FORECASTING THE BALTIC FREIGHT INDEX: BOX-JENKINS REVISITED

KEVIN P B CULLINANE
Maritime Studies
Hong Kong Polytechnic University
Hung Hom, Kowloon, Hong Kong

KEITH J MASON
College of Aeronautics
Cranfield University
Cranfield, Beds, MK43 0AL, UK

MATTHEW CAPE
University of Plymouth
Drake Circus
Plymouth PL4 8AA, UK

Abstract

In this paper, the Box-Jenkins approach is adopted to produce forecasting models for the Baltic Freight Index (BFI). Accurate short-term forecasts facilitates the development of a model for use as a tool for either speculation or aggressive hedging strategies in the futures market, the Baltic International Freight Futures Exchange (BIFFEX). Changes in the structure of the BFI enabled the re-appraisal of an earlier forecasting model, and an assessment of whether the fundamental characteristics of the BFI had changed. An AR(2) model was derived that performed well in the short term (two to ten days).

INTRODUCTION

The Baltic Freight Index (BFI) was initiated on 11 June 1984 at a base value of 1000 index points. It was designed to represent the *freight rates* (prices) charged by shipowners for the *chartering* (hire) of their ships. As freight rates are related to the income accruing to the shipping industry, fluctuations in the financial health of that industry can (at least, to some extent) be assessed from movements in the BFI time-series.

The value of the BFI at any given point in time is determined by the freight rates charged on eleven key trade-routes. Each morning, the panel members of the Baltic Exchange in London submit their assessment of freight rates on each of these trade-routes. Where available, these assessments are based on actual freight rates or, alternatively, where there are no actual *fixtures* (specific contracts of carriage known as *charterparties*), then they are based on informed estimates of what the freight rates might be. As a precautionary measure the highest and lowest assessments for each trade-route are excluded from the calculation and an average is taken of those that remain. These average freight rate values are then weighted in accordance with a prescribed system reflecting the relative importance of each trade route to the dry bulk shipping sector and then added together to form the BFI, which is published at one o'clock (London time) each working day.

The trade-routes and their respective weightings are under constant review to ensure that the index remains representative. In addition to many comparatively minor alterations to the composition of the BFI that have been implemented in the past, on 3 November 1993 all trade-routes plied by ships in the size range 25-50,000 Deadweight Tonnes (Dwt) (i.e. all *handysize* routes) were eliminated from the index and the number of component routes in the BFI's composition reduced from thirteen to eleven. The new index then comprised a weighted average of seven panamax routes (50-75,000 Dwt) and four capesize routes (over 75,000 Dwt), Cochran (1993).

This change in the composition of the BFI was implemented mainly in response to pressure from traders on the Baltic International Freight Futures Exchange (BIFFEX) which not only relies on the BFI to underpin and define its standard traded commodity, but was also the original motivation for the initial development of the BFI. In practice, the vast majority of hedgers on this futures market were seeking to hedge positions on panamax and capesize routes. Throughout 1993, however, whilst the capesize and panamax market segments remained depressed, the relative strength of the handysize sector had inflated the BFI. As a result, futures market hedges against adverse freight rates in the panamax and capesize dry bulk sectors were not working adequately as there was only a very low correlation between prices in the physical and futures markets. The hope and expectation of implementing this amendment to the composition of the BFI is that a better service will be provided for traders that hedge on BIFFEX and that, in consequence, the ambition exists that this will attract a greater number of hedgers onto the market. This is of particular importance since a lack of liquidity during 1993 threatened the existence of BIFFEX (Moloney, 1993).

OBJECTIVES

Because of shipping's narrow profit margins, even the slightest fluctuation in freight rates can have major implications for profits. An ability to accurately predict the market has the obvious potential advantage of allowing shipowners to maximise profits and/or minimise losses incurred in the shipping market. Inherent difficulties in obtaining accurate long-term forecasts undermine the usefulness of predictive methods in the physical market. This is because investments in physical shipping market assets are so long-term that forecasts are only useful with extremely long-term time horizons. Accurate short-term forecasts, however, have the capability of being harnessed in the futures market. Such short-term predictions can then be used as part of either a hedging or a speculative strategy.

This analysis will attempt to construct a model for predicting movements in the BFI which is based on data collected since the omission of handysize routes. The underlying assumption is that movements in the BFI are not purely random, but can be predicted and that all relevant information concerning the market's future performance is embodied within historical patterns in the BFI data. Such an approach conforms very closely to that proposed by the advocates of the technical analysis of stock and derivative prices who believe in the Efficient Markets Hypothesis that both the expectations of the market and its perception of market fundamentals have already been discounted and incorporated into current values and prices. Since there is a reasonably high correlation between the BFI and the value of the nearest spot contract traded on BIFFEX (Chang and Chang (1996) suggest that it is as high as 0.89 for one month ahead), forecasts of the BFI can be employed to develop a strategy for speculation.

Although a predictive model of the BFI based on a time series analysis of data covering the period between 1985 and 1988 was conducted by Cullinane (1992), it is hypothesised that the subsequent expulsion of handysize routes from the index has fundamentally altered the underlying nature of the data-set. The fundamental restructuring of the index in 1993 with the exclusion of the handysize trades, may mean that this model is no longer valid and that a new specification or estimation of the model may be required. By analysing the index using data which covers the period 3rd November 1993 to 29th March 1996, a model which is representative of the new time-series can be derived. One of the objectives of this paper is to reconstruct such a model using the same methodology as Cullinane (1992) not only to derive new predictions of the BFI, but also to assess and compare the revised specification of the model against the original, particularly the predictions it produces. The paper shows clearly how the methodology is applied to derive a Box-Jenkins model of the BFI which is then validated by comparing generated ex-ante forecast values against actual BFI data generously supplied by the London Commodity Exchange.

