NORWEGIAN SCHOOL OF ECONOMICS AND BUSINESS ADMINISTRATION (NHH)
Bergen, spring 2010

Master Profile: Financial Economics
Advisors: Professor Siri Pettersen Strandenes and Associate Professor Jonas Andersson

The Price Discovery Properties of Clean Tanker Freight Futures
- Unbiasedness, Causality and Forecasting

Audun Houmb Sjøli

Abstract

The scope of this thesis is to examine the price discovery properties of clean tanker freight futures. This is conducted by testing the unbiasedness hypothesis, the lead-lag relationship between freight futures and spot rates and the forecasting properties of freight futures with regards to the underlying spot rates. The research focuses on the most liquid clean tanker freight futures, which are those written on the routes TC2, TC4 and TC5. The results indicate that unbiasedness depends on the route in question and time to maturity. For a one-month horizon of TC2 and one-, two- and three-month horizons of TC5, the unbiasedness hypothesis is found to hold. Unbiasedness is also indicated for the two- and three-month horizons of TC2, but due to weak evidence no conclusions are drawn. For TC4 the unbiasedness hypothesis is rejected. The results from testing the lead-lag relationship indicate that futures prices lead spot rates for all the routes, but the relationship is found to be bi-directional for TC4. When investigating the forecasting performance of end-of-month freight futures it is found that univariate models generally are outperformed by a random walk, indicating that forecasts should not be based on historic spot prices alone. The multivariate models confirm this finding as they generally produce more accurate forecasts than their univariate cousins. Multivariate time-series models were generally found able to outperform forecasts indicated by outright futures prices for one- and two-month horizons, but for a three-month horizon the futures performed as well as or better than the multivariate models. These results imply that the investigated freight futures contain valuable information about future spot rates. Problems regarding the stationarity of the series were experienced throughout the thesis. Because of this it is recommended that the tests performed in this thesis are repeated in a few years when more data is available.

Acknowledgements

I would like to thank my advisors, Professor Siri Pettersen Strandenes and Associate Professor Jonas Andersson, for contributing to this thesis by sharing their extensive knowledge within shipping and econometrics, respectively. At Imarex I would like to thank Nils Arnesen and the tanker desk, Lene Refvik and Erlend Engelstad for providing data and insight on freight futures contracts and the freight futures market. At NOS I would like to thank Anette Bergvoll Nilsen for clarifying historic changes in the clearing procedures and for helping to make sure that the data used in this thesis is accurate. I would also like to thank Eirik Maanum and the clean tanker desk at Inge Steensland for their constructive inputs regarding the physical clean tanker market. At the Baltic Exchange I would like to thank William Lyth and Michael Ackerman for providing spot prices for the route TC2, and at Platts William Bathurst for answering questions regarding the spot prices of TC4 and TC5. I also appreciate the great service the staff at NHH library has provided. I am also very thankful to Tor Svelland and the commodity desk at Pareto Securities for sharing their broad experience on freight futures, and for making it possible for me to continue to explore this fascinating market.

I would also like to thank Åge for providing the best atmosphere imaginable for writing this thesis and Henning Bjørneng Sagen for excellent technical support.

Finally, I would like to express my gratitude to my family for supporting me through my whole education.

This thesis is dedicated to Line Sofie.

Table of Contents

1. General Notes

1.1 Introduction

The shipping industry is of great global importance as it is responsible for the carriage of approximately 90 percent of world trade (International Maritime Organization, 2006). The main reason for this is that the production of commodities and goods often does not take place in the same region as the consumption, and the most affordable way of transportation is by sea. This is also the case for the refined petroleum industry, where the output produced by the refineries does not mirror the regional consumption. Trade flow patterns have therefore evolved over time, making it possible to write futures contracts on commonly traded routes.

Futures are generally thought to have two economic functions. These are risk management through hedging and price discovery (Black, 1976). The shipping industry involves substantial business risk (Stopford, 1997). To mitigate this risk the BIFFEX freight futures market was developed in the 1980s (Kavussanos and Nomikos, 2000). The purpose of BIFFEX was to provide a hedging tool for those exposed to dry bulk freight rates. The BIFFEX market was however phased out in 2002, but by this time other markets such as Imarex had emerged offering freight futures on specific routes which reflected the trade flow patterns of both dry bulk and tanker shipping[1].

The object of this thesis is to investigate the price discovery properties of clean tanker freight futures. If price discovery properties exist these may be used to guide physical supply and demand decisions in ways that may contribute to a more efficient allocation of economic resources (Kavussanos and Nomikos, 1999). As such, the research conducted in this thesis should be of interest to everyone exposed to clean tanker freight rates. Additionally, the thesis is thought to contribute to the existing literature as the price discovery properties of clean tanker freight futures have not been investigated in earlier research to the author's knowledge.

[1] A complementary explanation is provided in footnote 11.

1.2 Objective

The objective of this thesis is to investigate the price discovery properties of clean tanker freight futures. These price discovery properties are desirable in an economic perspective because they enable the futures market to be used to guide physical supply and demand decisions in ways that contribute to a more efficient allocation of economic resources (Kavussanos and Nomikos, 1999), a function best performed if the unbiasedness hypothesis holds. The unbiasedness hypothesis is therefore tested. The lead-lag relationship between clean tanker freight futures and spot rates is also investigated as this may provide further insight to the interaction between the markets, and thereby the price discovery process. An appealing application of this analysis is that the futures, if found to lead the spot rates, may be used as a sentiment indicator for physical shipping. Finally, the forecasting properties of clean tanker freight futures, with regards to the underlying spot rates, will be investigated. This is done to reveal whether the futures prices contain information which is useful when building forecasting models, and to test the performance of forecasts implied by the futures themselves compared to those which can be obtained using time-series models.

1.3 Involved Parties

The parties which are relevant for the thesis are presented in this part. These are Imarex, NOS, The Baltic Exchange, Platts and The Worldscale Association.

1.3.1 IMAREX

Imarex - The International Maritime Exchange – opened in 2001 and is the only regulated marketplace offering trading of freight derivatives with instant clearing. The underlying indices for the freight derivatives are provided by the Baltic Exchange and Platts. Most of the tanker freight derivatives and their underlying indices are quoted using the Worldscale system. The exchange is regulated by The Financial Supervisory Authority of Norway (Finanstilsynet). Clearing and settlement of the Imarex derivatives is performed through NOS (Norsk Oppgjørssental).

When referring to Imarex in this thesis I refer to the Exchange. The reason for pointing this out is that Imarex also is a group of companies which facilitate both trading of salmon and energy derivatives, as well as research, and clearing and settlement of derivatives. The group is publicly listed at the Oslo Stock Exchange under the name IMAREX ASA.

1.3.2 NOS

NOS – NOS Clearing ASA - was established in 1987 and is a central clearing house for freight, seafood, power and UK gas derivatives. It is licensed through the Norwegian Ministry of Finance and is regulated by the Financial Supervisory Authority of Norway. NOS is the major clearing central for Imarex freight derivatives.

1.3.3 The Baltic Exchange

The Baltic Exchange is a well-renowned source of maritime market information. It is formed as a membership organisation and the members are responsible for a large proportion of all dry cargo and tanker fixtures. The first Baltic freight index was launched in 1985 and this has been supplemented by more indices up to today. The quotes of the majority of these indices are formed by having a panel of shipbrokers providing daily freight rate assessments. In the context of this thesis, the TC2 freight futures uses the Baltic index as the underlying.

1.3.4 Platts

Platts is a company which collects and publishes information relevant for the energy, metal, petro-chemical and shipping markets. The company was founded in 1909 and acquisitioned by McGraw-Hill in 1953. In the context of this thesis Platts is the publisher of the spot freight rate indices of which the Imarex TC4 and TC5 freight futures are settled against.

1.3.5 The Worldscale Association

The Worldscale Association consists of the two non-profit making organisations Worldscale Association (London) Limited and Worldscale Association (NYC) Inc. These two companies are responsible for publishing the Worldwide Tanker Nominal Freight Scale, also known as Worldscale, which originated under World War Two. The Worldscale system is used to express freight rates in tanker shipping. In the context of this thesis, all the involved routes are quoted using the Worldscale system[2].

[2] A further explanation of the Worldscale system is provided in chapter 2.3.

1.4 Outline

In **chapter one**, a brief introduction to this thesis is provided. The objective, which is to investigate the price discovery properties of clean tanker freight futures, is presented. The involved parties are also introduced.

Chapter two contains an introduction to freight futures. The basics of forward freight agreements (FFAs) and freight futures are explained, and the liquidity of the freight futures market is discussed. In addition a presentation of the price quotation system for tanker shipping, Worldscale, and its implications in the context of this thesis are elaborated on.

In **chapter three**, the link between freight futures and spot rates is introduced. This part contains an explanation of how the prices of freight futures are formed, which is the backbone of the price discovery properties of freight futures.

Chapter four contains a review of existing literature. As the price discovery properties of clean tanker freight futures were found to be a relatively unexplored area, this literature is on dry bulk freight futures and FFAs.

Chapter five contains an explanation of how to test the unbiasedness hypothesis for stationary and non-stationary data. In line with this, the OLS and Johansen methodology is presented and employed, before the results from testing the unbiasedness hypothesis are provided.

In **chapter six** the lead-lag relationship between futures prices and spot rates is investigated. To arrive upon correct model specifications the Johansen's methodology is employed. VECMs and a VAR in levels are then used to conduct Granger causality tests and impulse response analyses.

In **chapter seven** the forecasting abilities of freight futures are investigated by comparing the forecasts implied by the futures themselves to those produced by various uni- and bi-variate time-series models.

In **chapter eight** the main findings of this thesis are summarised and the conclusion is presented.

2. An introduction to freight futures

Before examining the price discovery properties of freight futures it is crucial to fully understand the basics of the futures. I therefore provide a general introduction to freight futures in this part, beginning with the closely related Forward Freight Agreements. I also present the price quotation system, Worldscale, which is used for both the freight futures and their underlying spot indices, and elaborate on the implications of this system in the context of this thesis.

2.1 Forward Freight Agreements (FFAs)

A Forward Freight Agreement (FFA) is a financial contract, where a buyer and a seller agree upon a freight rate for some specified route over a future time period. The contract does not involve any actual delivery of freight, and it is settled financially at maturity. There are essentially two types of FFAs: Over-The-Counter (OTC) FFAs and “Hybrid” FFAs.

The OTC FFAs have the same characteristics as normal forwards. As such they might be customized to fit the specific needs of the parties and do not involve margins or a clearing house. Consequently, the contracts can be difficult to close out and a credit risk arises because the counterparty may default. This makes the participants of the market rather careful when choosing counterparties, so companies without a “name” in the world of shipping or finance might experience problems when attempting to take an FFA position. Despite the fact that OTC FFAs may be customized they often have the same specifications as freight futures. The reason for this might be that the specifications of the freight futures serve the market well and that using these specifications lead to better liquidity in the FFA market. These markets therefore lend a hand to each other when it comes to correcting prices and liquidity. OTC FFAs are currently offered by at least 20 different brokerage houses (Baltic Exchange).

"Hybrid" FFAs have the same properties as OTC FFAs, except from being cleared through a clearing house for a fee[3]. This enables them to maintain the flexibility of the FFAs and at the same time to effectively remove the counterparty risk. These contracts also help correct the prices of the freight futures market and thereby add liquidity.

2.2 Freight Futures

Kavussanos and Visvikis (2006a) define a futures market as a market which "…must trade a uniform, standardized contract, in standard quantities, for delivery on specified dates in the future, with good price availability (transparency of pricing)". The freight futures market satisfies these conditions. The marketplace for freight futures consists of several exchanges where financial contracts on freight, which do not include physical delivery, are traded. Unlike FFAs, freight futures are marked-to-market (settled) daily by a clearing house, which also acts as the counterparty to each contract. This means that each contract can be closed out at any time as long as there is a liquid market and that the counterparty risk is with the clearing house, reducing the default risk substantially compared to FFAs. Having a clearing house acting as counterparty to each contract also enables all traders to stay anonymous. Because the default risk is virtually non-existing and that traders may stay anonymous, using a clearing house is thought to increase liquidity.

Because of the mark-to-market procedure, it is required that a company which trades freight futures is member of the exchange and clearing house where the freight futures are traded and cleared. Upon acceptance of membership the company has to deposit cash to a margin account with the clearing house. Taking positions may also require additional cash to be deposited, but to keep these amounts as low as possible, all active positions are netted by the clearing house. If the positions of a trader make the margin account drop below a level called the maintenance margin, the clearing house will make a margin call asking for cash to be deposited to the account. If the

[3] The clearing procedure is the same one presented in chapter 2.2.

trader does not meet the clearing house's request, the positions of the trader will be closed out by the clearing house. This procedure is also believed to increase liquidity, as it enables companies which participants in the FFA market do not recognize as credible counterparties to take positions. A third element which increases liquidity is the presence of general clearing members (GCMs) which are allowed to trade on behalf of their own clients as they themselves are financially liable with regards to the exchange and the clearing house.

There are several marketplaces where freight futures are traded, but Imarex is the only regulated marketplace that offers voice and electronic trading with instant clearing[4]. Because prices on equivalent contracts which differ between the marketplaces presents arbitrage opportunities, the existence of several marketplaces should not be an obstacle with regards to Imarex providing efficient prices. As such, the freight futures prices used in this thesis should be representative for the freight futures market as a whole[5].

Imarex offer shipping related derivatives for both the tanker and dry bulk markets. An exhaustive list of their products can be found at their websites. With regards to the scope of this thesis, the clean tanker freight futures are the most relevant[6]. Imarex is currently offering such futures for several routes, with the contracts written on TC2, TC4 and TC5 being the most liquid[7] (Imarex). Liquidity is important in futures markets because it facilitates correct pricing (Thompson, Garcia and Wildman, 1996). A presentation of Imarex and NOS trading volumes can be found in figure 2.1 below.

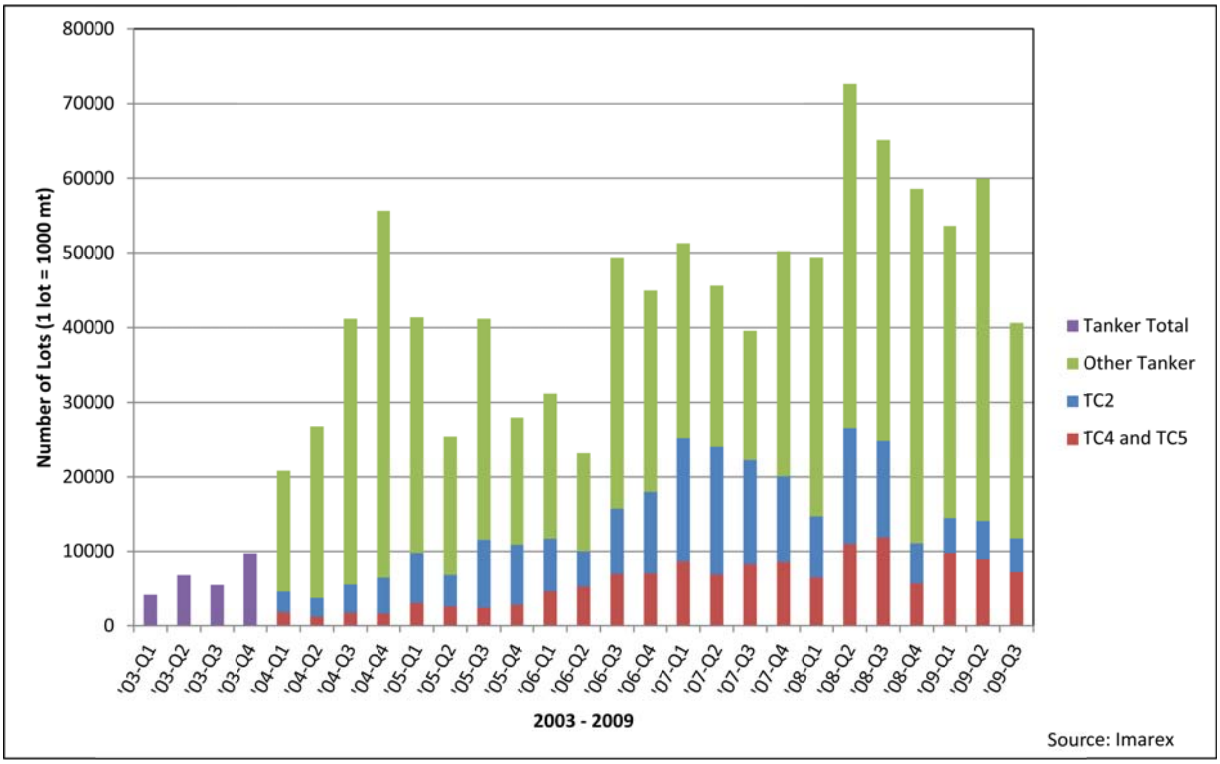
[4] Other companies currently offering freight futures with voice and electronic trading and access to clearing include GFI, ICAP and SSY.

[5] Imarex experienced a market share of approximately 40-45% in the tanker market and 10-15% in the dry bulk market from 2004 to 2009 (Imarex).

[6] Clean here refers to refined oil products. The names of the contracts therefore include the abbreviation TC, which is short for Tanker Clean.

[7] A brief description of these contracts can be found in table 5.1.

Figure 2.1: Tanker FFAs and Freight Futures traded via Imarex and NOS



The full height of the columns in this figure represents the total volume for tanker FFAs and freight futures. The material which the figure is based upon has been provided by Imarex. For 2003 only totals are available. From 2004 onwards the volume is divided into TC2, TC4 and TC5 and other tanker. The reason for representing TC4 and TC5 together is that separate data for these routes only exists for 2009. For this year TC4 amounted to approximately 29 percent of the total. TC4 is also reported by brokers to be the route with the lowest liquidity for the whole period in this context. The category 'Other Tanker' mainly consists of tanker dirty, but other tanker clean routes are also included in this category. For the total period TC2, TC4 and TC5 amounted to approximately 32 percent of the total tanker lots traded. Most of the liquidity is found in the front of the forward curves (Imarex).

Kavussanos and Visvikis (2004) point out that the prices of OTC FFAs and freight futures contracts might differ because of different underlying assets, transaction costs, default risk and divisibility. Since OTC FFAs and freight futures often have the same underlying assets, transaction costs and divisibility properties the prices might however differ solely because of factors connected with the default risk. Following the same argument it is clear that the prices of "hybrid" FFAs and the corresponding freight futures might differ slightly. As the markets for FFAs and freight futures are interconnected, it can be difficult to distinguish between these instruments. This, however, does not impose any significant problems with regards to this thesis.

2.3 Worldscale

New Worldwide Tanker Nominal Freight Scale, "Worldscale", is a price quotation system commonly used to calculate the cost of seaborne carriage of oil in bulk. The system is based on nominal rates which are published by the Worldscale Association. There are approximately 320,000 such rates and these cover virtually all possible voyages. The published rates are all based on the same principle, which is that the net daily revenue (or the time charter equivalent) from a round voyage of a standard vessel should be identical for all similar voyages after allowing for voyage specific costs, such as bunker costs, ports costs and canal dues. Because the published rates provided by Worldscale only serve as a basis for the freely negotiated actual rates, it is important to stress the word *nominal*. The actual rates are quoted as a percentage of the published rates and the price of freight is thereby represented by points of scale, known as Worldscale points. A quote of 100 Worldscale points is usually represented as Worldscale (WS) 100, which corresponds to the nominal rate itself. This is often referred to as Worldscale flat, or the flat rate. A quote of WS 200 corresponds to 200 percent of the published rate, while WS 50 corresponds to 50 percent of the published rate. Using this system, the price of freight is calculated in the following way;

$$Price\ of\ freight\ (USD) = \frac{Worldscale\ points}{100} \times Flat\ Rate\ (USD/MT) \times Cargo\ Size\ (MT)$$

where the first term expresses the market level of freight in terms of a percentage of the current nominal freight rate (flat rate), which is based on USD per metric ton (MT) units.

The flat rates are published by Worldscale in November every year and are effective for a one-year period from January 1st to December 31st the subsequent year. The calculations of all flat-rates for a given year are based on an average worldwide bunker price for fuel oil during the period October 1st two years before to September 30th the year before, port cost, canal dues and foreign exchange rates (Worldscale, 2008). If the flat rate of a voyage is increased (decreased), the number of Worldscale points will have to decrease (increase) for the net daily revenue of the ship owner to stay unchanged, all other factors held constant. When the flat rates are changed, the Worldscale rates which are agreed upon in the open market therefore tend to shift, and the magnitude of these shifts depends on to what extent the flat rates are changed, all other factors held constant.

The Worldscale system works well when used in the spot market, because the new flat rates are released before the fixing window where the new rates that are to be used appear. This means that the spot market always knows the flat rates of which physical freight is agreed upon, and thereby the actual price of a voyage. The system is also convenient for some types of voyages because a charterer does not always know where a cargo is to be loaded or discharged when agreeing on freight. The system thereby reduces the number of rates which are necessary to be agreed upon to one. This is in line with the principle that the daily net revenue shall be the same for all voyages. It should be mentioned that for voyages which are more of an arbitrary operation, a lump sum system is used, but this is not relevant for this thesis.

The changing flat rates do however represent a problem in the freight futures market. This occurs when trading freight futures of which the prices of the factors involved in the flat rate calculations are unknown. Market participants who trade such contracts are exposed to two risk factors: the underlying spot rates and the relevant flat rate. The futures market will of course try to approximate the future flat rates, and this approximation is likely to improve as maturity approaches because the underlying factors of which the flat rates are based upon are then revealed. In line with this, long term forecasts based on freight futures may be less precise than for example for the dry bulk market where a lump sum system is used.

