Economic significance of market timing rules in the Forward Freight Agreement markets

Nikos K Nomikos[*] and Kaizad Doctor [+]

Abstract

Quantitative market timing strategies have been traditionally tested in liquid commodity and financial futures, often with mixed results with respect to their performance. We extend this methodology to a non-storable commodity, freight, where hitherto this analysis has not been carried out. The freight futures market is mature and increasingly liquid, making it a good case for diversification and trading opportunities. We carry out a comprehensive study of quantitative trading strategies in the FFA (Forward Freight Agreements) market on a wide variety of contracts and maturities with a number of trading rules. We find that in spite of robustness checks, trading rules do outperform the buy-and-hold benchmark in general. We also explore the possibility that illiquidity and a small sample size may impact the results of the tests and therefore offer an intuitive approach to mitigate their effects. A procedure that augments the Hansen (2005) SPA (Superior Predictive Ability) methodology and allows us to use it for smaller sample sizes with increased confidence is also proposed.

Keywords: Forward Freight Agreements, Market Timing, Trading Strategies, Data Snooping, Multiple Hypothesis Testing, Reality Check, SPA test.

JEL Classification: G12, G14

[*] **Corresponding Author:** Nikos Nomikos, Cass Business School, City University, 106 Bunhill London EC1Y 8TZ, UK, Tel: +44 (0) 2070400104, Fax: +44 (0) 2070408681, e-mail: n.nomikos@city.ac.uk.

[+] Kaizad Doctor, Cass Business School, City University, 106 Bunhill London EC1Y 8TZ, UK; e-mail: kaizad.doctor.1@cass.city.ac.uk

1. Introduction

The shipping industry is an integral part of the global economy as it moves up to 90% of the international trade in a very cost-efficient way; the cost of ocean freight represents, on average, 6% of either the import value or the shelf price of imported consumer goods, and as such affords genuine economies of scale to consumers. (UNCTAD Report, 2010). With large volumes shipped at relatively low costs, the shipping industry is vital to the sustainability and growth of the global economy and therefore increasingly seen as a barometer for its health. For instance, the Baltic Dry Index (BDI), which is an assessment of the cost of moving dry bulk commodities by sea, is considered by the investment community as a leading indicator of economic activity (see Kilian 2009 and Bakshi et al. 2011).

The cost of seaborne transportation is reflected by the level of freight rates which are determined freely in competitive international markets; as such, freight markets exhibit price characteristics similar to those of other commodity markets such as high volatility, volatility clustering, seasonality, cyclicality and dependence on global commodity and financial markets. Operating in the freight market therefore poses significant price risk to participants who often mitigate this risk either through long-term time-charter contracts or, increasingly, with the use of derivatives products, such as Forward Freight Agreements (FFA), which afford them greater operational flexibility. FFA are cash-settled forward contracts that trade the value of expected freight rates, the underlying asset being one of the freight indices published by the Baltic Exchange. Initially, this market was primarily used for risk management by participants that had direct exposure to freight rates, such as shipowners, ship-operators, charterers and commodity traders. The growth in seaborne trade and freight rates from the first years of the new millennium - driven primarily by the increasing demand for raw materials from developing countries - led to an increase in the participation of market players outside of the shipping industry such as investment banks, fund managers and hedge funds. As a result, according to market sources, it is estimated that as of 2011 the value of trading for speculative purposes is more than twice that for hedging, with a 70/30% split. With the majority of participants using FFA contracts for speculative trading, it is interesting to investigate the effectiveness of trading strategies in the FFA markets, in keeping with contemporary studies in the finance literature but also allowing for the unique nature of those contracts both in terms of their characteristics as well as the way they are traded.

Trading based on technical analysis, in particular, is widely utilized by traders and analysts in the FFA markets either to generate trading signals or to support their trading decisions. The main objective for doing so is to identify trend- and momentum-based patterns in prices though it is not uncommon to encounter strategies aiming to identify trend reversals, market cycles and channel break-outs. The economic significance of these trading rules is generally tested out of sample. Nevertheless, there are various impediments to trading FFA's in the manner common to trading more liquid contracts. First, despite the fact that freight rates are determined freely in competitive markets, adjustments to information are not immediately impounded in current prices. Trades are primarily carried out through brokers in the OTC market and are therefore subject to 'friction' in terms of the speed of execution, liquidity and transaction costs, which may significantly affect profitability. Second, one also has to consider the nature of the underlying asset, which in this case is a non-storable service. Generating trading signals in storable commodities is generally carried out using fundamental analysis of the extant supply and demand. It is therefore expected that inefficiencies in the underlying markets, caused by temporary shocks either to the supply or the demand side, are arbitraged away in the futures markets with equilibrium being restored in the longer run. The same argument however, is difficult to apply to the freight market, which is considered a non-storable commodity owing to the nature of the underlying being a service. Though freight rates are determined through the interaction of supply (availability of fleet) and demand (seaborne trade), the absence of a storage relationship means that the link between spot and forward prices may not be as strong as in the case of storable commodities. In addition, the forward market includes non-shipping market participants, such as investment banks and hedge funds, whose motivations for trading may be more complex and driven by other parameters, in addition to market fundamentals. This may make fundamental analysis less effective in detecting market signals, while technical analysis may still be able to uncover underlying market trends.

Therefore, in this study we test the economic significance of quantitative trading strategies in a robust but intuitive manner and in a form that broadly represents general market practices. We consider 11,548 different parameterizations of trading strategies applied to the FFA markets both across vessel sizes as well as the term structure. In order to make this a realistic exercise all rules are applied on a forward looking basis and thus the results reported indicate the performance that a trader would have achieved in the market if he had followed the same approach. The evaluation of active trading strategies is made on the basis of mean outperformance as well as risk-adjusted outperformance measure - Sharpe ratio - over the benchmark position which is a buy-and-hold portfolio of FFAs. The analysis is carried out in the Capesize, Panamax and Supramax FFA routes, across a range of contract maturities and considering all other parameters that may affect the performance of trades such as delays in the execution of the orders (slippages) and transaction costs. The robustness of the outperformance of trading strategies is automatically questioned in technical analysis studies owing to “data snooping” or selection bias brought into the analysis by means of re-using the same data set over and over again, while testing a host of trading rules. In order to quantify the effect of a data snooping bias across sectors and contract maturities, we incorporate White's (2000) reality check (WRC) methodology and the improvements to the same in the form of the Superior Predictive Ability (SPA) test by Hansen (2005).

Previous attempts to apply technical analysis in the freight markets were generally restricted to physical shipping markets such as tanker freight rates (Adland and Strandenes, 2006) or to work out the optimal investment decisions in the sale and purchase of vessels (Alizadeh and Nomikos, 2007). Despite the growth in the number of studies looking at statistically predictable patterns in commodity and currency futures markets, notably by Marshall et al., (2008a, 2008b) for index futures and Miffre and Rallis (2007) for commodity and financial futures, there has been no attempt at investigating the same within the FFA markets. The freight market offers a very interesting area for the application of technical analysis especially in the rally years before the recent financial crisis (from 2005 to the second half of 2008), as freight markets had a positive trend which would have made it difficult to generate any timing signals for active strategies on the basis of modeling fundamentals alone. It is anticipated that this environment would force traders to look more towards technical strategies, particularly in order to generate trading signals over shorter horizons.

By investigating the performance of trading strategies in the FFA markets, we also aim to add to the existing body of literature regarding the distinction between “young” and “mature” or “established” markets. For instance, Hsu and Kuan (2005) find that trading rules are able to generate profits in relatively “young” markets (the Russel 2000 index being used as a proxy) as opposed to Sullivan et al. (1999) (hereafter STW) and Aronson (2011) who find that the same set of rules are unable to significantly outperform the benchmark in “established” markets, such as the S&P 500. They attribute this to younger markets attracting newer investors and arbitrageurs who then go on to exploit the inefficiencies using a host of trading strategies, eventually reducing these opportunities in time. This apparent “self-destruction” of profitable trades when more and more investors begin to use them has also been studied by Timmermann and Granger (2004) and is used to explain the decline of the predictive power of technical rules over time, which is also in keeping with Lo’s (2004) adaptive markets hypothesis (AMH); AMH is especially relevant to our study as it posits a Darwinian “natural selection” process amongst strategies and market participants, in which competition initially flourishes and eventually opportunities become well publicized and yield lower outperformance as markets mature. In most studies, the indicator for outperformance of active strategies is a low WRC or SPA test p-value. In line with the findings for younger markets, we find that certain technical trading rules consistently outperform the benchmark in the FFA markets even after accounting for transaction costs and order delays induced by thin trading.

Therefore, our study contributes to the literature in the following ways: Firstly, this is the first study of its kind that investigates the economic significance of technical trading rules within the FFA market. Our findings have direct implications towards weak-form market efficiency and as a consequence towards asset pricing models, while being practically relevant from a market practitioner’s perspective. Secondly, we contribute to the body of literature that investigates the observed inefficiencies in “younger” markets, which are relatively illiquid, and concur with the view that these inefficiencies may indeed yield economically significant results in the interim period as markets mature and trading opportunities become scarcer. Finally, we also find that the two tests (WRC and SPA) employed for robustness checks are sensitive to the choice of block-length parameter, which is often arbitrarily set at $w = 10$. We propose mitigating the sensitivity of the tests to the arbitrary imposition of a constant mean parameter by using a data driven method, thereby improving the small sample properties of the test.

This paper is divided into the following sections. Section 2 contains a description of the trading strategies employed in the study. Section 3 presents the underlying contract and the statistical properties of the dataset, followed by Section 4 for the methodology for mitigating the effects of data snooping. Empirical results are presented in Section 5. Finally, Section 6 concludes and elaborates on the implications of our findings.

2. Technical analysis and trading strategies

Technical analysis is a collection of rules that help to identify trends or patterns in the market price in order to generate buy or sell signals. The use of these rules has been well documented for over a century. A comprehensive literature review of studies analyzing a multitude of strategies in a host of markets with a summary of their finding can be found in Park and Irwin (2007). Early academic studies of technical analysis, such as Fama and Blume (1966) and Jensen and Benington (1970), conclude that technical analysis is generally of no economic significance. In contrast, contemporary research - starting from Brock et al. (1992), who demonstrate that a relatively simple set of technical trading rules possesses significant forecast power for changes in the Dow Jones Industrial Average (DJIA) over a long sample period - indicates that under certain conditions technical trading analysis can generate significant abnormal returns. As a result, Neely et al. (1997) represents the contemporary (and contrarian) view, where the economic significance of the technical trading strategies challenges the broadly held contention that financial markets are (highly) efficient. In addition to our approach being in line with current market practice, two main theoretical arguments have also been put forward in order to explain the usefulness of technical analysis. The first argument by Brown and Jennings (1989) deals with the importance of practitioners using technical analysis in disseminating information in the markets. The second argument proposed by Treynor and Ferguson (1985) deals with exploiting the transmission of information through prices in competitive markets.

Owing to the lack of a formal survey on the use of technical indicators amongst practitioners involved in trading FFA, justifying the use of any technical trading rules out of the millions available in the academic literature introduces a subjective bias. Also STW imply that the trading rules should be in existence and widely used by market practitioners during the sample period, for the results to be valid and provide conclusive evidence for the viability of trading strategies. As such, we base our analysis on the strategies used in the preliminary investigation carried out by Alizadeh and Nomikos (2009), as we believe these are sufficiently representative of the strategies popularly used for trading the FFA market and are seconded by various brokers' market reports who often present those as useful trading tools. The parameterizations and the execution of those strategies are then augmented based on the two major studies of STW and Hsu and Kuan (2005), the latter being more comprehensive in terms of the number of strategies tested. Overall, this results in a total portfolio of 11,548 strategies, which are applied and tested in the FFA markets.

