data at the 2.5th and 97.5th percentile, effectively reducing the impact of outliers. The effects of winsorization have been visualised in graph 4.5, in the appendix (histogram 11.5), and summarised using a z-test in the appendix (table 11.5). A comparison of pre- and post winsorization reveals a reduction in the magnitude of residuals, while the visualisation indicates a more normally distributed residual distribution. The Q-Q plot shows a closer alignment with the theoretical normal distribution after the winsorization, while the histogram indicates a more symmetrical distribution with reduced skewness. After these results, we have decided to winsorize all the data.



Graph 4.5: Normal Q-Q plots of the Box-Cox transformed and differenced variables (pre- and post-winsorization)

To summarize, having reduced the kurtosis, skewness, heteroskedasticity and the impact of outliers, while having established sufficient evidence of stationarity in the transformed time series through the ADF and KPSS test, we will in the next chapter move forward with the model selection.

5. **Forecasting Models**

In this section, we will first explore the univariate ARIMA and SARIMA models to identify the most optimal model based on metrics such as AIC, BIC, R^2, heteroskedasticity, and more, before we proceed to the multivariate ARIMAX and SARIMAX models. These have been found using the auto_arima function in Python. Please note that in all the following models $d = 0$, because differencing has been applied in advance. Differencing removes the first observation, leaving us with 282 observations.

A typical approach is to partition the data into training and test sets (Hyndman & Athanasopoulos, 2021, ch. 5.8). For all models, data from 13.01.2019 – 10.03.2024 (270 weeks/training points) will be used for training and the data from 17.03.2024 – 02.06.2024 (the last 12 weeks/testing points) will be used for testing. This lets us assess the model's ability to predict future values on unseen data.

5.1 Univariate Models

ARIMA MODELS			
Model	ARIMA	ARIMA	ARIMA
TRANSFORMATION	NO	LOG	BOXCOX
Components	(1,0,2)	(2,0,0)	(2,0,0)
Training Sample (periods)	12	12	12
Observations	270	270	270
AIC	4124,32	-1010,94	-2931,10
BIC	4138,71	-1000,14	-2920,31
Adjusted R-square	0,61	0,63	0,64
Ljung-box (Q)	0,00	0,10	0,02
Prob (Q)	0,97	0,75	0,88
Heteroskedasticity (H)	5,99	1,84	1,27
Prob (H)	0,00	0,00	0,26
Jarque-Bera (JB)	130,56	45,87	92,6
Prob (J-B)	0,00	0,00	0,00
Skew	0,64	0,47	0,75
Kurtosis	6,16	4,79	5,45
COEFFECIENTS & THEIR RESPECTIVE P-VALUES			
ar.L1	0,3541	1,0499	1,0553
P-value	0,0000	0,0000	0,0000
ar.L2		-0,4013	-4,03E-01

P-value		0,0000	0,0000
ma.L1	0,7026		
P-value	0,0000		
ma.L2	0,2654		
P-value	0,0000		
sigma2	2,44E+05	0,0013	1,10E-06
P-value	0,0000	0,0000	0,0000

Table 5.1: The ARIMA models with their respective performance metrics, from left to right: ARIMA (1) with no transformation; ARIMA (2) with log transformation; and ARIMA (3) with Box-Cox transformation

Overall, we consider the ARIMA (3) model to be the most suitable with the lowest AIC and BIC scores of -3069,071 and -3058,145 respectively. It was also the only model with a p-value[2] for heteroskedasticity above the 0,05 mark, indicating homoskedasticity. Additionally, it has the highest Adjusted R^2. The Ljung-Box test (Dritsaki, 2015, p. 742) for all models showed a p-value higher than 0,05, failing to reject the null hypothesis, implying that the residuals are independently distributed and behave as white noise (i.e. no significant autocorrelation). The JB test shows that the models are not normally distributed. The autoregressive lag order terms of 1 and 2 (i.e. ar.L1, ar.L2) explain to what degree the current value of the series relies on its immediate past value and two lags past. These are first positive, then negative, and are statistically significant. The moving average term (i.e. ma.L1, ma.L2) represents the influence of past forecast errors on the current value of the time series, with ma.L1 indicating the error term from the most recent lag and ma.L2 reflects the error two lags ago.

Now we will introduce seasonality into the model (i.e. SARIMA).

SARIMA MODEL	
Model	SARIMA
Components	(1,0,1)(2,0,0)(52)
Variables	ALL
Regional/Global S&D	N/A
Training Sample (periods)	12
Observations	270

[2] We were unable to find information on how the auto_arima function estimated heteroskedasticity; therefore, we entered in the results into a statsmodels ARIMA/SARIMAX manually for verification. In statsmodels, the null hypothesis is heteroskedasticity (https://www.statsmodels.org/dev/generated/statsmodels.tsa.statespace.sarimax.SARIMAXResults.test_heteroskedasticity.html).

MASE	0,95
MAE	0,000630
RMSE	0,0008197
AIC	-2926,31
BIC	-2908,32
Adjusted R-square	0,636
MAPE	171,08 %
Ljung-box (Q)	0,14
Prob (Q)	0,71
Heteroskedasticity (H)	1,24
Prob (H)	0,32
Jarque-Bera (JB)	45,61
Prob (J-B)	0,00
Skew	0,54
Kurtosis	4,69
COEFFECIENTS & THEIR RESPECTIVE P-VALUE	
ar.L1	0,5740
P-value	0,0000
ma.L1	0,4870
P-value	0,0000
ar.S.L52	0,2461
P-value	0,0000
ar.S.L104	0,0784
P-value	0,2900
sigma2	0,0000
P-value	0,0000

Table 5.2: Summary of the SARIMA model

Contrary to our expectations, adding seasonality led to a marginally worse AIC and BIC. Nevertheless, we observe an improvement to the model's explanatory power and that the seasonal autoregressive term (ar.S.L52) is statistically significant. This term sets out to explain the relationship between an observation and its corresponding value from the same point in the previous season. There is a weak, positive seasonal influence, which could be related to the Chinese New Year holiday we mentioned earlier.

5.2 Multivariate Models

For the multivariate models we will utilise different models with only the public data, models with only the private data, and then a combination. Additionally, seasonality will be included in the analysis.

ARIMAX MODELS				
Model	ARIMAX	ARIMAX	ARIMAX	ARIMAX
Components	(2,0,0)	(2,0,0)	(2,0,1)	(2,0,0)
Variables	ALL	ALL	AV_TON & DEMAND	OIL & DOLLAR
Regional/Global S&D	Regional 12	Global 7	Regional 12	N/A
Training Sample (periods)	12	12	12	12
Observations	270	270	270	270
MASE	0,72	0,73	0,75	0,75
MAE	0,00048	0,00049	0,00050	0,00050
RMSE	0,0006417	0,0006613	0,00068021	0,000664959
AIC	-2912,07	-2913,77	-2925,88	-2918,05
BIC	-2886,88	-2888,58	-2904,29	-2900,06
Adjusted R-square	0,620	0,622	0,634	0,621
MAPE	95,34 %	95,40 %	107,72 %	99,90 %
Ljung-box (Q)	0,91	0,92	0,07	0,60
Prob (Q)	0,34	0,34	0,78	0,44
Heteroskedasticity (H)	1,23	1,19	1,31	1,25
Prob (H)	0,32	0,40	0,20	0,29
Jarque-Bera (JB)	67,16	62,06	103,73	56,87
Prob (J-B)	0,00	0,00	0,00	0,00
Skew	0,61	0,57	0,62	0,33
Kurtosis	5,11	5,05	5,77	5,15
COEFFECIENTS & THEIR RESPECTIVE P-VALUES				
DEMAND	-0,0252	-4,78E-06	-1,07E-05	
P-value	0,025	0,397	0,802	
Available Tonnage	-9,36E-06	-1,74E-06	-1,53E-02	
P-value	0,842	0,964	0,145	
OIL	-1,42E-05	-1,46E-05		-1,18E-05
P-value	0,102	0,114		0,172
DOLLAR	1,14E+05	1,16E+05		1,62E+05
P-value	0,000	0,000		0,000
ar.L1	0,9737	0,9836	1,0204	0,9875
P-value	0,0000	0,0000	0,0000	0,0000
ar.L2	-0,3359	-0,3457	-0,3741	-0,3496
P-value	0,0000	0,0000	0,0000	0,0000

ma.L1			0,0341	
P-value			0,8180	
Sigma2	1,14E-06	1,14E-06	1,10E-06	1,22E-06
P-value	0,0000	0,0000	0,0000	0,0000

Table 5.3*:* A summary of ARIMAX models with public and private data, and a combination of the two

Table 5.3 shows that models using only private or public data perform marginally worse in terms of forecasting accuracy than the ARIMAX model incorporating all variables; they both show an improved (lower) MAE, MAPE, and MASE. The similarity between the two likely stems from the fact that a substantial portion of the global data is derived from the regional data from our chosen region.

The ARIMAX R_12 model's DEMAND coefficient shows statistical significance. As Jean-Paul Rodrigue (2010, p. 3) suggests, demand is one of the main driving forces for carriers to have their services integrated into the commodity chains, and that carriers have adopted the 'just-in-time' inventory practise (i.e. demand drives the production of commodities). Additionally, it seems plausible that the regional demand, being in closer proximity to our targeted route, would show statistical significance in comparison to global demand.

SARIMAX MODELS			
Model	SARIMAX	SARIMAX	SARIMAX
Components	(2,0,0)(2,0,0)(52)	(1,0,1)(2,0,0)(52)	(2,0,1)(2,0,0)(52)
Variables	OIL & DOLLAR	AV_TON & DEMAND	ALL
Regional/Global S&D	N/A	Regional 9	Regional 12
Training Sample (periods)	12	12	12
Observations	270	270	270
MASE	0,75	0,88	0,71
MAE	0,00050	0,00059	0,00047
RMSE	0,000698168	0,000757847	0,000679
AIC	-2912,95	-2927,06	-2903,86
BIC	-2887,76	-2901,87	-2867,88
Adjusted R-square	0,622	0,639	0,706
MAPE	118,03%	153,83 %	112,02 %
Ljung-box (Q)	0,03	0,47	0,11
Prob (Q)	0,86	0,49	0,74
Heteroskedasticity (H)	1,29	1,19	1,34
Prob (H)	0,23	0,4	0,17

Jarque-Bera (JB)	41,04	59,56	40,81
Prob (J-B)	0,00	0,00	0,00
Skew	0,52	0,61	0,54
Kurtosis	4,6	4,95	4,57
COEFFECIENTS & THEIR RESPECTIVE P-VALUES			
DEMAND		-0,0002	-0,0252
P-Value		0,5560	0,0350
Available Tonnage		-2,14E-05	-1,31E-05
P-Value		0,2480	0,8050
OIL	-1,10E-05		-1,23E-05
P-Value	0,2440		0,1680
DOLLAR	1,18E+05		1,14E+05
P-Value	0,0000		0,0000
ar.L1	-1,10E-05	0,5616	1,0660
P-value	0,2440	0,0000	0,0000
ar.L2	-0,3496		-0,4036
P-value			0,0010
ma.L1		0,4594	-0,1072
P-value		0,0000	0,5550
ar.S.L52	0,2440	0,1493	0,2511
P-value	0,0010	0,0470	0,0000
ar.S.L104	0,0635	0,0486	0,0663
P-value	0,4590	0,0000	0,4330
Sigma2	1,22E-06	1,07E-06	1,13E-06
P-value	0,0000	0,0000	0,0000

Table 5.4: SARIMAX models with their respective performance metrics, from left to right: SARIMAX (2,0,0)(2,0,0,52) with oil and dollar variables, SARIMAX (1,0,1)(2,0,0)(52) with AVAILABLE_TON and DEMAND variables, and SARIMAX (2,0,1)(2,0,0)(52) with all variables

We will start by only adding the publicly available exogenous variables. The DOLLAR variable with a p-value below the 5% threshold, is statistically significant. In contrast, the OIL, is not statistically significant. This could be due to the fact that regardless of the bunker price, or the price of oil, the shipping sector’s main source for energy is oil, and there are, currently, no quick and cheap fixes to make a vessel run on anything else.