SELECTION OF METHODOLOGY

Because movements in the demand for the BFI's main constituent cargo trades (coal, iron ore and grain) are well documented, a multivariate approach to modelling the BFI is possible. Such an analysis would certainly deepen our understanding of the fundamental mechanism by which movements in the value of the BFI are produced. Given the number of variables considered and consequent parameters included in the resulting model, however, its complexity and size will be such that it would be both difficult and costly to specify and estimate. In addition, since the derivation of forecast values for the input variables of a multivariate model is a problem as intractable as the original, it is difficult even to apply it. In any case, the predictive qualities of multivariate models are rarely found to be better in practice than those of univariate models (Makridakis and Wheelwright, 1978). According to Harvey (1981), in univariable (or univariate) analysis, movements in the variable of interest, and for which a model is desired, are explained solely in terms of its own past and its position with respect to time. This minimises data requirements and greatly simplifies the practical application of the model.

There are any number of univariate methods of analysis but Box-Jenkins is one of the major forms. Applying this same methodology to a more recent data-set than that of Cullinane (1992) facilitates an assessment of the robustness of that model even in the face of a fundamental alteration in the underlying data-set. The Box-Jenkins approach deals primarily with discrete time-series within the time-domain, and is based upon the development of ARMA (Autoregressive Moving Average) models, the foundations of which can be attributed to the work of Yule (1927) and Wold (1938). The general form of an ARMA model is given by the formula:

$$X_t = \sum_{i=1}^{p} \alpha_i . X_{t-i} + \sum_{j=1}^{q} \beta_j . Z_{t-j} + Z_t \qquad (1)$$

Where:

Xt	=	Value of an observation at time t
Xt-i	=	Value of an observation at time t-I
Zt	=	Error at time t
Zt-j	=	Error at time t-j
αi	=	Constant at time t-I
βj	=	Constant at time t-j

p = The order of the autoregressive element
q = The order of the moving average element

The ARMA model is composed of two separate and distinct elements; the AR(p) and the MA(q) submodels. The formula for each of these submodels is shown in equations 2 and 3 respectively. The autoregressive model - AR(p) - is given by:

$$X_t = \sum_{i=1}^{p} \alpha_i . X_{t-i} + Z_t \quad (2)$$

and the moving average model - MA(q) - by:

$$X_t = \sum_{j=1}^{q} \beta_j . Z_{t-j} + Z_t \quad (3)$$

When the parameter q is equal to zero in equation 1, the ARMA model is reduced to the AR(p) model shown in equation 2 and, likewise, when p is equal to zero in equation 1, it reduces to an MA(q) model as specified in equation 3.

EMPIRICAL ANALYSIS

Although ARMA modelling lies at the heart of the Box-Jenkins procedure, its usefulness as a modelling technique in its own right is limited. This is because it relies upon solely *stationary* data. This *stationarity* condition (at least in its weak form) amounts, in practice, to a description of a time-series exhibiting neither trend nor any sort of cyclical or seasonal variation. Very few time-series meet these strict constraints, so that the number of instances when strict ARMA modelling could be applied is severely limited. The Box-Jenkins approach surmounts this problem by utilising differentiation as a means of converting non-stationary data into a stationary data-set. This is achieved by simply differentiating the original base data so that a new and stationary time-series is produced. In fact, if X0, X1, X2, ..., Xn represent observations from the original time-series, at times t = 0, 1, 2, ..., n, then the new time-series, after differentiation, is denoted by:

$$Yt = Xt - Xt\text{-}1, \qquad t = 1, \ldots, n \quad (4)$$

The differentiation process may have to be implemented more than once, but it is reasonably certain to produce a stationary set of data. The transformation of a time-series through differentiation, permits the application of modelling in this general form on a much wider variety of data (Cullinane, 1992). Differentiation does, however, introduce another parameter. The result is the evolution of the ARIMA (p,d,q) technique, or Autoregressive Integrated Moving Average, where the term "integrated" refers to the degree of differentiation (the value of d) required to transform the data into a stationary time-series.

Testing Stationarity

The first step in any Box-Jenkins analysis, therefore, is to assess the stationarity of the data. This assessment can sometimes be achieved by plotting the time-series and, through simple visual recognition, determining whether any trends and/or seasonal variation exists. Plotting time-series also facilitates the identification of outliers (observations which are inconsistent with the rest of the data) and possible turning points (Chatfield, 1984). In essence, this graphical analysis of market indices is akin to basic *chartism*, a technique explained more fully by, for example, Mills (1992) and Stewart (1986) and may provide a range of useful information. At the most basic level, if a given time-series contains either trends or seasonal variation, it is not stationary (for a comprehensive explanation of stationarity see Jenkins and Watts, 1968). A plot of the BFI data considered in this study is shown in Figure 1.