We divide the strategies into two broad classes, the first consisting of simple trading strategies and the second of complex (learning and voting) rules based on the simple strategies. The simple trading strategies are further divided into four distinct categories: trend, momentum, volatility, and envelopes or channel breakouts. Then, we parameterize individual trading rules based on general market practice as well as availability of data. A description of the technical strategies is discussed below along with their parameterizations.

a) Trend following strategies mainly utilize a combination of moving averages to exploit trends in the underlying series and generate a trading signal. We choose to formulate three strategies within this class. We start with the basic moving average crossover [MAX] followed by the moving average crossover consensus [MAXC] and finally the moving average convergence/divergence [MACD]. The first strategy is the moving average crossover rule [MAX] where the trading signal is generated when there is a crossover between the fast (x-period) and slow (y-period) moving averages, where $x < y$. A long (short) position is initiated when x crosses y from below (above), and the position is closed when the reverse happens. An additional parameter, e, indicates whether the averaging is arithmetic or exponential. The parameterizations are: $x = (3, 5,\ldots,29$ days); $y = (31, 33,\ldots,99$ days) and $e = 0$ or 1 depending on whether the averaging is arithmetic or exponential, respectively. This results in 980 combinations of trading strategies [$x * y * e = 980$]. The second strategy is the moving average consensus [MAXC] that uses three moving averages representing short-run, x, medium-term, y, and long-run trends, z. This strategy generates a buy (sell) signal only when the price is greater (less) than all three moving averages and closes out the position when the price goes below (above) any of the moving averages. The parameterizations are: x = (3, 7,…,19 days); y = (21, 24,…,60 days), z = (61, 64,…,100 days) and e = 0 or 1 depending on the averaging method employed. The total numbers parameterizations for this strategy are 1960 (= $x * y * z * e$). The final strategy in this class is the moving average convergence divergence [MACD] which consists of constructing an ‘oscillator’ by taking the difference between the fast, x, and the slow, y, moving average. A ‘*signal line’* which is a moving average, z, of this oscillator is constructed and used to generate a buy (sell) signal if the oscillator is above (below) the signal line. The parameterizations are: x = (10,14,…,30 days); y = (31, 35,…,99 days), z = (6, 7,…,12 days) and e = 0 or 1. The total number of parameterizations for this strategy is 1512 (=$x * y * z * e$).

b) Momentum strategies are explained in detail by Colby and Meyers (1988) who recommend the use of three popular momentum based oscillators: the Relative Strength Indicator (RSI); the Stochastic-RSI, and the ‘Aroon’. The RSI is constructed in the following manner [1]:

$$RSI_{t+1} = 100 - \left[\frac{100}{1 + \left(\frac{1}{n}\sum_{t=1}^{n} u_t \Big/ \frac{1}{n}\sum_{t=1}^{n} d_t \right)} \right]$$

Where u_t and d_t are upward (upticks) and downward (downticks) daily price movements over the previous n days (referred to as the lookback period), respectively. If the number of upward movements is equal to the number of downward movements, the RSI will get a value of 50 indicating no momentum in price. If the there is an upward momentum in price

[1] Additionally, momentum strategies based on volume are popular inclusions in studies investigating liquid markets such as equities and currencies. However, volume data in the FFA markets is aggregated over sectors on a weekly basis and therefore is not appropriate for use in our study.

the RSI will start increasing. A long signal is initiated when the RSI breaches a pre-specified filter, k, from below and is closed out when the momentum "runs out" i.e. crosses 50. A short signal is initiated in a similar manner when the RSI crosses the filter from above and is closed out when it crosses 50, which is the midpoint. The total combinations of upper filter bands, U = 100 – k, and lower filter bands, L = k, where n = (3,6,…,27 days) and k = (20, 21,…,36), results in 2601 unique parameterizations (2601 = n * U * L).

Stochastic RSI rules and various combinations of those are collectively known as stochastics and indicate the relative position of the current price with respect to maxima/minima contained within a pre-specified period. Similar to the RSI, a look-back period (n_1) is specified, and the minimum (Low) and maximum price (High) of the FFA series within this look-back period is then computed to form the oscillator, K. The resulting oscillator is generally volatile and if used by itself would create "noisy" trading signals, which is why it is smoothed using a n_2-period (simple or exponential) moving average. This smoothed oscillator D, can be used to generate trading signals. The oscillator is restricted between a maximum of 100 and a minimum value of 0, $(0 \le D \le 100)$ and is calculated as follows:

$$K = \left(\frac{p_t - Low}{High - Low} \right) * 100$$

$$D = Ma_{n_2}(K)$$

The buy/sell signals are generated using the smoothened oscillator D by defining a filter k; a signal to go long (short) is generated when D breaches the upper (lower) band of the filter and a sell signal when it breaches the bottom band of the filter. The total combinations consist of upper filter bands, U = 100 – k, and lower filter bands, L = k, where n_1 = (10,15,…,40 days), n_2 = (3, 9, …, 27 days), k = (20,25,…,35) and e = 0 or 1. The total number of parameterizations for this strategy is 1,120 [= $(n_1 * n_2 * e) * (U * L)$].

Finally, the strategy named 'Aroon' detailed by Colby and Meyers (1988), detects changes in price momentum by utilizing the number of periods elapsed since the previous high and low within a lookback period, n. This consists of constructing two oscillators, namely "Aroon Up", $A_{up} = 100((n - n_H)/n)$ and "Aroon Down", $A_{dn} = 100((n - n_L)/n)$ where

$n_H\,(n_L)$ is the number of days elapsed since the previous maximum (minimum) price within the lookback period, n. On the one hand, an indication of a strong trend is made when a new maximum price is achieved within the look back period and A_{up} resets to 100. On the other hand, if a new maximum is not achieved within the look back period then A_{up} is equal to 0, indicating the end of the "bullish run". A long (short) position is triggered when $A_{up} > A_{dn}\,(A_{up} < A_{dn})$. The look-back period is the only pre-specified parameter in this strategy, where n = (6,8,…,100 days), resulting in 48 separate strategies.

c) Signals utilizing the volatility of the underlying are popularly referred to as Bollinger bands. They were developed in the 1980's and grew in popularity with traders who saw the obvious advantage of defining a maximum or minimum in price over a pre-specified lookback period that takes into consideration the dynamic instead of static volatility. These bands increase (decrease) in width with the increase (decrease) in the volatility over the lookback period (n_1 = 6,10,…,30 days). The bands are then applied to a smoothed price series, typically an arithmetic or exponentially weighted moving average of the price, n_2 = 6,9…,30 days, and a trading signal is generated when the smoothed price series crosses these bands. The bands therefore act as targets with a breach indicating a trend reversal. The upper Bollinger band Bu_t, is obtained by adding a pre-specified number of standard deviations (k =1.2, 1.4, …, 2.8) to the n_2-period moving average of the price series, $Pn_{2,t}$, as follows: $Bu_t = Pn_{2,t} + k\sigma_t$. The lower band conversely is obtained by subtracting $k\sigma_t$ from $Pn_{2,t}$ to give $Bl_t = Pn_{2,t} - k\sigma_t$. A long trading signal is generated when the price breaches Bl_t from below and is closed when the price then crosses Pn_2 from below yet again. A short trading signal is generated when the price series cuts the Bu_t from above and is closed out when it subsequently cuts $Pn_{2,t}$. Overall, this results in 1134 parameterizations [$(n_1 * n_2 * e) * (k) = 1134$].

d) A moving average envelope consists of two bands following a pre-specified moving average series and act as trading indicators. A signal is generated with the same rationale and manner as that of the rules comprising Bollinger bands. The bands are constructed by adding and subtracting a pre-specified percentage k to a smoothed price series. The smoothing is carried out using a n-period (n = 10, 14, …, 98 days) simple- or exponentially-weighted moving average. The upper envelope band (Eu_t) is obtained by adding k percent (k = 0.01, 0.012, …, 0.10) to the moving average of the price series, Pn_t ; thus $Eu_t = Pn_t + (k * Pn_t)$. The lower band conversely is obtained by subtracting $k * Pn_t$ from Pn_t to give $El_t = Pn_t - (k * Pn_t)$. A long trading signal is generated when the price breaches El_t from below and is closed when the price then crosses Pk_t from below yet again. A short signal is generated when the price series crosses Eu_t from above and is closed out when it subsequently crosses Pk_t . In total, there are 2116 individual strategies ($n * k * e$ = 2116).

Finally, channel breakouts or trading ranges are said to occur when the current price is higher than the high over a look-back period of n = 6, 8, …, 100 days, or lower than the low over the previous n days. The rule generates a buy signal when the price exceeds the channel and closes the position when the price moves below the channel. The channel is defined by the minimum and the maximum range over the look-back period of n days; the total number of strategies is 48.

Additionally, we follow the suggestion provided by Hsu and Kuan (2005) and supported by market reports, which indicate that practitioners may also use combinations of simple rules to support their trading decisions. In doing so they rely on the information gathered from a variety of simple rules to formulate a consensus. Therefore, we also consider two classes of complex rules: the complex voting trading rule (CVT) and the complex learning rule (CLR). The CVT rule generates a signal based on the majority amongst all the parameterizations of a particular rule; e.g., if the majority of the 1,960 parameterizations of the MAXC rule at time t generate a 'buy' signal and the remaining, either a 'sell' or 'neutral', then the voting rule will go with the majority position and generate a "buy" signal. Since we have 9 independent simple strategies, we have 9 parameterizations of the CVT rule. The complex learning strategy assumes that the practitioner has no fixed affinity to one particular strategy or class of strategies but instead changes over from one parameterization of a simple strategy to another based on a performance criterion (m =1 or 2) over a look-back period (n = 10, 20, …, 100 days). This rule allows the investor to switch their positions by utilizing the trading signal generated by the best performing rule amongst all the strategies over a look-back period. The performance criterion (m) is either the cumulative return (m = 1) or the Sharpe ratio (m = 2) resulting in 20 parameterizations [n * m].

The total number of rules to test on each of the FFA price series is 11,548, obtained by aggregating the combinations within each of the 9 simple trading strategies and the 2 complex strategies above. This can easily be extended to a larger number by simply reducing the size of the steps in the parameterizations or by adding additional parameters such as a minimum period for the (buy or sell) signal to be valid before an order is placed. However, finer granularity of trading parameters would increase the computational requirements without necessarily having an incremental effect on the performance of the outcomes. Overall, we feel that the combinations used provide a comprehensive yet parsimonious representation of market practices. A description of the contracts and the methodology for construction the continuous series in order to test these contracts is explained in Section-3.