In the context of this thesis the implications of the changing of the flat rates are the following: Firstly, testing the unbiasedness hypothesis and investigating the causality relationship between spot and futures can be performed without adjusting for the changing flat rates. This is because the futures prices series are then lagged to match the underlying spot series. Not adjusting for the changing flat rates does however imply allowing for shifts in the series which are well known. Because the underlying price of freight does not shift - all other factors held constant - this creates a shift which is not created by economic factors, but rather how these are measured. This might lead econometric tests to not reflect the properties of the real underlying prices series. Secondly, building forecasting models without adjusting for the changing flat rates is not reasonable. This is because forecasts will be based on historic observations, and the well known flat rate changes will thereby lead to less accurate forecasts than achieved when adjusting for the changing flat rates.

The problem which the changing flat rates induce can be dealt with in at least two ways: One is to employ a dummy variable for each year's flat rate and the other is to rebase the time series to the flat rate level of a given year. As the underlying flat rates are known at the time of the shift, these methods should be equivalent. To keep the econometric models used in this thesis as simple as possible, I have chosen to use the latter approach. All spot and futures time series have therefore been rebased to the flat rate level of 2009. The reason for rebasing all series instead of those related to forecasting only, is to avoid confusion throughout the thesis. An implication of this rebasing is that the futures market is thought to be able to estimate the correct underlying flat rates of January each year for the three-month futures price series. This is because the actual flat rates will be used before they are released for these series. The reason for doing this is that the alternative of using the market estimates of the future flat rates is very difficult, as these estimates cannot be observed. Assuming that the futures market is able to correctly estimate the future flat rates might induce a small bias in the results when series of futures prices which are collected three months prior to maturity are used. The solution is however thought to be better than allowing for seemingly unanticipated shifts which in reality are both anticipated and inevitable.

3. Freight futures and future spot rates

In this part, theory on how the prices of freight futures are formed is presented. This is the backbone of the price discovery properties of freight futures, and thereby also this thesis. The term *price discovery property is* also elaborated on.

3.1 How the prices of freight futures are formed

Futures can be divided into two categories: those written on storable commodities and those on non-storable commodities. This division is important because futures on non-storable commodities cannot be priced in the same way as futures on storable commodities. The reason for this is that futures on storable commodities often are priced using arbitrage arguments which are based on the possibility of storing the underlying commodity. Such arbitrage is not possible for non-storable commodities, and the prices of this type of futures are therefore solely based on expectations with regards to future spot prices[8].

Because seaborne freight is a service which is produced while carried out, and capacity which is not utilized cannot be stored, shipping is a non-storable commodity. Arbitrage between the spot and futures market is therefore not possible. Hence, freight futures cannot be priced using the cost-of-carry relationship which involves storage of the underlying commodity (Kavussanos and Nomikos, 2003). The prices of freight futures must therefore reflect the aggregated expectations of the market with regards to the underlying spot rates at the time of settlement (Cullinane, 1992). This relationship can be expressed in the following way;

$$F_{t,T} = E_t(S_T)$$

[8] An introduction to the pricing of forwards and futures on storable and non-storable commodities can be found in e.g. MacDonald (2006) or Hull (2009).

where $F_{t,T}$ is the price of a future with settlement at time t = T formed at time t. $E_t(S_T)$ is the expected spot rate of the underlying at time t=T, formed at time t. This pricing relationship is the backbone of the unbiasedness hypothesis, which implies that the futures prices are unbiased estimators of the underlying future spot prices. Assuming that the market is efficient and rational expectations this means that all available information must be reflected in the price of a future, and that this price will only be affected by new information which is typically referred to as news. In financial markets, news are generally thought to be random and available to all markets participants simultaneously. This means that it should be impossible to consistently outperform the market, except via luck. News are however not available simultaneously in the freight futures market and it might therefore be possible to consistently outperform this market. The reason for this originates from the fact that participants of the freight futures market may trade in the spot market of which the underlying indices are based on. Deals made in the spot market are not meant to be made publicly available and news therefore does not reach all the participants of the futures market simultaneously. A trader with first hand information on spot deals might therefore make a profit by acting on information not available to the whole market. Brokers do however report that attempts of trying to influence the futures market by making spot deals at rates which do not reflect the economic factors of the spot market are usually not successful. A further investigation of the trading possibilities connected to first hand information and fixing ships at artificial levels would be very interesting, but unfortunately the necessary data are not available. Assuming that that the latter is not possible, this phenomenon will only induce timing related trading opportunities for well informed market participants. It will therefore not prevent freight futures from being unbiased estimates of the underlying spot rates.

The unbiasedness hypothesis might, however, not hold in reality due to a mismatch between the supply and demand sides. This will attract speculators which are willing to balance the market if offered a risk premium. If this is the case, the futures prices will be biased estimates of the underlying future spot prices.

20th century economists like Keynes (1930) and Hicks (1939)[9] argued that biased futures prices might be a consequence of an overweight of producers which are selling futures. These are willing to lower the futures price below the expected spot price of the underlying commodity to hedge their physical exposure. If the market is right in its expectations of the future price rising to meet the future spot price, speculators which take a long position (buy) the future will earn the offered risk premium.

On the other hand, if the market is net long (more buyers) futures, the futures price will exceed the expected underlying spot price. The market will then expect the futures price to fall to meet the future spot price. Speculators which short (sell) futures will make a profit on this phenomenon if the expectation materializes. An empirical example of this can be found in Adam and Fernando (2006), where it is showed that producers of gold have earned a significant risk premium by shorting gold futures. A further investigation on the importance of hedging forces in futures markets for non-storable commodities can be found in Gray and Tomek (1970).

Because futures prices might include a risk premium it is not possible conduct an isolated test of whether the market agents have rational expectations. Test of the unbiasedness hypothesis is therefore a joint test of no risk premium and rationality of expectations (Fama, 1991). These two cannot be separated without making further assumptions regarding how expectations are formed and the risk preferences of the market agents. Because such assumptions probably will cause simplifications which do not account for the complexity of the market, I will perform the joint test when testing the unbiasedness hypothesis.

[9] As presented in Bodie, Kane and Marcus (2008).

3.2 The price discovery function of freight futures

Futures have two economic functions. These are price discovery and risk management through hedging[10] (Black, 1976). Freight futures have the very same functions in the shipping markets (Kavussanos and Visvikis, 2006a).

From the way prices of freight futures are formed it is clear that they reveal information on the expectations of the market participants with regards to future spot rates. The prices of futures may thereby contain more information about future spot rates than the current and past spot prices alone. Freight futures may therefore have price discovery properties. These price discovery properties are desirable in an economic perspective because they enable the futures market to be used to guide physical supply and demand decisions in ways that contribute to a more efficient allocation of economic resources (Kavussanos and Nomikos, 1999), a function best performed if the unbiasedness hypothesis holds. Then, anyone interested in the spot prices of the future can use freight futures prices as unbiased estimates of future spot prices.

[10] Risk management refers to hedgers using futures to control their price risk in the spot market. More on risk management in the shipping industry using futures (including freight futures) may be found in Kavussanos and Visvikis (2006a).

4. A Review of existing literature

Existing literature on the three main topics of this thesis is presented here. This literature is mainly on the BIFFEX future[11] and dry bulk FFAs. As the focus of this thesis is on clean tanker futures, it is thought to contribute to the existing literature by examining a market segment which, to the author's knowledge, has not been researched until now.

4.1 The unbiasedness hypothesis

Chang and Chang (1996) use OLS regression analyses to test whether the BIFFEX futures can be used to predict the BFI. They find that the futures may be used as an unbiased estimate of future spot rates up to a one-month horizon.

Kavussanos and Nomikos (1999) employ cointegration techniques in the form of Johansen (1988) to examine whether the unbiasedness hypothesis holds for the BIFFEX contract. This framework enables them to incorporate long run equilibrium information into the unbiasedness test. They find that futures prices one and two months before maturity are unbiased forecasts of the underlying spot prices. Futures prices three months before maturity are however found to be biased.

Haigh (2000) uses cointegration techniques in the form of Johansen (1988) to test the unbiasedness hypothesis with regards to the BIFFEX futures and spot prices. He finds evidence of unbiasedness for current-, one- and two-month, as well as quarterly contract horizons. He also

[11] BIFFEX (Baltic International Financial Futures Exchange) was a future launched on the BFI (Baltic Freight Index) in May 1985. The BFI is a daily published index based on a basket of dry bulk spot voyage routes and time charter routes. The composition of this basket was changed during the life of BIFFEX (see Kavussanos and Nomikos, 2003) to reflect the hedging needs of the dry bulk market and thereby attract trading activity. Haigh (2000) and Kavussanos and Nomikos (2003) find that the changes of the index composition also helped improve the price discovery function of the futures. The fall of the BIFFEX trading volume is therefore argued not to be due to lack of price efficiency, but rather the lack of hedging efficiency and the growth of the FFA market. Haigh argues that this happened because FFAs can be tailored to fit the individual needs of each market participant, which mended the cross-hedging problem. The BIFFEX contract ceased trading in April 2002. A figure representing the yearly trading volumes of the BIFFEX contract may be found in Kavussanos and Visvikis (2006b).

finds that for these contract horizons, the future price is the one to adjust if the prices fall out of their long run equilibrium. Haigh's results contradict those of Kavussanos and Nomikos (1999), who find that the quarterly BIFFEX contract is long-term biased. Haigh suggests that this is due to the fact that they used a small number of observations when testing the unbiasedness hypothesis for longer horizons.

Kavussanos, Visvikis and Menachof (2004) test the unbiasedness of OTC FFAs using cointegration techniques proposed by Johansen (1988). The routes investigated are the same four dry bulk routes as used in Kavussanos and Visvikis (2004). The results of the tests are that the FFA prices one- and two months before maturity are unbiased estimators of the underlying spot prices. For the three month contracts, the FFAs of the two Pacific routes are showed to be unbiased, while the two Atlantic routes are biased. This indicates that unbiasedness depends on the characteristics of the route investigated and time to maturity.

4.2 The lead-lag relationship between futures prices and spot rates

Kavussanos and Nomikos (2003) employ Johansen's framework (1988) on daily for BIFFEX futures and the underlying BFI to perform causality tests and impulse response analyses. They find that the futures prices help discover future spot prices, and that futures prices discover new information more rapidly than the current spot prices. This is reported to be in line with the lower costs associated with trading futures than trading in the spot market.

Kavussanos and Visvikis (2004) use daily data to examine the lead-lag relationship between OTC FFAs and spot returns. They focus on the same four dry bulk routes as investigated in Kavussanos, Visvikis and Menachof (2004). Using Johansen's framework (1988) they find that both price series respond to shocks to correct for deviations from the long-term equilibrium, and that FFAs have a leading role. Impulse response test and tests concerning the volatility of the price series lend support to this conclusion. The reason that FFAs have a leading role is thought to be that lower transaction costs and easier access to take short positions favour transactions in the FFA market relative to the spot market. They conclude that FFAs can be used as a price discovery vehicle.

Kavussanos, Visvikis and Menachof (2004) find a bi-directional relationship between spot and FFA prices collected one month prior to maturity. When increasing the time to maturity to two and three months, only the FFA prices correct a disequilibrium created by the previous period's deviations. FFA prices thereby are showed to lead spot prices.

Bessler, Drobetz and Seidel (2008) investigate the most liquid dry bulk route in terms of FFA trading using daily data. They test for autocorrelation in spot and FFA rates and find evidence of autocorrelation in spot rates, while FFAs prices on the other hand seem not to be much autocorrelated. This is seen as an implication of the price discovery properties of the FFAs. They also employ the Johansen (1988) framework and find that spot and forward rates are cointegrated, with spot rates converging to forward rates. This is seen as an implication of forward rates containing more information on future spot rates than the current spot rates.

4.3 Freight futures and their ability to forecast the underlying spot rates

Chang and Chang (1996) employ OLS regression analyses to test the predictability properties of the BIFFEX future with regards to the BFI. They find that the BIFFEX future predicts the underlying spot rates accurately for a one-month horizon. However, the performance significantly decreases for horizons up to six months. The explanation power is found to range from 90% one month ahead to 23% six months in advance. For longer horizons the BIFFEX is found to fail predicting the underlying spot rates.

Kavussanos and Nomikos (1999) investigate the ability of BIFFEX freight futures to forecast realized BFI spot prices one-, two- and three months before maturity. The performance of freight futures is compared to that of time-series models which are based on daily data. Freight futures are found to outperform all the time-series models, with the exception of a one month forecast performed by a Vector Error Correction Model (VECM). The performance of the futures is found to diminish as the forecasting horizon is increased. When comparing the time-series models they find that a VECM yields the best performance. An ARIMA (Box-Jenkins) model is found to outperform a random walk for one and two month forecasts, but not for three months, while a Holt-Winters exponential smoothing model is found to be the worst model for all forecasting horizons. They conclude that market participants can use the futures prices as indicators of the future course of the BFI prices.

Haigh (2000) incorporates monthly cointegrating information between spot rates and the BIFFEX future in a VECM, which is used to forecast future spot rates. The use of monthly data stands in contrast to earlier research on error-correction models where daily data was used. The reason for using monthly data is that a fixed time to maturity is thought to yield more robust results than daily data for error-correction models, if not incorporating a differential between the two price series (basis), due to the fact that futures and cash prices should converge when a contract approaches maturity. Haigh compares the forecasting abilities of the VECM to those of a Vector Autoregressive Model (VAR) in levels and a VAR in first differences. The forecasting ability of futures prices is also tested. He finds that the futures price provides the best predictor of the underlying spot rates for the current contract. The forecasting abilities of the future prices do however diminish for longer horizons. Time-series models are found able to outperform the futures contract for longer contract horizons. One- and two-month contracts are found to be explained best by the VAR in first differences. This is thought to be due to the fact that spot rates do not seem to help correct for deviations from the long-term equilibrium.

Kavussanos and Nomikos (2003) investigate the short-term forecasting abilities of futures prices and find that these, when incorporated in a VECM, forecast spot prices better than when used in VAR in first differences, ARIMA or random walk models. The cointegration relationship between future prices and spot prices is thereby showed to help provide the most accurate forecasts.

Batchelor, Alizadeh and Visvikis (2007) investigate FFA prices of the nearest contract and the corresponding spot rates for the same four dry bulk routes as mentioned above. They use time-series models to generate short-term forecasts of spot rates and FFA prices. The models used are an ARIMA model, a VAR model, a VECM and a restricted VECM. For out-of sample tests they estimate non-overlapping forecasts of spot rates up to 20 days ahead. They find that the models which incorporate the cointegrating relationship provide the best short-term forecasts for the spot rates. Finally, they conclude that forward prices help forecast future spot rates and that spot rates do not help forecasting FFA prices. FFA prices are therefore thought to contain more information than the spot rates.

5. Testing the unbiasedness hypothesis

In the context of this thesis, investigating the unbiasedness hypothesis is important because it may provide insight to whether the futures can be used to guide physical market decisions. This part starts off by presenting the methodology which is used for testing the unbiasedness hypothesis. Descriptive statistics of the data series, stationarity tests and tests for cointegration are then reported. Finally, results from testing the unbiasedness hypothesis are elaborated on. The main findings are that the unbiasedness hypothesis is found to hold for TC2 freight futures one month prior to maturity. For this route, the evidence is weaker when it comes to the two- and three month horizons. For TC4, the unbiasedness hypothesis is rejected for future prices collected one-, two- and three months prior to maturity. For TC5 the unbiasedness hypothesis is found to hold for all investigated time-horizons. Whether the unbiasedness hypothesis holds for clean tanker freight futures thereby seems to depend on the route in question and time to maturity.

5.1 How to test the unbiasedness hypothesis

In chapter 3.1 the unbiasedness hypothesis was presented. It was showed that the prices of futures and their underlying spot prices are connected through the following pricing relationship:

$$F_{t,T} = E_t(S_T),$$

given the joint hypothesis of no risk-premium and rationality of expectations. This implies that the price of a future at a time prior to maturity (t-i) differs from the underlying realized spot price at the time of maturity (T) only by a random error, ε_T.

$$S_T = F_{t-i,T} + \varepsilon_T; \qquad \varepsilon_T \sim iid(0, \sigma^2)$$

In line with, this the unbiasedess hypothesis has therefore traditionally been tested empirically using the following equation;

$$S_T = \beta_1 + \beta_2 F_{t-i,T} + \varepsilon_T; \qquad \varepsilon_T \sim iid(0, \sigma^2)$$

Employing this equation, the unbiasedness hypothesis may be investigated using a Wald test. This is done by testing whether the coefficients are statistically different from $\beta_1 = 0$ and $\beta_2 = 1$ simultaneously. If this is the case, the hypothesis of unbiasedness is rejected.

To use this OLS methodology the time series do however need to be stationary. A covariance stationary time series is defined as a time series which has a constant mean, constant variance and constant autocovariances for each given lag. If this is not the case the series is non-stationary. Using the OLS methodology on non-stationary time series induces problems. The coefficient estimates are then inconsistent and their test statistics do not follow standard distributions, leading to invalid inferences and spurious results (Granger and Newbold, 1974).

There is, however, one exception which can be obtained by differencing and testing for cointegration. Differencing is an operation which is performed by calculating the difference between the current and past value of all the observations of a time-series.

$$\Delta Y_t = Y_t - Y_{t-1}$$

Differencing results in losing one observation, but a non-stationary time series may in this manner be transformed to a stationary one. The minimum number of times a time series needs to be differenced to be made stationary is referred to as its order of integration. A non-stationary time series which can be made stationary after differencing once is therefore often denoted I(1), integrated of order 1, and said to have one unit root. If two time series are I(1) their difference is also usually expected to be I(1) as well. There is however one special case where the linear combination of two time-series are I(0), or stationary. The time series do then move together and are said to be cointegrated. Research on this topic was pioneered by Engle and Granger (1987). The reasoning behind testing for cointegration to test the unbiasedness hypothesis is that if spot and futures prices are I(1), they need to be cointegrated to avoid drifting apart. Cointegration is therefore a necessary condition for the unbiasedness hypothesis to hold.

The Engle and Granger methodology does however experience some problems related to their method of estimating a cointegration system. In the context of this thesis two problems are relevant. The first is that if the causality between the series runs both ways there may be a simultaneous equations bias. Secondly, statistical inference on the coefficient estimates of the cointegrating regression cannot be drawn. The reason for this is that the method consists of two steps of regression and the cointegrating regression which is the first regression conducted is based on I(1) variables. In line with this it is difficult to perform any hypothesis tests about the actual cointegrating relationship[12]. To test the unbiasedness hypothesis I will therefore use the vector error correction model (VECM) proposed by Johansen (1988). This method remedies the problems reported above. Besides providing more efficient estimates of the cointegration relationship (Gonzalo, 1994), it has also been showed to be fairly robust if the time series show signs of non-normality (Cheung and Lai, 1993) or heteroskedastic disturbances (Lee and Tse, 1996).

The VECM framework can be regarded as an extension of Vector Autoregressive models (VAR). VAR models are system regression models where the current value of each variable is explained by lagged values of all the variables in the system. Because all variables in the model have to be stationary to facilitate hypothesis testing, VAR models may have to be based on first differenced terms. Such models do however not allow for long run relationships and information on a possible cointegrating relationship between the variables will be lost[13]. A VAR model thereby allows the series to wander apart without bound, as only short term information of the series is incorporated in the model.

To account for both short and long run relationships a VECM combines first differenced and lagged levels of the series. The short run information is captured by the first differenced terms in the same way as in a VAR in first differences, while the long run information is captured by an

---

[12] For a more explanatory presentation of these problems see Brooks (2008).

[13] A long run solution implies that the variables have reached some long term equilibrium and do not change. The differenced terms will therefore be 0, and all the terms in the model cancel out.

error correction term. The error correction term regarding spot and futures prices can be represented in the following way.

$$Z_{t-1} = S_{t-1} + \beta_1 + \beta_2 F_{t-1,T}$$

where Z_{t-1} is a linear combination of the spot and futures price series. S_{t-1} is the spot price. $F_{t-1,T}$ is the futures price. β_1 is a constant and β_2 is the cointegrating vector describing the long run relationship between the variables.

A VECM can be presented by adding the error correction term to a VAR model in first differences. A VECM which can be employed to test for cointegration between spot and futures prices is presented below.

$$\Delta S_t = \mu_s + \sum_{i=1}^{p} a_{s,i} \Delta S_{t-i} + \sum_{i=1}^{p} \theta_{s,i} \Delta F_{t-i,T} + \gamma_s Z_{t-1} + \varepsilon_{st}$$

$$\Delta F_t = \mu_f + \sum_{i=1}^{p} a_{f,i} \Delta S_{t-i} + \sum_{i=1}^{p} \theta_{f,i} \Delta F_{t-i,T} + \gamma_f Z_{t-1} + \varepsilon_{ft}$$

where μ_s and μ_f are constants. γ_s and γ_f are parameters measuring the proportion of last period's equilibrium error which is corrected for, also known as the adjustment speed of the spot and futures prices to their long run equilibrium. $a_{s,i}$, $a_{f,i}$, $\theta_{s,i}$ and $\theta_{f,i}$ are parameters. ε_{st} and ε_{ft} are white-noise error terms. This model corresponds to a non-stationary VAR in levels with lag length p+1.

In the framework of Johansen this model can be represented using the following notation.

$$\Delta X_t = \mu + \sum_{i=1}^{p-1} \Gamma_i \Delta X_{t-i} + \Pi X_{t-1} + \varepsilon_t, \qquad \varepsilon_t \sim iid(0.\Sigma)$$

where μ is a 2×1 vector representing deterministic components which may include an intercept term and/or a linear trend. X_t is a 2×1 vector comprising S_t and $F_{t,T}$, each being I(1) such that the first differenced series is I(0). Γ_i and Π are 2×2 coefficient matrices representing respectively the short and long run adjustment to changes in X_t. ε_t is a 2×1 vector of white noise residuals which have the 2×2 variance-covariance matrix Σ.

