NORGES HANDELSHØYSKOLE

Bergen, 19.12.2006

Cointegrated Energy:

Can energy company shares be used to create Error Correction Model forecasts for the oil price? If so, are there certain sectors within the industry that produce better forecasts?

Ole Husebø Schøyen

Thesis Advisor: Associate Professor Jonas Andersson

Master's Thesis: Financial Economics

NORGES HANDELSHØYSKOLE

This thesis was written as a part of the Master of Science in Economics and Business Administration program - Major in International Business. Neither the institution, nor the advisor is responsible for the theories and methods used, or the results and conclusions drawn, through the approval of this thesis.

Abstract

Oil company shares are closely related to the oil price. This paper examines if this relationship is strong enough to conclude that the historical spot prices for oil are cointegrated with individual oil company share price time series, arguing that investors, who buy shares in companies whose business is related to the exploration, production or marketing of oil and oil based products, are forward looking. In turn this implies that the share prices embody information about expected future oil prices, much like oil futures. The paper will also attempt to explore, if there are certain sectors within the oil industry that are more appropriate to use as forecasting tools for oil. The paper answers whether there is a cointegrated relationship between oil shares and the spot price for crude, but fails to answer the second part because no such cointegrated relationship seems to exist.

Contents

1. Introduction

The Error Correction Model (ECM), developed as a result of Engle and Granger's investigations into non-stationarity, is often used to forecast non-stationary time series. This is because it corrects for the non-stationarity and, unlike models that differentiate, such as Box-Jenkins ARIMA, ECMs reflect long-term information embodied in time series. They can be used whenever one has two or more time series that move toward long-run equilibrium and popular models include forecasting exchange rates with the help of forward rates, commodity prices with the help of futures prices and equity prices, using dividends[1].

This paper aims to empirically examine whether ECM forecasts of the Brent Spot Crude Oil price are plausible i.e. is there cointegration between an individual oil company's share price and the oil price. The idea being that oil company share prices, much like oil futures, embody a lot of information, based on investor expectations regarding the oil price. The forecasts will be made using a selection of groups of oil related companies' historical share price data. The goal will be to determine whether ECM forecasts, using data from companies whose share prices should theoretically be more sensitive to oil price movements, also tend to produce better forecasting results for crude prices.

The hypothesis is that companies that are theoretically more sensitive to oil price changes, the ones whose fortunes are tied more directly to the crude oil price, rather than gas or petrochemicals or other portfolio items, should have share prices that embody more of the information that is pertinent to forecasting the oil price. Such an insight might be useful when determining whether or not to use ECM, with share prices, as a forecasting method for commodities.

[1] Brooks, C. *Introductory Econometrics for Finance*, Cambridge University Press, 2002 p. 388

2. Research Methodology

2.1 Cointegration and the Error Correction Model

Cointegration is a special feature of certain non-stationary time series. It means that a linear combination of the variables in a time series model yields stationary residuals, which can be used to correct the spurious regressions produced by an OLS model on non-stationary data. The stationarity of the residuals is the result of the series' moving together over time, implying that they are bound by some long-run relationship. In the case of share prices and dividends, the long-run relationship is the dividend discount model[2]. With commodities and futures it is that they respond similarly to available information in the market and are bound by cost of carry[3].

Crude oil prices and oil company shares also have a tendency to respond similarly to new information in the market. The symmetric growth of the oil price and energy company shares over the past two years has illustrated this. The figures presented in this paper will also demonstrate that this is the case.

There is a question of causality that must be addressed in designing these models. One might argue that the causality runs as follows: oil prices are determined exogenously, based on political tension, or worldwide demand and that company shares simply respond to this as a result. At first glance it is thus inappropriate to model oil price as the dependent variable. However, the causality need not run only one way. Oil companies represent the supply side in determining the oil price, through their technological improvements and exploration activities they also impact the oil price, in a more long-term time frame. The choice of oil as the dependent variable is based on the idea is that share prices embody information about expectations for future oil prices, unlike the spot price for oil which is determined by current demand and supply. Since the causality is not clear, this paper argues that the expectations of

---

[2] Brooks, C. *Introductory Econometrics for Finance*, Cambridge University Press, 2002 p. 389

[3] *ibid*

future oil based cash flows in the energy companies' shares is sufficient reason for using them as an independent variable to forecast oil prices. In a sense this assumes that share prices are determined in a more forward looking manner than oil spot prices. The grounds for this assumption are that oil purchased at spot price is for consumption, whereas shares are purchased for investment. Both are also purchased for speculation, but that is short-term. The investment motive in share purchases lead me to conclude that share prices embody more expectations about the future than commodity spot purchases.

2.1.1 Testing for non-stationarity

In order to examine if crude prices and energy stock prices are cointegrated, they must first be tested to ensure that they are I(1) and then tested for cointegration. The Augmented Dickey-Fuller (ADF) test will be utilized to test for initial non-stationarity. The ADF test is an improved version of the Dickey-Fuller test, which is an examination of a data set with the following hypotheses:

H_0 = the series has a unit root

H_A = the series is stationary

The test uses an autoregressive model to see if the root of the first lag is a unit root (=1) or less than 1.