3. Data and descriptive statistics

FFA's are cash-settled contracts for difference to settle a floating freight rate for a fixed quantity for one or more combinations of major shipping trade routes. The underlying asset of the FFA contracts are the basket trip-charter routes for the Baltic Capesize Index (BCI), the Baltic Panamax Index (BPI) and the Baltic Supramax Index (BSI) published by the Baltic Exchange. These indices reflect the cost of hiring a ship across a range of indicative shipping routes that reflect the typical trading patterns for each type of vessel (known as 4 trip-charter average (4TC) for the BCI and BPI and 6TC for the BSI) [2]. We focus on these routes owing to the majority of trading volume being concentrated there. FFA volume data, published by the

[2] A trip-charter contract is a shipping contract under which the charterer (shipper) agrees to hire the vessel from the owner (shipowner) for the duration of a specified trip. Normally, the charterer takes charge of the vessel from the point of delivery to the point of redelivery (after transportation and discharge of cargo) and pays the freight on a dollar per day basis ($/day). Under this type of contract, the shipowner has the operational control of the vessel, while the charterer is responsible for the voyage costs during the trip. For a description of the constituent shipping routes of the Baltic Exchange indices the reader is referred to Alizadeh and Nomikos (2009)

Baltic Exchange, indicate that Capesize contracts account for 44.2% of the market share in 2011 with Panamax contracts at 39.4% and Supramax and Handysize at a mere 14.6% and 1.9% respectively, which could in part be attributed to the higher volatility of the Capesize contracts, with the consequent need to hedge this risk and the attractiveness of trading opportunities from a speculator's perspective.

The data set used in this study consists of daily settlement prices from January 2005 to September 2011, thus adequately incorporating both the bullish period in freight rates from 2005 to the first half of 2008 as well as the subsequent crash from the second half of 2008 to the middle of 2009. Each route is further analyzed across contract maturities, namely the three nearest quarters (1Q, 2Q and 3Q) along with the two nearest calendars (1C and 2C); a quarter in this case refers to a group of three monthly contracts covering a respective quarter on a rolling basis. Similarly, a calendar contract consists of 12 individual monthly contracts covering a (calendar) year. In all, 5 maturities per route are investigated across the 3 routes.

[Insert Figure 1 about here]

In order to use the FFA price series for technical analysis, we need to construct a continuous series of FFA prices. This involves rolling over individual contracts to the next nearest in terms of maturity before their settlement date. Since FFA contracts settle using the average spot price over the maturity month, we construct the continuous series by rolling over the contract before the commencement of the averaging period. The motivation for this approach stems from the fact that participants who enter the averaging period will do so mainly with an intention to settle, and therefore if we encounter a trading signal (buy or sell, say) it would, in theory, be difficult to find a counterparty to trade. That means, for the quarterly contracts we close out the position a month prior to the settlement date of the first monthly contract for the quarter and a continuous series is therefore created by closing out the position a day before the averaging of the first contract and opening a new one in the next nearest contract; e.g., when trading 1Q(2011) we roll-over to 2Q(2011) in late December 2010 since 1Q starts averaging in the first week of January. Doing so results in both roll returns and transaction costs every time a trader switches between contracts. It is common practice in futures research to adjust the price level ex-post in order to remove the impact the roll-returns. We do not favor this approach as we feel it may significantly distort the results. Following Miffre and Rallis (2007) we take a correction-free approach since a proportion of the total returns from trading FFA's may be due to rolling-over between contracts of different maturities [3].

From the plot of the continuous series in Figure 1, we can see that the series across routes and maturities are highly correlated with each other. The period chosen for the in-sample analysis consists of both the bullish period of 2005-2008 as well as the subsequent bearish period following a correction where rates plunged by over 90% in less than six months; for instance the BCI 1Q(2009) contract was trading at 184,719 $/day on 4 June 2008, and by 1 December 2008 rates had fallen to 7,219 $/day. This price pattern is ideal for testing a host of trading strategies which individually perform better with certain regimes and patterns within the dataset. For example, trading strategies that involve moving averages are expected to perform well if there is a trending market. Conversely, less so in the case of a "flat" market, which exhibits minor price fluctuations without any recognizable trend. The validity of these assumptions is tested empirically in the next sections.

[Insert Table 1 about here]

Summary statistics for the daily logarithmic returns of the various continuous series are presented in Table 1. It can be seen from the Jarque-Bera test across all series/sectors that returns show significant departure from normality, unsurprisingly due to negative skewness and excess kurtosis. The ADF test (Dickey and Fuller, 1979) indicates that all series are non stationary in levels (results not reported in the table) but stationary in first differences. Serial correlation tests (Ljung and Box, 1978) indicate significant autocorrelation in the returns, and

---

[3] Roll return or Roll yield, is the price gap that arises when one switches between contract maturities. For instance, when we change over from the 1Q(2010) to the 2Q(2010) contracts and the market is in backwardation, then the 1Q contract is sold at a higher price and the 2Q contract bought at a lower price, thus resulting in a negative return for the trader. The significance of roll-returns in both the mean as well as the variance is also confirmed by following the testing methodology of Marshall et al. (2008b), which indicates that the inclusion of roll returns in the continuous series is appropriate. Results from these tests are available from the authors. In addition, transaction costs are generally fixed at 25 basis points of the contract value and are included in the subsequent calculations. Therefore the continuous price series include both transaction costs as well as roll yields.

as such any methodology to evaluate the significance of the trading strategies run on these series would have to incorporate this temporal dependence, as discussed in the Section 4. We can also see a well established relationship between vessel size and the volatility of the FFA contracts, as discussed in Alizadeh and Nomikos (2009). In the underlying freight markets, a larger vessel (Capesize, say) will offer greater economies of scale but is inflexible owing to its size in terms of the ports she can visit or the commodities carried. Smaller ships are more versatile in terms of the number of commodities they can carry, as well as being subject to fewer geographical restrictions and therefore, when freight markets are depressed, have better employment prospects. The same is evidenced in the FFA markets with Capesize contracts being more volatile compared to Panamax or Supramax contracts of similar maturities. We also evidence a downward sloping volatility term structure, which is attributed to the fact that contracts with maturity of up to two years are less volatile than ones with nearer maturity dates and is consistent with mean reverting behavior of the underlying spot series. Finally, we have 15 individual time series on which trading rules are tested. Issues concerning the effectiveness of these strategies are discussed next.

4. Methodology for detecting data snooping bias

According to STW, 'data snooping' occurs when the same dataset is used more than once for data selection and inference. Jensen and Benington (1970) refer to it as 'selection bias' and mention: "given enough computer time, we are sure that we can find a mechanical trading rule which "works" on a table of random numbers-provided of course that we are allowed to test the rule on the same table of numbers which we used to discover the rule" (p. 470). As a result, there is a possibility of certain trading rules demonstrating outperformance simply due to chance and not because of any inherent merit of the rules themselves. STW apply a large number of trading rules (7,846) to daily returns of the DJIA (Dow Jones Industrial Average) and find that none of the rules demonstrate any significant outperformance once data snooping bias is accounted for. There are two broad approaches in literature to mitigate the effect of data snooping. The first one, and simplest, is to use out-of sample performance tests as suggested by Hausman et al., (1992). The second is to use White's (2000) Reality Check (WRC), first used by STW, along with later improvements by Hansen (2005). Although we use both methods in this paper, we emphasize the findings of WRC as it allows us to quantify the effects of data snooping bias, despite being computationally demanding.

The intuition behind using the reality check is that the best model encountered during a specification search will have no predictive superiority over a benchmark model. This also happens to be the null hypothesis for this test. In doing so, the WRC test evaluates the distribution of a performance measure over the universe of trading rules. The performance statistic is a $l x 1$ vector;

$$\overline{f} = n^{-1}\sum_{t=d}^{T} \hat{f}_{t+1} \qquad (1)$$

where l is the number of trading rules specified in Section 2, n is the number of prediction periods from d to T, where d is the time taken for the rule to 'learn' and T = d + n - 1. For instance, a moving average crossover rule MA[3, 13] takes 13 periods to generate a signal and therefore $d_{[3,13]} = 14$. In our case, the maximum value of d = 101. $\hat{f}_{t+1} = f(Z_t, \hat{\beta}_t)$ is the observed performance measure one period ahead (t + 1), Z_t is a matrix containing the vector of dependent and explanatory variables and $\hat{\beta}_t$ is a vector of parameters estimated from an econometric model. In the case of the trading strategies employed in this study, returns are generated directly without the need to estimate additional parameters, and as such the parameterizations of those trading rules are given by β_k , k = 1, …, l. The observed performance measure for an individual rule k, $\hat{f}_{k,t+1}$, can then be specified as:

$$\hat{f}_{k,t+1} = \ln\left[1 + \frac{P_{t+1} - P_t}{P_t} X_k(Z_t, \beta_k)\right] - \ln\left[1 + \frac{P_{t+1} - P_t}{P_t} X_0(Z_t, \beta_0)\right], \ k = 1, \ldots, l \qquad (2)$$

where P_t is the price of the FFA contract at time t, β_0 indicates the benchmark model and X_k and X_0 are "signal" functions that indicate the market positions for active and buy & hold strategies, respectively. The function X_k may take one of three values; 1 indicating a long position in the market, 0 for a neutral position and -1 for a short position. The buy and hold strategy against which we benchmark the active strategies will always be long $\{X_0 = 1\}_{t=d}^{t=T}$. The null hypothesis that the performance of the best rule is no better than the performance of the benchmark can then be formulated as follows:

$$H_0 : \max_{k=1,\dots,l}\{E(f_k)\} \leq 0 \tag{3}$$

Rejecting the null at conventional levels of significance implies that the best performing rule indeed outperforms the benchmark and this outperformance is not due to chance alone. To evaluate the null hypothesis one can use the recursive method proposed by White (2000) to calculate the WRC p-value.

In order to obtain meaningful estimates of the extent of data snooping bias we need to have an expanded universe of technical rules, as using a small number of rules could cause biased statistical inference owing to the very effect we are trying to mitigate. However, according to Hansen (2005), introducing a lot of irrelevant rules could also reduce the power of WRC test due to the fact that its null distribution is obtained under the least favorable configuration, making it sensitive to the addition of a large number of irrelevant rules. As such, he proposes an alternative procedure known as the Superior Predictive Ability (SPA) test, which is the benchmark test we use here. The SPA test employs a studentized test statistic and also invokes a sample dependent distribution under the null, making it more powerful and less sensitive to the inclusion of irrelevant alternatives. The test statistic is denoted as:

$$T_n^{SPA} = \max\left[\max_{k=1,\dots,m}\{\frac{\sqrt{n}\,\bar{f}_k}{\hat{\omega}_k}, 0\}\right] \tag{4}$$

Where $\hat{\omega}_k^2$ is a consistent estimator of $\omega_k^2 = \mathrm{var}(\sqrt{n}\,\bar{f}_k)$. Hansen (2005) proposes the following estimator $\hat{\mu}_k^c$ which is robust in separating outperforming from poorly performing or irrelevant models:

$$\hat{\mu}_k^c = \bar{f}_k \mathbf{1}_{\{\sqrt{n}(\bar{f}_k/\hat{\omega}_k) \leq -\sqrt{2\ln\ln n}\}} \tag{5}$$

As there may be alternative threshold rates producing SPA p-values, Hansen (2005) also determines the upper (superscript u) and lower bounds (superscript l) of the estimators denoted by $\hat{\mu}_k^l = \min(\bar{f}_k, 0)$ and $\hat{\mu}_k^u = 0, \quad k = 1, \ldots, l$, respectively. We report all three SPA p-values in our results.

