UNIVERSITY OF PIRAEUS

DEPARTMENT OF BANKING AND FINANCIAL MANAGEMENT

M.Sc. IN FINANCIAL MANAGEMENT AND BANKING (FULL TIME)

MASTER'S THESIS

Can the Term Structure of Petroleum Futures be forecasted in Weekly Horizons?

Dimitrios Michalopoulos

SUPERVISOR: Dr. George Skiadopoulos

COMMITTEE:
Angelos Antzoulatos
Nikitas Pittis
George Skiadopoulos

TABLE OF CONTENTS

CHAPTER 1: INTRODUCTION

1.1 PURPOSE

Modeling the futures curve of various assets is a matter of extensive investigation in financial literature. In the following study we will examine the evolution of the term structure of petroleum futures. Especially, our research will be focused on crude oil, heating oil and gasoline futures traded on New York Mercantile Exchange (NYMEX) and crude oil futures traded on Intercontinental Exchange (ICE). Furthermore, we will use two different approaches in order to derive two models which could have explaining and predictive power over our data set.

The procedure of description the oil futures term structure is a point of great interest because it can be used by the energy market participants who aim to use this kind of analysis for hedging and speculative purposes, assuming some specific trading strategies. In a more theoretical academic level there is also a great number of applications including pricing of derivatives and the valuation of parameters related to oil and its byproducts. So, a model suitable to describe the dynamics of future prices curve could have serious impact on the estimation of factors like convenience yields and storage costs, that are of great importance in testing already existing theories and the derivation of new methodologies. As far as our paper is concerned, we will try to estimate the weekly term structure of the petroleum futures mentioned before so as to offer some evidence that would be useful to the pursuit of similar purposes both on economic and academic environment.

1.2 LITERATURE REVIEW

In accordance with our introduction it is widely recognized that predictive models and their efficient performance consist an important volume of the financial literature and a variety of different assets are involved in relative researches. The obtained results and their theoretical background often motivate a lot of applications, which provide meaningful information to analysts. These applications, in a more general form, have to do with assets pricing, trading strategies planning and risk valuation. Consequently, the researches try to identify the dynamics of the term structure of studied assets introducing different methodologies and approaches. Usually, the subsequent step is to check the forecasting power of a theoretically well-established model.

Of course, it is not possible to have a full review of the literature and this is not the target of the forthcoming study. Instead of this we decide to present a restricted reference to some articles, which are representative of different methodologies or are somehow connected with the object of the paper. To organize better the structure of the review we can specify two main categories or types of forecasting models. The discretization of the models is related to the kind of independent variables that each model takes into consideration. In this section a distinction between the two approaches will be provided.

The first approach is known as time series forecasting approach and involves trying to predict the future prices of series taking its previous values and/or previous values of an error term as the explanatory variables. Then, using statistic methods and some kind of data manipulation, the researcher tends to come out with a model assumed to have the necessary characteristics. These models have the advantage to be less time and effort consuming as far as data collection is concerned, because they demand only the records for the variable under consideration, but the retained from the mathematical procedure parameters are often of no meaning in economic terms.

The second approach, named economic (structural) approach, identifies a relation between the dependent variable and some other economic or financial factors, like the risk free interest rate, the market's portfolio return, commodities prices and everything else could influences the studied asset's price. Significant drawbacks of these models are the volume of data to be handled and the identification of the appropriate causal variables.

Hereafter, a variety of individual works will be commented in a comprehensive way. The following researches introduce different methods which are applied on a spectrum of assets.

Moving on to models based on the first approach we can start with Taylor's (1980) article. He proved that the random walk model doesn't stand for the daily returns of some commodities. Moreover, he introduced a price trend model which seemed to be more accurate in the explanation of the observed autocorrelations and the description of the observed returns.

Jegadeesh (1990) used cross sectional regressions in monthly stock returns and found them positive or negative correlations related to monthly returns at time-varying lags. Then, he used his findings to consider trading strategies which were profitable.

The monthly US stock returns were the studied asset in Conrad's and Kaul's (1989) paper. They concluded that as much as 25% of the variation in monthly portfolio returns could be explained by the behavior of the previous month's returns. Similarly, Chelley-Steeley (2001) presented an ARMA (1,1) model which could determine up to 15% of the variation in the monthly returns of a UK stock portfolio. The innovation of these two papers is the fact that the authors used weekly returns as explanatory variables in order to describe monthly returns. So, they were able to impose increased weights and, therefore, explanatory value to most recent returns.

Models similar to the two mentioned above are simple versions of the first approach which have been used throughout the literature. A review of these methods was made by Cochrane (1999). We should note that Cochrane's paper refers also to occasions of the second approach.

Continuing with more complex models we provide some further references to Principal Components Analysis (PCA), which, otherwise, is one of the methodologies to be used in the following dissertation.

Schwartz and Cortazar (1994) studied the curve of two copper future contracts traded on Commodity Exchange of New York (COMEX). They used daily futures returns and apply PCA on their data so as to obtain a three factor model adequate to reproduce the stochastic movement of futures prices. The three factors were capable of capturing the 93%, 4% and 1% of the total variance of the returns, respectively. The first of them was fairly stable across the contracts' maturities represents a level factor responsible for shifts or falls of the futures curve. The second one, characterized by steepness, drives the short term and long term returns to opposite directions. The third factor illustrates the effect of shocks that influence medium term futures in opposite way compared to short and long term futures. It is worth to be mentioned that this evidence is in line with the terminology introduced by Litterman and Scheinkman (1991) about the retained principal components.

Tomalsky and Hindanov (2002) extended this work to the petroleum futures markets. To be more specific they used the method to describe the term structure of log returns of crude and heating oil futures traded on NYMEX. They come up with the same three factors when they performed PCA separately in crude and heating oil futures. These factors were again able to explain 99.89% and 99.63% of the total variance respectively. The joint application of the technique resulted in the derivation of four factors that explain 99.36% of the total variance. They also tested the seasonality effect on these contracts without to derive a rigorous statistical conclusion.

Jarvinen (2003) added to these researches the cases of Brent crude oil and pulp. The data set is collected by monthly observations. The point where this paper is primarily differentiated is the use of par swap rates rather than the futures prices for the estimation of the forward curve through PCA. As a result of this practice, how Jarvinen assumes, the first three obtained factors show a more complex behavior than the one introduced by Litterman and Scheinkman (1991). Furthermore, these factors cannot explain up to 90% of the total variance either for crude oil or pulp.

The former papers are not related to the effort of forecasting future prices but their main target has to do with the identification of the forward curve. A forecasting model has examined by Cabiddo and Fiorenzani (2004), who applied PCA on the time series of the prices of twelve Brent future contracts traded on ICE. The results are comparable to previous researches as the first three components represent level, slope and convexity changes, respectively, and they explain more than the 99% of the total variance. Something that deserves to be mentioned is the conclusions of the authors that there is some kind of interaction between the components, but the use of vector autoregressive models to exploit this evidence doesn't improve the forecasting performance of the method, at least when the research is restricted to the macromovements of the curve which are described by the slope, steepness and curvature.

Chantziara and Skiadopoulos (2006) studied the term structure of petroleum futures trying also to examine the predictive power of PCA. Their data set consists of futures traded on NYMEX and on IPE and they found similar results in accordance with the existing literature and they formed some conclusions about the spillover effects between these markets.

On the other hand the second approach affects the methodology of several papers motivating the researchers to test the presence of numerous variables for explanatory purposes.

A brief categorization of this kind of models was made by Connor (1995). His study is associated with models designed to capture sources of predictability, like macroeconomic, statistical and fundamental factors, in stock returns.

Pesaran and Timmerman (1995) examined the robustness of the evidence on predictability of US stock returns by simulating the decision process of an open minded investor who uses historically available information to select a set of independent economic factors.

Min Qi (1999) extended this report using as independent variables nine economic and financial variables in order to examine the predictability of S&P 500 index. His major innovations lay on the fact that his approach allows the investor to select not only between different causal parameters, but also between various functional forms through which stock returns could be forecasted. Min Qi involves the use of neural networks to the derivation of his model.

Feedforward neural networks have been incorporated by Gencay and Stengos (1998). Their article proposes that feedforward network models gain in predictability when simple buy-sell signals, based on return and volume indicators, are included in them. Their data panel consists of the Dow Jones Industrial Average (DIJA) daily returns.

The second approach includes two articles written by Stock and Watson (2002a, 2002b). These researches are based on Principal Components Analysis but they used this method in a different way than the one we have already mentioned. Instead of applying PCA straight on their dependent variables, Stock and Watson estimated their principal components from a dataset of 215 and 149 monthly macroeconomic variables respectively. So, these variables are, substantially, their explanatory variables and through their decomposition they introduce models, which use the obtained factors so as to create forecasts for eight basic macroeconomic indicators of the US, in the first paper, and for the Reserve Board's Index of Industrial Production in the second study.

Moving to commodity futures case we should refer to Schwartz (1997). He developed one, two and three factor models, where the first factor is the spot price of the commodity, the second factor the instantaneous dividend yield and the third factor the stochastic interest rates, in order to examine each model's contribution to the derivation of futures curve of his dependent data set. This set included weekly prices of future contracts on oil, copper and gold. Schwartz concluded that the two and three factor models outperform the one factor model for various futures maturities.

Later, Miltersen and Schwartz (1998) developed a differentiated form of Schwartz's three factor model. They distinguished between futures and forwards and their subsequent stochastic convenience yields. They apply the new model on pricing options which had futures on copper as underlying asset. This discrimination of the parameters and also the time lag between the maturities of the options and the underlying futures resulted in different prices for the options.

Clewlow and Strickland (1999) introduced a one factor model, similar with the one derived by Schwartz (1997). They take into account the spot price, the initial forward curve and the volatility function parameters and used the model to price derivatives. The authors described a way of pricing American and exotic energy derivatives using trinomial trees.

Audet, Heiskanen, Keppo and Vehviläinen (2004) derived a model oriented to the identification of the term structure of electricity futures and forwards traded on Nordic market. Their data consisted of weekly prices from Nord Pool and they ended up with three applications of the model: conditional forecasting of the forward curve, when forecasts for the spot price is available, pricing of options on electricity forwards and futures, checking the accuracy of a forward curve model that uses only a finite number of curve points and therefore is disable to capture all the uncertainties between these points.

In 2004 Ribeiro and Hodges developed Schwartz's two factor model by ruling out arbitrage opportunities and considering time varying spot price and convenience yield volatilities. Their data set consisted of weekly observations of light crude oil futures traded on NYMEX and they found that their model slightly outperform Schwartz's one.

More recently Khan, Khokhor and Simin (2006) contribute a model, which purpose is to describe the dynamics of futures on commodities term structure in order to estimate the correct risk premia of these various commodities. They used a data set with composition similar with that studied by Schwartz (1997) consisted of weekly prices of NYMEX futures on crude oil, copper, gold and natural gas and they derive a regression model with the inventory levels, the spot price and the net hedging pressure as explanatory variables, which examine the relationship between these factors and futures weekly returns.

An interesting result as far as predictability is concerned, has been presented by Sadorsky (2002). He found that macroeconomic risk factors have significant forecasting power in petroleum future markets. Using an ARMAX-ARCH model and four macroeconomic indicators as causal factors he managed to capture both the correct sign and magnitude of monthly returns of crude oil, heating oil and unleaded gasoline futures traded on NYMEX.

The major target of the following paper is to contribute a comparison between two aspects of the approaches mentioned above. We will use two specific articles as guide lines for the models that will be derived. The first model will be based on Principal Components Analysis in the form that is proposed in Chantziara, Skiadopoulos (2006) and the second model will be consistent with the parametric method introduced by Sadorsky (2002). The data set will be collected from futures on crude oil, heating oil and gasoline traded on NYMEX and crude oil futures traded on ICE, referring to the variables to be forecasted. For the explanatory parameters we will follow the same patterns as proposed by Sadorsky (2002).

As a result of the investigation we will examine and compare the explaining and forecasting performance of each model. Furthermore, the data set will be obtained from weekly observations so as to differentiate the sample compared to daily or monthly prices used in Chantziara, Skiadopoulos (2006) and Sadorsky (2002) respectively. Additionally, the presence of spillover effects throughout the petroleum markets will be testified, just like in Chantziara's etc. (2006) paper.

1.3 PETROLEUM MARKETS

Firstly we will derive the model inspired by Chantziara, Skiadopoulos (2006) approach. To continue with this research it would be helpful we provide an overview of the underlying commodities and their trading behaviour, aiming to offer a more complete understanding of the petroleum market properties. Obviously, this practice will be extended to a more detailed reference to the oil futures and the relative markets (NYMEX, ICE), including the trading and settlement procedures.

To be more consistent with the previous literature we will present a restricted reference to patterns that have identified through oil and other commodity futures markets. This analysis will give us the opportunity later to test the results of our study with widely recognised properties of commodity futures and to see if they agree with patterns like backwardation and contango. We will also offer the theoretical background to check the presence of spillover effects.

1.4 TIME SERIES APPROACH

As an introduction to our parametric construction we would like to note that the data set will contain the future contracts which are available every time in the market and have classified with respect to their maturity horizon. To be more specific we will preserve the data obtained in Chantziara and Skiadopoulos (2006) paper. Thus, we will use the same generics allowing the rolling over contracts so as to keep in the same series futures with almost fixed maturities. These generics will be obtained from the first fifteen, by shortest maturities, crude oil futures traded on NYMEX (CL1 - CL15), the first twelve heating oil futures traded on NYMEX (HO1 – HO12), the first eleven gasoline futures traded on NYMEX (HU1 – HU11) and the first nine crude oil futures traded on ICE (CO1 – CO9). These series will be statistically tested in order to estimate their correlations for the contracts with the same maturities and to identify their price movements. Of course, the statistical interpretation will include the test of stationarity, which is necessary for the application of PCA. This way we will decide if we can use the futures prises or we will need to take as variables futures first differences or returns. When we have ended up with our final data series we will check them in terms of normality, skewness, kurtosis, etc.

The following section will be a general description of PCA. Main result of this analysis is the reduction of data's dimension. This can be done by transforming the correlated variables of our data panel to a new set of uncorrelated variables (principal components, PC's) and then dropping out the PC's which explain a very small amount of our original data variance.

After that we will perform PCA both separately and jointly to our futures time series so as to come up with the PC's that we will keep in order to derive the regression models which will describe the dynamics of the term structure of our dependent variables. These regressions will take into account the already obtained PC's at time t-1, as independent variables, so as to identify their forecasting adequacy over the dependent variables (futures differences or returns) at time t. From the forthcoming results it would be crucial we form our conclusions through the comparison with Chantziara, Skiadopoulos (2006) results, so as to detect possible disagreements emerging from the use of weekly observations instead of daily.

1.5 ECONOMIC APPROACH

At the second part of our research we will reproduce a parametric approach similar to the one introduced by Sadorsky (2002). According to previous literature asset returns in other markets, like stock or bond, can be related to macroeconomic factors. This idea had been extended to commodity future markets by Bessembinder and Chan (1993), Baum and Barkoulas (1996), Bjornson and Carter (1997). Sadorsky's contribution to this methodology was the derivation of an ARMAX-ARCH model which was designed to face the residuals heteroskedasticity, non-normality and serial correlation. This model describes the dynamics of crude oil, heating oil and unleaded gasoline futures on NYMEX monthly returns assuming as independent variables some macroeconomic factors. Sadorsky managed to construct a model which seemed to be reliable as far as the explanation point is concerned and also effective for predictive purposes over some specific time horizons.

Moving on, we will construct a similar model and use the same macroeconomic indicators as explanatory parameters. These factors are the monthly return on the annual dividend yield on the S&P 500 common stock portfolio, the monthly return on the annual yield on Moody's long term BAA-rated bond minus the yield on AAA-rated bonds, as an interpretation of the default risk premium, the monthly return on the annual yield on the 90-day Treasury bill, as an estimation of the risk free interest rate and, as a fourth parameter, an approximation of the monthly excess return of the market portfolio. As a matter of fact, it would be of great importance for our study we use the before mentioned parameters in weekly time steps and take into consideration similar variables for the UK economy in order to identify spillover effects, but mainly to explain in amore adequate way the dynamics of crude oil futures traded on ICE.

Consequently we need to come up with a statistical analysis of the independent variables. A same analysis for the futures returns will have been already done during the derivation of the first model in this paper, so it doesn't make sense to repeat the procedure and we have to check the explanatory variables for normality and stationarity.

The next step will be a brief presentation of models that could be used to describe the dynamics of conditional futures returns using as explanatory factors the lagged macroeconomic variables. Then, our study will take into account the procedure introduced by Sadorsky and we will conclude to the most appropriate model. Sadorsky proposed an ARMAX-ARCH model, which resulted in properly specification of the residuals.

The author tested the out-of-sample performance of this model and he discovered that it consistently outperforms a driftless random walk model. In addition to this, Sadorsky's model succeeds in forecasting returns' direction and magnitude for various time steps. So, we will test the performance of our model in a similar way trying to check its forecasting power considering its root-mean-square-error compared to

RMSE of a driftless random walk model and another criterion to check the model's ability in direction predictions. Last test for the model's efficiency will be the planning of a simple trading strategy, following Sadorsky's example, so as to see if this strategy could be profitable. Finally we will compare our findings with Sadorsky's results and see if his model's forecasting power is confirmed by our effort.

1.6 CONCLUSIONS

To conclude with the results of the research we will summarize the two models we derived. We will focus on each model's advantages and disadvantages in order to compromise with the challenging topic of explaining and forecasting the weekly dynamics of the term structure of petroleum futures. This way we could have the appropriate evidence to come up with the more effective between these two models, which would have been evaluated under various criteria. Our findings will be compared with the literature on commodity futures or other assets in general in a theoretical level. Finally, we will concentrate our study in the weaknesses of the research. Moreover, we will refer to possible extensions which could be targets of further investigation and will not be included in our paper.

CHAPTER 2: PETROLEUM PROPERTIES

2.1 INTRODUCTION

Oil is the world economy's most important source of energy and is, therefore, critical to economic growth. Its value is driven by demand for petroleum products which are derived from the different types of crude oil through the refining process. The largest sources of supply are Saudi Arabia, Russia, the United States, Iran, Mexico, China, and Europe's North Sea.

The Organization of Petroleum Exporting Countries (OPEC), an international cartel of oil-producing countries, is the single most important production-related entity. It produces about 40 percent of the world's daily consumption of crude oil.

Petroleum products power virtually all motor vehicles, aircraft, marine vessels, and trains around the globe. In total, products derived from oil, such as motor gasoline, jet fuel, diesel fuel, and heating oil, supply nearly 40 percent of the energy consumed by households, businesses, and manufacturers worldwide. Natural gas and coal, by comparison, each supply less than 25 percent of the world's energy needs.

So the crude oil value and characteristics are of great importance for the global economy and it is crucial we present a brief analysis of these elements that are determined by two main factors: location and quality and this way the crude oil "markers" or "benchmarks" are introduced.

Crude oil price Benchmarks were first introduced in the mid 1980's. The most widely used crude oil price benchmarks in the world are West Texas Intermediate (WTI), used primarily in the U.S; Brent, used primarily in Europe; and the Organization of Petroleum Exporting Countries (OPEC) market basket, used around the world. (Other benchmarks, like Dubai, also known as Fateh, are used in Asia.) WTI is very light and very sweet. This makes it ideal for producing products like low-sulfur gasoline and low-sulfur diesel. Brent is not as light or as sweet as WTI but it is still a high-grade crude. The OPEC basket is slightly heavier and sourer than Brent.

2.2 CRUDE OIL CHARACTERISTICS

The quality of crude oil is determined by a number of characteristics that affect the proportions of transportation fuels and petroleum products produced when the oil is refined. The two most common measurements of crude oil quality are the specific gravity (which is measured in degrees) and the sulfur content of the oil. Acid content is also a factor in determining the corrosive properties of the crude oil entering the refinery.

Specific Gravity

The specific gravity is typically measured using the American Petroleum Institute (API) standard or the API gravity of the crude oil. The API gravity is the measure of the weight of crude oil in relation to the weight of water (water has an API gravity of 10 degrees). Crude oil is characterized as heavy, intermediate, or light with respect to its API gravity.

Heavy Crude: Crude oils with API gravity of 18 degrees or less is characterized as heavy. The oil is viscous and resistant to flow, and tends to have a lower proportion of volatile components. Fifty one percent of California crude oil has an average API of 18 degrees or less.

Intermediate Crude: Crude oils with an API greater than 18 and less than 36 degrees are referred to as intermediate. Forty eight percent of California crude oil has an average API between 18 and 36 degrees.

Light Crude: Crude oils with an API gravity of 36 degrees or greater. Light crude oil produces a higher percentage of lighter, higher priced premium products.

Sulfur Content

Crude oil is defined as “sweet” if the sulfur content is 0.5 percent or less by weight and “sour” if the sulfur content is greater than 1.0 percent. Sulfur compounds in crude oil are chemically bonded to hydrocarbon molecules in the oil. Additional equipment in the refinery is required to remove the sulfur from crude oil, intermediate hydrocarbon feedstocks, and finished products. Transportation fuel specifications require extremely low sulfur contents, usually less than 80 parts per million (ppm).

Acid Content

Another characteristic of crude oil is the total acid number (TAN). The TAN represents a composite of acids present in the oil and is measured in milligrams (mg). A TAN number greater than 0.5 mg is considered high. However, some acids are relatively inert. Thus, the TAN number does not always represent the corrosive properties of the crude oil. Further, different acids will react at different temperatures – making it difficult to pinpoint the processing units within the refinery that will be affected by a particular high TAN crude oil. Nonetheless, high TAN crude oils contain naphthenic acids, a broad group of organic acids that are usually composed of carboxylic acid compounds. These acids corrode the distillation unit in the refinery and form sludge and gum which can block pipelines and pumps entering the refinery. High TAN oils account for an increasing percentage of the global crude oil market. Crude oil with a TAN greater than 1.0 mg increased in the world market from 7.5 percent in 1998 to 9.5 percent in 2003.

2.3 PETROLEUM MARKETS

The principal activities involved in moving crude oil from its source to the ultimate consumer are:

- Production, which involves finding, extracting, and transporting crude oil;
- Refining, the process by which crude oil is turned into products such as gasoline
- Distribution and marketing, which focus on moving those products to final consumers

These activities occur within a global marketplace – an extensive physical infrastructure that connects buyers and sellers worldwide, all supported by an international financial market. It links an international network of thousands of producers, refiners, marketers, brokers, traders, and consumers buying and selling physical volumes of crude oil and petroleum products throughout this chain of production. The international market also includes futures and other financial contracts that allow buyers and sellers to efficiently insure themselves against significant price and other business risks, thereby minimizing the impact of price volatility on their operations. In sum, the global oil market comprises thousands of participants who help facilitate the movement of oil from where it is produced, to where it is refined into products, to where those products are ultimately sold to consumers.

Over the last 25 years, the global oil industry has seen a transformation in the contractual structures used to purchase and sell crude oil. A market structure formerly based on rigid long-term, commercial arrangements has been replaced by a more efficient one that allows buyers and sellers greater flexibility in establishing commercial relationships that better meet their respective needs.

Whereas "spot" and "futures" markets have been long-established institutional structures for many commodities, they are relatively new to the oil industry. Their uses, however, have grown rapidly and are now a well-developed part of the market. Today it is from the spot and futures markets that the global oil market – producers, refiners, marketers, traders, consumers, investment banks, hedge funds, and so forth – receives competitively determined market signals that inform buyers and sellers on current and future supply and demand conditions allowing these market participants to form their own trading strategy in order to pursue speculative or risk hedging purposes.

The term "spot markets" is used to describe transactions which involve the near-term purchase and sale of a commodity, such as crude oil and refined products. In the crude oil market, "spot" contracts typically involve delivery of crude over the coming month, e.g., a contract signed in June for delivery in July. Spot markets are often referred to as the "physical market" since they entail the buying and selling of physical volumes. These markets provide the benefit of allowing buyers and sellers, e.g., refiners and marketers, to more easily adjust their crude supplies to reflect near-term supply and demand conditions in both the product markets and the crude oil markets.

