Forecasting container freight rates for the major trade routes: a comparison of artificial neural network and conventional models

Ziaul Haque Munim
Faculty of Technology, Natural and Maritime Sciences, University of South-Eastern Norway
E-mail: ziaul.h.munim@usn.no

Hans-Joachim Schramm
Institute for Transport and Logistics Management, Vienna University of Economics and Business
Department of Operations Management, Copenhagen Business School
E-mail: hschramm@wu.ac.at

Abstract

Major players in the maritime business such as shipping lines, charterers, shippers, and others rely heavily on container freight rate forecasts for operational decision making. The non-existence of a formal forward market in the container industry makes it necessary for them to rely on forecasts for their hedging strategy purposes, too. Thus, to identify better performing forecasting approaches, we compare three models, namely, Autoregressive Integrated Moving Average (ARIMA), Vector Autoregressive (VAR) or Vector Error Correction (VEC) and Artificial Neural Network (ANN) models. We examine the China Containerised Freight Index (CCFI) as a collection of weekly freight rates published by the Shanghai Shipping Exchange (SSE) in four major trade routes. Overall, VAR/VEC models outperform ARIMA and ANN in training-sample forecasts, but ARIMA outperforms VAR and ANN taking test-samples. On route level, we observe two exceptions to this. ARIMA performs better for the Far East to Mediterranean in the training-sample, and the VEC model did the same in the Far East to US East Coast route in the test-sample. Hence, we advise the industry players to use ARIMA for forecasting container freight rates for major trade routes ex-China except for VEC in the case of the Far East to US East Coast route.

Keywords: *artificial neural network; vector error correction; forecast performance; ocean freight; backpropagation algorithm*

Cite as: Munim, Z.H., Schramm, H-J. (2020). Forecasting container freight rates for the major trade routes: a comparison of artificial neural network and conventional models. *Maritime Economics and Logistics*, *https://doi.org/10.1057/s41278-020-00156-5*

1. Introduction

This study is the result of an ongoing quest to develop state-of-the-art forecasting approaches for container freight rates for main trade routes. As we present in Table 1, East Asia to North America, i.e. US West Coast (USWC) and US East Coast (USEC), and Europe, i.e. Northern Europe (NEU) and Mediterranean (MED) are by far the strongest trade routes in container shipping. Corresponding freight rate data is available in the Shanghai Containerized Freight Index (SCFI) and China Containerized Freight Index (CCFI), collected and published by SSE on a weekly basis.

CCFI and SCFI are highly regarded in the container shipping industry and therefore often used as the underlying asset in forward rate agreements or floating element in index-linked container contracts (Drewry, 2012; Kavussanos *et al.*, 2015; Miller *et al.*, 2015). SCFI collects spot rates (CIF, CY/CY including all major seaborne surcharges) of the export container market on 15 shipping routes ex Shanghai, denominated in USD per TEU or FEU (SSE, 2019a). CCFI, however, is more comprehensive than SCFI as its index reflects the overall freight level (including spot and long-term rates) of China's export container market, from 10 Chinese hub ports (Dalian, Tianjin, Qingdao, Shanghai, Nanjing, Ningbo, Xiamen, Fuzhou, Shenzhen and Guangzhou) on 14 individual shipping routes (base = 1,000 equivalent to the freight rate of a TEU or FEU at 01/01/1998 according to SSE, 2019b). Given the prominent position of these 10 Chinese hub ports in world-wide container throughput rankings and their inclusion in port rotations of outbound container services from Far East (FE), the CCFI can be regarded as a good proxy for the whole region.

Table 1: Containerised trade on major east-west trade routes (Mio. TEUs)

Year	East Asia to North America	North America to East Asia	NEU/MED to East Asia	East Asia to NEU/MED	North America to NEU/MED	NEU/MED to North America
2014	16.2	7.0	6.3	15.4	2.8	3.9
2015	17.5	6.9	6.4	15.0	2.7	4.1
2016	18.3	7.3	6.8	15.4	2.7	4.2
2017	19.5	7.3	7.1	16.5	3.0	4.6
2018	20.9	7.4	7.0	17.4	3.1	4.9

Source: Review of Maritime Transport (UNCTAD, 2019).

Following Stopford (2009) and Munim and Schramm (2017), there is a complex interplay between supply and demand in container shipping today. Over the last decade, ship upsizing and cascading of capacity have affected containerized trade largely (Clarkson, 2019; UNCTAD, 2019). Since 2006, almost all containership new buildings of more than 15,000

TEU capacity are employed on the FE to NEU or MED trade lanes. In 2014 to 2018 alone, 110 of such ultra large container ships (ULCS) joined the market and have led to a significant upgrade of available shipping capacity there (Clarkson, 2019). So far, these ULCS do not yet appear to be economically viable on other trade lanes but due to the cascading effect, large container ships formerly deployed on the FE to NEU or MED trade lanes add significant shipping capacity there.

In addition to this, the expanded Panama Canal began commercial operations on June 26, 2016 (Zamorano and Martinez, 2016), and soon thereafter the earlier Panamax vessels of around 5,000 TEU became less attractive. Many of them left the market after the canal expansion (Wackett, 2017) and were replaced by new capacity ranging from 8,000 to 12,000 TEUs on the FE to USEC trade route via the Panama Canal (Clarksons, 2017), with Neo-Panamax container vessels up to 14,863 TEU being able to transit the new locks nowadays (PanCanal, 2017).

During the period in question, a massive inflow of shipping capacity, persistent trade lane imbalances together with rather weak demand put pressure on container freight rates in line with common liner shipping theory (Haralambides *et al*., 2003). Moreover, competitive but untransparent freight markets with an inherent persistence of pricing habits based on a multitude of surcharges from the liner conference era contributed to container freight rate volatility (Slack and Gouvernal, 2011; Munim and Schramm, 2017). This is especially true when we look at Bunker Adjustment Factor (BAF, see Slack and Gouvernal, 2011; Wang *et al.*, 2011; Notteboom and Cariou, 2013) announced one month in advance or regularly occurring but mostly unsuccessful General Rate Increases (GRI, see Munim and Schramm, 2017; Chen *et al*., 2017) set by each liner shipping operator individually without taking actual supply demand balance into consideration. Meanwhile, on the two transpacific trade lanes from FE to UEWC and USEC, GRIs and BAFs were officially set on a quarterly basis in line with the Transpacific Stabilization Agreement (TSA)[1], which led to less competition in these routes until TSA's shut down on February 8, 2018 (Signals DPI, 2018). This might have contributed to a higher freight price stability than on the FE to NEU and MED trade lanes after total abolishment of the shipping conference system there in 2008 (Haralambides *et al.*, 2003).