The most appropriate number of lags to include in the model can be determined using the Schwartz information criterion (SIC) (Schwartz, 1978). For each cointegrating relationship which is to be tested I will estimate VAR models with lag lengths 1 to 12 to find the number of lags which yields the minimum value of SIC in the VECM. The lag length p for a VAR will correspond to the lag length p-1 for a VECM. I will then perform residual testing. If the residuals show signs of autocorrelation I will increase the lag length.

The result of a cointegration test using a VECM depends on the assumptions made with respect to deterministic components. Five different assumptions can be made[14]. I will comment on the choice of deterministic components when performing the test.

Johansen and Juselius (1990) show that the rank of Π contains information on the cointegrating relationship between the time series. It can therefore be used to choose the model specification which most appropriately reflects the relationship between the time series within this framework. If $\text{rank}(\Pi) = 0$, Π is a 2×2 zero matrix and there is no cointegrating relationship between the time series. The VECM then reduces to a VAR in first differences. If $\text{rank}(\Pi) = 1$, the time series have a single cointegrating relationship. ΠX_{t-1} is then an error correction term and Π can be factorised into two separate matrices, α and β, of dimensions 2×1. Using the representation $\Pi = \alpha\beta'$, α can be interpreted as the vector of the error correction coefficients. α thereby measures the speed of convergence to the long run equilibrium. β represents the vector of cointegrating parameters. If $\text{rank}(\Pi) = 2$ all the variables in X_{t-1} are I(0). The appropriate model is then a VAR in levels. Johansen (1988) provides the test statistics λ_{trace} and λ_{max} which

[14] A listing of these assumptions may be found in Johansen (1995).

can be used to determine the rank of Π, and thereby whether cointegration exists. The critical values determining the results of the tests will be those calculated by MacKinnon-Haug-Michelis (1999).

If the time series have a single cointegrating relationship I will use the VECM framework to test the unbiasedness hypothesis. When testing the unbiasedness hypothesis the object is to test whether the future prices are unbiased predictors of the future spot prices. This means that the futures prices must equal the underlying spot prices on average. Because the future prices are observed at a point in time prior to the realized underlying spot prices the futures prices will be lagged. If the series are found to be I(1) and cointegrated, the unbiasedness hypothesis is tested by restricting the error correction term of each estimated VECM to $\beta_1 = 0$ and $\beta_2 = -1$, making the error correction term at time t-1;

$$Z_{t-1} = S_{t-1} - F_{t-2,t-1}$$

In terms of the VECM framework this is done by testing whether the cointegrating vector is statistically different from $(1\ 0\ -1)$. A test based on the maximum log-likelihood of an unrestricted ($\mathrm{L_u}$) and restricted ($\mathrm{L_r}$) model will be employed. These maximum log-likelihoods are compared using the test statistic $-2(\mathrm{L_r} - \mathrm{L_u})$, which under the null hypothesis follows a chi-square distribution asymptotically with degrees of freedom equal to the number of restrictions (m) placed on the cointegrating vector[15].

[15] A more detailed explanation of this framework can be found in Brooks (2008).

5.2 Properties of the data series

In this chapter I investigate whether the most liquid clean tanker futures traded at Imarex are unbiased predictors of the underlying spot rates at maturity. I concentrate on the routes TC2, TC4 and TC5 as these are the most liquid clean tanker futures traded at Imarex[16]. A presentation of the relevant futures contracts can be found in table 5.1 below.

Table 5.1: Overview of the relevant freight futures contracts

Route	Trade	Size	Lot Size per Month	Price Quotation	Index Provider
TC2	Rotterdam - New York	37,000 mt	1000 mt	Worldscale	Baltic Exchange
TC4	Singapore - Chiba	30,000 mt	1000 mt	Worldscale	Platts
TC5	Ras Tanura - Yokohama	55,000 mt	1000 mt	Worldscale	Platts

Daily prices for the Imarex TC2, TC4 and TC5 freight futures contracts have been provided by Imarex. These contracts are written on indices published by the Baltic Exchange and Platts[17]. Contracts with maturities reaching from the current month and up to three years forward are traded, with delivery periods being months for the front contracts, quarters for more distant delivery periods and years at the back end of the curve. In practice, futures with a delivery period of a calendar year divided into four contracts with delivery periods of the four respective quarters, and these contracts are again divided into contracts with delivery periods of the respective months as maturity approaches, making all settled contracts monthly contracts. Traders have the option to trade contracts at any stage of this process, with the monthly contracts being tradable until the last day of each respective settlement period[18]. At maturity the monthly futures are settled against the arithmetic average of the spot prices of the delivery period of the

[16] Trading volumes were presented in chapter 2.2.

[17] The Baltic Exchange and Platts were introduced in chapter 1.3.

[18] The last trading day was changed from the 15th to the 20th of the month in question with effect from 15th February 2006 (NOS Rulebook Notice 01/2006), and from the 20th to the last day in the delivery period with effect from 18th July 2008 (NOS Rulebook Notice 06/2008).

relevant underlying index[19]. January 2004 is chosen as the first month of collecting futures prices because the liquidity of the futures market was low prior to this time. As the liquidity in this futures market is found in the front of the forward curves, I focus on the prices of futures three, two and one months before maturity. In line with this, three data sets of futures prices which match the underlying realized spot rates have been generated for each route. Relevant futures prices have been collected the last trading day of each relevant month.

Daily spot rates for TC2 have been provided by the Baltic Exchange and cross-checked against data sets provided by Imarex and NOS. The first observation of spot data available from the Baltic Exchange is March 1st 2004. March has therefore been chosen as the first month of spot observations for all routes to ensure that the results are comparable. The daily spot rates for TC4 and TC5 have been collected from various sources including Imarex and NOS, as Platts have a policy of not releasing historic rate assessments for academic purposes. These data sets have also been cross-checked and all irregularities have been investigated and corrected. As the Imarex futures are settled against the monthly arithmetic average of the underlying spot prices, series consisting of the relevant monthly average spot prices have been generated. Because futures prices have been collected from January 2004 onwards, the spot price data sets one and two months before maturity start March 2004, while the data sets matched with the futures prices three months prior to maturity start April 2004. The last included spot observations are collected in September 2009 for all the spot series, while the series stop at the last trading day of August, July and June that year for the one-, two-, and three month futures series, respectively.

To ensure that this thesis as a whole is as easy to understand as possible, all data has been transformed in the same way throughout the thesis. In line with this, the price series have been rebased to the flat rate level of 2009 and transformed in natural logarithms. Descriptive statistics on the logarithmic first differences of the rebased data series are presented in table 5.2.

[19] Further product specifications for the futures contracts investigated can be found in appendix I.

Table 5.2: Descriptive statistics and tests for normality for the return series

One-Month Price Series								
	N	Mean	Max	Min	StdDev	Skewness	Kurtosis	Jarque-Bera
TC2AVG1	66	-0.010057	0.603702	-0.451017	0.195082	0.480973	4.131543	6.065754
TC2-1M	66	-0.009265	0.535244	-0.444381	0.190534	0.296840	3.354318	1.314491
TC4AVG1	66	-0.009988	0.546654	-0.598901	0.218650	0.199318	3.262530	0.626541
TC4-1M	66	-0.010793	0.464708	-0.378654	0.192577	0.353277	2.972952	1.374866
TC5AVG1	66	-0.002746	0.487594	-0.484901	0.221478	-0.063043	2.589896	0.506229
TC5-1M	66	-0.006184	0.500463	-0.529193	0.211683	-0.030450	3.172750	0.092266

Two-Month Price Series								
	N	Mean	Max	Min	StdDev	Skewness	Kurtosis	Jarque-Bera
TC2AVG2	66	-0.010057	0.603702	-0.451017	0.195082	0.480973	4.131543	6.065754
TC2-2M	66	-0.005492	0.459532	-0.392344	0.140556	0.121374	4.544942	6.725876
TC4AVG2	66	-0.009988	0.546654	-0.598901	0.218650	0.199318	3.262530	0.626541
TC4-2M	66	-0.008161	0.389465	-0.333773	0.152149	0.290517	2.921381	0.945397
TC5AVG2	66	-0.002746	0.487594	-0.484901	0.221478	-0.063043	2.589896	0.506229
TC5-2M	66	-0.003524	0.514899	-0.535905	0.191040	0.303027	3.888186	3.179482

Three-Month Price Series								
	N	Mean	Max	Min	StdDev	Skewness	Kurtosis	Jarque-Bera
TC2AVG3	65	-0.004037	0.603702	-0.451017	0.190321	0.567370	4.271005	7.862533
TC2-3M	65	-0.003340	0.366281	-0.373625	0.119582	0.014601	4.962902	10.437480
TC4AVG3	65	-0.006870	0.546654	-0.598901	0.218867	0.171282	3.278384	0.527714
TC4-3M	65	-0.006159	0.367038	-0.473875	0.134584	-0.319205	4.464279	6.910801
TC5AVG3	65	0.002707	0.487594	-0.484901	0.218691	-0.074878	2.661922	0.370293
TC5-3M	65	-0.000844	0.451545	-0.518163	0.153742	-0.180916	4.745122	8.602677

N is the number of observations of each time series. Max is the maximum value of the series and min is the minimum value.
Mean, standard deviation[2], skewness and kurtosis are the first, second, third and forth standardized moments of the time series.
The Jarque-Bera is a test for normality. The null hypothesis is that the series have skewness and kurtosis similar to a normal distribution.
The test statistic follows a chi-square distribution with two degrees of freedom under the null hypothesis. The 5% critical value is 5.9915.
The null hypothesis is rejected if the test statistic is greater than the critical value. Rejection indicates sign of non-normality.
Values representing rejected null hypothesis are reported in red.
All series have been rebased to the flat rate levels of 2009 and reflect the logarithmic first differences of these series.

The means are as expected close to 0 for all series. The fact that all except the three-month spot series of TC5 are negative show that freight rates have fallen during the sample period. The maximum and minimum values seem fairly well balanced around the mean. They do however indicate that the freight rates are quite volatile. This is also reflected by the standard deviations.

An explanation for highly volatile spot rates might be found in the elastic expectation hypothesis of Zannetos (1966). For the short run this hypothesis states that when the market expects spot rates to increase (decrease) ship owners hold back (offer) tonnage while charterers hurry (wait) to fix ships. The volatility of the monthly average spot rate series are probably exposed to this behaviour through the momentum it induces to the daily rates. The standard deviations are also higher for the spot series than the futures series for all routes and maturities. A reason for this might be that the futures market is not able to foresee the peaks and troughs of the spot rates, hence the futures are traded at price levels which are perceived “normal”. As maturity approaches the accuracy regarding the expectations of the future spot rates should however improve. This will lead to increasing volatility as the futures approach the pricing period. When the pricing period is reached the volatility should then fall as the underlying spot rates will be reported gradually. Looking at the three, two and one-month futures series the standard deviation is in fact increasing as pricing approaches. These results are in line with the findings of Adland et al. (2009), which investigate forward curve dynamics of the tanker futures market. They do however argue that an implication of the elastic expectation hypothesis is that short term forward freight rates should be more volatile than spot rates. The reason why this is not found here might be that monthly data has been used. Adland et al. (2009) also argue that the volatility of freight futures across routes should converge in the back of the forward curve because the newbuilding prices of all ships are highly correlated. This argument is based on the fact that long term freight rates and the newbuilding prices of ships are interdependent under the assumption of integrated freight and newbuilding markets (Strandenes, 1984). Converging volatility levels can however not be observed in this case as a time-horizon of three months prior to pricing does not reflect long-term prices.

Looking at the standard deviations of the separate routes it may be observed that the volatility of TC5 is generally higher than that of TC4 and TC2. TC4 does at the same time experience higher volatility than TC2. The reason for the former is in line with the findings of for instance Kavussanos (1996, 2003). He shows that the freight rates of large ships are more volatile than those of smaller ships. The reason is thought to be that small ships are more versatile regarding trades and ports, and therefore are not as exposed to changes in demand as larger ships. When discussing this topic with shipbrokers which concentrate on the physical TC2, TC4 and TC5 markets they verified this phenomenon. An explanation of the latter may be that shipping markets are geographically separated in the short term. Because the routes TC2 and TC4 are served by the same type of ships the freight rates and thereby also volatility would be the same if these markets were perfectly integrated. There is, however, a barrier between the TC2 and TC4 market due to the distance between the two routes. This distance implies that for a ship to switch markets the ship operator must believe that doing so will yield a profit. The expected earnings must therefore exceed the cost of switching markets, which consists of the cost of the voyage plus an alternative cost due to lost income resulting from the days in ballast. Shipbrokers report that the ships usually sail through the Suez Canal if switching between these markets. The cost of ballasting from the European Continent to the Arabian Gulf for a ship serving TC2 is then a result of approximately USD 200 000 in canal dues, burning 700 tonnes of bunkers and the alternative cost of approximately 20 days at sea. Going the opposite way is usually a bit cheaper as such operations often involve transporting jet fuel or gas oil at discounted prices. The result of this geographical separation is that the TC4 and TC5 markets are more integrated in the short run than the TC2 and TC5 market. As the products shipped by the vessels operating TC5 may be shipped by two smaller ships, or the ships operating TC5 might do a typical TC4 voyage if demand is low, this will lead the rates and volatility of TC4 and TC5 to be more interconnected than those of TC2 and TC5. The volatility of TC5 is therefore thought to spill over to TC4 in the short run, making the volatility of TC4 greater than that of TC2.

The skewness and kurtosis measure the distributions of the price series relative to the normal distribution. Negative (positive) skewness indicates the distribution of a series has a long tail to the left (right) relative to the right (left). The series do not experience a high degree of skewness.

Kurtosis is a measure of the peakedness of distribution relative to a normal distribution. High kurtosis data tend to have a distinct peak near the mean, decline rapidly and have heavy tails. This measure is always positive, and the normal distribution has a kurtosis of three. Most of the series have excess kurtosis, and are thus leptokurtic, which is typical for financial data.

The Jarque-Bera tests (Bera and Jarque, 1980) are included to illustrate the distributions of the series. These indicate that the logarithmic first differences of all the one and two month series are normally distributed, except for the one month TC2 spot series and both the two-month TC2 series. The main reason for this is the high kurtosis of this series, indicating a distinct peak near the mean and heavy tails. The distribution of the three-month spot series of TC2 also show signs of non-normality, while the spot series of TC4 and TC5 do not deviate much from the normal distribution. Normal distributions with regards to the three-month futures series are rejected. These series seem to experience excess kurtosis.

As explained when presenting the theory on how to test the unbiasedness hypothesis it is important to determine whether the price series are stationary. Augmented Dickey Fuller (Dickey and Fuller, 1979) tests, PP tests (Phillips and Perron, 1988) and KPSS tests (Kwiatkowski, Phillips, Schmidt and Shin, 1992) are employed for this purpose. The PP test builds on the Dickey-Fuller test, but instead of including lagged variables it makes a non-parametric adjustment to the t-test statistic to correct for autocorrelation. The KPSS test differs from the ADF and PP test by formulating the null hypothesis as the series being stationary. The results from the tests are presented in table 5.3.

Table 5.3: Tests for stationarity

One-Month Price Series						
	ADF lvl (lags)	ADF 1st diff (lags)	PP lvl (BW)	PP 1st diff (BW)	KPSS lvl (BW)	KPSS 1st diff (BW)
TC2AVG1	-2.767434 (0)	-7.643165 (1)	-2.739663 (2)	-9.230865 (5)	0.203298 (5)	0.093470 (5)
TC2-1M	-3.286860 (0)	-9.570154 (0)	-3.340097 (3)	-10.40438 (5)	0.177519 (5)	0.086369 (4)
TC4AVG1	-2.342152 (3)	-8.252589 (1)	-2.556760 (0)	-7.937212 (7)	0.398637 (5)	0.084727 (7)
TC4-1M	-2.528757 (0)	-8.649365 (0)	-2.584869 (1)	-8.905438 (5)	0.406269 (5)	0.095812 (6)
TC5AVG1	-3.858930 (1)	-6.978811 (1)	-3.261801 (1)	-7.208832 (7)	0.091047 (5)	0.045983 (6)
TC5-1M	-3.248518 (0)	-8.486261 (0)	-3.360983 (1)	-9.111979 (7)	0.078108 (5)	0.102922 (9)

Two-Month Price Series						
	ADF lvl (lags)	ADF 1st diff (lags)	PP lvl (BW)	PP 1st diff (BW)	KPSS lvl (BW)	KPSS 1st diff (BW)
TC2AVG2	-2.767434 (0)	-7.643165 (0)	-2.739663 (2)	-9.230865 (5)	0.203298 (5)	0.093470 (5)
TC2-2M	-2.669912 (0)	-8.272453 (0)	-2.736482 (2)	-8.325780 (3)	0.165568 (5)	0.089277 (3)
TC4AVG2	-2.342152 (3)	-8.252589 (1)	-2.556760 (0)	-7.937212 (7)	0.398637 (5)	0.084727 (7)
TC4-2M	-2.181507 (0)	-8.066081 (0)	-2.181507 (0)	-8.150579 (5)	0.348428 (5)	0.147763 (6)
TC5AVG2	-3.858930 (1)	-6.978811 (1)	-3.261801 (1)	-7.208832 (7)	0.091047 (5)	0.045983 (6)
TC5-2M	-3.426529 (0)	-8.657439 (0)	-3.555020 (2)	-10.54446 (9)	0.084087 (4)	0.173831 (11)

Three-Month Price Series						
	ADF lvl (lags)	ADF 1st diff (lags)	PP lvl (BW)	PP 1st diff (BW)	KPSS lvl (BW)	KPSS 1st diff (BW)
TC2AVG3	-2.727194 (0)	-7.466431 (0)	-2.804396 (1)	-8.677414 (5)	0.202710 (5)	0.188196 (6)
TC2-3M	-2.282242 (0)	-6.728874 (0)	-2.550383 (1)	-6.720359 (7)	0.170888 (5)	0.128262 (5)
TC4AVG3	-2.336603 (3)	-8.175283 (1)	-2.683615 (1)	-7.779586 (7)	0.393138 (5)	0.102835 (6)
TC4-3M	-1.967071 (0)	-7.156638 (0)	-1.967071 (0)	-7.133414 (6)	0.285283 (5)	0.159256 (5)
TC5AVG3	-3.810726 (1)	-6.836806 (1)	-3.046421 (0)	-6.847288 (7)	0.092134 (5)	0.065176 (5)
TC5-3M	-3.096361 (0)	-7.873225 (0)	-3.309304 (2)	-8.008255 (6)	0.116235 (5)	0.116132 (7)

The ADF tests refer to Augmented Dickey Fuller tests with one intercept included. The null hypothesis is non-stationary.
Lag length set to max 12 to control for seasonality factors and automatically selected using Schwarz Information Criterion.
ADF is a one-tailed t-test where the test statistic follows the distribution calculated by Dickey and Fuller under the null hypothesis.
A 5% (10%) level of significance has a critical value of approximately -2.90 (-2.59) (depending on the number of lags) (MacKinnon, 1996)
The null hypothesis is rejected if the test statistic is smaller than the critical value, and if this is the case the series is said to be stationary.

The PP tests refer to the Phillips-Perron test with one intercept included. The null hypothesis is that a series is non-stationary.
The spectral estimation method is the Berlett kernel method and the bandwidth selection is the Newey-West Bandwidth.
The test statistic follows the same asymptotic distributions as the ADF test statistic and normalized bias statistics under the null hypothesis.
The critical values are therefore the same as for the ADF test.
The null hypothesis is rejected if the test statistic is smaller than the critical value, and if this is the case the series is said to be stationary.

The KPSS tests refer to the Kwiatkowski, Phillips, Schmidt and Shin test with an intercept included.
The spectral estimation method is the Berlett kernel method and the bandwidth selection is the Newey-West Bandwidth.
KPSS is a right tailed test, and the test statistic is a LM statistic which converges to a function of a standard Brownian motion under the null hypothesis.
The critical value is 0.436 (0.347) at a 5% (10%) level of significance and is calculated from the asymptotic distribution of a standard Brownian motion.
The null hypothesis is rejected if the test statistic is larger than the critical value, and if this is the case the series is said to be non-stationary.

Values representing rejected null hypothesis using the 5 % critical value are reported in red.
Values representing rejected null hypothesis using the 10 % critical value are reported in purple.

The ADF tests at a five percent level of significance indicate that all the TC2 and TC4 series are stationary in first differences, except form the TC2 one-month futures series. The TC5 series are all stationary in levels at this level of significance. Using the 10 percent level of significance only the TC2 three month futures series and the TC4 series are indicated stationary in first differences. These results are mirrored by the PP tests with on one exception, the TC4 three month spot series is indicated to be stationary in levels when using a 10 percent level of significance. This shows that the tests are sensitive to the observations included as the only difference between the one and two month TC4 spot price series and the three month TC4 spot price series is the observation of March 2004. The small sample size might therefore be a problem when conducting the tests. The KPSS tests indicate that all the series are stationary in levels at a 5 percent significance level. All the TC4 series except the three-month futures series do however show signs of being stationary in first differences at a 10 percent significance level.

As the TC2 series show signs of being stationary in levels I will put weight on the OLS methodology for this route. In addition to the OLS methodology I will however also use the Johansen methodology because the results from the stationarity tests are inconclusive. For TC4 which showed signs of being stationary in first differences, I will put weight on the Johansen methodology, but also look at the OLS results. Because all three test methods indicate that the TC5 series are stationary in levels, I will focus on the OLS methodology when testing the unbiasedness hypothesis for this route. As cointegration is a prerequisite for using Johansen’s framework to test the unbiasedness hypothesis I will now go on to test for cointegration.