A futures contract, in contrast to a spot transaction, concerns the future purchase or sale of crude oil or petroleum products. Specifically, it is a contract that carries the obligation for delivery of a given quantity of crude in the future. The contract specifies the volume, type or grade of crude oil, the price, the future time in which the crude is bought or sold, and the particular location to which it is to be delivered. The buying and selling of futures contracts occurs on organized exchanges.

At present, major global petroleum futures markets include the New York Mercantile Exchange (NYMEX), London-based ICE Futures (formerly the International Petroleum Exchange) and the emerging Tokyo Commodity Exchange, of which the former two are the most influential. Some information for the two first markets will be mentioned taking into account that this study is concerned about contracts traded in these specific platforms.

The NYMEX, the largest commodity exchange in the world, was established in 1872. It launched the first crude oil futures contract in 1983, based on West Texas Intermediate crude oil. By 2001, it had become the world's highest-volume futures contract. Because of its excellent liquidity and price transparency, the contract is used as a principal international pricing benchmark.

The leading European marketplace for regulated trading of energy contracts, the IPE is a London-based energy futures and options exchange. The IPE lists the benchmark Brent crude oil futures contract, which is relied upon for pricing an estimated two-thirds of the world's traded oil products. The IPE offers fully electronic trading in futures and options on Brent crude oil and gas oil, and futures on emissions, natural gas, and electricity. In June 2001, the IPE became a wholly owned subsidiary of Intercontinental Exchange (ICE), an electronic marketplace for trading both futures and over-the-counter (OTC) contracts in natural gas, power, and oil founded in 1999 and headquartered in Atlanta. In April 2005, the entire ICE portfolio of energy futures became fully electronic, ending the open-outcry form of trading for Brent crude futures, diesel futures and options that had lasted for a quarter century. The ICE is the world's leading electronic and over-the-counter trading market for energy futures contracts, with offices in the United States, Europe and Singapore. In London, Brent crude is traded on both spot and futures markets. This light, sweet crude oil sourced from Britain's North Sea forms a benchmark for crude oil trade in the region and oil exports to Northwest Europe, the North Sea, the Mediterranean, Africa and the Middle East.

2.4 CRUDE OIL FUTURES TRADING

NYMEX Light, Sweet Crude Oil Futures Contract

Crude oil began futures trading on the NYMEX in 1983. It is the most heavily traded futures contract based on a commodity worldwide. One NYMEX Division light, sweet crude oil futures contract consists of 1,000 U.S. barrels (42,000 gallons) and the currency used for pricing purposes is the U.S. dollar, which is used as the medium for most oil futures transactions. Open outcry trading is conducted from 9:00 a.m. until 2:30 p.m. New York time and electronic trading is conducted from 6:00 p.m. until 5:15 p.m. via the CME Globex® trading platform, Sunday through Friday. There is a 45-minute break each day between 5:15PM (current trade date) and 6:00 PM (next trade date). Minimum price fluctuation corresponds to $0.01 (1¢) per barrel or $10.00 per contract whereas maximum daily price fluctuation is $10.00 per barrel ($10,000 per contract) for all maturities. If any contract is traded, bid, or offered at the limit for five minutes, trading is halted for five minutes. When trading resumes, the limit is expanded by $10.00 per barrel in either direction. If another halt were triggered, the market would continue to be expanded by $10.00 per barrel in either direction after each successive five-minute trading halt. The existing crude oil futures are listed nine years forward using the following listing schedule: consecutive months are listed for the current year and the next five years; in addition, the June and December contract months are listed beyond the sixth year. Additional months will be added on an annual basis after the December contract expires, so that an additional June and December contract would be added nine years forward, and the consecutive months in the sixth calendar year will be filled in. As far as the duration of the trading procedure for each contract is concerned we have to mention the following information, trading terminates at the close of business on the third business day prior to the 25th calendar day of the month preceding the delivery month. If the 25th calendar day of the month is a non-business day, trading shall cease on the third business day prior to the business day preceding the 25th calendar day. Settlement is arranged through physical delivery in Cushing, Oklahoma. The deliverable commodities vary because in addition to WTI crude there are also other types of crude assumed to be acceptable[1]. All deliveries are ratable over the course of the month and must be initiated on or after the first calendar day and completed by the last calendar day of the delivery month.

ICE Brent Crude Futures Contract

The trading of Brent crude futures contract began in London International Petroleum Exchange (IPE) in 1988. Its trading unit is 1,000 net barrels (42,000 US gallons) of Brent crude oil and it is quoted in U.S. dollars and the trading hours in the ICE trading platform extend from 01:00 until 23:00 daily London time (20:00 to 18:00 New York time) daily, except Monday morning/Sunday evening when the opening time is 00:00 London (local time) / 19:00 New York (EST). There is a minimum price fluctuation calculated up to one cent per barrel or $10 per contract, but there are no upper limits at daily price fluctuation. The number of the different contract maturities traded each day fluctuates from 61 to a maximum of 72 consecutive months. This happens because twelve additional contract months will be added each year on the expiry of the prompt December contract month. Trading ceases at the close of business on the business day immediately preceding the 15th day prior to the first day of the delivery month, if such 15th day is a banking day in London. If the 15th day is a non-banking day in London (including Saturday), trading terminates at the business day immediately preceding the first business day prior to the 15th day. ICE Brent Futures is a deliverable contract having as underlying asset the current pipeline export quality Brent blend as supplied at Sullom Voe. There is also an option to cash settle against the published settlement price i.e. the ICE Futures Brent Index[2] price for the day following the last trading day of the futures contract.

---

[1] Specific domestic crudes with 0.42% sulfur by weight or less, not less than 37° API gravity nor more than 42° API gravity. The following domestic crude streams are deliverable: West Texas Intermediate, Low Sweet Mix, New Mexican Sweet, North Texas Sweet, Oklahoma Sweet, South Texas Sweet.
Specific foreign crudes of not less than 34° API nor more than 42° API. The following foreign streams are deliverable: U.K. Brent, for which the seller shall receive a 30 cent per barrel discount below the final settlement price; Norwegian Oseberg Blend is delivered at a 55¢–per–barrel discount; Nigerian Bonny Light, Qua Iboe, and Colombian Cusiana are delivered at 15¢ premiums

[2] The Exchange issues, on a daily basis at 12 noon local time, the ICE Futures Brent Index which is the weighted average of the prices of all confirmed 21 day Brent/Forties/Oseberg (BFO) deals throughout the previous trading day for the appropriate delivery months. These prices are published by the independent price reporting services used by the oil industry.
The ICE Futures Brent Index is calculated as an average of the following elements:
i)First month trades in the 21 day BFO market.
ii)Second month trades in the 21 day BFO market plus or minus a straight average of the spread trades between the first and second months.
iii)A straight average of all the assessments published in media reports.

2.5 REFINED PETROLEUM PRODUCTS AND MARKETS

Crude oil needs to be refined in order to produce the gasoline and other products demanded by consumers. Refining a barrel of crude oil involves a series of complex processes. The first stage for all refineries focuses on the initial distillation in which the barrel of crude oil is heated and broken down into its component parts. Subsequent processes, often referred to as "conversion," focus on transforming lower-valued products into higher-valued products by either removing impurities, such as sulfur, or further transforming lower-valued products, such as bunker fuel suited for ships, into higher-valued products, such as gasoline for automobiles. It is the size and scope of these various "conversion" processes that typically distinguish differences in refineries. As a result, different refineries will prefer different types of crude oil.

Conceptually, the market for refined petroleum products is very similar to the crude oil market in that there is widespread buying, selling, and trading of products in both the physical market (e.g., spot market) and the futures market. And just as with crude oil, there are significant international flows of refined products. The United States, for example, imports approximately 3.5 million and exports approximately one million barrels per day of refined products.

Trade in petroleum products reflects the international market's efforts to match what is produced (supply) with what consumers prefer (demand). In the United States, for example, the majority of exports tend to involve products for which there is little or no domestic demand. This would include commodities produced as by-products of the refining process and that are no longer consumed domestically, such as petroleum coke; products for which there is little seasonal demand, such as heating oil sent to the Southern Hemisphere during our summer season; and products for which there is no domestic market due to environmental specifications, such as residual fuel and gasoline that fails regional fuel specifications. Imports, in contrast, reflect domestic demand for products such as gasoline and winter heating oil, i.e., products demanded by U.S. consumers that cannot otherwise be met by domestic refiners.

In addition, petroleum products and futures are also traded on organized exchanges, such as NYMEX and the Chicago Mercantile Exchange, just like crude oil. Thus, the interactions of traders on organized exchanges establish transparent prices for petroleum products, as well as crude oil. Petroleum product deliveries in particular areas will often be at prices based on those determined on an organized exchange, with adjustments for differences in location and the precise type of petroleum product being traded.

This paper will focus on two specific refined petroleum products, gasoline and heating oil and, especially, their futures which are traded on NYMEX. So it is helpful we refer to the characteristics of these products and the contracts on them.

Heating oil

Heating oil, also known as No. 2 fuel oil, accounts for about 25% of the yield of a barrel of crude, the second largest "cut" after gasoline. Its NYMEX futures trading unit is 42,000 U.S. gallons (1,000 barrels) and price quotation is in U.S. dollars. Open outcry trading is conducted from 9:00 a.m. until 2:30 p.m and electronic trading is conducted from 6:00 p.m. until 5:15p.m. via the CME Globex® trading platform, Sunday through Friday. There is a 45-minute break each day between 5:15 p.m. (current trade date) and 6:00 p.m. (next trade date). The minimum price fluctuation is \$0.0001 (0.01¢) per gallon (\$4.20 per contract) and the maximum comes up to \$0.25 per gallon (\$10,500 per contract) for all months. If any contract is traded, bid, or offered at the limit for five minutes, trading is halted for five minutes. When trading resumes, the limit is expanded by \$0.25 per gallon in either direction. If another halt were triggered, the market would continue to be expanded by \$0.25 per gallon in either direction after each successive five-minute trading halt. The futures maturities correspond to 36 consecutive months which trading ceases at the close of business on the last business day of the month preceding the delivery month. The settlement type is the physical delivery of heating oil fulfilling the industry standards for fungible No. 2 heating oil and the delivery may only be initiated the day after the fifth business day and must be completed before the last business day of the delivery month.

Gasoline

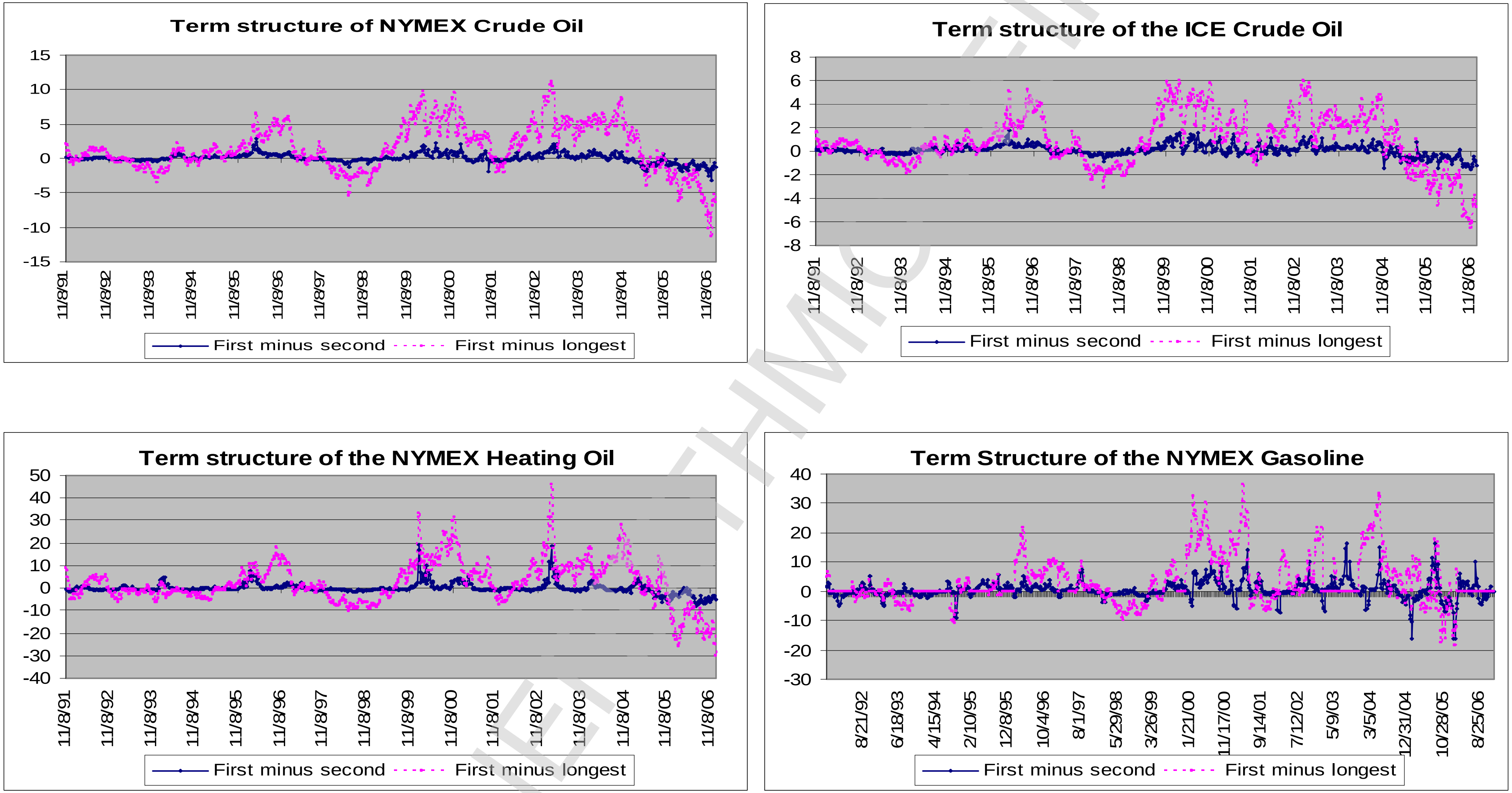
Gasoline is the largest single volume refined product sold in the United States and accounts for almost half of national oil consumption. It is a highly diverse market, with hundreds of wholesale distributors and thousands of retail outlets, making it subject to intense competition and price volatility. In NYMEX gasoline futures contract has a trading unit of 1,000 barrels or 42,000 U.S. gallons and it is quoted in U.S. dollars and cents. The open outcry trading takes place from 10:05 a.m. until 2:30 p.m. (New York time)and the Electronic trading is conducted from 6:00 PM until 5:15 p.m. via the CME Globex® trading platform, Sunday through Friday. There is a 45-minute break each day between 5:15 p.m. (current trade date) and 6:00 p.m. (next trade date). The price fluctuation follows exactly the same procedure with the one we mentioned above for heating oil. There are contracts covering maturities for twelve consecutive months and their trading terminates at the close of business on the last business day of the month preceding the delivery month. The settlement type is again physical delivery which may only be initiated the day after the fifth business day and must be completed before the last business day of the delivery month. The deliverable asset's grade and quality specifications have to conform to industry standards for Phase II Complex Model Reformulated Gasoline in accordance with Colonial Pipeline Co. specifications for fungible A grade, 87 octane index gasoline.

2.6 THE DATA SET

After the description of petroleum products and markets and the reference to its specific futures trading we move on with the definition of our dissertation data set. The dataset includes the weekly settlement futures prices of the aforementioned contracts. These elements have been obtained using Bloomberg data platform and the corresponding tickers which are CL, CO, HO and HU for the NYMEX Crude oil, the ICE Crude oil, the NYMEX Heating oil and Gasoline respectively. Bloombberg gives us the possibility to roll over contracts keeping the same time to maturity thus we are capable of creating generic series of contracts with fixed maturities at any point in time. For example CL1 denotes the NYMEX Crude oil future which is the nearest to maturity each time, CL2 the second nearest to maturity NYMEX Crude oil future and so on. For the purposes of our study we didn't choose the whole spectrum of the existing contracts so as to exclude some illiquid futures of the long term maturities. So we restrict our dataset to the first fifteen generics of the NYMEX Crude oil futures (CL1-CL15), the first nine of the ICE Crude oil futures (CO1-CO9), the first twelve of the NYMEX Heating oil (HO1-HO12) and eleven of NYMEX Gasoline (HU1-

HU11). Unfortunately, we have to say that we didn't succeed in dropping out all the illiquid contracts especially in the case of Gasoline which futures settlement prices include a lot of missing data. As we have already mentioned our data consist of weekly prices covering the period from 11/08/1991 to 12/29/2006. To visualise the term structure evolution of the aforementioned commodity futures we derive Figure 1. These Figures show the differences between the first shortest and the second shortest to maturity futures as well as between the first and the contract with the longest maturity that is included in our data set. Obviously, the latter one presents more extreme prices showing a more unstable behaviour, we can also note the changes from backwardation to contango and vice versa. This is a common observation among the two series and the "backwarded" prices are indicated by the positive prices of the differences whereas contango phenomenon appears with the presence of negative sign.

Figure 1: Evolution of all four commodities Term structure futures prices. The solid line denotes the difference of thefirst shortest minus the second shortest future contract and the dotted one the difference of the first minus the last expiring future contract. Sample period: 8/11/1991-29/12/2006



Before we move on to our analysis we had to check the series for stationarity having in mind that Principal Components Analysis presupposes that we have to do with stationary series. In order to come up with this necessity we applied an Augmented Dickey-Fuller (ADF) test on the generic series of our dataset. Our findings (see Table 1) ensure that all level series are non-stationary at the 1% significance level, as it is partially shown by the term structures evolution diagrams. To overcome this problem we create new time series using the first differences of the futures prices trying to keep in touch with the former bibliography (see Chantziara, Skiadopoulos 2005). Therefore, the new obtained Augmented Dickey-Fuller (ADF) tests indicate that we construct stationary series at the 1% significance level (Table 1). These time series, which are used from now and then in this paper, are denoted by the tickers DCL, DCO, DHO and DHU for the first differences of NYMEX and ICE Crude oil, NYMEX Heating oil and Gasoline futures prices, respectively. These series summary statistics are reported in Table 2 and through the application of Jarque-Bera test it is proved that we have to do with not normally distributed time series

Table 1: Unit Root Tests fot he weekly futures prices (level) and their first differences. They are reported by commodity (NYMEX Crude Oil, ICE Crude Oil, Heating Oil and Gasoline). Sample period: 8/11/1991-29/12/2006

Unit Root Tests

ADF test statistic
NYMEX Crude Oil generic contracts

	Level	First differences
CL1	-0.50982	-29.70871
CL2	-0.25964	-29.15258
CL3	-0.05102	-28.90070
CL4	0.13880	-28.74740
CL5	0.30830	-28.65841
CL6	0.46177	-28.65093
CL7	0.59832	-28.70490
CL8	0.73415	-28.73774
CL9	0.84330	-28.75565
CL10	0.96137	-28.69782
CL11	1.07461	-28.65739
CL12	1.16544	-28.65988
CL13	1.25568	-28.58120
CL14	1.33604	-28.52576
CL15	1.41392	-28.46684

ADF test statistic
NYMEX Heating Oil generic contracts

	Level	First differences
HO1	-0.44121	-24.17412
HO2	-0.17390	-23.92135
HO3	-0.01574	-23.52586
HO4	0.06538	-22.84939
HO5	-0.14613	-21.99653
HO6	-0.00244	-29.08501
HO7	0.16784	-29.23616
HO8	0.35299	-29.74054
HO9	0.54612	-30.24050
HO10	0.72037	-30.43903
HO11	0.88380	-29.87467
HO12	1.07008	-29.62780

ADF test statistic
ICE Crude Oil generic contracts

	Level	First differences
CO1	-0.21015	-28.37426
CO2	0.05817	-28.46260
CO3	0.26240	-28.49099
CO4	0.42306	-28.42729
CO5	0.56481	-28.46912
CO6	0.69786	-28.49164
CO7	0.81487	-28.58786
CO8	0.93778	-28.50681
CO9	1.05226	-28.44341

ADF test statistic
NYMEX Gasoline generic contracts

	Level	First differences
HU1	-1.45441	-15.62695
HU2	-0.87789	-29.19844
HU3	-0.98056	-28.95774
HU4	-0.46244	-29.16088
HU5	-0.46244	-29.16088
HU6	0.85058	-21.63205
HU7	1.77608	-12.34339
HU8	0.93479	-26.65156
HU9	1.38784	-26.69466
HU10	0.96270	-25.47753
HU11	0.62094	-23.57118

Table 2: Summary Statistics of the first differences of the futures prices. The elements include each expiry for each one of the four commodities (NYMEX Crude Oil, ICE Crude Oil, Heating Oil and Gasoline).

Summary statistics of the first differences of the futures prices

Sample: 11/08/1991 12/29/2006

NYMEX Crude Oil generic contracts															
	DCL1	DCL2	DCL3	DCL4	DCL5	DCL6	DCL7	DCL8	DCL9	DCL10	DCL11	DCL12	DCL13	DCL14	DCL15
Mean	0.048	0.050	0.051	0.053	0.054	0.055	0.055	0.056	0.057	0.057	0.058	0.058	0.058	0.059	0.059
Std. Dev.	1.444	1.339	1.252	1.178	1.118	1.067	1.027	0.990	0.963	0.934	0.907	0.888	0.869	0.852	0.837
Skewness	-0.316	-0.333	-0.311	-0.287	-0.243	-0.194	-0.156	-0.098	-0.081	-0.037	0.013	0.041	0.066	0.089	0.119
Kurtosis	6.579	6.241	6.277	6.535	6.782	7.022	7.237	7.452	7.573	7.736	7.951	8.071	8.286	8.487	8.687
Jarque-Bera	434.740	360.340	366.270	422.070	478.590	537.500	594.230	653.790	689.170	738.500	806.780	846.730	920.290	991.950	1066.500
Probability	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Observations	790	790	790	790	790	790	790	790	790	790	790	790	790	790	790

Sample: 11/08/1991 12/29/2006

ICE Crude Oil generic contracts									
	DCO1	DCO2	DCO3	DCO4	DCO5	DCO6	DCO7	DCO8	DCO9
Mean	0.049	0.051	0.053	0.054	0.055	0.056	0.056	0.057	0.057
Std. Dev.	1.307	1.205	1.136	1.083	1.040	1.003	0.974	0.945	0.919
Skewness	-0.223	-0.185	-0.175	-0.148	-0.109	-0.076	-0.048	-0.018	0.002
Kurtosis	6.383	6.301	6.425	6.669	6.910	7.125	7.266	7.484	7.731
Jarque-Bera	383.237	363.210	390.149	445.974	504.793	560.902	599.307	661.960	736.878
Probability	0	0	0	0	0	0	0	0	0
Observations	790	790	790	790	790	790	790	790	790

Table 2: Summary Statistics of the first differences of the futures prices (Cont'd)

Summary statistics of the first differences of the futures prices

Sample: 11/08/1991 12/29/2006

NYMEX Heating Oil generic contracts												
	DHO1	DHO2	DHO3	DHO4	DHO5	DHO6	DHO7	DHO8	DHO9	DHO10	DHO11	DHO12
Mean	0.114	0.120	0.127	0.134	0.140	0.145	0.149	0.152	0.155	0.157	0.160	0.163
Std. Dev.	4.653	4.222	3.957	3.737	3.532	3.348	3.158	2.996	2.863	2.775	2.716	2.645
Skewness	0.025	-0.056	-0.161	-0.206	-0.167	-0.186	-0.100	0.034	0.121	0.082	0.057	0.132
Kurtosis	7.516	8.227	8.875	9.202	8.530	8.451	8.507	8.653	8.615	8.403	8.364	8.486
Jarque-Bera	671.538	899.772	1139.719	1271.606	1010.218	982.790	999.503	1052.177	1039.616	961.918	947.405	992.943
Probability	0	0	0	0	0	0	0	0	0	0	0	0
Observations	790	790	790	790	790	790	790	790	790	790	790	790

Sample: 11/08/1991 12/29/2006

NYMEX Gasoline generic contracts											
	DHU1	DHU2	DHU3	DHU4	DHU5	DHU6	DHU7	DHU8	DHU9	DHU10	DHU11
Mean	0.113	0.132	0.107	0.126	0.126	0.165	0.175	0.152	0.172	0.165	0.166
Std. Dev.	5.417	4.508	4.080	3.614	3.614	3.175	3.112	3.082	2.964	3.049	2.979
Skewness	-0.259	-0.385	-0.362	-0.389	-0.389	-0.400	-0.269	-0.402	-0.551	-0.440	-0.674
Kurtosis	9.041	8.172	7.621	7.939	7.939	8.377	10.290	10.247	10.320	11.692	10.417
Jarque-Bera	1210.214	895.298	712.890	809.389	809.389	942.996	1674.121	1617.078	1582.136	1974.851	1325.985
Probability	0	0	0	0	0	0	0	0	0	0	0
Observations	790	786	782	777	777	766	752	730	693	621	560

CHAPTER 3: PRINCIPAL COMPONENTS ANALYSIS

In this Chapter we present the mathematical theory for the Principal Components Analysis, and we analyze the retained PCs interpretation compared with the previous literature findings. The last part of the Chapter consists of the regression definition and the forecasting power checking of our model.