Secondly, we incorporate only AVAILABLE_TON_R_9 and DEMAND_R_9 into the SARIMAX model to assess any correlation, before we move further to the model with all our variables. The DEMAND and AVAILABLE_TON variables have a p-value below the 5% threshold, suggesting the variables are not statistically significant. The resulting model exhibits higher MASE and MAE compared to the model utilising only publicly available data. This suggests that the first model relying solely on publicly available information provides a better forecast by better capturing the underlying patterns of the time series. However, it should be noted that second model is still significantly better than the naïve method in terms of MAPE, MASE, and RMSE with a score of 0,88, 112,02%, and 0,000679, respectively.

Lastly, by including all the variables into the model, the auto_arima function opted for (2,0,1)(2,0,0)(52). Two variables showed statistical significance at the 5% threshold: one public and one private. The DOLLAR was considered significant in both instances, likely due to shipping costs typically being denominated in U.S. dollar, resulting in a positive coefficient to FREIGHT_RATE. DEMAND was significant with all the variables included, showing an inverse correlation. This may be because when DEMAND increases, shipowners respond by reallocating their vessels to the location, thereby increasing available tonnage and lowering FREIGHT_RATES. Combined, they produce the best in-sample MASE result overall at 0,71.

The third SARIMAX model demonstrates the lowest (i.e. most optimal) out-of-sample accuracy, as suggested by the forecasting accuracy metrics, while also having the highest adjusted-R^2. This suggests that the SARIMAX model is superior in both forecasting accuracy and explanatory power compared to the ARIMA, SARIMA, and ARIMAX models discussed in the previous section.

6. Forecasting

In this chapter, we intend to measure the forecast accuracy between the most suitable – based on the given metrics – univariate and multivariate models, namely ARIMA, SARIMA, ARIMAX and SARIMAX. As a benchmark, we will implement the naïve forecast model (Random Walk) on the pre-processed (Box-Cox, differenced, and winsorized) FREIGHT_RATE data.

6.1 Comparison of Models

BENCHMARK: Box-Cox-transformed, differenced, 95% winsorized FREIGHT_RATE endogenous variable

Model	Components	Variables	Regional/Global	MAE	MAPE	MASE	RMSE	Adjusted R^2	AIC	BIC
ARIMA	(2,0,0)	FREIGHT_RATE	N/A	0,00049	103,44 %	0,74	0,00068	0,636	-2931,99	-2920,30
SARIMA	(1,0,1)(2,0,0)(52)	FREIGHT_RATE	N/A	0,00063	171,08 %	0,95	0,00082	0,636	-2926,31	-2908,32
ARIMAX 1	(1,0,1)	ALL	Regional 7	0,00055	114,01 %	0,83	0,00074	0,584	-2892,09	-2866,90
ARIMAX 2	(1,0,1)	ALL	Regional 9	0,00054	118,99 %	0,81	0,00069	0,614	-2911,55	-2886,37
ARIMAX 3	(2,0,0)	ALL	Regional 12	0,00048	95,34 %	0,72	0,00064	0,615	-2912,07	-2886,88
ARIMAX 4	(1,0,1)	ALL	Regional 15	0,00056	124,48 %	0,84	0,00068	0,649	-2901,76	-2876,57
ARIMAX 5	(2,0,0)	ALL	Global 7	0,00049	95,40 %	0,73	0,00066	0,616	-2913,77	-2888,58
ARIMAX 6	(2,0,0)	ALL	Global 9	0,00055	127,01 %	0,83	0,00073	0,590	-2895,56	-2870,37
ARIMAX 7	(2,0,0)	ALL	Global 12	0,00057	143,30 %	0,86	0,00073	0,593	-2897,64	-2872,45
ARIMAX 8	(1,0,1)	ALL	Global 15	0,00057	137,50 %	0,85	0,00070	0,616	-2913,33	-2888,14
ARIMAX 9	(2,0,0)	AV_TON & DEMAND	Regional 7	0,00051	110,05 %	0,76	0,00069	0,632	-2926,20	-2908,20
ARIMAX 10	(2,0,0)	AV_TON & DEMAND	Regional 9	0,00050	108,64 %	0,75	0,00069	0,634	-2927,97	-2909,98
ARIMAX 11	(2,0,1)	AV_TON & DEMAND	Regional 12	0,00050	107,72 %	0,75	0,00068	0,634	-2925,88	-2904,29
ARIMAX 12	(2,0,0)	AV_TON & DEMAND	Regional 15	0,00053	120,63 %	0,80	0,00069	0,619	-2917,35	-2899,36
ARIMAX 13	(2,0,0)	AV_TON & DEMAND	Global 7	0,00052	117,09 %	0,78	0,00069	0,630	-2924,83	-2906,83
ARIMAX 14	(2,0,0)	AV_TON & DEMAND	Global 9	0,00055	129,48 %	0,83	0,00071	0,630	-2925,06	-2907,07
ARIMAX 15	(2,0,0)	AV_TON & DEMAND	Global 12	0,00057	144,47 %	0,86	0,00071	0,633	-2926,95	-2908,96
ARIMAX 16	(2,0,0)	AV_TON & DEMAND	Global 15	0,00054	130,92 %	0,81	0,00069	0,635	-2928,88	-2910,89
ARIMAX 17	(2,0,0)	OIL	N/A	0,00050	111,63 %	0,76	0,00066	0,638	-2930,77	-2916,38
ARIMAX 18	(2,0,0)	DOLLAR	N/A	0,00049	99,93 %	0,74	0,00068	0,629	-2925,21	-2910,82
ARIMAX 19	(2,0,0)	OIL & DOLLAR	N/A	0,00050	99,90 %	0,75	0,00066	0,621	-2918,05	-2900,06

SARIMAX 1	(2,0,0)(2,0,0)(52)	ALL	Regional 9	0,00049	120,38 %	0,74	0,00069	0,620	-2910,74	-2878,35
SARIMAX 2	(2,0,1)(2,0,0)(52)	ALL	Regional 12	0,00047	112,02 %	0,71	0,00068	0,706	-2903,86	-2867,88
SARIMAX 3	(2,0,0)(2,0,0)(52)	ALL	Regional 15	0,00050	121,59 %	0,76	0,00061	0,604	-2900,17	-2867,78
SARIMAX 4	(2,0,0)(2,0,0)(52)	ALL	Global 9	0,00054	123,73 %	0,81	0,00076	0,587	-2888,37	-2855,99
SARIMAX 5	(2,0,0)(2,0,0)(52)	ALL	Global 12	0,00058	143,67 %	0,87	0,00077	0,589	-2889,83	-2857,45
SARIMAX 6	(1,0,1)(2,0,0)(52)	ALL	Global 15	0,00056	150,38 %	0,85	0,00075	0,619	-2910,64	-2878,25
SARIMAX 7	(1,0,1)(2,0,0)(52)	AV_TON & DEMAND	Regional 7	0,96000	173,35 %	0,96	0,00082	0,634	-2922,36	-2897,17
SARIMAX 8	(1,0,1)(2,0,0)(52)	AV_TON & DEMAND	Regional 9	0,88000	153,83 %	0,88	0,00076	0,639	-2927,06	-2901,87
SARIMAX 9	(1,0,1)(2,0,0)(52)	AV_TON & DEMAND	Regional 12	0,92000	165,98 %	0,92	0,00081	0,628	-2918,12	-2892,94
SARIMAX 10	(1,0,1)(2,0,0)(52)	AV_TON & DEMAND	Regional 15	0,98000	177,32 %	0,98	0,00082	0,619	-2911,19	-2886,00
SARIMAX 11	(1,0,1)(2,0,0)(52)	AV_TON & DEMAND	Global 7	0,98000	178,17 %	0,98	0,00083	0,630	-2919,31	-2894,12
SARIMAX 12	(1,0,1)(2,0,0)(52)	AV_TON & DEMAND	Global 9	1,00000	179,61 %	1,00	0,00084	0,628	-2917,96	-2892,71
SARIMAX 13	(1,0,1)(2,0,0)(52)	AV_TON & DEMAND	Global 12	1,02000	186,05 %	1,02	0,00085	0,630	-2919,68	-2894,49
SARIMAX 14	(1,0,1)(2,0,0)(52)	AV_TON & DEMAND	Global 15	0,97000	177,53 %	0,97	0,00083	0,634	-2922,65	-2897,46
SARIMAX 15	(1,0,2)(2,0,0)(52)	OIL	N/A	0,90000	164,80 %	0,90	0,00077	0,643	-2929,43	-2904,25
SARIMAX 16	(2,0,0)(1,0,0)(52)	DOLLAR	N/A	0,77000	124,20 %	0,77	0,00072	0,625	-2924,66	-2903,07
SARIMAX 17	(2,0,0)(2,0,0)(52)	OIL & DOLLAR	N/A	0,75000	118,03 %	0,75	0,00070	0,622	-2912,95	-2887,76

Table 6.1: A tabulated summary of all variations of models we have tested using the auto_arima function

As illustrated by table 6.1, the results of a total of 38 univariate (1 ARIMA, 1 SARIMA) and multivariate (19 ARIMAX, 17[3] SARIMAX) models are compared. The latter models take either public data, private data – with varying degrees of spatial and temporal limitations employed – or a combination of exogenous variables. The MAE, MAPE, MASE, and RMSE are used to compare the forecasting performance of the different models. The AIC, BIC, and adjusted R^2 metrics are included to show the models' goodness of fit and explanatory power.

Overall, the results are mixed. The MAPE scores can be considered to be, across the *table*, very poor. Contrariwise, the MASE and MAE[4] scores reveal how 36 models outperformed the naïve forecast benchmark model; the best one being SARIMAX 2. The same model also has the highest adjusted R^2. The RMSE prefers another model (i.e. SARIMAX 3) to the MASE, though they both select a model with a seasonal autoregressive term.

However, the results are consistent with respect to model characteristics. Models with an exogenous variable or more are favoured by all forecasting metrics. Particularly, models incorporating both public and private data are ranked highest. Additionally, regional, private data generally score better than their global counterpart, with the lone exception of ARIMAX 5.

We also remark the fact that by only observing the results from models using either the public data or the private data and not both ('ALL'), all forecasting accuracy metrics select the ones incorporating public data, specifically ARIMAX 19 (MAPE), ARIMAX 18 (MASE), and ARIMAX 17 (RMSE).