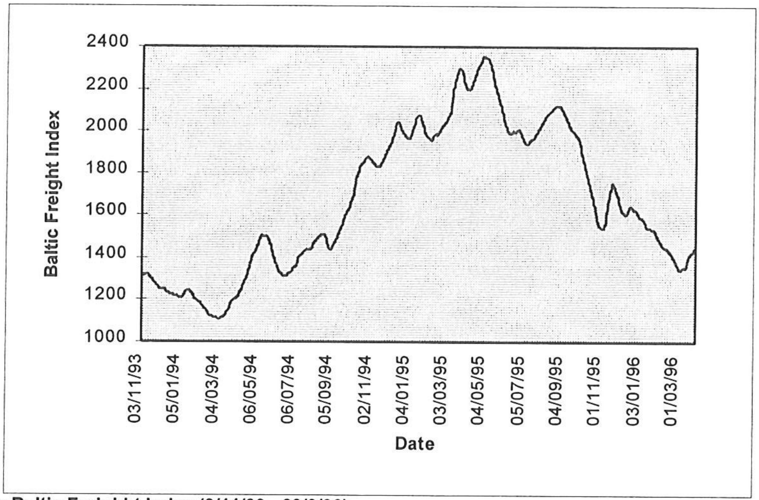

Figure 1: The Baltic Freighht Index (3/11/93 - 29/3/96)

The graph shows the existence of two distinct market trends. The first (a bull trend) lasted for approximately fourteen months following January 1994, during which time the value of the index doubled. In contrast, as characterised by the distinctive "head & shoulders" pattern in the data (see Stewart, 1986), a turning point can be seen at the beginning of May 1995 as the market enters a distinct bearish trend which reverses 75% of the rise in the index value realised under the previous bullish period. Using merely this most basic form of analysis, the market can be shown to be subject to trends and the underlying data can, therefore, be deemed non-stationary.

In many cases, the stationarity of a data-set cannot be ascertained from a simple visual inspection of a plot of the data and further tests are required. One approach is to conduct a simple linear regression of the data against time. If a time-series is stationary, then in the absence of any seasonality, the average trend, as represented by the coefficient of the slope β will be zero or insignificantly different from zero. Running such a regression analysis on the data yields a β coefficient of 0.9785 with an associated t-ratio of 14.023. The coefficient is, therefore, strongly significant, suggesting that a trend in the data is present. Although this is a crude method and cannot be considered definitive proof either way, it does provide some evidence that the data is non-stationary.

A second method of testing for stationarity is through the application of an AR(1) model to the time series whereby:

$$X_t = \alpha . X_{t-1} + Z_t \qquad (5)$$

If the constant $|\alpha|$ is estimated as being significantly different from unity, the time-series can be regarded as stationary. Applying this model to the data, an absolute value of 0.99945 is derived which is well within an arbitrarily assumed cut-off point of 0.9 as the limit of confidence of equality with unity. The fact that $|\alpha|$ is not significantly different from unity again adds weight to the supposition of non-stationarity.

The strongest test of stationarity is based on the autocorrelation function (ACF) of the time-series which at each lag (denoted by the parameter *k*) is compared to the significance statistic of $2/\sqrt{n}$ (equating to a 95% confidence interval where *n* is the number of observations). The sample ACF is denoted by rk where:

$$r_k = \frac{\sum_{t=1}^{n-k}(x_t - \overline{x})(x_{t+k} - \overline{x})}{\sum_{t=1}^{n}(x_t - \overline{x})^2} \qquad k = 1, 2, \ldots \qquad (6)$$

A plot of this function for the original BFI time-series under consideration is shown in Figure 2.

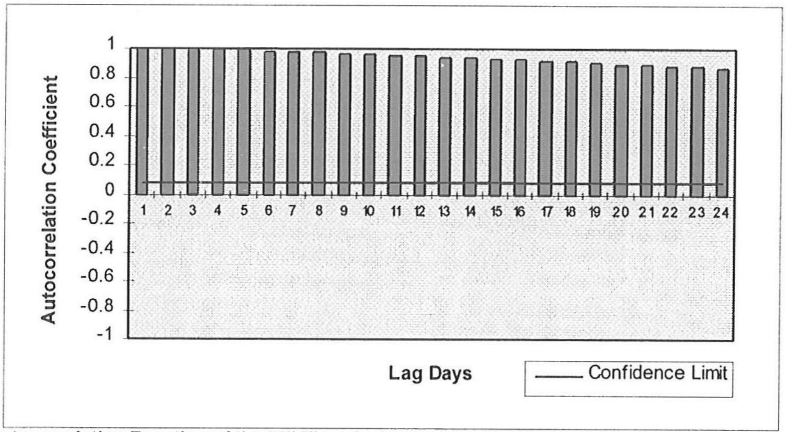

Figure 2: Autocorrelation Function of the BFI Time Series (3/11/93 - 29/3/96)

The graph clearly shows the failure of the ACF at any lag *k* to fall below the established significance level. Therefore, the data is highly autocorrelated irrespective of the lag analysed. This is final and overwhelming evidence that the time-series is non-stationary and, in accordance with the Box-Jenkins methodology, suggests that the data contained in the original time-series requires differentiation as defined by equation (4). Figure 3 shows a plot of the first differential of the time-series. At first inspection, the trend which was perceived to exist in the original data would seem to have been eliminated. To be certain that this new time-series is stationary, however, the preceding tests for stationarity are repeated.