3.1 PCA DESCRIPTION

Principal Components Analysis (PCA) is a technique for simplifying a dataset, by reducing multidimensional datasets to lower dimensions for analysis. Technically speaking, PCA is an orthogonal linear transformation that transforms the data to a new coordinate system such that the greatest variance by any projection of the data comes to lie on the first coordinate (called the first principal component), the second greatest variance on the second coordinate, and so on. The fact that PCA proposes an orthogonal linear transformation means that these factors are perpendicular to each other. PCA can be used for dimensionality reduction in a dataset of correlated variables because it maintains its variance-covariance structure and gives us the opportunity to retain those characteristics of the dataset that contribute most to its variance, by keeping lower-order principal components and ignoring higher-order ones. Such low-order components often contain the "most important" aspects of the data. But this is not necessarily the case, depending on the application. Moreover, PCA allows us to compress our data making possible the visualization of them, assuming a small PCs number, and the identification of some specific general patterns among our variables. In this dissertation, we follow Chantziara, Skiadopoulos (2005, 2007) consideration applying PCA on the first differences of petroleum futures weekly prices.

For a more formal view of the method we refer briefly to the mathematical background. Considering our data set we can assume p variables which are observed across time steps t by t=1,…,T. So, each one variable constitutes a (T×1) vector. The purpose of the PCA is to transform this data set to a new one consisting of p artificial variables (Principal Components - PCs hereafter) uncorrelated to each other reproducing the original variance-covariance structure. So, we have the folowing equation:

$$Z = XA \quad \textbf{(1)}$$

where **Z** is the (T×p) matrix of principal components, **X** is the (T×p) matrix of the original variables, and **A** is a (p×p) matrix of coefficients called loadings. The first order condition of this maximization problem described as

$$(X'X - l_i I)A = 0 \quad \textbf{(2)}$$

l_i are the Lagrange multipliers and I is a (p×p) identity matrix. $X'X$ represents the original variance covariance matrix and make obvious that the for the equation solution we have just to find this matrix eigenvalues l_i and eigenvectors A. Instead of using the variance-covariance matrix we often use the correlation matrix in order to standardize our variables avoiding to include the measurement units of them in our analysis. As far as the retained results are concerned we have to note that we obtain the variance of the ith PC from the ith eigenvalue and the total variance. Moreover, the total variance of the X variables equals the sum of the PCs variances.

Until now we have reproduced the original variance cvariance matrix and we have retained uncorrelated artificial variables from our data, but we have not yet reduce our data dimensionality. To do that we can simply exclude some of the retained principal components keeping the ones that explain the greater part of the total variance. Doing that we maintain r PCs which are less than the estimated p from the p number of the original variables (r<p). Then we come up with the following equation

$$X = Z_{(r)}A_r^{'} + e_{(r)} \quad \textbf{(3)}$$

All matrices are defined as before, but we have reduce the number of columns from p to r, $e_{(r)}$ is a (T×p) matrix of residuals because now we have excluded some PCs responsible for some information. The remaining PCs explain a specific percentage of the total variance. This percentage is called communality and it is calculated from the loadings. The loadings show the way that a change in the corresponding PC affects each variable.

As far as the determination of the retained PCs number is concerned we could refer to some formal or informal criteria. There are statistical base tests we can find in Jackson (1991) or Basilevsky (1994) which make assumptions about the original variables distribution and there are also rules which are simple in use but they don't enjoy theoretical support. The fact that even the statistical robust methods work under assumptions and can give misleading results we decide to choose the PCs number using interpretation evidence from previous literature and considering the percentage of the total variance each additional factor explain.

3.2 PCA RESULTS AND DISCUSSION

Separate PCA

After the extraction of stationary series from our data, we have already described, we perform PCA on the first differences of the weekly settlement prices separately for every one of the four commodities. So our data are the stationary time series DCL1-DCL15, DCO1-DCO9, DHO1-DHO12 and DHO1-DHO11 for the time period 11/15/1991 to 12/29/2006 on which we applied PCA using the correlation matrix analysis.

Table 3 shows the descriptive statistics of the first three retained PCs for each commodity. As it was the case with the differences of the futures settlement prices we observe from the Jarque-Bera test results that their PCs are also non-normally distributed.

Table 3: Summary statistics of the first three standardized principal components obtained from the separate PCA. The results are reported by commodity (NYMEX Crude Oil, ICE Crude Oil, Heating Oil and Gasoline)

Panel A: Separate PCA - Standardised PCs

Sample: 11/08/1991 12/29/2006

NYMEX Crude Oil	CLPC1	CLPC2	CLPC3
Mean	0.000	0.000	0.000
Std. Dev.	1	1	1
Skewness	0.443	-0.527	-0.270
Kurtosis	11.024	6.712	7.584
Jarque-Bera	2145.106	490.217	701.346
Probability	0	0	0
Observations	790	790	790

NYMEX Heating Oil	HOPC1	HOPC2	HOPC3
Mean	0.000	0.000	0.000
Std. Dev.	1	1	1
Skewness	0.323	0.292	-0.260
Kurtosis	7.823	11.047	11.659
Jarque-Bera	779.342	2142.643	2476.942
Probability	0	0	0
Observations	790	790	790

ICE Crude Oil	COPC1	COPC2	COPC3
Mean	0.000	0.000	0.000
Std. Dev.	1	1	1
Skewness	0.133	-0.492	0.195
Kurtosis	8.752	7.092	5.658
Jarque-Bera	1091.364	583.009	237.575
Probability	0	0	0
Observations	790	790	790

NYMEX Gasoline	HUPC1	HUPC2	HUPC3
Mean	0.000	0.000	0.000
Std. Dev.	1	1	1
Skewness	-0.415	-0.525	-0.336
Kurtosis	11.327	8.581	5.747
Jarque-Bera	1709.801	787.510	195.271
Probability	0	0	0
Observations	586	586	586

Table 4 shows the cumulative variance that is explained by the first four Principal Components. For the first three commodities we observe that there are not important differences between the weekly data analysis and former bibliography (see Tomalsky and Hindanov 2002 and Chantziara, Skiadopoulos 2005). The first three PCs explain more than 98.5% of the total variance while the fourth PC adds less than 1% offering only a marginal increase to the cumulative percentile. The slightly different result comes from the NYMEX Gasoline PCs. We find out that the first three components are now "responsible" for the 96.2% of the total variance. This disagreement has its justification when we look at the pairwise correlations between the different gasoline maturities. These are smaller compared to the other commodities. As a proof of this statement the correlations matrices are reported in Appendix . Although we had this evidence it is not a sign of a possible inadequacy of the first three PCs, so, we continue our analysis using these components trying to be in accordance with the previous literature. Furthermore, we should refer to the fact that even in this extreme case the variance explained by the first three PCs sums up to a percentile greater than 96% of the total.

Table 4: Total percentage of variance explained by the first four principal components obtained from the separate PCA. Results are reported by commodity (NYMEX Crude Oil, ICE Crude Oil, Heating Oil and Gasoline). Sample period: 8/11/1991-29/12/2006.

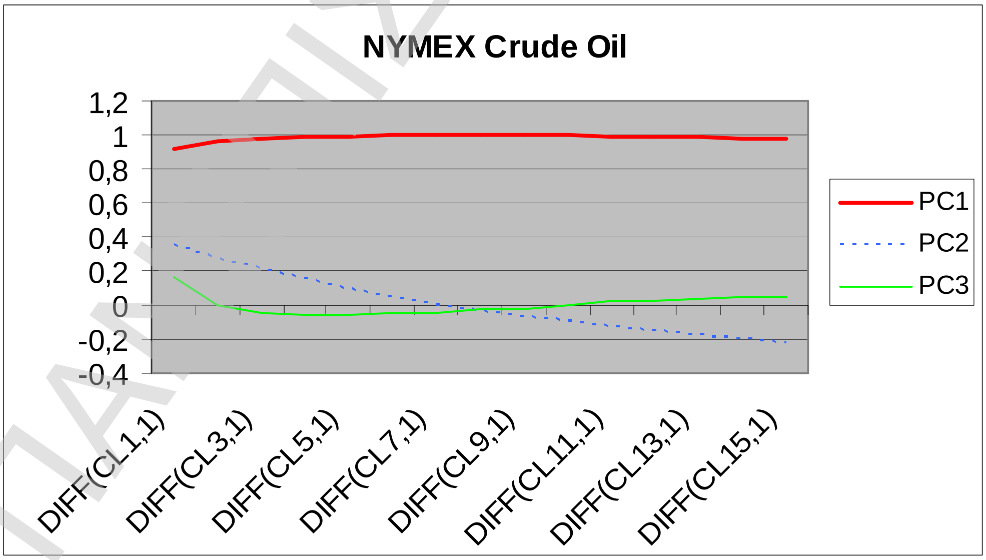
Principal components and explained Variance

Panel A: Separate PCA

Pincipal component	NYMEX Crude Oil	IPE Crude Oil	NYMEX Heating Oil	NYMEX Gasoline
1	96.481	97.626	92.232	88.195
2	99.523	99.597	96.637	93.721
3	99.856	99.853	98.671	96.173
4	99.939	99.913	99.482	97.361

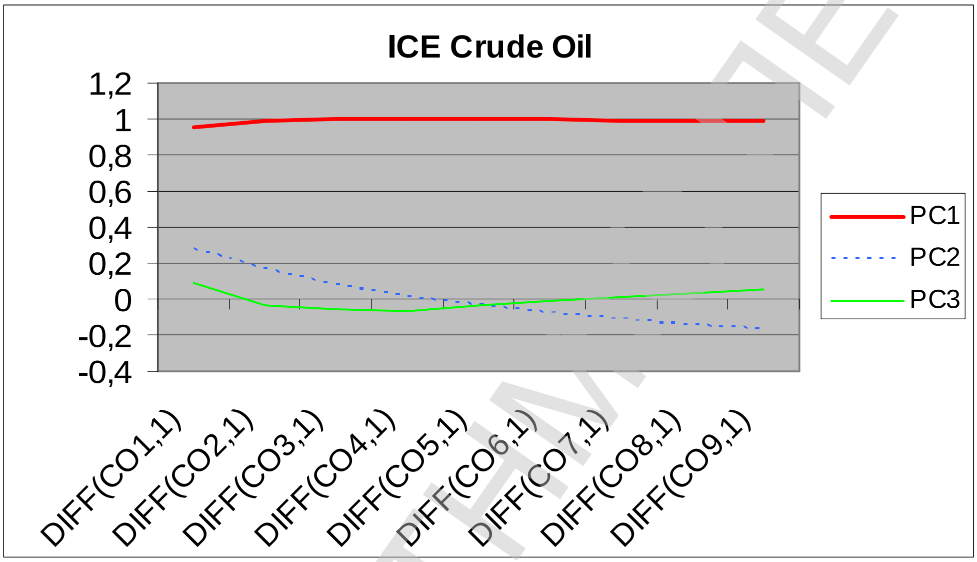
Then we analyze the behaviour of the correlation loadings for the first three PCs. Starting with NYMEX Crude Oil in Figure 2 we can see that the first component affects the term structure of this contract by the same amount and to the same direction. We can say that it offers a parallel shift to NYMEX Crude Oil Futures first differences. The second PC can be mentioned as a slope factor. It corresponds to different reactions between short and long term futures. When the second component changes it influences the contracts with the seven shortest maturities in the same way and the next eight futures in the exactly opposite way. This influence is greater in absolute magnitude for the very short and very long term futures. The third component corresponds to a curvature effect. It moves short and long term maturities to one direction and mid term to another. As far as this commodity is concerned the shortest maturity future is in line with the latest four maturities futures from the fifteen of our sample. The rest contracts move to the opposite direction for a given change of the third PC. The fourth component shows a noisy behaviour which doesn't allow us to shape stable conclusions. So, as we have already mentioned, it is excluded from our analysis for all four commodities.

Figure 2: NYMEX Crude Oil correlation loadings of the first three separate Principal Components. Sample period:8/11/1991-29/12/2006



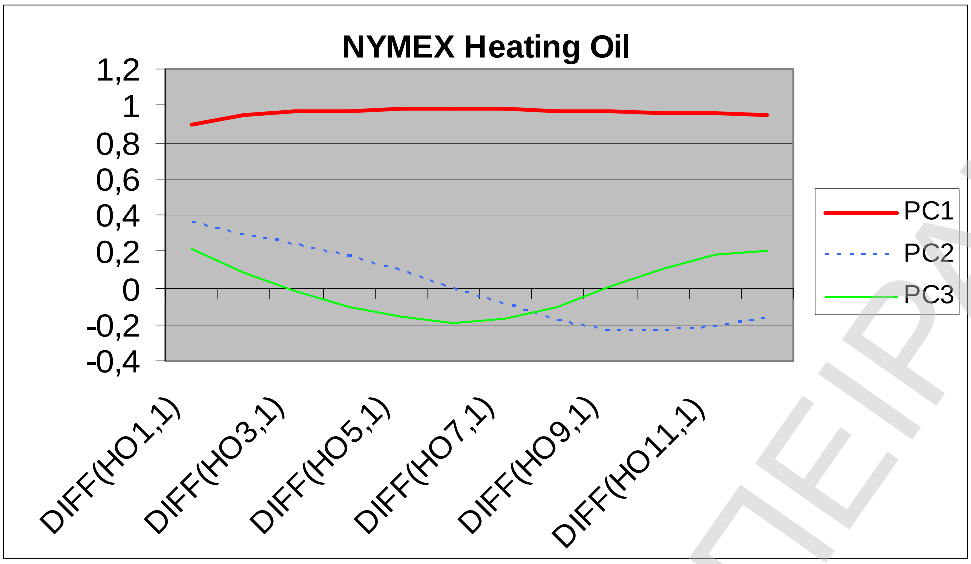
Moving to ICE Crude Oil we present Figure 3 which plots its PCs correlation loadings. This picture has obvious similarities with the previous one. The first three factors represent a parallel shift, a slope and a curvature effect respectively. A common characteristic has to do with the steeper third PC for the short expiries compared to the long ones.

Figure 3: ICE Crude Oil correlation loadings of the first three separate Principal Components. Sample period:8/11/1991-29/12/2006



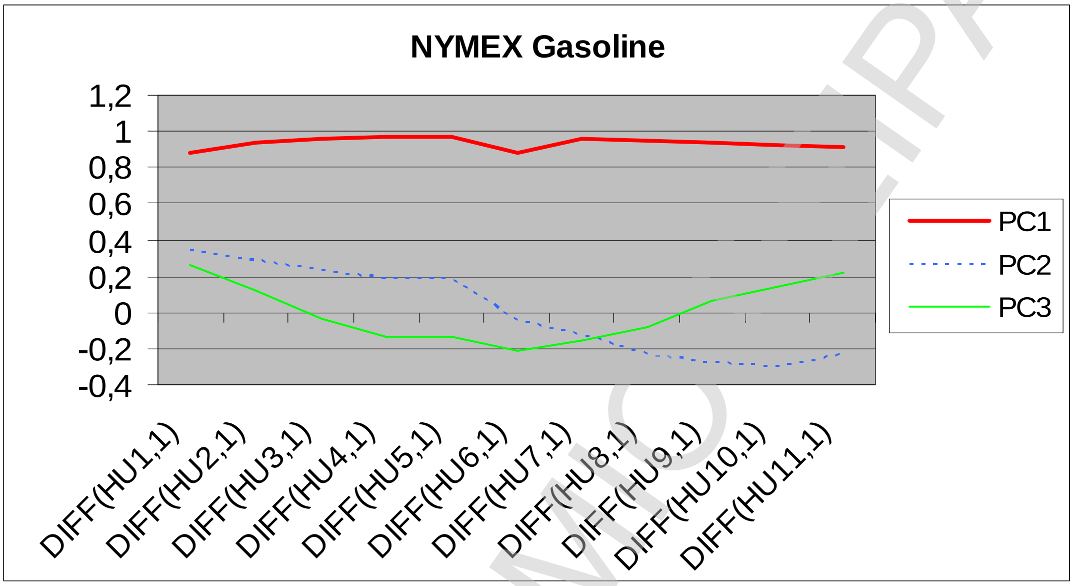
The case of NYMEX Heating Oil is very near to the aforementioned commodities. Of course we should note that the first three PC's explain slightly less variance than they did for the Crude Oils but again it is more than 98.5% of the total (see Table 4). Figure 4 shows the same performance for the PCs. Again we have the first PC as a level factor, the second affects differently the four shortest maturity futures than the rest eight which expire five to twelve months from the given day and the third can be interpreted as a curvature factor which has more symmetrical effects because it is not so steep as we observed in Crudes case.

Figure 4: NYMEX Heating Oil correlation loadings of the first three separate Principal Components. Sample period:8/11/1991-29/12/2006



So far our results were generally in agreement with the former studies. Cortazar and Schwarz 1994, Clewlow and Strickland 1999b, Tomalsky and Hindanov 2002, Chantziara and Skiadopoulos 2005 found similar results as far as the dynamics of the PCs and the variance they explain are concerned. The fact that our dataset includes weekly prices instead of daily which was the object in Chantziara, Skiadopoulos paper didn't seem to influence our PCA's performance. So it would make sense to maintain the same evidence from the NYMEX Gasoline. Figure 5 shows a picture which can be regarded as similar with the other plots. Gasoline's first three PCs present a little more complex behaviour when we first look at their correlation loadings. A more comprehensive study of the findings ensures that these PCs have some elements in common with the other commodities' components. The first one of them can be interpreted as a parallel shift for all maturities and still drives the futures to the same direction with slightly different volumes throughout the maturities. To make it clear we can say that a change in the first PC moves all the contracts to the same direction but it affects to a greater change for mid term maturities and a smaller for short term maturities futures. The second factor is positive for the first five months expiries and negative for the next seven keeping the characteristic of being greater in absolute magnitude for very short or very long term maturities. It only lose its slope characteristic for the forth and fifth shortest maturity contracts, which have exactly the same correlation loading. Finally the third factor represents again a curvature effect for the third PC which is omitted for the same aforementioned contracts.

Figure 5: NYMEX Gasoline correlation loadings of the first three separate Principal Components. Sample period:8/11/1991-29/12/2006



Joint PCA

Our next step to the PCA performance on our data is the simultaneous application of the method on all commodities and across the different maturities. Our target is to obtain PCs which are able to describe the dynamics of all four commodities and the joint evolution of their weekly prices term structure (see Tomalsky and Hindanov 2002, Chantziara and Skiadopoulos 2005).

Table 5 shows the descriptive statistics for these PCs, whereas Table 6 informs us about the variance explained by the first four PCs. Easily, we can detect from the Jarque-Bera test that these components have a non-normal distribution. Furthermore, we have exclude our data listwise when we have to do with missing values in any commodity so as to avoid non synchronous effect in the derivation of our analysis.

Table 5: Summary statistics of the first three standardized principal components obtained from the joint PCA.

Panel B: Joint PCA - Standardised PCs

Sample: 11/08/1991 12/29/2006

	PC1	PC2	PC3
Mean	0.000	0.000	0.000
Std. Dev.	1	1	1
Skewness	-0.103	-0.505	-0.384
Kurtosis	8.234	12.741	8.063
Jarque-Bera	669.926	2341.766	640.349
Probability	0	0	0
Observations	586	586	586

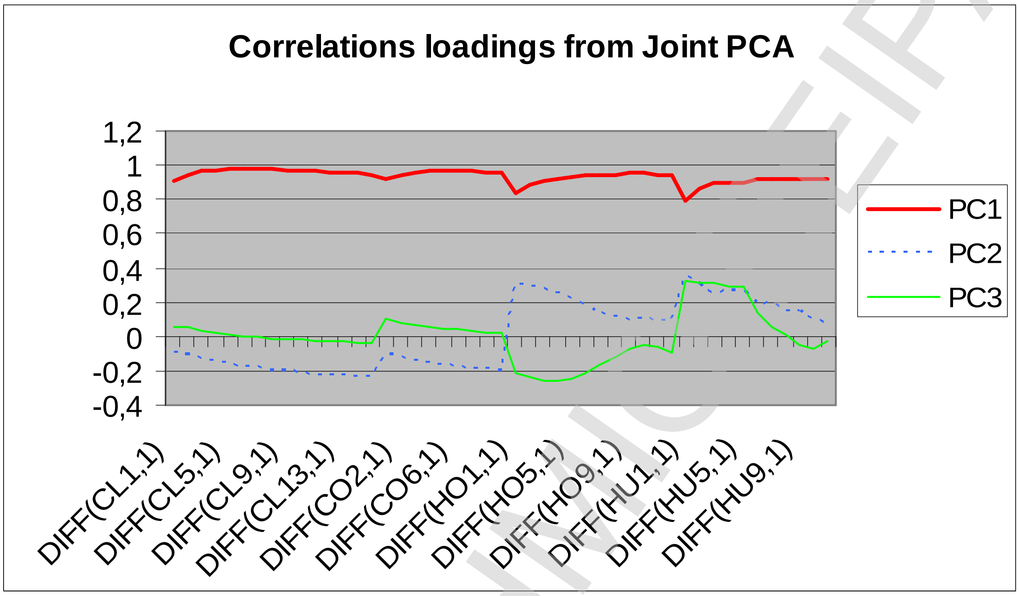
Table 6: Total percentage of variance explained by the first four principal components obtained from the joint PCA. Sample period: 1/1/1993-31/12/2003.

Principal components and explained Variance	
Panel B: Joint PCA	
Pincipal component	
1	87.613
2	91.450
3	93.462
4	95.039

Table 6 shows that the cumulative variance explained by the first three or four PCs is smaller compared to the separate PCs. We should note that this amount is even smaller than the percentile explained by the Gasoline's PCs, but assuming the first three of them we still can explain as much as 95% of the total variance, which is a critical value often used as a rule of thump for the derivation of the appropriate PCs number.

Figure 6 plots the correlation loadings of the first three components retained from the joint Analysis. We can see that the first PC continues to present similar behaviour, at least for each one of the four commodities separately (NYMEX Crude, Heating Oil and Gasoline, and ICE Crude). In this case it can be interpreted as a parallel shift. In the shortest maturities of each one commodity we can observe lower values compared to the longer ones. The second PC keeps its slope interpretation and these effects are more intense in Heating Oil and Gasoline case. The only differentiation has to do with Gasoline where the slope effect shows more complex behaviour across the maturities. We should note the second PC moves Heating Oil and Gasoline towards the same direction and the rest two commodities (Crude Oils) to the opposite according to the quantities determined by the separate slope effects, as they are interpreted by the loadings. The third PC has the same slope interpretation as we mentioned for the second, as far as Crude Oil futures are concerned. Changes of this component provoke similar changes to the Crude Oils, whereas they affect in amore unstable way the two other commodities. Heating Oil contracts show a more complex reaction depending on their expiry months. This is comparable to the curvature effects of the separately obtained third PCs, with the differentiation of the loading getting again smaller for the last examined maturities. Gasoline PCs seem to be similar with the two Crudes presenting slope characteristics, but the last maturities give a slight curvature effect, because the correlation loading of the eleven month expiring contract is greater than the previous one.