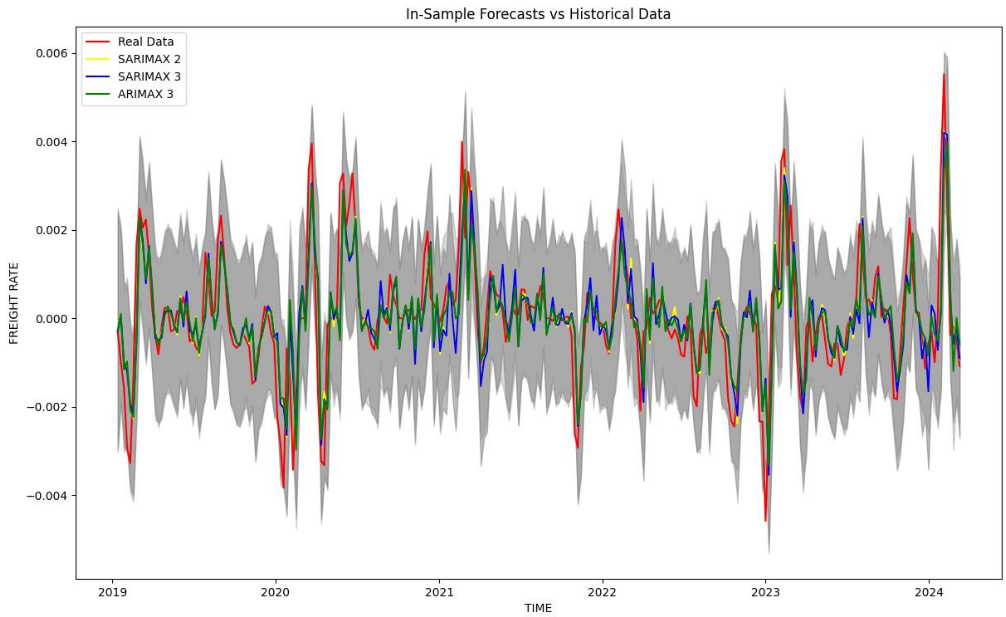
	Highest Ranking based on Metric		
Model	MAPE	MASE	RMSE
ARIMAX 3	1st	2nd	2nd
ARIMAX 5	2nd	3rd	3rd
ARIMAX 19	3rd	7th	5th
SARIMAX 2	10th	1st	7th
SARIMAX 3	17th	13th	1st
ARIMA	5th	5th	9th
SARIMA	32nd	32nd	32nd
ARIMAX 11	6th	8th	7th

[3] For some combinations of SARIMAX, the auto_arima function only returned what was, in fact, an ARIMAX model.
[4] The MAE is not included further as it gives, unsurprisingly, the same selection of models as the MASE.

Table 6.2: Table showcasing the models to be used in forecasting

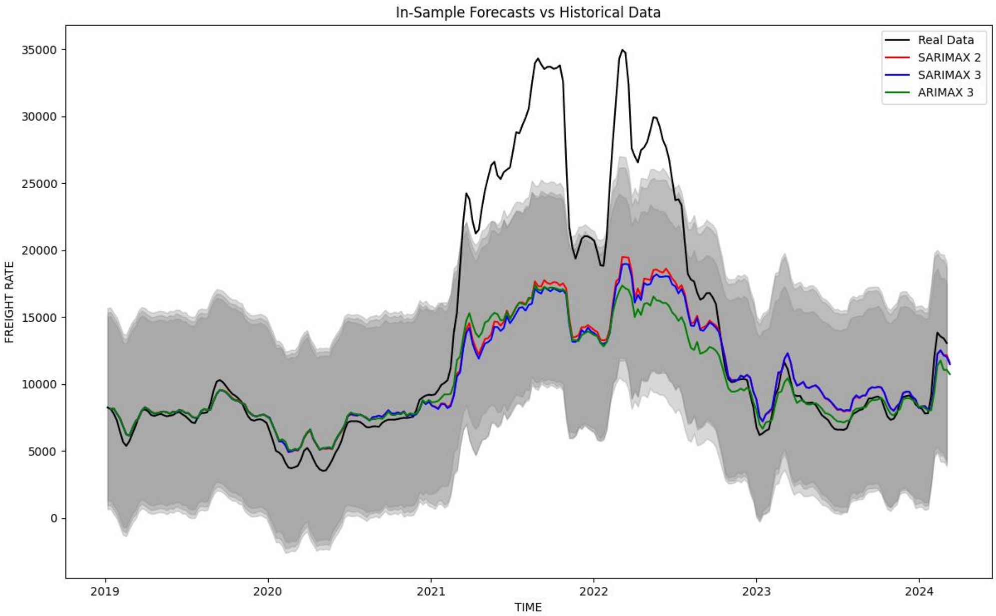
Table 6.2 shows a summary of the highest-ranking models based on each forecasting accuracy metric. To ensure that all model types are represented, the ARIMA and SARIMA models from table 6.1 are included as well. Since ARIMAX 11 is the model that scored the best using only AVAILABLE_TON and DEMAND variables, it has been included too. These we will use in the following sections to show in-sample and out-of-sample forecasting.

6.2 In-Sample Forecast



Graph 6.1: In-sample forecasts on transformed scale versus historical data with 95% confidence interval for SARIMAX 2 in yellow, SARIMAX 3 in blue, and ARIMAX 3 in green

In graph 6.1, the in-sample forecasts are compared to the real data with all three seeming to fluctuate around zero (i.e. they do not consistently increase or decrease over time). Although the in-sample forecasts capture the overall pattern of the data (note the red line), it fails to capture the severity of the fluctuations belonging to the period (i.e. the peaks and troughs in 2020, the troughs in 2022 and 2023, and the peak in 2024).



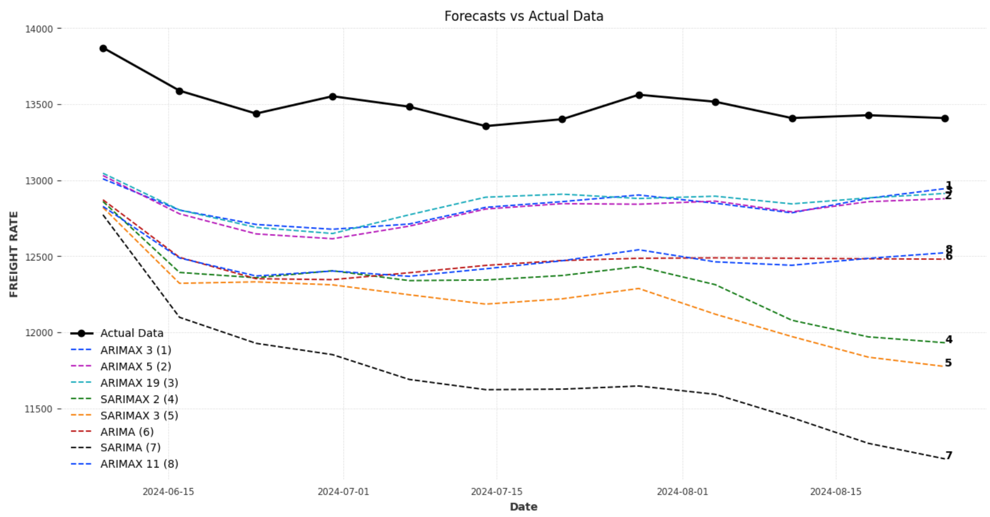
Graph 6.2: in-sample forecasts on the original scale vs real data with 95% confidence interval, all variables included for the SARIMAX 2 in red, SARIMAX 3 in blue, and ARIMAX 3 in green.

For interpretational purposes, we have included the back-transformed in-sample forecasts compared to the real data, where one can better see the difference between the models.

SARIMAX 2 and 3 show better performance in capturing the trend and seasonality of the data compared to the ARIMAX 3. While ARIMAX 3 adequately follows the general trend, visual inspection shows a lower degree of accuracy compared to the real data. For the two first years, all models show similar patterns, however, SARIMAX 2 more effectively captures the magnitude of fluctuations observed between mid-2021 to mid-2022. While in the latter portion of the forecast, the SARIMAX models more effectively captures the time series' volatility, while the ARIMAX 3 underestimates the volatility, suggesting the inclusion of seasonal parameters enhances the forecasting performance.

Similar to graph 6.1, the magnitude of the fluctuations has not been captured, because the sharpest spikes are so large as to fall outside the 95% confidence interval (i.e. the shaded areas). The shaded areas reach values beneath zero, which is unrealistic as in no case would a shipowner pay the charterer to hire his/her vessel. Despite missing out on the short-term movements, the long-term trends have been incorporated well.

6.3 Final Forecasts

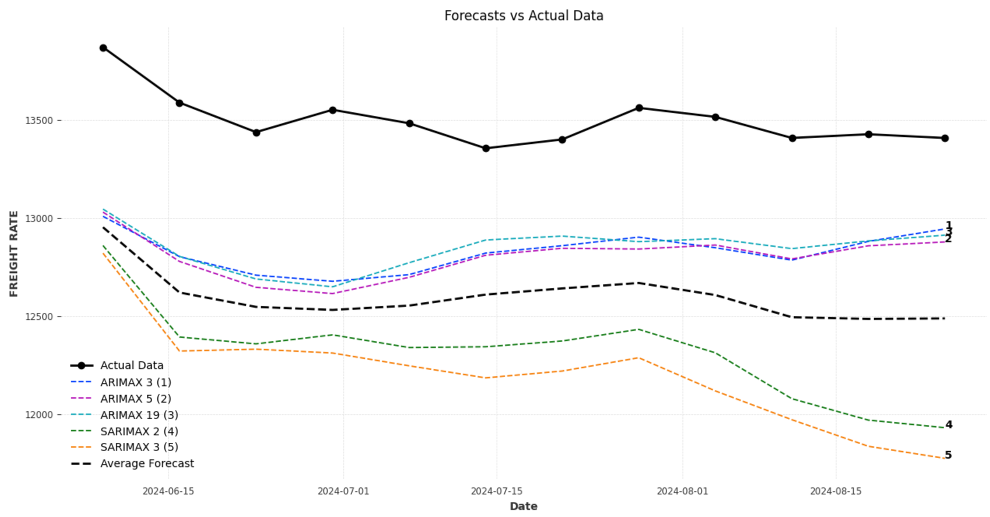


Graph 6.3: Eight models' actual out-of-sample forecast versus actual data

Graph 6.3 shows the eight selected models from table 6.2 beneath a dark dotted line representing the actual data. Each line's number corresponds to the parenthesised number behind each model in the legend, placed bottom left.

The actual data exhibits a thrice-repeated decrease lasting for two periods (i.e. weeks) at a time: first a two-week decrease followed by a short one-week upturn; then another two-week decrease followed by a two-week increase; and finally a two-week decline before it stabilises at just beneath $13 500.

The models, in general, manage to emulate the fluctuations in the actual data. Lines (1) – (3) follow the trends, but fail to capture short-term movements. Lines (4) and (5) are relatively similar and better mimic the upturns. Lines (6) and (8) are also similar, but the former is an ARIMA model that becomes almost a straight line after first three weeks, and the latter is ARIMAX 11 that better captures the turns. Line (7) fails to capture anything but the initial drop.



Graph 6.4: The most promising five models from table 6.1 based on the forecast accuracy criteria and their combined average

Graph 6.4 shows the same first five models included in table 6.2 and the previous graph with the average of the predictions marked as the black, dashed line that fall in between the forecasts.

Only lines (4) and (5) successfully capture the upturns, but like the other lines fall off between the seventh and eighth week. The averaged forecast, basing itself on the remaining models, follow the trends well.

7. Discussion

This chapter discusses the results from the preceding chapters in light of what we set out to examine: can incorporating private data improve the forecasting models' accuracy in terms of forecasting the targeted HS7_38 route, and can this give indication of inefficiency in freight rates.