Finally, this rather depressive market environment resulted in overall low profitability of liner shipping operators which, among others, resulted in a market shakeup with the bankruptcy of Hanjin in 2016, followed by a rapid sequence of mergers and acquisitions (Porter, 2016). As a consequence, the top league of container liner operators with more than 200,000 TEU fleet capacity declined from 20 to 15 within two years (see Table 2). This trend of consolidation went even further with the merger of the three Japanese operators to ONE (Ocean Network

[1] TSA (March 5, 1989-February 8, 2018) was a freight rate discussion forum supported by the Federal Maritime Commission of the USA comprised of 15 major container shipping lines operating on the transpacific trade lane. During the 29 years of operation, TSA had significant influence over the transpacific shipping freight rates including GRIs, BAF and seasonal surcharges.

Express) finalized on April 1, 2018. Furthermore, COSCO acquired OOCL on June 30, 2018 and thus at the beginning of 2019 there were only 12 major container liner operators left. Moreover, fleet capacity and the average size of ships grew significantly, with fewer vessels in operation as a result (Table 2).

From the international trade perspective, Haralambides (2019) argued that economic distance (as proxied by ocean freight rates) is nowadays more important than geographical distance in determining trade between countries. Under such circumstances, forecasting container shipping freight rates accurately or at least better than naïve forecasts could help liner companies make the right decisions at the right time as their profitability and investments depend heavily on such forecasts. Moreover, major players in the global trade industry, from shippers to ship owners, ports authorities, banks and others rely extensively on freight rate forecasts (Gharehgozli *et al.*, 2018). Also, as there exist no active forward markets for container shipping freight rates, major players rely on rate forecasts as an alternative approach to hedging while making different contracts.

Table 2: Top League of Operators in Container Liner Shipping

Date	No. of operators in Top League	Fleet capacity in TEU	No. of Ships in fleet	Average Ship Size in TEU
07/01/2015	20	16,308,488	3,447	4,729
01/01/2016	20	17,085,521	3,402	5,022
04/01/2017	17	16,802,291	3,172	5,297
07/01/2018	15	18,410,621	3,340	5,512
02/01/2019	12	19,230,573	3,314	5,803

Source: Top 100 Alphaliner[2]

The remainder of the paper is as follows: Section 2 unfolds the literature of freight rate forecasting approaches, with a focus on the container shipping industry. We present the properties of data and selected forecasting methods in Section 3. Empirical results are presented in Section 4, including an assessment of forecasting accuracy in the given context. Section 5 presents key findings and concludes with highlighting the relevance of our work for practice and future research.

2. Literature review

Forecasting freight rates has been an interesting topic in the shipping industry for a long time, and researchers have tested different types of econometric models in the last decades. Among

[2] https://alphaliner.axsmarine.com/PublicTop100.

those, the ARIMA model (Box and Jenkins, 1976), the Vector Autoregressive (VAR) model (Sims, 1980), the Vector Equilibrium Correction (VEC) model (Engle and Granger, 1987) and its variants, including exogenous variables like ARIMAX and VARX are the most explored in this domain. However, most applications are from the bulk shipping sector (e.g. Veenstra and Franses, 1997; Bachelor *et al.*, 2007; Chen *et al.*, 2012; Zhang and Zeng, 2015; Tsioumas *et al.*, 2017; Yin *et al.*, 2017) due to greater data availability and market maturity.

Kavussanos *et al.* (2004) and Li *et al.* (2014) have also used Generalized Autoregressive Conditional Heteroskedasticity (GARCH) models, initially proposed by Engle (1982), and then generalized by Bollerslev (1986). A rather new approach in shipping freight rate forecasting is Artificial Neural Networks (ANN). Yun *et al* (2018) use ANN to determine the future realization value of options in time charter markets. Recently, Zhang *et al.* (2018) use three different autoregressive ANN algorithms to predict the Baltic Dry Index.

Whereas the literature on forecasting freight rates could be described as *mature* in the case of the bulk shipping sector, maritime economists have been focusing on container shipping only recently (e.g. Luo *et al.*, 2009; Nielsen *et al.*, 2014; Munim and Schramm, 2017). In addition to this, the relevant studies are about freight index construction (Karamperidis *et al.*, 2013; Yifei *et al.*, 2018), seasonality effects (Yin and Shin, 2018), GRIs (Chen *et al*, 2017; Avdasheva *et al*, 2018), or overall freight rate structure and surcharges (e.g. BAFs) in general (Slack and Gouvernal, 2011; Wang *et al.*, 2011; Notteboom and Cariou, 2013; De Oliveira, 2014).

Luo *et al.* (2009) have made one of the pioneer attempts to forecasting freight rates in the container shipping industry as a function of fleet capacity for the period 1980-2008. The explanatory power of their model is high, which could be simply due to the fact that their dataset is based on yearly data, which is usually less volatile than weekly and monthly data.

Nielsen *et al.* (2014) develop a forecasting model for container freight rates, investigating the relationship between aggregated market rates (i.e. SCFI) and individual liner rates. Their model focuses on performance and robustness based on observation fit and forecasting horizon. However, the model has limitations in explanatory power, suggesting a possible inconsistency in freight rate governing mechanisms over time.

Munim and Schramm (2017) apply ARIMA and ARIMARCH (i.e. a combination of ARIMA and ARCH) models on SCFI and CCFI, at monthly and weekly levels, respectively, for the FE to NEU trade lane and for the period 2009-2015. The ARIMARCH model provides better results than other ARIMA group models while performing short-term forecasts on a weekly level.

Fan (2011) and Nielsen *et al.* (2014) state that container freight rates are cyclical in nature and can fluctuate widely in the course of a single week. Yifei *et al.* (2018) described how to make a freight rate index even on a daily basis. However, considering the significance of weekly container freight rate forecasts for liner companies, shippers and other maritime stakeholders, we forecast the weekly CCFI for the four major trade routes for 13 weeks out-sample period.

3. Data and methodology

3.1. Dataset

Apart from freight rate indexes like SCFI or CCFI, data on container shipping is rarely available on a weekly basis, and most data sources report on a monthly to quarterly basis. However, some indicators are available on weekly basis and can be derived from shipping databases such as Clarksons (n.d.).

As a representative container freight rate, weekly announcements of CCFI for the four trade lanes examined are taken from the SSE (n.d.). To reflect recent developments in the supply side, week-to-week changes in the capacity of vessels of more than 14,000 TEU is derived from fleet data available in Clarksons (n.d.). Newbuilding prices of containerships in the 16,000 - 16,500 TEU capacity are taken from Clarksons (n.d.), too. Containership charter indexes reflect earnings on the side of tonnage providers. For this, we use the HARPEX shipping index from Harper Petersen (n.d.), as it reflects better bigger container vessel sizes than others like the Howe Robinson Index (HRI, cf. Howe Robinson, n.d.) or the Hamburg Index (HAX, cf. VHBS, n.d.).