Figure 6: Correlation loadings of the first three joint Principal Components. Sample period:8/11/1991-29/12/2006



3.3 PCA AND FORECASTING POWER

In this chapter we test the predictive power of two multivariate regression models which are based on our PCA results. These models try to identify the forecasting skills of simple regression models, when we are occupy as our independent variables the retained PCs we have already analyse. Obviously, we have to do with a time series forecasting approach because we try to predict our dependent variable using its own dynamics which we have summarize in a restricted amount of PCs. We move on with the derivation of our models. Firstly, the weekly futures prices changes are regressed on the twelve PCs (three for each commodity) retained by the Separate PCA. The second part of the chapter includes the regression of the contracts changes on the three PCs retained from the joint PCA.

Separate PCA: Regression and Results

As we have mentioned this procedure is substantially similar to a time series forecasting approach. The main advantage of our method is the possibility to handle a significantly smaller amount of causal variables. Now we have as dependent variables the futures differences for all four commodities across 47 different maturities in contract (15 for NYMEX Crude, 9 for ICE Crude, 12 for NYMEX Heating Oil and 11 for NYMEX Gasoline). If we would like to detect a simple regression model we should calculate 47 regressions with 47 explanatory variables for each one of them. Instead of this, we use the twelve retained PCs (CLPC1-3, COPC1-3, HOPC1-3, HUPC1-3) as causal parameters because they explain the evolution of the term structures for each one commodity. Moreover, our analysis offers the advantage of checking for spillover effects across commodities allowing a specific contract to have as explanatory variable PCs that are obtained from the other commodities. Of course this would not be applicable if we had chosen a univariate autoregression model, for example, in order to reduce our parameters.

To be more specific we introduce the regressions we estimate and the corresponding parameters. These are as follows:

$$DF_t^j = c + \sum_{k=1}^{3} a_k CLPC_{k,t-1} + \sum_{k=1}^{3} b_k COPC_{k,t-1} + \sum_{k=1}^{3} c_k HOPC_{k,t-1} + \sum_{k=1}^{3} d_k HUPC_{k,t-1} + u_t$$

(4)

DF_t^j denotes the weekly changes of futures prices at time t .

Indicator j represents the different generics (CL1-CL15, CO1-CO9, HO1-HO12, HU1-HU11) of the contracts on the four commodities, NYMEX Crude Oil, IPE Crude Oil, NYMEX Heating Oil and NYMEX Gasoline respectively

$CLPC_{k,t-1}$, $COPC_{k,t-1}$, $HOPC_{k,t-1}$ and $HUPC_{k,t-1}$ are the time series of the three per commodity principal components we have extracted from the PCA at time t-1

a_k, b_k and c_k are the retained from the regression coefficients

We should also note that we detect heteroskedasticity in the regression residuals, so we used Newey-West standard errors, which correct both for heteroskedasticity and autocorrelation, to obtain our coefficients t-statistics.

Table 7 shows the results for regressions per commodity across all maturities. The first column includes the dependent variables per commodity for all the corresponding maturities. The next thirteen columns present the constant term and the PCs coefficients with their t-statistics underneath them. In addition to this we have put into frames the coefficients which t-statistics indicate that they are statistically significant at the 5% significance level. The following two columns show the R^2 statistic and the Durbin –Watson statistic for first order serial correlation. The last one presents the probability estimated by the F-statistic. This element allows us to reject the null hypothesis that all coefficients have zero values for the confidence level we want to examine. In our analysis we refer to a confidence level of 5%.

Table 7: Forecasting Power of the Separate PCs

NYMEX Crude Oil

j		C	a1	a2	a3	b1	b2	b3	c1	c2	c3	d1	d2	d3	R-squared	Durbin-Watson stat	Prob(F-stat)
DCL1	Coefficient	0.072	-0.221	-0.368	0.024	0.169	0.439	-0.078	-0.322	-0.242	-0.255	0.392	0.145	0.166	0.034	2.031	0.002
	t-Statistic	1.305	-0.786	-1.540	0.289	0.754	2.256	-1.120	-1.501	-1.564	-1.609	2.188	1.272	1.014			
DCL2	Coefficient	0.075	-0.260	-0.351	-0.042	0.233	0.430	-0.036	-0.278	-0.229	-0.255	0.368	0.108	0.127	0.024	2.041	0.012
	t-Statistic	1.439	-0.967	-1.575	-0.530	1.071	2.340	-0.554	-1.390	-1.616	-1.702	2.184	1.018	0.807			
DCL3	Coefficient	0.076	-0.252	-0.326	-0.051	0.247	0.412	-0.020	-0.257	-0.230	-0.256	0.358	0.103	0.105	0.023	2.035	0.016
	t-Statistic	1.540	-0.974	-1.590	-0.690	1.170	2.376	-0.322	-1.360	-1.704	-1.804	2.271	1.011	0.705			
DCL4	Coefficient	0.076	-0.265	-0.315	-0.052	0.269	0.398	-0.008	-0.235	-0.222	-0.256	0.353	0.101	0.087	0.023	2.032	0.015
	t-Statistic	1.631	-1.068	-1.625	-0.748	1.326	2.415	-0.146	-1.300	-1.712	-1.878	2.372	1.022	0.613			
DCL5	Coefficient	0.077	-0.279	-0.308	-0.053	0.283	0.383	-0.005	-0.210	-0.211	-0.251	0.347	0.096	0.072	0.024	2.031	0.013
	t-Statistic	1.713	-1.168	-1.671	-0.792	1.447	2.428	-0.088	-1.203	-1.675	-1.896	2.460	0.987	0.526			
DCL6	Coefficient	0.077	-0.287	-0.299	-0.052	0.285	0.364	-0.005	-0.187	-0.201	-0.245	0.343	0.092	0.061	0.025	2.031	0.010
	t-Statistic	1.789	-1.245	-1.692	-0.819	1.513	2.398	-0.100	-1.096	-1.632	-1.886	2.550	0.961	0.454			
DCL7	Coefficient	0.077	-0.296	-0.288	-0.053	0.287	0.345	-0.006	-0.162	-0.188	-0.241	0.338	0.084	0.048	0.026	2.035	0.009
	t-Statistic	1.854	-1.327	-1.682	-0.869	1.567	2.354	-0.127	-0.973	-1.549	-1.883	2.612	0.901	0.370			
DCL8	Coefficient	0.077	-0.290	-0.275	-0.047	0.264	0.321	-0.005	-0.141	-0.182	-0.234	0.341	0.085	0.039	0.027	2.036	0.007
	t-Statistic	1.927	-1.340	-1.647	-0.805	1.488	2.261	-0.102	-0.859	-1.543	-1.869	2.733	0.920	0.309			
DCL9	Coefficient	0.077	-0.299	-0.281	-0.059	0.277	0.325	-0.005	-0.127	-0.170	-0.226	0.322	0.075	0.034	0.028	2.032	0.006
	t-Statistic	1.976	-1.428	-1.747	-1.030	1.621	2.354	-0.097	-0.789	-1.449	-1.829	2.656	0.819	0.269			
DCL10	Coefficient	0.078	-0.284	-0.261	-0.056	0.259	0.303	-0.005	-0.115	-0.164	-0.223	0.317	0.072	0.028	0.028	2.032	0.006
	t-Statistic	2.040	-1.400	-1.653	-1.013	1.564	2.250	-0.120	-0.724	-1.433	-1.848	2.720	0.804	0.229			
DCL11	Coefficient	0.078	-0.274	-0.250	-0.051	0.242	0.290	-0.006	-0.104	-0.163	-0.217	0.315	0.071	0.023	0.029	2.028	0.005
	t-Statistic	2.096	-1.396	-1.621	-0.940	1.505	2.206	-0.143	-0.663	-1.449	-1.834	2.789	0.807	0.192			
DCL12	Coefficient	0.077	-0.265	-0.244	-0.062	0.239	0.284	-0.002	-0.097	-0.157	-0.215	0.304	0.065	0.018	0.030	2.028	0.004
	t-Statistic	2.128	-1.382	-1.607	-1.159	1.512	2.201	-0.052	-0.630	-1.404	-1.850	2.765	0.751	0.154			
DCL13	Coefficient	0.077	-0.250	-0.231	-0.064	0.228	0.274	-0.001	-0.093	-0.154	-0.214	0.297	0.063	0.014	0.031	2.025	0.004
	t-Statistic	2.169	-1.336	-1.544	-1.227	1.477	2.159	-0.017	-0.612	-1.399	-1.864	2.761	0.730	0.120			
DCL14	Coefficient	0.077	-0.240	-0.227	-0.067	0.222	0.270	0.002	-0.089	-0.152	-0.212	0.290	0.061	0.010	0.031	2.018	0.003
	t-Statistic	2.199	-1.316	-1.545	-1.314	1.464	2.177	0.053	-0.594	-1.397	-1.863	2.762	0.721	0.089			
DCL15	Coefficient	0.077	-0.233	-0.220	-0.069	0.219	0.264	0.004	-0.088	-0.148	-0.210	0.284	0.061	0.007	0.032	2.014	0.003
	t-Statistic	2.235	-1.304	-1.517	-1.385	1.471	2.164	0.084	-0.593	-1.384	-1.868	2.766	0.719	0.063			

Table 7: Forecasting Power of the Separate PCs(Cont'd)

Panel B: ICE Crude Oil

j		C	a1	a2	a3	b1	b2	b3	c1	c2	c3	d1	d2	d3	R-squared	Durbin-Watson stat	Prob(F-stat)
DCO1	Coefficient	0.080	-0.024	0.052	-0.027	-0.070	-0.041	-0.025	-0.213	-0.170	-0.176	0.375	0.088	0.117	0.025	1.994	0.301
	t-Statistic	1.473	-0.103	0.260	-0.344	-0.311	-0.208	-0.401	-1.206	-1.172	-1.160	2.355	0.804	0.679			
DCO2	Coefficient	0.079	-0.045	0.016	-0.019	-0.029	0.031	-0.051	-0.200	-0.201	-0.204	0.375	0.088	0.102	0.030	2.002	0.149
	t-Statistic	1.574	-0.196	0.081	-0.264	-0.137	0.170	-0.867	-1.226	-1.507	-1.465	2.622	0.870	0.652			
DCO3	Coefficient	0.078	-0.028	0.005	-0.030	-0.044	0.061	-0.054	-0.160	-0.198	-0.217	0.369	0.078	0.076	0.037	2.003	0.049
	t-Statistic	1.656	-0.121	0.024	-0.453	-0.224	0.362	-0.983	-1.010	-1.563	-1.628	2.693	0.798	0.511			
DCO4	Coefficient	0.077	0.008	0.015	-0.033	-0.066	0.073	-0.050	-0.141	-0.189	-0.220	0.352	0.067	0.056	0.043	2.011	0.018
	t-Statistic	1.736	0.034	0.085	-0.523	-0.352	0.458	-0.958	-0.914	-1.550	-1.689	2.674	0.691	0.392			
DCO5	Coefficient	0.077	0.008	-0.003	-0.036	-0.073	0.079	-0.040	-0.124	-0.186	-0.228	0.360	0.068	0.048	0.048	2.006	0.008
	t-Statistic	1.796	0.039	-0.015	-0.592	-0.405	0.516	-0.797	-0.821	-1.583	-1.807	2.842	0.723	0.349			
DCO6	Coefficient	0.076	0.010	-0.014	-0.039	-0.080	0.081	-0.030	-0.108	-0.180	-0.226	0.356	0.065	0.043	0.050	2.008	0.005
	t-Statistic	1.837	0.050	-0.081	-0.647	-0.453	0.548	-0.617	-0.719	-1.562	-1.824	2.887	0.703	0.326			
DCO7	Coefficient	0.076	0.025	-0.022	-0.044	-0.093	0.084	-0.019	-0.096	-0.169	-0.220	0.345	0.059	0.033	0.051	2.016	0.004
	t-Statistic	1.884	0.120	-0.129	-0.746	-0.541	0.583	-0.398	-0.638	-1.483	-1.805	2.831	0.644	0.251			
DCO8	Coefficient	0.076	0.001	-0.037	-0.043	-0.083	0.084	-0.017	-0.081	-0.164	-0.223	0.354	0.065	0.031	0.054	2.012	0.002
	t-Statistic	1.953	0.006	-0.224	-0.757	-0.500	0.596	-0.369	-0.547	-1.470	-1.863	3.038	0.718	0.247			
DCO9	Coefficient	0.075	0.006	-0.041	-0.046	-0.088	0.083	-0.011	-0.072	-0.156	-0.219	0.350	0.061	0.025	0.056	2.010	0.001
	t-Statistic	1.988	0.033	-0.257	-0.829	-0.542	0.608	-0.252	-0.488	-1.416	-1.867	3.085	0.695	0.199			

Table 7: Forecasting Power of the Separate PCs(Cont'd)

Panel C: NYMEX Heating Oil

j		C	a1	a2	a3	b1	b2	b3	c1	c2	c3	d1	d2	d3	R-squared	Durbin-Watson stat	Prob(F-stat)
DHO1	**Coefficient**	0.167	1.188	1.473	-0.014	-0.334	-0.721	-0.172	-1.281	-0.940	-1.009	0.646	0.100	0.162	0.035	2.050	0.069
	t-Statistic	1.000	1.261	2.074	-0.056	-0.407	-1.100	-0.718	-1.634	-1.830	-2.022	0.938	0.238	0.282			
DHO2	**Coefficient**	0.188	0.833	1.122	0.032	-0.119	-0.497	-0.206	-1.147	-0.841	-1.083	0.750	0.180	0.091	0.033	2.048	0.093
	t-Statistic	1.182	0.894	1.581	0.134	-0.149	-0.796	-0.899	-1.511	-1.734	-2.205	1.126	0.420	0.165			
DHO3	**Coefficient**	0.187	0.574	0.768	0.002	0.107	-0.238	-0.188	-1.039	-0.658	-1.051	0.725	0.192	0.056	0.030	2.040	0.155
	t-Statistic	1.213	0.627	1.107	0.008	0.142	-0.411	-0.858	-1.424	-1.396	-2.204	1.142	0.468	0.109			
DHO4	**Coefficient**	0.191	0.324	0.514	0.008	0.318	-0.004	-0.174	-0.993	-0.623	-1.052	0.791	0.216	0.064	0.031	2.032	0.129
	t-Statistic	1.260	0.373	0.777	0.034	0.450	-0.008	-0.823	-1.438	-1.372	-2.239	1.332	0.545	0.130			
DHO5	**Coefficient**	0.198	0.278	0.445	0.003	0.327	0.075	-0.152	-0.986	-0.668	-1.052	0.843	0.249	0.104	0.034	2.036	0.083
	t-Statistic	1.356	0.342	0.704	0.013	0.504	0.148	-0.750	-1.526	-1.516	-2.274	1.516	0.645	0.217			
DHO6	**Coefficient**	0.201	0.246	0.415	0.012	0.351	0.173	-0.151	-0.933	-0.749	-1.074	0.852	0.204	0.068	0.041	2.043	0.025
	t-Statistic	1.458	0.323	0.693	0.056	0.600	0.361	-0.809	-1.546	-1.745	-2.347	1.617	0.548	0.148			
DHO7	**Coefficient**	0.210	0.203	0.330	0.002	0.341	0.233	-0.141	-0.847	-0.740	-1.000	0.816	0.161	0.048	0.044	2.045	0.015
	t-Statistic	1.634	0.287	0.586	0.008	0.639	0.517	-0.830	-1.499	-1.787	-2.309	1.656	0.447	0.110			
DHO8	**Coefficient**	0.223	0.097	0.209	-0.026	0.359	0.293	-0.117	-0.762	-0.705	-0.918	0.798	0.141	0.048	0.044	2.062	0.014
	t-Statistic	1.848	0.147	0.392	-0.148	0.721	0.690	-0.761	-1.432	-1.731	-2.264	1.704	0.407	0.116			
DHO9	**Coefficient**	0.235	0.050	0.095	-0.053	0.359	0.307	-0.094	-0.727	-0.661	-0.818	0.770	0.132	0.090	0.043	2.064	0.018
	t-Statistic	2.049	0.079	0.185	-0.322	0.754	0.761	-0.664	-1.416	-1.645	-2.166	1.732	0.394	0.231			
DHO10	**Coefficient**	0.246	0.020	0.074	-0.072	0.413	0.312	-0.089	-0.722	-0.641	-0.775	0.729	0.107	0.088	0.045	2.060	0.013
	t-Statistic	2.229	0.035	0.153	-0.478	0.884	0.806	-0.658	-1.438	-1.613	-2.110	1.706	0.334	0.233			
DHO11	**Coefficient**	0.248	0.081	0.145	-0.105	0.418	0.270	-0.098	-0.723	-0.618	-0.730	0.660	0.073	0.073	0.046	2.038	0.010
	t-Statistic	2.325	0.147	0.322	-0.706	0.918	0.724	-0.735	-1.449	-1.583	-2.050	1.633	0.240	0.201			
DHO12	**Coefficient**	0.244	0.149	0.192	-0.104	0.361	0.210	-0.112	-0.715	-0.586	-0.685	0.602	0.049	0.045	0.047	2.023	0.009
	t-Statistic	2.358	0.281	0.442	-0.699	0.814	0.577	-0.849	-1.469	-1.575	-2.015	1.575	0.172	0.131			

Table 7: Forecasting Power of the Separate PCs(Cont'd)

j		C	a1	a2	a3	b1	b2	b3	c1	c2	c3	d1	d2	d3	R-squared	Durbin-Watson stat	Prob(F-stat)
									Panel D: NYMEX Gasoline								
DHU1	Coefficient	0.187	0.235	0.937	-0.002	0.001	-0.089	-0.224	-0.511	-0.881	-0.634	0.627	-0.433	0.045	0.008	2.030	0.180
	t-Statistic	0.859	0.236	1.066	-0.006	0.002	-0.113	-0.635	-0.399	-1.313	-0.890	0.749	-0.622	0.049			
DHU2	Coefficient	0.212	0.129	0.585	0.109	-0.035	-0.059	-0.245	-0.759	-0.949	-0.902	1.071	0.061	0.236	0.010	2.077	0.127
	t-Statistic	1.150	0.160	0.759	0.488	-0.049	-0.093	-0.853	-0.703	-1.536	-1.414	1.433	0.112	0.312			
DHU3	Coefficient	0.197	0.299	0.480	0.007	-0.055	0.051	-0.192	-0.796	-0.818	-0.963	1.047	0.218	0.201	0.009	2.077	0.151
	t-Statistic	1.188	0.408	0.696	0.036	-0.081	0.089	-0.829	-0.922	-1.467	-1.724	1.511	0.460	0.310			
DHU4	Coefficient	0.182	0.355	0.682	0.040	0.024	-0.113	-0.221	-0.799	-0.692	-0.966	0.902	0.114	0.050	0.016	2.049	0.048
	t-Statistic	1.237	0.517	1.023	0.229	0.039	-0.206	-1.094	-1.049	-1.454	-1.898	1.501	0.285	0.088			
DHU5	Coefficient	0.182	0.355	0.682	0.040	0.024	-0.113	-0.221	-0.799	-0.692	-0.966	0.902	0.114	0.050	0.016	2.049	0.048
	t-Statistic	1.237	0.517	1.023	0.229	0.039	-0.206	-1.094	-1.049	-1.454	-1.898	1.501	0.285	0.088			
DHU6	Coefficient	0.185	0.209	0.219	-0.162	0.291	0.257	-0.169	-0.633	-0.308	-0.569	0.485	-0.095	-0.058	0.012	1.999	0.090
	t-Statistic	1.444	0.343	0.403	-0.901	0.505	0.552	-0.926	-0.964	-0.703	-1.255	0.890	-0.275	-0.123			
DHU7	Coefficient	0.171	0.016	0.139	-0.040	0.306	0.225	-0.129	-0.636	-0.424	-0.769	0.755	0.026	-0.017	0.013	2.030	0.089
	t-Statistic	1.276	0.024	0.250	-0.209	0.560	0.493	-0.767	-1.032	-1.040	-1.693	1.412	0.082	-0.038			
DHU8	Coefficient	0.156	0.545	0.354	-0.142	0.118	0.205	-0.135	-0.696	-0.286	-0.761	0.458	-0.078	-0.220	0.020	1.982	0.024
	t-Statistic	1.192	0.799	0.649	-0.713	0.226	0.444	-0.764	-1.072	-0.753	-1.733	0.963	-0.247	-0.467			
DHU9	Coefficient	0.145	0.144	0.211	-0.122	0.436	0.371	-0.149	-0.591	-0.187	-0.678	0.315	-0.205	-0.324	0.021	2.054	0.021
	t-Statistic	1.154	0.226	0.391	-0.631	0.795	0.808	-0.876	-0.985	-0.517	-1.546	0.643	-0.668	-0.715			
DHU10	Coefficient	0.157	-0.062	-0.069	-0.143	0.351	0.492	-0.093	-0.335	-0.194	-0.699	0.547	-0.139	-0.303	0.020	2.049	0.028
	t-Statistic	1.237	-0.092	-0.120	-0.672	0.595	0.997	-0.546	-0.543	-0.493	-1.535	1.164	-0.423	-0.653			
DHU11	Coefficient	0.175	0.216	-0.187	-0.044	-0.036	0.344	-0.020	-0.577	-0.355	-0.653	0.671	0.176	0.088	0.006	2.060	0.229
	t-Statistic	1.333	0.339	-0.344	-0.245	-0.072	0.738	-0.117	-0.903	-0.886	-1.494	1.379	0.582	0.186			

Results from regressing ΔFt^{j} (j = CL1,…, CL15, CO1,…, CO9, HO1,…, HO12, HU1,…, HU11) on the twelve principal components obtained from the separate PCA on each one of the four commodities. Sample period: 8/11/1991-29/12/2006

Starting with NYMEX Crude Oil we should refer to the low R^2 values (0.022-0.033) which are consistent with Chantziara and Skiadopoulos results for the daily changes in these futures prices. So we conclude that PCs have no forecasting power in this case as far as the mid term horizon of the weekly futures prices is concerned. The first differences of the weekly NYMEX Crude Oil futures settlement prices seem to have as statistically significant explanatory variables the second ICE Crude PC (COPC2) and the first NYMEX Heating Oil PC (HUPC1), from the variables examined in our study. This result is the same across all fifteen maturities and the coefficients have always a positive sign and prices fluctuating between 0.26 and 0.44 for both regressors. Of course it would not make sense to assume that NYMEX Crude prices are driven by changes which had taken place in the ICE Crude one week ago. So we can just mention the existence of spillover effects, which affect the evolution of the weekly term structures between these markets. This is the case also for Gasoline, but the persistency of the results across maturities is an encouraging sign regarding the model's interpretation. One more statistically significant factor at the 5% level is the constant term, as far as contracts expiring in nine months and later are concerned. This remark could be interpreted as a specific mean reversion of these particular contracts first differences. The F-statistics indicate the rejection of null hypothesis across all maturities.

Considering the regression on ICE Crude results we observe in Table 7 the continuing low R^2 values (0.024-0.056) and the persistent significance of the first Heating Oil PC, (HUPC1), across all maturities. These coefficients have positive sign. All other coefficients are statistically insignificant except from the constant term which has forecasting power on first differences of the expiring last ICE Crude Oil futures weekly prices. So, as an obvious interference, we could note the importance of HUPC1 so as to derive some restricted predictions of the next weeks NYMEX and ICE Crude futures prices. Moreover, the contract with the longest maturity seems to have the more statistically significant predictors we have identified all over our research.