5.3 Results from testing the unbiasedness hypothesis

The VECM framework proposed by Johansen will now be employed to test whether spot and futures prices one, two and three months prior to maturity are cointegrated. In order to obtain a well specified VECM, starting points regarding the lag length of the models are estimated using VAR models. This is in line with the fact that the lag length p of an unrestricted VAR can be re-parameterised to the lag length p-1 in a VECM of first differences of the dependent variable plus the levels terms (Kavussanos and Nomikos, 1999). The lag length p of the unrestricted VARs proposed by SIC was found to be 1 for all the one month series and the two month TC2 and TC5 series, 2 for the two month TC4 series and the three month TC2 series, and 3 for the three month TC4 and TC5 series.

Using the lag lengths p-1, VECMs for all the cointegrating relationships were estimated. These models included no deterministic trend or intercept in μ and only intercept in the cointegrating relationship. Regarding the deterministic components this reflects that a trend in the price series is neither likely theoretically, nor when observing the series. The intercept in the cointegrating relationship is included to enable testing of the unbiasedness hypothesis. Some of these models did however yield autocorrelated residuals. The lag length of the models was therefore increased to obtain models with satisfactory residual diagnostics. The results from employing these models to test for cointegration is presented in table 5.4 below.

Table 5.4: Results from cointegration tests using VECMs.

One-Month Price Series - Trace Statistics			
r	**TC2 (1 lag)**	**TC4 (1 lag)**	**TC5 (1 lag)**
0	24.04513	28.99162	39.82123
	(20.26184)	(20.26184)	(20.26184)
1	5.26007	3.64622	11.11962
	(9.16455)	(9.16455)	(9.16455)
Rank	1	1	2

One-Month Price Series - Max Statistics			
r	**TC2 (1 lag)**	**TC4 (1 lag)**	**TC5 (1 lag)**
0	18.78506	25.34540	28.70162
	(15.89210)	(15.89210)	(15.89210)
1	5.26007	3.64622	11.11962
	(9.16455)	(9.16455)	(9.16455)
Rank	1	1	2

Two-Month Price Series - Trace Statistics			
r	**TC2 (4 lags)**	**TC4 (2 lags)**	**TC5 (3 lags)**
0	30.87681	23.54791	34.80822
	(20.26184)	(20.26184)	(20.26184)
1	6.57580	3.86595	13.61228
	(9.16455)	(9.16455)	(9.16455)
Rank	1	1	2

Two-Month Price Series - Max Statistics			
r	**TC2 (4 lags)**	**TC4 (2 lags)**	**TC5 (3 lags)**
0	24.30101	19.68196	21.19594
	(15.89210)	(15.89210)	(15.89210)
1	6.57580	3.86595	13.61228
	(9.16455)	(9.16455)	(9.16455)
Rank	1	1	2

Three-Month Price Series - Trace Statistics			
R	**TC2 (3 lags)**	**TC4 (2 lags)**	**TC5 (2 lags)**
0	25.13691	21.92552	36.11275
	(20.26184)	(20.26184)	(20.26184)
1	8.08913	6.10219	11.27679
	(9.16455)	(9.16455)	(9.16455)
Rank	1	1	2

Three-Month Price Series - Max Statistics			
r	**TC2 (3 lags)**	**TC4 (2 lags)**	**TC5 (2 lags)**
0	17.04778	15.82333	24.83596
	(15.89210)	(15.89210)	(15.89210)
1	8.08913	6.10219	11.27679
	(9.16455)	(9.16455)	(9.16455)
Rank	1	0	2

r is the number of cointegrating vectors.
For the λ_{trace} statistic the null is that rank is smaller than or equal to r, while the alternative is that rank exceeds r.
For the λ_{max} statistic the null is that rank equals r, while the alternative is that rank equals r+1
The 5% critical values are those of MacKinnon-Haug-Michelis (1999) and are reported in parenthesis
The rejection rule is to reject the null hypothesis if the test statistic is larger than the critical value
Red colour denotes rejection of the null hypothesis
Rank describes the number of cointegrating vectors implied by the hypothesis tests

The results from the cointegration tests using Johansen's procedure indicate that the spot and futures prices one, two and three months before maturity are cointegrated for TC2 and TC4[20]. Because the λ_{trace} and λ_{max} statistics of Johansen have been showed to imply that the variables are cointegrated too often in small samples, I applied the small sample correction proposed by Reimer (1992). Using this correction did not alter the results, except from the λ_{max} statistic of the

[20] For the TC2 three-month series the λ_{trace} and λ_{max} statistics yield conflicting results.

three month series of TC2 where the rank changed to 0[21]. In line with Cheung and Lai (1993) I will however put more weight on the λ_{trace} statistic, because this has been found to be more robust to non-normality than the λ_{max} statistic. The Johansen tests for cointegration thereby indicate that the TC2 and TC4 series are cointegrated, while the TC5 series are not. I will therefore not use the Johansen approach when testing the unbiasedness hypothesis for the TC5 series.

When comparing these results to those of the stationarity tests it can be observed that the TC2 series are suggested to be I(1) by the cointegration tests, which contradicts the findings of the stationarity tests to some extent. For the TC4 series the results from the cointegration tests are in line with those of the stationarity tests, suggesting that it is reasonable to put more weight on the Johansen methodology than on the OLS methodology. The ranks of the TC5 series are found to be two by the cointegration tests, which is in accordance with the tests for stationarity. Focusing on the OLS tests is therefore further supported by the cointegration tests for the TC5 series.

Given the assumption of no deterministic trend or intercept in the VECMs which are to be estimated the term μ is superfluous. The VECMs therefore take the following form.

$$\Delta X_t = \sum_{i=1}^{p-1} \Gamma_i \Delta X_{t-i} + \Pi X_{t-1} + \varepsilon_t, \qquad \varepsilon_t \sim iid(0.\Sigma)$$

A presentation of the models with one lag included is provided below. These models can easily be expanded to reflect more lags, and should therefore yield an easy understandable picture of the structure of the VECMs. The reason for including only one lag here is to conserve space.

[21] Results from these tests are provided in Appendix II.

One-month futures price and matching spot price series (1 lag):

$$\begin{pmatrix} \Delta S_T \\ \Delta F_{t-1,T} \end{pmatrix} = \begin{pmatrix} \Gamma_{11} & \Gamma_{12} \\ \Gamma_{21} & \Gamma_{22} \end{pmatrix} \begin{pmatrix} \Delta S_{t-1} \\ \Delta F_{t-2,t-1} \end{pmatrix} + \begin{pmatrix} \alpha_1 \\ \alpha_2 \end{pmatrix} (1\ \beta_1\ \beta_2) \begin{pmatrix} S_{t-1} \\ 1 \\ F_{t-2,t-1} \end{pmatrix} + \begin{pmatrix} \varepsilon_{sT} \\ \varepsilon_{fT} \end{pmatrix}, \begin{pmatrix} \varepsilon_{sT} \\ \varepsilon_{fT} \end{pmatrix} \sim iid(0, \Sigma)$$

Two-month futures price and matching spot price series (1 lag):

$$\begin{pmatrix} \Delta S_T \\ \Delta F_{t-2,T} \end{pmatrix} = \begin{pmatrix} \Gamma_{11} & \Gamma_{12} \\ \Gamma_{21} & \Gamma_{22} \end{pmatrix} \begin{pmatrix} \Delta S_{t-1} \\ \Delta F_{t-3,t-1} \end{pmatrix} + \begin{pmatrix} \alpha_1 \\ \alpha_2 \end{pmatrix} (1\ \beta_1\ \beta_2) \begin{pmatrix} S_{t-1} \\ 1 \\ F_{t-3,t-1} \end{pmatrix} + \begin{pmatrix} \varepsilon_{sT} \\ \varepsilon_{fT} \end{pmatrix}, \begin{pmatrix} \varepsilon_{sT} \\ \varepsilon_{fT} \end{pmatrix} \sim iid(0, \Sigma)$$

Three-month futures price and matching spot price series (1 lag):

$$\begin{pmatrix} \Delta S_T \\ \Delta F_{t-3,T} \end{pmatrix} = \begin{pmatrix} \Gamma_{11} & \Gamma_{12} \\ \Gamma_{21} & \Gamma_{22} \end{pmatrix} \begin{pmatrix} \Delta S_{t-1} \\ \Delta F_{t-4,t-1} \end{pmatrix} + \begin{pmatrix} \alpha_1 \\ \alpha_2 \end{pmatrix} (1\ \beta_1\ \beta_2) \begin{pmatrix} S_{t-1} \\ 1 \\ F_{t-4,t-1} \end{pmatrix} + \begin{pmatrix} \varepsilon_{sT} \\ \varepsilon_{fT} \end{pmatrix}, \begin{pmatrix} \varepsilon_{sT} \\ \varepsilon_{fT} \end{pmatrix} \sim iid(0, \Sigma)$$

where Δ denotes the first difference. S_T is the realized monthly average spot rate. $F_{t-1,T}$, $F_{t-2,T}$ and $F_{t-3,T}$ are the futures prices with maturity at time T respectively one, two and three months before maturity. The vector of Γ_i measures the short run adjustment to changes in the vector of the differenced spot and futures prices from the previous period. α_1 and $\alpha_2$form the vector of the error correction coefficients, measuring the speed of convergence to the long run equilibrium. The vector $(1\ \beta_1\ \beta_2)$ represents the cointegrating parameters, which together with the subsequent vector form the error correction term. ε_{sT} and ε_{fT} are white noise residuals which have the variance-covariance matrix Σ. The tables 5.5 – 5.7 include relevant coefficients of the estimated models and results from testing the unbiasedness hypothesis.

Table 5.5: Model specifications and unbiasedness tests for the one-month VECMs.

One-Month Price Series - Model Specification							
	Coefficients				Hypothesis Tests		
	α1	α2	β1	β2	H0: β1 = 0	H0: β2 = -1	H0: β1=0, β2= -1
TC2							
Coeff	0.28364	0.72517	1.14689	-1.22411	-	-	-
Std Error	(0.32714)	(0.20022)	(0.38984)	(0.07664)	-	-	-
Test Stat	0.86701	3.62191	2.94194	-15.97310	5.21732	5.15298	5.45348
5% CV	2.00172	2.00172	2.00172	2.00172	3.84146	3.84146	5.99146
TC4							
Coeff	-0.73564	0.33665	0.72328	-1.13523	-	-	-
Std Error	(0.41747)	(0.29134)	(0.21596)	(0.04245)	-	-	-
Test Stat	-1.76213	1.15552	3.34921	-26.74240	5.85550	5.45506	10.31702
5% CV	2.00172	2.00172	2.00172	2.00172	3.84146	3.84146	5.99146

For the coefficients the null hypothesis is that they are not statistically different from zero.
The test statistics of the coefficients follow a student t-distribution with N-r degrees of freedom under the null hypothesis.
r is the number of regressors.
The test statistics of the hypothesis tests to the right follow a chi-square distribution asymptotically under the null hypothesis.
The degrees of freedom of these tests is equal to the number of restrictions imposed.
5% CV is the critical value of each respective test.
Red colour denotes rejection of the null hypothesis.
Green colour denotes indication of unbiasedness.

One-Month Price Series - Residual Diagnostics						
	Normality	Normality*	LM(1)*	Q(12)	Q(12)*	White*
TC2						
Est	0.22208	3.39955	1.03163	12.23100	42.73139	15.85601
Eft	3.17747			11.24900		
5% CV	5.99146	9.48773	9.48773	21.02607	60.48089	28.86930
TC4						
Est	2.01040	2.05704	5.15033	18.69500	51.52841	10.82943
Eft	0.04664			18.31800		
5% CV	5.99146	9.48773	9.48773	21.02607	60.48089	28.86930

Normality refers to the Doornik and Hansen (1994) test for normality.
LM(1) is the Breusch-Godfrey Lagrange Multiplier test for multivariate autocorrelation of order 1.
Q(12) denotes the Ljung-Box test for autocorrelation of the first 12 lags.
Q(12)* is the Ljung-Box statistic with a small sample correction.
White refers to White's test for heteroskedasticity.
All test statistics follow a X^2 distribution asymptotically.
Asterixes denote that the respective test is bivariate.

Table 5.6: Model specifications and unbiasedness tests for the two-month VECMs.

Two-Month Price Series - Model Specification							
	Coefficients				Hypothesis Tests		
	α1	α2	β1	β2	H0: β1 = 0	H0: β2 = -1	H0: β1=0, β2= -1
TC2							
Coeff	-0.24179	0.483045	1.860879	-1.363451	-	-	-
Std Error	(0.24437)	(0.09612)	(0.80073)	(0.15683)	-	-	-
Test Stat	-0.98943	5.02527	2.32399	-8.69378	4.06616	4.03990	4.17228
5% CV	2.01537	2.01537	2.01537	2.01537	3.84146	3.84146	5.99146
TC4							
Coeff	-0.13928	0.44525	2.32606	-1.44174	-	-	-
Std Error	(0.23650)	(0.09720)	(0.58076)	(0.11339)	-	-	-
Test Stat	-0.58891	4.58051	4.00520	-12.71520	9.47916	9.15100	12.66946
5% CV	2.00488	2.00488	2.00488	2.00488	3.84146	3.84146	5.99146

See notes in table 5.5.

Two-Month Price Series - Residual Diagnostics						
	Normality	Normality*	LM(1)*	Q(12)	Q(12)*	White*
TC2						
Est	7.25395	16.21789	3.05997	7.39980	28.88028	54.33703
Eft	8.96394			12.46800		
5% CV	5.99146	9.48773	9.48773	21.02607	46.19426	72.15322
TC4						
Est	5.61638	9.46898	7.02705	11.51400	41.85113	23.28098
Eft	3.85260			16.91900		
5% CV	5.99146	9.48773	9.48773	21.02607	55.75848	43.77297

See notes in table 5.5.

Table 5.7: Model specifications and unbiasedness tests for the three-month VECMs.

Three-Month Price Series - Model Specification							
	Coefficients				Hypothesis Tests		
	α1	α2	β1	β2	H0: β1 = 0	H0: β2 = -1	H0: β1=0, β2= -1
TC2							
Coeff	0.01032	0.19527	6.68975	-2.30855	-	-	-
Std Error	(0.09955)	(0.04681)	(2.09916)	(0.41138)	-	-	-
Test Stat	0.10361	4.17101	3.18686	-5.61170	5.38000	5.35408	5.52272
5% CV	2.0095752	2.0095752	2.0095752	2.0095752	3.84146	3.84146	5.99146
TC4							
Coeff	0.15598	0.28907	3.57459	-1.68375	-	-	-
Std Error	(0.16226)	(0.07151)	(1.14158)	(0.22257)	-	-	-
Test Stat	0.96127	4.04222	3.13127	-7.56491	5.52204	5.34668	8.06708
5% CV	2.005746	2.005746	2.005746	2.005746	3.84146	3.84146	5.99146

See notes in table 5.5.

Three-Month Price Series - Residual Diagnostics						
	Normality	Normality*	LM(1)*	Q(12)	Q(12)*	White*
TC2						
Est	6.33239	13.12012	2.77333	8.10870	52.53067	42.61549
Eft	6.78773			20.59700		
5% CV	5.99146	9.48773	9.48773	21.02607	50.99846	58.12404
TC4						
Est	5.95984	8.49306	6.71602	12.83400	35.21473	28.10006
Eft	2.53321			11.54200		
5% CV	5.99146	9.48773	9.48773	21.02607	55.75848	43.77297

See notes in table 5.5.

Results from using the OLS methodology to test for unbiasedness are reported in table 5.8 – 5.10. For the tests where autocorrelation in the residuals is indicated, Newey-West heteroskedasticity-autocorrelation consistent standard errors have been employed.

Table 5.8: Model specifications and unbiasedness tests for the one-month OLS models.

		One-Month Price Series - Model Specification				
		β1	β2	H0: β1 = 0	H0: β2 = 1	H0: β1=0, β2= 1
TC2						
	Coeff	-0.18555	1.03454	-	-	-
	Std Error	(0.32664)	(0.06430)	-	-	-
	Test Stat	-0.56805	16.08939	-0.56805	0.53710	0.38338
TC4						
	Coeff	-0.62172	1.11503	-	-	-
	Std Error	(0.25101)	(0.04888)	-	-	-
	Test Stat	-2.47689	22.81108	-2.47689	2.35327	6.88812
TC5						
	Coeff	-0.41230	1.08341	-	-	-
	Std Error	(0.21833)	(0.04468)	-	-	-
	Test Stat	-1.88843	24.25003	-1.88843	1.86691	1.84038
5% CV		1.99714	1.99714	1.99714	1.99714	3.13814

For the coefficients the null hypothesis is that they are not statistically different from zero.
The test statistics of the coefficients follow a student t-distribution with N-r degrees of freedom under the null hypothesis.
The test statistics of the hypothesis tests to the right follow a chi-square distribution asymptotically under the null hypothesis.
The degrees of freedom of these tests is equal to the number of restrictions imposed.
5% CV is the critical value of each respective test.
Red colour denotes rejection of the null hypothesis.
Green colour denotes indication of unbiasedness.

		One-Month Price Series - Residual Diagnostics			
		Normality	LM(1)	Q(12)	White
TC2					
	Test Stat	0.56048	0.00085	8.69330	0.10172
TC4					
	Test Stat	0.44148	0.09075	21.68800	0.00110
TC5					
	Test Stat	0.38309	0.30412	4.42010	1.68618
5% CV		(5.99146)	(3.84146)	(21.02607)	(3.84146)

Normality refers to the Jarque-Bera test for normality.
LM(1) is the Breusch-Godfrey Lagrange Multiplier test for multivariate autocorrelation of order 1.
Q(12) denotes the Ljung-Box test for autocorrelation of the first 12 lags.
White refers to Whites test for heteroskedasticity.
All test statistics follow a X^2 distribution asymptotically.

Table 5.9: Model specifications and unbiasedness tests for the two-month OLS models.

Two-Month Price Series - Model Specification	β1	β2	H0: β1 = 0	H0: β2 = 1	H0: β1=0, β2= 1
TC2					
Coeff	1.08500	0.78333	-	-	-
Std Error	(0.77116)	(0.15063)	-	-	-
Test Stat	1.40697	5.20026	1.40697	-1.43841	1.27033
TC4					
Coeff	-0.37063	1.05827	-	-	-
Std Error	(0.57947)	(0.11344)	-	-	-
Test Stat	-0.63960	9.32895	-0.63960	0.51363	3.85633
TC5					
Coeff	0.62310	0.86720	-	-	-
Std Error	(0.69862)	(0.14332)	-	-	-
Test Stat	0.89190	6.05092	0.89190	-0.92662	0.64093
5% CV	1.99714	1.99714	1.99714	1.99714	3.13814

See notes in table 5.8.

Two-Month Price Series - Residual Diagnostics	Normality	LM(1)	Q(12)	White
TC2				
Test Stat	0.90419	11.14051	16.84200	0.68033
TC4				
Test Stat	0.98720	3.20292	13.69400	0.00445
TC5				
Test Stat	0.77851	12.50689	14.60400	0.04452
5% CV	(5.99146)	(3.84146)	(21.02607)	(3.84146)

See notes in table 5.8.

Table 5.10: Model specifications and unbiasedness tests for the three-month OLS models.

Three-Month Price Series - Model Specification					
	β1	β2	H0: β1 = 0	H0: β2 = 1	H0: β1=0, β2= 1
TC2					
Coeff	1.83763	0.63480	-	-	-
Std Error	(1.05298)	(0.20518)	-	-	-
Test Stat	1.74517	3.09386	1.74517	-1.77991	1.92661
TC4					
Coeff	0.07113	0.97013	-	-	-
Std Error	(1.05061)	(0.20293)	-	-	-
Test Stat	0.06771	4.78072	0.06771	-0.14720	2.17886
TC5					
Coeff	1.71173	0.64505	-	-	-
Std Error	(0.89563)	(0.18428)	-	-	-
Test Stat	1.91120	3.50036	1.91120	-1.92617	2.53805
5% CV	1.99773	1.99773	1.99773	1.99773	3.14044

See notes in table 5.8.

Three-Month Price Series - Residual Diagnostics				
	Normality	LM(1)	Q(12)	White
TC2				
Test Stat	0.76630	25.59714	40.89200	0.45610
TC4				
Test Stat	0.79921	18.08121	23.95100	0.02345
TC5				
Test Stat	0.69135	30.87508	43.79700	1.20369
5% CV	(5.99146)	(3.84146)	(21.02607)	(3.84146)

See notes in table 5.8.

In the following paragraphs I comment on the results from testing the unbiasedness hypothesis. As mentioned earlier, findings from employing both the OLS and Johansen methodology will be used. More specifically, I will use the results of both methods for TC2. For TC4 I will do the same, but will put more weight on the Johansen tests. For TC5 I will use OLS exclusively. This is in line with the results from testing for stationarity and cointegration.

The results for the one-month price series can be found in table 5.5 and 5.8. For TC2 the Johansen methodology test indicates that the unbiasedness hypothesis cannot be rejected. The OLS test provides the same result. The unbiasedness hypothesis is therefore thought to hold for the one-month TC2 series. For TC4 the Johansen test suffers from signs of autocorrelated residuals at lag 6 and 7 (these Ljung-Box Q-statistics are not reported in the table). The test statistic values are however only slightly above the critical values, so the model should at least be able to provide a useable indication with regards to whether the unbiasedness hypothesis holds. The results from using both the Johansen and OLS methodology suggests that the unbiasedness hypothesis does not hold for this route. For TC5 the unbiasedness hypothesis indicated to hold using OLS methodology.

Table 5.6 and 5.9 show the results for the two-month price series. For TC2 the Johansen methodology suggests that the unbiasedness hypothesis holds. The VECM for TC2 does however suffer from non-normal residuals. The OLS methodology suggests that the unbiasedness hypothesis holds. Both methodologies thereby indicate that the unbiasedness hypothesis holds for the two-month TC2 series, but no conclusion will be drawn due to the problems with the residual diagnostics. Using the Johansen methodology the unbiasedness hypothesis is indicated not to hold for TC4. The result is the same when using the OLS methodology. For TC5 the unbiasedness hypothesis is showed to hold when using the OLS methodology.