Other indicators on a weekly basis could be GRIs and freight surcharges like BAF. However, both are set individually by each operator for specific trade routes so that data is difficult to collect apart from the transpacific trade lane under TSA reported by Signals DPI until February 8, 2018 (Signals DPI, 2018).

Time series data of the CCFI container freight rates on the four major trade routes are depicted in Figure 1. Descriptive statistics are presented in Table 3. In-sample and out-sample (also referred to as training-sample and test-sample, respectively) time series are defined in Table 4, where the in-sample period consists of 142 weeks from 1/16/2015 to 9/29/2017 and the out-sample period consists of 13 weeks from 10/6/2017 to 12/29/2017. Prior to forecasting, all data is log-transformed and checked for stationarity. Data becomes stationary in first difference, see PP test (Phillips and Perron, 1988) results in Table 4. For multivariate modelling purposes, multicollinearity among the exogenous variables (i.e. week-to-week change in fleet capacity, newbuilding prices and HARPEX index) has been checked, and no major problem exists as the variance inflation factor (VIF) value of all variables are below 2 (Hair *et al.*, 2010).

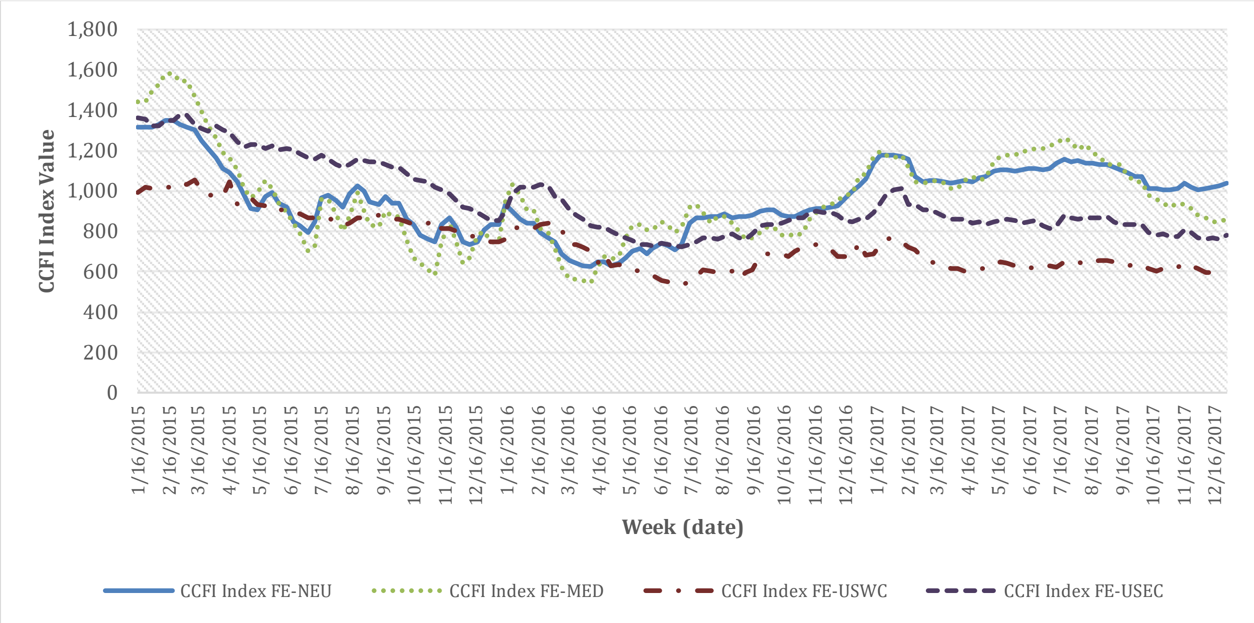


Figure 1: Time series data of container freight rates on the major trade routes

Table 3: Descriptive statistics

Variables	N	Mean	Std. dev.	Min.	Max.	J-B test
CCFI: FE to NEU	155	962.62	172.80	625.12	1349.27	1.49
CCFI: FE to MED	155	963.90	225.87	543.31	1581.25	10.80**
CCFI: FE to USWC	155	1581.25	135.66	534.47	1054.24	13.98***
CCFI: FE to USEC	155	951.20	181.45	722.65	1377.88	21.11***
ΔWoW capacity (TEUs)	155	12056.15	13877.82	0	74198	74.70***
Newbuilding Prices (mi. USD)	155	130.00	4.05	123	134.5	18.61***
HARPEX index	155	439.04	98.82	314	645.82	11.22**

*Here, ΔWoW capacity: week-to-week change in of the fleet capacity of vessels with more than 14,000 TEU capacity; Newbuilding Prices: newbuilding prices in million USD for containership with 16,000 - 16,500 TEU capacity. Descriptive statistics and Jarque-Bera test results are of the time series in levels.*p<0.05, **p<0.01, ***p<0.001; For J-B test, P < 0.05 indicates non-normality of time series.*

Table 4: Sample definition and unit root test for stationarity

Variables	In-sample			Out-sample
	Time series	PP-test (level)	PP-test (1st difference)	Time series
CCFI: FE to NEU	1/16/2015 – 9/29/2017	-5.93	-78.27**	10/6/2017 – 12/29/2017
CCFI: FE to MED	1/16/2015 – 9/29/2017	-8.49	-78.06**	10/6/2017 – 12/29/2017
CCFI: FE to USWC	1/16/2015 – 9/29/2017	-7.55	-156.44**	10/6/2017 – 12/29/2017
CCFI: FE to USEC	1/16/2015 – 9/29/2017	-5.03	-82.63**	10/6/2017 – 12/29/2017
ΔWoW capacity (TEUs)	1/16/2015 – 9/29/2017	-131.98**	-174.71**	10/6/2017 – 12/29/2017
Newbuilding Prices (mi. USD)	1/16/2015 – 9/29/2017	-3.90	-117.81*	10/6/2017 – 12/29/2017
HARPEX index	1/16/2015 – 9/29/2017	-2.14	-44.83**	10/6/2017 – 12/29/2017

*Note that all the time series were log transformed. *p<0.05, **p<0.01, ***p<0.001; for PP-test, P < 0.05 indicates stationarity of time series.*

3.2. Forecasting models

Based on the discussion in Section 2, we adopt three forecasting models — one univariate and two multivariate. We choose ARIMA as the univariate model due to its proven effectiveness in freight rate forecasting in the context of container shipping. After ARIMA, VAR/VEC is the most promising one, and ANN is recently gaining attention. Therefore, VAR/VEC and ANN are chosen as the two multivariate models.