The implementation of the SPA test is fairly similar to that of the WRC. We start by generating a resample of $\{f_{k,t+1}\}_{t=R}^{t=T}$ for each rule $k = 1, \ldots, l$ using a block bootstrap method, such as Politis and Romano (1994). The resample series is denoted by $\{f^*_{k,t+1,i}\}$, where i indicates the i-th bootstrap repetition. The next step involves calculating $Z^*_{k,t+1,i} = f^*_{k,t+1,i} - \bar{f}_k \mathbf{1}_{\{\sqrt{n}(\bar{f}_k/\hat{\omega}_k) \geq -\sqrt{2 \ln \ln n}\}}$, $\forall k = 1, \ldots, l, t = d, \ldots, T$. Subsequently, the test statistics are computed in the following manner:

$$V_{SPA} = \max[\max_{k=1,\ldots,l} \frac{\sqrt{n}\,\bar{f}_k}{\hat{\omega}_k}, 0], \quad \text{and} \tag{6}$$

$$V^*_{SPA,i} = \max\left[\max_{k=1,\ldots,l} \frac{\sqrt{n}\bar{Z}^*_{k,i}}{\hat{\omega}_k}, 0\right], \quad i = 1, \ldots, B \tag{7}$$

where $\bar{Z}^*_{k,i} = n^{-1} \sum_{t=d}^{T} Z^*_{k,t+1,i}$: V_{SPA} is then compared with the quantiles of $V^*_{SPA,i}$ and finally the p-value for the SPA is given by;

$$P_{SPA} = \sum_{i=1}^{B} \frac{\mathbf{1}\{V^*_{SPA,i} > V_{SPA}\}}{B} \tag{8}$$

P_{SPA} is the consistent SPA p-value, and replacing $\hat{\mu}_k^c$ with $\hat{\mu}_k^l$ and $\hat{\mu}_k^u$ we can compute its lower and upper bounds as SPA_L and SPA_U respectively; $\hat{\mu}_k^c$ is equivalent to the (studentized) WRC test p-value.

Since the vectors of outperformance, $\hat{f}_k$, are serially-dependent stationary time series, the SPA test is conducted using a block time series bootstrap. In order to carry out the block bootstrap one has to specify a mean block length which opens up room for subjective bias. White (2000) suggests the selection of the mean block length w in a data dependent manner, although in most empirical studies it is common practice to (jointly) specify w as 10 and subsequently carry out parameter sensitivity checks towards robustness. Romano and Wolf (2005) incorporate this suggestion in an extension of the WRC test by fitting a semi-parametric model to the vectors $\hat{f}_k$ and computing a joint confidence region. They claim good finite sample properties of the WRC test using this approach which, however, is computationally intensive and involves a further subjective bias in the form of the selection of an appropriate semi-parametric model. To the best of our knowledge no other study involving the WRC/SPA test invokes this procedure. We replace the suggestion made by Romano and Wolf (2005) with a methodology proposed by Politis and White (2004), which produces estimators of the optimal block sizes w_{opt} using the notion of spectral estimation via the flat-top lag-windows of Politis and Romano (1995)[4]. All results in this paper are therefore computed using 10,000 bootstrap repetitions based on the optimal block lengths for each outperforming strategy, $w_{opt,k}$, which makes redundant the need for carrying out and reporting sensitivity checks to varying values of w.

5. Empirical Results

We tested 11,548 trading rules on three dry FFA contracts (Capesize 4TC, Panamax 4TC and Supramax 6TC) across five maturities (3 nearest quarters 1Q, 2Q and 3Q and 2 nearest calendars 1C and 2C) for a sample period starting from January 2005 to September 2011. We present the performance metrics of the "winning" (outperforming) strategy across all contracts and maturities with various combinations that act as robustness checks. The outperformance of the trading strategies is measured in terms of their mean annualized returns over the benchmark buy and hold strategies. We also identify the outperforming strategies in terms of risk-adjusted returns, i.e. using Sharpe ratios. We find that although the detailed parameterization of the outperforming strategies is different, all instances are drawn from the same class of trading strategies as when mean returns were used.

We test five alternative scenarios across our strategies; the first is the long/short strategy, where the trader can take both long and short positions depending on the market signal without any

[4] We incorporate the procedure coded by Dr. Andrew Patton, which corrects for theoretical errors in Politis and White (2004) and returns the optimal 'w' by minimizing a loss function of the variance/autocovariance of the bootstrapped data given a primary random draw. The code is available at: http://www.economics.ox.ac.uk/members/andrew.patton/code.html .

short selling restrictions. The next two scenarios mimic participants that could either take a long position or short position but not both. We assume that the short only participant is a ship-owner who is hedging his position in the physical market by selling FFA and in addition takes (short-only) speculative positions; a charterer is considered the counterparty to this transaction and is allowed to take long positions only. In the fourth strategy, we recalculate the performance metrics with the introduction of an order delay parameter in order to account for illiquidity in the market. The final scenario assumes that the participant is allowed to go both long and short but exits the markets for 190 days during the turbulent period from 1 April, 2008 to 24 December, 2008. This is carried out in order to discount the possibility of a particular trading strategy making excessive returns during this period by taking a short position on the market but in general exhibiting poor performance overall. Thus the 5 scenarios outlined above effectively represent the majority of the participation mix.

The performance metrics of the buy-and-hold benchmark strategy are presented in Table 1. This illustrates the profitability of the net position of a participant who is effectively long throughout the sample period; this passive position still incurs transaction costs when the participant “rolls” his position forward to the next nearest maturity, effectively closing-out his position before the contract enters the settlement period and buying the next nearby contract. The benchmark returns across contracts and term structure are characterized by high annualized volatility and therefore relatively smaller risk-adjusted performance, the Sharpe ratios being less than 1 in all cases. In addition, we utilize the procedure proposed by Opdyke (2007) in adjusting the Sharpe ratios for higher moments (skewness and kurtosis), which also allows us to make statistical inferences. The adjusted Sharpe ratios are lower than the corresponding un-adjusted ones and are also not statistically significant at conventional levels.

[Insert Table 2 about here]

The results from the long-short strategy are presented in Table 2. We can see that the rule that provides the best outperformance, in terms of mean returns, is the Moving Average Crossover Consensus [MAXC] Rule. The rule is consistent in outperforming the benchmark across vessel sizes and the term structure with alternative parameterizations of the rule yielding qualitatively similar performance metrics. The rule allows the practitioner to trade on strong trends only and exit the market when the trend weakens, as opposed to the other trend following rules which are less strict with respect to the entry and exit conditions.

The mean annualized returns of the active strategies, when compared with the respective benchmark metrics, show a significant outperformance across contracts as well as maturities [5]. One can also see a clear trend in outperformance across vessel sizes since outperformance is greater in the Capesize followed by the Panamax and Supramax contracts. Since Capesize vessels are relatively constrained in terms of the cargoes they can carry (primarily coal and iron ore) as well as the ports they can visit due to size and depth restrictions, price signals emanating from the underlying commodity markets are less diffused compared to smaller vessels, who are more versatile in terms of commodities they may carry and have a wider geographical reach. As such, trend-following rules seem to provide superior outperformance for Capesize vessels.

The relative outperformance of trend-following strategies may also be attributed to the presence of "momentum traders" as described in Hong and Stein (1999), who identify and profit from changing trends due to the arrival of price-sensitive news that may come in the market in the form of supply/demand shocks to the underlying commodity or freight markets. Their actions may also be motivated by the slow diffusion of information through trade reports or the network of brokers, owing to the lack of a centralized market mechanism as is the case in most financial markets. In addition, non-linearities and deviations from normality, as evidenced by the descriptive statistics reported in Table 1, may also contribute to the superior (risk-adjusted) performance of technical trading strategies, as discussed by Neftci (1991).

The active trading strategies also have a lower standard deviation compared to the benchmark indicating a reduction in the riskiness of the exposure to the FFA markets. As a result, risk-adjusted returns show an improvement for the active strategies, as evidenced by the Sharpe and Sortino Ratios, which measure the excess return per unit of total risk and downside risk,

---

[5] The annualized returns especially for the active strategies appear to be very high; this is primarily owing to the reported returns being the largest across the group of 11,548 parameterizations. At this stage returns are reported without any robustness checks or consideration for their significance, which will follow later in this section.

respectively[6]. In terms of skewness and kurtosis, we observe that the active strategy is positively skewed (with the exception of Capesize 2C) in contrast to the passive strategy, which is negatively skewed across all contracts and maturities (with the exception of Panamax 1Q). Aside of the superior risk-adjusted outperformance, the positive skewness is yet another attractive feature of actively trading FFA. Finally, the active strategy also shows excess kurtosis and is more leptokurtic compared to the benchmark, with the exception of two instances (Panamax and Supramax 2C).

The use of "two dimensional" performance metrics like the Sharpe ratio or the increasingly popular Sortino ratio is common across literature as well as financial markets. These measures incorporate the first and second moments of the returns series only, and as such do not consider the impact that adverse changes in higher moments may have on the performance of trading strategies. Thus, we also report adjusted Sharpe ratios (Opdyke, 2007) to illustrate the bias induced when higher order moments are ignored. We can see that adjusted Sharpe ratios show a much steeper decline across the term structure compared to un-adjusted Sharpe ratios, consistent with the larger magnitude of higher order moments. The metrics are reported along with their p-values, which show that all active trading strategies have positive and significant Sharpe ratios, which are also significantly superior to those of the benchmark strategy.

Additional metrics commonly used by market practitioners to evaluate trading strategies include total market exposure, average days per trade and ratio of winning to losing trades. It would be unrealistic for active trades to have very little total market exposure as well as large periods of inactivity, even if the rule is an outperforming rule. We can see that the outperforming rules have a market exposure of about 42% with an average exposure of 7 days per trade and an average maximum period out of the market of 14 days, across all vessel sizes and maturities. Similarly, some trading rules may generate noisy trades, i.e. buy or sell signals that are too frequent; we therefore compute the ratio of number of 'winning' trades to the total

[6] Sortino ratio measures the actual rate of return in excess of the investor's target rate of return per unit of downside risk; the target is kept at 0%. The downside risk is slightly larger for the active strategies however Sortino ratios are higher than the benchmark due to superior outperformance.

number of trades, which on average across contracts and maturities is around 89%, and thus strongly in favor of active trading.

It is interesting to note that from all the various trading strategies employed, the highest outperformance is attributed to various parameterizations of the Moving Average Crossover Consensus rule [MAXC]. This is also confirmed visually in Figure 2, which presents the outperformance, measured in terms of Sharpe ratios and mean returns, of the different parameterizations of trading strategies across vessel sizes and contract maturities. Each point on the surface of the graphs represents the mean across all parameterizations of a single strategy for a particular contract maturity. For example the average of the mean annualized returns across 1,960 parameterizations of the MAXC rule for Capesize 1Q constitutes a single point on the graph surface.

[Insert Figure 2 about here]

The superiority of the three trend following strategies (MACD, MAX and MAXC) is immediately apparent in the graphs. We can also note that with the exception of the channel break-out rules, momentum- and volatility-based strategies fail to provide positive mean returns or Sharpe ratios. Complex voting strategies (CVT) do not seem to outperform the benchmark, in contrast to complex learning strategies (CLR), which provide the highest average outperformance in terms of Sharpe ratios. Despite the poor average performance of some strategies, certain parameterizations of those strategies may provide significant outperformance, and hence may be used by practitioners possibly in conjunction with other rules to support their trading positions and allow for a richer and more comprehensive investigation of trading performance.