As we can see in Table 7 the NYMEX heating Oil case presents clear results, which forecasting performance is very restricted because of the low R^2 values (0.0359-0.047) the regression gives us. The contract with the shortest maturity seems to have two statistically significant predictors. The second PC of NYMEX Crude Oil (CLPC2) has predictive power on this with coefficient of positive sign. Another characteristic is the statistically significant third PC of the NYMEX Heating Oil itself (HOPC3), which affects the dependent variables through a negative coefficient. This result is common for all Heating Oil maturities. Moving to first price differences of the four contracts (DHO9, DHO10, DHO11, DHO12) which expire last we observe one more statistical significant coefficient, which is the constant term of the calculated equation. Until now the additive importance of the constant term for the last maturities constitutes a common characteristic across the three aforementioned commodities. Same attitude is observed for the F-statistics, at least for ICE Crude and NYMEX Heating Oil, with p-values declining as we move to later expiry dates.

The last contracts settlement price differences we use as dependent variables to run regressions are the NYMEX Gasoline futures. These regressions have statistically insignificant coefficient for all PCs, but they also come up with the lowest R^2 values (0.007-0.020). This picture is rather disappointing and this particular differentiation across Gasoline contracts should make us more critical towards the results in our effort to detect the PCs predictive power. This complex behaviour is repeated when we study F-statistic p-values across various maturities.

To conclude with the results mentioned above we have to pay attention to the small R^2 values found throughout the four commodities. This evidence suggests the limited forecasting power for PCs as independent variables on the weekly futures prices, which happens also with daily prices (see Chantziara and Skiadopoulos 2005). Moreover, the number of statistically significant PCs, which is very restricted, even zero for the whole term structure of NYMEX Gasoline futures, denotes the insufficient predictability possibilities of the model. Finally, we should refer to the effects appearing in the NYMEX Crude regressions. These express the only clear relation between two of the studied commodities, but it is very disorienting to assume clear spillover effects, because of our specific time horizon. Something that deserves to be mentioned is the increasing forecasting power of our regressions as we move to longer maturities. We have already paid attention to the constant terms which become statistically important at the 5% level for the weekly differences of the later expiring futures. Of course this is not the rule in Gasoline case. Something that could be an object for further investigation is the finding that there is no statistically significant PC in NYMEX Heating Oil regressions even though a Heating Oil PC (HUPC1) has important forecasting power in both NYMEX and ICE Crude Oil futures maturities.

Joint PCA: Regression and Results

The next step in our analysis has to do with the testing of the joint PCs forecasting performance. This approach makes more obvious the advantage of the small number of regressors, because it summarizes the term structure of all four commodities in three independent variables. The weakness of the method, compared to the previous one, is related with the fact we cannot search for spillover effects across commodities.

Our model's parameters are the similar with the aforementioned equation, but to be more specific we would like to introduce the following regression:

$$DF_t^{\,j} = c + a_1 PC_{1,t-1} + a_2 PC_{2,t-1} + a_3 PC_{3,t-1} + u_t \quad \textbf{(5)}$$

We have the differences of the futures weekly prices at time t and the corresponding generics as dependent variables

$PC_{1,t-1}$, $PC_{2,t-1}$ and $PC_{3,t-1}$ are the time series of the principal components that have been obtained from the joint PCA at time t-1.

a_1, a_2 and a_3 are the coefficients that come from the regression analysis

We also used Newey-West standard errors again, which correct both for heteroskedasticity and autocorrelation, to obtain our coefficients t-statistics.

Table 8: Forecasting Power of the Joint PCs

j		C	a1	a2	a3	R-squared	Durbin-Watson stat	Prob(F-statistic)
				Panel A: NYMEX Crude Oil				
DCL1	Coefficient	0.062	-0.006	-0.063	0.028	-0.003	2.055	0.718
	t-Statistic	1.107	-0.061	-0.762	0.459			
DCL2	Coefficient	0.066	0.007	-0.066	0.005	-0.003	2.041	0.699
	t-Statistic	1.264	0.079	-0.881	0.086			
DCL3	Coefficient	0.068	0.023	-0.066	0.004	-0.002	2.044	0.611
	t-Statistic	1.376	0.268	-0.940	0.080			
DCL4	Coefficient	0.069	0.032	-0.063	0.005	-0.002	2.046	0.557
	t-Statistic	1.471	0.385	-0.942	0.094			
DCL5	Coefficient	0.070	0.037	-0.058	0.003	-0.001	2.050	0.533
	t-Statistic	1.558	0.451	-0.916	0.061			
DCL6	Coefficient	0.071	0.039	-0.054	0.001	-0.001	2.054	0.517
	t-Statistic	1.635	0.497	-0.875	0.019			
DCL7	Coefficient	0.072	0.040	-0.050	-0.002	-0.001	2.061	0.527
	t-Statistic	1.706	0.521	-0.817	-0.046			
DCL8	Coefficient	0.073	0.039	-0.047	-0.003	-0.001	2.062	0.535
	t-Statistic	1.774	0.516	-0.788	-0.065			
DCL9	Coefficient	0.073	0.037	-0.046	-0.006	-0.001	2.058	0.527
	t-Statistic	1.831	0.508	-0.797	-0.142			
DCL10	Coefficient	0.074	0.039	-0.044	-0.007	-0.001	2.056	0.503
	t-Statistic	1.896	0.554	-0.773	-0.150			
DCL11	Coefficient	0.074	0.040	-0.043	-0.007	-0.001	2.050	0.481
	t-Statistic	1.946	0.574	-0.768	-0.154			
DCL12	Coefficient	0.074	0.040	-0.044	-0.009	-0.001	2.047	0.450
	t-Statistic	1.982	0.595	-0.783	-0.212			
DCL13	Coefficient	0.074	0.042	-0.044	-0.010	0.000	2.039	0.411
	t-Statistic	2.021	0.635	-0.794	-0.230			
DCL14	Coefficient	0.074	0.042	-0.044	-0.010	0.000	2.029	0.382
	t-Statistic	2.049	0.649	-0.818	-0.247			
DCL15	Coefficient	0.074	0.042	-0.044	-0.011	0.000	2.021	0.368
	t-Statistic	2.081	0.662	-0.820	-0.257			

Table 8: Forecasting Power of the Joint PCs (Cont'd)

Panel B: ICE Crude Oil									
j		**C**	**a1**	**a2**	**a3**	**R-squared**	**Durbin-Watson stat**	**Prob(F-statistic)**
DCO1	**Coefficient**	0.070	0.000	0.002	-0.028	-0.005	2.004	0.966
	t-Statistic	1.289	-0.001	0.027	-0.516			
DCO2	**Coefficient**	0.070	0.022	-0.021	-0.019	-0.004	2.017	0.915
	t-Statistic	1.399	0.274	-0.277	-0.371			
DCO3	**Coefficient**	0.071	0.041	-0.025	-0.015	-0.003	2.023	0.765
	t-Statistic	1.514	0.535	-0.355	-0.308			
DCO4	**Coefficient**	0.072	0.056	-0.022	-0.014	-0.002	2.034	0.589
	t-Statistic	1.615	0.761	-0.338	-0.284			
DCO5	**Coefficient**	0.072	0.056	-0.024	-0.016	-0.001	2.036	0.530
	t-Statistic	1.678	0.795	-0.386	-0.342			
DCO6	**Coefficient**	0.072	0.056	-0.025	-0.017	-0.001	2.038	0.501
	t-Statistic	1.721	0.806	-0.407	-0.367			
DCO7	**Coefficient**	0.072	0.051	-0.025	-0.018	-0.001	2.039	0.529
	t-Statistic	1.771	0.761	-0.420	-0.386			
DCO8	**Coefficient**	0.072	0.051	-0.022	-0.019	-0.001	2.039	0.514
	t-Statistic	1.835	0.773	-0.386	-0.421			
DCO9	**Coefficient**	0.072	0.052	-0.019	-0.021	-0.001	2.035	0.483
	t-Statistic	1.871	0.804	-0.340	-0.475			

Table 8: Forecasting Power of the Joint PCs (Cont'd)

j		C	a1	a2	a3	R-squared	Durbin-Watson stat	Prob(F-statistic)
		Panel C: NYMEX Heating Oil						
DHO1	**Coefficient**	0.142	0.241	-0.369	-0.265	0.007	2.069	0.067
	t-Statistic	0.840	0.738	-1.582	-1.329			
DHO2	**Coefficient**	0.165	0.198	-0.351	-0.219	0.006	2.071	0.088
	t-Statistic	1.031	0.607	-1.662	-1.079			
DHO3	**Coefficient**	0.169	0.193	-0.269	-0.172	0.003	2.060	0.181
	t-Statistic	1.089	0.615	-1.329	-0.880			
DHO4	**Coefficient**	0.174	0.222	-0.224	-0.127	0.003	2.061	0.212
	t-Statistic	1.144	0.752	-1.092	-0.678			
DHO5	**Coefficient**	0.181	0.242	-0.223	-0.088	0.004	2.073	0.164
	t-Statistic	1.233	0.865	-1.053	-0.496			
DHO6	**Coefficient**	0.186	0.274	-0.273	-0.107	0.009	2.093	0.046
	t-Statistic	1.348	1.039	-1.253	-0.646			
DHO7	**Coefficient**	0.196	0.265	-0.275	-0.121	0.011	2.104	0.028
	t-Statistic	1.532	1.063	-1.310	-0.762			
DHO8	**Coefficient**	0.210	0.237	-0.260	-0.122	0.010	2.129	0.031
	t-Statistic	1.755	0.992	-1.303	-0.807			
DHO9	**Coefficient**	0.221	0.205	-0.233	-0.129	0.009	2.136	0.047
	t-Statistic	1.953	0.893	-1.234	-0.866			
DHO10	**Coefficient**	0.233	0.203	-0.224	-0.142	0.010	2.130	0.038
	t-Statistic	2.124	0.929	-1.234	-0.984			
DHO11	**Coefficient**	0.237	0.227	-0.207	-0.151	0.011	2.091	0.026
	t-Statistic	2.202	1.106	-1.171	-1.077			
DHO12	**Coefficient**	0.233	0.212	-0.200	-0.177	0.012	2.054	0.021
	t-Statistic	2.196	1.093	-1.182	-1.312			

Table 8: Forecasting Power of the Joint PCs (Cont'd)

Panel D: NYMEX Gasoline								
j		C	a1	a2	a3	R-squared	Durbin-Watson stat	Prob(F-stat)
DHU1	Coefficient	0.173	0.306	-0.309	-0.402	0.007	2.077	0.083807
	t-Statistic	0.825	0.778	-0.954	-1.402			
DHU2	Coefficient	0.187	0.187	-0.304	-0.203	0.003	2.116	0.205778
	t-Statistic	1.047	0.573	-1.185	-0.969			
DHU3	Coefficient	0.177	0.244	-0.224	-0.057	0.001	2.098	0.288509
	t-Statistic	1.103	0.799	-1.027	-0.284			
DHU4	Coefficient	0.165	0.247	-0.202	-0.161	0.004	2.078	0.144973
	t-Statistic	1.149	0.888	-1.117	-0.912			
DHU5	Coefficient	0.165	0.247	-0.202	-0.161	0.004	2.078	0.144973
	t-Statistic	1.149	0.888	-1.117	-0.912			
DHU6	Coefficient	0.180	0.237	-0.030	-0.211	0.004	1.982	0.149872
	t-Statistic	1.373	0.965	-0.188	-1.251			
DHU7	Coefficient	0.160	0.176	-0.105	-0.203	0.002	2.003	0.227279
	t-Statistic	1.164	0.735	-0.622	-1.340			
DHU8	Coefficient	0.155	0.256	-0.081	-0.257	0.008	1.932	0.061839
	t-Statistic	1.144	0.986	-0.502	-1.637			
DHU9	Coefficient	0.145	0.175	-0.101	-0.316	0.009	2.019	0.042742
	t-Statistic	1.116	0.736	-0.623	-2.068			
DHU10	Coefficient	0.159	0.178	-0.073	-0.257	0.005	2.009	0.117894
	t-Statistic	1.219	0.728	-0.418	-1.610			
DHU11	Coefficient	0.168	0.066	-0.109	-0.092	-0.003	2.030	0.686796
	t-Statistic	1.265	0.279	-0.588	-0.658			

Results from regressing ΔFt^j (j = CL1,…, CL15, CO1,…, CO9, HO1,…, HO12, HU1,…, HU11) on the three principal components obtained from the joint PCA on each one of the four commodities. Sample period: 8/11/1991-29/12/2006

Table 8 presents this method's results. The first column shows the dependent variables per commodity for all the corresponding maturities. The next four columns present the constant term and the PCs coefficients with their t-statistics underneath them. We have put into frames the coefficients which t-statistics prove them to be statistically significant at the 5% significance level. The following column shows the R^2 statistic, the last two present the Durbin –Watson statistic for first order serial correlation and the F-statistic, respectively.

The results this time are very poor. R^2 values don't exceed 1.1% and we usually come up with negative values which indicate that our models are poorly fitting our data. We should note that we use the adjusted R-squared statistic, as we have done also for the separate PCs. Furthermore, the three retained PCs have no forecasting power for almost all contracts. The only statistical significant coefficients are the constant terms of four NYMEX Crude Oil and three Heating Oil regressions which dependent variables are the weekly price differences of the futures that have the longest maturities across the two commodities. Moreover, Heating Oil is the only commodity which longer maturities futures regressions result in some F-statistics able to reject the null hypothesis. Consequently, PCs obtained from the joint PCA have no forecasting power on the weekly futures prices we studied. Furthermore their performance is even worse than the one estimated by Chantziara and Skiadopoulos for the daily prices predictability.

CHAPTER 4: ECONOMIC ANALYSIS

The second approach studied in this paper is an economic (structural) approach, which aims to identify the relation between the same as before dependent variables and some other economic or financial factors. This work is based on the methodology followed by Sadorsky whose article offers encouraging results as far as the forecasting power of his mode is concerned. Some distinct points of this research have to do with the different time horizon (weekly instead of monthly) and the introduction of some more variables in our dataset. The dependent variable, which is added in Sadorsky's data, is the first differences of ICE Crude Oil future settlement prices. So, there is no reason to be more comprehensive about our dependent variables which are already analyzed for the purposes of our PCA approach. On the other hand we have to write some remarks in order to present our explanatory parameters and this, exactly, section of the study follows. As a final observation we should say that we use the first differences of all the incorporated variables instead of Sadorsky who takes the returns of the data prices. This change is of little importance for the forthcoming results, but it is necessary in our effort to be consistent with the first part of our own analysis (PCA).

4.1 ECONOMIC DATA

Constructing our independent variables data set we followed Sadorsky's work and the variables he used as guidelines. In addition to this, the introduction of ICE Crude Oil futures first differences as dependent variable induced the necessity to enrich our explanatory data. We should take into account some economic factors indicative of the U.K. economy in order to search a possible relation between them and, especially, this specific commodity.

Firstly, we have to find two elements representative of the U.S. and U.K. market portfolios. For the fist one we picked up the S&P500 index and as an equivalent to the U.K. portfolio we chose the FTSE 100 index.

The S&P 500 contains the stocks of 500 Large-Cap corporations, most of which are American. The index is the most notable of the many indices owned and maintained by Standard & Poor's, a division of McGraw-Hill, and was introduced in 1957. All of the stocks in the index are those of large publicly held companies and trade on the two largest US stock markets, the New York Stock Exchange and Nasdaq. After the Dow Jones Industrial Average, the S&P 500 is the most widely watched index of large-cap US stocks. It is considered to be an indicator of the US economy.

The FTSE 100 Index is a share index of the 100 most highly capitalised companies listed on the London Stock Exchange. The index is seen as a barometer of the British economy and is the leading share index in Europe. It is maintained by the FTSE Group, a now independent company which originated as a joint venture between the *Financial Times* and the London Stock Exchange. According to the FTSE Group's website the FTSE 100 companies represent about 80% of the UK share market.

We also tried to include in our predictors the dividend yields of the aforementioned markets. Consequently, we qualify as indicators the weekly differences of the S&P 500 and FTSE 100 annual dividend yields (S&P500DY and FTSE100DY). These dividend yields represent the annual dividend levels of stock portfolios simulating the two indices divided by the market values of the two portfolios, estimated for each week.

Sadorsky introduces the 90-day Treasury Bill annual yield as a benchmark for the interest risk free rate. Treasury bills, or T-Bills, are like zero-coupon bonds in that they do not pay interest prior to maturity; instead they are sold at a discount of the par value to create a positive yield to maturity. Treasury bills are considered to be the most risk-free investment for U.S. investors, so it is a suitable indicator for the U.S. economy. As far as U.K. conditions are concerned we have to contribute another variable which we chose to be the annual yield of the 90-day London Interbank Offered Rate (LIBOR). LIBOR is a daily reference rate based on the interest rates at which banks offer to lend unsecured funds to other banks in the London wholesale money market (or interbank market). So, it is regarded as the premium equal to the risk free rate which banks are willing to pay so as to secure their assets. Of course for both rate, like all our variables, we use their weekly differences.

As a last variable we incorporate another one of Sadorsky's factors. This the monthly return on the annual yield on Moody's long term BAA-rated bond minus the yield on AAA-rated bonds. This parameter is considered to be an interpretation of the default risk premium which is adequate for large companies and corporations.

We have already present summary statistics for petroleum futures so now we only have to present the statistics of the economic factor's weekly differences.

Table 9: Summary Statistics of the first differences of the economic and financial variables

Panel E: Economic Variables

Sample: 11/08/1991 12/29/2006

	DSP500	DFTSE100	DSP500DY	DFTSE100DY	DTBILL	DLIBOR	DBAAAAA
Mean	1.298	4.641	0.000	0.000	0.000	0.000	0.000
Std. Dev.	22.521	100.720	0.000	0.001	0.001	0.001	0.000
Skewness	-0.625	-0.297	0.053	-1.073	-1.983	-1.713	2.161
Kurtosis	9.187	5.530	5.768	12.893	20.136	18.449	30.773
Jarque-Bera	1313.023	222.547	252.835	3377.245	10195.730	8253.412	26037.020
Probability	0	0	0	0	0	0	0
Observations	791	791	791	791	791	791	791

Table 9 shows these summary statistics for the seven causal variables and Table 10 the ADF tests for these series which indicate that we have to do with stationary series at the 1% significance level.

Table 10: Unit Root Tests for the first differences of the weekly prices of the economic and financial variables. Sample period: 8/11/1991-29/12/2006

Panel E: Economic Variables

	ADF test statistic
	First differences
DSP500	-32.093
DFTSE100	-29.482
DSP500DY	-32.750
DFTSE100DY	-28.222
DTBILL	-17.185
DLIBOR	-10.171
DBAAAAA	-25.036

4.2 ECONOMIC VARIABLES AND FORECASTING POWER

In this section we perform a multivariate regression model using as regressors the differences of the seven economic variables. This model tries to identify the forecasting skills of some economic variables in a simple regression model. The derivation of this model is based on an economic (structural) approach and, as a matter of fact, it is affected by the major drawback of this kind of models which is connected with the choice of suitable causal variables. As we have already mentioned, in our case we used as guideline Sadorsky's article, but there is no proof that we could repeat his important results in the interpretation of our model.

Economic analysis: The regression settings and results

We move on with the presentation of our basic model starting with the regression equation which follows

$$DF_t^{\,j} = c + a_1(DS\,\&\,P500DY_{t-1}) + a_2(DBAA - AAA_{t-1}) + a_3(DTBILL_{t-1}) +$$
$$a_4(DS\,\&\,P500_{t-1}) + a_5(DFTSE100_{t-1}) + a_6(DLIBOR_{t-1}) + a_7(DFTSE100DY_{t-1}) + u_t \quad \textbf{(6)}$$

$DF_t^{\,j}$ represents the weekly changes of futures prices at time t, as before, and j stands again for the different generics.

The independent variables are also differences of the weekly prices of the following economic indicators

$DS\,\&\,P500_{t-1}$ and $DS\,\&\,P500DY_{t-1}$ are the weekly difference of the S&P500 Index and the weekly difference of the same Index's annual dividend yield, respectively at time t-1

$DFTSE100_{t-1}$ and $DFTSE100DY_{t-1}$ describe variables that have been similarly constructed and have as subject the weekly differences of FTSE100 Index and its annual dividend yield at time t-1

$DTBILL_{t-1}$ and $DLIBOR_{t-1}$ indicate the weekly differences of the two risk free interest rates at time t-1. $DTBILL_{t-1}$ is the weekly difference of the three month T-Bill annualised rate (representative for the US risk free rate) and for the UK economy we use the $DLIBOR_{t-1}$ (weekly difference of the 3-month LIBOR annualised rate at time t-1).

$DBAA - AAA_{t-1}$ denotes the default risk premium and is calculated from the weekly difference of the annual yield on Moody's long term BAA-rated corporate bonds minus the return of the yield on AAA-rated bonds

a_1, a_2, etc. are the regression retained coefficients

4.3 ECONOMIC VARIABLES MODEL

To start with the examination of the forecasting skills of the factors mentioned above we don't make any attempt to model our errors at this stage. We should also note that we detect heteroskedasticity in the regression residuals, so in this simple regression equation we used Newey-West standard errors, which correct both for heteroskedasticity and autocorrelation, to obtain our coefficients t-statistics.

Table 11 presents this regression results per commodity and across the different maturities. The first column shows again our dependent variables and the next eight columns show the coefficients and the corresponding t-statistics of the constant term and the seven economic variables. The last three columns contain the R^2 statistic, the Durbin –Watson statistic for first order serial correlation and the F-statistic, respectively.

NYMEX Crude Oil regressions give very low R^2 values (0.002-0.010) which is a common characteristic among all four commodities and indicates that our model doesn't present forecasting power. On the other hand we observe that there are some statistically significant factors. These are the T-Bill rate and the S&P 500 for almost all contract maturities, the FTSE 100 Index for the six futures which expire the next ten to fifteen months and the constant terms for the last two maturities (DCL14, DCL15). Furthermore, this commodity is the only one which majority of regressions give us F-statistics high enough to reject the null hypothesis at the 5% significance level.