Table 5.7 and 5.10 contain the results for the three-month price series. For TC2 the results from using the Johansen methodology support the unbiasedness hypothesis. The residuals of the VECM do however show signs of non-normality and autocorrelation for lag 12. The OLS results also support the unbiasedness hypothesis. I interpret this as signs of the unbiasedness hypothesis holding for the TC2 three-month series, but again do not draw any conclusions. For TC4 the unbiasedness hypothesis does not hold using the Johansen methodology. It does however hold when using the OLS methodology, but as mentioned I will not put too much weight on this result. For TC5 the unbiasedness hypothesis cannot be rejected when using the OLS methodology.

These results suggest that whether the unbiasedness hypothesis holds or not for clean tanker FFAs depends on the route and time to maturity. For TC2 the unbiasedness hypothesis is found to hold for the one-month price series, but the evidence is weaker when it comes to the two and three-month series. The reason for this is problems in the residual diagnostics of the VECMs. For TC4 the unbiasedness hypothesis does not hold for any of the investigated times to maturity. For TC5 the unbiasedness hypothesis is found to hold for all the investigated time series.

As mentioned earlier, VECMs may be used to gain knowledge about how the spot and futures series interact in both the short and long run. Information on the speed of convergence to the long run equilibrium can be found by looking at α_1 and α_2. Because the TC5 series are not cointegrated I will only comment on this relationship for the TC2 and TC4 series. From the estimated models in table 5.5 – 5.7 it can be observed that all coefficients of α_1 are insignificant. All the coefficients of α_2 except for the TC4 one-month series are however significant and positive. α_2 being positive implies that a positive error at period t-1 (i.e. $S_{t-1} > F_{t-2,t-1}$.) will be followed by a relative increase in the price of the futures in the next period. The futures thereby help restore the long-run equilibrium. This means that past errors affect the current forecasts of the underlying realised spot rates, i.e. the future prices, but not the spot prices themselves. Because any disequilibrium from the previous period is not carried forward to the current period, there is no sign of a systematic bias for either TC2 or TC4. This implies that the reason for rejecting the unbiasedness hypothesis for TC4 cannot be attributed to a consistent risk premium. The low liquidity of the TC4 futures might however be the reason for this shortcoming[22].

[22] The liquidity of TC4 relative to the other routes may be found in figure 2.1.

In line with the fact that the stationarity tests are inconclusive and that low liquidity might be the reason for rejecting the unbiasedness hypothesis for TC4, I recommend that the exercise of testing the unbiasedness hypothesis is repeated at a later time. A longer time period might then result in more consistent results with regards to the stationarity tests. The liquidity of the freight futures market will also hopefully increase with time.

6. The lead-lag relationship between futures prices and spot rates

In this part the price discovery properties of freight futures are further explored by investigating the lead-lag relationship between futures prices and spot rates. Relevant methodology is presented first, followed by descriptive statistics and stationarity tests for the data series. Johansen's framework is then employed to find that a VECM specification should be used for TC2 and TC4, while a VAR in levels is preferred for TC5. The lead lag relationship is then investigated by testing the significance of the coefficients of the estimated models, performing Granger causality tests and conducting impulse response analyses. The main findings are that futures prices are indicated to lead the spot rates for TC2 and TC5. For TC4 a bi-directional relationship is found, but the futures seem to have a leading role.

6.1 How to test the lead-lag relationship

The lead-lag relationship between future prices and spot rates refers to how well the two markets are linked and how fast one of the markets reflects new information relative to the other. Assuming that new information is available to both markets at the same time, the markets should theoretically react simultaneously. This might however not be the case in the real world as market frictions such as transaction costs, short-sale restrictions or flexibility might favour trading in one of the markets. In line with this, one market might lead the other, and thus work as a price discovery vehicle.

To assess the lead-lag relationship between futures prices and spot rates I will first use Johansen's test for cointegration to find the most appropriate model to use for each route[23]. If $\text{rank}(\Pi) = 0$ a VAR in first differences will be used. $\text{Rank}(\Pi) = 1$ implies that a VECM will be

[23] The Johansen framework and its implications regarding model selection was presented in chapter 5.1.

preferred, while $\mathrm{rank}(\Pi) = 2$ implies that using a VAR in level is the best model specification. The VECM specification is provided below[24].

$$\Delta S_t = \sum_{i=1}^{p} a_{s,i} \Delta S_{t-i} + \sum_{i=1}^{p} \theta_{s,i} \Delta F_{t-i,T} + \gamma_s Z_{t-1} + \varepsilon_{st}$$

$$\Delta F_t = \sum_{i=1}^{p} a_{f,i} \Delta S_{t-i} + \sum_{i=1}^{p} \theta_{f,i} \Delta F_{t-i,T} + \gamma_f Z_{t-1} + \varepsilon_{ft}$$

where $a_{s,i}$, $a_{f,i}$, $\theta_{s,i}$ and $\theta_{f,i}$ are coefficients for the lagged terms. γ_s and γ_f are coefficients measuring the proportion of last period's equilibrium error which is corrected for, also known as the adjustment speed of the spot and futures prices to their long-term equilibrium. These cross-sectional coefficients of the model which contain information on the lead-lag relationship of the variables. $Z_{t-1}$is the error correction term. ε_{st} and ε_{ft} are white-noise error terms which have the 2×2 variance-covariance matrix Σ.

If the spot rates and futures prices are cointegrated, the variables must either move simultaneously or one must lead the other (Granger, 1988). Granger causality must therefore exist in at least one direction. If the series are not found to be cointegrated a lead-lag relationship might not exist.

To test the lead-lag relationship I will first look at the significance of the relevant coefficients individually. I will then employ Granger causality tests. These are Wald-tests which investigate the joint significance of the lagged terms of variables other than the dependent variable with respect to the dependent variable.

After investigating the lead-lag relationship I will conduct impulse response analyses. These will measure the reaction of the spot and futures prices to imposed shocks of one standard error in the

[24] This model can easily be transformed into a VAR in first differences by removing the error correction term. A VAR in levels may be represented by additionally removing the differencing operators.

models. These analyses are conducted to gain more insight of the causal relationship between spot and futures prices.

6.2 Properties of the data series

The data used to investigate the lead-lag relationship consist of daily data of spot and futures prices from the period March 1st 2004 to September 30th 2009. Two series of futures prices were collected together with the spot prices for each route. The two futures price series contain futures prices of the contract closest to maturity and second closest to maturity, respectively. Because the futures expire as time goes, the contracts have been rolled over at the 15th of the expiry month and the month before expiry, respectively. The 15th was chosen because this is the last trading day of the futures in the first part of the observation period[25]. According to Imarex brokers the liquidity of the contracts is at its best in the first half of the expiry month. It then falls sharply as maturity approaches. The rollover procedure thereby ensures that the futures prices used in this analysis are the most liquid available. Dates where the spot and/or futures price is not reported have been removed from the dataset. This is done to obtain spot and futures series with corresponding observations, and thereby enable cointegration methods to be used. The dataset of TC2 consists of 1398 observations, while the datasets of TC4 and TC5 include 1403 observations.

When conducting the analysis I first tried to base it on the series of futures prices closest to maturity. ARIMA (3,1,0) regressions which included a dummy variable taking the value one each time the futures were rolled and zero ordinarily did however show that all the futures price series then contained structural breaks[26]. As this might lead to biased results, perpetual series of futures prices were generated for all routes[27]. These series were based on the weighted average of

[25] See footnote 18.

[26] The ARIMA model is presented in chapter 7.1.1. The results from these tests are provided in Appendix II.

[27] Kavussanos and Visvikis (2004) suggest this method to avoid the problem of price jumps.

the two futures series referred to in the paragraph above for each route. The series were weighted according to their days to maturity, yielding one series of futures prices with a constant time to maturity for each route.

The fact that the flat rates change every year might bias the results of the lead-lag analyses toward future prices leading spot rates. This is because the future prices will reflect the new flat rates before the turn of a year, while the spot prices will not reflect these until a new year has begun[28]. To avoid the result being biased all futures and spot prices have therefore been rebased to the flat rate of 2009. All series were also transformed in natural logarithms. Descriptive statistics of the first differences of the series used in this analysis are presented in table 6.1.

Table 6.1: Descriptive statistics and tests for normality

	Daily Spot Rates and Forward Prices							
	N	Mean	Max	Min	StdDev	Skewness	Kurtosis	Jarque-Bera
TC2 DSPOT	1397	-0.000592	0.164150	-0.200990	0.027833	0.773980	10.803550	3684.10
TC2 PERPRB	1397	-0.000370	0.147980	-0.249990	0.028555	-0.303841	9.403303	2408.17
TC4 DSPOT	1402	-0.000480	0.180320	-0.189240	0.025136	0.701089	14.305040	7580.74
TC4 PERPRB	1402	-0.000377	0.123060	-0.143530	0.025681	0.168367	5.164608	280.34
TC5 DSPOT	1402	-0.000075	0.223140	-0.143100	0.025076	0.866310	13.244540	6306.23
TC5 PERPRB	1402	-0.000116	0.185710	-0.190440	0.028797	0.190137	8.927550	2060.97

TCX DSPOT are daily spot prices series for TC2, TC4 and TC5.
TCX PERPRB are perpetual futures prices series with a constant time to maturity of 33 days which are rebased to the flat rate level of 2009.

N is the number of observations of each time series. Max is the maximum value of the series and min is the minimum value.
Mean, standard deviation[2], skewness and kurtosis are the first, second, third and forth standardized moments of the time series.
The Jarque-Bera is a test for normality. The null hypothesis is that the series have skewness and kurtosis similar to a normal distribution.
The test statistic follows a chi-square distribution with two degrees of freedom under the null hypothesis. The 5% critical value is 5.9915.
The null hypothesis is rejected if the test statistic is greater than the critical value. Rejection indicates sign of non-normality.
Values representing rejected null hypothesis are reported in red.
All series have been rebased to the flat rate levels of 2009 and reflect the logarithmic first differences of these series.

As expected, the mean is close to zero for all the series. It is also negative which implies that prices have fallen over the period. The standard deviation is almost 3% for all the series, and both the minimum and maximum values are large. As the series are based on daily data, this is

[28] This phenomenon is explained in chapter 2.3.

not negligible and can be seen as an indication of the high volatility in tanker shipping. The standard deviations of the futures series are higher than those of the spot series for all routes. This is line with the results of Adland et al. (2009), and is thought to be an implication of the elastic expectations hypothesis[29]. For the spot prices TC2 has the highest standard deviation. The standard deviations of TC4 and TC5 are virtually equal. The daily spot rates of TC2 therefore seem to be more unstable than those of TC4 and TC5. The reason for this might be that TC2 is often very unstable during the Atlantic hurricane season. Another reason suggested by physical brokers is that the market seems more concerned with the last done rates when discussing freight for TC4 and TC5 than for TC2. For the futures series TC5 has the largest standard deviation. This might be because TC5 is based on larger ships than TC2 and TC4. The volatility of TC5 is theoretically thought to spill over to TC4 in the spot market[30], but for the futures market this effect seems to be smaller than the forces which drive the volatility of TC2.

The skewness is slightly positive for all the series except for the TC2 futures. Most of the series are thereby right skewed, meaning that the distributions of all the time series except for the TC2 futures series have a relatively long tail to the right. The high kurtosis shows that the distributions are leptokurtic, meaning that they have a distinct peak near the mean, decline rapidly and have heavy tails. This can be attributed to the many observations around zero and some outliers. The high Jarque-Bera statistics can be regarded as a consequence of the factors influencing the skewness and kurtosis. This test indicates that none of the series are normally distributed. I have employed ADF, PP and KPSS tests to investigate the stationarity of the series. The results of these tests are provided in table 6.2.

[29] This hypothesis was presented in chapter 5.2.

[30] A more detailed explanation is provided in chapter 5.2.

Table 6.2: Tests for stationarity

	Daily Spot Rates and Forward Prices					
	ADF lvl (lags)	ADF 1st diff (lags)	PP lvl (BW)	PP 1st diff (BW)	KPSS lvl (BW)	KPSS 1st diff (BW)
TC2 DSPOT	-3.436003 (1)	-19.06937 (0)	-3.297080 (18)	-19.24167 (5)	0.505115 (30)	0.032151 (18)
TC2 PERPRB	-2.689422 (1)	-29.37192 (0)	-2.263046 (1)	-29.16697 (9)	0.549396 (30)	0.062950 (4)
TC4 DSPOT	-4.165379 (8)	-7.460681 (7)	-2.976547 (25)	-35.37327 (24)	0.993787 (30)	0.024874 (25)
TC4 PERPRB	-1.810884 (1)	-31.44120 (0)	-2.322801 (18)	-33.05249 (17)	1.064627 (30)	0.051403 (18)
TC5 DSPOT	-3.577689 (4)	-11.12418 (3)	-3.305316 (24)	-31.97382 (22)	0.187215 (30)	0.023155 (24)
TC5 PERPRB	-2.395263 (1)	-33.46488 (0)	-2.940822 (16)	-34.64453 (14)	0.217058 (30)	0.032300 (16)

The ADF tests refer to Augmented Dickey Fuller tests with one intercept included. The null hypothesis is non-stationary.
Lag length set to max 21 which equals the average trading days of a month and is automatically selected using Schwarz Information Criterion.
ADF is a one-tailed t-test where the test statistic follows the distribution calculated by Dickey and Fuller under the null hypothesis.
A 5% (10%) level of significance has a critical value of approximately -2.8634 (-2.5678) (depending on the number of lags) (MacKinnon, 1996)
The null hypothesis is rejected if the test statistic is smaller than the critical value, and if this is the case the series is said to be stationary.

The PP tests refer to the Phillips-Perron test with one intercept included. The null hypothesis is that a series is non-stationary.
The spectral estimation method is the Berlett kernel method and the bandwidth selection is the Newey-West Bandwidth.
The test statistic follows the same asymptotic distributions as the ADF test statistic and normalized bias statistics under the null hypothesis.
The critical values are therefore the same as for the ADF test.
The null hypothesis is rejected if the test statistic is smaller than the critical value, and if this is the case the series is said to be stationary.

The KPSS tests refer to the Kwiatkowski, Phillips, Schmidt and Shin test with an intercept included.
The spectral estimation method is the Berlett kernel method and the bandwidth selection is the Newey-West Bandwidth.
KPSS is a right tailed test, and the test statistic is a LM statistic which converges to a function of a standard Brownian motion under the null hypothesis.
The critical value is 0.463 (0.347) at a 5% (10%) level of significance and is calculated from the asymptotic distribution of a standard Brownian motion.
The null hypothesis is rejected if the test statistic is larger than the critical value, and if this is the case the series is said to be non-stationary.

Values representing rejected null hypothesis using the 5 % critical value are reported in red.
Values representing rejected null hypothesis using the 10 % critical value are reported in purple.

For TC2 the ADF and PP tests indicate that the spot series is stationary in levels. The KPSS test however contradicts this result. The futures series is found to be stationary in levels using the ADF test at a 10 percent level of significance. It is however found to be non-stationary in levels using the PP and KPSS tests. In line with this it possible that the TC2 series are both I(0) or I(1). For TC4 the results are the same as for TC2 with one exemption, the futures series is not indicated to be stationary in levels using the ADF test. The results for the TC4 series are therefore also not clear. The spot series of TC5 is found to be stationary in levels using the ADF, PP and KPSS test. The futures series is found to be non-stationary in levels using the ADF and KPSS. This result is however contradicted by the PP test. For TC5 the spot series is therefore probably I(0) while the futures series is indicated to be I(1). Overall, these results are in line with the results of the monthly data series used when testing the unbiasedness hypothesis. In line with this I will use Johansen's test for cointegration to obtain the most appropriate model for each route. I will then use the models suggested by this test to investigate the lead-lag relationship between spot and futures prices.

6.3 Investigating the lead-lag relationship between spot and futures prices

Johansen's cointegration framework is now employed to test for cointegration. The models used include a constant in the cointegrating relationship, but not outside and no trend. The relevance of the constant in the cointegrating relationship will be tested. To make sure the VECMs are well specified in terms of lag length VAR models are estimated. I use SIC (Schwartz, 1978) to determine the number of lags to include. Using SIC and transforming the lag length it is found that 2, 2 and 3 lags are appropriate in the VECMs for TC2, TC4 and TC5, respectively[31]. These lag lengths did however turn out to yield autocorrelated residuals. To remedy this problem the lag lengths were gradually increased to 5, 9 and 8 lags, respectively. Using these lag-lengths the Johansen (1988) procedure was employed to test the cointegrating relationship between the spot and futures prices for the three routes. The results are presented in the table below.

[31] A further explanation is provided in chapter 5.3.

Table 6.3: Cointegration tests for spot prices and the perpetual futures contracts

	Trace Statistics		
r	**TC2 (5 lags)**	**TC4 (9 lags)**	**TC5 (8 lags)**
0	34.83938	46.09745	58.19454
	(20.26184)	(20.26184)	(20.26184)
1	5.27433	5.80075	12.29626
	(9.16455)	(9.16455)	(9.16455)
Rank	1	1	2

	Max Statistics		
r	**TC2 (5 lags)**	**TC4 (9 lags)**	**TC5 (8 lags)**
0	29.56505	40.29670	45.89829
	(15.89210)	(15.89210)	(15.89210)
1	5.27433	5.80075	12.29626
	(9.16455)	(9.16455)	(9.16455)
Rank	1	1	2

r is the number of cointegrating vectors.

For the λ_{trace} statistic the null is that rank is smaller than or equal to r, while the alternative is that rank exceeds r.

For the λ_{max} statistic the null is that rank equals r, while the alternative is that rank equals r+1.

The 5% critical values are those of MacKinnon-Haug-Michelis (1999) and reported in parenthesis.

The rejection rule is to reject the null hypothesis if the test statistic is larger than the critical value.

Red colour denotes rejection of the null hypothesis.

Rank describes the number of cointegrating vectors implied by the hypothesis tests.

For TC2 and TC4 the results of the cointegration tests indicate that there is one cointegration relationship between the spot and perpetual futures series. A VECM will therefore be used to investigate the lead-lag relationship for these routes. For TC5 the cointegration test indicates that a VAR in levels should be employed.

To obtain a correct specification for the VECM of TC2 and TC4 a test of whether the cointegrating relationship can be restricted to be the lagged basis, i.e. $\beta_1 = 0$ and $\beta_2 = -1$, was conducted. This was done using a log-likelihood framework[32].

[32] A further explanation can be found in chapter 5.1.

Tests of restricting ECT to lagged basis

	H0: β1=0, β2= -1
TC2	4.33600
	(5.99146)
TC4	11.87000
	(5.99146)

5% critical value reported in parenthesis.
Green colour denotes evidence of unbiasedness.
Red colour denotes rejection of the null hypothesis.

The test showed that the cointegrating relationship of the TC2 VECM is not significantly different from the lagged basis at a 5 percent level of significance. For TC4 the cointegrating relationship is significantly different from being the lagged basis. In the following analysis I therefore employ this restriction only to TC2. The results regarding model specifications are in line with those obtained when testing the unbiasedness hypothesis. The estimated models and residual diagnostics are presented below.

Table 6.4: Estimated coefficients for the models.