(2.1) $y_t = \phi y_{t-1} + \varepsilon_t$

The model is rearranged for ease of computation and interpretation[4]:

Let $\beta \equiv \phi - 1$
then (2.1) is $y_t = (1+\beta) y_{t-1} + \varepsilon_t$
$y_t = y_{t-1} + \beta y_{t-1} + \varepsilon_t$
$y_t - y_{t-1} = \beta y_{t-1} + \varepsilon_t$
which leaves us with
(2.2) $\Delta y_t = \beta y_{t-1} + \varepsilon_t$

[4] Brooks, C. *Introductory Econometrics for Finance*, Cambridge University Press, 2002 p. 377

This is easier to compute using software packages using OLS, because H0 $\phi = 1$ is tested as $\beta = 0$, which is the standard coefficient test in statistical software packages. The coefficients produced are not t-distributed. Several authors have used simulated data to find critical values for the test statistic: $\frac{\hat{\beta}}{SE\left(\hat{\beta}\right)}$. The R statistical package, employed in this work, uses Banerjee et. al[5] to produce a "p-value" to show whether or not the test statistic is greater in absolute value than the critical value for the data.

In practice the Dickey-Fuller test tends to overstate the coefficients in cases where there is autocorrelation in the time series. The solution to this problem is the ADF test, which adds lags, which is a common method for dealing with autocorrelation problems. The number of lags is usually determined by some information criterion. R has a time series package with an ADF test that uses $\sqrt[3]{n-1}$ where n is the number of obervations in the series, to set the upper bound for the number of lags.

The choice of ADF as the test for non-stationarity is based on its ease of computation. It does have weaknesses. It really tests if a one cannot conclude that a series is stationary, which is the wrong way round in statistical methodology. However, as long as the data point to a clear non-rejection of the null hypothesis, the author will be content with the ADF results.

2.1.2 Testing for Cointegration

Cointegration is present when a linear combination of two or more non-stationary time series yields residuals that are stationary. The test for this is fairly obvious from the above definition. One must build a linear model of two or more time series and determine if the residuals that model produces are stationary. This work aims to determine if individual company characteristics can influence the quality of forecasts so the most appropriate type of model is a simple univariate one where the oil price is y_t and the company share price is x_t. The linear model is thus:

---

[5] A. Banerjee, J. J. Dolado, J. W. Galbraith, and D. F. Hendry: *Cointegration, Error Correction, and the Econometric Analysis of Non-Stationary Data*, Oxford University Press, Oxford. (1993)

(2.3) $y_t = \alpha + \beta x_t + \varepsilon_t$

The residuals produced $\hat{\varepsilon}_t$ are substituted into (2.2) and the ADF test is used to determine stationarity. ADF is certainly appropriate here, as it is not statistically counterintuitive, because it is used to determine stationarity, rather than non-stationarity. However, because we are performing a regression on residuals the distribution of critical ADF t-statistics differ[6]. The analysis here will employ Engle and Granger's critical ADF values, instead of the default in R's time series package.

2.1.3 Building Error Correction Models

The last section described the first step of the method this analysis will employ. It is the Engle and Granger two step method. Step 1 was to determine that the time series are both I(1), and then to determine if they are cointegrated. Step two is to build a model using the linear combination of the two series in addition to the differenced series themselves to create a model that is stationary and compensates for deviations from the long-term equilibrium suggested by the cointegrated relationship. This paper will utilize the method presented in Brooks[7]:

(2.4) $\Delta y_t = \beta_1 \Delta x_t + \beta_2 \left(\hat{u}_{t-1}\right) + \varepsilon_t$

Where $\hat{u}_{t-1} = y_{t-1} - \hat{\tau} x_{t-1}$ which shows a linear combination of the two time series, known as the cointegrating vector $(1-\tau)$. The key insight into this model is that the β_2 coefficient will correct for deviations from the long-run equilibrium that result from shocks in previous periods. Thus we have established a stationary model for OLS estimation that also incorporates long-run information from the time series.

The analysis here will proceed to analyze the quality of the forecast models by Mean Squared Error (MSE) statistics, both within sample and out of sample, and then compare the results

---

[6] Engle, R. and Granger, C. (eds.) *Long-Run Economic Relationships*, Oxford University Press 1991 p. 327

[7] Brooks, C. *Introductory Econometrics for Finance*, Cambridge University Press, 2002 p. 394

across the oil company groups to see if there is a pattern of improved forecasting results for the companies whose fortunes are more directly linked to movements in the oil price.

2.2 Data Selection

As this analysis requires cross comparisons of forecasting model there are some criteria that must be met to ensure that the results are comparable.

1) The data must all have the same number of observations

This is to ensure that no model has more, or less, information than the others.

2) The data must run over the same time period

To compare models, it is desirable to have data based on the exact same information.

3) The data must be in the same format

i.e. the share prices and crude price must all be stated in USD because it is the commonest denomination for available share prices and the oil price. This is really a matter of practicality, but a useful guide when searching for data.

This analysis wishes to explore four groups of oil-related companies, under the assumption that they are linked respond, alongside the oil price, to new information in varying magnitudes. The groups, selected before data collection, are:

1) Exploration and Drilling: these are relatively small companies that specialize in areas that are directly related to the extraction of oil. As such they should be very responsive to new information that also impacts the oil price and tend strongly towards a long-run equilibrium with crude.

2) Independent Integrated Oil Companies: These are larger and more diversified than Exploration and Drilling companies, but usually have their fortunes connected to a few specified production facilities producing oil or gas, and are therefore also very responsive to new information that impacts the oil price.

3) Super Majors: These are the largest oil companies in the world by market capitalization. Oil is obviously a substantial part of their portfolios, but they are also invested in Chemicals and, the oil-related, gas exploration and production. In theory they should be less responsive to new information impacting oil price, because the markets expectations for future cash flows will also be based on how they perform in the other sectors they are invested in.

4) Oil Services: They provide engineering services and manufactured equipment used in oil exploration and production. As such their fortunes are indirectly linked to the oil price. They usually have large long-term contracts with oil companies they provide services to and their future cash flows depend more on these deals, than the direct changes in the oil price.

