Judgmental Forecasting in the Dry Bulk Shipping Business: Statistical vs. Judgmental Approach

Article *in* Asian Journal of Shipping and Logistics · December 2009

DOI: 10.1016/S2092-5212(09)80002-3

2 authors:

Okan Duru
Nanyang Technological University
64 PUBLICATIONS **370** CITATIONS
SEE PROFILE

Shigeru Yoshida
Kobe University
30 PUBLICATIONS **231** CITATIONS
SEE PROFILE

The Asian Journal of Shipping and Logistics ● Volume 25 Number 2 December 2009 pp.189-217 ●

Judgmental Forecasting in the Dry Bulk Shipping Business: Statistical vs. Judgmental Approach

Okan DURU* · Shigeru YOSHIDA**

Contents

I. Introduction
II. Qualitative Factors of Shipping Markets
III. Decision-Supported Structure of Methodology in Dry Bulk Freight Forecasting
IV. Empirical Studies
V. Empirical Results
VI. Conclusion

Abstract

This paper investigates the accuracy of judgmental forecasting methods for dry bulk freight market. Judgmental forecasting is a method that implements the purely judgmental aspects of a system. The research investigates judgmental forecasting for the Baltic Dry Index, and compares it with statistical extrapolation methods. Post-sample forecasting accuracy indicates the superiority of the proposed approach over the conventional statistical methods. Empirical studies are performed with an expert group and the accuracy of judgmental point estimates is compared with exponential smoothing, the X12 ARIMA, and the TRAMO/SEATS algorithms. The results of expert predictions outperformed traditional time-series methods.

Key Words: Judgmental forecasting, freight market, expert opinion, Delphi

* PhD candidate in Kobe University, Japan and research staff in Istanbul Technical University, Turkey. Email: duruokan@yahoo.com
**Professor of Maritime Logistics Science in Kobe University, Japan, Email: syoshida@maritime.kobe-u.ac.jp

I. Introduction

Forecasting aids decision-making in the shipping business and shipping investment management. For managers, analysts, and other practitioners in the shipping industry, the decision is the raison d'être and is based on forecasts even if forecasts are inferior.

Forecasts are important inputs of investment management and are necessarily included in shipping business decisions.

In economics, statistical forecasting tools are used by many researchers. Statistical forecasting methods currently available assume that constituted patterns or relationships will not change in the post-sample forecasting phase. However, this is not a realistic assumption for economic objectives. Pattern changes, structural brakes, and purely judgmental behavior exist in the shipping business environment. The forecasting of repetitive events is accurately managed by quantitative extrapolation methods. In shipping business matter, the historical data consists of many unusual cycles characterized by crowd feed, mass psychology, or other judgmental impacts of the market.[1]

Several methodologies are developed accordingly to account for exogenous factors that affect market route.

Judgmental assessment of forecasting studies are necessary with whatever approach is used. Some of the more complicated quantitative methodologies require subjective assessment and experienced judgment to apply the method appropriately. Moreover, criteria for selecting the most appropriate methodology may be largely subjective.

Goodwin and Fildes[2] divide time-series into two parts: a normal period and a special period. The normal period is the time with no unusual events. Therefore, statistical methods may provide an appropriate solution for this case. On the other hand, the special period includes unusual and sometimes unexpected events. In this case, statistical methods give only baseline forecast based on historical patterns. A judgment-

1 Stopford (1997).
2 Goodwin and Fildes (1999), pp.37-53.

sensitive method is necessary for understanding subjective trends.

The judgmental process of management plays a determinative and diagnostic role in balancing countering expectations and financial aspects. Most of the market is driven by judgmental impacts of individual shareholders of the shipping business. Therefore, the inability of quantitative methods to account for purely judgmental objects is one of the most important motivations of the subjective forecasting discussed in this paper.

Quantitative analysis and forecasting of shipping markets has been an attractive research stream in the last fifty years as an interdisciplinary field of research. Most of the research on the quantitative analysis of shipping markets has focused on time-series analyses and econometric models. Studies show that the shipping markets appear to be influenced mainly by the trade flow, fleet capacity, shipbuilding schedule, scrapings, and industrial production.[3] The importance of trade flows on the freight market is repeatedly accentuated in variety of studies. Econometric models[4], statistical extrapolation works[5], co-integration[6], and volatility forecasting methods[7] are suggested by many researchers.

Although recent theoretical and empirical developments have improved our understanding of the role of quantitative methods in shipping market forecasts, gaps still exist in the shipping market forecast literature. One gap stems from the focus that most studies have on a limited set of statistical methods, rather than including a pure judgmental or judgmentally-adjusted set of theoretically relevant forecasts for the shipping market. The present study aims to identify a theoretical and empirical approach for judgmental forecasting in the freight market. Specifically, we proposed that forecasts in shipping markets are influenced by two situational factors: historical background, which is mainly time-series data of the market; and behavioral attribution, which has been found to be the most important appraisal dimension. These situational factors lead to the determination of shipping markets.

3 Tinbergen (1959); Hawdon (1978), pp.203-217; Shimojo (1979); Norman (1979); Beenstock & Vergottis (1993).
4 Tinbergen (1959); Norman (1979); Charemza & Gronicki (1981), pp.21-30; Beenstock & Vergottis (1993).
5 Cullinane (1992), pp.91-114.
6 Hale & Vanags (1992), pp.31-39; Glen (1997), pp.245-260; Veenstra (1997), pp.447-458.
7 Kavussanos (1996, 1997).

The present study contributes to the literature with a unique analysis of judgmental forecasting in the dry bulk freight market and a comparative analysis with time-series methods. It explores judgmental forecasting of shipping markets, discusses a theoretical framework for the judgmental forecasting of shipping markets, and extends the recent literature to behavioral aspects of the shipping sector and an available model of forecasting by expert judgments. Empirical studies utilize the Baltic Dry Index (BDI)[8] as an indicator of the dry bulk freight market.

In the bulk of this paper, we review various univariate time-series forecasting methods in freight market practice, with particular reference to their short-term accuracy, and present the results of empirical studies designed to assess the performance of some of these methods on real data relative to judgmental methodology. Section 2 contains a brief description of the qualitative factors in shipping markets. Section 3 presents decision support for shipping forecasting. Section 4 explains methodology used in empirical studies and application tasks. Section 5 introduces the results of the procedures and out-of-sample performance. Finally, Section 5 presents concluding remarks and suggestions for future research.