3.2.1. Autoregressive integrated moving average (ARIMA)

A non-seasonal ARIMA model has two parts: autoregressive (AR) and moving average (MA). The AR part captures the association between the value of a variable at time t and its value at a previous time $(t - i)$. The MA part captures the association between the value of a variable at time t and its residual value at a previous time $(t - i)$. When a time series is stationary in first differences, then, its ARIMA (p, d, q) model can be represented by equation (1) as follows:

$$\Delta z_t = \sum_{i=1}^{p} \emptyset_i \Delta z_{t-i} + \sum_{i=1}^{q} \theta_i \varepsilon_{t-i} + \varepsilon_t \qquad (1)$$

Here, $\Delta z_t = z_t - z_{t-1}$; z_t is the container shipping freight rate at time t; $\emptyset_i$ is the coefficient of Δz_{t-i}; θ_i is the coefficient of error terms at time $t-1$, ε_{t-i}; and ε_t is the error term at time t.

3.2.2. Vector autoregressive (VAR) and Vector error correction (VEC) models

VAR is stochastic process model and can capture linear interdependencies among multiple variables. VAR differs from ARIMA by allowing more than one variable in the AR process. That is, it models the association between the value of a variable at time t, and its value, as well as the value of some other variables, at a previous time $(t-i)$. A VAR model with a vector of $k\ x\ k$ variables can be represented by equation (2) as follows:

$$\Delta z_t = \beta + \sum_{i=1}^{p} \varphi_i \Delta z_{t-i} + \varepsilon_t \qquad (2)$$

Here, Δz_t is a $k\ x\ 1$ vector of variables in first difference; β is a $k\ x\ 1$ vector of constants; φ_i is a time-invariant $k\ x\ k$ matrix of the coefficients of Δz_{t-i}; and p refers to the lag structure.

VEC models are similar to VAR models but with an error correction term, and can be represented by equation (3) as follows:

$$\Delta z_t = \beta + \sum_{i=1}^{p-1} \varphi_i \Delta z_{t-i} + \delta\lambda' z_{t-p} + \varepsilon_t \qquad (3)$$

Here, β is a $k\ x\ 1$ vector of constants; $\sum_{i=1}^{p-1} \varphi_i \Delta z_{t-i}$ is the component of VAR terms in first difference; $\delta\lambda' z_{t-p}$ represents the error-correction terms in levels; p refers to the lag structure; δ and λ are $k\,x\,k$ matrices representing the weights and parameters of co-integrating relationships, respectively.

3.2.3. Artificial neural network (ANN)

Unlike ARIMA and VAR/VEC, ANN does not require assumptions of linearity and stationarity, so that it can be used to estimate any complex functional relationship (Günther and Fritsch, 2010). ANN is completely data-driven and developed based on *multi-layer perceptions* (MLP). Typically, an ANN model has three layers: input, hidden and output and as such it models the relationship between input and output covariates (see Günther and Fritsch, 2010 for detail). An ANN (x, m, s) model (where x represents the number of input covariates, m the number of hidden layers and s the number of output covariates) can be represented by equation (4) as follows:

$$z_t = \mathrm{f}(X_t, \gamma, \omega)$$

$$= \gamma_0 + \sum_{j=1}^{n} \gamma_j F\left(\sum_{i=1}^{x} \omega_{ij} X_{it} + \omega_{oj}\right) + \varepsilon_t \qquad (4)$$

Here, z_t is the container shipping freight rate at time t; γ_0 denotes the intercept of the output neuron z_t; n denotes the number of middle layer units; γ_j denotes a vector of weights (or coefficients) from the middle to output layer units; F is a logistic function, where, $F(\gamma) = 1/(1 + \exp(-\gamma))$; ω_{ij} represents a matrix of weights from the input to middle layer at time t; and X_{it} is a vector of input variables that are same as the ones used in VAR and VEC models.

4. Empirical analysis and results

We use the static forecasting approach as it is more relevant for practical purposes and decision making. First, forecasts of weekly container shipping freight rates for four major trade routes are made using two traditional models — ARIMA and VAR. Then these forecasts are benchmarked with the forecasts of ANN. For ARIMA, we use the R package *forecast* (Hyndman and Khandakar, 2007), for VAR/VEC *tsDyn* (Narzo *et al.*, 2018) and for ANN *neuralnet* (Günther and Fritsch, 2010).

The ARIMA models for each trade route are selected based on the lowest AIC (Akaike Information Criterion, Akaike, 1974) and the PP test for stationarity. The selected ARIMA models for the in-sample period and their respective parameters are presented in Table 5. Results from several residual diagnostic tests of selected models for each trade route are posted in Table 5, too. The Ljung-Box test (L-B, Ljung and Box, 1978) confirms the non-existence of autocorrelation among the residuals, but the Jarque-Bera test (J-B, Jarque and Bera, 1980) failed to confirm normality of the residuals. However, the Q-Q plot of the residuals shows a rather normal distribution pattern but with fat tails (available upon request). Furthermore, L-B test of residual squared confirms the non-existence of ARCH effect.

Table 5: ARIMA model parameters

ARIMA(p,d,q) model parameters	FE-NEU. ARIMA(0,1,3).	FE-MED. ARIMA(3,1,2).	FE-USWC. ARIMA(0,1,0).	FE-USEC. ARIMA(1,1,0)
$\mathbf{AR_1}$	-	0.590***	-	0.421***
$\mathbf{AR_2}$	-	-1.028***	-	-
$\mathbf{AR_3}$	-	0.335***	-	-
$\mathbf{MA_1}$	0.457***	-0.161*	-	-
$\mathbf{MA_2}$	0.084	0.782***	-	-
$\mathbf{MA_3}$	-0.161	-	-	-
AIC	-542.46	-399.12	-565.53	-691.68
BIC	-530.66	-381.42	-562.58	-685.78
Residual diagnostic:				
L-B test (lag 10)	7.249	7.805	2.477	6.527
Residual^2 L-B test (lag 10)	17.070	18.428	9.767	3.414
J-B test	49.010***	114.729***	60.890***	38.535***

**p<0.05, **p<0.01, ***p<0.001*

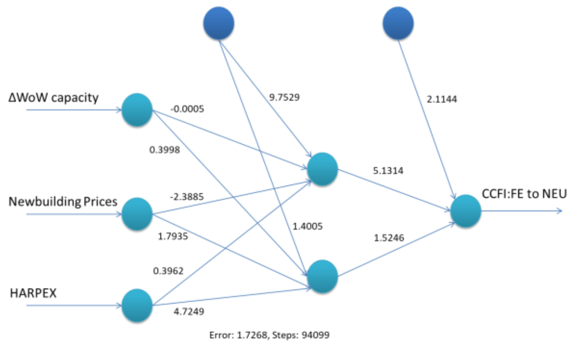
For multivariate modelling, we consider three explanatory variables: (1) week-to-week containership capacity change; (2) containership newbuilding prices; and (3) the HARPEX charter price index. Stationarity of the weekly time series data of these variables is checked through the PP test. All of them confirmed stationarity at first difference, and we thus proceed with the Johansen co-integration test (Johansen, 1991). Based on the Hannan-Quinn (HQ) and Schwarz criterion (SC) of VAR lag selection (Lütkepohl, 1985), two lags are used for the co-integration test. Interestingly, at 5% significance level, we find one co-integration equation for three of the trade routes, except for the FE-MED. Therefore, VEC modelling is considered for all routes, and VAR modelling for the FE-MED. The selected VAR/VEC models for the in-sample period and their respective parameters are presented in Table 6. Regarding model residual diagnostic tests, the L-B test confirms the non-existence of autocorrelation for three trade routes, except FE-MED, while the J-B test fails to confirm normality of the residuals. Again, the Q-Q plot of the residuals show a rather normal distribution pattern but with fat tails (available upon request).