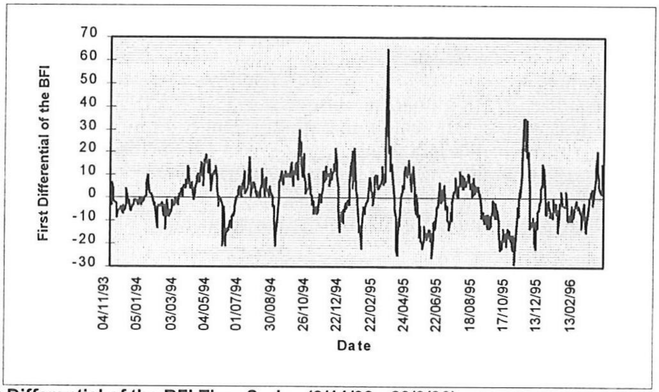

Figure 3: The First Differential of the BFI Time Series (3/11/93 - 29/3/96)

The gradient β which is derived from a simple linear regression of the differentiated data against time is found to be -0.009326 with an associated t-ratio of -3.96. While this test reveals that a considerable amount of linear trend has been removed, the coefficient remains significantly different from zero at the 5% level, thus suggesting that the time-series is still non-stationary. Although the value of the β coefficient for the differentiated data-set is very small, its respective t-ratio does not reflect a statistical proximity to zero. In fact, this is a statistical incongruity due to the fact that linear regression assumes that individual data values are independent of one another. It can be shown that the differentiated BFI data are not independent of one another by considering the Durbin-Watson statistic (Anderson, Sweeney and Williams, 1993).

When applied to the simple linear regression model of the differentiated data, a Durbin-Watson statistic of 0.29 is found, indicating a strong degree of positive serial correlation within the data. This means that although the estimators are unbiased they are not efficient and that estimates of their variance and standard errors will be biased downwards. The ultimate effect of this is that t-ratios will be biased upwards and that the least-squares estimators will appear more significant than they actually are (Haines, 1979). Given this qualification to the results of the simple linear regression test of stationarity, the gradient of the differentiated data can be regarded as not significantly different from zero, thereby providing some support, even though not exactly persuasive, for the hypothesis that the data transformation has led to a stationary time-series.

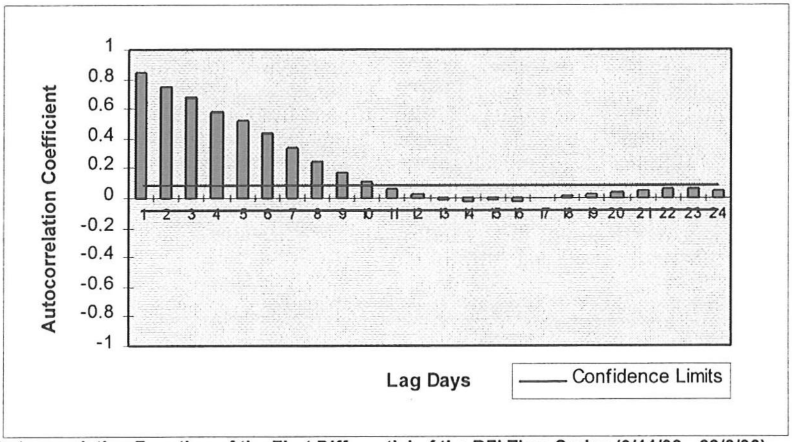

Figure 4: Autocorrelation Function of the First Differential of the BFI Time Series (3/11/93 - 29/3/96)

When an AR(1) model is applied to the differentiated data-set, a value for the constant $|\alpha|$ of 0.8355 is derived. Again using a cut-off of 0.9 for being significantly different from unity, this finding lends further weight to the argument that stationarity in the data has been achieved.

The definitive test of stationarity is provided by the ACF of the differentiated data which is shown in Figure 4. As can clearly be seen, the value of the autocorrelation coefficient dies away fairly quickly as the lag k increases to a point where it falls below the significance statistic of $2/\sqrt{n}$. Although it has been borne out, to a greater or lesser extent, by the previous tests, this result provides the most persuasive evidence that a stationarity data-set has been attained. This transformed data can now be used to take the Box-Jenkins methodology to its natural conclusion.

Model Selection

Having obtained a stationary set of data, the next step in the Box-Jenkins methodology is to calculate and analyse the sample ACF as previously defined in equation (6) and sample partial autocorrelation function (PACF) denoted as r_{kk} and given by:

$$r_{kk} = \frac{r_k - \sum_{j=1}^{k-1} r_{k-1,j}\, r_{k-j}}{1 - \sum_{j=1}^{k-1} r_{k-1,j}\, r_j}, \qquad k = 2,3,\ldots \tag{7}$$

where;

$r_{k,j} \;=\; r_{k-1,j} - r_{kk}\, r_{k-1,k-j} \qquad j=1,\ldots,k-1$

$r_{kk} = r_1 \qquad k = 1$

By comparing the actual properties of the ACF and PACF of the differentiated time-series with the theoretical properties of these functions that are associated with particular generic ARMA model forms and which are shown in Tables 1 and 2, then it is possible to identify a model which fits the characteristics of the time-series.

These theoretical properties of the ACF and PACF provide the basis not upon which a particular model is selected, but rather by which alternative models are rejected. Once all the models which fail to meet the criteria are rejected, the selection of a specific model from a range of surviving feasible options is made on the basis of the principle of parsimony. Cost and complexity increase with the number of parameters required by a model. Hence, the simpler the model the cheaper and easier is its application.

As can be seen from the plot of the sample ACF shown in Figure 4, the autocorrelation coefficients are significantly different from zero up until lag 11. At this point it falls below the 95% confidence limit where it is no longer regarded as significantly different from zero. The gradual fall of the ACF as lag k increases is compared to the theoretical ACF characteristics presented in Table 1. From this comparison, the MA(1) and MA(q) models can be immediately rejected since both models are characterised by a sudden drop in the value of their respective ACFs. The pattern in the ACF also casts doubt on the appropriateness of the AR(1) and ARMA(1,1) models, though they cannot be categorically eliminated. Therefore, four generic model specifications remain under consideration: ARMA(1,1), ARMA(p,q), AR(1) and AR(p).