Table 5.1 can provide insights into answering the above. The results showed that incorporating regional available tonnage and demand variables provided increased forecasting accuracy, notably from the MASE. DEMAND and DOLLAR exhibited a statistically significant relationship at the 5% threshold with FREIGHT_RATE, suggesting their importance. While AIC and BIC decreased, goodness of fit will not necessarily translate into (more) accurate forecasts (Hyndman & Athanasopoulos, 2021, ch. 5.8).

Moreover, table 6.1 results showed how all forecasting accuracy metrics favoured models that incorporated both private and public data. The MAPE scores were, however, unconvincing, which could give rise to the notion that our results are inconclusive. Notwithstanding that MAPE is widely used and we have included it in our table, there are too many drawbacks related to it to give any credence to its results: (1) being a percentage, then at $Y_t = 0$ for any t in the particular period, MAPE is either infinite or undefined; (2) when any value of Y_t is close to zero, the distribution will become extremely skewed; (3) it is scale-sensitive; and (4) it is biased, favouring models that under-forecast the data and is neither minimised by mean nor median (Hyndman & Koehler, 2006, p. 683; Svetunkov, 2023, ch. 2.1). Therefore, it is better to rely on the MASE and RMSE results, which ranked SARIMAX 2 and 3 at top, and ARIMAX models with all variables included at second and third place, as shown in table 6.2.

In continuation of the preceding paragraph, the fact that SARIMAX and ARIMAX models rank at the top is interesting. It suggests that there is seasonality in the data, which we noticed in graph 4.4 and noted in subsubsection 4.3.1, and attributed to the Chinese New Year holiday because of China's prominence in the dry bulk market and proximity to our route. Warehouses stock up and temporarily halt imports in advance. Then, as the holiday ends, imports (demand) rebounds. Although their paper relates to the container shipping freight rate market, Yin and Shi (2018, p. 171) point out that trading volumes are at their lowest in January and February (i.e. when Chinese New Year typically falls) and that factories can keep closed up until March, leading to a decrease in freight rates. Additionally, the fact that these models all contain exogenous variables illustrate the point that multivariate models usually produce better forecasts, as suggested by Cullinane (1992, p. 92).

Out of the 38 models included in table 6.1, 36 models beat the naïve forecast method, of which 28 models used the email circulars datasets. If we disregard time-varying risk premia, these results indicate that freight rates are, in general, inefficient, at least in the short run. This is in line with the empirical evidence on the EMH in freight rates being, on balance, unsupportive of it, and that freight rates are thus inefficient (Karakitsos et al., 2014, pp. 126, 132). There are several theories as to this inefficiency in the weak type of market efficiency, for instance that economic agents do not fully comprehend the ramifications of new information; irrationality; heterogenous expectations; differing forecasts and understanding of the market participants; that a lag period exists before information is absorbed into the market, and more (Karakitsos et al., 2014, p. 120; Malkiel, 2003, p. 33; Shostak, 1997, p. 29).

To further contextualise these findings, in table 6.1, from the perspective of public data versus private data, the three metrics for forecasting accuracy all opted for the models leveraging only public data. This is suggestive of the relevance and informativeness of the public data, in particular our choice of variables, in our forecast. Considering that they are also much easier to obtain and require less effort in terms of filtering, they are a viable alternative to private data.

Graphs 6.1 and 6.2 illustrate the in-sample forecast; graphs 6.3 and 6.4 illustrate the out-of-sample forecast. In both cases, the best models catch the overall trend well, but fail to fully incorporate the sharper movements. The period in question exhibits marked volatility, impacting our results. The volatility in the original datasets were described in subsubsections 4.2.1 and 4.2.2, and shown in graphs 4.1 and 4.3. Major geopolitical events occurred during that time, e.g. the COVID-19 pandemic (lockdowns, port congestion, etc.) and Russia cutting off gas to Europe in 2022 and launching the invasion of Ukraine the same year. World Health Organization (n.d.) states that the pandemic outbreak started December 2019 and it was on May 5th, 2023, declared to no longer be a public health emergency of international concern. These times roughly correspond to when FREIGHT_RATE, OIL, and DOLLAR started exhibiting clear trends and volatility, and when they eventually stabilised. FREIGHT_RATE was particularly volatile from 2021-2023. A higher forecasting accuracy could have been achieved, we believe, were it not for the training set that featured this pronounced volatility.

However, in extenuation of the above, we cannot expect future periods not to be volatile either. It seems to be the case then that the usefulness of not only the public data but also the email circular datasets in forecasting is dependent on the particular times in which they are employed.

8. Conclusion

The purpose of our research is to determine if using private datasets, specifically email circulars for available tonnage (supply) and cargo orders (demand), in forecasting the targeted BHSI HS7_38 route, Far East to Southeast Asia region, can improve forecasts. In extension of that, we try to assess if freight rates are efficient. We have also used datasets for Brent crude oil prices and the Nominal Broad U.S. Dollar Index that are publicly available online.

While prior research exists into the topic of forecasting freight rates, private data have predominantly not been included. Previous research on efficiency in freight rates mainly regard time-varying risk premia and not actual testing for market efficiency (Karakitsos et al., 2014, p. 99). This has left gaps in the literature that this thesis endeavours to bridge.

The first part of our thesis explores the original dataset (see subsection 4.2) and applying transformations to our data. We took note of the different pronounced trends and volatility. To remedy them, we aggregated the data into weekly data and found cause to believe that the Box-Cox transformation (and differencing) coupled with winsorization was the most suited to our dataset (see subsection 4.3). This was a necessary part to establish a benchmark and be able to assess the purpose of this thesis.

In the subsequent part, we analysed and compared the univariate and multivariate models (see chapter 5). We discovered in our model comparison in table 5.1 that the Box-Cox transformation reduced skewness and kurtosis while also being the only model to exhibit homoskedasticity.

In the penultimate part, chapter 6, we compared the different variations of multivariate models using public data and/or private data based on the forecasting accuracy metrics MAPE, MASE, and RMSE. We found that out of the 38 models included in table 6.1, 36 models beat the naïve forecast method, of which 28 models used the email circulars datasets. The models with the most positive scores in terms of the aforementioned metrics were then used in further forecasting. We found that the SARIMAX models gave the most accurate forecasts, followed closely by the ARIMAX models.

In the last part, we discussed our findings, which showed that models with all variables (available tonnage, demand, oil, and dollar datasets) included were the favourites based on the metrics used. We then discussed the reasons as to why freight rates might be inefficient.

In concluding, using email circulars as predictive signals for available tonnage (supply) and demand, based on our results and metrics, does improve forecasting accuracy. Furthermore, these results also give indication of inefficiency in freight rates as most of the models provided results that surpassed those of a naïve benchmark model. Our research contributes to the topic of forecasting freight rates, especially in the dry bulk market, and participates in laying the groundwork for further studies incorporating private datasets and testing efficiency in markets.

8.1 Limitations and Further Research

During our research, we have become aware of some limitations that could have a bearing on our results.

First, our analysis focuses solely on a single shipping route, whereas signals from all routes within this vessel class, as well as from neighbouring vessel classes, may exert significant influence on freight rate developments.

Second, the methodology is limited to single-equation models, which fail to incorporate potential cross effects or interactions between the variables. Using more advanced models, one can capture how changes in one variable affect others, providing a more comprehensive analysis.

Third, more advanced analytical methods have not been utilised, which could provide deeper insights into the data and improve the robustness of our results. Utilising more advanced models may account for the large fluctuations in seasonality caused by extraordinary geopolitical events that this study could not capture.

Further research could employ more advanced analytical methods like machine learning models or multi-equation models. This may lead to better results that capture the underlying patterns in the data. Another avenue of further research could be to expand the scope to include the adjacent routes and examine the interactive effects between them and the Handysize vessel signals as well as other, neighbouring vessel classes' signals. Additionally, further research could factor in particular geopolitical events to minimise their potential disruptive impact on the forecasting results.

9. AI Decleration

Name and version of the AI-tool: ChatGPT, 4.0.

The purpose behind the use of ai-tool:

1. Brainstorming;
2. To assist with coding, both in writing and troubleshooting;
3. As a Thesaurus.

We are aware that we are responsible for all content of this master's thesis, including the parts where AI tools are used. We are responsible for ensuring that the thesis complies with ethical rules for privacy and publication.

10. Bibliography

Acharyya, R. (2023). 'Gains from Trade'. In *International Economics: An Introduction to Theory and Policy*. Oxford University Press. https://doi.org/10.1093/oso/9780192865144.003.0002

Akatsuka, K., & Leggate, K. (2001). Perceptions of foreign exchange rate risk in the shipping industry. *Maritime Policy & Management*, *28*(3), 235-249. https://doi.org/https://doi.org/10.1080/03088830110055684

Berry, W. D. (1985). *Multiple regression in practice*. Sage.

Bertaut, C., Beschwitz, B. v., & Curcuru, S. (2023, June 23, 2023). *"The International Role of the U.S. Dollar" Post-COVID Edition*. Federal Reserve Board. https://www.federalreserve.gov/econres/notes/feds-notes/the-international-role-of-the-us-dollar-post-covid-edition-20230623.html

Betzer, A., & Theissen, E. (2009). Insider trading and corporate governance: The case of Germany. *European Financial Management*, *15*(2), 402-429.

Board, F. R. (2024, November 18, 2024). *Foreign Exchange Rates - H.10*. FRB. https://www.federalreserve.gov/releases/h10/summary/

Box, G. E. P., & Jenkins, G. M. (1970). *Time Series Analysis Forecasting and Control*. Holden Day.

Chen, S., Meersman, H., & Voorde, E. v. d. (2012). Forecasting spot rates at main routes in the dry bulk market. *Maritime Economics & Logistics*, *14*(4), 498-537. https://doi.org/10.1057/mel.2012.18

Clarksons Research. (n.d.-a, 04.12.2024). *542365: BHSI7_38: N.China - S.Korea - Japan trip to SE Asia*. Clarksons Research. https://www.clarksons.net/n/#/sin/timeseries/browse;e=%5B1716,10312,11016,79544%5D;c=%5B29245%5D/(ts:data/100/latest;t=%5B542365%5D;l=%5B542365%5D;listMode=false;viewMode=Grid)

Clarksons Research. (n.d.-b). *Handysize Orderbook % Fleet*. Clarksons Research. https://www.clarksons.net/n/#/sin/timeseries/browse;e=%5B1716,10312,11020,71099%5D;c=%5B43782%5D/(ts:data/100/Latest;t=%5B534514%5D;l=%5B534454%5D;exportAsPNG=false;listMode=false;viewMode=Chart)

Clarksons Research. (n.d.-c). *Maritime & Shipping Research Services*. https://www.clarksons.com/research/

Cullinane, K. (1992). A short-term adaptive forecasting model for BIFFEX speculation: a Box—Jenkins approach. *Maritime Policy & Management*, *19*(2), 91-114. https://doi.org/https://doi.org/10.1080/03088839200000018

Degutis, A., & Novickytė, L. (2014). The efficient market hypothesis: A critical review of literature and methodology. *Ekonomika*, *93*(2), 7-23.

Dickey, D. A., & Fuller, W. A. (1979). Distribution of the Estimators for Autoregressive Time Series With a Unit Root. *Journal of the American Statistical Association*, *74*(366), 427-431. https://doi.org/10.2307/2286348

Dritsaki, C. (2015). Box-Jenkins modeling of Greek stock prices data. *International Journal of Economics and Financial Issues*, *5*(3), 740-747.