Table 6: VAR/VEC model parameters

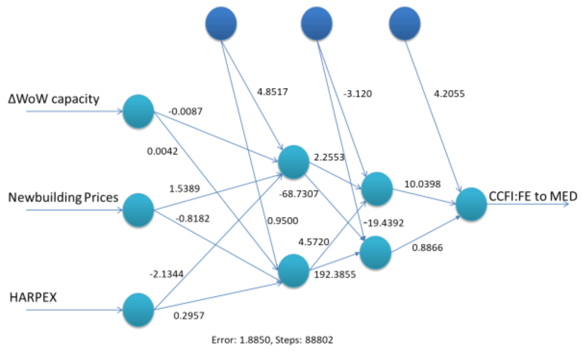
VEC(p) model parameters	FE-NEU. VEC(1).	FE-MED. VAR(2).	FE-USWC. VEC(1).	FE-USEC. VEC(1).
ECT	-0.0002	-	-0.0019	0.0004
Intercept	0.0066	2.1406	0.0127	-0.0087
$\mathbf{CCFI}: \mathbf{Y_1}$	0.3950***	1.2730***	-0.0628	0.4070***
$\mathbf{CCFI}: \mathbf{Y_2}$	-	-0.3645***	-	-
$\mathbf{\Delta WoW\ capacity_1}$	-0.0012*	-0.0017	6.0e-05	0.0001
$\mathbf{\Delta WoW\ capacity_2}$	-	0.0032**	-	-
$\mathbf{N.B.Prices_1}$	-0.2018	2.3472	1.6897	0.5496
$\mathbf{N.B.Prices_2}$	-	-2.6738	-	-
$\mathbf{HARPEX_1}$	-0.0054	0.3349	-0.0171	0.0284
$\mathbf{HARPEX_2}$	-	-0.3236	-	-
HQ (2)	-23.868	-22.657	-24.000	-25.000
SC (2)	-23.350	-22.139	-23.482	-24.482
AIC	-3392.493	-3237.681	-3412.829	-3542.625
BIC	-3313.069	-3131.782	-3333.404	-3463.200
Residual diagnostic:				
L-B test (lag 10)	13.538	46.647***	1.425	5.533
$Residual^2$ L-B test (lag 10)	17.439	37.551***	10.617	4.238
J-B test	40.076 ***	67.443***	44.946***	29.163***
Result of Johansen co-integration test at 5% significance level	One co-integrating equation exists.	No co-integrating equation exists.	One co-integrating equation exists.	One co-integrating equation exists.

**p<0.05, **p<0.01, ***p<0.001*

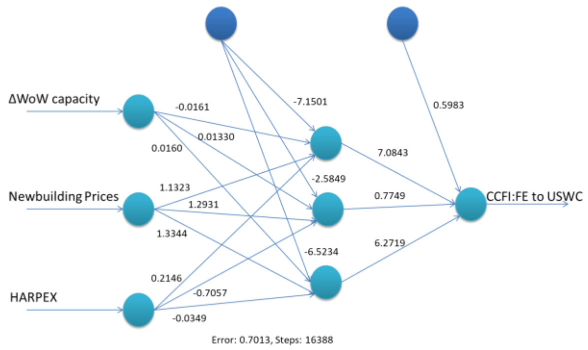
For ANN, the same explanatory variables as in the VAR/VEC models are used. As mentioned above, prior assumptions about the underlying data are not necessary in ANN modelling. Issues such as stationarity and co-integration are not decisive either in ANN, but convergence difficulties may occur due to using a large number of covariates and response variables (Günther and Fritsch, 2010). Note that we assume linear interdependencies among variables when scrutinizing the appropriate ANN models. The converged ANN models for each of the trade routes are presented in Figure 2 with their respective parameters. The resilient backpropagation with backtracking (*rprop+*) ANN algorithm (Günther and Fritsch, 2010) is employed. For ANN model selection, starting with the rule of thumb of x-1 hidden layers (x is number of input covariates), we examine 11 models with different number of hidden layer combinations[3]. For each of four trade routes, we select the model with the best in-sample and out-sample forecasting performance.



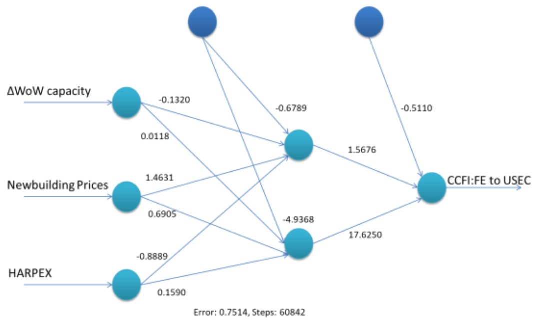
(a) Far East to NEU route. ANN(3,2,1).



(b) Far East to MED route. ANN(3,2,2,1).



(c) Far East to USWC route. ANN(3,3,1).



(d) Far East to USEC route. ANN(3,2,1).

Figure 2: Parameters of selected ANN models

After selecting the appropriate ARIMA, VAR/VEC and ANN models based on the training sample and estimating their respective parameters, we forecast out-sample container freight rates, employing those parameters. For both in-sample and out-sample forecasts, the performance of forecasting models is of great interest to practitioners and academics. Therefore, we use three measures of forecasting accuracy: root mean square error (RMSE), mean absolute percent error (MAPE) and mean absolute scaled error (MASE). Not many

[3] Examined ANN models are (3,1,1), (3,2,1), (3,3,1), (3,4,1), (3,5,1), (3,1,1,1,), (3,2,1,1), (3,2,2,1), (3,3,1,1), (3,3,2,1), (3,3,3,1).

maritime studies have used MASE to compare forecasting performance, but Hyndman and Koehler (2006) argue that traditional accuracy measures, such as RMSE and MAPE, tend to be biased because of its dependency on the number of the out-sample forecasting period. The equations to calculate RMSE, MAPE and MASE are presented below:

$$RMSE = \sqrt{\frac{1}{n}\sum_{t=1}^{n}(d_t - z_t)^2} \tag{5}$$

$$MAPE = \frac{100}{n}\sum_{t=1}^{n}\left|\frac{(d_t - z_t)}{d_t}\right| \tag{6}$$

$$MASE = mean\left|\frac{e_t}{\frac{1}{n-1}\sum_{t=2}^{n}|z_t - z_{t-1}|}\right| \tag{7}$$

Here, d_t is the actual container freight rate at time t; z_t is the forecasted rate at time t; n is the total number of observations; e_t is the forecasting error calculated as $(d_t - z_t)$; and $z_t - z_{t-1}$ is the forecasting error of the naïve forecast.