II. Qualitative Factors of Shipping Markets

Behavior and psychology of shipping markets plays an orienting role in determining prices. When Hampton[9] mentions the psychology of the marketplace, the psychology of crowd feeds and human emotions are described as important drivers of the shipping market. Stopford also clarifies shipping cycles that have a distinctive character and are affected by crowd psychology. In recent experiences with market cycles, different factors are considered that influence shipping markets and maritime trade in the context of economic theory. Shipping market history consists of miscellaneous events that

8 Baltic Dry Index:BDI is a composite index of several dry bulk routes and cargoes. It is determined by the panelist members of the Baltic Exchange (www.balticexchange.com) regarding the standard vessel, cargo, and route specifications.
9 Hampton (1991).

affected the development of the market. Furthermore, shipping is concerned with international relations and politics, which are prominent drivers of global trade and seaborne transportation.[10]

Judgmental factors play a major role in influencing market behavior in general and freight price determination in particular. Judgmental factors may influence the freight market in two primary ways. First, judgments of shipping stakeholders influence overall expectations in the shipping market. The aggregate expectation of freight rates is constituted by individual players and leads to price identification. Second, historical data of shipping markets are compiled through a cognitive process of stakeholders and applied to practical business using subjective reasoning.

<Figure 1> Price determination environment of freight market

The price determination environment of freight rates is integrated by internal and external factors of the shipping business. The freight negotiation process is a behavioral part of the internal factors and is affected by industrial issues and investment aspects.

The psychology of negotiation influences task results, and also domination of counterparts and personal psychology constitute overall interaction. Other internal

10 Stopford (1997).

factors are mainly caused by trading balances of supply and demand of shipping.[11]

External factors span the entire shipping business and lead to subjective changes or provide frontiers of trade changes. As opposed to a strictly economic viewpoint, the overall picture consists of many subjective variables that can be handled merely by judgmental forecasting. Fig. 1 shows the structure of factors that affect price determination.

In 2004 and 2005, the dry bulk freight market was faced with unusual changes by various policy implementations by China. The Prime Minister of the People's Republic of China, Wen Jiabao, declared new restrictive investment policies on main industries (e.g., steel, real estate; the major impact was considered to be on steel production volume) on April 28, 2004. The declaration influenced the BDI to fall sharply from 4,229 to 2,622 over a period of two months. However, in actual numbers, the change in Chinese crude steel production was not more than 1%. Crude steel and steel products hold and increase the current volume of production. In 2005, a similar declaration was released and the BDI decreased from 4,804 in April to 1,769 in August. This type of unusual and enormous price variation could not be described by trade volume or a simple supply-demand perspective. Economic theory was unable to explain freight price behavior from the perspective of subjective reasoning. The judgmental forecasting approach is one of the unique solutions for subjective reasoning and the extraction of market perception and crowd feed behavior.

An empirical study of subjective analysis of the shipping market was performed by Ariel.[12] Ariel used a Delphi-based forecast for the shipping industry for 1985-2000.

This study collected a variety of considerations from a number of practitioners from the international maritime society, commodity trading, ship-broking, ship management, port authorities, shipping research institutes, etc. The study built a long-term scenario for the shipping industry to obtain Delphi group forecasts. Forecasting items were

11 Supply-demand equilibrium theory is investigated in the context of maritime transportation by various publications. For further reading, please refer to McConville(1999), Stopford(1997), or Veenstra(1999).
12 Ariel (1989), pp.305-336.

mainly stated in terms of shipping fleet trade volume and marine technology. Fleet and trade focused items partially provided reasonable performance on forecasts. On the other hand, technology-focused items were not accurate for forecasting the marine technology cycle. This study was a long-term rather than a short-term based understanding of the freight market.

Wars, canal disputes, oil shocks, and economic boom and bust cycles were all experienced in the Twentieth Century. Their effects could not be forecasted by quantitative methods, but the society of the shipping sector expected coming shocks by the frontiers of shocks (declarations, policy changes, news etc.). The response and behavior of the shipping society determine how the freight market should be structured and valued. If an unexpected event occurs, quantitative methods may not support a quick response to market dynamics and behavior. Because, quantitative methods need a historical background and similar events must be experienced before. However, there are empirical studies that suggest judgmental forecasts outperform statistical techniques, particularly in unforeseeable situations.[13]

III. Decision-Supported Structure of Methodology in Dry Bulk Freight Forecasting

A decision-supported structure of forecasting methodologies for the shipping business is suggested by Duru and Yoshida.[14] Forecasting studies focus on quantitative solutions for problems in maritime transportation. Qualitative equivalents and composite structure means quantitative methods can be adjusted by decision-supported methods (qualitative methods), or a qualitative method itself is entirely accurate for forecasting rather than a hybrid model or adjustment approach.

Forecasting methods for shipping markets (in this paper, dry bulk freight market)

13 Clemen (1989), pp.559-583; Goodwin and Wright (1993),pp.147-161; Lawrence and O'Connor (1992),pp.15-26; Sanders (1992), pp.353-364.
14 Duru and Yoshida (2008a,b).

have two major divisions. The first is a pure judgmental approach formed by consulting expertise. Expert opinion and Delphi panel methods are popular judgmental approaches to prediction tasks. The second category is quantitative methods, which include many different statistical and econometric procedures. Fig. 2 presents divisions of forecasting and main processes of forecasting tasks. The present paper focuses on single method applications and comparative analysis.