The Crude price chosen for this has been Brent Spot, taken from the American Energy Information Administration[8]. The historical share price data was found on Yahoo Finance[9], a website that provides free historical time series, and organises these in industries, much like the search criteria demand.

Monthly data was selected. This was to avoid calendar problems, but also because ECM uses at long-term relationships and 100+ monthly observations covers nearly ten years. In order to get a sufficient number of observations the companies selected needed to have more than a ten year history of public trading. The period chosen was May 1993- August 2006: 160 observations. This was done in late November 2006 leaving three months for out of sample MSE calculations.

[8] http://tonto.eia.doe.gov/

[9] finance.yahoo.com

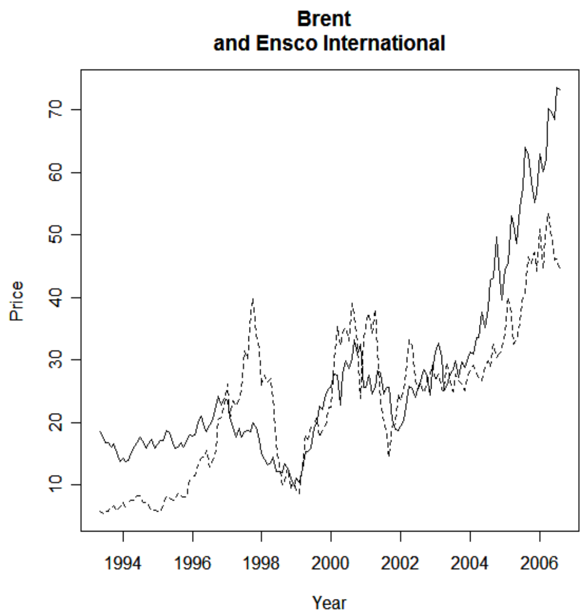
The companies chosen were as follows:

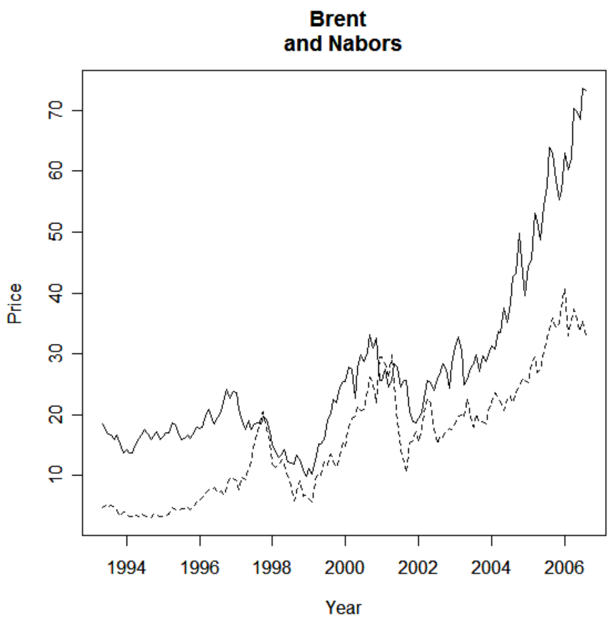
Exploration and Drilling	**Independent Integrated**
Ensco International	Norsk Hydro A.S.
Nabors Ltd.	ConocoPhillips
Marathon Oil	Repsol
Parker Drilling	Hess
Transocean Rig	Occidental
Super Majors	**Oil Services**
ExxonMobil	BJ Services
Shell	Haliburton
Total	Schlumberger
Chevron	Weatherford International
BP	Baker Hughes