The in-sample and out-sample forecasting accuracy of each of the trade routes, for their respective forecasting models, are presented in Table 7 and Table 8, respectively. On the aggregate level, in the in-sample forecasting accuracy measures, VEC/VAR models outperform ARIMA and ANN. But on route level, there is one exception, that is, ARIMA performs better than VAR and ANN for the FE-MED trade route. On the other hand, while comparing out-sample forecasting accuracy on the aggregate level, ARIMA models outperform VEC/VAR and ANN. However, on the route level, VEC model outperforms others for the FE-USEC route.

Table 7: In-sample forecasting performance

Forecasting models	FE-NEU.	FE-MED.	FE-USWC.	FE-USEC.	Average
ARIMA(p,d,q)	ARIMA(0,1,3)	ARIMA(3,1,2)	ARIMA(0,1,0)	ARIMA(1,1,0)	
RMSE	0.034	**0.056**	0.032	0.020	0.036
MAPE	0.336	**0.563**	0.328	0.224	0.363
MASE	0.842	**0.795**	0.995	0.913	0.886
VAR(p)/VECM(p)	VEC(1)	VAR(2)	VEC(1)	VEC(1)	
RMSE	0.017	0.060	0.005	0.010	**0.023**
MAPE	0.189	0.645	0.061	0.113	**0.252**
MASE	0.475	0.912	0.183	0.462	**0.508**
ANN(x,m,s)	ANN(3,2,1)	ANN(3,2,2,1)	ANN(3,3,1)	ANN(3,2,1)	
RMSE	0.156	0.163	0.099	0.103	0.130
MAPE	1.681	1.886	1.177	1.158	1.475
MASE	4.251	2.691	3.569	4.721	3.808

In-sample period is 142 weeks from 1/16/2015 to 9/29/2017.

Table 8: Out-sample forecasting performance

Forecasting models	FE-NEU.	FE-MED.	FE-USWC.	FE-USEC.	Average
ARIMA(p,d,q)	ARIMA(0,1,3)	ARIMA(3,1,2)	ARIMA(0,1,0)	ARIMA(1,1,0)	
RMSE	0.020	0.029	0.019	0.023	**0.023**
MAPE	0.206	0.315	0.255	0.285	**0.265**
MASE	0.552	0.472	0.768	1.122	**0.729**
VAR(p)/VECM(p)	VEC(1)	VAR(2)	VEC(1)	VEC(1)	
RMSE	0.024	0.040	0.021	**0.022**	0.027
MAPE	0.254	0.429	0.280	**0.283**	0.311
MASE	0.681	0.642	0.846	**1.120**	0.822
ANN(x,m,s)	ANN(3,2,1)	ANN(3,2,2,1)	ANN(3,3,1)	ANN(3,2,1)	
RMSE	0.127	0.280	0.023	0.075	0.126
MAPE	1.810	3.960	0.305	1.104	1.795
MASE	4.854	5.910	0.920	4.368	4.013

Out-sample period is 13 weeks from 10/6/2017 to 12/29/2017.

The forecasting accuracy measures in Table 7 and 8 (RMSE, MAPE and MASE) enables to determine, which forecasting models perform better or worse based in terms of out-sample forecasting accuracy but do not allow to establish whether there is a statistically significant difference present. In the existing shipping freight rate forecasting literature, while there have been many attempts to compare forecasting models (Chen *et al.*, 2012; Duru, 2012; Munim and Schramm, 2017), to the best of the authors' knowledge, the statistical significance of forecasting accuracy difference has rarely been examined. To address this, we use the DM test (Diebold and Mariano, 1995) represented below in the test statistic equation (8).

$$\mathrm{DM} = \bar{d}/\sqrt{2\pi\widehat{f_d}(0)T} \tag{8}$$

Here, $\bar{d} = \frac{1}{T}\sum_{t=1}^{T}[g(e_{it}) - g(e_{jt})]$ is the sample mean loss differential, where e_{it} and e_{jt} represent forecasting errors from two different models; and $g(e_{it})$ and $g(e_{jt})$ denote their respective loss functions. Furthermore, $\widehat{f_d}(0)$ is a consistent estimate of $f_d(0)$; in the null hypothesis of zero mean loss differential ($\mu = 0$) and assuming covariance stationarity, the asymptotic distribution of the sample mean loss differential $\sqrt{T}(\bar{d} - \mu)$ converges in distribution to $N(0,2\pi f_d(0))$ (Diebold and Mariano, 1995: 135).

The results of DM test are presented in Table 9. We find that forecasting results of ARIMA models are significantly better than others for the FE-NEU and FE-MED trade lanes. For the FE-USWC route, forecasting results of all models are identical, while for the FE-USEC route, forecasts of ARIMA and VAR are identical, but both are better than ANN.

Table 9: DM-test for comparing forecasting accuracy

Forecasting results compared	FE-NEU	FE-MED	FE-USWC	FE-USEC
(M1) ARIMA versus (M2) VEC (or VAR)	-1.790*	-1.875*	-1.462	0.831
(M1) ARIMA versus (M2) ANN	-11.304***	-7.273***	-0.897	-8.893***
(M1) VEC (or VAR) versus (M2) ANN	-10.931***	-7.127***	-0.431	-8.587***

**p<0.05, **p<0.01, ***p<0.001; for DM-test, P < 0.05 indicates forecasting results of the first method (M1) is better than the second method (M2).*

5. Discussion and conclusion

This study presents an investigation of competing forecasting models of container shipping freight rates, for four major trade routes, namely, FE-NEU, FE-MED, FE-USWC and FE-USEC. Three forecasting models are scrutinised: ARIMA, VAR/VEC and ANN. For robustness of forecasting accuracy checks, we use three accuracy measures, including the mean absolute scaled error (MASE). This measure is free from sample-size bias, and a value of MASE below one in the training-sample forecasts indicates better accuracy than naïve forecasts. Furthermore, we employ the DM test (Diebold and Mariano, 1995) to scrutinise the superiority of forecasting results achieved from different models in the context of container freight forecasting.