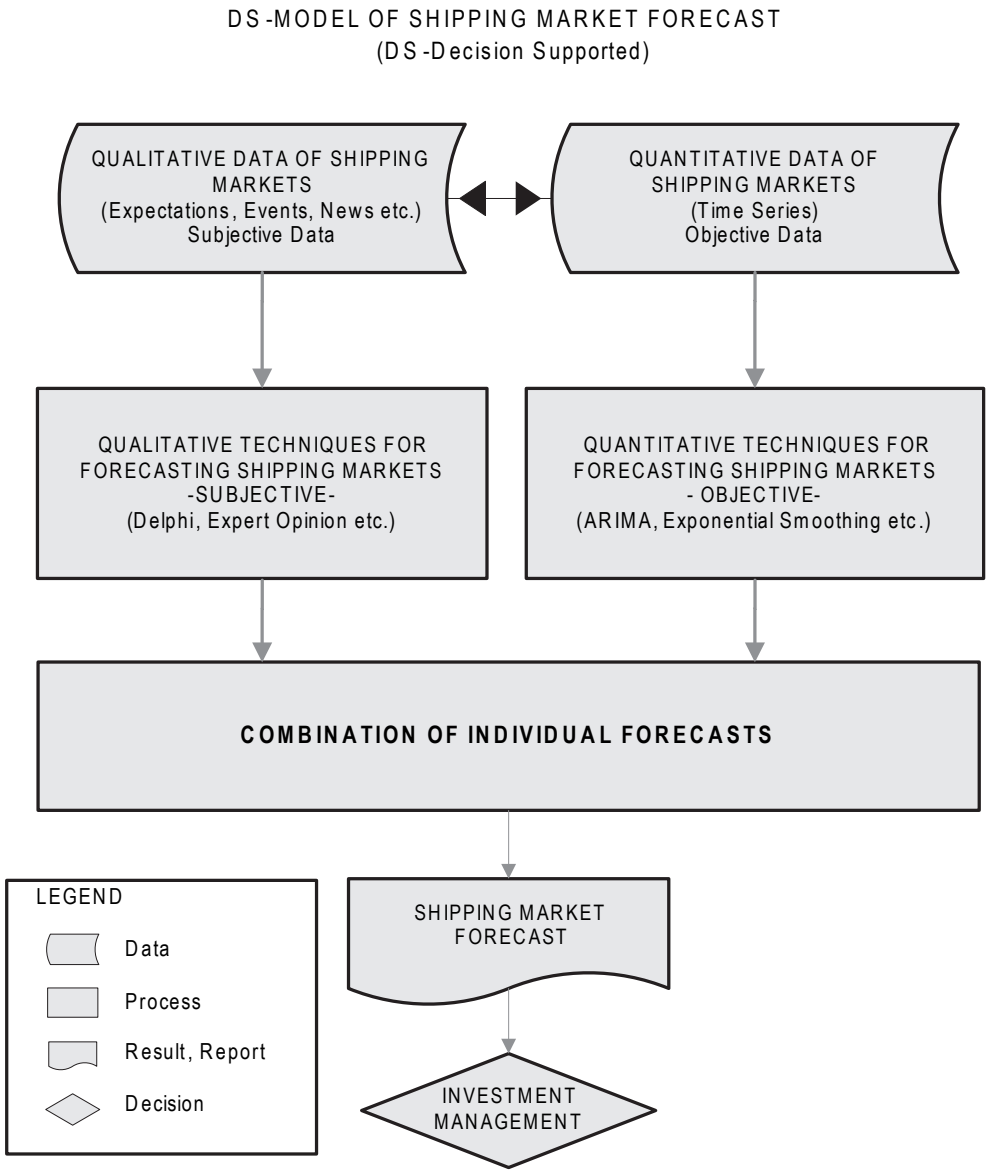
The critical role of judgment in the forecasting of economic and financial issues is frequently emphasized by existing research.[15] The main problem is bias and judgment heuristics. For that reason, task formats are designed to decrease bias and heuristics, and decision methods are tested for accuracy in forecasting. Accurate decision methods and effective task formats are key factors in forecast performance. Expert opinion and Delphi consensus decision methods are chosen for our forecasting study due to empirical studies. The dichotomous task format is defined as an expert-based experiment.

For quantitative research, a variety of parametric methods are available for forecasting. The shipping business literature provides many different approaches to forecasting problems, most of which are approached with statistical time-series methods. Furthermore, this field can be extended by other possible approaches.

Hybrid forecasting models are proposed by combining judgmental and statistical extrapolations. The practical and particularly accurate combination algorithm is still not suggested for freight markets by any researcher. In classical forecasting theory, simple average composition or judgmental adjustment of statistical results are the most referenced strategies for combination tasks.

15 i.e., Clemen (1989), pp.559-583; Bunn and Wright (1991); Goodwin and Wright (1993), pp.147-161; Önkal and Muradoglu (1996), pp.9-24; Goodwin (2005), pp.8-12.

<Figure 2> Decision-supported model of freight market forecast
(classification of three phases: objective, subjective and combining of both results)[16]

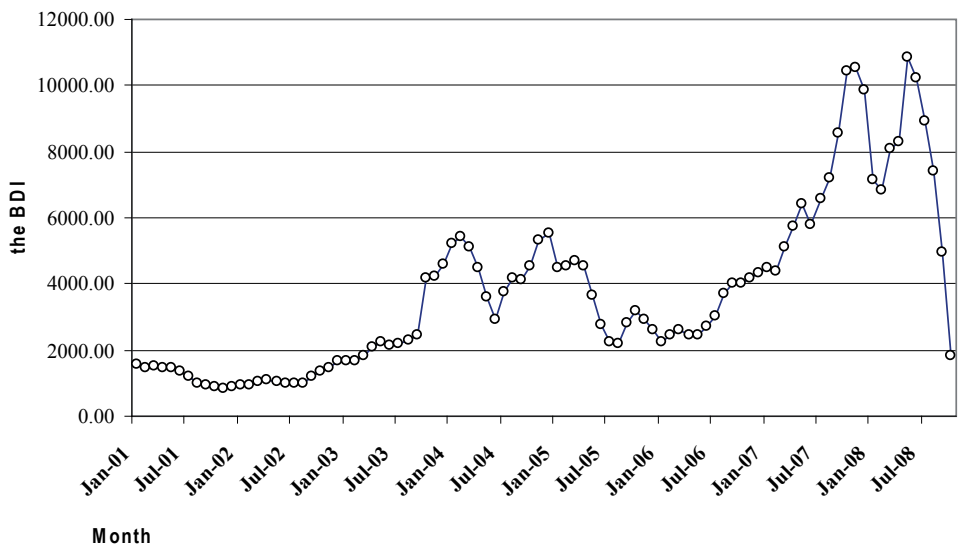


16 Classification of three phases: objective approach, subjective approach, and a combination of both. The objective approach is sourced by statistical data and mainly uses time-series or econometric modeling methods. The subjective approach is sourced by judgments of an expert group, board of directors, or other possible subjective information providers. Recent forecasting literature suggests a combination for a hybrid solution. As defined in the results of this paper, judgmental forecasts provide better accuracy in most cases. A combination will result in worse predictions. Therefore, a hybrid solution is not suggested at this stage of research.

IV. Empirical Studies

The performance of decision support in freight forecasting is tested using two judgmental forecasting methods. For freight forecasting, the BDI is assumed as a reference for the dry bulk shipping market. The BDI is a composite index of freight rates in various dry bulk shipment routes. It is defined by Baltic Exchange panelists on every trading day. BDI indicates the overall situation of the dry bulk freight market. In this paper, monthly BDI data, which includes data between January 2001 and October 2008, is used for empirical analysis. Fig. 3 shows the BDI series provided by Tramp Data Services Co. Ltd. and NYK Line Research Group. BDI is used as a reference and participants are asked to predict BDI in a predetermined forecasting horizon and response duration.

<Figure 3> Baltic Dry Index: BDI, January 2001 – October 2008



The objective of this empirical study is divided into two parts. First, the study investigates the superiority of methods and hypothesis defined as:

H_1: Judgmental forecasting provides better forecast accuracy than statistical extrapolation methods in several accuracy metrics (Holt-Winters exponential smoothing, X12 ARIMA, and TRAMO/ SEATS).