Table 11: Forecasting Power of the Economic Variables

			Panel A: NYMEX Crude Oil									
Independent Variables			**S&P500DY**	**BAA-AAA**	**T-Bill**	**S&P500**	**FTSE100**	**LIBOR**	**FTSE100DY**			
j		**C**	**a1**	**a2**	**a3**	**a4**	**a5**	**a6**	**a7**	**R-squared**	**Durbin-Watson stat**	**Prob(F-stat)**
DCL1	**Coefficient**	0.052	-18.487	-84.851	100.637	-0.007	0.001	10.916	72.613	0.002	2.132	0.262
	t-Statistic	1.128	-0.121	-0.632	2.049	-2.025	1.240	0.308	0.658			
DCL2	**Coefficient**	0.054	36.492	-102.187	104.910	-0.007	0.002	18.270	46.539	0.006	2.100	0.098
	t-Statistic	1.248	0.265	-0.858	2.368	-2.271	1.501	0.540	0.458			
DCL3	**Coefficient**	0.056	39.120	-108.075	103.670	-0.007	0.002	19.536	48.115	0.009	2.084	0.051
	t-Statistic	1.341	0.313	-0.996	2.535	-2.463	1.703	0.627	0.518			
DCL4	**Coefficient**	0.057	43.169	-121.779	96.852	-0.007	0.002	19.978	49.069	0.010	2.074	0.037
	t-Statistic	1.429	0.378	-1.204	2.538	-2.562	1.831	0.699	0.572			
DCL5	**Coefficient**	0.058	44.954	-125.137	89.798	-0.006	0.002	20.723	45.113	0.010	2.069	0.034
	t-Statistic	1.512	0.426	-1.313	2.492	-2.608	1.875	0.785	0.559			
DCL6	**Coefficient**	0.058	45.870	-125.292	84.680	-0.006	0.001	19.926	44.270	0.010	2.068	0.033
	t-Statistic	1.586	0.466	-1.381	2.460	-2.642	1.914	0.813	0.579			
DCL7	**Coefficient**	0.059	41.392	-120.684	79.505	-0.006	0.001	19.566	41.488	0.010	2.071	0.033
	t-Statistic	1.660	0.447	-1.389	2.414	-2.709	1.945	0.846	0.568			
DCL8	**Coefficient**	0.060	43.080	-118.949	75.296	-0.005	0.001	17.281	43.288	0.010	2.072	0.042
	t-Statistic	1.721	0.496	-1.412	2.361	-2.660	1.936	0.806	0.621			
DCL9	**Coefficient**	0.061	37.531	-112.346	70.796	-0.006	0.001	18.775	36.185	0.010	2.072	0.038
	t-Statistic	1.791	0.454	-1.392	2.304	-2.802	1.949	0.902	0.536			
DCL10	**Coefficient**	0.061	35.227	-106.665	66.941	-0.006	0.001	18.735	35.461	0.010	2.067	0.040
	t-Statistic	1.852	0.447	-1.365	2.238	-2.830	1.968	0.945	0.544			
DCL11	**Coefficient**	0.061	40.557	-102.203	64.140	-0.005	0.001	16.443	37.732	0.009	2.064	0.050
	t-Statistic	1.902	0.539	-1.346	2.181	-2.772	1.984	0.874	0.602			
DCL12	**Coefficient**	0.062	41.545	-95.398	59.833	-0.005	0.001	16.991	36.654	0.009	2.063	0.045
	t-Statistic	1.955	0.573	-1.287	2.092	-2.867	2.060	0.921	0.598			
DCL13	**Coefficient**	0.062	41.455	-91.319	56.541	-0.005	0.001	16.616	33.439	0.009	2.057	0.050
	t-Statistic	1.996	0.592	-1.261	2.024	-2.874	2.043	0.926	0.559			
DCL14	**Coefficient**	0.062	43.370	-87.008	53.492	-0.005	0.001	16.351	30.558	0.009	2.052	0.056
	t-Statistic	2.034	0.642	-1.226	1.948	-2.859	2.033	0.931	0.522			
DCL15	**Coefficient**	0.062	42.991	-82.980	51.035	-0.005	0.001	16.089	28.896	0.008	2.047	0.063
	t-Statistic	2.070	0.659	-1.193	1.890	-2.853	2.032	0.930	0.503			

Table 11: Forecasting Power of the Economic Variables (Cont'd)

Independent Variables			S&P500DY	BAA-AAA	T-Bill	S&P500	FTSE100	LIBOR	FTSE100DY			
			Panel B: ICE Crude Oil									
j		C	a1	a2	a3	a4	a5	a6	a7	R-squared	Durbin-Watson stat	Prob(F-stat)
DCO1	Coefficient	0.052	21.303	-115.319	105.493	-0.006	0.002	8.789	52.046	0.005	2.055	0.133
	t-Statistic	1.154	0.167	-1.034	2.562	-1.817	1.469	0.272	0.524			
DCO2	Coefficient	0.054	5.239	-101.156	97.999	-0.006	0.001	12.426	44.420	0.007	2.062	0.087
	t-Statistic	1.288	0.046	-0.987	2.520	-2.230	1.560	0.419	0.477			
DCO3	Coefficient	0.055	14.454	-105.100	99.670	-0.005	0.001	15.017	38.910	0.008	2.065	0.067
	t-Statistic	1.395	0.140	-1.085	2.821	-2.163	1.574	0.540	0.453			
DCO4	Coefficient	0.057	18.233	-107.248	100.420	-0.005	0.001	15.142	29.197	0.009	2.058	0.045
	t-Statistic	1.503	0.191	-1.180	2.987	-2.331	1.538	0.571	0.363			
DCO5	Coefficient	0.058	18.455	-111.010	89.410	-0.005	0.001	15.460	42.544	0.008	2.062	0.059
	t-Statistic	1.570	0.207	-1.264	2.749	-2.325	1.693	0.642	0.554			
DCO6	Coefficient	0.058	19.687	-110.846	83.787	-0.005	0.001	13.221	40.312	0.008	2.063	0.065
	t-Statistic	1.639	0.235	-1.305	2.669	-2.363	1.720	0.587	0.552			
DCO7	Coefficient	0.059	18.799	-112.490	80.573	-0.005	0.001	14.627	37.327	0.008	2.068	0.070
	t-Statistic	1.713	0.237	-1.362	2.650	-2.358	1.700	0.678	0.534			
DCO8	Coefficient	0.060	19.264	-112.645	74.221	-0.005	0.001	14.938	35.372	0.007	2.063	0.085
	t-Statistic	1.774	0.255	-1.406	2.524	-2.355	1.694	0.732	0.525			
DCO9	Coefficient	0.060	8.658	-104.444	71.085	-0.004	0.001	12.446	40.874	0.006	2.058	0.106
	t-Statistic	1.819	0.120	-1.351	2.483	-2.323	1.725	0.640	0.636			

Table 11: Forecasting Power of the Economic Variables (Cont'd)

Independent Variables			Panel C: NYMEX Heating Oil									
			S&P500DY	BAA-AAA	T-Bill	S&P500	FTSE100	LIBOR	FTSE100DY			
j		C	a1	a2	a3	a4	a5	a6	a7	R-squared	Durbin-Watson stat	Prob(F-stat)
DHO1	Coefficient	0.119	-41.736	-760.581	222.256	-0.020	0.004	13.225	-127.670	0.004	2.120	0.197
	t-Statistic	0.846	-0.094	-1.422	1.771	-1.710	1.188	0.117	-0.364			
DHO2	Coefficient	0.130	111.061	-704.624	203.312	-0.022	0.005	36.955	23.742	0.006	2.113	0.108
	t-Statistic	0.989	0.274	-1.510	1.510	-2.324	1.575	0.373	0.077			
DHO3	Coefficient	0.138	207.348	-674.463	203.563	-0.021	0.005	44.676	87.697	0.007	2.092	0.081
	t-Statistic	1.086	0.559	-1.591	1.551	-2.307	1.640	0.480	0.296			
DHO4	Coefficient	0.147	225.991	-658.537	217.669	-0.020	0.005	50.148	93.589	0.008	2.078	0.064
	t-Statistic	1.176	0.671	-1.700	1.774	-2.284	1.592	0.577	0.332			
DHO5	Coefficient	0.151	169.746	-670.665	235.140	-0.018	0.004	41.335	92.494	0.009	2.081	0.055
	t-Statistic	1.239	0.546	-1.902	2.074	-2.233	1.538	0.522	0.352			
DHO6	Coefficient	0.154	152.706	-631.297	230.134	-0.016	0.004	25.779	75.409	0.007	2.093	0.075
	t-Statistic	1.308	0.532	-1.920	2.171	-2.151	1.471	0.356	0.311			
DHO7	Coefficient	0.155	91.492	-524.283	211.229	-0.014	0.003	1.665	83.316	0.005	2.104	0.148
	t-Statistic	1.375	0.344	-1.752	2.136	-2.033	1.447	0.024	0.372			
DHO8	Coefficient	0.157	25.778	-412.598	186.281	-0.013	0.003	-8.705	94.855	0.003	2.138	0.229
	t-Statistic	1.473	0.104	-1.534	2.068	-2.076	1.516	-0.135	0.456			
DHO9	Coefficient	0.160	19.812	-340.115	173.996	-0.012	0.003	-14.373	82.371	0.002	2.173	0.279
	t-Statistic	1.604	0.084	-1.380	2.049	-2.084	1.479	-0.236	0.423			
DHO10	Coefficient	0.161	39.234	-320.219	166.917	-0.012	0.003	-17.648	54.960	0.002	2.189	0.289
	t-Statistic	1.704	0.177	-1.352	2.031	-1.999	1.432	-0.304	0.298			
DHO11	Coefficient	0.163	59.677	-338.786	156.847	-0.012	0.003	-9.174	71.106	0.003	2.154	0.235
	t-Statistic	1.791	0.281	-1.456	1.967	-2.018	1.573	-0.163	0.381			
DHO12	Coefficient	0.166	74.351	-329.662	141.837	-0.010	0.003	2.975	110.076	0.002	2.135	0.281
	t-Statistic	1.880	0.357	-1.438	1.795	-1.867	1.619	0.055	0.582			

Table 11: Forecasting Power of the Economic Variables (Cont'd)

			Panel D: NYMEX Gasoline									
Independent Variables			S&P500DY	BAA-AAA	T-Bill	S&P500	FTSE100	LIBOR	FTSE100DY			
j		C	a1	a2	a3	a4	a5	a6	a7	R-squared	Durbin-Watson stat	Prob(F-stat)
DHU1	Coefficient	0.118	-11.082	-480.568	102.298	-0.025	0.005	-61.037	65.414	-0.001	2.121	0.523
	t-Statistic	0.649	-0.024	-0.836	0.585	-2.063	1.421	-0.494	0.174			
DHU2	Coefficient	0.136	82.539	-270.608	167.812	-0.020	0.004	-37.874	-40.393	0.000	2.101	0.447
	t-Statistic	0.847	0.200	-0.611	1.049	-2.007	1.309	-0.355	-0.127			
DHU3	Coefficient	0.108	-167.010	-94.920	158.099	-0.019	0.004	-13.609	56.780	0.000	2.089	0.431
	t-Statistic	0.753	-0.453	-0.266	1.231	-2.102	1.539	-0.149	0.195			
DHU4	Coefficient	0.131	-68.175	-64.204	126.232	-0.014	0.003	35.996	43.726	-0.003	2.113	0.676
	t-Statistic	1.048	-0.217	-0.209	1.010	-1.740	1.286	0.429	0.172			
DHU5	Coefficient	0.131	-68.175	-64.204	126.232	-0.014	0.003	35.996	43.726	-0.003	2.113	0.676
	t-Statistic	1.048	-0.217	-0.209	1.010	-1.740	1.286	0.429	0.172			
DHU6	Coefficient	0.166	151.893	-77.080	152.370	-0.015	0.005	5.124	138.373	0.003	2.068	0.210
	t-Statistic	1.607	0.542	-0.294	1.299	-1.996	1.898	0.068	0.573			
DHU7	Coefficient	0.184	90.039	-332.106	176.586	-0.013	0.003	49.578	92.038	0.002	2.070	0.319
	t-Statistic	1.717	0.334	-1.258	1.509	-1.898	1.341	0.675	0.412			
DHU8	Coefficient	0.166	129.052	-410.016	159.021	-0.014	0.003	55.534	79.807	0.002	2.015	0.290
	t-Statistic	1.520	0.466	-1.462	1.381	-1.917	1.120	0.727	0.331			
DHU9	Coefficient	0.187	93.477	-294.680	162.576	-0.015	0.002	62.073	82.146	0.003	2.080	0.248
	t-Statistic	1.697	0.372	-1.147	1.366	-2.348	1.382	0.853	0.403			
DHU10	Coefficient	0.181	107.411	-353.611	187.005	-0.015	0.003	63.090	118.601	0.003	2.103	0.248
	t-Statistic	1.541	0.398	-1.302	1.531	-2.260	1.433	0.737	0.480			
DHU11	Coefficient	0.176	140.845	-419.331	177.075	-0.016	0.003	62.586	179.792	0.005	2.065	0.192
	t-Statistic	1.382	0.476	-1.326	1.453	-2.349	1.636	0.703	0.798			

Results from regressing ΔFt^{j} (j = CL1,…, CL15, CO1,…, CO9, HO1,…, HO12, HU1,…, HU11) on the first differences of the seven economic and financial factors weekly prices. Sample period: 8/11/1991-29/12/2006

Moving on to ICE Crude we retain similar results with low R^2 values, they don't even reach 1%. The similarities continue to exist as far as the significant regressors are concerned. So, T-Bill rate and S&P 500 seem to have some predictability on all but the first one contract. Major difference lies on the low F-statistic values, which indicate non zero coefficient only for the future expiring in four months.

Exactly the same picture appears in NYMEX Heating Oil futures first differences. The R^2 statistic fluctuates between 0.2% and 0.9% proving the poor forecasting power of our model, whereas the statistically significant parameters consist of the same variables (T-Bill and S&P 500) in the case of Heating Oil.

The Gasoline results are more disappointing. The R^2 values remain below 0.5% for all regressions and they often become negative. The coefficients show also disappointing results and only the S&P 500 weekly differences shows predictability on seven futures series. Assuming the last two commodities we should refer to the fact that F-statistic ensures the null Hypothesis (that all coeficients are statistically insignificant at the 5% level) across the regressions of all maturities.

To sum up these regression results we can refer to the previous literature. Our research is made in a weekly horizon, instead of monthly as it was Sadorsky's case, but in general we come up with the same findings. We ensured that a simple multivariate regression model using economic variables as independent variables has no explanatory or forecasting power on petroleum futures first differences, but we also extended the former conclusions demonstrating that the inclusion of more regressors cannot improve this model's performance. Moreover, the introduction of the U.K. economic factors didn't offer any enhancement of the predictability and not a single one of them was statistically significant as a predictor of the ICE Crude or any other commodity price differences at the 5% significance level. To conclude with this model analysis, we showed that we have to seek for a different regression model instead of enriching this one with more explanatory variables.

4.4 ARMA-GARCH MODELS

As an extension of the aforementioned analysis we try to enhance our model by making it more accurate so as to correct our residuals serial autocorrelation and the ARCH effects that are observed in our regressions. To face these difficulties in the previous analysis we derive the equations using the Newey-West heteroskedasticity and autocorrelation consistent covariance matrix, now we introduce ARMA-GARCH models in order to include autoregressive factors in our residuals distribution and to estimate the residuals conditional variance. The rest independent variables remain constant as before in the simple regression settings with the economic factors.

Specification

To be more consistent with our forthcoming regressions we present a brief description of an ARMA-GARCH model.

The Autoregressive Moving Average (ARMA_ part of the model has to do with the autoregressive characteristics we use to model our residuals. First of all we assume that our regression is of the form:

$$R_t = b * X_{t-1} + e_t \quad \textbf{(7)}$$

R_t denotes our dependent variable and $b * X_{t-1}$ our independent variable X_{t-1} along with the estimated coefficient b. As we can observe R_t is defined at time t whereas the regressor is retained one time step before at time t-1. This is the case also for our data. Of course we have more than one independent variables, but this is not the matter for the time being. e_t represents the error term which distribution we would like to model. The first tool is the autoregressive, or AR, term. Each AR term denotes a lagged autoregressive factor in our error equation. An AR(1) model uses only the first-order term, but in favour of our analysis we use also higher-order AR terms. So, an AR(p) model can be described by the following equation:

$$e_t = r_1 e_{t-1} + r_2 e_{t-2} + ... + r_p e_{t-p} + u_t \quad \textbf{(8)}$$

The r terms denote the coefficients and of course there is no constant term because the error mean still is zero. u_t is the error term of our residuals regression.

The next part is the MA, or moving average term. A moving average forecasting model uses lagged values of u_t to improve the current estimation of the main error regression term, e_t. So our error equation for a MA(q) model is:

$$e_t = u_t + q_1 u_{t-1} + q_2 u_{t-2} + ... + q_p u_{t-q} \quad \textbf{(9)}$$

The autoregressive and moving average specifications can be combined to form an ARMA(p, q) specification:

$$e_t = r_1 e_{t-1} + r_2 e_{t-2} + ... + r_p e_{t-p} + u_t + q_1 u_{t-1} + q_2 u_{t-2} + ... + q_p u_{t-q} \quad \textbf{(10)}$$

The GARCH part of the suggested models has to do with the estimation of the residuals variance. Assuming that this variance is h_t we can present the following equation for the conditional variance of a GARCH(m, n) model:

$$h_t = w + \sum_{i=1}^{n} a_i e_{t-n}^2 + \sum_{j=1}^{m} b_j s_{t-m}^2 \quad \textbf{(11)}$$

As we can see w stands for a constant term and we also have a term (e_{t-n}^2) for volatility from the previous period, measured as the lags of the squared residuals (the ARCH term). The last factor is the GARCH term which estimates the last period's variance effect.

Regressions and Results

Moving on to the specific regressions we have to note that our effort is to find ARMA-GARCH models fitted well on each future we examine. We choose this approach in order to have the most appropriate model for each contract's weekly differences and to avoid misleading generalizations across futures on the same commodities or with similar periods to maturity. As a critical value so as to decide the order of our models we use the Schwarz information criteria we obtained from the examination of a variety of cases. Table 12 comes up with the usual statistics we include in the former approaches, but there are two additional columns indicating the orders of the ARMA-GARCH model which was qualified for each contract differences. To be more specific, Table shows our dependent variables in the first column and the next eight columns show the coefficients and the corresponding z-statistics of the constant term and the seven economic variables, then we have the ARMA and the GARCH order we preferred for each commodity future. The last three columns contain the R^2 statistic, the Durbin –Watson statistic for first order serial correlation and the F statistic, respectively.

Moving on to the examination of our results we start with NYMEX Crude Oil. Taking into account the very low or even negative values of the retained adjusted R^2 we should be seriously questioned about the improvement of our models, because their, this way, estimated forecasting power declines. On the other hand, the number of statistically significant regressors remains almost constant and we cannot ignore that all the ARMA terms included are statistically significant at the 5% level. Like in the simple regression case the S&P 500 Index is still significant, but now we have the default risk premium (BAA-AAA) as an additional factor. T-Bill is important only for the sixth shortest to maturity contract. The last disappointing result comes from the fact that all F-statistics ensure the null hypothesis for all coefficients to be equal to zero.

ICE Crude regressions give us worse results. All but one R^2 values are negative declaring that the models are rather mispecified which is proved by the presence of only two statistically significant factors. One of them is derived from the regression of the future expiring in four months, which offers also the only one positive adjusted R^2 value. Again F-statistics suggest zero value for all retained coefficients. So, it is again very ambiguous to be satisfied from the statistical significance of the ARMA terms and the eradication of the ARCH effects.

Table 12: Forecasting Power of the Economic Variables tested by ARMA-GARCH models

Independent Variables			Panel A: NYMEX Crude Oil											
			S&P500DY	BAA-AAA	T-Bill	S&P500	FTSE100	LIBOR	FTSE100DY					
j		C	a1	a2	a3	a4	a5	a6	a7	ARMA	GARCH	R-squared	Durbin-Watson stat	Prob(F-stat)
DCL1	Coefficient	0.030	110.630	-173.075	37.877	-0.004	0.001	37.652	12.324	(6,0)	(1,1)	0.000	2.141	0.473
	z-Statistic	0.971	1.174	-1.495	1.187	-1.546	0.900	1.497	0.142					
DCL2	Coefficient	0.030	80.628	-203.526	41.771	-0.004	0.001	38.596	46.182	(6,0)	(1,1)	0.007	2.107	0.131
	z-Statistic	1.153	1.022	-2.047	1.554	-2.029	1.847	1.836	0.637					
DCL3	Coefficient	0.015	90.055	-172.429	40.574	-0.005	0.001	27.747	36.487	(6,0)	(1,1)	0.009	2.085	0.089
	z-Statistic	0.622	1.216	-1.784	1.563	-2.138	1.920	1.401	0.525					
DCL4	Coefficient	0.001	20.380	-206.147	39.141	-0.003	0.001	26.436	-7.328	(6,0)	(1,1)	0.006	2.074	0.152
	z-Statistic	0.032	0.321	-2.520	1.759	-1.593	1.588	1.630	-0.135					
DCL5	Coefficient	0.012	48.845	-144.645	37.566	-0.004	0.001	20.467	-10.070	(6,0)	(1,1)	0.009	2.059	0.080
	z-Statistic	0.630	0.827	-1.874	1.855	-2.152	1.447	1.428	-0.198					
DCL6	Coefficient	0.008	46.247	-132.150	42.254	-0.004	0.001	13.933	-23.232	(6,0)	(1,1)	0.008	2.055	0.095
	z-Statistic	0.441	0.827	-1.813	2.344	-2.109	1.227	1.062	-0.508					
DCL7	Coefficient	0.012	26.544	-134.350	33.865	-0.003	0.001	14.805	7.721	(6,0)	(1,1)	0.008	2.059	0.099
	z-Statistic	0.673	0.518	-1.985	1.685	-2.153	1.483	1.152	0.158					
DCL8	Coefficient	0.011	25.637	-133.692	33.876	-0.003	0.001	12.922	2.883	(6,0)	(1,1)	0.007	2.059	0.119
	z-Statistic	0.681	0.530	-2.090	1.814	-2.027	1.327	1.050	0.064					
DCL9	Coefficient	0.011	18.721	-127.998	31.254	-0.003	0.001	12.044	-2.124	(6,0)	(1,1)	0.007	2.058	0.124
	z-Statistic	0.683	0.399	-2.044	1.741	-2.177	1.306	1.009	-0.050					
DCL10	Coefficient	0.010	15.654	-123.828	29.547	-0.003	0.000	12.199	-6.896	(6,0)	(1,1)	0.006	2.052	0.148
	z-Statistic	0.659	0.346	-2.039	1.710	-2.166	1.214	1.045	-0.171					
DCL11	Coefficient	0.009	18.545	-119.486	28.539	-0.003	0.000	10.733	-8.454	(6,0)	(1,1)	0.006	2.048	0.164
	z-Statistic	0.649	0.430	-2.035	1.765	-2.047	1.095	0.926	-0.219					
DCL12	Coefficient	0.007	-8.801	-83.121	20.728	-0.003	0.001	11.204	5.851	(4,1)	(1,1)	-0.014	1.848	1.000
	z-Statistic	0.450	-0.189	-1.391	1.217	-1.978	1.432	0.913	0.134					
DCL13	Coefficient	0.007	-8.485	-83.158	18.633	-0.003	0.000	10.233	2.441	(4,1)	(1,1)	-0.017	1.828	0.990
	z-Statistic	0.459	-0.189	-1.439	1.123	-1.976	1.373	0.857	0.058					
DCL14	Coefficient	0.008	17.011	-113.979	17.219	-0.003	0.000	10.759	-8.168	(6,1)	(1,1)	-0.006	1.810	0.848
	z-Statistic	0.718	0.425	-2.129	1.125	-2.002	1.211	0.959	-0.223					
DCL15	Coefficient	0.008	16.081	-111.875	14.306	-0.002	0.000	10.975	-8.340	(6,1)	(1,1)	-0.009	1.795	0.947
	z-Statistic	0.732	0.412	-2.154	0.955	-1.990	1.184	0.996	-0.233					

Table 12: Forecasting Power of the Economic Variables tested by ARMA-GARCH models (Cont'd)

			Panel B: ICE Crude Oil											
Independent Variables			S&P500DY	BAA-AAA	T-Bill	S&P500	FTSE100	LIBOR	FTSE100DY					
j		C	a1	a2	a3	a4	a5	a6	a7	ARMA	GARCH	R-squared	Durbin-Watson stat	Prob(F-stat)
DCO1	Coefficient	0.018	142.695	-174.360	22.533	-0.003	0.001	24.566	58.273	(6,6)	(1,2)	-0.008	2.047	0.925
	z-Statistic	0.664	1.943	-1.827	0.868	-1.330	1.515	1.150	0.776					
DCO2	Coefficient	0.012	84.096	-175.146	28.983	-0.003	0.001	21.424	27.624	(6,6)	(1,2)	-0.006	2.051	0.827
	z-Statistic	0.496	1.223	-1.989	1.167	-1.731	1.383	1.043	0.405					
DCO3	Coefficient	0.017	55.183	-128.292	32.934	-0.003	0.001	20.586	23.809	(6,6)	(1,2)	-0.003	2.050	0.649
	z-Statistic	0.729	0.867	-1.517	1.431	-1.596	1.312	1.109	0.380					
DCO4	Coefficient	0.015	33.563	-92.088	41.105	-0.003	0.001	21.130	20.766	(6,6)	(1,1)	0.000	2.043	0.426
	z-Statistic	0.701	0.566	-1.153	1.959	-1.962	1.503	1.248	0.378					
DCO5	Coefficient	0.016	30.976	-99.115	31.754	-0.003	0.001	14.939	18.769	(6,6)	(1,1)	-0.001	2.044	0.513
	z-Statistic	0.783	0.550	-1.332	1.610	-1.755	1.521	0.927	0.372					
DCO6	Coefficient	0.013	31.603	-98.392	26.935	-0.003	0.001	10.534	10.394	(6,6)	(1,1)	-0.001	2.042	0.532
	z-Statistic	0.716	0.615	-1.472	1.507	-1.854	1.432	0.705	0.227					
DCO7	Coefficient	0.006	-30.275	-83.255	23.653	-0.002	0.000	10.949	16.965	(6,6)	(1,1)	-0.004	2.040	0.699
	z-Statistic	0.314	-0.587	-1.120	1.200	-1.682	1.067	0.578	0.319					
DCO8	Coefficient	0.011	9.968	-95.553	24.674	-0.002	0.000	11.371	5.124	(6,6)	(1,1)	-0.002	2.041	0.590
	z-Statistic	0.636	0.204	-1.526	1.439	-1.744	1.209	0.815	0.121					
DCO9	Coefficient	0.004	-33.026	-77.864	21.826	-0.003	0.001	9.309	30.772	(6,6)	(1,1)	-0.005	2.032	0.766
	z-Statistic	0.248	-0.719	-1.311	1.439	-1.916	1.503	0.764	0.702					