Federal Reserve Economic Data. (n.d.-a, 02.12.2024). *Nominal Broad U.S. Dollar Index (DTWEXBGS)*. FRED. https://fred.stlouisfed.org/series/DTWEXBGS

Federal Reserve Economic Data. (n.d.-b). *What is FRED?* FRED. https://fredhelp.stlouisfed.org/fred/about/about-fred/what-is-fred/

Feng, C., Wang, H., Lu, N., Chen, T., He, H., Lu, Y., & Tu, X. M. (2014). Log-transformation and its implications for data analysis. *Shanghai Arch Psychiatry*, *26*(2), 105-109. https://doi.org/10.3969/j.issn.1002-0829.2014.02.009

Geman, H., & Smith, W. O. (2012). Shipping markets and freight rates: an analysis of the Baltic Dry Index. *Journal of Alternative Investments*, *15*(1), 98-109.

Hyndman, R. J. (2010). *The ARIMAX model muddle*. Rob J Hyndman. https://robjhyndman.com/hyndsight/arimax/

Hyndman, R. J., & Athanasopoulos, G. (2018). *Forecasting: principles and practice* (2nd ed.). OTexts: Melbourne, Australia. https://otexts.com/fpp2/

Hyndman, R. J., & Athanasopoulos, G. (2021). *Forecasting: principles and practice* (3rd ed.). OTexts: Melbourne, Australia. https://otexts.com/fpp3/

Hyndman, R. J., & Koehler, A. B. (2006). Another look at measures of forecast accuracy. *International Journal of Forecasting*, *22*(4), 679-688. https://doi.org/https://doi.org/10.1016/j.ijforecast.2006.03.001

Institute of Chartered Shipbrokers. (2023a). *Dry Cargo Chartering*. ICS. www.ics.org.uk

Institute of Chartered Shipbrokers. (2023b). *Introduction to Shipping*. ICS. www.ics.org.uk

International Chamber of Shipping. (n.d.). *Shipping and World Trade: World Seaborne Trade*. ICS. https://www.ics-shipping.org/shipping-fact/shipping-and-world-trade-world-seaborne-trade/

Jing, L., Marlow, P. B., & Hui, W. (2008). An analysis of freight rate volatility in dry bulk shipping markets. *Maritime Policy & Management*, *35*(3), 237-251.

Karakitsos, E., Varnavides, L., Karakitsos, E., & Varnavides, L. (2014). The Efficiency of Shipping Markets. *Maritime Economics: A Macroeconomic Approach*, 95-152.

Kwiatkowski, D., Phillips, P. C., Schmidt, P., & Shin, Y. (1992). Testing the null hypothesis of stationarity against the alternative of a unit root: How sure are we that economic time series have a unit root? *Journal of econometrics*, *54*(1-3), 159-178.

Liu, T., Liu, S., & Shi, L. (2020). ARIMA Modelling and Forecasting. In *Time Series Analysis Using SAS Enterprise Guide* (pp. 61-85). Springer Singapore. https://doi.org/10.1007/978-981-15-0321-4_4

Malkiel, B. G. (2003). The efficient market hypothesis and its critics. *Journal of economic perspectives*, *17*(1).

Mushtaq, R. (2011). Augmented Dickey Fuller Test. *SSRN*, 19. https://doi.org/http://dx.doi.org/10.2139/ssrn.1911068

Neath, A. A., & Cavanaugh, J. E. (2012). The Bayesian information criterion: background, derivation, and applications. *Wiley Interdisciplinary Reviews: Computational Statistics*, *4*(2), 199-203.

Nkongolo, M. (2023). Using ARIMA to Predict the Growth in the Subscriber Data Usage. *Eng*, *4*(1), 92-120. https://www.mdpi.com/2673-4117/4/1/6

Osborne, J. (2010). Improving your data transformations: Applying the Box-Cox transformation. *Practical Assessment, Research, and Evaluation*, *15*(1).

Petropoulos, F., Apiletti, D., Assimakopoulos, V., Babai, M. Z., Barrow, D. K., & Taieb, S. B., ... & Ziel, F, et al. (2022). Forecasting: theory and practice. *International Journal of Forecasting*, *38*(3).

Rodrigue, J.-P. (2010). *Maritime transportation: drivers for the shipping and port industries*.

Shostak, F. (1997). In defense of fundamental analysis: A critique of the efficient market hypothesis. *The Review of Austrian Economics*, *10*(2), 27-45.

Stopford, M. (2009). *Maritime Economics* (3 ed.). Routledge.

Svetunkov, I. (2023). *Forecasting and Analytics with the Augmented Dynamic Adaptive Model (ADAM)* (C. a. Hall/CRC, Ed. 1 ed.). https://doi.org/https://doi.org/10.1201/9781003452652

The Baltic Exchange. (n.d.). *Who We Are*. The Baltic Exchange. https://www.balticexchange.com/en/who-we-are.html

Timmermann, A., & Granger, C. W. (2004). Efficient market hypothesis and forecasting. *International Journal of Forecasting*, *20*(1), 15-27.

Ţiţan, A. G. (2015). The efficient market hypothesis: Review of specialized literature and empirical research. *Procedia Economics and Finance*, *32*, 442-449.

U.S. Energy Information Administration. (n.d.-a). *ABOUT EIA*. EIA. https://www.eia.gov/about/

U.S. Energy Information Administration. (n.d.-b, 04.12.2024). *Weekly Europe Brent Spot Price FOB (Dollars per Barrel)*. EIA. https://www.eia.gov/dnav/pet/hist/LeafHandler.ashx?n=PET&s=RBRTE&f=W

United Nations Conference on Trade and Development. (n.d.). *Review of Maritime Transport 2024*. UNCTAD. https://unctad.org/publication/review-maritime-transport-2024

Veson Nautical. (n.d.). *Shipfix*. Veson Nautical. https://veson.com/shipfix-veson/

Willige, A. (2024). *These are the world's most vital waterways for global trade*. World Economic Forum. https://www.weforum.org/stories/2024/02/worlds-busiest-ocean-shipping-routes-trade/

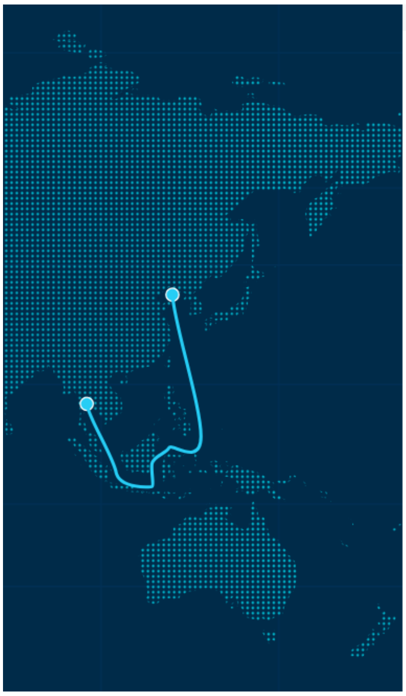
Willmott, C. J., & Matsuura, K. (2005). Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance. *Climate research*, *30*(1), 79-82.

World Health Organization. (n.d.). *Coronavirus disease (COVID-19) pandemic*. WHO. https://www.who.int/europe/emergencies/situations/covid-19

Yilmazkuday, H. (2024). Geopolitical Risks and Shipping. *Social Science Research Network (SSRN)*.

Yin, J., & Shi, J. (2018). Seasonality patterns in the container shipping freight rate market. *Maritime Policy & Management*, *45*(2), 159-173.

11. Appendix



HS7

Route

Delivery North China-South Korea-Japan range, laydays/cancelling 5/10 days from index date, for a 25-30 day trip, redelivery Krabi-Campha range including Malaysia, Indonesia & Philippines. Basis the Baltic handysize vessel. 5% total commission.

Baltic Handysize 38 vessel for Timecharter routes is a non-scrubber fitted vessel based on the following description:

Singledeck self trimming geared bulk carrier
38,200mt dwt on 10.538m SSW
Max Age 15 Years
LOA 180m / Beam 29.8m / TPC 49
47,125 cbm grain / 45,300 cbm bale
5 holds / 5 hatches
4 x 30 ton cranes
Speed & consumptions including main engine & auxibrary engines:
14 knots on 26mt IFO (380 CST) laden/24mt IFO (380 CST) ballast + 0.1 MDO at sea
12 knots on 18mt IFO (380 CST) laden/17mt IFO (380 CST) ballast + 0.1 MDO at sea

Image 11-1: Description of vessel type used for BHSI. Taken from: https://www.balticexchange.com/en/data-services/routes.html.

Handysize Index

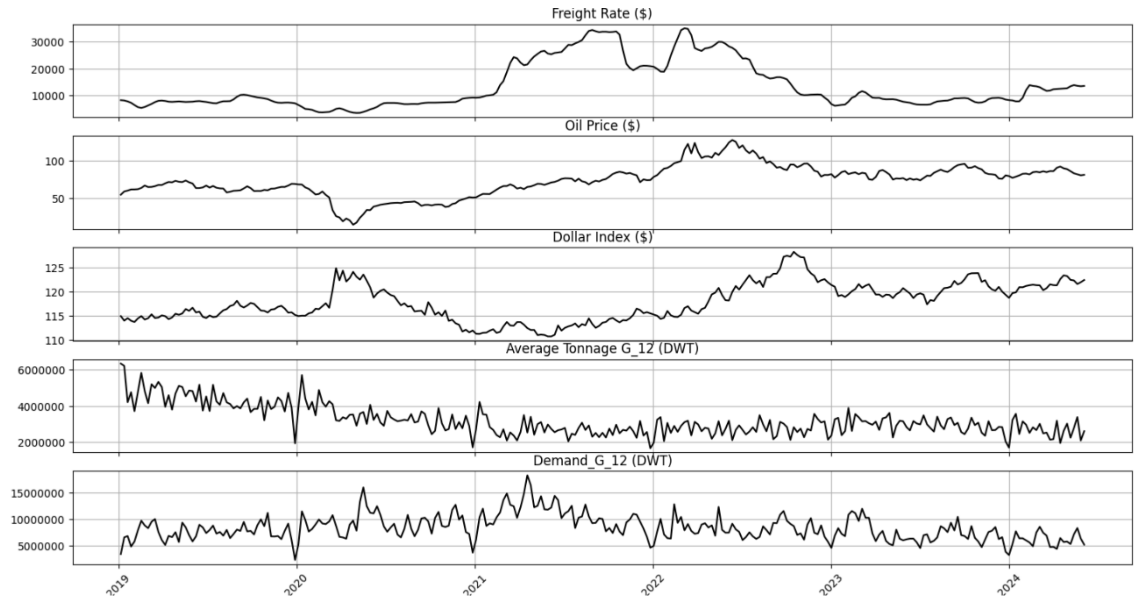
HS1_38 - Skaw-Passero trip to Rio de Janeiro-Recalada
HS2_38 - Skaw-Passero trip to Boston-Galveston
HS3_38 - Rio de Janeiro-Recalada trip to Skaw-Passero
HS4_38 - US Gulf trip via US Gulf or north coast South America to Skaw-Passero
HS5_38 - South East Asia trip to Singapore-Japan
HS6_38 - North China-South Korea-Japan trip to North China-South Korea-Japan
HS7_38 - North China-South Korea-Japan trip to south east Asia