Table 1: Characteristics of the ACF of Alternative Model Specifications

Model Name	Model Form	Lag k=1	All other Lags (k>|1|)
AR(1)	$X_t = \alpha . X_{t-1} + Z_t$	$p(1) = \alpha$	$p(k) = \alpha^{\|k\|}$
MA(1)	$X_t = \beta . Z_{t-1} + Z_t$	$p(1) = \frac{\beta}{\beta^2 + 1}$	$p(k) = 0$
AR(p)	$X_t = \sum_{i=1}^{p} \alpha_i . X_{t-i} + Z_t$	Indeterminate	Indeterminate
MA(q)	$X_t = \sum_{j=1}^{q} \beta_j Z_{t-j} + Z_t$	$p(1) = \frac{\beta_1 + \sum_{j=1}^{q-1} \beta_j . \beta_{j+1}}{1 + \sum_{j=1}^{q} \beta_j^2}$	$p(k) = 0 \ \forall \|k\| > q$
ARMA(1,1)	$X_t = \alpha . X_{t-1} + \beta . Z_{t-1} + Z_t$	$p(1) = \frac{(1 - \alpha . \beta).(\alpha + \beta)}{1 + \beta^2 + 2.\alpha . \beta}$	$p(k) = \alpha . p.(k-1)$
ARMA(p,q)	$X_t = \sum_{i=1}^{p} \alpha_i . X_{t-i} + \sum_{j=1}^{q} \beta_j . Z_{t-j} + Z_t$	Indeterminate	Indeterminate

i If α>0, then the ACF dies away exponentially. If α<0 the ACF dies away exponentially in magnitude, oscillates in sign.
ii The ACF dies off after lag 1.
iii The ACF is a mixture of damped exponentials and sinusoids which does not cut off, rather dies away slowly.
iv The ACF dies off after lag q and $p(1)$ has a maximum value of $q/(q+1)$.
v Each lagged correlation after lag 1 is reduced as k increases, by a constant factor α.
vi The ACF function does not cut off, but dies away slowly.

Table 2: Characteristics of the PACF of Alternative Model Specifications

Model Name	Model Form	Lag k=1	All other Lags (k>|1|)
AR(1)[i]	$X_t = \alpha . X_{t-1} + Z_t$	$p_{11} = \alpha$	$p_{kk} = 0$, $\forall_k > 1$
MA(1)[ii]	$X_t = \beta . Z_{t-1} + Z_t$	Indeterminate	Indeterminate
AR(p)[iii]	$X_t = \sum_{i=1}^{p} \alpha_i . X_{t-i} + Z_t$	Indeterminate	$p'(k) = 0$, $\forall \|k\| > p$
MA(q)[iv]	$X_t = \sum_{j=1}^{q} \beta_j Z_{t-j} + Z_t$	Indeterminate	Indeterminate
ARMA(1,1)[v]	$X_t = \alpha . X_{t-1} + \beta . Z_{t-1} + Z_t$	Indeterminate	Indeterminate
ARMA(p,q)[vi]	$X_t = \sum_{i=1}^{p} \alpha_i . X_{t-i} + \sum_{j=1}^{q} \beta_j . Z_{t-j} + Z_t$	Indeterminate	Indeterminate

[i] The partial autocorrelation function cuts off after lag 1.
[ii] The PACF of this model dies out slowly in the same way as the ACF of an AR(1) model with α>0.
[iii] The PACF of an AR(p) model will cut off after lag p in the same way as the ACF of an MA(q) model.
[iv] The PACF dies away slowly, and possibly sinusoidally.
[v] and [vi] Difficult to define but does not cut off, rather it dies away slowly and possibly sinusoidally.

The next characteristics to be examined are those of the PACF which can be seen in Figure 5 and which are to be compared to the theoretical properties presented in Table 2. The observed partial autocorrelation coefficients are such that values drop off suddenly rather than dying away gradually as the lag k increases. Since an inherent theoretical property of the PACF of ARMA (p,q) models in general and the ARMA(1,1) model specifically is that they decline gradually as the lag k is increased, the sharp cut-off exhibited by the observed PACF of the sample data provides strong support for dismissing all ARMA(p,q) models (including the ARMA(1,1) model) from consideration. Although there is a sharp fall in the partial autocorrelation coefficient at lag 1, the AR(1) model can also be rejected because the cut-off to a value below the assumed significance level does not occur immediately after lag 1.

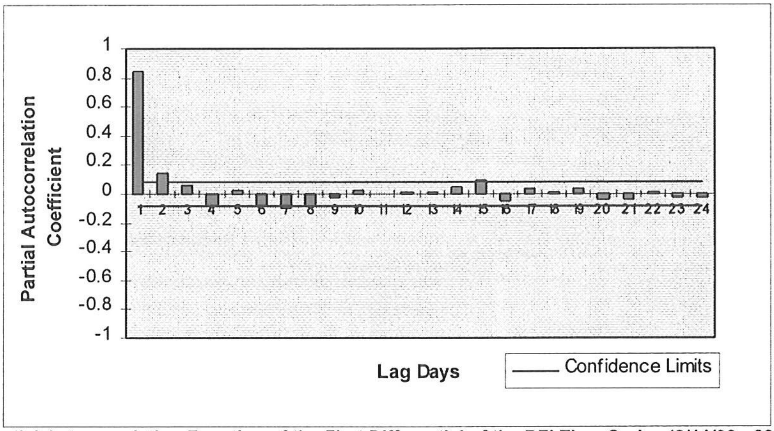

Figure 5: Partial Autocorrelation Function of the First Differential of the BFI Time Series (2/11/93 - 29/3/96)