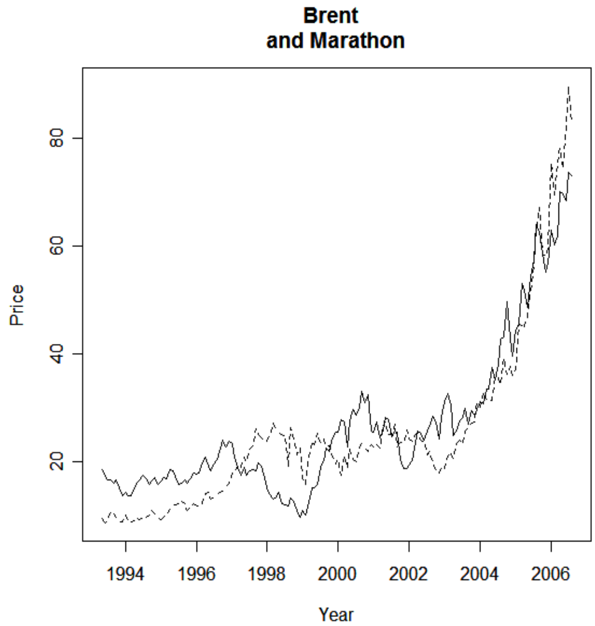
3. Data Analysis

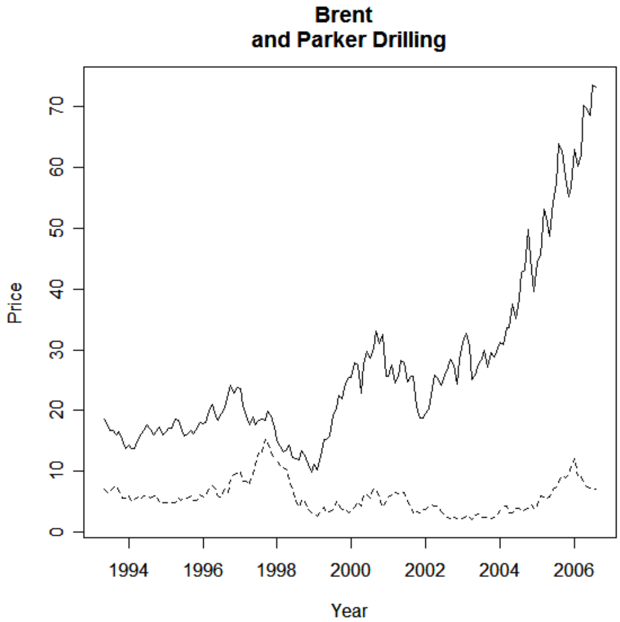
3.1 Non-Stationarity

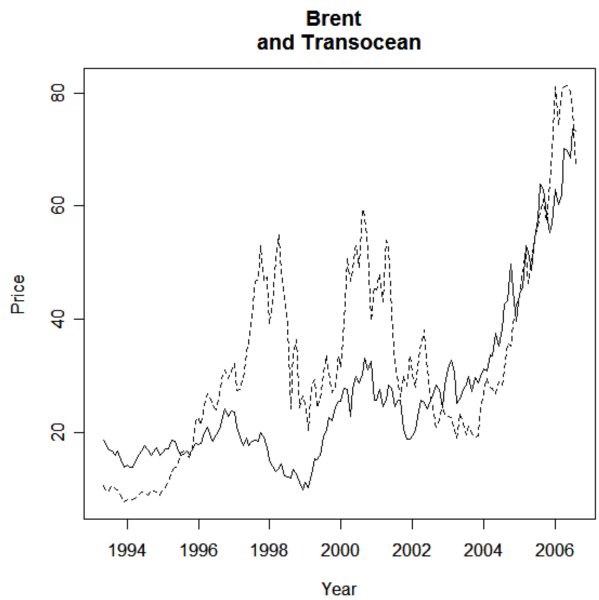
The first step was to determine that all the variables were non-stationary. Visual inspection seems to indicate this as the time series plots below show (Brent Crude is the heavy line in each):