[Insert Table 3 about here]

In Table 3 we also calculate performance metrics for two additional scenarios, where the traders are allowed to take long-only (Panel A) and short-only (Panel B) trading positions. As expected, long-short trading strategies significantly outperform long-only or short-only positions, with the outperformance of the long-only participant being greater than his short-only counterpart. This is also reflected in the Sharpe ratios, where long-short positions have the largest risk-adjusted return followed by the long-only and short-only participants owing to superior returns but similar risk profiles. Comparing next the skewness and kurtosis of the strategies across the 3 scenarios, we find that, as expected, the long-only scenario has the largest positive skewness overall. The skewness is higher in the calendar contracts, particularly in 1C, possibly owing to these contracts being less liquid than their near-term counterparts. These contracts may also be subject to price "jumps", possibly resulting from large orders placed by participants, thus significantly moving the markets which are thinly traded. In addition, the roll-returns in these markets contribute significantly to the "jump" component, thereby affecting both skewness and kurtosis. Similarly the excess kurtosis of the long-only and short-only scenarios is larger than that of the others. Overall, judging by the increase in the Jarque and Bera (1987) test statistic (not reported), the non-normality of the active returns is higher than that of the benchmark though in all cases the series are not normal at conventional significance levels.

Another significant advantage of utilizing active trading strategies in the FFA markets is the reduction in size and duration of the (maximum) drawdown. The drawdown measures the decline in the cumulative returns from a historical peak. The maximum drawdown experienced in the FFA markets was during the market crash of 2008; in particular, the period from 4 June 2008 to 1 December 2008, saw FFA rates fall by over 90%. All the active strategies show that the duration of the drawdown as well as its intensity is largely reduced when compared to the benchmark. Upon analyzing the cumulative returns of the active strategies overlaid with the buy/sell market timing signals, we find that all the outperforming rules had short positions during the crash due to the strong downward trend. Both the magnitude and the duration of the maximum drawdown for the active strategies is considerably smaller, which is consistent with the superior risk-adjusted performance of the active strategies.

Following the methodology described in Section 4 we report the results of the WRC/SPA data snooping test in Table 4. We aim to investigate whether the superior outperformance shown by the active trading strategies is attributed to the genuine properties of the trading rule or is due to a data snooping bias. [7]

[Insert Table 4 about here]

Results from the joint hypothesis test are presented in the SPA test column, where the lower, consistent and upper p-values of the SPA test are presented. The WRC p-value is not reported separately as the upper SPA p-value is the same as the studentized WRC p-value. After 10,000 bootstrap repetitions across 11,548 different parameterizations, we find that all the SPA consistent p-values are highly significant (lower than 0.01), indicating superior outperformance after allowing for data snooping bias for the long- short strategies [see Panel A, Table 4] . We can see that Supramax p-values are marginally greater than 0 but still highly significant, especially in the near maturities.

Up to now the analysis has been carried out assuming perfect liquidity and no order delay. Therefore, when a signal to buy or sell is generated by the rules described above, it is assumed that the order placed will get executed immediately. However, this may not be the case with FFA as execution is primarily carried out through a network of brokers and may generally be subject to an order execution delay.[8] To accommodate this, we introduce a variable $o_d \in \{0,1,..,5\}$ which represents order execution delays, in days, from the time the signal is generated up to the time the order is executed, and is randomly drawn from a geometric

[7] Table 4 reports the SPA p-values on the basis of the best performing rule in terms of the in-sample performance, as presented in Table 2 and Table 3. Prior to carrying out the WRC/SPA tests, we perform unit root tests on the vectors of outperformance $\hat{f}_k$ to ensure that they are strictly stationary; this follows Assumption-1 of Hansen (2005) and is a necessary requirement for the use of stationary bootstrap, on which the test is based on. ADF unit root tests indicate that for the mean return criterion all outperformance vectors follow an I(0) process, however when Sharpe ratios are used as outperformance criteria, less than 10 percent of the vectors are found to be I(0). As such WRC/SPA values are only reported for the mean returns criterion.

[8] For instance, it is estimated that only 10% of the FFA trades is effected via online trading screens, the remaining being traded in the OTC market via voice broking. When a trade signal (long or short) is generated, the trader will get on the phone to the broker who will then facilitate the execution of the trade. We mimic the execution delay by introducing a partial fill for larger orders (1000 lots are filled in over three days incurring different prices) and also a randomized order execution delay for smaller orders or single lots; the order delay is randomly selected from a geometric distribution. Finally, it is also worth noting that more than 98% of the contracts are subsequently novated via a Clearing House thus eliminating counterparty risk.

distribution. Doing so, we specify the probability of the order going through in the first day (without any delays) to $P(o_0)=0.5$. This assumption is convenient as it translates to a probability of the order not going through even on the fifth day at around $P(o_5)=0.016$, which is a fairly realistic assumption to make in the FFA markets. The price is then averaged across the delay period, which mimics the case of a large order that is filled in at different rates. Results from this case are presented in Panel B, and are largely similar with the perfect liquidity trades, with no change in the significance in any of the cases. The outperformance therefore remains un-diluted with the introduction of order delays in trades.

Due to the very large price movements during the financial crisis of 2008 it is likely that the overall profitability of the trading rules may have been affected over that period. It is equally likely, that certain opportunities would not be realized owing to a lack of suitable counterparties at the time of the market crash. In order to test the robustness of our results to these arguments, we eliminate the turbulent period from 1 April, 2008 to 24 December, 2008 and re-estimate the significance of the outperformance (Panel C of Table 4). We can see that the Supramax 2C contract becomes insignificant at 1% with no perceivable changes in the significance of the remaining contracts; therefore the results are largely robust to extreme market movements experienced during this turbulent period.

The results are further assessed for long-only (Panel D of Table 4) and short-only (Panel E of Table 4) positions. For the long-only analysis we find that the Supramax 1C and 2C contracts are not significant at 1%. The most striking results are seen in the SPA p-values of the short-only positions (Panel E of Table 4) where none of the 15 series assessed are significant at 5% with the highest p-values observed in the Supramax sector; it is also interesting to note that short only positions have the lowest market exposures on average, ranging from 8% to 17% across contracts and maturities. This observation has clear implications for market participants who are naturally short on the FFA market - such as ship owners - and are looking to trade beyond their hedging exposure, as they would be unable to time the market in a significantly profitable manner. Charterers on the other hand, who are naturally long on FFA would be able to successfully time the market based on trading rules. This could be partly responsible for the relatively smaller participation of ship owners in this market. Investigating this issue further, we eliminate the market crash period from the short only scenario and find that the p-values are larger than before and as such not sensitive to this turbulent period. Therefore, we are able to highlight the significance of market timing strategies after allowing for several robustness checks in most cases except for a participant who would be short-only.[9]

6. Conclusion

In this study, we show that technical trading rules can be profitably applied to FFA contracts. To the best of our knowledge, a rigorous study of this sort has not been carried out before. Our results indicate that trading strategies based on momentum and trends in prices are capable of generating significant excess returns over the benchmark of buy-and-hold, within the dry markets even after accounting for data snooping bias. This is further confirmed by incorporating an order execution delay parameter, by testing different variants of the basic strategy such as long-only trades as well as by excluding the market crash period of 2008. In addition, the results are robust to the choice of block-length parameter for the bootstrap replications, transaction costs and roll-yields.

We also find that employing active trading strategies reduces risk and improves Sharpe ratios over buy-and-hold. Other advantages include a reduced drawdown and reduced market exposure, which is a result of appropriately timing exit signals during periods of sharp trend changes. These benefits coupled with evidence of trading strategies generating significant outperformance make a strong case for active trading in the FFA markets.

There is evidence of a certain "coming of age" for the FFA dry markets in terms of liquidity, transparency and credit risk mitigation with increased participation from both hedging and

[9] Out of sample tests were also carried out in order to further substantiate our findings across contracts and maturities. We utilized 66% of the total sample in order to identify the best rules in terms of market outperformance. We find that these rules significantly outperform the benchmark during the out-of-sample period in terms of both mean returns as well as risk-adjusted returns. These results are available from the authors.

speculative interests. This generally implies that opportunities for profitable trades using trading strategies would gradually diminish. However, we find that has not been the case, as of yet. On the contrary, since the FFA market is considered a "young" market, we are still able to find a large amount of economically significant market inefficiencies across contracts and maturities in spite of robustness checks and a host of perceived market proxies.

Bibliography

Adland, R. & Strandenes, S. 2006, "Market efficiency in the bulk freight market revisited", Maritime Policy & Management, vol. 33, no. 2, pp. 107-117.

Alizadeh, A.H. & Nomikos, N.K. 2009, Shipping Derivatives and Risk Management, Palgrave Macmillan.

Alizadeh, A.H. & Nomikos, N.K. 2007, "Investment timing and trading strategies in the sale and purchase market for ships", Transportation Research Part B: Methodological, vol. 41, no. 1, pp. 126-143.

Aronson, D. 2011, Evidence-based technical analysis: applying the scientific method and statistical inference to trading signals, Wiley.

Bakshi, G., Panayotov, G. & Skoulakis, G. 2011, "The Baltic Dry Index as a predictor of global stock returns, commodity returns, and global economic activity".

Brock, W., Lakonishok, J. & LeBaron, B. 1992, "Simple technical trading rules and the stochastic properties of stock returns", Journal of Finance, vol. 47, no. 5, pp. 1731-1764.

Brown, D.P. & Jennings, R.H. 1989, "On technical analysis", Review of Financial Studies, vol. 2, no. 4, pp. 527.

Colby, R.W. & Meyers, T.A. 1988, The encyclopaedia of technical market indicators, Dow Jones-Irwin.

Dickey, D.A. & Fuller, W.A. 1979, "Distribution of the estimators for autoregressive time series with a unit root", Journal of the American statistical association, , pp. 427-431.

Fama, E.F. & Blume, M.E. 1966, "Filter rules and stock-market trading", Journal of Business, vol. 39, no. 1, pp. 226-241.

Geman, H. 2008, Risk Management in Commodity Markets: From Shipping to Agriculturals and Energy, John Wiley & Sons.

Hansen, P.R. 2005, "A test for superior predictive ability", Journal of Business and Economic Statistics, vol. 23, no. 4, pp. 365-380.

Hausman, J.A., Lo, A.W. & MacKinlay, A.C. 1992, "An ordered probit analysis of transaction stock prices", Journal of Financial Economics, vol. 31, no. 3, pp. 319-379.

Hong, H., & Stein, J. 1999, “A unified theory of under reaction, momentum trading, and overreaction in asset markets”, Journal of Finance, 54, pp. 2143-2184.

Hsu, P.H. & Kuan, C.M. 2005, "Re-examining the profitability of technical analysis with data snooping checks", Journal of Financial Econometrics, vol. 3, no. 4, pp. 606.

Jarque, C.M. & Bera, A.K. 1987, "A test for normality of observations and regression residuals", International Statistical Review, vol. 55, no. 2, pp. 163-172.

Jensen, M.C. & Benington, G.A. 1970, "Random walks and technical theories: Some additional evidence", The Journal of Finance, vol. 25, no. 2, pp. 469-482.

Kilian, L. 2009, "Not all oil price shocks are alike: Disentangling demand and supply shocks in the crude oil market", The American Economic Review, vol. 99, no. 3, pp. 1053-1069.

Ljung, G.M. & Box, G.E.P. 1978, "On a measure of lack of fit in time series models", Biometrika, vol. 65, no. 2, pp. 297-303.