Table 12: Forecasting Power of the Economic Variables tested by ARMA-GARCH models (Cont'd)

Independent Variables			S&P500DY	BAA-AAA	T-Bill	S&P500	FTSE100	LIBOR	FTSE100DY					
			Panel C: NYMEX Heating Oil											
j		C	a1	a2	a3	a4	a5	a6	a7	ARMA	GARCH	R-squared	Durbin-Watson stat	Prob(F-sta)
DHO1	Coefficient	0.014	90.555	-621.628	36.436	-0.013	0.002	52.034	-68.543	(1,1)	(1,1)	0.013	1.969	0.032
	z-Statistic	0.235	0.332	-1.980	0.338	-1.891	1.191	0.755	-0.307					
DHO2	Coefficient	0.023	143.191	-562.289	61.576	-0.012	0.002	50.289	-102.407	(6,6)	(1,1)	-0.005	2.102	0.746
	z-Statistic	0.306	0.606	-2.110	0.719	-1.993	0.803	0.830	-0.507					
DHO3	Coefficient	0.023	230.790	-588.200	79.052	-0.010	0.001	49.888	-60.190	(6,6)	(1,1)	-0.006	2.073	0.854
	z-Statistic	0.330	1.165	-2.503	1.052	-1.735	0.771	0.946	-0.315					
DHO4	Coefficient	0.025	278.329	-684.288	61.560	-0.009	0.002	61.313	-9.994	(6,6)	(1,1)	-0.006	2.058	0.848
	z-Statistic	0.385	1.686	-3.082	0.804	-1.613	0.913	1.245	-0.055					
DHO5	Coefficient	0.026	25.725	-582.932	53.354	-0.007	0.001	36.774	-36.995	(6,6)	(1,1)	-0.006	2.060	0.814
	z-Statistic	0.408	0.158	-2.901	0.809	-1.306	0.538	0.721	-0.205					
DHO6	Coefficient	0.026	-4.552	-492.410	62.378	-0.005	0.000	15.580	-70.631	(6,6)	(1,1)	-0.006	2.070	0.841
	z-Statistic	0.450	-0.029	-2.641	1.045	-1.100	0.255	0.338	-0.430					
DHO7	Coefficient	0.022	-34.719	-382.382	62.648	-0.005	0.000	-5.949	-80.204	(6,6)	(1,1)	-0.006	2.078	0.841
	z-Statistic	0.409	-0.232	-2.085	1.202	-1.118	0.167	-0.154	-0.597					
DHO8	Coefficient	0.032	-85.095	-378.601	44.337	-0.005	0.001	-10.585	1.326	(6,6)	(1,1)	-0.006	2.119	0.831
	z-Statistic	0.594	-0.564	-2.020	0.821	-1.314	0.652	-0.268	0.010					
DHO9	Coefficient	0.025	-84.879	-308.546	38.903	-0.005	0.000	-17.090	-54.515	(6,6)	(1,1)	-0.008	2.152	0.917
	z-Statistic	0.491	-0.574	-1.714	0.760	-1.368	0.312	-0.460	-0.446					
DHO10	Coefficient	0.022	-97.474	-253.065	40.064	-0.005	0.000	-24.096	-65.542	(6,6)	(1,1)	-0.009	2.169	0.966
	z-Statistic	0.449	-0.681	-1.417	0.807	-1.282	0.324	-0.669	-0.563					
DHO11	Coefficient	0.035	58.806	-402.126	15.720	-0.004	0.000	-4.985	-18.589	(1,1)	(3,1)	-0.027	1.788	0.837
	z-Statistic	1.104	0.402	-2.884	0.337	-1.094	0.091	-0.161	-0.140					
DHO12	Coefficient	-0.063	59.015	-238.272	-1.989	0.001	0.000	26.550	-34.101	(0,1)	(3,1)	-0.045	1.686	0.786
	z-Statistic	-2.352	0.439	-2.464	-0.046	0.264	-0.287	0.909	-0.325					

Table 12: Forecasting Power of the Economic Variables tested by ARMA-GARCH models (Cont'd)

			Panel D: NYMEX Gasoline											
Independent Variables			**S&P500DY**	**BAA-AAA**	**T-Bill**	**S&P500**	**FTSE100**	**LIBOR**	**FTSE100DY**					
j		**C**	**a1**	**a2**	**a3**	**a4**	**a5**	**a6**	**a7**	**ARMA**	**GARCH**	**R-squared**	**Durbin-Watson stat**	**Prob(F-stat)**
DHU1	Coefficient	0.119	173.206	398.161	21.923	-0.007	0.003	-10.975	22.240	(6,6)	(1,2)	-0.001	1.967	0.476
	z-Statistic	1.081	0.681	1.335	0.191	-1.060	1.147	-0.109	0.085					
DHU2	Coefficient	0.014	-327.754	75.529	-115.448	-0.012	0.004	-11.930	186.404	(6,6)	(1,1)	0.008	1.954	0.175
	z-Statistic	0.160	-1.356	0.227	-1.201	-2.568	2.606	-0.105	0.816					
DHU3	Coefficient	0.068	-238.659	66.731	-124.531	-0.012	0.004	90.540	193.801	(6,6)	(1,1)	0.001	1.886	0.408
	z-Statistic	0.719	-0.945	0.191	-1.281	-2.581	2.493	0.703	0.739					
DHU4	Coefficient	0.076	-225.491	-148.532	-64.088	-0.009	0.004	41.567	312.163	(6,6)	(1,1)	-0.004	2.018	0.635
	z-Statistic	0.926	-1.025	-0.473	-0.678	-1.855	2.555	0.412	1.336					
DHU5	Coefficient	0.016	-501.264	-156.838	-114.443	-0.002	0.002	-85.263	235.844	(6,6)	(1,1)	-0.010	2.136	0.909
	z-Statistic	0.172	-2.439	-0.542	-1.097	-0.419	1.099	-0.928	1.022					
DHU6	Coefficient	-0.018	-306.452	157.491	-85.909	-0.003	0.002	-5.362	160.293	(6,6)	(1,3)	-0.009	2.104	0.833
	z-Statistic	-0.217	-1.494	0.540	-1.019	-0.725	1.494	-0.062	0.719					
DHU7	Coefficient	0.034	-325.182	56.199	-128.239	-0.002	0.001	-2.682	185.932	(4,4)	(1,1)	-0.008	2.074	0.811
	z-Statistic	0.478	-1.660	0.190	-1.407	-0.564	0.875	-0.025	0.835					
DHU8	Coefficient	0.022	-464.224	251.295	-93.284	-0.005	0.001	-6.972	138.343	(4,4)	(1,1)	-0.003	2.055	0.601
	z-Statistic	0.340	-2.123	0.976	-1.111	-1.297	0.928	-0.072	0.646					
DHU9	Coefficient	0.085	-223.910	331.213	12.395	-0.007	0.003	-28.206	254.006	(6,6)	(1,2)	-0.013	2.051	0.955
	z-Statistic	1.118	-0.886	1.470	0.132	-1.769	2.029	-0.294	1.258					
DHU10	Coefficient	0.059	387.950	33.379	-118.636	0.003	0.001	-77.544	70.705	(6,6)	(1,1)	-0.006	1.977	0.718
	z-Statistic	0.427	1.181	0.106	-0.785	0.407	0.439	-0.586	0.212					
DHU11	Coefficient	-0.041	81.183	-606.835	-139.031	0.007	0.000	-98.471	-311.806	(2,2)	(1,5)	-0.028	2.094	1.000
	z-Statistic	-0.420	0.309	-1.989	-1.242	0.996	-0.163	-1.207	-1.414					

Results from regressing ΔFt^j (j = CL1,…, CL15, CO1,…, CO9, HO1,…, HO12, HU1,…, HU11) on the first differences of the seven economic and financial factors weekly prices using ARMA-GARCH regression models. Sample period: 8/11/1991-29/12/2006

The picture remains disappointing when we look at NYMEX Heating Oil results. The negative R^2 values constitute a general phenomenon again with the only one exception of the first contract. The default risk premium (BAA-AAA) takes the place of S&P 500 and T-Bill as valid explanatory variables in cooperation with the proposed ARMA terms, whereas F-statistic insists to qualify zero coefficients values.

Our last commodity, NYMEX Gasoline, is the only one that the ARMA-GARCH models results are comparable with the previous findings. Unfortunately, this is not because of an impressive performance of this specific approach, but it is due to the same very poor results we had obtained from our previous efforts. We continue to observe negative R^2 values and the F-statistics correspond to zero coefficients values. Moreover, the statistically important factors show extreme instability and their set contains four different variables (S&P 500, S&P 500DY, BAA-AAA and FTSE100).

The aforementioned findings indicate an inadequate model poorly fitting in our data set with no predictive skills. These remarks can be by far worse if we try to get them in a comparison with former literature. Sadorsky derived a similar model using similar explanatory variables and he managed to present some results which indicate forecastability over three different contracts of the shortest maturities in a monthly time horizon. Our research cannot be characterised by equivalent conclusions as far as the weekly time interval is concerned. Of course, we should remind that the negative R^2 values cannot be the reason to exclude a model by themselves, especially in our case because we have to do with ARCH models which can result in negative R^2. Consequently, the final valuation of this model will be done in the out of sample section of this paper, where we can shape more stable opinion about the forecasting performances of all models mentioned before.

CHAPTER 5: VAR MODEL

As a last approach in order to identify predictability across the petroleum futures term structure we take into consideration a Vector Autoregression (VAR) model. Obviously, this approach is an alternative of the two PCA methods giving the fact that we have to do again with a time series forecasting approach which uses as explanatory factors the lags of the dependent variables. To be more specific our model occupies only the first lagged values of the weekly differences of contracts which represent the various maturities of the futures on the same commodity as our dependent variable belongs. So, the VAR equation has the following form:

$$DF_t^l = c^l + \Phi^l DF_{t-1}^l + u_t^l \quad \textbf{(12)}$$

The only substantial differentiation from previous equations, as far as the variables representation is concerned, has to do with the fact that the equation terms consist of vectors and matrices. So,

DF_t^l denotes a (J*1) vector of the weekly changes of futures prices at time t. Indicator j represents as before the different generics (CL1-CL15, CO1-CO9, HO1-HO12, HU1-HU11) of the contracts, whereas l refers to commodities NYMEX Crude Oil, IPE Crude Oil, NYMEX Heating Oil and NYMEX Gasoline, so as to determine four different vectors and the corresponding VAR equations.

Furthermore, Φ^l is the J*J matrix of coefficients of the l commodity to be estimated, c^l and u_t^l are J*1 vectors of constant terms and errors, respectively. Error terms may be contemporaneously correlated but are uncorrelated with their own lagged values and uncorrelated with all of the right-hand side variables.

The results for each commodity are presented in Table 13. The first three commodities come up with results very similar to separate PCA model results. The adjusted R-squared values are very low indicating no forecasting power across the maturities even though t- and F-statistics confirm the null hypothesis for all ICE Crude regressions. NYMEX Crude and Heating Oil futures have statistically significant coefficients, but R^2 values remain lower than 2.5%.

Panel A: NYMEX Crude Oil

	DCL1	DCL2	DCL3	DCL4	DCL5	DCL6	DCL7	DCL8	DCL9	DCL10	DCL11	DCL12	DCL13	DCL14	DCL15
C	0.053	0.054	0.054	0.055	0.055	0.056	0.056	0.057	0.057	0.057	0.058	0.058	0.058	0.058	0.058
	[1.02886]	[1.13653]	[1.22417]	[1.31464]	[1.39701]	[1.46950]	[1.54040]	[1.61148]	[1.66856]	[1.73029]	[1.78949]	[1.83448]	[1.87767]	[1.91837]	[1.95729]
DCL1(-1)	-0.547	-0.216	-0.227	-0.208	-0.189	-0.170	-0.149	-0.131	-0.107	-0.096	-0.085	-0.062	-0.049	-0.036	-0.026
	[-2.63210]	[-1.11652]	[-1.25505]	[-1.22435]	[-1.17468]	[-1.10141]	[-1.00765]	[-0.91939]	[-0.76655]	[-0.71179]	[-0.65022]	[-0.48104]	[-0.39398]	[-0.29608]	[-0.21581]
DCL2(-1)	1.112	0.470	0.649	0.515	0.470	0.459	0.409	0.422	0.333	0.328	0.358	0.302	0.299	0.287	0.271
	[1.30655]	[0.59356]	[0.87727]	[0.73966]	[0.71201]	[0.72713]	[0.67378]	[0.72142]	[0.58461]	[0.59381]	[0.66836]	[0.57544]	[0.58396]	[0.57132]	[0.54838]
DCL3(-1)	-0.340	0.179	-0.140	0.037	-0.218	-0.446	-0.507	-0.725	-0.641	-0.749	-0.945	-0.941	-1.038	-1.124	-1.162
	[-0.16142]	[0.09133]	[-0.07623]	[0.02141]	[-0.13348]	[-0.28551]	[-0.33744]	[-0.50096]	[-0.45518]	[-0.54876]	[-0.71201]	[-0.72477]	[-0.81756]	[-0.90283]	[-0.95016]
DCL4(-1)	-4.749	-5.092	-4.380	-3.919	-2.897	-2.337	-1.980	-1.486	-1.368	-1.079	-0.751	-0.641	-0.444	-0.260	-0.117
	[-1.36016]	[-1.56961]	[-1.44357]	[-1.37273]	[-1.06965]	[-0.90327]	[-0.79558]	[-0.61977]	[-0.58585]	[-0.47668]	[-0.34171]	[-0.29817]	[-0.21135]	[-0.12619]	[-0.05794]
DCL5(-1)	8.607	8.707	7.614	6.841	5.844	5.626	5.150	4.709	4.437	4.160	4.000	3.872	3.723	3.631	3.460
	[1.95875]	[2.13259]	[1.99414]	[1.90423]	[1.71434]	[1.72820]	[1.64450]	[1.56050]	[1.51054]	[1.46116]	[1.44574]	[1.43069]	[1.40662]	[1.39858]	[1.35763]
DCL6(-1)	1.285	1.121	1.452	1.313	1.254	0.783	1.091	0.929	0.977	1.048	0.854	0.907	0.923	0.805	0.842
	[0.26859]	[0.25228]	[0.34933]	[0.33571]	[0.33784]	[0.22099]	[0.32017]	[0.28266]	[0.30542]	[0.33805]	[0.28338]	[0.30768]	[0.32052]	[0.28486]	[0.30331]
DCL7(-1)	-6.580	-6.291	-5.955	-5.469	-5.088	-4.677	-4.788	-4.085	-4.162	-4.229	-3.999	-4.058	-4.003	-3.831	-3.757
	[-1.69096]	[-1.74004]	[-1.76089]	[-1.71921]	[-1.68559]	[-1.62242]	[-1.72664]	[-1.52848]	[-1.59992]	[-1.67722]	[-1.63185]	[-1.69316]	[-1.70817]	[-1.66626]	[-1.66465]
DCL8(-1)	-0.714	-0.863	-1.143	-1.330	-1.431	-1.521	-1.547	-2.084	-1.477	-1.487	-1.567	-1.446	-1.443	-1.435	-1.445
	[-0.31352]	[-0.40775]	[-0.57753]	[-0.71383]	[-0.80971]	[-0.90125]	[-0.95256]	[-1.33198]	[-0.96943]	[-1.00727]	[-1.09259]	[-1.03023]	[-1.05143]	[-1.06649]	[-1.09379]
DCL9(-1)	1.414	1.653	1.752	1.880	1.955	1.935	2.011	2.020	1.442	1.802	1.787	1.729	1.708	1.583	1.524
	[0.48969]	[0.61615]	[0.69831]	[0.79653]	[0.87281]	[0.90490]	[0.97742]	[1.01898]	[0.74703]	[0.96327]	[0.98297]	[0.97258]	[0.98230]	[0.92830]	[0.91016]
DCL10(-1)	3.425	2.935	2.786	2.464	2.198	1.994	1.697	1.607	1.722	1.312	1.474	1.362	1.218	1.190	1.138
	[0.83698]	[0.77198]	[0.78359]	[0.73664]	[0.69247]	[0.65767]	[0.58207]	[0.57195]	[0.62934]	[0.49492]	[0.57220]	[0.54041]	[0.49423]	[0.49206]	[0.47968]
DCL11(-1)	1.159	1.745	2.152	2.458	2.597	2.665	2.729	2.793	2.570	2.528	2.173	2.444	2.419	2.402	2.371
	[0.32686]	[0.52956]	[0.69809]	[0.84769]	[0.94367]	[1.01416]	[1.07961]	[1.14636]	[1.08388]	[1.09982]	[0.97278]	[1.11881]	[1.13249]	[1.14629]	[1.15252]
DCL12(-1)	-8.605	-9.139	-9.882	-10.166	-10.220	-9.979	-9.712	-9.552	-9.036	-8.733	-8.522	-8.688	-8.256	-8.110	-7.939
	[-1.44222]	[-1.64839]	[-1.90595]	[-2.08397]	[-2.20786]	[-2.25742]	[-2.28395]	[-2.33104]	[-2.26534]	[-2.25884]	[-2.26819]	[-2.36419]	[-2.29743]	[-2.30078]	[-2.29385]
DCL13(-1)	-2.105	-1.109	0.180	1.034	1.596	1.792	1.967	2.169	1.970	1.946	2.086	2.083	1.781	1.984	1.955
	[-0.29467]	[-0.16708]	[0.02902]	[0.17697]	[0.28793]	[0.33857]	[0.38628]	[0.44207]	[0.41248]	[0.42035]	[0.46371]	[0.47329]	[0.41376]	[0.47007]	[0.47177]
DCL14(-1)	5.522	4.398	3.584	3.137	2.837	2.730	2.606	2.673	2.479	2.505	2.582	2.449	2.389	2.014	2.096
	[0.74076]	[0.63500]	[0.55323]	[0.51465]	[0.49048]	[0.49422]	[0.49049]	[0.52202]	[0.49743]	[0.51855]	[0.54995]	[0.53333]	[0.53206]	[0.45735]	[0.48479]
DCL15(-1)	1.066	1.461	1.533	1.399	1.290	1.150	1.031	0.753	0.874	0.760	0.572	0.708	0.796	0.925	0.814
	[0.25235]	[0.37210]	[0.41751]	[0.40512]	[0.39346]	[0.36733]	[0.34255]	[0.25952]	[0.30932]	[0.27771]	[0.21513]	[0.27219]	[0.31300]	[0.37056]	[0.33212]
R-squared	0.022	0.018	0.018	0.018	0.017	0.017	0.019	0.018	0.017	0.018	0.017	0.019	0.020	0.020	0.021
F-statistic	2.196	1.953	1.973	1.958	1.932	1.932	1.997	1.952	1.903	1.946	1.915	2.035	2.094	2.094	2.113

Table 13: Forecasting Power of the Vector Autoreggresion model (Cont'd)

Panel B: ICE Crude Oil

	DCO1	DCO2	DCO3	DCO4	DCO5	DCO6	DCO7	DCO8	DCO9
C	0.055	0.057	0.057	0.058	0.058	0.059	0.059	0.060	0.060
	[1.18056]	[1.31168]	[1.40501]	[1.49808]	[1.56791]	[1.63507]	[1.70736]	[1.76745]	[1.83551]
DCO1(-1)	-0.139	0.111	0.088	0.081	0.085	0.080	0.070	0.065	0.056
	[-0.62435]	[0.54215]	[0.45544]	[0.44147]	[0.48169]	[0.46833]	[0.42336]	[0.40158]	[0.35540]
DCO2(-1)	0.069	-0.214	0.188	0.165	0.193	0.200	0.225	0.247	0.262
	[0.11246]	[-0.38056]	[0.35380]	[0.32677]	[0.39733]	[0.42751]	[0.49540]	[0.56077]	[0.61043]
DCO3(-1)	1.410	1.175	0.596	0.909	0.646	0.519	0.449	0.369	0.327
	[1.33512]	[1.20720]	[0.65043]	[1.04320]	[0.77079]	[0.64163]	[0.57212]	[0.48436]	[0.44139]
DCO4(-1)	-0.901	-0.710	-0.516	-0.836	-0.278	-0.225	-0.241	-0.224	-0.254
	[-1.07734]	[-0.92160]	[-0.71080]	[-1.21057]	[-0.41836]	[-0.35202]	[-0.38717]	[-0.37080]	[-0.43296]
DCO5(-1)	-0.843	-0.621	-0.643	-0.610	-1.026	-0.568	-0.653	-0.655	-0.540
	[-0.55731]	[-0.44580]	[-0.49007]	[-0.48904]	[-0.85579]	[-0.49047]	[-0.58169]	[-0.60060]	[-0.50873]
DCO6(-1)	0.651	0.356	0.133	0.000	-0.010	-0.472	0.042	0.005	-0.167
	[0.41484]	[0.24606]	[0.09788]	[-0.00018]	[-0.00778]	[-0.39354]	[0.03634]	[0.00432]	[-0.15154]
DCO7(-1)	-0.702	-0.657	-0.565	-0.540	-0.514	-0.542	-1.018	-0.539	-0.383
	[-0.81785]	[-0.82990]	[-0.75876]	[-0.76241]	[-0.75411]	[-0.82412]	[-1.59551]	[-0.86994]	[-0.63533]
DCO8(-1)	-0.163	-0.045	0.032	0.164	0.249	0.290	0.542	-0.018	0.241
	[-0.13127]	[-0.03960]	[0.02974]	[0.16061]	[0.25331]	[0.30534]	[0.58921]	[-0.02029]	[0.27692]
DCO9(-1)	0.546	0.541	0.640	0.630	0.625	0.694	0.557	0.726	0.433
	[0.62278]	[0.66940]	[0.84151]	[0.87092]	[0.89892]	[1.03370]	[0.85492]	[1.14766]	[0.70317]
R-squared	-0.002	-0.002	0.000	0.005	0.004	0.004	0.005	0.003	0.001
F-statistic	0.782	0.808	1.043	1.415	1.351	1.340	1.441	1.224	1.098

Table 13: Forecasting Power of the Vector Autoreggresion model (Cont'd)