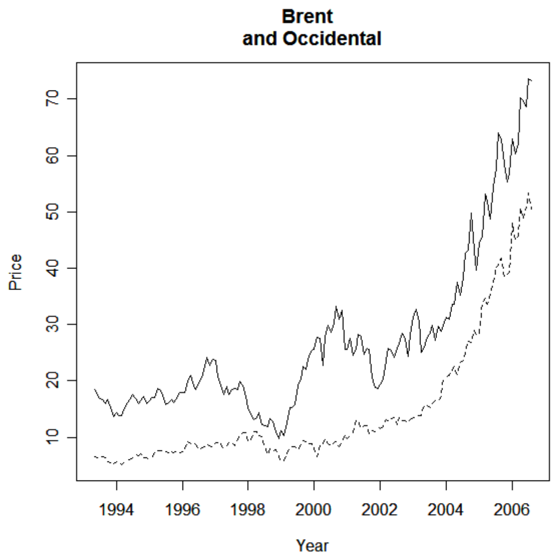


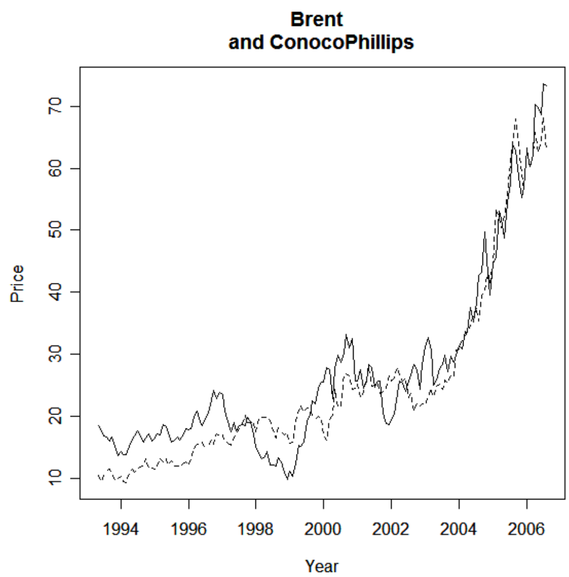


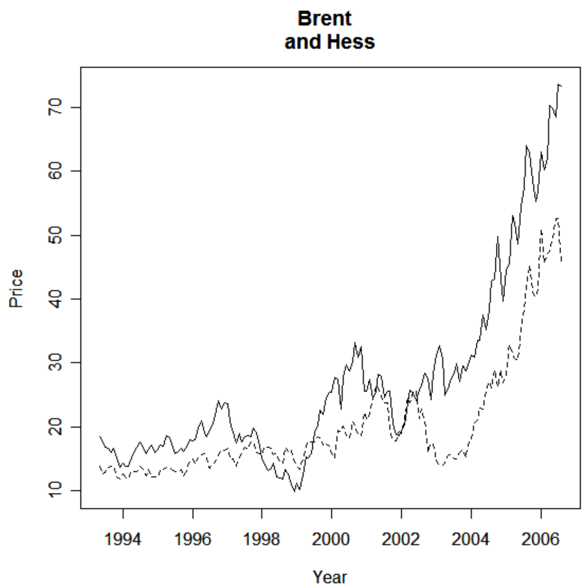


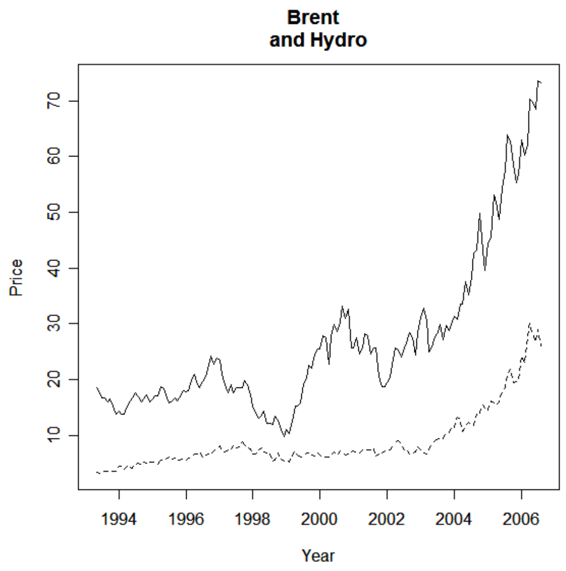


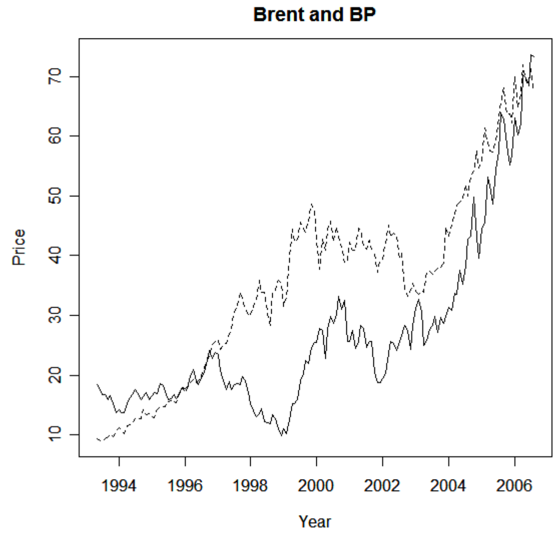

Brent
and Transocean
Price
Year

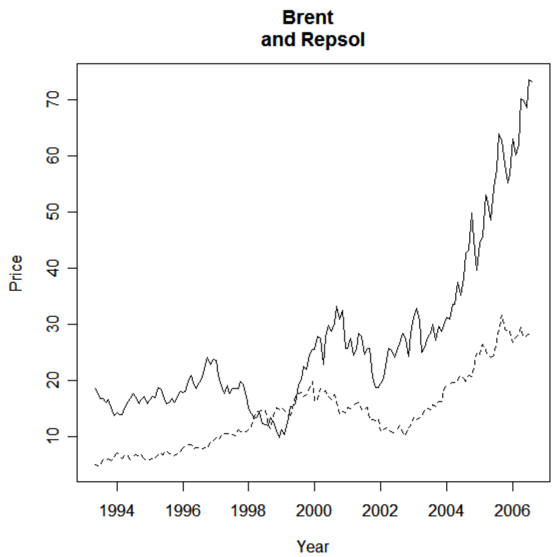

Brent
and Occidental
Price
Year

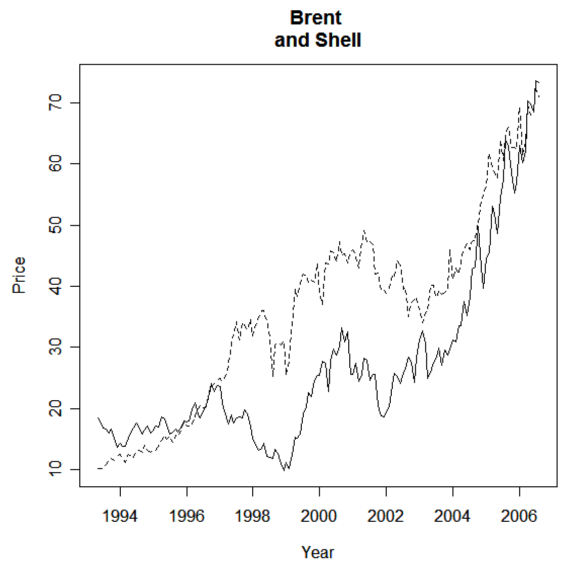

Brent
and ConocoPhillips
Price
Year

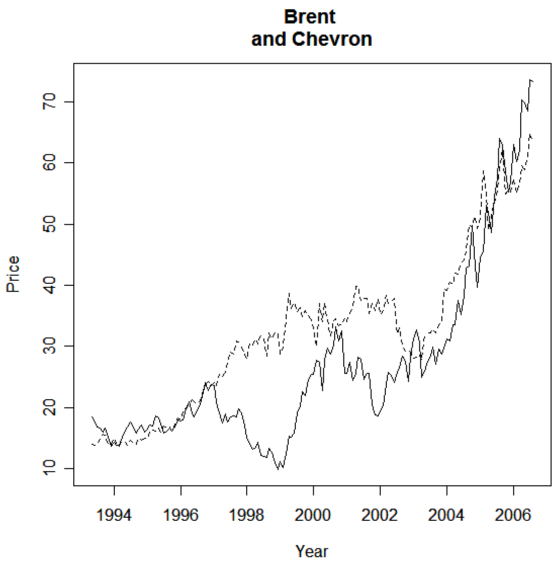

Brent
and Hess
Price
Year

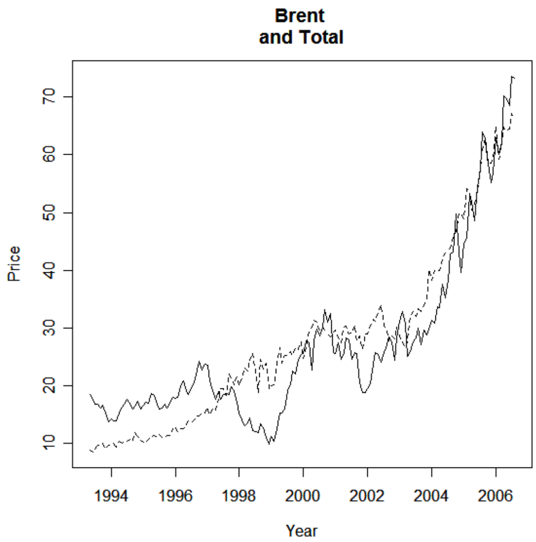

Brent
and Hydro
Price
10
20
30
40
50
60
70
1994
1996
1998
2000
2002
2004
2006
Year