The expert-based and Delphi-based studies are compared with statistical benchmark methods. Second, research analyzes the performance of judgmental methods under different specifications. For example, results of the expert-based study can be extracted by a simple average, freight negotiation experience weighted, or self-confidence weighted approach. Hypotheses are constructed under these approaches as:

H_{2a}: Judgmental forecasting provides better forecast accuracy using freight negotiation experience weighting than simple average of individual judgments.

H_{2b}: Judgmental forecasting provides better forecast accuracy using self-confidence level weighting than simple average of individual judgments.

The last objective is to understand the accuracy level under single- and multi-iteration methods. The research investigates:

H_{2c}: Multi-iteration group-based Delphi methodology improves forecast accuracy over the single-iteration individual-based expert opinion.

We try to suggest the best judgmental process by testing different configurations.

1. Methodology

1) Statistical Forecasts

There are many scholars who investigated forecasting dry bulk shipping markets, and ARIMA methodology is suggested among the more advanced equivalents. For instance, Batchelor et. al.[17] performed a series of forecasting exercises for spot and forward

17 Batchelor et. al. (2007), pp.101-114.

freight prices, and the results indicated serious reductions of the RMSE (root mean squared error) than the random walk benchmark model in most of the horizons. For instance, the 20-day horizon results ensured reductions about 36-61% for four different routes, and these results were superior than the VAR (vector autoregression) , VECM (vector error correction model) and S-VECM (a restricted VECM) as well.[18] Kavussanos and Alizadeh[19] also briefly review such an ARIMA application for stochastic seasonality case, and point out spurious regression issue without a seasonal differencing and related theoretical works. Seasonal unit roots are also investigated in the present paper among the intended data series of the BDI.

Judgmental forecasts are compared with statistical benchmarking methods, including fully automatic algorithms and exponential smoothing approaches. Quantitative forecasting is conducted using three different methods and forecasting programs. The Box-Jenkins type ARIMA with Census Bureau X12 forecasting algorithm is one of the statistical methods used in this study.[20] The X12 ARIMA program is the primary method used for seasonal adjustment of governmental and economic time-series in the United States, Canada, and the European Union. Originally developed by the U.S. Census Bureau, it is based on a ratio-to-moving-average decomposition and includes a number of improvements developed through empirical testing since its release in 1996.[21] The Census Bureau's X12 ARIMA program essentially includes all the capabilities of the latest version of the X11 ARIMA and the X11. The X12 ARIMA also provides extensive model selection diagnostics based on out-of-sample forecast performance.[22]

TRAMO (Time-series Regression with ARIMA noise, Missing values, and Outliers) is a program for estimating and forecasting regression models with errors that follow non-stationary ARIMA processes, when there may be missing observations in the series,

18 This study compares ARIMA, VAR, VECM and S-VECM to random walk model in different horizons (beginning with 1-day ahead and lasting with 20-day ahead predictions). The results show a gradually increasing treatment by the ARIMA model when the horizon increases, and the ARIMA type model outperforms other methods. The VAR model is only accurate in the first 3-day ahead period than the conventional ARIMA for the spot freight rates.
19 Kavussanos and Alizadeh (2001), pp.443-467.
20 Box and Jenkins (1970).
21 Makridakis, Wheelwright, and Hyndman (1998).
22 Ladiray and Quenneville (2001, 2004).

and when there might be contamination by outliers and other special (deterministic) effects.[23] Program SEATS (Signal Extraction in ARIMA Time-series) uses the ARIMA-Model-Based (AMB) methodology to estimate unobserved components in series that follow ARIMA models. In SEATS, the unobserved components are the trend-cycle, p(t), seasonal, s(t), transitory, c(t), and irregular, u(t), components as in Eq. (1)

$$x(t) = p(t) + c(t) + u(t) + s(t) = n(t) + s(t), \quad (1)$$

where n(t) denotes the seasonally adjusted series. Broadly speaking, the trend-cycle captures the spectral peak around the zero frequency; the seasonal component captures the spectral peaks around the seasonal frequencies; the irregular component picks up white-noise variation; and the transitory component captures highly transitory variation that differs from white noise.[24]

Exponential smoothing is a method to estimate the local level of a time-series data as a weighted average of the most recent observations and the preceding estimate of level. X_t series shall be smoothed by Eq. (2).

$$F_t = \alpha X_t + (1-\alpha)F_{t-1} \qquad 0 < \alpha < 1 \quad (2)$$

Eq. (2) is the general form of exponential smoothing. The smoothed series, F_t, is a weighted average of current and previous values of X_t with weights, α, decreasing exponentially. Eq. (3) shows expanded form of F_t with its components as follows

$$F_t = \alpha \sum_{j=0}^{\infty} (1-\alpha)^j X_{t-j} \quad (3)$$

After the latest smoothed value, forecast of X_t in m forecasting horizon is

$$F_{t+m} = F_t \qquad m = 1, 2, 3\ldots \quad (4)$$

In the long-run, forecast function is in an absolute form (Eq. (4)) (Makridakis, Wheelwright, and Hyndman, 1998).

Holt and Winters[25] extended this simple approach and consolidated some improvements. The Holt-Winters forecasting method has been widely used in business and

23 Gómez and Maravall (1996).
24 Gómez and Maravall (1996).
25 Holt (1957); Winters (1960), pp.324-342.

financial applications. The structure of the Holt-Winters exponential smoothing is made up by local trend (T_t), level of series (F_t) (smoothed value) and seasonality (S_t). Eq. (5) shows the level of series smoothed by the constant α. Eq. (6) shows the seasonality component smoothed by the constant β. Eq. (7) shows the trend component smoothed by the constantg. Eq. (8) is the Winters's forecasting function that includes the components.

$$F_t = \alpha X_t / S_{t-p} + (1-\alpha)(F_{t-1} + T_{t-1}) \quad (5)$$

$$S_t = \beta X_t / F_t + (1-\beta) S_{t-p} \quad (6)$$

$$T_t = \gamma (F_t - F_{t-1}) + (1-\gamma) T_{t-1} \quad (7)$$

$$W_{t+m} = (F_t + mT_t) S_{t+m-p} \quad (8)$$

X_t : Actual value for period t

m : Number of period ahead to be forecasted

p : Number of periods in the seasonal cycle

The seasonality term can be written as additive. The Holt-Winters method assumes an additive trend, but a multiplicative time-series can be fitted to the additive trend pattern by logarithmic transformation or other valid transformations. Smoothing constants, α, β, γ, are determined by the best fitting configuration that minimizes average squared forecast errors.