7TC Weighted Time Charter Average: Sum(HS1_38*0.125, HS2_38*0.125, HS3_38*0.125, HS4_38*0.125, HS5_38*0.20, HS6_38*0.20, HS7_38*0.10)

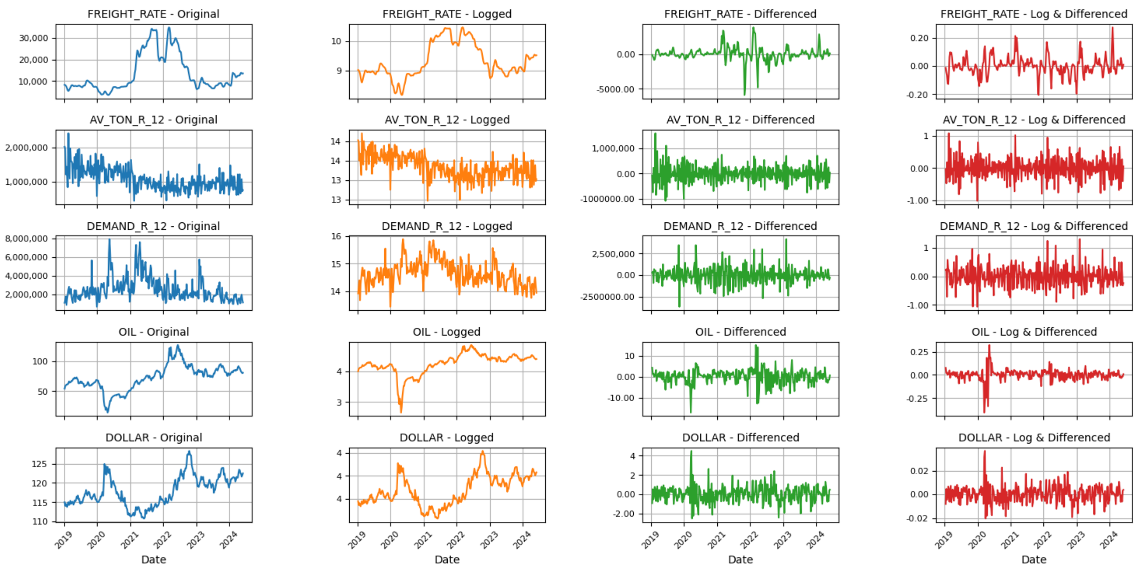
Baltic Handysize Index (BHSI): Sum(HS1_38*0.006944444, HS2_38*0.006944444, HS3_38*0.006944444, HS4_38*0.006944444, HS5_38*0.011111111, HS6_38*0.011111111, HS7_38*0.005555556) Baltic Handysize 38 vessel for Timecharter routes is a non-scrubber fitted vessel based on the following description:

Singledeck self trimming geared bulk carrier
38,200mt dwt on 10.538m SSW
Max Age 15 Years
LOA 180m / Beam 29.8m / TPC 49
47,125 cbm grain / 45,300 cbm bale
5 holds / 5 hatches
4 x 30 ton cranes
Speed & consumptions including main engine & auxibrary engines:
14 knots on 26mt IFO (380 CST) laden/24mt IFO (380 CST) ballast + 0.1 MDO at sea
12 knots on 18mt IFO (380 CST) laden/17mt IFO (380 CST) ballast + 0.1 MDO at sea

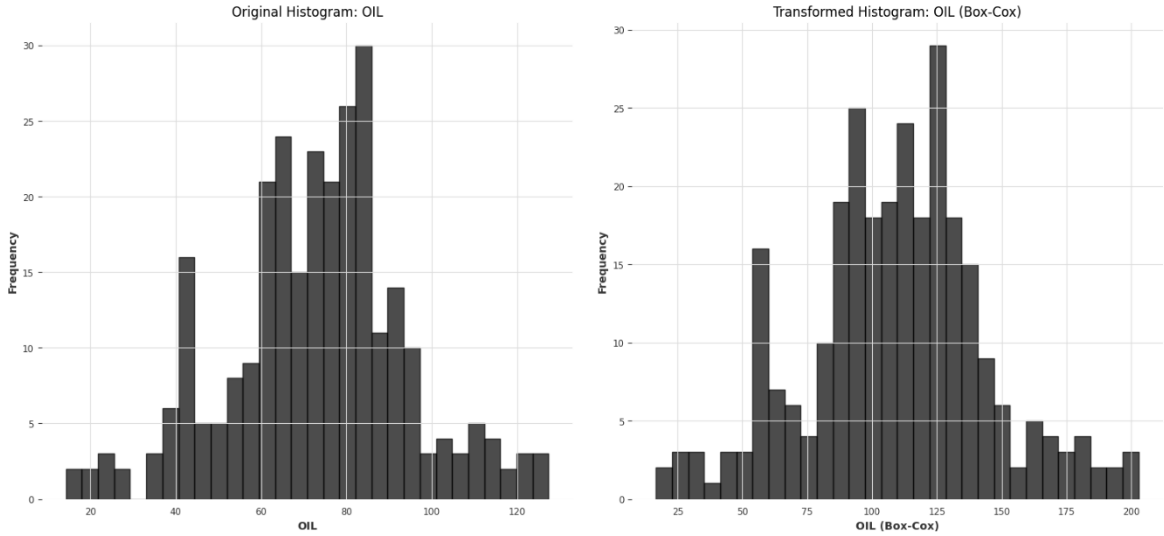
Image 11-2: The weightings of the different routes comprising the Baltic Handysize Index as well as vessel specifications for the index. Taken from: https://www.balticexchange.com/en/data-services/market-information0/dry-services.html.



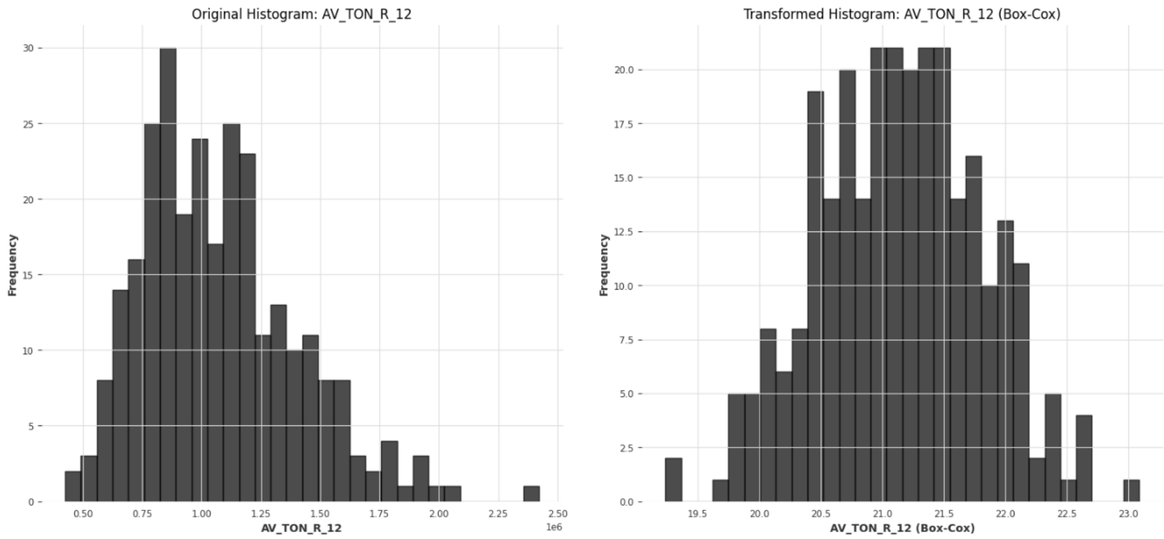
Graph 11.1*:* Graph showing weekly global available tonnage and demand, using a 12 days' limit for fixtures



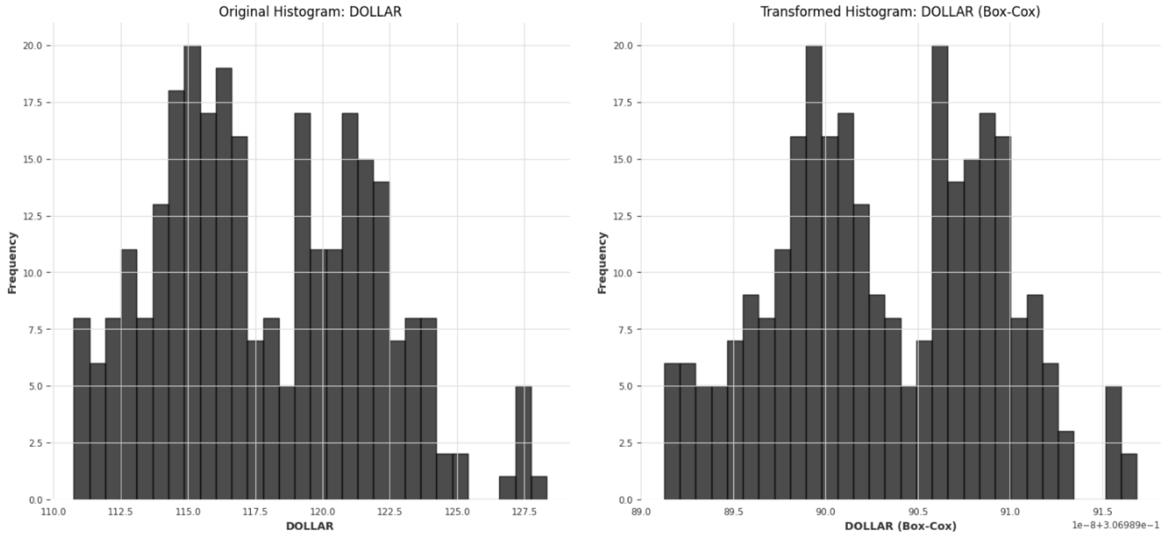
Graph 11.2: The graph shows the different variables subjected to log transformation and differencing, then combined



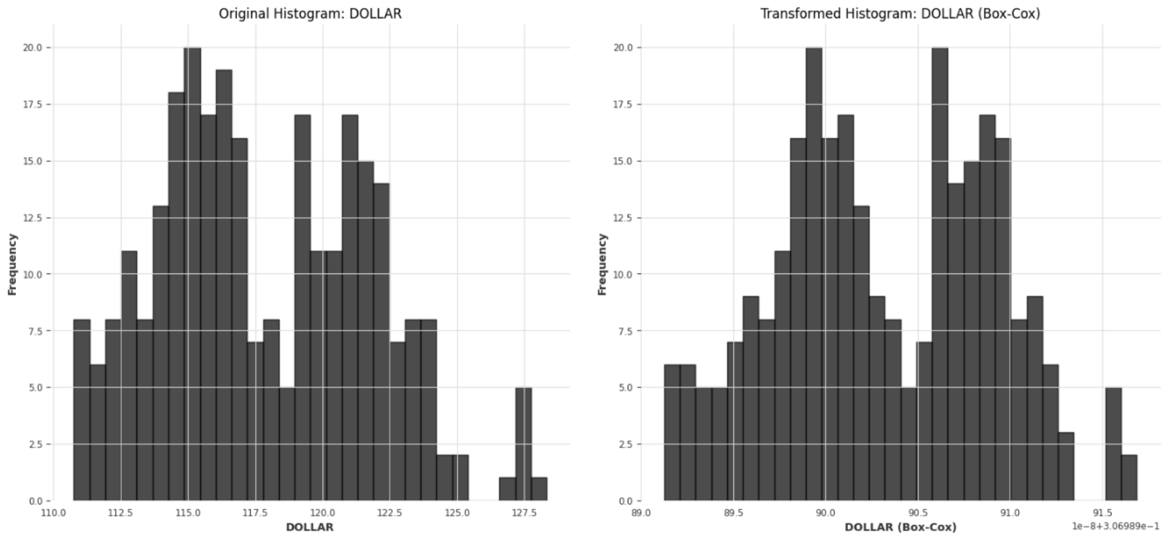
Histogram 11.1: OIL before and after the Box-Cox transformation