Lo, A. 2004, "The adaptive markets hypothesis: Market efficiency from an evolutionary perspective", Journal of Portfolio Management, vol. 30, no. 5, pp. 15-29.

Marshall, B.R., Cahan, R.H. & Cahan, J.M. 2008, "Can commodity futures be profitably traded with quantitative market timing strategies?", Journal of Banking & Finance, vol. 32, no. 9, pp. 1810-1819.

Marshall, B.R., Cahan, R.H. & Cahan, J.M. 2008, "Does intraday technical analysis in the US equity market have value?", Journal of Empirical Finance, vol. 15, no. 2, pp. 199-210.

Miffre, J. & Rallis, G. 2007, "Momentum strategies in commodity futures markets", Journal of Banking & Finance, vol. 31, no. 6, pp. 1863-1886.

Neely, C., Weller, P. & Dittmar, R. 1997, "Is technical analysis in the foreign exchange market profitable? A genetic programming approach", Journal of Financial and Quantitative Analysis, vol. 32, no. 04, pp. 405-426.

Neftci, S. 1991, “Naive trading rules in financial markets and Wiener-Kolmogorov prediction theory: A study of technical analysis”, Journal of Business **64**, pp. 549-571.

Opdyke, J. D. 2007, “Comparing Sharpe ratios: So where are the p-values?”, Journal of Asset Management, **8**(5), pp. 308-336.

Park, C.H. & Irwin, S.H. 2007, "What do we know about the profitability of technical analysis?", Journal of Economic Surveys, vol. 21, no. 4, pp. 786-826.

Politis, D.N. & Romano, J.P. 1995, "Bias - Corrected Non Parametric Spectral Estimation", Journal of Time Series Analysis, vol. 16, no. 1, pp. 67-103.

Politis, D.N. & Romano, J.P. 1994, "The Stationary Bootstrap", Journal of the American Statistical Association, vol. 89, no. 428.

Politis, D.N. & White, H. 2004, "Automatic block-length selection for the dependent bootstrap", Econometric Reviews, vol. 23, no. 1, pp. 53-70.

Romano, J.P. & Wolf, M. 2005, "Stepwise multiple testing as formalized data snooping", Econometrica, vol. 73, no. 4, pp. 1237-1282.

Sullivan, R., Timmermann, A. & White, H. 1999, "Data-snooping, technical trading rule performance, and the bootstrap", The journal of Finance, vol. 54, no. 5, pp. 1647-1691.

Sweeney, R.J. 1986, "Beating the foreign exchange market", Journal of Finance, vol. 41, no. 1, pp. 163-182.

Timmermann, A. & Granger, C.W.J. 2004, "Efficient market hypothesis and forecasting", International Journal of Forecasting, vol. 20, no. 1, pp. 15-27.

Treynor, J.L. & Ferguson, R. 1985, "In defense of technical analysis", Journal of Finance, vol. 40, no. 3, pp. 757-773.

White, H. 2000, "A reality check for data snooping", Econometrica, vol. 68, no. 5, pp. 1097-1126.

Table 1: Descriptive statistics and performance metrics for the benchmark (Buy and hold) log-returns.

Criterion	Capesize 4TC					Panamax 4TC					Supramax 6TC				
	1Q	2Q	3Q	1C	2C	1Q	2Q	3Q	1C	2C	1Q	2Q	3Q	1C	2C
Mean Returns (annualized)	0.1545	0.1273	0.0699	0.1295	0.0524	0.1713	0.2903	0.2508	0.1972	0.0933	0.3814	0.3921	0.2885	0.2223	0.1269
Standard Deviation of Returns (annualized)	0.7544	0.6024	0.544	0.5152	0.3738	0.6212	0.5212	0.4516	0.4573	0.3345	0.4743	0.4186	0.3672	0.3968	0.2728
Sample size (N)	1699	1699	1658	1699	1699	1699	1699	1658	1699	1699	1438	1438	1438	1438	1438
Skewness	-0.33	-0.58	-0.28	-3.62	-2.93	0.03	-0.61	-0.62	-3.25	-3.03	-0.75	-0.92	-0.70	-3.18	-4.12
Kurtosis	8.04	9.34	12.14	50.53	41.11	9.79	8.04	9.51	42.61	40.42	10.49	10.60	12.15	47.87	63.94
Jarque-Bera	1832 [0.0000]	2949 [0.0000]	5937 [0.0000]	163669 [0.0000]	105307 [0.0000]	3273 [0.0000]	1905 [0.0000]	3119 [0.0000]	114114 [0.0000]	101743 [0.0000]	3529 [0.0000]	3699 [0.0000]	5188 [0.0000]	124103 [0.0000]	228519 [0.0000]
ADF-Test	-30 [0.0000]	-26 [0.0000]	-26 [0.0000]	-33 [0.0000]	-31 [0.0000]	-31 [0.0000]	-26 [0.0000]	-29 [0.0000]	-27 [0.0000]	-27 [0.0000]	-26 [0.0000]	-29 [0.0000]	-27 [0.0000]	-32 [0.0000]	-30 [0.0000]
Q(24)	222 [0.0000]	230 [0.0000]	246 [0.0000]	199 [0.0000]	264 [0.0000]	187 [0.0000]	240 [0.0000]	213 [0.0000]	165 [0.0000]	257 [0.0000]	229 [0.0000]	225 [0.0000]	256 [0.0000]	198 [0.0000]	257 [0.0000]
Sharpe Ratio	0.2048	0.2113	0.1286	0.2514	0.1401	0.2757	0.5569	0.5553	0.4313	0.2790	0.8040	0.9368	0.7856	0.5602	0.4654
Sharpe Ratio (adjusted)	0.1378 [0.8168]	0.1310 [0.7729]	0.0605 [0.8728]	0.1076 [0.6878]	0.0348 [0.8313]	0.1761 [0.7002]	0.4013 [0.3247]	0.3762 [0.2636]	0.1916 [0.4052]	0.1032 [0.5283]	0.4004 [0.2120]	0.4609 [0.0983]	0.3254 [0.1645]	0.1014 [0.4331]	0.0315 [0.6826]
Sortino ratio (0%)	0.2856	0.2893	0.1729	0.3627	0.1986	0.3731	0.7869	0.7776	0.6329	0.3911	1.147	1.3595	1.117	0.8557	0.6791
Maximum Drawdown	3.25	2.58	3.03	2.42	2.02	2.64	1.97	1.76	1.93	1.43	2.17	1.75	1.59	1.86	1.30
Duration of Drawdown	97	78	791	129	712	133	195	169	169	97	126	126	179	160	179
% of positive Returns	0.5	0.4994	0.4944	0.5006	0.4925	0.5225	0.5231	0.5081	0.5069	0.5106	0.5274	0.5274	0.5207	0.5141	0.5104

Notes: Table 1 shows descriptive statistics for the log differences of FFA rates in three vessel size categories (Capesize, Panamax and Supramax) across the term structure (5 maturities - 1Q, 2Q, 3Q, 1C and 2C - each). Data are daily for the period 4 January 2005 to 30 September 2011 (Capesize and Panamax) and 3 January 2006 to 30 September 2011 (Supramax). Since the exposure to the individual contract is essentially the benchmark strategy (always long), we present the performance metrics for the benchmark returns. These returns are adjusted for roll-yields and associated transaction costs as described in the text. N denotes the number of daily observations. Skewness and kurtosis are the centralized third and fourth moments of the data. J-B is the Jarque-Bera (1980) test for normality; Q(24) is the Ljung and Box (1978) Q statistic on the first 24 lags of the sample autocorrelation function of the level series, distributed as $\chi^2(24)$. ADF is the Augmented Dickey and Fuller (1981) test, 5% critical value for this statistic being -2.8794. Figures in [ ] are p-values. The Sharpe ratio $S = [R_A - R_f]/\sigma$ is a measure of excess annualized returns $[R_A - R_f]$ (realized returns minus risk free rate) divided by the annualized standard deviation of returns which is then adjusted for third (skewness) and fourth (kurtosis) moments using the procedure by Opdyke (2007); p-values for adjusted Sharpe ratios are in [ ]. The Sortino ratio is similar to the Sharpe ratio but the volatility measure is constructed using negative returns only. Maximum drawdown is the largest decline in cumulative returns of the series from a historical peak; formally: $\mathrm{MD} = \max_{\tau\in(0,\mathrm{T})}\left[\max_{\mathrm{t}\in(0,\tau)}\left(\mathrm{P_t} - \mathrm{P_\tau}\right)\right]$ where $\max_{\mathrm{t}\in(0,\tau)}\left(\mathrm{P_t} - \mathrm{P_\tau}\right)$ is defined as the drawdown at time τ and $\mathrm{P_t}$ is the price at time t. The duration of the drawdown is simply the difference between $\mathrm{t} - \tau$, expressed in days. The ratio of positive returns to total returns is expressed as a percentage.

Table 2: Performance measures for outperforming active strategies: Long & Short trades.

Criterion	Capesize 4TC					Panamax 4TC					Supramax 4TC				
	1Q	2Q	3Q	1C	2C	1Q	2Q	3Q	1C	2C	1Q	2Q	3Q	1C	2C
Description of rules	MAXC § (3/51/61)	MAXC (3/60/61)	MAXC (3/60/61)	MAXC (3/51/61)	MAXC (3/60/61)	MAXC (3/60/61)	MAXC (3/60/61)	MAXC (3/60/61)	MAXC (3/60/61)	MAXC (3/57/61)	MAXC (3/60/61)	MAXC (3/60/61)	MAXC (3/57/61)	MAXC (3/60/61)	MAXC (3/57/61)
Mean Returns (ann)	3.2758	2.541	2.3019	1.9985	1.3972	2.4822	2.1455	1.8876	1.7576	1.2485	2.0263	1.7976	1.5514	1.5284	1.0459
Standard Deviation of Returns (ann)	0.4691	0.3777	0.3469	0.326	0.2683	0.3769	0.3329	0.301	0.287	0.209	0.3305	0.2841	0.2496	0.2666	0.1679
Sharpe Ratio	6.9834	6.7269	6.6346	6.1296	5.2068	6.5859	6.4441	6.2702	6.1232	5.9749	6.1315	6.3284	6.2154	5.7328	6.2286
Sharpe Ratio (adjusted)	4.5135 [0.0000]	4.2844 [0.0000]	4.0165 [0.0000]	2.6189 [0.0000]	0.9310 [0.0000]	4.4296 [0.0000]	4.3272 [0.0000]	3.6436 [0.0000]	2.4867 [0.0000]	3.1707 [0.0000]	3.0609 [0.0000]	3.0894 [0.0000]	2.9302 [0.0000]	1.3793 [0.0000]	2.9952 [0.0000]
Sortino ratio (0%)	35.2024	29.4777	30.218	28.4747	9.5471	23.6828	23.0177	18.5978	27.9717	18.6766	18.1028	24.0062	24.9339	42.6476	50.5509
Skewness	2.8493	2.8822	3.1098	4.1081	-3.1141	2.4915	2.5471	2.3138	4.4323	2.3241	2.6131	3.2741	3.5754	6.5016	4.2532
Kurtosis	15.7256	16.3432	18.5422	34.8562	124.6024	14.1002	14.1645	20.399	40.3495	24.4431	23.1079	24.5586	26.8963	75.5197	28.5794
% of positive Returns	0.9281	0.9287	0.9163	0.9237	0.9087	0.9213	0.9275	0.9256	0.9306	0.9175	0.9207	0.9178	0.9244	0.9274	0.9178
Maximum Drawdown	0.1752	0.1034	0.073	0.1397	0.3471	0.1136	0.0838	0.1949	0.0845	0.1932	0.1505	0.1011	0.0975	0.0393	0.0279
Duration of Drawdown	3	2	1	16	1	18	1	11	3	4	1	1	9	10	2
Total Market exposure	0.424	0.423	0.456	0.415	0.442	0.416	0.424	0.441	0.428	0.425	0.422	0.421	0.435	0.419	0.426
Average days per trade	6.98	6.83	7.11	7.11	7.34	6.77	6.77	6.98	7.01	7.23	6.95	7.10	7.25	7.10	7.45
Ratio Winning/Losing trades	0.893	0.916	0.929	0.897	0.849	0.900	0.870	0.884	0.911	0.879	0.890	0.899	0.902	0.884	0.896

Notes: shows the performance of 11,548 trading strategies tested on each of the 15 series after allowing for transaction costs and yields arising out of rolling over the contracts. The total market exposure is the number of days the strategy was active (either long or short) divided by the total number of days in the sample expressed as a ratio. The average number of days per trade is the ratio of the number of days the strategy was active in the market divided by the number of trades. Finally, the ratio of winning to losing trades is a measure of profitable trades divided by the loss making ones. See also the notes in Table 1 for further definitions.