Since the present paper attempts to use the mentioned statistical methods, the detection of whether seasonal factors have unit roots has a critical role. Many scholars indicated about the drawbacks of routine adjustment of seasonality.[26] Barsky and Miron[27] pointed out that the routine elimination of the seasonal cycles is concluded by losing important informations about the fluctuations. In existence of stochastic seasonality, Box-Jenkins type ARIMA modelling can be proper for an univariate series. However, it is reported that spurious regression results are probable if the series is not differenced at the seasonal frequency. Hyllberg et. al.[28] suggested a testing procedure

26 Barsky and Miron (1989), pp.503-534, Hyllberg et. al. (1990), pp.215-238.
27 Barsky and Miron (1989), pp.503-534.
28 Hyllberg et. al. (1990), pp.215-238.

for seasonal unit roots (HEGY test), and that method is also developed for monthly frequencies by Franses, and Beaulieu and Miron.[29]

A seasonal unit root test is performed for the series of the BDI, and the results are prospective. The historical backgrounds of two empirical study have seasonal unit roots, and adjustment or seasonal differencing will be rational. The tests are conducted by monthly HEGY test for the models that is based on intercept only, intercept and trend, and intercept, trend and seasonal dummies. Constant, trend and seasonal dummy variables are found insignificant in most. The significance tests of seasonal frequencies are conducted by the reference t-test and F-test tables of Franses and Hobijn.[30] The single and joint existence of seasonal unit roots are indicated present in the applied datasets. Therefore, the X12 ARIMA, Tramo/Seats and Holt-Winters' methods are defined to be consistent for comparative analysis.

2) Judgmental Forecasts

For judgmental forecasting, expert opinion and Delphi panel methods are proposed to improve the subjective exercise of the forecasting study. These methods are implemented in many judgmental forecasting studies. The dry bulk freight market is the first application area for short-term forecasting in shipping business literature.

Expert opinion is a basic method based on individual decisions and a single iteration application. Expert decisions are composed of individual forecast of percentage change of the BDI and personal confidence of participants for their indication.

The Delphi panel method is somewhat more complicated than expert opinion for improving consensus among panel members. It is a multi-iteration exercise (a minimum of two iterations are required and, in the most cases, practice has two iterations) and in every iteration, a feedback system ensures a brief understanding of overall response for participants. Through the feedback system, panel members have the ability to revise

29 Franses (1990); Beaulieu and Miron (1993), pp.305-328.
30 Franses and Hobijn (1997), pp.25-47.

their responses in between iterations. A moderator leads the Delphi session and keeps all information anonymous (identification of participants, individual responses, etc.).

One of the most important advantages of Delphi is the anonymous structure that eliminates social pressure and the drawbacks of meeting. It also balances individual attendance by using a simple average of responses. In the final iteration, judgments of every participant have an equal weight and effect on the outcome of the Delphi session.

3) Accuracy check of results

The comparative analysis of forecasting methods is accomplished using the absolute percentage error (APE), mean absolute percentage error (MAPE) and root mean squared error (RMSE) metrics. The APE gives percentage error rate of a single prediction (Eq. 9).

$$APE = \left| \frac{X_i - Y_i}{X_i} \right| \times 100 \qquad (i=1,\ldots, n) \tag{9}$$

The MAPE is an indicator commonly used in quantitative and qualitative forecasting studies. The major attribute of MAPE is that metric provides an average percentage deviation from actual series. Eq. (10) defines the MAPE function.

$$MAPE = \frac{1}{n} \sum_{i=1}^{n} \left| \frac{X_i - Y_i}{X_i} \right| \times 100 \qquad (i=1,\ldots, n) \tag{10}$$

The RMSE metric gives an average deviation interval, and increases effects of larger errors by squares of them. Eq. (11) indicates the RMSE function.

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n} \left(X_i - Y_i \right)^2}{n}} \qquad (i=1,\ldots, n) \tag{11}$$

2. Nature of forecasters

A number of experts were selected from the shipping industry and were mainly practitioners of freight negotiation or had some other freight negotiation background. Titles are included shipbroker, chartering manager, sales manager, or general manager. The origins of the experts are Turkey, U.K. and Singapore. For the purposes of the expert-based study, eight subjects attended and for the Delphi-based study, nine subjects participated. The efficient group size is also a judgmental decision. Rowe and Wright[31] reviews several Delphi studies on a seminal paper, and this paper provides a broad list of empirical studies which are presented in refereed scientific journals. The group size of the Delphi based studies are various, and a number subjects about 5-10 is a common situation. On account of the evidences of applied Delphi studies, the group size of the present research will be proper and consistent for our analysis.

Participants had freight negotiation expertise between two and ten years (average is 5 years). All participants were required to fill out a multiple interval forecast form, which were collected by a predetermined deadline. In the Delphi-based study, the forecasting task was completed in two iterations, so participants were required to provide their judgmental forecasts twice. In contrast, the expert-based study was single iteration and forecasts were provided only one time.

3. Task for expert-based study

The forecasting task involved two main stages. The first involved the derivation of the statistical (model-based) forecasts. Depending on the forecasting support system, this was either done automatically (with algorithms used to estimate the optimal forecasting method) or the user chose the forecasting method. In this study, the X12 ARIMA Census Bureau program, Holt-Winter's Exponential Smoothing, and TRAMO/SEATS ARIMA

31 Rowe and Wright (1999), pp.353-375.

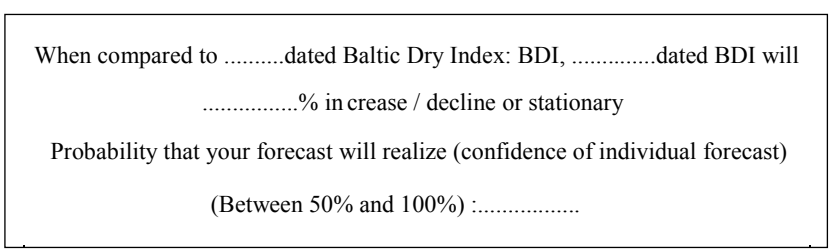
seasonal adjustment and decomposition programs were used to establish statistical extrapolation. The second stage involved judgmental forecasts of the expert group.

Judgmental forecasts were collected individually two weeks before the first forecasting point (two-week forecasting horizon). These forecasts included two-week, one month, three months, and six months horizons as point forecasts. Forecasts were anonymous and not revealed to each participant. In every completion of the forecast horizon, a summary result was sent to experts, including results on the mean and freight negotiation experience weighted and self-confidence grade weighted forecasts, along with their own prediction to compare their personal performance.

The expert group was required to fill out a personal information questionnaire to scale their experience in freight negotiation practice. Using experience ratings (years in practice), forecasts were adjusted to weight their individual ratings. Moreover, a self-confidence percentage was required from participants to grade their own confidence for provided judgmental predictions. Results were provided on pure forecast mean, freight negotiation experience weighted (FNE), and self-confidence level weighted base.