This leaves only the generic AR(p) model open for consideration. With the PACF cutting-off to below the point of significance at lag 3 and on the basis that a theoretical property of an AR(p) model is that its cut-off point occurs at lag $p+1$, it would appear that the most appropriate specific option is the AR(2) model defined as:

$$Y_t = a_1.Y_{t-1} + a_2.Y_{t-2} + Z \qquad (8)$$

Parameter Estimation

Having selected a model, the next task is to calculate the values of its parameters (α1 and α2) using the ARMA (p,q) modelling element of SPSS. With input values of $p=2$ and $q=0$, the programme produces values for the coefficients α1 and α2 shown in the estimated version of the model given by:

$$Yt = 0.721.Yt\text{-}1 + 0.141.Yt\text{-}2 + Zt \qquad (9)$$

Model Evaluation

An assessment of the model is conducted through an examination of the residuals (or errors) produced by the model; residuals being the differences between the forecast and actual values of the time-series. Box and Jenkins (1976) emphasise how the visual inspection of a plot of the residuals is an indispensable first step in the checking process. By plotting the residuals in histogram form, the examination of their distribution is facilitated. For the model to be deemed acceptable, particularly in terms of its specification, the residuals produced by the model need to be independent and normally distributed about a mean of zero. This histogram of residuals for the model revealed that, by visual inspection, the residuals seem to conform to this desired pattern.

A more mathematical examination of residual structure is provided by the application of the Box-Pierce statistic. This is calculated as follows:

$$Q(m) = n(n+2)\sum_{k=1}^{m}(n-k)^{-1}\left[r_k(\varepsilon)\right]^2 \qquad (10)$$

where;
rk(ε) is the sample autocorrelation of the residuals at lag k
m is an arbitrary value, usually 12, 24, 36, etc.

Although this test is more associated with graphical procedures, if the data-set is large enough, a modified Box-Pierce statistic (or Box-Ljung) statistic provides the basis of a valid test. The statistic constructed from the ARMA model may be considered appropriate if it is taken to have a Chi-square distribution (χ^2_{k-p-q}) under the null hypothesis. By taking the number of fitted parameters as the degrees of freedom, limits of confidence can be taken from the statistic and the significance of the residuals assessed (Harvey, 1981).

These tests are to ensure that the 'fitted' model is adequate. Should the chosen model fail in some way, the results of the test will indicate how the model should be modified, and the cycle of identification, estimation, and verification is started again (Anderson, 1976). Thus, it can be seen that Box-Jenkins analysis is not simply a modelling technique, but rather a modelling strategy.

Table 3: Standard Deviation and Mean of the Residuals Produced from a Range of Models

Model	Standard Deviation (σ)	Mean
AR(1)	5.59	0
AR(2)	5.53	0
AR(3)	5.52	0
AR(4)	5.50	0
AR(5)	5.50	0
ARMA(1,1)	5.53	0

At this stage, other AR(p) models were examined for comparison. As one would expect, forecasting ability improved with the number of parameters, although the scale of the improvement was minimal. The standard deviation and mean values of the various models are shown in Table 3. From an examination of these figures, the most accurate models are the AR(4) and AR(5). For a gain of only 0.03 in standard deviation over that produced by the AR(2) model, the cost is the time required to calculate twice as many parameters. These higher specification models are, therefore, rejected on the basis of parsimony. It is interesting to note that even though the ARMA(1,1) model was rejected on the basis of a comparison of theoretical and sample characteristics of the ACF and PACF, this model produces residuals with a similar distribution to the AR(2) model. Although it may be as accurate as the AR(2) model, its dismissal can still be justified on the grounds that despite possessing only two parameters, each relates to different formulaic elements within what is a more complex model form than the AR(2) model. The model itself being more complex, the two parameters are consequently more expensive to obtain.

The Box-Ljung Statistics at lag 12 and lag 24 for the AR(2) model were found to be 18.5 (with 10 degrees of freedom) and 36.5 (with 22 degrees of freedom) respectively. From an examination of Chi-square tables, these statistics were found to be marginally outside the 95% limit of confidence (Miller and Powell, 1979). The first model form that satisfies this test is the AR(4). In the case of the AR(2) model, it could be argued that the size of the data-set is too small for the test statistic. As has already been stated, the Box-Pierce statistic is more commonly associated with graphical procedures and this modified form requires a large data-set to provide a valid result. It is possible that the data-set which underpins this analysis is insufficient in size for this test to be used. The AR(4) may fit within Box-Pierce's limits of confidence, but as a more specified model it should only be substituted if it is thought absolutely necessary.

Again, the final tests of the appropriateness of the derived model involves the application of the autocorrelation and partial autocorrelation functions. The functions are applied to the residuals from the AR(2) model. The plots of the two sample functions showed that less than 5% of the lagged auto- and partial auto-correlation coefficients are significant and, therefore, it can be assumed that there is no remaining structure in the residuals produced by the model.