§ MAXC(3/51/61): Moving Average Crossover Consensus Rule, for an explanation regarding trading strategy parameterizations, refer to Section-2.

Table 3: Performance measures for outperforming active strategies: Long-only & Short-only trades.

Criterion	Capesize 4TC					Panamax 4TC					Supramax 4TC				
	1Q	2Q	3Q	1C	2C	1Q	2Q	3Q	1C	2C	1Q	2Q	3Q	1C	2C
Panel – A [Long Only]															
Description of rules	MAXC (3/51/61)	MAXC (3/39/61)	MAXC (3/60/61)	MAXC (3/60/61)	MAXC (3/60/61)	MAXC (3/60/61)	MAXC (3/60/61)	MAXC (3/60/61)	MAXC (3/60/61)	MAXC (3/60/61)	MAXC (3/60/64)	MAXC (3/60/61)	MAXC (3/60/64)	MAXC (3/60/64)	MAXC (3/57/61)
Mean Returns (ann)	1.7377	1.397	1.2816	1.1117	0.7279	1.3519	1.2087	1.0997	0.9451	0.6763	1.2632	1.1091	0.9248	0.8688	0.5829
Standard Deviation of Returns (ann)	0.356	0.2742	0.2615	0.2392	0.1539	0.2727	0.2516	0.217	0.213	0.1463	0.2342	0.2119	0.1805	0.2004	0.1148
Sharpe Ratio (adjusted)	2.1541 [0.0000]	2.4544 [0.0000]	2.2618 [0.0000]	1.2367 [0.0000]	2.1959 [0.0000]	2.3670 [0.0000]	2.4231 [0.0000]	2.1144 [0.0000]	0.8872 [0.0000]	1.6265 [0.0000]	1.8461 [0.0000]	1.5557 [0.0000]	1.5414 [0.0000]	0.4141 [0.0000]	1.3579 [0.0000]
Sortino ratio (0%)	25.3535	31.8358	26.2804	32.3765	19.766	24.1421	18.8135	27.2715	24.2456	11.4978	46.8944	36.5897	25.8994	36.0127	36.3196
Skewness	4.5023	4.2973	4.435	6.4965	4.0224	4.3245	3.9497	4.742	7.7294	1.3842	5.0242	5.3206	5.0904	11.0947	5.7079
Kurtosis	35.0747	30.048	32.3898	75.1969	32.0568	30.454	27.4381	38.5883	110.1046	50.5773	40.0501	49.7305	48.9389	203.7579	58.5337
% of positive Returns	0.96	0.9656	0.9544	0.9575	0.9519	0.9525	0.9625	0.9619	0.9587	0.9556	0.9437	0.9378	0.9437	0.9563	0.9489
Maximum Drawdown	0.1303	0.0789	0.0713	0.0547	0.0752	0.0916	0.0861	0.0953	0.0553	0.1548	0.0601	0.0484	0.062	0.038	0.0309
Duration of Drawdown	1	48	188	42	87	32	43	18	1	2	82	2	69	27	159
Total Market exposure	0.211	0.198	0.218	0.213	0.221	0.214	0.215	0.217	0.216	0.225	0.259	0.256	0.274	0.247	0.260
Average days per trade	12.0301	12.5984	12.8	12.4031	12.9032	12.1212	12.2137	12.8	12.5	12.9032	10.8	10.8	10.6299	10.8	11.7391
Ratio Winning/Losing	0.8968	0.9153	0.9407	0.9008	0.8684	0.904	0.88	0.9091	0.8917	0.8707	0.9115	0.913	0.8772	0.9138	0.8692
Panel – B [Short Only]															
Description of rules	MAXC (3/60/67)	MAXC (3/60/64)	MAXC (3/60/70)	MAXC (3/60/64)	MAXC (3/60/70)	MAXC (3/42/64)	MAXC (3/54/61)	MAXC (3/60/64)	MAXC (3/54/61)	MAXC (3/57/61)	MAXC (3/54/61)	MAXC (3/60/61)	MAXC (3/57/61)	MAXC (3/54/73)	MAXC (3/60/61)
Mean Returns (ann)	1.5934	1.1704	1.0482	0.9077	0.6915	1.1592	0.9633	0.8207	0.8364	0.5976	0.7885	0.7129	0.6571	0.6928	0.49
Standard Deviation of Returns (ann)	0.34	0.2801	0.2482	0.2382	0.228	0.2772	0.2364	0.2138	0.2006	0.159	0.2478	0.2036	0.1853	0.1862	0.1308
Sharpe Ratio (adjusted)	2.4177 [0.0000]	1.8429 [0.0000]	1.6821 [0.0000]	1.2002 [0.0000]	0.3099 [0.0000]	2.1367 [0.0000]	1.9796 [0.0000]	1.5006 [0.0000]	1.6677 [0.0000]	1.4470 [0.0000]	1.1254 [0.0000]	1.3428 [0.0000]	1.2029 [0.0000]	1.1363 [0.0000]	1.4229 [0.0000]
Sortino ratio (0%)	24.9778	18.1599	19.4654	14.8294	4.904	15.5707	14.6013	12.0921	28.9072	18.8062	7.5294	10.9829	13.3882	59.2289	37.2546
Skewness	4.021	4.6626	5.0209	5.4209	-5.4794	3.8351	4.016	4.137	5.4792	5.4652	3.5686	4.6119	5.5647	6.8553	6.368
Kurtosis	25.023	35.1765	41.5669	57.4753	228.8225	26.7788	29.3286	33.6424	42.736	42.7054	48.289	44.2299	53.8722	63.1644	52.5191
% of positive Returns	0.9688	0.9663	0.9644	0.9663	0.9594	0.9663	0.9694	0.9644	0.9762	0.9606	0.9733	0.9741	0.9756	0.9793	0.9674
Maximum Drawdown	0.0981	0.1034	0.073	0.1397	0.3471	0.0967	0.0928	0.0875	0.0459	0.0609	0.1837	0.0961	0.0784	0.0151	0.0253
Duration of Drawdown	1	2	1	16	1	1	39	58	1	91	213	1	1	1	70
Total Market exposure	0.151	0.150	0.169	0.138	0.151	0.133	0.142	0.159	0.138	0.142	0.086	0.092	0.091	0.081	0.094
Average days per trade	16.8421	15.8416	16.3265	16.6667	17.3913	15.8416	15.2381	15.534	16.6667	16.3265	19.8529	20.7692	22.5	23.2759	20.4545
Ratio Winning/Losing	0.9121	0.9167	0.9247	0.8925	0.8315	0.9043	0.9394	0.8511	0.9667	0.8495	0.8824	0.8906	0.9	0.9655	0.8788

Notes: The table presents the performance of 11,548 trading strategies after allowing for transaction costs and yields arising out of rolling over the contracts. Market participants are allowed to take long-only (Panel A) or short-only (Panel B) positions. See also the notes in Table 2 for further definitions.

Table 4: Significance of the outperforming active strategies: WRC and SPA tests

Number of Models: k=11,548, Bootstrap repetitions B=10,000

$$H_0 : \max_{k=1,\ldots,l}\{E(f_k)\} \leq 0$$

Criterion	Capesize 4TC					Panamax 4TC					Supramax 4TC				
	1Q	2Q	3Q	1C	2C	1Q	2Q	3Q	1C	2C	1Q	2Q	3Q	1C	2C
Panel – A [Long-Short Strategies]															
Description of rules	MAXC § (3/51/61)	MAXC (3/60/61)	MAXC (3/60/61)	MAXC (3/51/61)	MAXC (3/60/61)	MAXC (3/60/61)	MAXC (3/60/61)	MAXC (3/60/61)	MAXC (3/60/61)	MAXC (3/57/61)	MAXC (3/60/61)	MAXC (3/60/61)	MAXC (3/57/61)	MAXC (3/60/61)	MAXC (3/57/61)
C†	0.000*	0.000*	0.000*	0.000*	0.000*	0.000*	0.000*	0.000*	0.000*	0.000*	0.001*	0.003*	0.003*	0.005*	0.005*
U	0.000*	0.000*	0.000*	0.000*	0.000*	0.000*	0.000*	0.000*	0.000*	0.000*	0.001*	0.004*	0.003*	0.006*	0.006*
L	0.000*	0.000*	0.000*	0.000*	0.000*	0.000*	0.000*	0.000*	0.000*	0.000*	0.001*	0.003*	0.002*	0.004*	0.004*
Panel – B [Long-Short – With Order Delay]															
Description of rules	MAXC (3/51/61)	MAXC (3/60/61)	MAXC (3/60/61)	MAXC (3/60/61)	MAXC (3/60/61)	MAXC (3/60/61)	MAXC (3/60/61)	MAXC (3/60/61)	MAXC (3/60/61)	MAXC (3/57/61)	MAXC (3/60/61)	MAXC (3/60/61)	MAXC (3/57/61)	MAXC (3/60/61)	MAXC (3/57/61)
C	0.000*	0.000*	0.000*	0.000*	0.000*	0.000*	0.000*	0.000*	0.000*	0.000*	0.002*	0.003*	0.000*	0.003*	0.000*
U	0.000*	0.000*	0.000*	0.000*	0.000*	0.000*	0.000*	0.000*	0.000*	0.000*	0.002*	0.003*	0.000*	0.003*	0.000*
L	0.000*	0.000*	0.000*	0.000*	0.000*	0.000*	0.000*	0.000*	0.000*	0.000*	0.002*	0.003*	0.000*	0.003*	0.000*
Panel – C [Long-Short – Without the market crash period]															
Description of rules	MAXC (3/60/61)	MAXC (3/60/61)	MAXC (3/60/70)	MAXC (3/60/61)	MAXC (3/60/61)	MAXC (3/60/61)	MAXC (3/54/61)	MAXC (3/54/61)	MAXC (3/60/61)	MAXC (3/57/61)	MAXC (3/60/61)	MAXC (3/60/64)	MAXC (3/57/61)	MAXC (3/54/70)	MAXC (3/57/61)
C	0.000*	0.000*	0.000*	0.000*	0.003*	0.000*	0.000*	0.000*	0.000*	0.000*	0.002*	0.002*	0.002*	0.005*	0.017*
U	0.000*	0.000*	0.000*	0.000*	0.003*	0.000*	0.000*	0.000*	0.000*	0.000*	0.002*	0.003*	0.002*	0.006*	0.017*
L	0.000*	0.000*	0.000*	0.000*	0.003*	0.000*	0.000*	0.000*	0.000*	0.000*	0.002*	0.002*	0.002*	0.004*	0.011*
Panel – D [Long-Only]															
Description of rules	MAXC (3/51/61)	MAXC (3/39/61)	MAXC (3/60/61)	MAXC (3/60/61)	MAXC (3/60/61)	MAXC (3/60/61)	MAXC (3/60/61)	MAXC (3/60/61)	MAXC (3/60/61)	MAXC (3/60/61)	MAXC (3/60/64)	MAXC (3/60/61)	MAXC (3/60/64)	MAXC (3/60/64)	MAXC (3/57/61)
C	0.000*	0.000*	0.000*	0.000*	0.003*	0.000*	0.000*	0.002*	0.000*	0.002*	0.000*	0.003*	0.003*	0.014*	0.019*
U	0.000*	0.000*	0.000*	0.000*	0.003*	0.000*	0.000*	0.002*	0.000*	0.002*	0.000*	0.003*	0.002*	0.017*	0.028*
L	0.000*	0.000*	0.000*	0.000*	0.003*	0.000*	0.000*	0.002*	0.000*	0.002*	0.000*	0.003*	0.000*	0.007*	0.014*
Panel – E [Short-Only]															
Description of rules	MAXC (3/60/67)	MAXC (3/60/64)	MAXC (3/60/70)	MAXC (3/60/64)	MAXC (3/60/70)	MAXC (3/42/64)	MAXC (3/54/61)	MAXC (3/60/64)	MAXC (3/54/61)	MAXC (3/57/61)	MAXC (3/54/61)	MAXC (3/60/61)	MAXC (3/57/61)	MAXC (3/54/73)	MAXC (3/60/61)
C	0.066	0.104	0.059	0.118	0.104	0.083	0.152	0.182	0.132	0.118	0.397	0.389	0.355	0.265	0.239
U	0.067	0.104	0.059	0.118	0.104	0.084	0.157	0.184	0.132	0.118	0.429	0.440	0.389	0.267	0.239
L	0.037*	0.067	0.038*	0.080	0.073	0.049*	0.071	0.093	0.064	0.075	0.226	0.248	0.210	0.162	0.132