A dichotomous task format was provided for decision interface. As demonstrated in Fig. 4, the format consisted of percentage forecast of variation (increase, decrease, or firm condition) and personal confidence of their individual prediction. Confidence level was limited between 50% and 100% because a minimum confidence should be over 50%. Otherwise, the prediction was kept out.

<Figure 4> Dichotomous task format for the BDI forecasts



4. Task for Delphi-based study

The Delphi-based judgmental forecasting of BDI was performed with a nine-participant group. Participants were required to predict the BDI for two-week and one month horizons. The test was completed in two iterations and participants provided their initial forecast for the first iteration and revised forecasts for the second iteration. In the first iteration, participants ensured a single forecast of BDI for two horizons. After the first iteration, a summary report was returned to the panel members and they were required to check their individual prediction and revise it if a correction was necessary.

In the second iteration, the Delphi experiment was terminated and a simple average of the second round was recorded as the final result of the study. For analysis of multi-iteration efficiency, both the second and first round outcomes were indicated in the result with statistical equivalents.

V. Empirical Results

1. Comparison of statistical and judgmental forecasting methods

1) Overall results

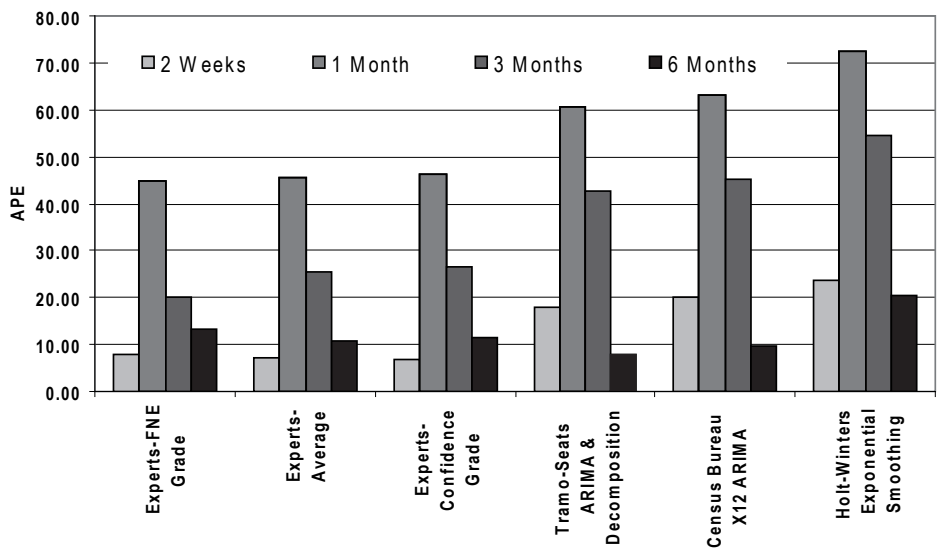
The results of Delphi-based forecasting and expert opinion-based forecasting indicate an improvement of forecasting accuracy (Table 1). In most of the empirical studies, statistical benchmark methods are inferior. Judgmental forecasting approaches provide lesser MAPE results even these methods do not require statistical data series as an input.

The RMSE results are also prospective. The overall RMSE metric denotes better performance results for judgmental studies.

2) H_1: Judgmental forecasting vs. statistical forecasting

The first hypothesis, H_1, compares the forecast accuracy of judgmental forecasting and statistical forecasting. The expert-based individual forecasting study investigated the prediction performance of experts under a no-consensus condition. As defined in Fig. 5, judgmental forecasts outperformed statistical benchmark methods in most prediction horizons. Experts provided higher efficiency in the two-week and six month horizons an error rate of less than 15%. The results reported in Table 1 show that statistical benchmark methods could only improve performance in the six month horizon, but APE values of judgmental forecasts were also close to statistical equivalents.

<Figure 5> The APE results of expert-based prediction of the BDI index.



The second part of this research examined the prediction accuracy for the group consensus task for the BDI forecasting objective. A Delphi panel group was gathered by the moderator under a no-interaction approach. Some of the members of the Delphi-aided study also participated in the first part of research (expert-based forecasting).

Every member of the group provided the initial (the first round) and revised (the second round) predictions with four days time lag. After the first round was completed, a summary report was returned to participants and revised predictions were required if a correction was needed. Iterations were terminated after the second round and a slight improvement of variance reduction was seen.

Results of the Delphi group forecasts are prospective. Fig. 6 and Table 2 show that in two different time scales, the prediction accuracy of the Delphi group is higher than that of statistical benchmark methods.

<Table 1> Judgmental and statistical forecast errors of the expert opinion-based study

	Experts-Mean[a]	Experts-FNE[b]	Experts-C. L.[c]	X12 ARIMA	Holt-Winters	Tramo/ Seats
APE						
Two weeks ahead	7.04	8.01	6.99	19.98	23.62	18.04
One month ahead	45.44	45.02	46.22	63.16	72.40	60.54
Three months ahead	25.52	20.04	26.42	45.20	54.46	42.85
Six months ahead	10.77	13.35	11.56	9.81	20.35	8.01
MAPE	22.19	21.61	22.80	34.54	42.71	32.36
RMSE	1887	1815	1936	2899	3491	2743

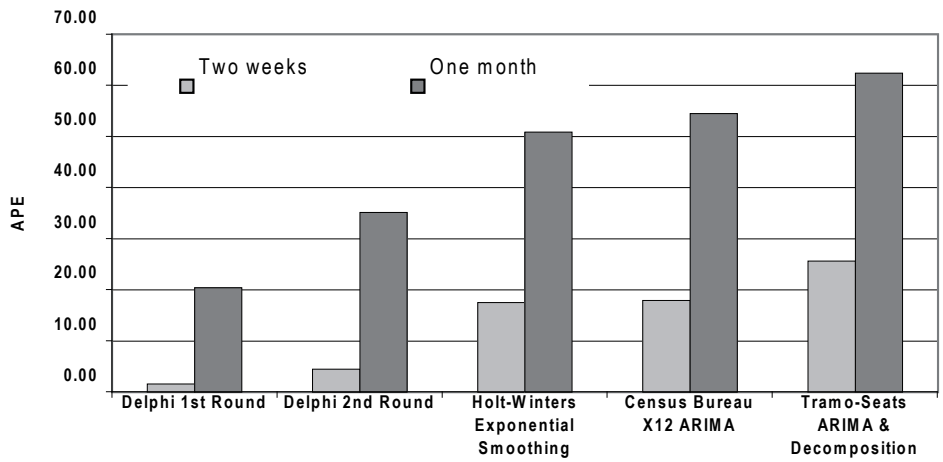
a Experts-mean indicates the simple average of individual forecasts.
b Experts-FNE indicates the freight negotiation experience weighted results.
c Experts-C. L. indicates the results after the weighting of self-confidence levels.