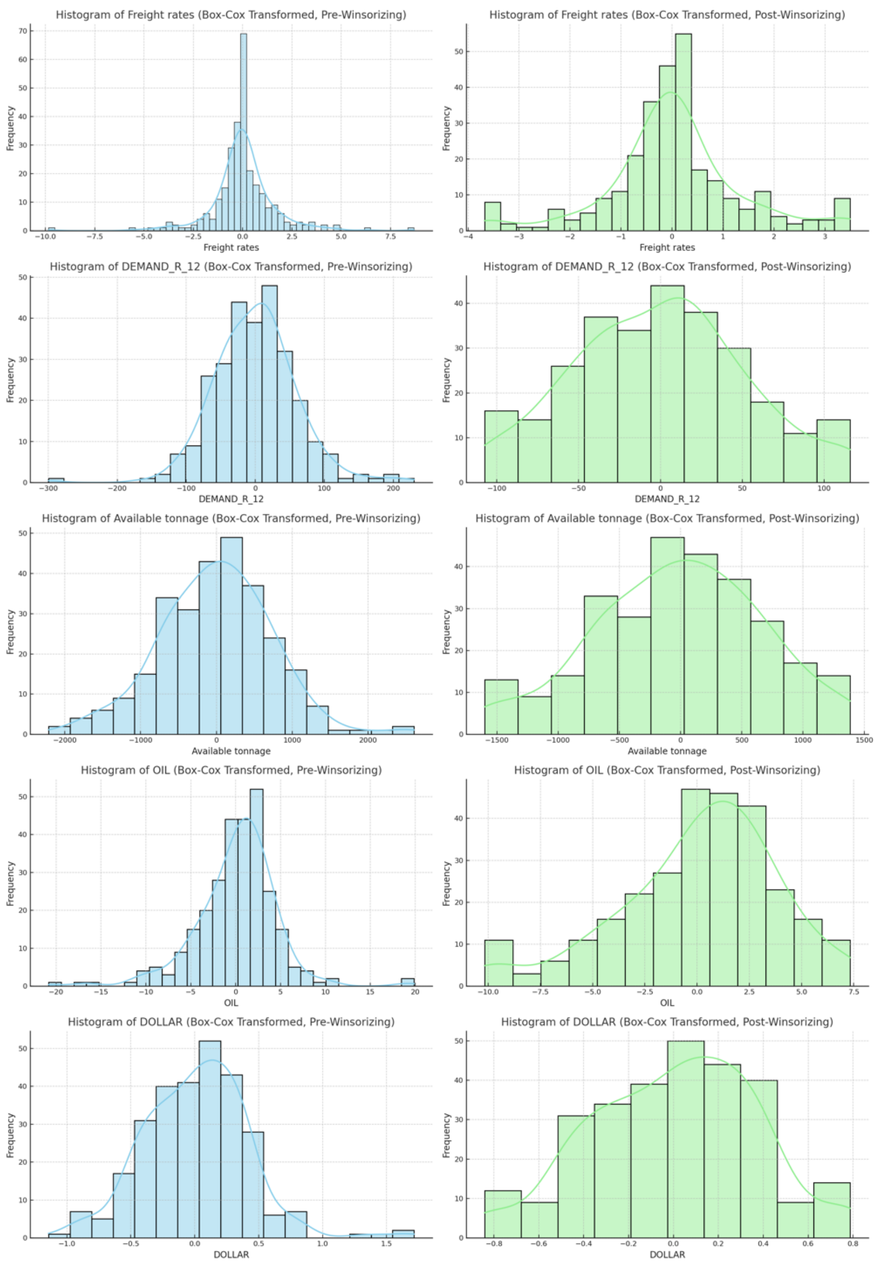
Histogram 11.2: AVAILABLE_TONNAGE before and after the Box-Cox transformation



Histogram 11.3: The Freight rate before and after the Box-Cox transformation



Histogram 11.4*:* DOLLAR before and after the Box-Cox transformation



Histogram 11.5: Histogram and density plot of Box-Cox transformed and differenced variables (pre- and post-winsorization)

	count	mean	std	min	max	Skewness	Kurtosis	J-B
FREIGHT_RATE_raw	283	13325,76	8631,041	3526	34966,4	1,102642	-0,11897	57,51302
OIL_raw	283	72,91618	21,12886	14,24	127,4	-0,12369	0,271296	1,589473
DOLLAR_raw	283	117,846	3,910295	110,76	128,33	0,28319	-0,6595	8,911211
AVAILABLE_TON_R_7_raw	283	381511,3	179198,3	60900	1043624	0,881239	0,773659	43,68668
AVAILABLE_TON_G_7_raw	283	1215821	421819,6	294400	2763464	0,793731	0,281417	30,64925
AVAILABLE_TON_R_9_raw	283	647026,8	249722,8	218000	1737149	0,869159	0,959464	46,48652
AVAILABLE_TON_G_9_raw	283	1977780	608823,4	720199	4180539	0,948302	0,897271	51,90933
AVAILABLE_TON_R_12_raw	283	1072923	326460,8	425600	2422976	0,767741	0,707416	33,70226
AVAILABLE_TON_G_12_raw	283	3197691	824594,1	1669599	6343489	1,051809	1,179602	68,58825
AVAILABLE_TON_R_15_raw	283	1478546	395765	447700	2958541	0,704081	0,72509	29,58149
AVAILABLE_TON_G_15_raw	283	4339335	999600,6	2498193	8550050	1,205615	1,57712	97,88664
DEMAND_R_7_raw	283	983371,5	598438,8	105998	3377561	1,595166	2,835506	214,8243
DEMAND_G_7_raw	283	3395150	1180329	845107	7259172	0,776409	0,602681	32,71558
DEMAND_R_9_raw	283	1566357	859719,5	300254	5137336	1,574646	2,923043	217,7003
DEMAND_G_9_raw	283	5376384	1683493	1240144	12124776	0,734857	0,848334	33,9568
DEMAND_R_12_raw	283	2552971	1284216	691080	7897845	1,572517	2,627144	198,019
DEMAND_G_12_raw	283	8422705	2423810	2331568	18337213	0,738384	1,025248	38,11042
DEMAND_R_15_raw	283	3508372	1661010	898479	10435164	1,624169	2,917141	224,7657
DEMAND_G_15_raw	283	11316085	3128877	3235245	22768600	0,830521	0,939134	42,93385

Table 11.1: Descriptive statistics of all original (raw) variables, both global and regional

ADF and KPSS tests for stationarity

Variable	Transformation	ADF Statistic	P-value	ADF Stationarity	KPSS Statistic	P-value	KPSS Stationarity
BHSI	Original	-1,7287	0,4164	Non-Stationary	0,5217	0,0368	Non-Stationary
OIL	Original	-1,6806	0,4412	Non-Stationary	1,3442	0,0100	Non-Stationary
DOLLAR	Original	-1,8006	0,3801	Non-Stationary	1,0305	0,0100	Non-Stationary
DEMAND_R_7	Original	-1,8494	0,3562	Non-Stationary	0,6996	0,0136	Non-Stationary
DEMAND_R_9	Original	-1,6953	0,4335	Non-Stationary	0,6772	0,0156	Non-Stationary
DEMAND_R_12	Original	-3,8246	0,0027	Stationary	0,5886	0,0237	Non-Stationary
DEMAND_R_15	Original	-3,7707	0,0032	Stationary	0,5205	0,0371	Non-Stationary
DEMAND_G_7	Original	-2,0201	0,2778	Non-Stationary	0,7598	0,0100	Non-Stationary
DEMAND_G_9	Original	-1,7820	0,3894	Non-Stationary	0,7632	0,0100	Non-Stationary
DEMAND_G_12	Original	-2,0769	0,2539	Non-Stationary	0,6707	0,0162	Non-Stationary
DEMAND_G_15	Original	-2,2725	0,1810	Non-Stationary	0,5652	0,0270	Non-Stationary
SUPPLY_R_7	Original	-1,7607	0,4001	Non-Stationary	2,2245	0,0100	Non-Stationary
SUPPLY_R_9	Original	-1,5155	0,5258	Non-Stationary	2,2603	0,0100	Non-Stationary
SUPPLY_R_12	Original	-1,6559	0,4539	Non-Stationary	2,3005	0,0100	Non-Stationary
SUPPLY_R_15	Original	-2,2540	0,1872	Non-Stationary	2,1429	0,0100	Non-Stationary
SUPPLY_G_7	Original	-2,0838	0,2511	Non-Stationary	2,0493	0,0100	Non-Stationary
SUPPLY_G_9	Original	-2,6484	0,0834	Non-Stationary	1,8313	0,0100	Non-Stationary
SUPPLY_G_12	Original	-3,0362	0,0317	Stationary	1,7623	0,0100	Non-Stationary
SUPPLY_G_15	Original	-3,7600	0,0033	Stationary	1,6261	0,0100	Non-Stationary
BHSI	Box-Cox	-8,6915	0,0000	Stationary	0,0960	0,1000	Stationary
OIL	Box-Cox	-6,1509	0,0000	Stationary	0,0923	0,1000	Stationary
DOLLAR	Box-Cox	-17,8808	0,0000	Stationary	0,1056	0,1000	Stationary
DEMAND_R_7	Box-Cox	-8,0648	0,0000	Stationary	0,1587	0,1000	Stationary
DEMAND_R_9	Box-Cox	-5,2409	0,0000	Stationary	0,3961	0,0788	Stationary
DEMAND_R_12	Box-Cox	-5,1353	0,0000	Stationary	0,4542	0,0538	Stationary
DEMAND_R_15	Box-Cox	-4,6623	0,0001	Stationary	0,3209	0,1000	Stationary

DEMAND_G_7	Box-Cox	-6,4292	0,0000	Stationary	0,2971	0,1000	Stationary
DEMAND_G_9	Box-Cox	-7,2898	0,0000	Stationary	0,3193	0,1000	Stationary
DEMAND_G_12	Box-Cox	-6,5388	0,0000	Stationary	0,2890	0,1000	Stationary
DEMAND_G_15	Box-Cox	-6,3745	0,0000	Stationary	0,3929	0,0802	Stationary
AVAILABLE_TON_R_7	Box-Cox	-10,9183	0,0000	Stationary	0,2035	0,1000	Stationary
AVAILABLE_TON_R_9	Box-Cox	-8,0336	0,0000	Stationary	0,3241	0,1000	Stationary
AVAILABLE_TON_R_12	Box-Cox	-9,4102	0,0000	Stationary	0,2345	0,1000	Stationary
AVAILABLE_TON_R_15	Box-Cox	-7,5214	0,0000	Stationary	0,1628	0,1000	Stationary
AVAILABLE_TON_G_7	Box-Cox	-6,1220	0,0000	Stationary	0,1245	0,1000	Stationary
AVAILABLE_TON_G_9	Box-Cox	-7,3070	0,0000	Stationary	0,2107	0,1000	Stationary
AVAILABLE_TON_G_12	Box-Cox	-8,1160	0,0000	Stationary	0,1604	0,1000	Stationary
AVAILABLE_TON_G_15	Box-Cox	-5,7649	0,0000	Stationary	0,0935	0,1000	Stationary