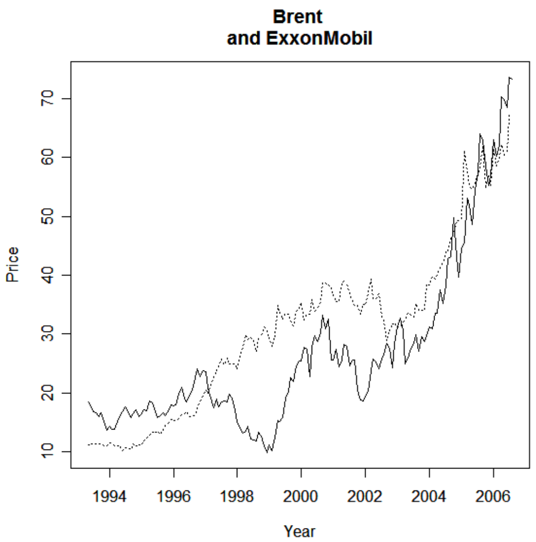

Brent and BP
Price
10
20
30
40
50
60
70
1994
1996
1998
2000
2002
2004
2006
Year

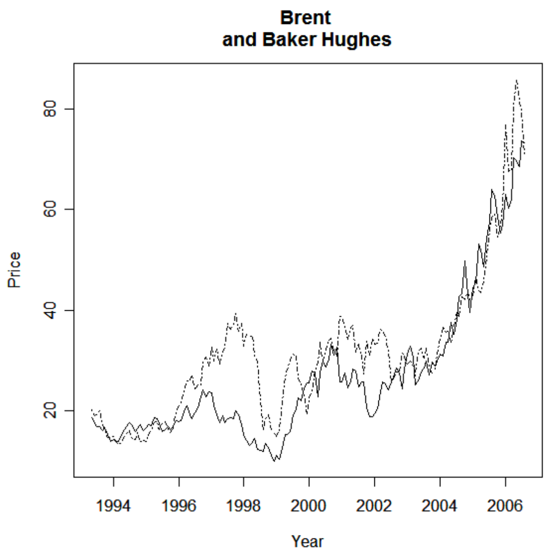

Brent
and Repsol
Price
10
20
30
40
50
60
70
1994
1996
1998
2000
2002
2004
2006
Year

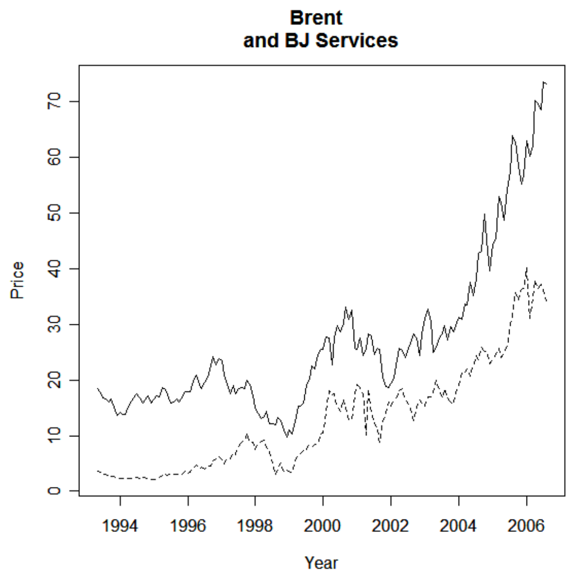

Brent
and Shell
Price
10
20
30
40
50
60
70
1994
1996
1998
2000
2002
2004
2006
Year

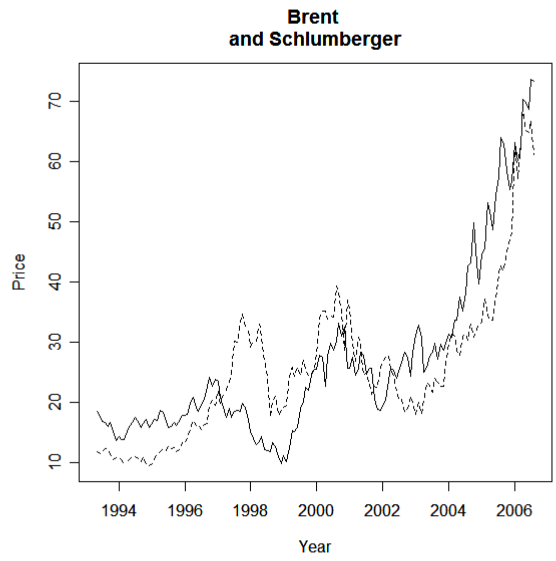

Brent
and Chevron
Price
Year
1994
1996
1998
2000
2002
2004
2006
10
20
30
40
50
60
70