<Table 2> Judgmental and statistical forecast errors of the Delphi group-based study

	Experts-2nd round[a]	Experts-1st round[b]	X12 ARIMA	Holt-Winters	Tramo/ Seats
APE					
25th July 2008	4.49	1.59	17.95	17.52	25.64
30th August 2008	35.14	20.40	54.49	50.85	62.38
MAPE	19.81	10.99	36.22	34.18	44.01
RMSE	1695	1020	2843	2672	3387

a Experts-2nd round indicates the results of the 2nd round in the Delphi based study.
b Experts-1st round indicates the results of the 1st round in the Delphi based study.

<Figure 6> The APE results of the Delphi aided group consensus prediction of the BDI index



The evidence indicated that judgmental forecasting improved the forecast accuracy and, thus, H_1 is accepted.

3) H_{2a}: FNE weighting vs. simple average of judgmental forecasts and, H_{2b}: Self-confidence weighting vs. simple average of judgmental forecasts

With the hypothesises, H_{2a} and H_{2b}, we wish to compare forecast performance between freight negotiation experience (FNE) weighting, self-confidence weighting and simple average results. Table 1 and table 2 show that the results of the FNE weighted average and self-confidence weighted average provided the lowest error rate in some cases. However, a significant existence is not found. In some prediction horizons, simple average strategy also improved over the FNE weighted forecast and self-confidence weighting method. Results of Wilcoxon/Mann-Whitney test indicate *U*:-0.1443 and *P*: 0.8852 for comparison of FNE and simple average conditions; *U*:-0.1443 and *P*: 0.8852 for comparison of self-confidence and simple average conditions.[32] Although, medians of conditions are different, ranks of results are same.

Therefore, non-parametric test results are also same since the Wilcoxon/Mann-Whitney test is based on rank correlations. These evidences lead us to reject H_{2a} and H_{2b}.

4) H_{2c}: Delphi forecasting vs. expert opinion forecasting

The hypothesis, H_{2c}, compares group consensus forecasting (Delphi) and individual single forecasting (expert opinion). An important caveat was noted regarding whether consensus was provided. We expected to improve accuracy using a multi-iteration method. The Delphi group outcome indicated an accuracy loss under a multi-iteration condition. In fact the first round of the Delphi method was the exact same procedure as in the expert-based methodology. Multi-iteration did not provide significant improvement over the single iteration method. Non-parametric test of equality indicates that there is no significant difference between methods (*U*: 0.2314 and *P*: 0.8170). This evidence leads us to reject H_{2c}.

32 The sample size is not enough for a consistent student-t test. Non-parametric test of equality is preferred as it is suggested by conventional basis.

2. Statistical forecasting results

Statistical results provided higher APE error rates in both studies. In the expert-based study, statistical extrapolation ensured better accuracy for the six month horizon, but the remaining forecasts were not more accurate than judgmental predictions. The Census Bureau's X12 ARIMA and the TRAMO/SEATS fully automatic forecasting algorithms were applied BDI time-series data to define the best-fit ARIMA configuration. The estimated model was a Seasonal ARIMA (0 1 1) (0 1 1) configuration in both the X12 ARIMA and TRAMO/SEATS algorithms.

The Holt-Winters method provided relatively higher accuracy than the X12 ARIMA and TRAMO/SEATS approaches in Delphi-based practice, but performance over the judgmental equivalent was weak. The difference between the errors of Holt-Winters exponential smoothing and judgmental methods was about 10-30% (APE). A downward tendency was found for the June-August term. The smoothing constant,b for seasonality was estimated to be 0.16.

VI. Conclusion

From our theoretical perspective, judgments gain importance based on unexpected movements that are caused by sporadic events, political attractions, industry news, and others. Another aspect of judgment is marketplace behavior and psychology. For instance, a situation may simply expose crowd psychology and behavior. Quantitative forecasting may never capture these types of purely judgmental concerns and response timing under quantitative methods often takes longer to implement. This judgmental forecasting approach attempts to fill this research gap.

The main objective of this study is to understand efficiency, applicability, and accuracy of judgmental prediction compared to conventional statistical methods.

Towards this purpose, two popular judgmental forecasting strategies, expert-opinion and Delphi, were tested with a voluntary expert group. All participants were selected from the shipping business and freight negotiation experience was required in order to provide validity and consistency to the research. Most participants had recently worked in the ship-broking field and had practiced in chartering. Some were administrators.

Results of expert-based and Delphi-based studies indicated that these methods represented a significant improvement over statistical methods. Accuracy was mostly accomplished. Practical application of judgmental methods was found to be relatively simple, straightforward, and reasonably time-efficient (less than one week). Notwithstanding the short computing duration required for most statistical methods, data availability is one major drawback of the statistical approach. Statistical extrapolation methods assume a continuous historical pattern, but historical patterns are subject to changes in trade condition, technology, and other factors. Judgmental forecasts do not require historical data in most cases, or, at the very least, only a limited number of past data is required for judgmental extrapolation. Furthermore, a judgmental method provides a suitable environment for combining historical and recent information in forecasting.

Thus, it ensures the latest perspective of predictors.

Future research could be extended to find an accurate way of combining statistical and judgmental results. Recently, the forecasting literature has included many different

solutions for this purpose. However, the shipping business has specific conditions that should be tested for the most suitable composite structure.*

Acknowledgements

The authors thank to two anonymous referees and participants at the IAME conference in Dalian, April 2008, and the 2nd International Conference of Asian Journal of Shipping and Logistics for their many helpful comments. All remaining errors are

* Date of Contribution: Aug. 10, 2009
Date of Acceptance: Nov. 30, 2009

the responsibility of the authors. The authors also gratefully acknowledge Tramp Data Services, NYK Line Research Group, V.Ships Ltd., Gedenlines AS, Pacmar Shipping PTE Ltd. and Armabulk Shipping Ltd. for providing input data. This research has been partly supported by the Ministry of Education, Culture, Sports, Science & Technology (MEXT) of the Japanese Government.

References

ARIEL, A. (1989), "Delphi forecast of the dry bulk shipping industry in the year 2000", *Maritime Policy and Management*, Vol.16, No.4, pp.305-336.

BARSKY, R.B. and MIRON, J.A. (1989), "The seasonal cycle and the business cycle", *Journal of Political Economy*, Vol. 97, pp.503-534.

BATCHELOR, R. and ALIZADEH, A. and VISVIKIS, I. (2007), "Forecasting spot and forward prices in the international freight market", *International Journal of Forecasting,* Vol.23, pp. 101-114.