Empirical results show mixed findings. On the aggregate level, considering the means of the forecasting accuracy measures of each model for each trade route, VAR/VEC outperforms others in the in-sample forecasts, but ARIMA outperforms others in the out-sample period. This is similar to Bachelor *et al.* (2007), who also find that simple ARIMA models perform better for forward freight rate forecasting. While this result holds for most of the routes, there are two exceptions. On the route level, interestingly, ARIMA performs better than others for the in-sample forecasts on the FE-MED route. For out-sample, the VEC model performs better than others on the FE-USEC route. Also, for the out-sample forecasts of the FE-USWC route, forecasting results of all models are identical (see Table 9). From Figure 1, it is evident that container freight rates for the FE-MED route in the investigated period are the most volatile, starting at the highest of 1,581 (2/13/2015), then dropping to the lowest of 543 (4/8/2016) and then again reaching the peak at 1,264 (7/21/2017). These freight rate movements may be triggered by changing service patterns in that region switching from transhipment to direct call operations followed by an influx of shipping capacity resulting in temporary supply demand imbalances. On contrary, freight rate movement on the FE-NEU route show volatility due to a persistent old habits of liner shipping operators dominated by recurrent announcements of GRI, BAF or other similar surcharges on top of the base rate. The FE-USWC and FE-USEC routes follow a similar pattern – but to a much lesser extent may be because the existence of TSA.

While the underlying reasons for mixed forecasting results should be further explored in future studies, similar to Munim and Schramm (2017), it is evident that the ARIMA is still the most suitable model for container freight rate forecasting (see Table 8 and Table 9). One reason for this could be that container freight rates are significantly associated with their previous values (accounted well by ARIMA) rather than exogenous variables like week-to-week change in the fleet capacity, newbuilding prices or charter price indexes. For instance, the multivariate modelling parameters in Table 6 show that the considered exogenous variables do not have any statistically significant association with container freight rates, except for the week-to-week fleet capacity change in the FE-NEU and FE-MED trade lanes.

Due to the high volatility of container freight rates over time, particularly on a weekly level, the structural pattern of data changes quickly. Therefore, in future research, expanding and rolling forecasting windows should be examined to validate the findings of the forecasting models of each route. Apart from the traditional models, fuzzy time series forecasting approaches should be explored for the modelling of container freight rate volatility. Researchers should always try to develop new forecasting methods, to improve the results of earlier models. Finally, as container freight rate volatility reflects also the psychological behaviour of shipowners and investors, combining judgemental and statistical models could be an option.

Acknowledgement

The authors would like to thank the anonymous reviewers and Professor Hercules Haralambides for useful suggestions. An earlier version of the manuscript has been presented at the International Association of Maritime Economists 2018 conference at Mombasa, Kenya.

References

Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, 19(6): 716–72.

Avdasheva, S., Golovanova, S. and Yusupova, G. (2018). Advance freight rate announcements (GRI) in liner shipping: European and Russian regulatory settlements compared, *Maritime Economics & Logistics*, 21(2), 192–206 .

Batchelor, R., Alizadeh, A. and Visvikis, I. (2007). Forecasting spot and forward prices in the international freight market. *International Journal of Forecasting*, 23(1), 101-114.

Bollerslev, T. (1986). Generalized autoregressive conditional heteroscedasticity. *Journal of Econometrics*, 31(3), 307-327

Box, G. E. and Jenkins, G. M. (1976). Time series analysis: forecasting and control. revised ed., San Francisco: Holden-Day.

Chen, S., Meersman, H. and Van de Voorde, E. (2012). Forecasting spot rates at main routes in the dry bulk market. *Maritime Economics & Logistics*, 14(4), 498-537.

Clarksons (n.d.). SIN - Shipping Intelligence Network, https://sin.clarksons.net/.

Clarksons (2017). Container Intelligence Quarterly Q2/2017, London: Clarksons Research.

Clarksons (2019). Container Intelligence Quarterly Q4/2019, London: Clarksons Research.

Diebold, F. X., & Mariano, R. S. (1995). Comparing predictive accuracy. *Journal of Business & Economic Statistics*, 13(3), 134-144.

Drewry (2012). Index-Linked Container Contracts. London: Drewry Shipping Consultants.

Duru, O. (2012). A multivariate model of fuzzy integrated logical forecasting method (M-FILF) and multiplicative time series clustering: A model of time-varying volatility for dry cargo freight market. *Expert Systems with Applications*, 39(4), 4135-4142.

Engle, R. F. (1982). Autoregressive Conditional Heteroskedasticity with Estimates of the Variance of UK inflation. *Econometrica* 50(4), 987-1008.

Engle, R. F. and Granger, C. W. (1987). Co-integration and error correction: representation, estimation, and testing. *Econometrica* 55(2), 251-276.

Fan, L. (2011). Econometric analyses of container shipping market and capacity development. PhD Thesis, The Hong Kong Polytechnic University, Hong Kong.

Gharehgozli, A., Duru, O., and Bulut, E. (2018). Input data range optimization for freight rate forecasting using the rolling window testing procedure. *International Journal of Transport Economics*, 45(3), 393-413

Günther, F. and Fritsch, S. (2010). neuralnet: Training of neural networks. *The R journal* 2(1), 30-38.

Hair, J. F., et al. (2010). Multivariate Data Analysis. Upper Saddle River, NJ, Pearson.

Haralambides, H.E., Fusillo, M., Hautau, U., Sjostrom, S., & Veenstra, A. (2003). The "Erasmus Report": Global Logistics and the Future of Liner Shipping Conferences. COMP/2002/16/SI2.346635. European Commission Competition Directorate General.

Haralambides, H. E. (2019). Gigantism in container shipping, ports and global logistics: a time-lapse into the future. *Maritime Economics & Logistics*, 21(1), 1-60

Harper Petersen (n.d.). HARPEX, http://www.harperpetersen.com/harpex/harpexRH.csv.

Howe Robinson (n.d.). https://www.howerobinson.com.

Hyndman, R. J., & Koehler, A. B. (2006). Another look at measures of forecast accuracy. *International Journal of Forecasting*, 22(4), 679-688.

Hyndman, R. J., & Khandakar, Y. (2008). Automatic time series for forecasting: the forecast package for R (No. 6/07). *Journal of Statistical Software*, 27(3), 1-22.

Jarque, C.M. and Bera, A.K. (1980). Efficient tests for normality, homoscedasticity and serial independence of regression residuals. *Economics Letters* 6(3), 255–259.

Johansen, S. (1991). Estimation and hypothesis testing of cointegration vectors in Gaussian vector autoregressive models. *Econometrica*, 59 (6), 1551-1580.

Karamperidis, S., Jackson, E., and Mangan, J. (2013). The use of indices in the maritime transport sector. *Maritime Policy & Management*, 40(4), 339-350.