	Model Specification											
	TC2				TC4				TC5			
	ΔSt		ΔFt		ΔSt		ΔFt		St		Ft	
Zt-1	γs	-0.02556	Γf	-0.01029	γs	-0.03747	γf	-0.02299	C	-0.05053	C	0.02921
		(0.00488)		(0.00649)		(0.00609)		(0.00699)		(0.01366)		(0.01865)
		[-5.23231]		[-1.58476]		[-6.15244]		[-3.29057]		[-3.69961]		[1.56596]
ΔSt-1	αs,1	0.32466	αf,1	0.07928	αs,1	0.05273	αf,1	0.14961	αs,1	1.05613	αf,1	-0.01774
		(0.02983)		(0.03965)		(0.02786)		(0.03196)		(0.02775)		(0.03790)
		[10.8845]		[1.99949]		[1.89269]		[4.68069]		[38.0627]		[-0.46822]
ΔSt-2	αs,2	0.00964	αf,2	0.01308	αs,2	0.06963	αf,2	-0.02983	αs,2	-0.01283	αf,2	-0.00265
		(0.03115)		(0.04141)		(0.02777)		(0.03187)		(0.04041)		(0.05519)
		[0.30951]		[0.31585]		[2.50699]		[-0.93598]		[-0.31744]		[-0.04795]
ΔSt-3	αs,3	0.01171	αf,3	-0.00519	αs,3	0.02380	αf,3	-0.01726	αs,3	0.01113	αf,3	-0.00390
		(0.03103)		(0.04125)		(0.02766)		(0.03174)		(0.04030)		(0.05505)
		[0.37725]		[-0.12590]		[0.86041]		[-0.54392]		[0.27613]		[-0.07083]
ΔSt-4	αs,4	-0.01137	αf,4	-0.04430	αs,4	0.02869	αf,4	-0.02346	αs,4	-0.04234	αf,4	0.06435
		(0.03076)		(0.04089)		(0.02765)		(0.03173)		(0.04020)		(0.05490)
		[-0.36976]		[-1.08356]		[1.03754]		[-0.73952]		[-1.05327]		[1.17207]
ΔSt-5	αs,5	0.04185	αf,5	-0.05248	αs,5	-0.00849	αf,5	-0.02683	αs,5	-0.03071	αf,5	-0.04460
		(0.02658)		(0.03533)		(0.02755)		(0.03160)		(0.04020)		(0.05490)
		[1.57477]		[-1.48530]		[-0.30825]		[-0.84910]		[-0.76404]		[-0.81235]
ΔSt-6	αs,6	-	αf,6	-	αs,6	0.00624	αf,6	0.03018	αs,6	0.01108	αf,6	-0.01229
						(0.02717)		(0.03117)		(0.04016)		(0.05485)
						[0.22982]		[0.96812]		[0.27590]		[-0.22400]
ΔSt-7	αs,7	-	αf,7	-	αs,7	0.06103	αf,7	0.00270	αs,7	-0.04926	αf,7	0.03630
						(0.02702)		(0.03100)		(0.04010)		(0.05477)
						[2.25903]		[0.08697]		[-1.22855]		[0.66280]
ΔSt-8	αs,8	-	αf,8	-	αs,8	0.13878	αf,8	0.07093	αs,8	0.01348	αf,8	-0.03894
						(0.02648)		(0.03039)		(0.02573)		(0.03514)
						[5.24012]		[2.33416]		[0.52402]		[-1.10809]
ΔSt-9	αs,9	-	αf,9	-	αs,9	0.01850	αf,9	-0.06135	αs,9	-	αf,9	-
						(0.02654)		(0.03045)				
						[0.69720]		[-2.01471]				
ΔFt-1	ϑs,1	0.26936	ϑf,1	0.19148	ϑs,1	0.06469	ϑf,1	0.10682	ϑs,1	0.15508	ϑf,1	1.08989
		(0.02322)		(0.03086)		(0.02581)		(0.02961)		(0.02033)		(0.02777)
		[11.6019]		[6.20416]		[2.50639]		[3.60744]		[7.62760]		[39.2494]
ΔFt-2	ϑs,2	0.11478	ϑf,2	-0.01365	ϑs,2	0.08379	ϑf,2	-0.08151	ϑs,2	-0.07315	ϑf,2	-0.03810
		(0.02420)		(0.03217)		(0.02575)		(0.02954)		(0.02959)		(0.04041)
		[4.74309]		[-0.42414]		[3.25439]		[-2.75957]		[-2.47244]		[-0.94282]
ΔFt-3	ϑs,3	0.04616	ϑf,3	-0.05423	ϑs,3	0.03040	ϑf,3	0.02005	ϑs,3	0.03520	ϑf,3	-0.04372
		(0.02424)		(0.03222)		(0.02597)		(0.02980)		(0.02966)		(0.04051)
		[1.90423]		[-1.68287]		[1.17044]		[0.67289]		[1.18680]		[-1.07923]
ΔFt-4	ϑs,4	0.05543	ϑf,4	0.00596	ϑs,4	0.11882	ϑf,4	0.05459	ϑs,4	-0.02589	ϑf,4	0.00269
		(0.02411)		(0.03205)		(0.02592)		(0.02974)		(0.02967)		(0.04052)
		[2.29911]		[0.18611]		[4.58355]		[1.83530]		[-0.87258]		[0.06646]
ΔFt-5	ϑs,5	0.08674	ϑf,5	0.02991	ϑs,5	0.08777	ϑf,5	0.02607	ϑs,5	0.07092	ϑf,5	0.05549

		(0.02359)		(0.03135)		(0.02606)		(0.02990)		(0.02964)		(0.04048)
		[3.67749]		[0.95407]		[3.36820]		[0.87192]		[2.39286]		[1.37095]
ΔFt-6	$\vartheta_{s,6}$	-	$\vartheta_{f,6}$	-	$\vartheta_{s,6}$	0.01723	$\vartheta_{f,6}$	-0.01568	$\vartheta_{s,6}$	-0.09066	$\vartheta_{f,6}$	-0.03131
						(0.02604)		(0.02988)		(0.02971)		(0.04058)
						[0.66164]		[-0.52480]		[-3.05161]		[-0.77150]
ΔFt-7	$\vartheta_{s,7}$	-	$\vartheta_{f,7}$	-	$\vartheta_{s,7}$	0.05261	$\vartheta_{f,7}$	0.00626	$\vartheta_{s,7}$	0.02630	$\vartheta_{f,7}$	-0.05915
						(0.02595)		(0.02978)		(0.02981)		(0.04071)
						[2.02707]		[0.21008]		[0.88227]		[-1.45290]
ΔFt-8	$\vartheta_{s,8}$	-	$\vartheta_{f,8}$	-	$\vartheta_{s,8}$	-0.00085	$\vartheta_{f,8}$	-0.06831	$\vartheta_{s,8}$	-0.04434	$\vartheta_{f,8}$	0.03761
						(0.02543)		(0.02918)		(0.02160)		(0.02951)
						[-0.03347]		[-2.34104]		[-2.05233]		[1.27445]
ΔFt-9	$\vartheta_{s,9}$	-	$\vartheta_{f,9}$	-	$\vartheta_{s,9}$	0.06472	$\vartheta_{f,9}$	0.08446	$\vartheta_{s,9}$	-	$\vartheta_{f,9}$	-
						(0.02547)		(0.02922)				
						[2.54157]		[2.89071]				
R2		0.44400		0.065964		0.278517		0.084248		0.995744		0.986979

Numbers in parenthesis are standard errors and numbers in brackets are t-statistics.
Red colour denotes rejection of the null hypothesis at a 5% level of significance.
For TC2 a VECM where the ECT is restricted to be the lagged basis is used; ECT = St-1 - Ft-1.
For TC4 a VECM with ECT = St-1 - 1.133095Ft-1 + 0.7337 is used. The t-values of the ECT are -19.4983 and 2.47984, respectively.
For TC5 a VAR in levels is used.

For TC2 and TC4 the coefficients of the ECT provide some insight to the adjustment towards the long-term equilibrium. For TC2 both these coefficients are negative, but the coefficient of the futures equation is not significant. This implies that only the spot prices react to correct a shock to the system in order to reach the long-term equilibrium. The negative coefficient means that a positive (negative) shock leads the spot price to decrease (increase). For TC4 both the coefficients of the spot and futures equation are negative and significant. Both the spot and futures price is thereby thought to react to a shock to the system in order to reach the long-term equilibrium.

The number of own lags which are statistically significant seem to be fairly similar for all the systems. At the same time more cross–market lags do however seem necessary for the spot equations than the futures equations. This may be interpreted as an indication of the futures market leading the spot market.

Table 6.5: Residual diagnostics

			Residual Diagnostics			
Route	**Residuals**	**Normality**	**Normality***	**Q(12)**	**Q(12)***	**White***
TC2	εst	1099	1503	5.20560	31.68370	219
	εft	404		6.81810		
	5%cv	(5.99146)	(9.48773)	(21.02607)	(41.33714)	(85.96491)
TC4	εst	1147	1297	4.95420	14.22299	230
	εft	150		0.50720		
	5%cv	(5.99146)	(9.48773)	(21.02607)	(21.02607)	(139.92077)
TC5	εst	908	1714	8.58650	20.97884	204
	εft	806		2.00140		
	5%cv	(5.99146)	(9.48773)	(21.02607)	(26.29623)	(119.87094)

Normality refers to the Doornik and Hansen (1994) test for normality.
Q(12) denotes the Ljung-Box test for autocorrelation of the first 12 lags.
Q(12)* is the Ljung-Box statistic.
White refers to Whites test for heteroskedasticity.
All test statistics follow a X^2 distribution asymptotically.
Asterixes denote that the respective test is bivariate.

The residual diagnostics show that the residuals are not normally distributed. Results from the Doornik and Hansen (1994) test for normality indicate that the reason for this is mostly excess kurtosis[33]. The Central Limit Theorem does however state that the deviation from normality has very little effect on inferences for large samples. Both the univariate and multivariate Ljung-Box tests for autocorrelation were performed up to 21 lags and no signs of autocorrelation were detected. White's test for general heteroskedasticity indicated that all the models suffer from heteroskedastic residuals. Plots of the residuals and explanatory variables did however indicate that the heteroskedasticity was not of the unconditional form. At the same time both correlograms of the squared residuals and the ARCH LM test of Engle (1982) indicated conditional heteroskedasticity. In line with the fact that conditional heteroskedasticity does not impose problems for large sample sizes, the heteroskedasticity problem was not prioritized because the software did not support using more advanced methods.

[33] These results are provided in Appendix II.

Granger causality tests were employed to test the joint significance of the lagged futures prices on the current spot price and vice versa. The results from these tests are presented below.

Table 6.6: Results from the Granger Causality Tests

Granger Causality Tests			
Dependent variable	**Excluded variable**	**Chi-sq**	**CV**
TC2 DSPOT	TC2 PERP	166.39160	(11.07050)
TC2 PERP	TC2 DSPOT	10.04639	(11.07050)
TC4 DSPOT	TC4 PERP	46.82795	(16.91898)
TC4 PERP	TC4 DSPOT	33.13487	(16.91898)
TC5 DSPOT	TC5 PERP	206.81160	(15.50731)
TC5 PERP	TC5 DSPOT	10.06608	(15.50731)

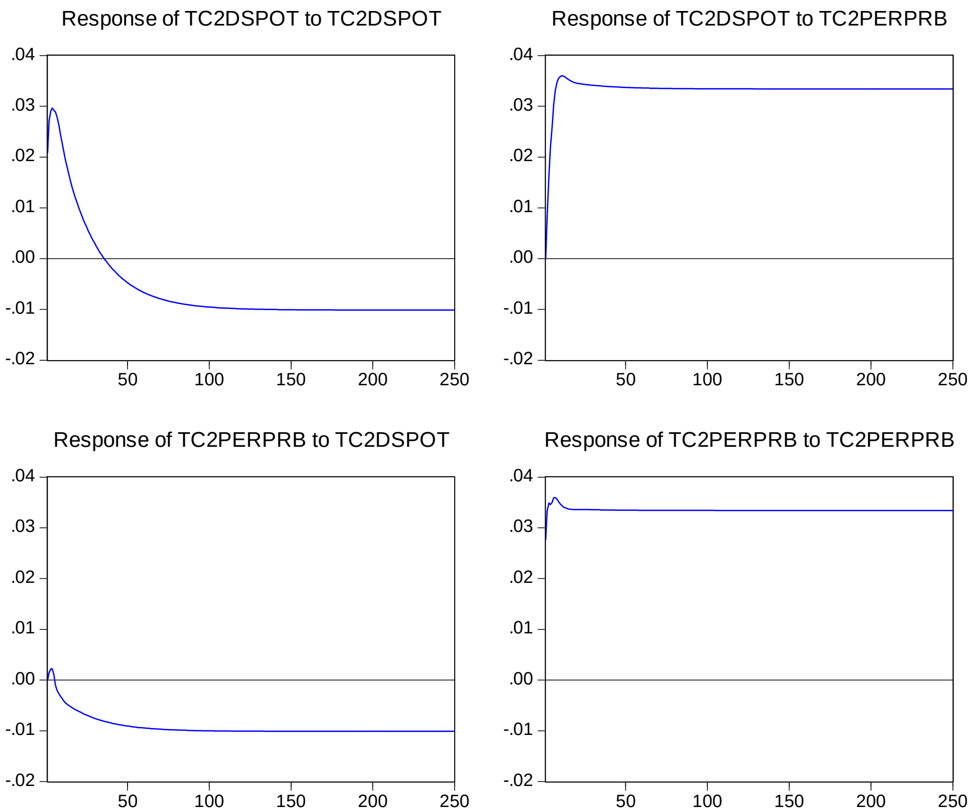
The null is that all the lagged terms of the excluded variable is insignificant.
The test statistic follows the X^2 distribution under H_0.
Numbers in parenthesis are critical values at 5% level of significance.
Red colour denotes rejection of the null hypothesis.

The Granger causality tests show that there is a bi-directional relationship between spot and futures prices for TC4. For TC2 and TC5, however, only the lagged futures prices seem to affect the current spot prices. Relatively, the test statistics also imply that the future to spot relationship is stronger than the spot to future relationship. This indicates that the futures prices lead the spot prices.

6.4 Impulse Response Analyses

Impulse response analyses were conducted to gain a further understanding of the causal relationship between spot and futures prices. These show the reaction of the spot and futures with regards to a shock in the estimated system. Impulse response functions for impulses of one standard deviation to the spot and futures series of TC2, TC4 and TC5 are provided in figure 6.1, 6.2 and 6.3, respectively. For these figures the horizontal axis represents number of days after the shock, while the vertical axes represent the magnitude of the shock. The reaction of the spot rates (upper windows) and futures prices (lower windows) to a shock in the spot rates can be observed to the left, while the reaction to a shock in the future prices can be observed to the right. Bands of ± 2 standard deviations are only available for TC5 as this function is not supported by the software for VECMs.

Figure 6.1: Impulse Response Functions for TC2

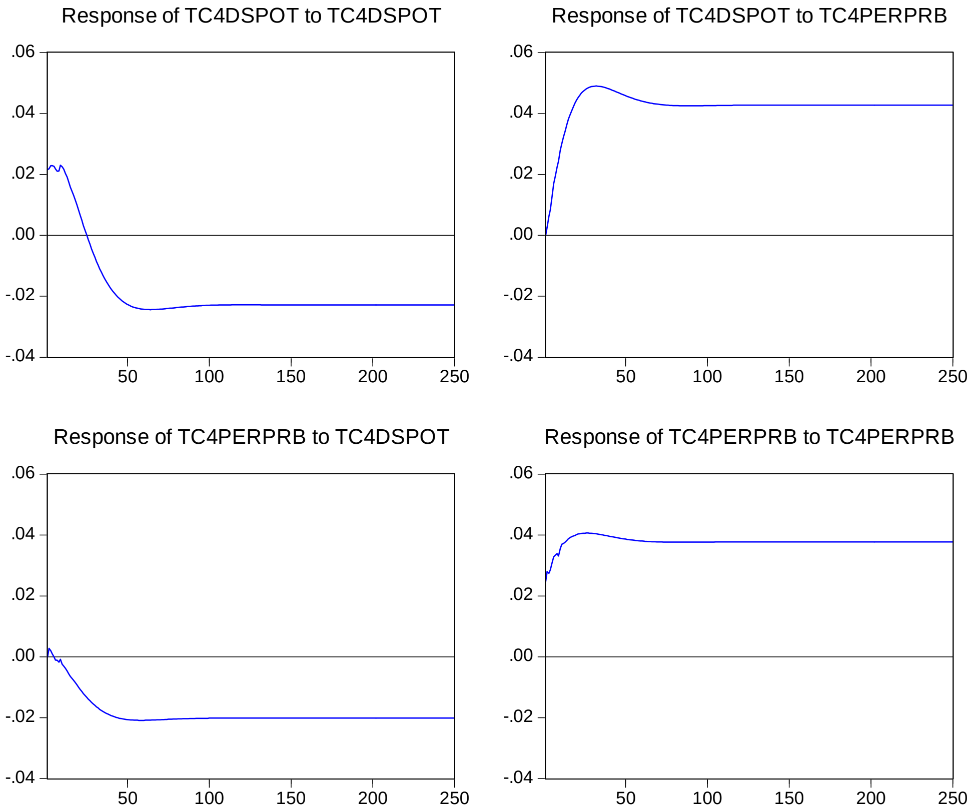


For TC2 a shock to the spot rates makes the spot rates increase before they start approaching the new long term equilibrium. This is reached after approximately 150 days. The futures prices however only slightly react to a shock in the spot rates, and start approaching the new long term equilibrium almost instantly. The futures prices also stabilize at the new equilibrium before the spot rates.

A shock to the futures prices makes the futures prices increase and overshoot the new equilibrium, before they start approaching it asymptotically after approximately 15 days. The spot rates follow the futures prices fairly quickly, and also overshoot the new equilibrium, but it does however take some more time for them to reach it.

The two paragraphs above indicate that the futures prices reach the new long run equilibrium before the spot prices irrespective of in which series the shock occurs. It also seems like spot prices copy the path of the futures prices, and do so in a lagged way. The futures are therefore thought to have a leading role for TC2. This is in line with the results of the Granger Causality tests.

Figure 6.2: Impulse Response Functions for TC4

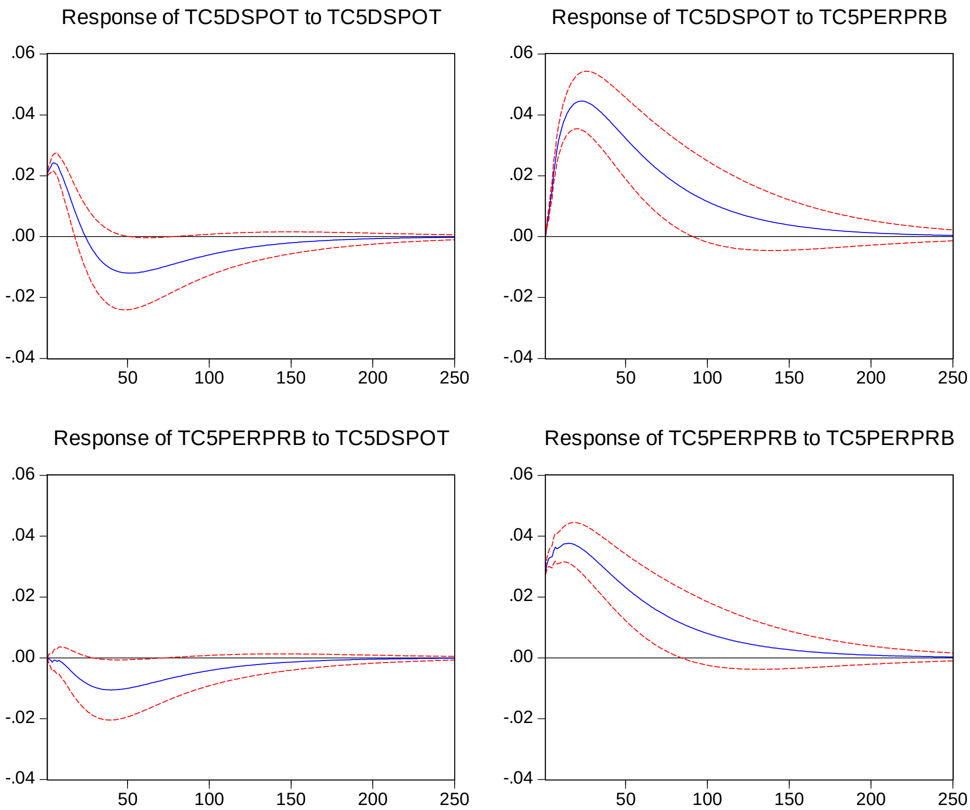


For TC4 a shock to the spot rates make the spot rates somewhat wobbly before they start to decrease after approximately 10 days. The spot rates then overshoot their new steady state before reaching it after approximately 100 days. A shock to the spot rates make the futures prices start declining before the spot prices do so. The futures prices also experience some overshooting of the new steady state, although less than the spot rates, and reach their new steady state after approximately 80 days. The spot rates and futures prices do not reach a new long term equilibrium, which implies that the rates do not converge upon each other. This is in line with the results from testing the unbiasedness hypothesis.

A shock to the futures prices makes the futures prices increase and overshoot their new long run steady state. They then decline a bit too much and therefore need to increase again to reach their steady state, which happens after approximately 100 days. The same seems to happen with the spot rates, but these overshoot their steady state to a larger extent than the futures prices. Before reaching their new steady state the spot and future prices seem to move in parallel. This might be an indication of a bi-directional relationship between the spot rates and futures prices. Again, spot rates and futures prices do not reach a new long-term equilibrium.

This indicates that the futures prices lead the spot rates. The futures prices do however spend more time reaching their new steady state for TC4 than TC2, and in addition to this also move in parallel with the spot rates when absorbing a shock to the futures prices. These facts are indications of a bi-directional relationship. The finding of a bi-directional relationship between spot rates and futures prices is in line with the Granger Causality tests. In line with the findings here, these do however also indicate that the futures prices have more of a leading role than the spot rates.

Figure 6.3: Impulse Response Functions for TC5



For TC5 a shock to the spot rates makes the spot rates slightly increase before they start decreasing. They overshoot the new steady state until approximately 50 days after the shock, and then asymptotically approach for approximately another 200 days. The futures price series does not seem to follow the spot rates after the shock, but rather starts decreasing a few days after it occurs. The futures overshoot the new steady state, but starts approaching it after approximately 40 days. Both the fact that the futures series do not follow the spot rates immediately after the shock and that the futures prices series start approaching the steady state before the spot rates, are indications of the futures prices leading the spot rates.

A shock to the futures prices makes the futures prices increase before they start decreasing and thereby approaching the steady state. This is reached after approximately 250 days. The spot rates follow the futures prices in the beginning, but experience a higher peak and start to decline after the futures prices. They then lag the futures prices down towards the steady state. This indicates that the futures prices lead the spot rates, which is in line with the Granger Causality tests.

The lead-lag analyses conducted here indicate that the futures prices lead the spot rates for TC2 and TC5. For TC4 a bi-directional relationship is found, but the futures seem to have a leading role. The futures are thought to lead the spot rates due to lower transaction costs, easier access to shorting and a higher degree of flexibility in the positions. Futures prices thereby seem informationally more efficient than spot rates in the clean tanker market, and may be used as price discovery vehicles. In line with this, the next chapter focuses on the abilities of freight futures when it comes to forecasting the underlying spot rates.

7. Freight futures and their ability to forecast the underlying spot rates

In this chapter the forecasting performance of end-of-month freight futures with regards to the underlying spot rates is investigated. This is done by comparing the forecasting accuracy of freight futures prices one-, two- and three months before maturity to the accuracy of forecasts generated by various time-series models and a random walk. The time-series models used are ARMA and ARIMA models, VARs in levels and first differences and VECMs. Multivariate time-series models are found to outperform the futures prices for the one-month horizon of all the routes and the two-month horizon of TC4 and TC5. For the two-month horizon of TC2 the results are inconclusive regarding the futures prices and a VECM. For the three-month horizon the futures prices provide the best forecasts for TC2, while the results are mixed between the futures prices and VECMs for TC4 and TC5. Overall the results indicate that future prices contain valuable information regarding future spot rates. The forecasts implied by futures prices are found to perform well compared to more complex time-series models, and also seem to improve relatively to those of time series-models when the forecasting horizon is increased.