BEENSTOCK, M. and VERGOTTIS, A. (1993), *Econometric Modeling of World Shipping*, London: Chapman and Hall.

BEAULIEU, J.J. and MIRON, J.A. (1993), "Seasonal unit roots in aggregate US data", *Journal of Econometrics*, Vol. 55, pp.305-328.

BOX, G.E.P. and JENKINS, G.M. (1970), *Time-series Analysis: Forecasting and Control*, San Francisco: Holden-Day.

BUNN, D. and WRIGHT, G. (1991), "Interaction of judgmental and statistical forecasting methods: issues and analysis", *Management Science*, Vol. 37, No.5, pp.501-518.

CHAREMZA, W. and GRONICKI, M. (1981), "An econometric model of world shipping and shipbuilding", *Maritime Policy and Management*, Vol. 8, pp.21-30.

CLEMEN, R.T. (1989), "Combining forecasts: a review and annotated bibliography", *International Journal of Forecasting*, Vol. 5, pp.559-583.

CULLINANE, K.P.B. (1992), "A short-term adaptive forecasting model for Biffex speculation: A Box—Jenkins approach", *Maritime Policy and Management*, Vol.19, pp.91−114.

DURU, O. and YOSHIDA, S. (2008a), "Composite forecast: a new approach for forecasting shipping markets", in: Proceedings for the International Association of Maritime Economists Conference, Dalian, China, April 2-4.

DURU, O. and YOSHIDA, S. (2008b), "*Market Psychology*", Lloyd's Shipping Economist Magazine Vol. 30, August 2008.

FRANSES, P.H. (1990), "Testing for seasonal unit roots in monthly data", *Econometric Institute Report 9032*, Rotterdam: Erasmus University Rotterdam.

FRANSES, P.H. and HOBIJN, B. (1997), "Critical values for unit root tests in seasonal time series", *Journal of Applied Statistics*, Vol. 24, pp.25-47.

GLEN, D.R. (1997), "The market for second-hand ships: further results on efficiency using

cointegration analysis", *Maritime Policy and Management*, Vol. 24, pp.245-260.

GOMEZ, V. and MARAVALL, A. (1996), "Programs seats and tramo: Instructions for the user", *Working paper no*. 9628, Bank of Spain.

GOODWIN, P. and WRIGHT G. (1993), "Improving judgmental time-series forecasting: a review of the guidance provided by research", *International Journal of Forecasting*, Vol. 9, pp. 147-161.

GOODWIN, P. and FILDES R. (1999), "Judgmental forecasts of time-series affected by special events: does providing a statistical forecast improve accuracy?", *Journal of Behavioral Decision-making*, Vol. 12, pp.37-53.

GOODWIN, P. (2005), "How to integrate management judgment with statistical forecasts", *Foresight: the International Journal of Applied Forecasting*, Vol. 1, pp.8-12.

HALE, C. and VANAGS, A. (1992), "The market for second-hand ships: some results on efficiency using cointegration", *Maritime Policy and Management*, Vol. 19, pp.31-39.

HAMPTON, M.J. (1991), *Long and short shipping cycles*, Cambridge: Cambridge Academy of Transport (3rd edition).

HAWDON, D. (1978), "*Tanker Freight rates in the short and long run*", Applied Economics, Vol. 10, pp.203-217.

HOLT, C. C. (1957), *Forecasting seasonal and trends by exponentially weighted moving averages*, Pittsburg: Carnegie Institute of Technology.

HYLLBERG, S., ENGLE, R.F., GRANGER, C.W.J. and YOO, B.S. (1990), "Seasonal integration and cointegration", *Journal of Econometrics*, Vol. 44, pp.215-238.

KAVUSSANOS, M.G. (1996), "Comparison of volatility in the dry-cargo ship sector", *Journal of Transport Economics and Policy*, Vol. 29, pp.67-82.

KAVUSSANOS, M.G. (1997), "The dynamics of time-varying volatilities in different size second-hand ship prices of the dry-cargo sector", *Applied Economics*, Vol. 29, pp.433-444.

KAVUSSANOS, M.G. and ALIZADEH, M.A.H. (2001), "Seasonality patterns in dry bulk shipping spot and time charter freight rates", *Transportation Research Journal*, Part E, Vol. 37, PP.443-467.

LADIRAY, D. and QUENNEVILLE, B. (2001), *Seasonal adjustment with the X-11 method*. New York: Springer.

LADIRAY, D. and QUENNEVILLE, B. (2004), "Implementation issues on shrinkage estimators for seasonal factors within the X-11 seasonal adjustment method", *International Journal of Forecasting*, Vol. 20, pp.557– 560.

LAWRENCE, M.J. and O'CONNOR, M.J. (1992), "Exploring judgmental forecasting", *International Journal of Forecasting*, Vol. 8, pp.15-26.

MAKRIDAKIS, S., WHEELWRIGHT, S.C. and HYNDMAN, R.J. (1998), *Forecasting Methods and Applications*, New York: John Wiley and Sons.

MCCONVILLE, J. (1999), *Economics of maritime transport: theory and practice*, London: Witherby.

NORMAN, V.D. (1979), *The economics of bulk shipping*, Bergen: Institute for Shipping Research.

ONKAL, D. and MURADOGLU, G. (1996), "Effects of task format on probabilistic forecasting of stock prices", *International Journal of Forecasting*, Vol. 12, pp.9-24.

ROWE, G. and WRIGHT, G. (1999), "The Delphi technique as a forecasting tool: issues and analysis", *International Journal of Forecasting*, Vol. 15, pp.353–375.

SANDERS, N.R. (1992), "Accuracy of judgmental forecasts: a comparison", *Omega International Journal of Management Science*, Vol. 20, No.3, pp.353-364.

SHIMOJO, T. (1979), *Economic Analysis of Shipping Freights*, Kobe: Kobe University Press.

STOPFORD, M. (1997), *Maritime Economics*, London: Routledge (2nd edition).

TINBERGEN, J. (1959), "Tonnage and freight", in L.H. Klaassen et al., ed., *Jan Tinbergen Selected Papers*, Amsterdam: North-Holland Publishing Company.

Tramp Data Services Co. Ltd., Baltic Dry Index monthly data between January 2001 and October 2008.

VEENSTRA, A.W. and FRANSES, P.H. (1997), "A co-integration approach to forecasting freight rates in the dry bulk shipping sector", *Transportation Research Part A*, Vol. 31, No. 6, pp. 447-458.

VEENSTRA, A. W. (1999), *Quantitative analysis of shipping markets*, T99/3, TRAIL Thesis Series, Netherlands: Delft University Press.

WINTERS, P.R. (1960), "Forecasting sales by exponentially weighted moving averages", *Management Science*, Vol. 6, pp.324-342.