Kavussanos, M. G., & Nomikos, N. K. (2003). Price discovery, causality and forecasting in the freight futures market. *Review of Derivatives Research*, 6(3), 203-230.

Kavussanos, M. G., Visvikis, I. D. and Batchelor, R. A. (2004). Over-the-counter forward contracts and spot price volatility in shipping. *Transportation Research Part E: Logistics and Transportation Review* 40(4), 273-296.

Kavussanos, M.G., Visvikis, I.D., and Dimitrakopoulos, D.N. (2015). Freight Markets and Products. In: A. Roncoroni, G. Fusai and M. Cumrins (eds.) Handbook of Multi-Commodity

Markets and Products: Structuring, Trading and Risk Management. John Wiley & Sons, pp. 355-398.

Lau, Y-Y., Ng, A. K.Y., Fu, X., and Li, K.X. (2013). Evolution and research trends of container shipping, *Maritime Policy & Management*, 40(7), 654-674.

Li, K. X., Qi, G., Shi, W., Yang, Z., Bang, H.-S., Woo, S.-H., and Yip, T. L. (2014). Spillover effects and dynamic correlations between spot and forward tanker freight markets. *Maritime Policy and Management* 41(7), 683-696.

Ljung, G.M. and Box, G.E. (1978). On a measure of lack of fit in time series models. *Biometrika*, 65(2), 297–303.

Lütkepohl, H. (1985). Comparison of criteria for estimating the order of a vector autoregressive process. *Journal of Time Series Analysis*, 6(1), 35-52.

Luo, M., Fan, L. and Liu, L. (2009). An econometric analysis for container shipping market. *Maritime Policy and Management* 36(6), 507-523.

Miller, K., Ward, R., Craig, A.W., Lowe, E.K., Kroll, C.A., Hale, T., Miller, E., Cassels, J., Schöne, S. and Höth, A. (2015). Risk Management and Applications. In O. Schinas, C. Grau and M. Johns (eds.) HSBA Handbook on Ship Finance. Springer: Berlin - Heidelberg, pp. 297-321.

Munim, Z.H. and Schramm, H.J., (2017). Forecasting container shipping freight rates for the Far East–Northern Europe trade lane. *Maritime Economics & Logistics*, 19(1), 106-125.

Narzo, Di A. F., Aznarte, J. L., and Stigler, M. (2018). Nonlinear Time Series Models with Regime Switching. Accessed on June 10, 2018 from https://cran.r-project.org/web/packages/tsDyn/tsDyn.pdf

Nielsen, P., Jiang, L., Rytter, N.G.M. and Chen, G., (2014). An investigation of forecast horizon and observation fit's influence on an econometric rate forecast model in the liner shipping industry. *Maritime Policy & Management*, 41(7), 667-682.

Notteboom, T. and Cariou, P., (2013). Slow steaming in container liner shipping: is there any impact on fuel surcharge practices?. *The International Journal of Logistics Management*, 24(1), 73-86.

De Oliveira, G.F. (2014). Determinants of European freight rates: The role of market power and trade imbalance. *Transportation Research Part E: Logistics and Transportation Review*, 62, 23-33.

PanCanal (2017) Panama Canal Welcomes Largest Capacity Container Vessel To-Date Through Expanded Locks, Press Release, August 22, 2017, http://www.pancanal.com/eng/pr/press-releases/2017/08/22/pr632.html

Phillips, P.C.B. and Perron, P. (1988). Testing for a Unit Root in Time Series Regression, *Biometrika*, 75(2), 335–346.

Porter, J. (2016). Four Weddings and a Funeral, *Containerisation International* November Issue, 21-24.

Shi, X. (2000). The Study on the Compilation of the China Container Freight Index. *Maritime Policy & Management* 27(3), 303–308. doi:10.1080/030888300411121.

Signals DPI (2018). TSA Carrier Group Cancels FMC Agreement, Closes Office and Deletes Websites, *Signals* 22(3), March 2, 2018, https://www.dpiusa.com/archives/signals/180302.html#B.

Slack, B., and Gouvernal, E. (2011). Container Freight Rates and the Role of Surcharges. *Journal of Transport Geography* 19(6), 1482–1489.

SSE (n.d.). Shanghai Shipping Exchange. http://en.sse.net.cn/index.jsp.

SSE (2019a). About SCFI. http://en.sse.net.cn/indices/introduction_scfi_new.jsp.

SSE (2019b). About CCFI. http://en.sse.net.cn/indices/introduction_ccfi_new.jsp.

Stopford, M. (2009). Maritime Economics (3rd ed.). New York: Routledge.

Tsioumas, V., Papadimitriou, S., Smirlis, Y. and Zahran, S.Z. (2017). A Novel Approach to Forecasting the Bulk Freight Market. *The Asian Journal of Shipping and Logistics*, 33(1), 33-41.

UNCTAD (2019). Review of Maritime Transport. Geneva: United Nations Conference on Trade and Development.

Veenstra, A. W., & Franses, P. H. (1997). A co-integration approach to forecasting freight rates in the dry bulk shipping sector. *Transportation Research Part A: Policy and Practice*, 31(6), 447-458.

VHBS (n.d.). HAX - HAMBURG Index, https://www.vhbs.de/index.php?id=13&L=1

Wackett, M. (2017). The ups and downs of container vessel scrapping as charter rates rise and fall. *The Loadstar* 14/06/2017, https://theloadstar.co.uk/ups-downs-container-vessel-scrapping-charter-rates-rise-fall/.

Wang D.H., Chen C.C., and Lai C.S. (2011). The rationale behind and effects of Bunker Adjustment Factors. *Journal of Transport Geography*, 19(4), 467-74.

Yifei, Z., Dali, Z. and Yanagita, T. (2018). Container liner freight index based on data from e-booking platforms, *Maritime Policy & Management*, 45(6), 739-755. DOI:10.1080/03088839.2018.1443224.

Yin, J., Luo, M. and Fan, L. (2017). Dynamics and interactions between spot and forward freights in the dry bulk shipping market. *Maritime Policy & Management*, 44(2), 271-288.

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

Yun, H., Lim, S., and Lee, K. (2018). The value of options for time charterparty extension: an artificial neural networks (ANN) approach. *Maritime Policy & Management*, 45(2), 197-210.

Zamorano, J. and Martinez, K. (2016). Panama Canal opens $5B locks, bullish despite shipping woes. The Big Story. Associated Press June 26, 2016. Retrieved April 13, 2017.

Zhang, X., Xue, T., & Stanley, H. E. (2019). Comparison of econometric models and artificial neural networks algorithms for the prediction of baltic dry index. *IEEE Access*, 7, 1647-1657.

Zhang, H. and Zeng, Q., (2015). A study of the relationships between the time charter and spot freight rates. *Applied Economics*, 47(9), 955-965.