7.1 Introducing the time-series models and measures of forecasting accuracy

Freight futures may have forecasting abilities regardless of whether the unbiased hypothesis holds. These price discovery properties may be investigated by comparing futures prices to forecasts generated by time-series models. A random walk, the ARMA model, the ARIMA model, the VAR model in levels, the VAR model in first differences and a VECM will be employed to perform this exercise. The ARIMA model will be introduced first, followed by the ARMA model and a random walk as these may be explained based on the ARIMA. The multivariate models are then presented. Finally, three measures of forecasting accuracy are introduced.

7.1.1 The models

The autoregressive integrated moving average model, ARIMA, is a univariate time-series model which combines an autoregressive process and a moving average process with a differencing factor. An autoregressive process is a process where the current value of the dependent variable is determined by the past values of that variable plus a white noise error term. A moving average process is a process where the current value of the dependent variable is determined by a constant plus the current and previous white noise error terms. The differencing is performed in order to make the time-series stationary. The model is often denoted ARIMA(p,d,q) where p, d and q refers to the order of the autoregressive, integrated and moving average parts of the model, respectively. An ARIMA (p,1,q) model for the spot price is presented below.

$$\Delta S_t = \alpha_{10} + \sum_{i-1}^{p} \alpha_{1i} \Delta S_{t-i} + \sum_{j-1}^{q} \beta_{1i} \varepsilon_{1t-j} + \varepsilon_{1t}, \qquad \varepsilon_{1t} \sim iid\ N(0, \sigma_1^2)$$

where the AR term is the first term which is summarized and the MA term is the second term which is summarized. The differenced spot prices signal that the model is differenced once, and therefore represent the integrated part of the model.

In addition to the ARIMA model I will employ an ARMA model. This is the same as an ARIMA (p,0,q), which corresponds to a univariate model in levels. This model specification is correct if a time series is stationary in levels. The reason for including ARIMA and ARMA models is to establish whether using historic spot rates only is sufficient when forecasting future spot rates, and thereby whether futures prices contain valuable information when generating forecasts.

The random walk (RW) is a special case of the ARIMA model, denoted ARIMA (0, 1, 0). Estimation is not necessary for this model as it assumes that the current price is the best estimate of the price of the next period. This model is included for evaluation purposes.

$$y_t = y_{t-1} + \varepsilon_t, \qquad \varepsilon_t \sim iid\ N(0, \sigma_2^2)$$

A VAR in levels is a multivariate model which simultaneously explains the current values of the included variables. This type of VAR model is preferred if the series are stationary in levels. Using a VAR in levels the spot and futures prices is explained by lagged values of the same variables. The VAR in levels which will be used to create forecasts is presented below.

$$S_t = \mu_s + \sum_{i=1}^{p} a_{s,i} S_{t-i} + \sum_{i=1}^{p} \theta_{s,i} F_{t-i,T} + \varepsilon_{st}$$

$$F_t = \mu_f + \sum_{i=1}^{p} a_{f,i} S_{t-i} + \sum_{i=1}^{p} \theta_{f,i} F_{t-i,T} + \varepsilon_{ft}$$

where μ_s and μ_f are constants. $a_{s,i}$, a, $\theta_{s,i}$ and $\theta_{f,i}$ are parameters. ε_{st} and ε_{ft} are white-noise error terms. The model specification used for the bivariate models in this part of the thesis differs slightly from that used when testing the unbiasedness hypothesis. This is in line with the fact that the data points which are to be forecasted should be withheld when estimating the models in this part. As the existence of a trend thereby cannot be excluded, a constant in the model is included.

A VAR in first differences is appropriate if differencing is necessary to make the variables stationary. The VAR in first differences is presented below.

$$\Delta S_t = \mu_s + \sum_{i=1}^{p} a_{s,i} \Delta S_{t-i} + \sum_{i=1}^{p} \theta_{s,i} \Delta F_{t-i,T} + \varepsilon_{st}$$

$$\Delta F_t = \mu_f + \sum_{i=1}^{p} a_{f,i} \Delta S_{t-i} + \sum_{i=1}^{p} \theta_{f,i} \Delta F_{t-i,T} + \varepsilon_{ft}$$

where μ_s and μ_f are constants. $a_{s,i}$, a, $\theta_{s,i}$ and $\theta_{f,i}$ are parameters. ε_{st} and ε_{ft} are white-noise error terms. Again a constant is included in the model. The reason for employing VAR models is that when compared to the univariate models these can demonstrate whether futures prices contain valuable information about future spot prices.

A VECM incorporates information regarding the long term relationship between spot and futures prices[34]. If this cointegration information helps explain the relationship between spot and futures prices and the futures have price discovery properties, a VECM should outperform the other time-series models. It should be pointed out that testing the forecasting abilities of freight futures is performed as an isolated task, and therefore has no direct connection to testing the unbiasedness hypothesis. The VECM used here will therefore include a constant in the model, but not in the cointegrating expression. This stands in contrast to the VECM used when testing the unbiasedness hypothesis, where a constant was included in the cointegrating expression, but not in the model. The reason for including the constant in the model is the same as for the VAR models presented above. The constant in the cointegrating expression is omitted as it is not expedient to test whether this constant is significant for all the VECMs which are to be estimated in this part. The VECM which will be employed here is presented below.

$$\Delta S_t = \mu_s + \sum_{i=1}^{p} a_{s,i}\Delta S_{t-i} + \sum_{i=1}^{p} \theta_{s,i}\Delta F_{t-i,T} + \gamma_s Z_{t-1} + \varepsilon_{st}$$

$$\Delta F_t = \mu_f + \sum_{i=1}^{p} a_{f,i}\Delta S_{t-i} + \sum_{i=1}^{p} \theta_{f,i}\Delta F_{t-i,T} + \gamma_f Z_{t-1} + \varepsilon_{ft}$$

$$Z_{t-1} = S_{t-1} + \beta_2 F_{t-1,T}$$

where μ_s and μ_f are constants. γ_s and γ_f are parameters measuring the proportion of last period's equilibrium error which is corrected for, also known as the adjustment speed of the spot and futures prices to their long run equilibrium. $a_{s,i}$, $a_{f,i}$, $\theta_{s,i}$ and $\theta_{f,i}$ are parameters. ε_{st} and ε_{ft} are white-noise error terms. Z_{t-1} is the cointegrating expression where S_{t-1} is the spot price, $F_{t-1,T}$ is the futures price and β_2 is the cointegrating vector describing the long run relationship between the variables.

[34] See chapter 5.1 for a further explanation.

All time-series models will be estimated recursively. This means that the coefficients of the parameters will be re-estimated for each forecast, thus the models will always reflect the most recent information available. The underlying assumptions regarding model specification will however not be changed.

7.1.2 Measures of forecasting accuracy

The forecasting accuracy of the models will be assessed using the mean error, the root mean square error and the mean absolute percentage error.

The mean error (ME) measures the mean of all the forecast errors. Because positive and negative errors will cancel each other out this is normally considered a uninformative measure. It does however have one interesting property, which is the ability to indicate whether the forecasts are generally biased. I will therefore include this measure in the analysis.

$$ME = \frac{1}{T-(T_1-1)}\sum_{t=T_1}^{T-S}(y_{t+s}-f_{t,s})$$

where T is the total sample size. T_1 is the fist out of sample forecast observation. $f_{t,s}$ is the s-step-ahead forecast made at time t. y_{t+s} is the actual value of the variable at time t.

The next measure which will be included is the root mean square error (RMSE). This measures the average deviation of the forecasts to the realized values and penalizes large error more than small ones. The RMSE is calculated in the following way:

$$RMSE = \left[\frac{1}{T-(T_1-1)}\sum_{t=T_1}^{T-S}(y_{t+s}-f_{t,s})^2\right]^{\frac{1}{2}}$$

I will also include the mean absolute percentage error (MAPE). This measures the average percentage error of the forecasts.

$$MAPE = \frac{100}{T-(T_1-1)} \sum_{t=T_1}^{T-S} \left| \frac{y_{t+s} - f_{t,s}}{y_{t+s}} \right|$$

All forecasts made in first differences will be transformed back to levels to ensure that the measures presented above are comparable for all models.

7.2 Properties of the data series

The data used to test the forecasting abilities of freight futures are monthly data starting March 2004 and ending September 2009. Monthly average spot prices have been calculated for each month. Rolling futures prices have been collected at the last day of each month. These prices are collected in such a way that three series of rolling futures prices are available for each route, containing futures prices with a constant time to maturity of one, two and three months.

To test the forecasting abilities of the models the observations have been split into two sub periods. The first starts March 2004 and ends February 2008. These 48 observations form the fit period, and is the period of which the models performing the first forecast is based on. The last 19 observations form the test period. The first forecast for the one, two and three month series will therefore be March, April and May 2008, respectively. The last forecasts will be September 2009. This yields 19, 18 and 17 testable forecasts for the one, two and three month prices series, respectively.

Because monthly prices are used in this exercise I have few observations to base the models on. To obtain good model specifications and avoid large forecasting errors which can be attributed to changes in the underlying flat rates I have therefore rebased all spot and futures prices to the flat rate level of 2009. Some words of caution are therefore appropriate. For the three-month futures prices collected in the month of October this implies that the futures market is able to estimate the underlying future flat rate correctly[35]. All price series are also transformed into natural logarithms. Descriptive statistics for the first differences from the fit period of the series is provided in table 7.1.

Table 7.1: Descriptive statistics and normality tests.

TC2 Price Series	N	Mean	Max	Min	StdDev	Skewness	Kurtosis	Jarque-Bera
SPOT AVG TC2	47	-0.00189	0.60370	-0.40141	0.18968	0.57968	4.15460	5.24288
TC2 RF1	47	0.00799	0.53524	-0.34377	0.18498	0.65046	3.35519	3.56138
TC2 RF2	47	0.00921	0.45953	-0.24335	0.13617	0.78659	4.27292	8.01987
TC2 RF3	47	0.01003	0.36628	-0.22314	0.11559	0.75548	4.29718	7.76608

TC4 Price Series	N	Mean	Max	Min	StdDev	Skewness	Kurtosis	Jarque-Bera
SPOT AVG TC4	47	-0.00528	0.54665	-0.59890	0.23605	0.23472	3.10775	0.45430
TC4 RF1	47	0.00086	0.46471	-0.37865	0.20174	0.35379	2.89287	1.00293
TC4 RF2	47	0.00373	0.38947	-0.27748	0.15255	0.48356	2.69644	2.01213
TC4 RF3	47	0.00542	0.36704	-0.22314	0.12360	0.45624	3.08879	1.64595

TC5 Price Series	N	Mean	Max	Min	StdDev	Skewness	Kurtosis	Jarque-Bera
SPOT AVG TC5	47	-0.00076	0.38879	-0.48490	0.20754	0.00918	2.63766	0.25776
TC5 RF1	47	0.00590	0.50046	-0.37729	0.20056	0.12461	2.73413	0.26007
TC5 RF2	47	0.00712	0.51490	-0.36101	0.18044	0.70049	3.89896	5.42626
TC5 RF3	47	0.00854	0.45155	-0.34295	0.13728	0.45988	4.32918	5.11653

All statistics are based on the first 48 observations of the series.

SPOT AVG TCX is series of the monthly average spot prices.
TCX RFY are series of rolling futures prices one, two and three months prior to maturity observed at the same time as the spot price series.

N is the number of observations of each time series. Max is the maximum value of the series and min is the minimum value.
Mean, standard deviation^2, skewness and kurtosis are the first, second, third and forth standardized moments of the time series.
The Jarque-Bera is a test for normality. The null hypothesis is that the series have skewness and kurtosis similar to a normal distribution.
The test statistic follows a chi-square distribution with two degrees of freedom under the null hypothesis. The 5% critical value is 5.9915.
The null hypothesis is rejected if the test statistic is greater than the critical value. Rejection indicates sign of non-normality.
Values representing rejected null hypothesis are reported in red.
All series have been rebased to the flat rate levels of 2009 and reflect the logarithmic first differences of these series.

[35] An explanation of this problem can be found in chapter 2.3.

The mean is as expected close to zero for all the series. It is however increasingly positive for all the routes when the time to maturity is increased. This might be a result of the market expecting future rates to rise in the end of the fit period. The minimum and maximum values show a tendency of declining in absolute terms as the time to maturity is increased, and this is also reflected by the standard deviations. The fact that the standard deviations are declining as the time to maturity is increased and that TC4 generally is more volatile than TC2 is expected[36]. The TC4 spot and RF1 series do however experience higher volatility than the corresponding TC5 series. This is not expected because the ships of TC5 are larger than those of TC4. It therefore seems like some route specific factors are influencing TC4 rates of the near future. All series are slightly positively skewed and the kurtosis is quite close to that of the normal distribution, three, for all series. The Jarque-Bera tests (Bera and Jarque, 1980) indicate that the time series follow a normal distribution, except from the TC2 RF2 and RF3 series which seem to be slightly more skewed and have a higher kurtosis than the other routes. It should be mentioned that these results are based on few observations collected over a fairly short period of time. These comments might therefore not reflect the long term properties of the series.

[36] See chapter 5.2 for a further explanation.

Table 7.2: Tests for stationarity

TC2 Price Series	ADF lvl (lags)	ADF 1st diff (lags)	PP lvl (BW)	PP 1st diff (BW)	KPSS lvl (BW)	KPSS 1st diff (BW)
SPOT AVG TC2	-3.690107 (0)	-6.509671 (1)	-3.835845 (1)	-9.913008 (12)	0.176877 (3)	0.149928 (9)
TC2 RF1	-3.980386 (0)	-7.916460 (0)	-3.999975 (1)	-10.87490 (8)	0.258038 (3)	0.104473 (7)
TC2 RF2	-3.421998 (0)	-6.727556 (0)	-3.460413 (2)	-6.936726 (5)	0.350893 (3)	0.089898 (5)
TC2 RF3	-1.198307 (9)	-7.220750 (8)	-3.074060 (5)	-6.059910 (9)	0.377461 (3)	0.101192 (5)

TC4 Price Series	ADF lvl (lags)	ADF 1st diff (lags)	PP lvl (BW)	PP 1st diff (BW)	KPSS lvl (BW)	KPSS 1st diff (BW)
SPOT AVG TC4	-4.789822 (1)	-8.449842 (1)	-4.032770 (4)	-14.41358 (45)	0.132870 (2)	0.500000 (46)
TC4 RF1	-4.016331 (0)	-7.919471 (0)	-4.023262 (2)	-11.73929 (9)	0.103555 (2)	0.182973 (13)
TC4 RF2	-3.552524 (0)	-7.280089 (0)	-3.582367 (3)	-7.489471 (5)	0.092042 (3)	0.109998 (6)
TC4 RF3	-5.339334 (4)	-7.098226 (8)	-3.275853 (3)	-5.948554 (5)	0.153464 (3)	0.084191 (4)

TC5 Price Series	ADF lvl (lags)	ADF 1st diff (lags)	PP lvl (BW)	PP 1st diff (BW)	KPSS lvl (BW)	KPSS 1st diff (BW)
SPOT AVG TC5	-5.053750 (4)	-6.326177 (1)	-3.319382 (5)	-8.003045 (17)	0.065133 (2)	0.147302 (16)
TC5 RF1	-3.737369 (0)	-7.547738 (0)	-3.695399 (4)	-9.997170 (9)	0.074843 (3)	0.183167 (13)
TC5 RF2	-3.615427 (0)	-6.891715 (8)	-3.545832 (4)	-7.951811 (6)	0.089840 (3)	0.180812 (8)
TC5 RF3	-4.472626 (2)	-7.260744 (0)	-3.333932 (3)	-6.261256 (4)	0.148868 (3)	0.095626 (4)

The ADF tests refer to Augmented Dickey Fuller tests with one intercept included. The null hypothesis is non-stationary.
Lag length set to max 12 to control for seasonality factors and automatically selected using Schwarz Information Criterion.
ADF is a one-tailed t-test where the test statistic follows the distribution calculated by Dickey and Fuller under the null hypothesis.
A 5% (10%) level of significance has a critical value of approximately -2.93 (-2.60) (depending on the number of lags) (MacKinnon, 1996)
The null hypothesis is rejected if the test statistic is smaller than the critical value, and if this is the case the series is said to be stationary.

The PP tests refer to the Phillips-Perron test with one intercept included. The null hypothesis is that a series is non-stationary.
The spectral estimation method is the Berlett kernel method and the bandwidth selection is the Newey-West Bandwidth.
The test statistic follows the same asymptotic distributions as the ADF test statistic and normalized bias statistics under the null hypothesis.
The critical values are therefore the same as for the ADF test.
The null hypothesis is rejected if the test statistic is smaller than the critical value, and if this is the case the series is said to be stationary.

The KPSS tests refer to the Kwiatkowski, Phillips, Schmidt and Shin test with an intercept included.
The spectral estimation method is the Berlett kernel method and the bandwidth selection is the Newey-West Bandwidth.
KPSS is a right tailed test, and the test statistic is a LM statistic which converges to a function of a standard Brownian motion under the null hypothesis.
The critical value is 0.436 (0.347) at a 5% (10%) level of significance and is calculated from the asymptotic distribution of a standard Brownian motion.
The null hypothesis is rejected if the test statistic is larger than the critical value, and if this is the case the series is said to be non-stationary.

Values representing rejected null hypothesis using the 5 % critical value are reported in red.
Values representing rejected null hypothesis using the 10 % critical value are reported in purple.

The three types of stationarity tests indicate that all the series, except from TC2 RF3, are stationary in levels at a 5 percent level of significance. For TC2 RF3 the ADF test indicates that the series is I(1), while the PP and KPSS tests indicate that it is I(0). At a 10 percent level of significance the KPSS tests however indicates that the TC2 RF2 and RF3 series are I(1). Because the results of the stationarity tests might be influenced by the fact that the time series are collected over a short time period and contain few observations, models reflecting both I(0) and I(1) time series will be employed to produce forecasts.

7.3 Evaluating the forecasting results

All the models which are used in this forecasting exercise were presented in part 7.1.1. To make sure that the models were well specified in terms of lag length SIC was first employed. Testing did however show that the accuracy of the forecasts declined when several lagged term were included. The reason for this might be that the data used are monthly data and combining this with serial correlated and highly volatile shipping rates[37] might imply large forecasting errors when the series move from one local trend to another. Another reason might be that estimating more parameters might yield less accurate forecasts if the estimates are uncertain. In line with this, one lagged term only is included in the models. Forecasts were generated as described in the presentation of the data. The accompanying forecasting accuracy measures are presented in the table below.

[37] Stopford (1997) contains an excellent explanation of the high volatility, including the underlying factors.

Table 7.3: Forecasting accuracy of the monthly spot rate forecasts

One-Month Forecasting of Spot Rates (Monthly)

	TC2			TC4			TC5		
	ME	RMSE	MAPE	ME	RMSE	MAPE	ME	RMSE	MAPE
RW	-0.03025	0.20841	3.12648	-0.02164	0.17010	2.81501	-0.00766	0.25209	4.29800
Futures	-0.03638	0.14622	2.39537	-0.06562	0.12188	2.08495	-0.03075	0.12695	2.34593
ARMA	-0.04809	0.23947	3.88040	-0.10809	0.23086	4.09870	-0.01780	0.26077	4.15691
ARIMA	-0.02958	0.22042	3.15701	-0.03034	0.24840	4.30651	-0.01931	0.26497	4.56591
VAR (levels)	-0.05360	0.17251	2.91140	-0.05837	0.15251	2.77010	-0.04555	0.15945	2.92990
VAR (diff)	-0.02222	0.14369	2.09827	-0.00935	0.11192	1.85800	-0.00240	0.14637	2.37349
VECM	-0.02472	0.13772	2.14070	-0.00331	0.09390	1.65560	-0.04110	0.12067	2.13080

Two-Month Forecasting of Spot Rates (Monthly)

	TC2			TC4			TC5		
	ME	RMSE	MAPE	ME	RMSE	MAPE	ME	RMSE	MAPE
RW	-0.06346	0.28299	4.81702	-0.05515	0.29919	5.34068	-0.02521	0.40046	6.76142
Futures	-0.05532	0.25784	4.37716	-0.12923	0.24417	4.51756	-0.05592	0.29335	5.03770
ARMA	-0.09006	0.32317	5.84949	-0.24387	0.44639	8.20038	-0.04248	0.41418	6.91100
ARIMA	-0.06537	0.30668	5.12082	-0.07564	0.35243	6.36949	-0.04739	0.41484	6.68721
VAR (levels)	-0.09199	0.31333	5.68330	-0.16927	0.35874	6.51470	-0.03981	0.35253	5.94470
VAR (diff)	-0.07258	0.28562	4.87973	-0.06567	0.34317	6.03462	-0.03256	0.41515	6.91253
VECM	-0.04629	0.25829	4.33680	-0.03625	0.18702	3.10170	-0.03194	0.28334	4.80470

Three-Month Forecasting of Spot Rates (Monthly)

	TC2			TC4			TC5		
	ME	RMSE	MAPE	ME	RMSE	MAPE	ME	RMSE	MAPE
RW	-0.11059	0.31799	5.30573	-0.09197	0.40895	7.53831	-0.04576	0.49068	8.80422
Futures	-0.07974	0.28796	4.95153	-0.14872	0.31401	5.80043	-0.04824	0.39225	6.44106
ARMA	-0.14128	0.35978	6.56489	-0.32045	0.54492	10.22334	-0.06753	0.47534	8.29006
ARIMA	-0.11744	0.34953	6.09804	-0.12175	0.44485	7.97522	-0.08422	0.51988	9.15774
VAR (levels)	-0.15058	0.36020	6.89330	-0.21962	0.44894	8.31550	-0.03694	0.44447	8.06570
VAR (diff)	-0.11868	0.32549	5.43327	-0.10003	0.42203	7.75057	-0.05606	0.49914	8.90149
VECM	-0.13607	0.33710	6.42970	-0.11833	0.31523	5.70910	-0.02973	0.38880	6.50870

As expected the results indicate that the forecasting accuracy of the models generally declines as the forecasting horizon is increased. There are, however, three exceptions to this regarding the mean error which represents the biasedness of the models. For TC5 the two-month forecasts of the futures are more biased than the three-month forecasts, and for the VAR models in level and VECMs the bias seems to decrease with the time horizon of the forecasts. The fact that it becomes more difficult to produce precise forecasts as the forecasting horizon is increased can also be observed from the performance of the random walk, which improves relatively to the other models as the time horizon is increased.