Table 11.2: KPSS and ADF unit root tests to test stationarity in the data and the results. All data exhibited stationarity following first-order differencing. Note that the Box-Cox transformed variables has been differenced and winsorized before applying the ADF and KPSS tests

	count	mean	std	min	max	Skewness	Kurtosis	J-B
FREIGHT_RATE_boxcox	283	2,493452	0,015203	2,457037	2,519633	0,060816	-0,6442	5,067968
OIL_boxcox	283	108,8066	35,13721	16,60425	202,8704	0,004888	0,231382	0,632425
DOLLAR_boxcox	283	0,30699	0	0,30699	0,30699	0,030141	-0,84588	8,479878
AVAILABLE_TON_R_7_boxcox	283	140,5024	20,2778	82,62452	194,333	-0,00651	-0,15554	0,287272
AVAILABLE_TON_G_7_boxcox	283	28,5258	1,251068	23,856	31,84001	0,00071	0,033313	0,01311
AVAILABLE_TON_R_9_boxcox	283	30,92375	1,726778	26,588	35,93403	-0,0052	-0,35614	1,496904
AVAILABLE_TON_G_9_boxcox	283	7,415733	0,064682	7,195355	7,582389	-0,00148	0,16512	0,321598
AVAILABLE_TON_R_12_boxcox	283	21,14329	0,666951	19,23886	23,08783	-0,00121	-0,21055	0,522821
AVAILABLE_TON_G_12_boxcox	283	1,93467	0,000107	1,934353	1,934941	0,003399	-0,02315	0,006862
AVAILABLE_TON_R_15_boxcox	283	103,2464	6,460308	78,39461	122,1863	0,00983	0,438568	2,272593
AVAILABLE_TON_G_15_boxcox	283	0,972043	0	0,972043	0,972043	0,018795	-0,1528	0,291955
DEMAND_R_7_boxcox	283	15,48601	0,731695	12,88223	17,29443	0,000551	0,153554	0,278047
DEMAND_G_7_boxcox	283	108,9761	8,413655	80,74726	130,2254	0,001145	0,016706	0,003353
DEMAND_R_9_boxcox	283	8,144831	0,147962	7,6701	8,513583	-0,00014	0,027174	0,008708
DEMAND_G_9_boxcox	283	474,6573	49,09649	296,0296	626,6481	0,017445	0,432922	2,224368
DEMAND_R_12_boxcox	283	3,172109	0,004579	3,157317	3,183048	0,008949	-0,13325	0,213134
DEMAND_G_12_boxcox	283	395,8136	34,59483	270,2472	505,4042	0,017323	0,540713	3,461695
DEMAND_R_15_boxcox	283	3,017553	0,002991	3,006339	3,024856	-0,00474	0,218619	0,564635
DEMAND_G_15_boxcox	283	70,47447	3,192912	57,44801	79,50412	0,010913	0,729157	6,274885

Table 11.3: Descriptive statistics of all variables when Box-Cox transformed

	count	mean	std	min	max	Skewness	Kurtosis	J-B
FREIGHT_RATE_boxcox_diff	282	0,0000	0,0018	-0,0061	0,0074	0,3020	1,8838	45,9839
OIL_boxcox_diff	282	0,1566	5,9228	-26,6194	26,4365	-0,4016	4,2203	216,8552
DOLLAR_boxcox_diff	282	0,0000	0,0000	0,0000	0,0000	0,4531	1,9755	55,5079
AVAILABLE_TON_R_7_boxcox_diff	282	-0,1633	24,0996	-95,3031	71,4578	-0,2210	0,9231	12,3078
AVAILABLE_TON_G_7_boxcox_diff	282	-0,0127	1,1687	-4,9885	3,2734	-0,2740	1,3229	24,0902
AVAILABLE_TON_R_9_boxcox_diff	282	-0,0147	1,9313	-5,8679	6,4296	0,0121	0,2941	1,0231
AVAILABLE_TON_G_9_boxcox_diff	282	-0,0007	0,0564	-0,2499	0,2010	-0,1200	2,1042	52,7008
AVAILABLE_TON_R_12_boxcox_diff	282	-0,0079	0,7453	-2,2501	2,4225	0,0174	0,3547	1,4921
AVAILABLE_TON_G_12_boxcox_diff	282	0,0000	0,0001	-0,0003	0,0003	-0,0051	0,9870	11,4474
AVAILABLE_TON_R_15_boxcox_diff	282	-0,0831	7,1502	-19,6558	27,3721	0,1956	0,8888	11,0801
AVAILABLE_TON_G_15_boxcox_diff	282	0,0000	0,0000	0,0000	0,0000	0,2166	0,9461	12,7230
DEMAND_R_7_boxcox_diff	282	0,0008	0,7092	-2,5552	2,4924	-0,1128	0,5710	4,4288
DEMAND_G_7_boxcox_diff	282	0,0425	7,9512	-28,5979	21,9111	-0,3593	0,4904	8,8940
DEMAND_R_9_boxcox_diff	282	-0,0001	0,1375	-0,5185	0,4595	0,0632	0,8117	7,9288
DEMAND_G_9_boxcox_diff	282	0,2038	42,3784	-168,9713	124,6403	-0,2311	0,8443	10,8855
DEMAND_R_12_boxcox_diff	282	0,0000	0,0039	-0,0137	0,0131	0,1913	0,8864	10,9514
DEMAND_G_12_boxcox_diff	282	0,1475	27,1731	-91,8402	92,5159	0,2964	0,6926	9,7660
DEMAND_R_15_boxcox_diff	282	0,0000	0,0024	-0,0092	0,0091	0,5286	1,9059	55,8124
DEMAND_G_15_boxcox_diff	282	0,0176	2,3844	-8,9860	9,9224	0,6307	1,7913	56,3989

Table 11.4: Descriptive statistics of all the variables when Box-Cox transformed and differenced

Outliers (Z-Score)		
	Original	95% Winsorizing
FREIGHT_RATE	6	0
OIL	5	0
DOLLAR	3	0
Supply_R_7	2	0
Supply_R_9	2	0
Supply_R_12	3	0
Supply_R_15	3	0
Demand_R_7	2	0
Demand_R_9	3	0
Demand_R_12	2	0
Demand_R_15	4	0
Supply_G_7	3	0
Supply_G_9	4	0
Supply_G_12	3	0
Supply_G_15	1	0
Demand_G_7	2	0
Demand_G_9	3	0
Demand_G_12	3	0
Demand_G_15	5	0

Table 11.5*:* Z-score for the Box-Cox differenced variables (pre- and post-winsorization)

Variables	
Endogenous	Exogenous
FREIGHT_RATE_raw	OIL_raw
	DOLLAR_raw
	AVAILABLE_TON_R_7_raw
	AVAILABLE_TON_G_7_raw
	AVAILABLE_TON_R_9_raw
	AVAILABLE_TON_G_9_raw
	AVAILABLE_TON_R_12_raw
	AVAILABLE_TON_G_12_raw

		AVAILABLE_TON_R_15_raw
		AVAILABLE_TON_G_15_raw
		DEMAND_R_7_raw
		DEMAND_G_7_raw
		DEMAND_R_9_raw
		DEMAND_G_9_raw
		DEMAND_R_12_raw
		DEMAND_G_12_raw
		DEMAND_R_15_raw
		DEMAND_G_15_raw
Number:	1	18

Table 11.6: Number of endogenous and exogenous variables

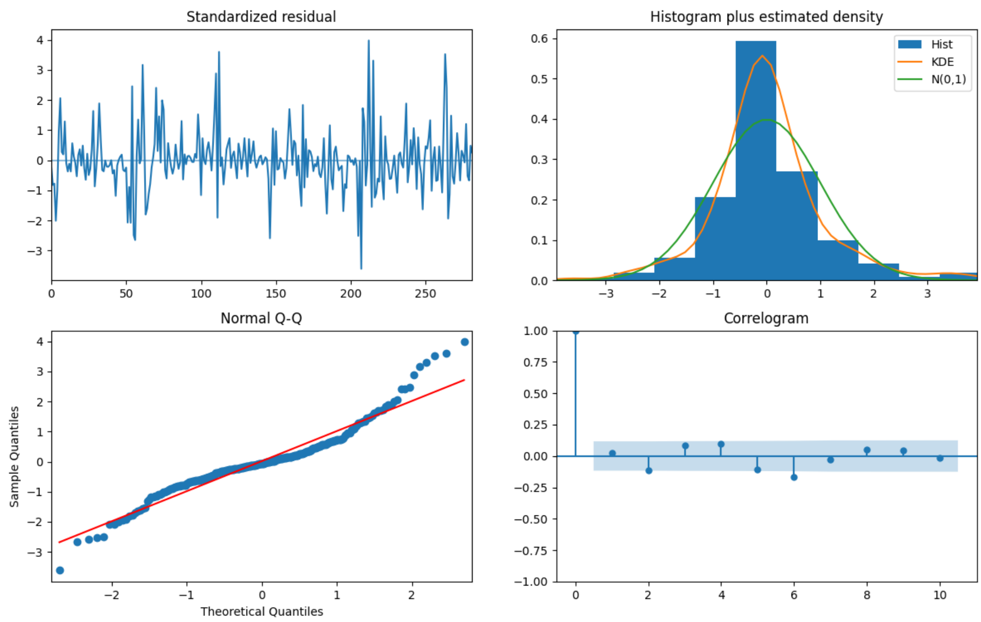


Figure 11.1: Information of the residuals of a Box-Cox transformed, differenced ARIMA model

11.1 Exhaustive List of Packages Used in Python

1. arch (from arch import arch_model);
2. darts:
   a. darts.dataprocessing.transformers (from darts.dataprocessing.transformers import Scaler);

   b. darts.models (from darts.models import ARIMA, AutoARIMA, ExponentialSmoothing);
   c. darts.utils.statistics (from darts.utils.statistics import stationarity_tests);
3. matplotlib:
   a. matplotlib.pyplot (import matplotlib.pyplot as plt);
   b. matplotlib.ticker (import matplotlib.ticker as ticker);
   c. matplotlib.dates (import matplotlib.dates as mdates);
4. numpy (import numpy as np);
5. pandas (import pandas as pd);
6. pmdarima (from pmdarima import auto_arima);
7. sklearn.metrics (from sklearn.metrics import mean_squared_error)
8. scipy.stats (from scipy.stats import skew, kurtosis, jarque_bera, boxcox);
9. scipy.stats.mstats (from scipy.stats.mstats import winsorize);
10. seaborne (import seaborne as sns);
11. statsmodels.api (import statsmodels.api as sm);
12. statsmodels.graphics.tsaplots (from statsmodels.graphics.tsaplots import plot_acf, plot_pacf);
13. statsmodels.stats.diagnostic (from statsmodels.stats.diagnostic import het_arch);
14. statsmodels.stats.outliers_influence (from statsmodels.stats.outliers_influence import variance_inflation_factor);
15. statsmodels.tsa.seasonal (from statsmodels.tsa.seasonal import seasonal_decompose, STL);
16. statsmodels.tsa.statespace.sarimax (from statsmodels.tsa.statespace.sarimax import SARIMAX);
17. statsmodels.tsa.stattools (from statsmodels.tsa.stattools import adfuller, kpss);
18. tabulate (from tabulate import tabulate)