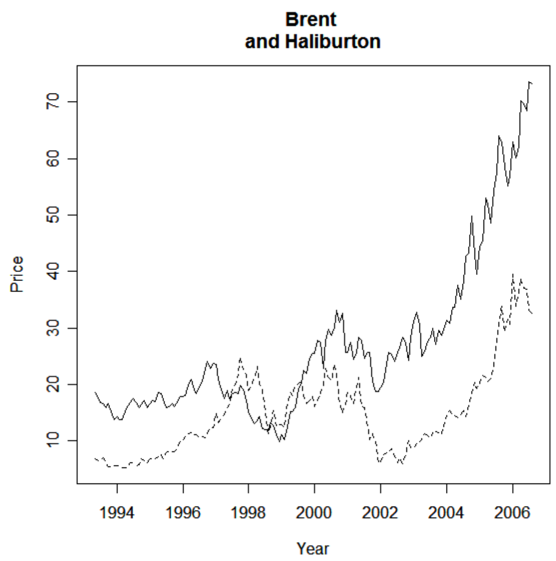

Brent
and Total
Price
Year
1994
1996
1998
2000
2002
2004
2006
10
20
30
40
50
60
70

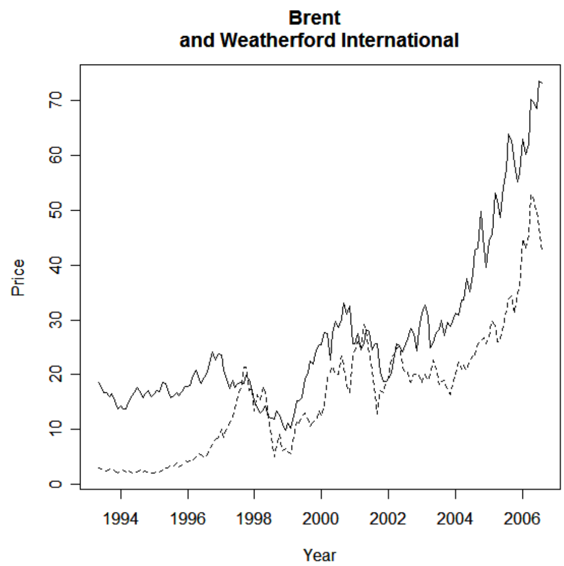

Brent
and ExxonMobil
Price
Year
1994
1996
1998
2000
2002
2004
2006
10
20
30
40
50
60
70


Brent
and Baker Hughes
Price
Year
1994
1996
1998
2000
2002
2004
2006
20
40
60
80









All data were found to be non-stationary using the ADF test. The results of these are summarized in the following table:

Data Series	**ADF Test Statistic**	**"P-Value" Interpreted by R**	**Test Result**
Ensco International	-2.6384	0.3101	Cannot discard H_0
Nabors Ltd.	-3.0937	0.1202	Cannot discard H_0
Marathon Oil	2.7348	0.9999	Cannot discard H_0

Parker Drilling	-2.1519	0.5130	Cannot discard H_0
Transocean Rig	-1.6146	0.7370	Cannot discard H_0
Norsk Hydro A.S.	2.9216	0.9999	Cannot discard H_0
ConocoPhillips	-0.3222	0.9886	Cannot discard H_0
Repsol	-1.0582	0.9262	Cannot discard H_0
Hess	-0.8373	0.7377	Cannot discard H_0
Occidental	2.935	0.9999	Cannot discard H_0
ExxonMobil	-0.578	0.9772	Cannot discard H_0
Shell	-1.4495	0.8060	Cannot discard H_0
Total	0.3993	0.9999	Cannot discard H_0
BP	-1.4016	0.9487	Cannot discard H_0
Chevron	-0.9164	0.9487	Cannot discard H_0
BJ Services	-2.1795	0.8259	Cannot discard H_0
Haliburton	-1.6131	0.7377	Cannot discard H_0
Weatherford Intl.	-2.6743	0.2951	Cannot discard H_0
Baker Hughes	-1.4912	0.7885	Cannot discard H_0
Schlumberger	-1.5611	0.7594	Cannot discard H_0
Brent Crude	0.2086	0.9999	Cannot discard H_0

The evidence that all the series are I(1) is strong, as the hypothesis that they are non-stationary cannot be discarded at a 5% significance level.

3.2 Cointegration

Visual inspection of the time series plots indicates strongly that the oil companies move with the oil price, but testing is required to make the models. The procedure for testing was described in the last chapter, but the procedure in R was to run a linear regression using the Brent time series as the dependent variable and company x as the independent variable, collecting the residuals from that model and running testing them for stationarity. In this case because we are performing regressions on residuals to test for stationarity the distributions for raw data found in the R package no longer applies. The test statistic band is the same for all series, as they are the same length and they all contain the same number of independent variables. It has been taken from Engel and Granger[10]. The tables did not include a statistic for 159 observations, so I used the critical statistic for 150 observations. I considered linear interpolation to be unnecessary as the difference between the statistic for 150 observations and 300 observations was 0.01. The results for each company are summarized in the table below:

Company Name	**ADF Test Statistic from residuals**	**Critical T-Statistic (upper and lower bands)**	**Test Result**
Ensco International	-1.3326	U: -3.26 L: -3.43	Cannot discard H_0
Nabors Ltd.	-1.0775	U: -3.26 L: -3.43	Cannot discard H_0
Marathon Oil	-2.4716	U: -3.26 L: -3.43	Cannot discard H_0
Parker Drilling	0.2105	U: -3.26 L: -3.43	Cannot discard H_0
Transocean Rig	-1.4694	U: -3.26 L: -3.43	Cannot discard H_0
Norsk Hydro A.S.	-3.0381	U: -3.26 L: -3.43	Cannot discard H_0
ConocoPhillips	-2.8645	U: -3.26 L: -3.43	Cannot discard H_0
Repsol	-1.2599	U: -3.26 L: -3.43	Cannot discard H_0
Hess	-3.2219	U: -3.26 L: -3.43	Cannot discard H_0

[10] Engle, R. and Granger, C. (eds.) *Long-Run Economic Relationships*, Oxford University Press 1991 p. 327