Since the derived AR(2) model is based on a differentiated data-set, before it can be used to forecast actual movements in the variable of primary interest (the BFI itself), it needs to be integrated so that forecasts relate to true values of the BFI rather than to daily changes in its value. This transformation is derived from the specification of an AR(2) model given in equation (8) and by the fact that the differentiated data-set was derived initially by applying the transformation defined in equation (4). The AR(2) model specification thereby inverts to give:

$$X_t - X_{t-1} = \alpha_1.(X_{t-1} - X_{t-2}) + \alpha_2.(X_{t-2} - X_{t-3}) + Z_t \qquad (11)$$

while the estimated version is given by: $X_t = 1.721.X_{t-1} - 0.581.X_{t-2} + 0.141.X_{t-3} + Z_t$

Model Comparisons

As shown in equation (12), the estimated model derived by Cullinane (1992) using the same methodology is in one sense markedly different from the model derived within this analysis. The model given in equation (12) is a customised AR(3) model, which includes the additional term X_{t-4} in its specification. At the same time, however, the estimated values of the coefficients associated with the other three independent variables appear to be very close in value to those determined in this modelling exercise where an AR(2) model resulted.

$$X_t = 1.556X_{t-1} - 0.556X_{t-2} + 0.189(X_{t-3} - X_{t-4}) + Z_t \qquad (12)$$

The validity of both models can only really be assessed by comparing the ex-ante forecasts produced by each of the models with actual values of the BFI. There are then a number of methods which can be employed to measure the accuracy of the two models. Those used within this study include the Mean Sum of Squares of the residual (MSS), Mean Average Deviation (MAD), Maximum Average Deviation (MAXAD) and a modified Theil's inequality coefficient. The last measure provides the clearest indication as to whether any systematic error exists within the forecast values (Theil, 1966) and is given by:

$$u = \frac{(1/n)\left(\sum (P_i - A_i)^2\right)}{\left\{(1/n)\sum P_i^2\right\}^{1/2} + \left\{(1/n)\sum A_i^2\right\}^{1/2}} \qquad (13)$$

where;

A_i are the actual values of the series being modelled

P_i are the predicted values of the series

n is the number of values over which actual and forecast values can be compared

So that a full measure of each model's accuracy can be assessed, tests were carried out over a range of forecast lead-times. The forecast lead times chosen for testing were 1 day, 5 days, 10 days, 15 days, and 20 days and the results of this part of the analysis are presented in Table 4.

Table 4: Comparative Error Measurements for the Models

Lead Time	Model	MSS	MAD	MAXAD	Theil's Coefficient
1 Day	1998 AR(2) Model	8.7	2.337	5.998	0.003
	1992 AR(3) Model	7.9	2.274	5.688	0.003
5 Days	1998 AR(2) Model	233.7	13.727	24.399	0.080
	1992 AR(3) Model	210.2	11.816	29.946	0.072
10 Days	1998 AR(2) Model	1410.4	30.827	66.914	0.482
	1992 AR(3) Model	562.8	18.214	68.805	0.194
15 Days	1998 AR(2) Model	7992.4	72.138	163.891	2.772
	1992 AR(3) Model	6505.0	66.116	152.363	2.264
20 Days	1998 AR(2) Model	19681	118.579	208.965	7.035
	1992 AR(3) Model	16251	108.695	179.251	5.796

The results of the MSS and MAD measures are similar between the two models and, as fully expected, increase monotonically with lead-times as the confidence interval which encompass forecasts also increase. Over all five lead-times, the Cullinane (1992) model proves to be the more accurate. This is hardly surprising, however, given the higher specification of this model.

Since it is closely associated with assessing maximum potential losses, the Maximum Average Deviation, MAXAD, is the most significant measure of error from the perspective of either hedging or risk averse speculation. Error must be expected when forecasting, but when making decisions on the basis of these forecasts, the absolute size of these errors is very significant, especially when considering the worst possible case. Examination of the MAXAD tells a slightly different story to the measures of error previously analysed. Over a lead-time of one day, the Cullinane (1992) model has the smallest MAXAD, but the spread between the MAXAD of both models is only marginal at 0.3 BFI points. With lead-times of 5 and 10 days, however, the Cullinane (1992) model produces the greater MAXAD. As lead-times increase to 15 and 20 days, the greater specification of the Cullinane (1992) model begins to show through. This measure seems to suggest that over the short-term, that is to say up to ten days, the AR(2) model developed in this analysis provides the most consistent forecasts with the lowest MAXAD.

Theil's coefficient gives the clearest picture of the existence of systematic error within the forecasts. With a lead-time of one day, it suggests that the degree of systematic forecast error is absolutely negligible for both models. Again, as one would expect, as the lead-time is increased, the forecasts produced become less and less accurate. It is apparent, however, that the extent of systematic error present in the forecasts produced by the Cullinane (1992) model is much less than that present in the current lower specification AR(2) model, with the divergence between the two models on the basis of this measure increasing with lead times.

CONCLUSIONS

This work hypothesised that the underlying statistical characteristics of the BFI may have fundamentally changed since the removal of handysize routes from its composition on 3rd November 1993. If these characteristics had altered, then past models of the BFI (for instance, that by Cullinane, 1992) would have lost some of their validity. The objective of the paper was, therefore, to determine the veracity of this argument.

By reapplying the same methodology used by Cullinane (1992) but to a more recent data-set, a comparison of the developed univariate time-series models would reveal whether any major differences exist. The two statistical signatures most influential to the derivation of the ARIMA models to be compared in this way are the autocorrelation and partial autocorrelation functions. Through a comparison of these functions for data-sets pertaining to before and after 3rd November 1993 (when a significant alteration to the composition of the BFI took place and when a fundamental change in the behaviour of the BFI is hypothesised), it is possible to demonstrate the extent to which the characteristics of the index have altered. Any changes in these characteristics will then manifest themselves in either the specification or estimation of the emergent ARIMA model.