† SPA test (p-values) where C (Consistent), L (Lower bound) and U (Upper bound) are defined in Section-4

§ MAXC(3/51/61): Moving Average Crossover Consensus Rule, for an explanation regarding trading strategy parameterizations, refer to Section-2.

* Denotes significance at 5%.

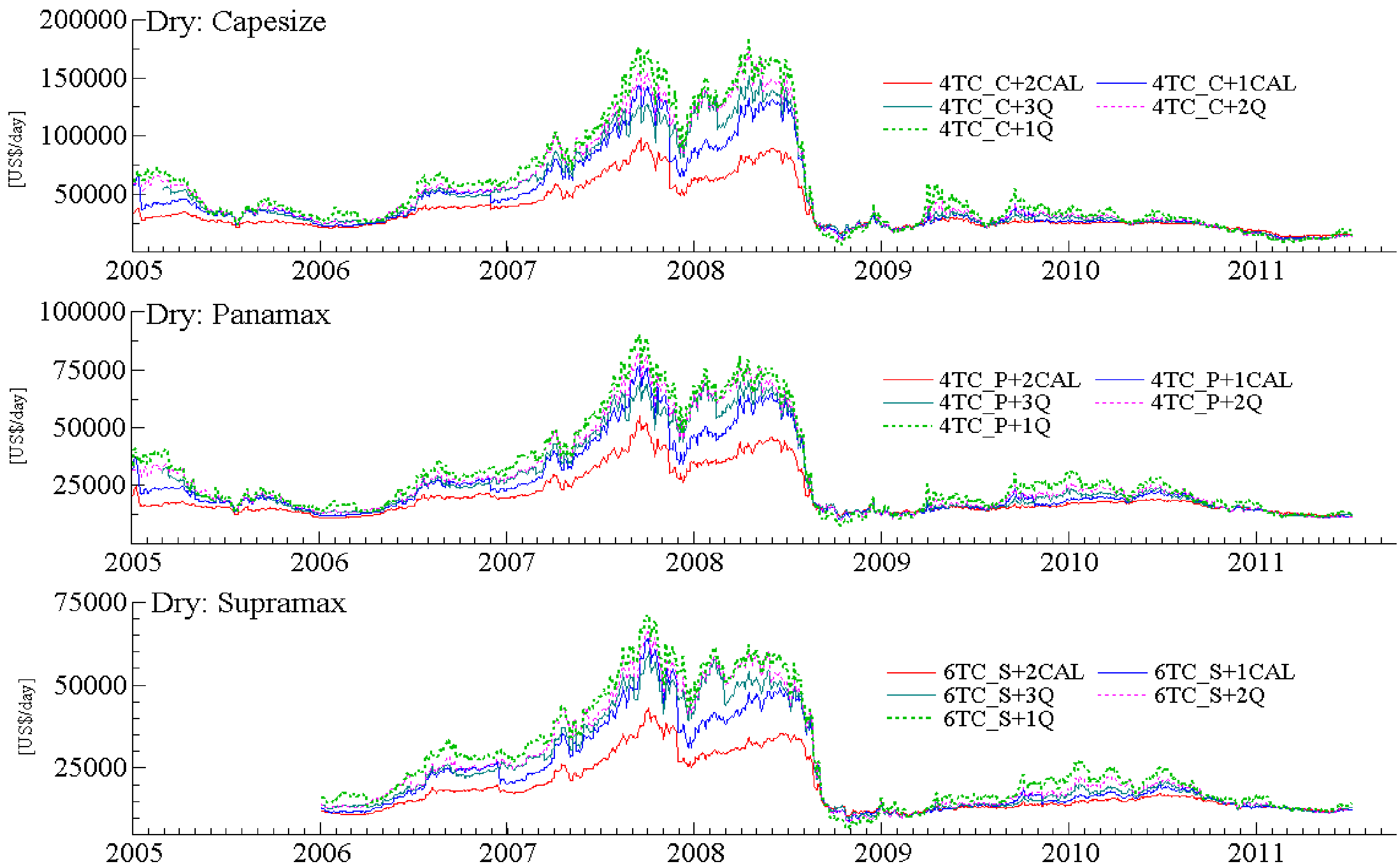
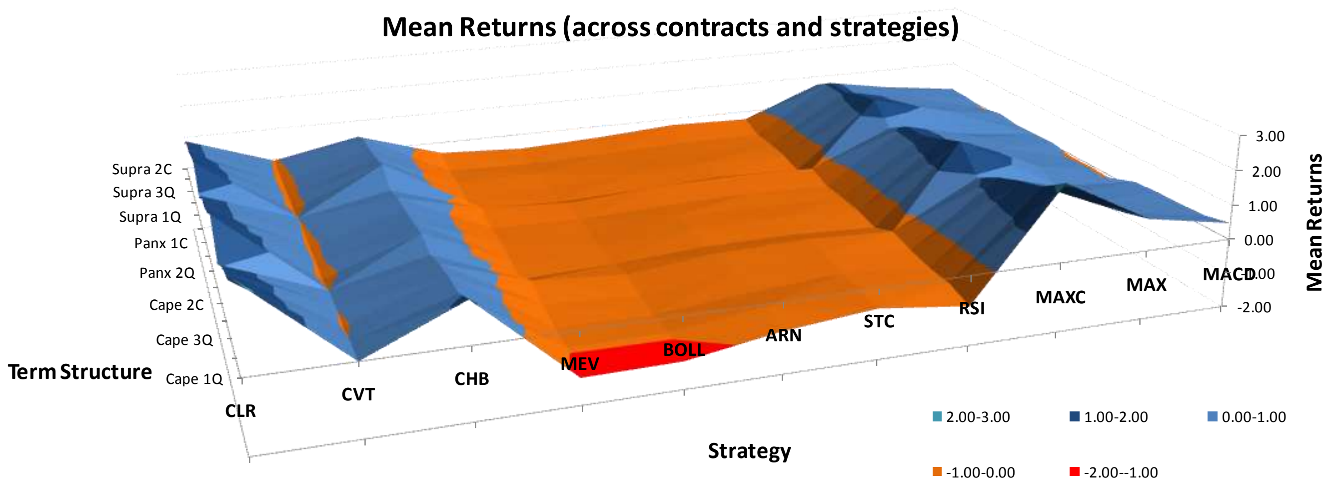


Figure 1: Evolution of the Baltic Capesize 4TC, Panamax 4TC and Supramax 6TC basket routes across 5 maturities (1Q, 2Q, 3Q, 1C and 2C) from January 2005 (except for Supramax 6TC which starts from January 2006) to September 2011. The Y-axis measures freight rates denominated in US$/day for all three vessel sizes.



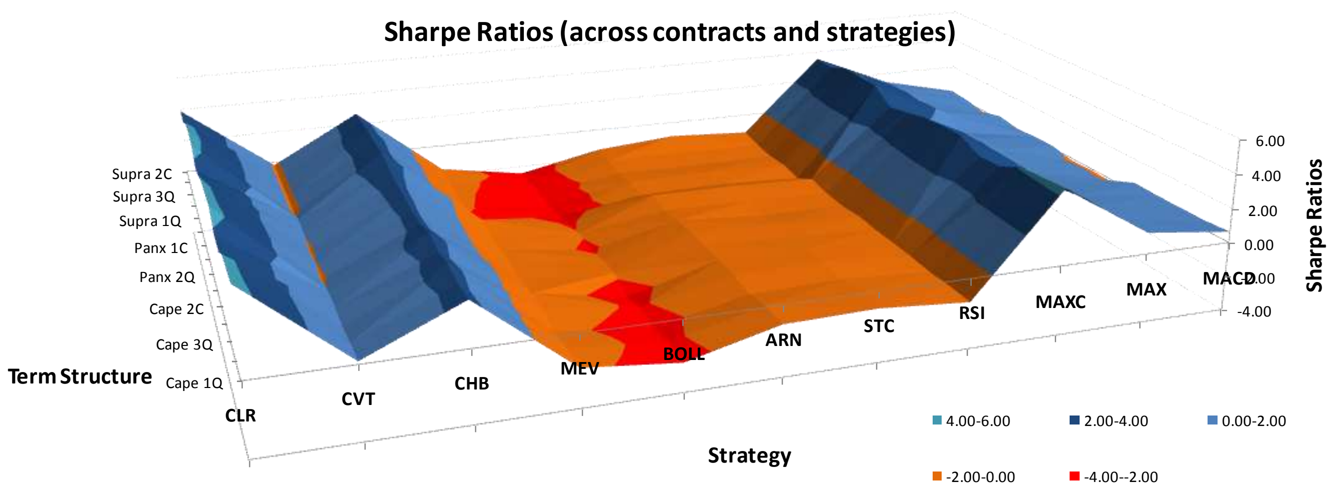


Figure 2: Surface plots of mean out-performance of each strategy based on average annualized returns as well as average annualized Sharpe ratios.