Panel C: NYMEX Heating Oil

	DHO1	DHO2	DHO3	DHO4	DHO5	DHO6	DHO7	DHO8	DHO9	DHO10	DHO11	DHO12
C	0.124	0.131	0.139	0.144	0.149	0.153	0.156	0.161	0.165	0.167	0.167	0.171
	[0.74829]	[0.87360]	[0.98722]	[1.08764]	[1.19113]	[1.28716]	[1.39788]	[1.51762]	[1.63087]	[1.70132]	[1.74176]	[1.83371]
DHO1(-1)	-0.173	0.026	0.082	0.139	0.170	0.189	0.173	0.137	0.104	0.065	0.039	0.025
	[-1.00186]	[0.16370]	[0.55895]	[1.00612]	[1.29614]	[1.52925]	[1.48114]	[1.24246]	[0.98723]	[0.63114]	[0.38470]	[0.25629]
DHO2(-1)	0.517	0.060	0.080	-0.081	-0.165	-0.203	-0.141	-0.032	0.074	0.159	0.167	0.117
	[0.87198]	[0.11119]	[0.15948]	[-0.17018]	[-0.36684]	[-0.47595]	[-0.35029]	[-0.08389]	[0.20405]	[0.45187]	[0.48516]	[0.34842]
DHO3(-1)	-0.980	-0.570	-0.623	-0.341	-0.342	-0.399	-0.501	-0.591	-0.654	-0.646	-0.576	-0.447
	[-0.94090]	[-0.60283]	[-0.70247]	[-0.40774]	[-0.43261]	[-0.53306]	[-0.70994]	[-0.88329]	[-1.02578]	[-1.04522]	[-0.94986]	[-0.75754]
DHO4(-1)	1.513	1.435	1.354	1.012	1.096	1.098	1.046	0.984	0.890	0.718	0.578	0.503
	[1.25277]	[1.30925]	[1.31705]	[1.04225]	[1.19518]	[1.26487]	[1.27985]	[1.26972]	[1.20303]	[1.00101]	[0.82352]	[0.73551]
DHO5(-1)	-1.032	-1.380	-1.283	-1.114	-1.171	-1.000	-0.764	-0.609	-0.449	-0.288	-0.159	-0.189
	[-0.87032]	[-1.28199]	[-1.27183]	[-1.16876]	[-1.30163]	[-1.17391]	[-0.95193]	[-0.80101]	[-0.61797]	[-0.40899]	[-0.23073]	[-0.28204]
DHO6(-1)	-1.144	-0.768	-0.695	-0.511	-0.440	-0.672	-0.870	-0.952	-1.039	-1.019	-1.041	-0.940
	[-0.88829]	[-0.65639]	[-0.63389]	[-0.49379]	[-0.44996]	[-0.72598]	[-0.99749]	[-1.15121]	[-1.31745]	[-1.33309]	[-1.38885]	[-1.28898]
DHO7(-1)	2.588	2.616	2.506	2.250	2.161	2.301	2.335	2.335	2.344	2.183	2.104	1.933
	[1.80433]	[2.00926]	[2.05326]	[1.95170]	[1.98433]	[2.23268]	[2.40481]	[2.53699]	[2.66846]	[2.56339]	[2.52193]	[2.38132]
DHO8(-1)	-1.907	-1.986	-1.991	-1.892	-1.768	-1.733	-1.625	-1.583	-1.503	-1.332	-1.237	-0.991
	[-1.36378]	[-1.56424]	[-1.67299]	[-1.68332]	[-1.66473]	[-1.72480]	[-1.71668]	[-1.76367]	[-1.75468]	[-1.60472]	[-1.52096]	[-1.25161]
DHO9(-1)	1.752	1.492	1.465	1.405	1.257	1.172	1.032	0.910	0.752	0.665	0.625	0.476
	[1.46699]	[1.37637]	[1.44195]	[1.46427]	[1.38648]	[1.36578]	[1.27648]	[1.18780]	[1.02777]	[0.93775]	[0.90045]	[0.70458]
DHO10(-1)	-1.960	-1.700	-1.620	-1.556	-1.475	-1.459	-1.353	-1.189	-1.076	-1.055	-1.058	-1.186
	[-1.64593]	[-1.57288]	[-1.59902]	[-1.62652]	[-1.63226]	[-1.70575]	[-1.67875]	[-1.55601]	[-1.47660]	[-1.49284]	[-1.52877]	[-1.76074]
DHO11(-1)	-0.263	-0.089	-0.018	0.022	0.075	0.139	0.107	0.030	0.034	0.040	0.071	0.360
	[-0.23575]	[-0.08810]	[-0.01880]	[0.02459]	[0.08901]	[0.17282]	[0.14159]	[0.04149]	[0.04925]	[0.06031]	[0.10979]	[0.56902]
DHO12(-1)	1.026	0.793	0.675	0.610	0.552	0.519	0.512	0.507	0.468	0.458	0.445	0.290
	[1.76935]	[1.50761]	[1.36940]	[1.31035]	[1.25375]	[1.24701]	[1.30581]	[1.36264]	[1.31894]	[1.33167]	[1.31898]	[0.88281]
R-squared	0.014	0.013	0.013	0.013	0.014	0.017	0.020	0.021	0.023	0.023	0.022	0.022
F-statistic	1.948	1.893	1.856	1.842	1.932	2.146	2.315	2.412	2.565	2.562	2.447	2.500

Table 13: Forecasting Power of the Vector Autoreggresion model (Cont'd)

Panel D: NYMEX Gasoline

	DHU1	DHU2	DHU3	DHU4	DHU5	DHU6	DHU7	DHU8	DHU9	DHU10	DHU11
C	0.040	0.039	0.094	0.073	0.204	0.178	0.205	0.160	0.158	0.198	0.193
	[0.48199]	[0.48495]	[1.00317]	[0.87848]	[1.06692]	[1.05341]	[1.36855]	[1.20444]	[1.17946]	[1.15834]	[1.17311]
DHU1(-1)	-0.425	-0.313	-0.175	-0.098	-0.214	-0.063	-0.036	-0.063	-0.144	-0.104	-0.362
	[-8.41048]	[-6.34127]	[-3.02236]	[-1.92318]	[-1.82724]	[-0.60335]	[-0.38846]	[-0.76719]	[-1.74505]	[-0.99394]	[-3.57642]
DHU2(-1)	0.198	-0.142	0.039	0.079	-0.262	0.137	-0.262	-0.259	-0.245	-0.239	-0.103
	[2.52294]	[-1.84308]	[0.43836]	[1.00170]	[-1.43447]	[0.84492]	[-1.82883]	[-2.03609]	[-1.90662]	[-1.46131]	[-0.65227]
DHU3(-1)	0.069	0.283	0.086	-0.167	0.283	-0.438	0.235	0.269	0.313	0.318	0.395
	[0.91824]	[3.87383]	[1.00573]	[-2.21871]	[1.63145]	[-2.85068]	[1.72962]	[2.22430]	[2.57101]	[2.04549]	[2.64459]
DHU4(-1)	0.149	0.122	0.016	0.064	0.005	0.222	-0.100	-0.058	-0.016	-0.023	-0.016
	[2.61583]	[2.19866]	[0.24015]	[1.10809]	[0.03877]	[1.89413]	[-0.97046]	[-0.63406]	[-0.16946]	[-0.19070]	[-0.14027]
DHU5(-1)	0.166	0.118	0.068	-0.062	0.056	0.193	0.183	0.136	0.166	0.134	0.193
	[3.39004]	[2.45577]	[1.22098]	[-1.26300]	[0.49163]	[1.90639]	[2.05416]	[1.70793]	[2.06881]	[1.31377]	[1.96377]
DHU6(-1)	-0.270	0.191	0.260	0.293	-0.400	-0.579	-0.571	-0.518	-0.457	-0.337	-0.369
	[-3.41032]	[2.46319]	[2.86962]	[3.66297]	[-2.17208]	[-3.54738]	[-3.95552]	[-4.03372]	[-3.53395]	[-2.04309]	[-2.32846]
DHU7(-1)	0.376	-0.204	0.066	0.117	0.052	-0.190	-0.011	0.130	0.010	0.051	0.389
	[4.59909]	[-2.54857]	[0.70629]	[1.41643]	[0.27618]	[-1.12959]	[-0.07130]	[0.97986]	[0.07554]	[0.30036]	[2.37297]
DHU8(-1)	-0.186	0.041	-0.115	-0.027	0.365	0.504	0.329	-0.002	0.163	0.013	-0.325
	[-1.97416]	[0.44717]	[-1.06417]	[-0.27862]	[1.66288]	[2.59021]	[1.91210]	[-0.01475]	[1.06090]	[0.06780]	[-1.71783]
DHU9(-1)	0.061	0.000	-0.175	-0.085	-0.297	-0.132	-0.098	0.045	-0.106	0.127	0.071
	[0.87895]	[0.00668]	[-2.21872]	[-1.21573]	[-1.85442]	[-0.92920]	[-0.77821]	[0.40271]	[-0.93908]	[0.88383]	[0.51265]
DHU10(-1)	0.325	0.266	0.254	0.149	0.177	0.230	0.199	0.187	0.173	-0.068	0.103
	[8.07064]	[6.74265]	[5.50210]	[3.64975]	[1.89227]	[2.77068]	[2.71298]	[2.85651]	[2.63188]	[-0.81232]	[1.27147]
DHU11(-1)	0.369	0.335	0.323	0.315	0.119	0.040	0.107	0.157	0.171	0.108	-0.091
	[8.93342]	[8.30625]	[6.83903]	[7.57113]	[1.23836]	[0.47242]	[1.42455]	[2.33933]	[2.53948]	[1.26340]	[-1.10225]
R-squared	0.718	0.673	0.576	0.601	0.066	0.116	0.109	0.137	0.127	0.001	0.062
F-statistic	135.966	110.058	73.027	80.931	4.726	7.951	7.513	9.411	8.745	1.056	4.528

Results from regressing ΔFt^{j} (j = CL1,…, CL15, CO1,…, CO9, HO1,…, HO12, HU1,…, HU11) on on each commodity contracts first lagged weekly price difference (VAR(1) model). Sample period: 8/11/1991-29/12/2006

On the other hand we face surprisingly different results considering NYMEX Gasoline. Gasoline futures lagged values seem to have extremely high forecasting power on the four shortest maturities contracts affecting in a quite impressive way R^2 values. These are as high as 70% for the shortest maturity and decline as we move to later expiring contracts, but R^2 values continue to present predictive skills for the most regressions. Running some univariate regressions we discovered particular forecasting power of the longest maturity contract. We have to remind that this specific future had the most missing values in its data set. As a result, the most observations we excluded listwise were due to this contract's illiquidity. Taking that into account and the fact that PCA results were very poor for this commodity, we assume that recent unexpected results were only a coincidence or even a malcalculation during the listwise prices rejection procedure.

To conclude we can take for granted that term structure dynamics of our first three petroleum futures (Crudes and Heating Oil) cannot be forecasted. VAR model provokes some doubts regarding Gasoline contracts, but still there is a great chance of default in these specific results. So, it may be necessary to repeat the research after carefully reconsideration of the data set. At the moment, we will try to declare the general spectrum of our models advantages and drawbacks through testing their out of sample performance.

CHAPTER 6: CONCLUSIONS

Forecasting assets prices is one of the most usual aspects of finance literature. In this study we tried to derive some evidence connected with predictability on petroleum futures. To do that we follow former papers so as to construct our data set and to choose our forecasting models. So, we used as guidelines Chantziara, Skiadopoulos 2006, 2007 and Sadorsky 2002 articles. The main differentiations of the present effort have to do with the separate models we considered and the time horizon we determined for our data observations. To achieve our fist target we focus on two different approaches widely used in literature. For the second one we used weekly price differences instead of daily or monthly that had already ivestigated.

Firstly, motivated by Chantziara Skiadopoulos work we estimate a time series model based on PCA analysis. Of course the method was introduced earlier in papers like Stock and Watson 2002, but our model and the proposed regression settings are similar to Chantziara, Skiadopoulos. So, we performed PCA separately on each one of our commodities and jointly on them. Regarding the retained Principal Components number and behaviour, we come up with findings consistent with former literature. We ensured that PCs can explain the dynamics of petroleum futures term structure so what it was left was to identify if they could forecast its evolution. We estimate multiple regressions with the PCs first lagged values as regressors, but results were rather disappointing staying in line with previous findings for daily oil futures prices. Low R^2 values were a common factor for the majority of the contracts and even though there were some statistically significant PCs, especially in the separate analysis, these were not enough to improve our forecasting results.

The economic (stuctural) models were influenced by Sadorsky and we were hoping to extend his findings in weekly time steps. We have to note that Sadorsky found a predictable pattern across some petroleum futures of the shortest maturities using monthly returns. To implement this goal we introduced some additional economic and financial factors as possible explanatory and forecasting variables. After the definition of our equations we derive two groups of regressions. The first one consisted of a simple multivariate regression model (Economic variable model) which we tried to enhance later by including ARMA and GARCH terms (ARMA-GARCH model). Unfortunately, the results were very poor failing to be comparable with Sadorsky's ARMAX-ARCH model.

Again we come up with particularly low R^2 values, but something that deserves to mention is that evaluations present the ARMA-GARCH models to be worse than simpler ones despite the fact that they describe better the residuals distributions.

As a last model we introduced a Vector Autoregressive model with the first lagged value of the same commodity futures as regressors. This one has the role of a benchmark for our previously obtained findings and in general it proves the poor forecasting power of the most sophisticated of our models. The PCA or economic analysis had only marginal advantages compared to the VAR model. On the contrary, this last model came up with surprisingly good results as far as some Gasoline contracts are concerned. On the other hand, this unique and non stable result provokes some skepticism about the robustness of our evidence. As a matter of fact we cannot be very optimistic for the real forecasting skills of a VAR(1) model.

The findings of this dissertation should be compared with two different categories of the former literature. The PCA results support the already existing studies regarding the PCs and the opinion that the evolution petroleum futures term structure cannot be forecasted. The retained three first PCs explain up to 95% of the total dependent variables variance and their correlation loadings present the same level, slope and curvature characteristics. These are common elements in the commodity futures term structure described by PCA models (see Schwartz and Cortazar 1994, Tomalsky and Hindanov 2002). The poor forecasting power of PCs is another result in line with previous evidence. We have already referred to Chantziara, Skiadopoulos work and there is also a similar result in Cabiddo and Fiorenzani (2004) at least when the research is restricted to the macromovements of the Brent futures curve which are described by the slope, steepness and curvature.

The second category of articles are those which contribute structural models to the bibliography. Derivation of this kind of models is a usual phenomenon across several assets and financial variables and we can note Stock and Watson (2002a, 2002b), Ribeiro and Hodges (2004) or Khan, Khokhor and Simin (2006). On the otherhand, the present paper is directly connected with Sadorsky 2002 article and its theoretical framework. Mainly, we tried to re-estimate this model using different time hotizon and a more convenient set of independent variables. As a result, we reproduced poor results in the first section of this approach and despite the fact we tested a variety of ARMA-GARCH models across all futures maturities we failed to discover forecasting patterns analogous to Sadorsky's conclusions. As we have remarked VAR model wasn't a straight priority of this dissertation so these results were used only for comparative purposes.

To end up with this study we would like to indicate some inefficiencies or matters for further investigation. We examined ARMA-GARCH structures in the economic variables model, but there is still the question whether this kind of errors specification could improve the PCA performance. Of course these models predictability should be tested in even longer time steps, if we take into account former studies on different assets which verify the presence of predictability for longer horizons. The slightly increasing predictive skills of our PCA regressions as we moved to longer maturity contracts is also an evidence towards this specific direction. The economic variable models should be re-estimated using monthly observations so as to conclude if our results are directly contradictive to Sadorsky ones or if the poor fitting of our models is due to the choice of weekly time horizon. Moreover, it would be crucial these models be tested out of sample so as to check their performance using not only the regression outputs but also some typical metrics which could give us the opportunity compare them considering their out of sample predictive evaluation. In addition to this, it could be derived some explanation why the VAR model clearly outperforms the rest as far as the four shortest maturities of NYMEX Gasoline futures are concerned. Finally, it would be crucial to identify the performance of more complex models, like neural networks or other non-linear approaches.

REFERENCES

Alexander, C. O., (2001). Principles of the Skew, *Risk,* 14, 29-32.

Ané, T., and Labidi, C. (2001). Implied Volatility Surfaces and Market Activity over Time, *Journal of Economics and Finance,* 25, 259-275.

Artis, M.,J., Banerjee, A., and Marcellino, M. (2005). Factor Forecasts for the UK, *Journal of Forecasting,* 24, 279-298.

Audet, N., Heiskanen, P., Keppo, J. and Vehviläinen, I. (2002), Modeling Electricity Forward Curve Dynamics in the Nordic Market, Modeling Prices in Competitive Electricity Markets, 251-265, *Wiley Series in Probability and Mathematical Statistics.*

Basilevsky, A., (1994). Statistical Factor Analysis and Related Methods, Theory and Applications, *John Wiley and Sons.*

Baum, C.F., Barkoulas, J. (1996), Time-varying risk premia in the foreign currency futures basis, *Journal of Futures Markets,* 16, 735–755.

Bessembinder, H., Chan, K. (1992), Time-varying risk premia and forecastable returns in futures markets, *Journal of Financial Economics* 32, 169–193.

Bjornson, B., Carter, C.A., 1997. New evidence on agricultural commodity return performance under time-varying risk. *American Journal of Agricultural Economics,* 79, 918–930.

Cabbibo, G., and Fiorenzani, S. (2004). String Theory, *Energy Risk,* March, 68-72

Chantziara, T. and Skiadopoulos, G. (2007), Can the Dynamics of Petroleum Futures be forecasted? Evidence from Major Markets, *Energy Economics,* forthcoming.

Chelley-Steeley P. (2001), Mean Reversion in the Short Horizon Returns of U.K. Portfolios, *Journal of Business Finance & Accounting,* 28(1), (2), 0306-686X

Chinn, M.D., LeBlanc, M., and Coibion, O. (2005). The Predictive Content of Energy Futures: An Update on Petroleum, Natural Gas, Heating Oil, and Gasoline. Working Paper, National Bureau of Economic Research.

Clewlow, L. and Strickland, C. (1999a). Valuing Energy Options in a One Factor Model Fitted to Forward Prices, Working Paper, School of Finance and Economics, University of Technology, Sydney.

Cochrane, J.H. (1999). New Facts in Finance, Economic Perspectives, *Federal Reserve Bank of Chicago,* 23, 36-58.

Conrad, J. and Kaul, G. (1988), Time Variation in expected returns, *Journal of Business,* 61, 409-425.

Cont, R., and Da Fonseca, J. (2002). Dynamics of Implied Volatility Surfaces, *Quantitative Finance,* 2, 45-60.

Cortazar, G. and Schwartz, E. (1994). The Valuation of Commodity-Contingent Claims, *Journal of Derivatives,* 1, 27-39.

Fama, E.F. (1970). Efficient Capital Markets: A Review of Theory and Empirical Work, *Journal of Finance,* 25, 383-417.

Fama, E.F. (1991). Efficient Capital Markets: II, *Journal of Finance,* 46, 1575-1617.

Fengler, M., Härdle, W., and Villa, C. (2003). The Dynamics of Implied Volatilities: A Common Principal Components Approach, *Review of Derivatives Research,* 6, 179-202.

Fleming, J. and Ostdiek, B. (1999). The Impact of Energy Derivatives on the Crude Oil Market, *Energy Economics,* 21, 135-167.

Frachot, A., Jansi, D. and Lacoste, V. (1992). Factor Analysis of the Term Structure: A Probabilistic Approach, Working Paper, Bank of France.

Gencay, R., and Stengos, T. (1998), "Moving Average Rules, Volume and the Predictability of Security Returns With Feedforward Networks," *Journal of Forecasting,* 17, 401414.

Gibson, R. and Schwartz, E. (1990), Stochastic Convenience Yield and the Pricing of Oil Contingent Claims, *Journal of Finance,* 45, 959-976.

Girma, P.B. and Paulson, A.S. (1999), Risk Arbitrage Opportunities in Petroleum Futures Spreads, *Journal of Futures Markets,* 19, 931-955.

Heath, D., Jarrow, R. and Morton, A. (1992), Bond Pricing and the Term Structure of Interest Rates: A New Methodology For Contingent Claims Valuation, *Econometrica,* 60, 77-105.

Hull, J. (2006), Options, Futures and other Derivatives, *Prentice Hall.*

Järvinen, S. (2003). Dynamics of Commodity Forward Curves, Working Paper, Helsinki School of Economics.

Jegadeesh N. (1990), Evidence of Predictable Behavior of Security Returns, *Journal of Finance,* Vol. 45, No. 3, 881-898

Khan, S., Khoker, Z. and Simin, T. (2006), Commodity Prices, Scarcity and Risk Premia, Working Paper, University of Western Ontario

Knez, P. J., Litterman, R., and Scheinkman, J. (1994). Explorations into Factors Explaining Money Market Returns, *Journal of Finance,* 49, 1861-1882.

Lambadiaris, G., Papadopoulou, L., Skiadopoulos, G., and Zoulis, Y. (2003). VAR: History or Simulation?, *Risk,* 3, 123-126.

Lautier, D. (2005a). Term Structure Models of Commodity Prices: A Review, *Journal of Alternative Investments,* 8, 42-64.

Lautier, D. (2005b). A Matter of Principal, *Energy RISK,* November, 58-62.

Lin, S.X., and Tamvakis, M.N. (2001). Spillover Effects in Energy Futures Markets, *Energy Economics,* 23, 43-56.

Litterman, R. and Scheinkman, J. (1991). Common Factors Affecting Bond Returns, *Journal of Fixed Income,* 2, 54-61.

Litzenberger, R.H. and Rabinovitz, N. (1995). Backwardation in Oil Futures Markets: Theory and Empirical Evidence, *Journal of Finance,* 50, 1517-1545.

Miffre, J. (2002), The predictability of futures returns: rational variation in required returns or market inefficiency?, *Applied Financial Economics,* 12, 715-724.

Miltersen, K. and Schwartz, E. (1998). Pricing of Options on Commodity Futures with Stochastic Term Structures of Convenience Yields and Interest Rates, *Journal of Financial and Quantitative Analysis,* 1, 33-59.

Panigirtzoglou, N., and Skiadopoulos, G. (2004). A New Approach to Modeling the Dynamics of Implied Distributions: Theory and Evidence from the S&P 500 Options, *Journal of Banking and Finance,* 28, 1499-1520.

Pesaran, M. H., and Timmermann, A. (1992), A Simple Nonparamet- ric Test of Predictive Performance, *Journal of Business & Economic Statistics,* 10, 461-465.

Pesaran, M. H., and Timmermann, A. (1994), A Generalization of the Nonparametric Henriksson-Merton Test of Market Timing, *Economics Letters,* 44, 1-7.

Pesaran, M. H., and Timmermann, A. (1995), Predictability of Stock Returns: Robustness and Economic Significance, *Journal of Finance,* 50, 1201-1228

Reisman, H. (1991). Movements of the Term Structure of Commodity Futures and Pricing of Commodity Claims, Working Paper, Haifa University.

Qi, M. (1999), Nonlinear Predictability of Stock Returns Using Financial and Economic Variables, *Journal of Business & Economic Statistics,* Vol. 17, No. 4. , 419-429

Ribeiro, D. and Hodges, S.D. (2004). A Two-Factor Model for Commodity Prices and Futures Valuations, Working Paper, Financial Options Research Centre, Warwick Business School.

Ribeiro, D. and Hodges, S.D. (2005). A Contango-Constrained Model for Storable Commodity Prices, *Journal of Futures Markets,* 25, 1025-1044.

Sadorsky, P. (2002). Time-Varying Risk Premiums in Petroleum Futures Prices, *Energy Economics,* 24, 539-556.

Schwartz, E. and Smith, J. (2000). Short-Term Variations and Long-Term Dynamics in Commodity Prices, *Management Science,* 7, 893-911.

Schwartz, E. (1997). The Stochastic Behavior of Commodity Prices: Implications for Valuation and Hedging, *Journal of Finance,* 3, 923-973.

Skiadopoulos, G., Hodges, S.D., and Clewlow, L. (1999). The Dynamics of the S&P 500 Implied Volatility Surface, *Review of Derivatives Research,* 3, 263-282.

Stock, J.H. and Watson, M.W. (2002a), Forecasting using Principal Components from a Large Number of Predictors, *Journal of the American Statistical Association,* 97, 1167-1179.

Stock, J.H. and Watson, M.W. (2002b). Macroeconomic Forecasting Using Diffusion Indexes, *Journal of Business and Economic Statistics,* 20, 147-162.

Taylor, S. J (1980), Conjectured Models for Trends in Financial Prices, Tests and Forecasts, *Journal of the Royal Statistical Society.* Series A (General), Vol. 143, No. 3.

Tolmasky, C. and Hindanov, D. (2002). Principal Components Analysis for Correlated Curves and Seasonal Commodities: The Case of the Petroleum Market, *Journal of Futures Markets,* 11, 1019-1035.

Wiener, R.J. (2004). Do Birds of a Feather Flock Together? Speculator herding in Derivatives Markets, Working Paper, G. Washington University.