The analysis by Cullinane (1992) was based on five data-sets covering the period between 1985 and 1988 (one for each of the four individual years and one for the whole period). With only the slight exception of 1986, when the market was turning and in a state of flux, the ACF and PACF for each of the five data-sets exhibit a great deal of consistency. The ACF for each data-set declines gradually towards the established 5% significance level while the PACF of each falls sharply below the 5% significance level immediately following lag 3. On the basis of the consistency of results achieved across all five data-sets, it could be assumed that these structural characteristics are inherent long-term features of the index which only change with a somewhat drastic alteration to the fundamental nature of the index.

The examination of the autocorrelation and partial autocorrelation functions undertaken within this updated analysis reveals only a very small deviation from these findings. The form of the ACF output is the same as that of Cullinane (1992), but the characteristics of the PACF differ slightly with it cutting-off to below the 5% significance level after lag 2 as opposed to lag 3. This can hardly be termed a revolutionary difference since it implies that the optimal Box-Jenkins model which balances predictive ability with parsimony is an AR(2) rather than a modified version of an AR(3) as found in Cullinane (1992). The mathematical content of both models is, in fact, remarkably similar, especially given the different time periods covered by the two analyses and the number of minor and major changes that have occurred to the composition of the BFI in the interim period.

With the exception of the MAXAD beyond a five-day lead time, on the basis of most measures, the Cullinane (1992) model is consistently more accurate than the AR(2) model developed within this analysis which relies upon a database of later observations of the time series. Given the fundamental change in the composition of the BFI which has taken place through the expulsion of handysize routes, this is an interesting conclusion and may appear slightly incongruous given the optimal model derived herein. The analysis contained herein, however, merely demonstrates that the greater accuracy gained through the estimation of a more highly specified model does not compensate for the associated loss in parsimony.

One can conclude, therefore, that even though the composition of the BFI is under continuous review, with neither the composite routes nor the weightings given to them remaining constant, these changes do not appear to significantly affect the fundamental characteristics of the index.

As was already stated in the introduction, ARIMA models are put to best use when producing short-term forecasts. Given this is the case, ARIMA models must be responsive to short-term market movements. As any highly specified model is handicapped by the number of its parameters and less specified models limited by lesser accuracy, there is an optimum level of specification. The Box-Jenkins approach provides an effective method of estimating that level. It ensures that the selected ARIMA model is sufficiently specified that the forecasts it produces replicate reasonably well the characteristics of the market but, at the same time, is not over specified and, thus, hindered by the number of its parameters.

The use of univariate time-series analysis provides a cost-effective and efficient technique for developing forecasts as the basis of a strategy for speculating on BIFFEX. A more comprehensive speculative strategy may be achieved by integrating this with other methods of technical analysis and further research needs to be conducted into this area. Individually, such techniques tell part of a story; combined they are likely to yield much more accurate short-term forecasts. Over longer-term forecast horizons, however, such as that necessary for hedging, the only truly effective predictive method is the application of fundamental analysis where the determination of market predictions is based on expected levels of future supply and demand in the market. The difficulty of undertaking such an analysis in an industry as volatile as shipping is well known but, if achievable, would not only enable a greater understanding of causality rather than mere mathematical relationship, it would also provide a link between fundamental and technical analysis, which are too often regarded as competing, rather than complementary, methods of analysis.

REFERENCES

Anderson, O.D. (1976) **Time Series Analysis and Forecasting**, Butterworths, London.

Anderson, D.R.; Sweeney, D.J. and Williams, T.A. (1993) **Statistics for Business and Economics**, fifth edition, West Publishing, New York.

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

Chang, Y-T. and Chang, H.B. (1996) Predictability of the dry bulk shipping market by BIFFEX, **Maritime Policy & Management, 23(2),** 103-114

Chatfield, C. (1984) **The Analysis of Time Series: An Introduction**, third edition, Chapman and Hall, New York.

Cochran, I. (1993) Two moves to lift interest in BIFFEX, **Lloyd's List**, 2 November, Lloyd's of London Press, Colchester.

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

Haines, B. (1979) **Introduction to Quantitative Economics**, George Allen & Unwin, London.

Harvey, A. C. (1981) **Time Series Models**, Philip Allan, Deddington, Oxford.

Jenkins, G. M. and Watts, D. G. (1968) **Spectral Analysis and its Applications**, Holden-Day, San Francisco.

Makridakis, S. and Wheelwright, S.C. (1978) **Forecasting Methods and Applications**, John Wiley & Sons, New York.

Miller and Powell (1979) **The Cambridge Elementary Mathematical Tables**, Cambridge University Press, Cambridge.

Mills, T.C. (1992) **Predicting the Unpredictable? Science and Guesswork in Financial Market Forecasting**, Institute of Economic Affairs, Occasional Paper No. 87.

Moloney, S. (1993) Falling volume threat to future of BIFFEX, **Lloyd's List**, 19 May, Lloyd's of London Press, Colchester.

Stewart (1986) **How Charts Can Make You Money: Technical Analysis for Investors**, Woodhead-Faulkner.

Theil, H. (1966) **Applied Economic Forecasting**, Rand McNally, Chicago.

Wold, H. O. (1938) **A Study in the Analysis of Stationary Time Series**, Almquist & Wiksell, Uppsala.

Yule, G. U. (1927) Why do we sometimes get nonsense correlations between time-series? -A study in sampling and the nature of time series, **Journal of the Royal Statistical Society, 89, 1-64**