The mean errors indicate that all the models generate negatively biased forecasts. The mean error, however, is usually not perceived as a sufficient measure of forecasting accuracy because positive and negative forecast errors have a tendency of cancelling each other out. In line with this I will focus on the root mean square errors and mean absolute percentage errors in the following paragraphs.

Looking at the results of the univariate time-series models it is clear that the ARIMA models outperform the ARMA models for all three time horizons for TC2 and for the two and three-month horizon for TC4. The ARMA model is preferred for the one month horizon of TC4 and generally for TC5. All ARMA and ARIMA models are however outperformed by a random walk, except from the three-month ARMA model for TC5. This illustrates that basing forecasts of future spot prices on historic spot prices only is not recommended.

Turning to the multivariate models, a VAR in first differences generates better forecasts than a VAR in levels for TC2, TC4 and the one-month horizon of TC5. On the other hand, a VAR in levels outperforms a VAR in first differences for the two and three-month horizon of TC5. At the same time the VARs in levels outperform the ARMA models for all routes and horizons, except for the three-month time-horizon of TC2. The VARs in first differences outperform the ARIMA models for all routes and horizons, except for a two-month horizon for TC5. These results imply that the accuracy of the forecasting models is increased when including futures prices. The futures prices are thereby showed to contain valuable information about future spot rates. The results also indicate that the TC2 and TC4 series probably are non-stationary, while the TC5 series are stationary. For the one-month horizon of TC4 and TC5 the indications are however not clear.

The VECMs produce the best forecasting results of all the time-series models, except for the three-month horizon for TC2 and partly the one-month horizon of TC2. This indicates the existence of a cointegrating relationship between spot rates and futures prices, and that the VECM is able to utilize this cointegrating relationship when producing forecasts. For TC5 this is quite puzzling because both the univariate and multivariate models, except the VECM, indicate that a model in levels produces better forecasts than a model in first differences. Models in levels are also used throughout this thesis for TC5. In line with this, the tests performed here should be repeated in a few years when more data is available.

Comparing the performance of the time series models to the forecasts based on the futures prices a VECM and VAR in first differences outperform the futures prices for the one-month time horizon of TC2 and TC4. For the one-month horizon of TC5 the futures prices are only outperformed by the VECM. The result for the two-month horizon of TC4 and TC5 is the same, while the result for TC2 is somewhat mixed between the futures prices and the VECM. For the three-month time-horizon the futures price outperform all time-series models for TC2, while the results are mixed between the futures prices and VECMs for TC4 and TC5. Overall the results indicate that forecasts based on the futures prices perform well compared to more complex time-series models. Considering that employing multivariate time-series models to forecast future spot rates is affiliated with higher costs than using the already available futures prices, some might actually prefer using the futures prices even for the routes and time-horizons where multivariate models provide the best forecasts.

8. Summary and conclusions

The price discovery properties of clean tanker freight futures have been investigated in this thesis, focusing on the most liquid clean tanker freight futures which are those written on the routes TC2, TC4 and TC5. The test conducted here concentrated on the unbiasedness hypothesis, the lead-lag relationship between freight futures and spot rates and the forecasting properties of freight futures with regards to the underlying spot rates.

The results from testing the unbiasedness hypothesis indicate that freight future prices one month prior to maturity are unbiased for TC2. For the two- and three month horizons of this route the evidence was, however, weaker, and in line with this no conclusion was arrived upon. For TC4 the unbiasedness hypothesis was rejected for future prices collected one-, two- and three months prior to maturity, and for TC5 the unbiasedness hypothesis was accepted for all investigated time-horizons. Whether the unbiasedness hypothesis holds for clean tanker freight futures thereby seems to depend on the route in question and time to maturity.

The lead-lag relationship between futures prices and spot rates was then analysed in order to gain further understanding of the interaction between the spot and futures markets. The main findings were that futures prices seem to lead the spot rates for TC2 and TC5. For TC4 a bi-directional relationship was found, but the futures also seem to have a leading role also for this route. Futures are therefore thought to lead spot rates in the clean tanker market, and might thereby be used as a sentiment indicator.

The final part of this thesis concentrated on the forecasting performance of end-of-month freight futures with regards to the underlying spot rates. As expected, the forecasting accuracy of all models generally declined as the forecasting horizon was increased from one- to two- and three months. When comparing the models it was found that univariate models were mostly outperformed by a random walk, indicating that forecasts should not be based on historic spot prices alone. The multivariate models confirmed this by generally providing more accurate forecasts than their univariate cousins. These results imply that the futures prices contain valuable information about future spot rates. Multivariate time-series models were generally found able to outperform the forecasts indicated by the futures prices themselves for the one- and two-month horizons, but for the three-month horizon the futures performed as well as or better than the multivariate models.

When looking at the forecasting results of TC5 a puzzling phenomenon was observed. The results from both the univariate and multivariate models, except the VECM, indicated that a model in levels provided the best forecasts, at least for the longer time-horizons. This indicates that the time series of TC5 are stationary in levels. A VECM was however found to be the time series-model which provided the most accurate forecasts, which indicates the existence of a cointegrating relationship between spot rates and futures prices. Throughout this thesis problems were experienced for all routes when trying to detect whether the time series are stationary in levels or first differences. In line with this the tests performed here should be repeated in a few years when more data is available.

The tests of the unbiasedness hypothesis, the lead-lag relationship and the forecasting properties of clean tanker freight futures conducted in this thesis indicate that the futures have price discovery properties. These price discovery properties imply that market participant may use clean tanker freight futures to guide decisions in the physical market, and the futures can thereby contribute to a more efficient allocation of economic resources.

As topics for future research I would like to suggest investigating the short-term forecasting performance of clean tanker freight futures with regard to both spot and futures prices. It would also be interesting to extend the research conducted in this thesis to also include the dirty tanker market, because most of the literature on freight futures up to now has been focused on dry bulk routes.

Readings

Adam, T.R. and Fernando, C.S., (2006). *Hedging, Speculation and Shareholder Value*. Journal of Financial Economics, 81, pp. 283-309.

Adland, R., Koekebakker, S. and Sødal, S., (2009). *Forward Curve Dynamics in the Tanker Futures Market*. Working Paper.

The Baltic Exchange <www.balticexchange.com> 15.01.2010.

Batchelor, R., Alizadeh, A., Visvikis, I., (2007). *Forecasting Spot and Forward Prices in the International Freight Market*. International Journal of Forecasting, 23, pp 101–114.

Bera, A.K. and Jarque C.M.J., (1980). *Efficient Tests for Normality, Homoscedaticity and Serial Independence of Regression Residuals*. Economics Letters, 6, pp. 255 – 259.

Bessler, W., Drobetz, W., Seidel, J., (2008). *Ship Funds as a New Asset Class: An Empirical Analysis of the Relationship Between Spot and Forward Prices in Freight Markets*. Journal of Asset Management, 9, pp. 102–120.

Black, F., (1976). *The Pricing of Commodity Contracts*. Journal of financial economics, 3, pp. 167-179.

Bodie, Z., Kane, A. and Marcus A.J., (2008). *Investments (International edition).* 7th ed. McGraw-Hill, NY.

Brooks, C., (2008). *Introductory Econometrics for Finance*. Cambridge.

Chang, Y. and Chang H., (1996). *Predictability of the Dry-Bulk Shipping Market by BIFFEX*. Maritime Policy and Management, 23, pp. 103-114.

Cheung, Y. and Lai, K., (1993). *Finite Sample Sizes of Johansen's Likelihood Ratio Test for Cointegration*. Oxford Bulletin of Economics and Statistics, 55, pp. 313-328.

Cullinane, K., (1992). *A Short-Term Adaptive Forecasting Model for BIFFEX Speculation: A Box-Jenkins Approach*. Maritime Policy and Management, 19, pp. 91-114.

Dickey, D.A. and Fuller W.A., (1979). *Distribution of the Estimators for Autoregressive Time Series With a Unit Root*. Journal of the American Statistical Association, 74, pp 427–431.

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

Fama, E.F., (1991). *Efficient capital markets II*. Journal of Finance, 46, pp. 1575-1617.

Granger, C.W.J. and Newbold, P. (1974). *Spurious Regressions in Econometrics*. Journal of Econometrics, 2, pp. 111 –120.

Gray, R. and Tomek, W., (1970). *Temporal Relationships Among Prices on Commodity Futures Markets: Their Allocative and Stabilizing Roles*. American Journal of Agricultural Economics, 52, pp. 372-380.

Gonzalo, J., (1994). *Five Alternative Methods of Estimating Long-Run Equilibrium Relationships*. Journal of Econometrics, 60, pp. 203-233.

Haigh, M., (2000). *Cointegration, Unbiased Expectations and Forecasting in the BIFFEX Freight Futures Market*. Journal of Futures Markets, 20, pp. 545-571

Hicks, J., (1939). *Value and Capital*. Oxford University Press, New York.

Hull, J., (2009). *Options Futures and other derivatives*. Upper Saddle River, N.J. Pearson Prentice Hall. 7th ed.

Imarex, <www.imarex.com> 15.01.2010

International Maritime Organization <http://www.marisec.org/worldtradeflyer.pdf>. 27.10.2010

Johansen, S., (1988). *Statistical Analysis of Cointegration Vectors*. Journal of Economic Dynamics and Control, 12, pp. 231-254.

Johansen, S., and Juselius, K. (1990). *Maximum Likelihood Estimation and Inference on Cointegration-With Applications to the Demand for Money*. Oxford Bulletin of Economics and Statistics, 52, pp. 169-210.

Johansen, S., (1995). *Likelihood-Based Inference in Cointegrated Vector-Autoregressive Models*. Oxford University Press Inc., NY.

Kavussanos, M. (1996). *Comparisons of Volatility in the Dry-Cargo Ship Sector: Spot versus Time Charters, and Smaller versus Larger Vessels.* Journal of Transport Economics and Policy, 1, pp 67-82.

Kavussanos, M., (2003). *Time Varying Risks Among Segments of the Tanker Freight Markets*. Maritime Economics & Logistics, 5, pp. 227-250.

Kavussanos, M. and Nomikos, N., (1999). *The Forward Pricing Function of the Shipping Freight Futures Market*. Journal of Futures Markets, 19, pp. 353-376.

Kavussanos, M. G. and Nomikos, N. K., (2000). *Hedging in the Freight Futures Market*. Journal of Derivatives, 8, pp. 41-58.

Kavussanos M.G., and Nomikos, N.K. (2003). *Price Discovery, Causality and Forecasting in the Freight Futures Market.* Review of Derivatives Research, 6, pp. 203-230.

Kavussanos, M. and Visvikis I., (2004). *Market Interactions in Returns and Volatilities Between Spot and Forward Shipping Markets.* Journal of Banking and Finance, 28, pp. 2015-2049.

Kavussanos, M. and Visvikis, I. (2006a). *Derivatives and Risk Management in Shipping*, Witherby Publishing, London.

Kavussanos, M. and Visvikis, I., (2006b). *Shipping Freight Derivatives: A Survey of Recent Evidence*. Maritime Policy & Management, 33, pp. 233-255.

Kavussanos, M., Visvikis, I. and Menachof, D., (2004). *The Unbiasedness Hypothesis in the Freight Forward Market: Evidence from Cointegration Tests.* Review of Derivatives Research, 7, pp. 241–266.

Keynes, J., (1930). *A Treatise on Money*. Macmillan, London.

Kwiatkowski, D., Phillips, P. C. B., Schmidt P., and Shin Y., (1992). *Testing the Null Hypothesis of Stationarity Against the Alternative of a Unit Root*. Journal of Econometrics 54, pp. 159–178.

Lee, T. and Tse, Y., (1996). *Cointegration Tests with Conditional Heteroskedasticity*. Journal of Econometrics, 73, pp. 401-410.

Mackinnon, J.G., Haug, A.A. and Michelis, L. (1999). *Numerical Distribution Functions of Likelihood Ratio Tests for Cointegration.* Journal of Applied Econometrics, 14, pp. 563-577

McDonald, R., (2006). *Derivatives Markets,* 2nd ed., Pearson. Boston.

NOS, <www.nos.no>, 15.01.2010.

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

Platts, <www.platts.com> 15.01.2010

Stopford, M., (1997). *Maritime Economics*, Taylor & Francis, London

Strandenes, S. P., (1984). *Price Determination in the Time Charter and Second Hand Markets*, Centre for Applied Research, NHH, Working Paper MU 06.

Thompson, S., Garcia, P. and Wildman, L. D. (1996). *The Demise of the High Fructose Corn Syrup Futures Contract: A Case Study*. Journal of Futures markets, 16, pp. 697-724.

Worldscale <www.worldscale.co.uk> 15.01.2010

Worldscale Association (London) Limited and Worldscale Association (NYC) Inc, (2008). *New Worldwide Tanker Nominal Freight Scale "Worldscale", As Revised Effective 1st January 2009*, Polestar Wheatons Ltd, Exeter, Great Britain.

Zannetos, Z. S., (1966). *The Theory of Oil Tankship Rates. M.I.T. Press.* Cambridge, MA.

Appendix I: IMAREX Freight Futures Product Specifications

The following rules regarding closing prices and relevant product specifications have been extracted from appendix 5 of the IMAREX rulebook as per 3^{rd} of August 2009.

Closing Price

Closing Price is set to:

The best bid, if last price < best bid

The best offer, if last price > best offer

or else use Last Price.

The Closing Price shall reflect the market value of the Product at the end of Clearing Hours.

Product Specifications

Underlying	**Index**	**Index Provider**	**Closing Price**
	TC2, MR, Continent – USAC, 37,000 mt	Baltic Exchange	Imarex
	TC4, MR, Singapore - Japan, 30,000 mt	Platts	Imarex
	TC5, LR1, AG – Japan, 55,000 mt	Platts	Imarex
Flat Rates	As published by the Worldscale Association (London) Limited and the Worldscale Association (NY) Inc.		
Price quotation	Worldscale points		
Minimum price fluctuation	0.25 Worldscale point		

Contract value	#Lots × Lot size × Worldscale Flatrate × (Worldscale points/100) (The Worldscale Flatrate applicable for each Index Day in the Delivery Period)
Delivery Period	Month: First Index Day of the month to last Index Day of the month. Quarter: First Index Day of the Quarter to last Index Day of the Quarter. A Quarter Contract will be split equally into 3 Month Contracts on the Trading Day and settled as Month Contracts. Year: First Index Day of the Year to last Index Day of the Year A Year Contract is split into equally into 12 Month Contracts on the Trading day and settled as Month Contracts.
Final Settlement Day	Last settlement day in the Delivery Period.
Settlement Price	The arithmetic average of the Spot Prices for the relevant Underlying Product over the number of Index Days in the Delivery Period.
Lot size	Month: 1 lot = 1,000 mt Quarter: 1 lot = 3,000 mt Year: 1 lot = 12,000 mt
Minimum lots per contract	1 lot in all Products
Product structure	Months: 6 consecutive months starting with the current month. A new month Product is introduced once the current month is no longer available for trading. Please refer to ”Last Trading Day” for details of Last Trading Day. Quarters: 6 consecutive quarters starting with the present quarter. A new quarter Product is introduced once the present quarter is no longer available for trading. Please refer to ”Last Trading Day” for details

		on Last Trading Day.
	Year:	2 year Products available. A new year Product commencing in the next full calendar year is introduced once the current year is no longer available for trading. Please refer to ”Last Trading Day” for details on Last Trading Day.
Last trading day	Month:	Last Trading Day is the last day of the Delivery Period for the month in question. If this date is a non-trading day, the Last Trading Day is defined as the nearest Trading Day prior to this.
	Quarter:	Last Trading Day is the last Trading Day of the first month of the quarter.
	Year:	Last Trading Day is the last Trading Day of the first month of the year.

Appendix II: Results that are not included in the text

Table footnote 21: Results from cointegration tests using Reimer's small sample correction.

	One-Month Price Series - Trace Statistics		
r	TC2 (1 lag)	TC4 (1 lag)	TC5 (1 lag)
0	23.30528	28.09957	38.59596
	(20.26184)	(20.26184)	(20.26184)
1	5.09823	3.53403	10.77748
	(9.16455)	(9.16455)	(9.16455)
Rank	1	1	2

	One-Month Price Series - Max Statistics		
r	TC2 (1 lag)	TC4 (1 lag)	TC5 (1 lag)
0	18.20706	24.56554	27.81849
	(15.89210)	(15.89210)	(15.89210)
1	5.09823	3.53403	10.77748
	(9.16455)	(9.16455)	(9.16455)
Rank	1	1	2

	Two-Month Price Series - Trace Statistics		
r	TC2 (4 lags)	TC4 (2 lags)	TC5 (3 lags)
0	26.89271	22.07617	31.49315
	(20.26184)	(20.26184)	(20.26184)
1	5.72731	3.62433	12.31587
	(9.16455)	(9.16455)	(9.16455)
Rank	1	1	2

	Two-Month Price Series - Max Statistics		
r	TC2 (4 lags)	TC4 (2 lags)	TC5 (3 lags)
0	21.16540	18.45184	19.17728
	(15.89210)	(15.89210)	(15.89210)
1	5.72731	3.62433	12.31587
	(9.16455)	(9.16455)	(9.16455)
Rank	1	1	2

	Three-Month Price Series - Trace Statistics		
r	TC2 (3 lags)	TC4 (2 lags)	TC5 (2 lags)
0	22.70431	20.53342	33.81988
	(20.26184)	(20.26184)	(20.26184)
1	7.30631	5.71475	10.56080
	(9.16455)	(9.16455)	(9.16455)
Rank	1	1	2

	Three-Month Price Series - Max Statistics		
r	TC2 (3 lags)	TC4 (2 lags)	TC5 (2 lags)
0	15.39799	14.81867	23.25907
	(15.89210)	(15.89210)	(15.89210)
1	7.30631	5.71475	10.56080
	(9.16455)	(9.16455)	(9.16455)
Rank	0	0	2

r is the number of cointegrating vectors.
For the λ_{trace} statistic the null is that rank is smaller than or equal to r, while the alternative is that rank exceeds r.
For the λ_{max} statistic the null is that rank equals r, while the alternative is that rank equals r+1.
The 5% critical values are those of MacKinnon-Haug-Michelis (1999) and are reported in parenthesis.
The rejection rule is to reject the null hypothesis if the test statistic is larger than the critical value.
Red colour denotes rejection of the null hypothesis.
Rank describes the number of cointegrating vectors implied by the hypothesis tests.

Table footnote 26: Results from the ARIMA tests.

ARIMA (3,1,0) Model Parameters Rebased TC2 Series				
		Coefficient	Standard Error	t-value
Constant		-0.00024	0.00026	-0.93204
AR	Lag 1	0.12490	0.02681	4.65904
	Lag 2	0.02979	0.02700	1.10308
	Lag 3	-0.02603	0.02681	-0.97082
Dummy	Lag 0	0.00293	0.00102	2.87062

ARIMA (3,1,0) Model Parameters Rebased TC4 Series				
		Coefficient	Standard Error	t-value
Constant		-0.00061	0.00024	-2.57067
AR	Lag 1	0.08951	0.02674	3.34680
	Lag 2	0.02164	0.02686	0.80545
	Lag 3	0.02990	0.02679	1.11628
Dummy		0.01085	0.00093	11.61257

ARIMA (3,1,0) Model Parameters Rebased TC5 Series				
		Coefficient	Standard Error	t-value
Constant		-0.00040	0.00026	-1.57864
AR	Lag 1	0.06598	0.02674	2.46738
	Lag 2	0.03957	0.02679	1.47711
	Lag 3	0.03468	0.02675	1.29624
Dummy		0.00796	0.00101	7.92024

ARIMA regressions were run on the square roots of the series to include jumps in both price levels and variances.

The null hypothesis is that there are no jumps in the series which may be attributed to rolling the series.

The test statistic follows a student t-distribution with 1392 (TC2) and 1397 (TC4 and TC5) degrees of freedom under the null hypothesis. The 5% critical value is 1.96.

The null hypothesis is rejected if the test statistic is greater than the critical value. Rejection indicates structural breaks.

Values representing rejected null hypothesis are reported in red.

Table footnote 33: Details from the Doornik and Hansen tests.

		Doornik and Hansen (1994) test for normality				
Route	Component	Skewness	Chi-sq	Kurtosis	Chi-sq	Jarque-Bera
TC2	Est	0.59398	71.37334	12.27018	1028.12500	1099.49800
	Eft	-0.18362	7.77766	6.95261	395.83420	403.61190
	Joint		79.15101		1423.95900	1503.11000
TC4	Est	0.37066	30.26621	11.80564	1116.44200	1146.70800
	Eft	0.07681	1.38015	5.02602	148.59470	149.97480
	Joint		31.64636		1265.03600	1296.68300
TC5	Est	0.55240	62.95192	10.94378	845.12760	908.07950
	Eft	0.07028	1.15762	9.36598	804.62260	805.78030
	Joint		64.10955		1649.75000	1713.86000