Occidental	-2.8776	U: -3.26 L: -3.43	Cannot discard H_0
ExxonMobil	-1.1647	U: -3.26 L: -3.43	Cannot discard H_0
Shell	-0.3310	U: -3.26 L: -3.43	Cannot discard H_0
Total	-1.0642	U: -3.26 L: -3.43	Cannot discard H_0
BP	-0.4794	U: -3.26 L: -3.43	Cannot discard H_0
Chevron	-1.3608	U: -3.26 L: -3.43	Cannot discard H_0
BJ Services	-1.9007	U: -3.26 L: -3.43	Cannot discard H_0
Haliburton	-1.8625	U: -3.26 L: -3.43	Cannot discard H_0
Weatherford Intl.	-1.9392	U: -3.26 L: -3.43	Cannot discard H_0
Baker Hughes	-2.7457	U: -3.26 L: -3.43	Cannot discard H_0
Schlumberger	-2.2667	U: -3.26 L: -3.43	Cannot discard H_0

As none of the test statistics are greater in absolute value that the critical values. There is no cointegration between the selected oil company shares and the Brent spot oil price.

3.3 ECM Modelling

Unfortunately none of the oil company shares were able to go beyond step 1 of the Engel and Granger two step method. This has rendered the second research question moot. I will address this further in the concluding remarks.

4. Conclusion

The research question in the subtitle contains two parts. It is clear now that the answer to the first part seems to be no. Oil company shares are not individually cointegrated with the oil price. The result is that the second question cannot be answered, because the models required to answer it cannot be built. However, in scientific research one must keep an open mind and be prepared for the rejection of a hypothesis. It does not necessarily mean that the research has been futile and there are always lessons to be learned form failure.

The scope of the empirical research has made some headway in terms of answering the question of using share prices to forecast the oil price. The selection of companies was presented a significant share of the available data for Energy Company shares traded in USD. As such it is fairly strong evidence to suggest that oil company shares are not individually cointegrated with the oil price. The research could of course be expanded to include all publicly traded companies, and in time, companies that have only recently gone public, whose data sets are too small to discern any long-term relationships.

One surprising lesson, at least for the author, was that oil services companies had share prices that tended to move with the oil price a lot more consistently than the super majors. It's a strong indication that the orders and expected future cash flows are quite dependent on the oil price. The expected cash flows for owners of the super majors seem to be quite independent of the oil price. ExxonMobil and BP in particular showed strong growth during the 1996-1999 oil price slump. The most likely explanation is their chemical business, which is countercyclical to the oil price. As the oil price falls industrial activity becomes cheaper because energy is cheaper and the demand for industrial chemicals grows. It seems to act as a hedge, which is an interesting insight for investors who might steer away from super majors if they expect higher oil prices, because some of the gains will be hedged away by the chemicals sector.

On the whole it appears that these companies aren't invested directly enough in oil to tend towards a long-term equilibrium with the oil price. Part of the explanation may be that they are involved in the exploration and production of natural gas. Oil and Gas are energy substitutes, but gas is less mobile. It usually has to be transported through pipelines, and converting it into liquid petroleum gas (LPG), for shipping and haulage, is costly. Price setting in gas markets is therefore usually in regional markets. So oil and gas will affect expected future cash flows differently. This is true for all the companies in the sample, even the drilling firms, as they drill for gas as well as oil. An interesting example to look at in the future is Rosneft, the Russian oil company formed mostly with the confiscated assets of Yukos. Its portfolio consists almost entirely of oil, but it hasn't been on the market for long enough to be a part of the sample in this analysis.

The failure to find cointegrating relationships along the lines outlined earlier in this paper may of course be the result of the causality problem. It may be that the shareholders are not more forward looking than oil purchasers; this would negate the whole premise for using share prices as the dependent variable. However, shares are purchased for investment and speculation and oil is purchased on the spot market for consumption and speculation. Ignoring the speculative motive, it exists for both and is anyway likely to be short sighted, it is highly likely that investment decisions are more forward looking than consumption decisions.

The fact remains, however, that energy company share prices are not cointegrated with the oil price. This paper must therefore conclude that cointegration requires that the data sets move together to a greater extent than what one can see in the time series plots. The significant sample of oil companies suggests that in general one will not find cointegration of oil company stock and oil price. One might expect this because there are so many other factors that play into the share price of a company, for example, even if a company only produces and sells crude, investors will also judge it based on how well it does so, and that is not related to the oil price. Oil company shares are not purely based on future oil price expectations in the way that oil futures are, so it is less likely that they are cointegrated with the oil price. There are still avenues to explore however, Rosneft has been mentioned, but one might look at other commodities and the share prices of its producers and sellers. Hershey and the price of cocoa would be an example. At any rate, the most important conclusion from this work is that the presence of cointegration requires a very strong relationship between the regressor and the regressand.

References

A. Banerjee, J. J. Dolado, J. W. Galbraith, and D. F. Hendry: *Cointegration, Error Correction, and the Econometric Analysis of Non-Stationary Data*, Oxford University Press, Oxford, 1993

Brooks, C. *Introductory Econometrics for Finance*, Cambridge University Press, 2002

Chatfield, C. *The Analysis of Time Series: An Introduction* Chapman & Hall/CRC 2003 Sixth Edition

Engle, R. and Granger, C. (eds.) *Long-Run Economic Relationships*, Oxford University Press 1991

Lanza, A. Manera, M. Grasso, M. and Giovanni, M *Long-Run models of oil stock prices* Science Direct 2005

Wårell, L *Market Integration in the International Coal Industry: An Error Correction Model* Luleå University of Technology 2002

Data sources:

http://finance.yahoo.com

http://tonto.eia.doe.gov/