Can the Dynamics of the Term Structure of Petroleum Futures be forecasted? Evidence from Major Markets*

Thalia Chantziara[a], George Skiadopoulos[b]

First Draft: 1/10/2005 – This Draft: 6/12/2005

Abstract

We investigate whether the evolution of the term structure of petroleum futures can be forecasted. To this end, the principal components analysis is employed. The retained principal components describe the dynamics of the term structure of futures prices parsimoniously and are used to forecast the subsequent daily changes of futures prices. Data on the New York Mercantile Exchange (NYMEX) crude oil, heating oil, gasoline, and the International Petroleum Exchange (IPE) crude oil futures are used. We find that the retained principal components have small forecasting power. Similar results are also obtained from standard univariate and vector autoregression models. Spillover effects between the four petroleum futures markets are also detected.

Keywords: Petroleum futures, Principal Components Analysis, Spillovers, Term Structure of futures prices.

JEL Classification: G10, G13, G14.

---

* We would like to thank Stefano Fiorenzani, Jeff Fleming, Daniel Giamouridis, Costas Milas, Leonardo Nogueira, and the participants at the 2nd Advances in Financial Forecasting Conference (Loutraki) and especially the discussant Orestis Soldatos for helpful discussions and comments. Any remaining errors are our responsibility alone.
[a] New York University, chantziara@hotmail.com
[b] Corresponding author. University of Piraeus, Department of Banking and Financial Management, and Financial Options Research Centre, Warwick Business School, University of Warwick, gskiado@unipi.gr

1. Introduction

Since their introduction, futures on petroleum products (crude oil and its by-products heating oil and gasoline) are gaining in importance because they have been designed to serve the oil industry's needs. In particular, the whole term structure of petroleum futures prices is of importance to practitioners. Petroleum futures are traded across a wide range of maturities; different maturities may be used for different purposes by investors (see Lautier, 2005, and the references therein). Furthermore, the petroleum term structure of futures prices evolves stochastically over time. It is typically characterised by alternating backwardation and contango states and high volatility. This attracts speculators and makes the hedging of these contracts a challenging task. Therefore, forecasting the evolution of the whole term-structure of futures is of great interest to the market participants. The previous literature has explored the predictive power of petroleum futures prices (see e.g., Chinn et *al*., 2005, and the references therein), the formation of the shape of the petroleum futures term structure (see e.g., Litzenberger and Rabinovitz, 1995), and its dynamics in the context of pricing petroleum derivatives[1]. Surprisingly, to the best of our knowledge, the question of whether the evolution of the petroleum futures term structure can be forecasted has received little attention. This paper fills this void.

In any context where forecasting needs to be performed, the primary question is what variables should be used as predictors in the forecasting regression equation. One approach would be to specify specific variables that have some clear economic interpretation. For the purposes of forecasting the dynamics of commodity futures prices, possible choices could be the underlying spot price, the interest rate, and the convenience yield. Alternatively, the previous day futures term structure could be employed. The former and latter choice of variables sets up tests of semi-strong and weak form market

[1] Two approaches have been developed to model the dynamics of the term structure of futures prices and price commodity derivatives that depend on the futures price (see Lautier, 2005, for an extensive survey). The first approach assumes that a number of factors (e.g., the underlying spot price, the convenience yield, the interest rate, the long term futures price) affect the futures price. An assumption is made about the process that governs their dynamics. Then, Itô's lemma is used to derive the dynamics of the futures price and the pricing model is built (see e.g., Gibson and Schwartz, 1990, Schwartz, 1997, Schwartz and Smith, 2000, Ribeiro and Hodges, 2004, 2005). However, most of the assumed factors are not observable. The second approach takes the current term structure as given and prices derivatives consistently with it (see e.g., Reisman, 1991, Cortazar and Schwartz, 1994, Clewlow and Strickland, 1999a, 1999b, Tolmasky and Hindanov, 2002). The latter approach is analogous to the Heath et *al*. (1992) methodology in the interest rate literature.

efficiency (Fama, 1970, 1991). However, in the case where one would not want to restrict himself by making a priori assumptions about the forecasting variables, an alternative and more general approach would be to let the data decide on the forecasting variables to be used. Stock and Watson (2002a) have shown that Principal Components Analysis (PCA) can be employed to this end. The principal components can be used as predictors in a linear regression equation since they are proven to be consistent estimators of the true latent factors under quite general conditions (see also Stock and Watson, 2002b, and Artis et *al*., 2005, among others, for empirical applications of this idea to macroeconomic variables). Moreover, the forecast constructed from the principal components is shown to converge to the forecast that would be obtained in the case where the latent factors were known. In our context, PCA can describe the dynamics of the term structure of futures prices parsimoniously by means of a small number of factors. Despite the fact that these factors may not have a clear economic interpretation, they contain all the information about the "hidden" variables that drive the dynamics of the futures term structure and hence they can be used as predictors.

Up to date, PCA has not been used in the finance literature in a forecasting context, as far as we are concerned. In particular, in the commodity futures literature, PCA has been used in most of the studies to investigate the dynamics of the term structures of commodity futures empirically for the purposes of pricing commodity derivatives[2]. Among others, Cortazar and Schwartz (1994) performed PCA on the term structure of copper futures over the period 1978-1990. Clewlow and Strickland (1999b) applied PCA to oil and gas futures traded in NYMEX over the period 1995-1997. Tolmasky and Hindanov (2002) applied PCA to crude oil and heating oil over the period 1983-2000. Järvinen (2003) has also applied PCA to Brent crude oil and pulp over the periods 1997-2002 and 1998-2001, respectively; the forward curve is estimated from the par swap quotes rather than taken directly from the futures market. All these studies have found that

[2] In general, PCA has been used in the option pricing and risk management literature, extensively, to model the dynamics of the variable under consideration. For instance, it has been used in the interest rate literature to explore the dynamics of the yield curve and to provide alternative hedging schemes to the traditional duration hedge (see among others, Litterman and Scheinkman, 1991, Knez et *al*., 1994). Kamal and Derman (1997), Skiadopoulos et *al*. (1999), Alexander (2001a), Ané and Labidi (2001), Fengler et *al*. (2003), and Cont and da Fonseca (2002) have applied PCA to investigate the dynamics of implied volatilities. Panigirtzoglou and Skiadopoulos (2004) have applied PCA to characterize the dynamics of implied distributions. Lambadiaris et *al*. (2003) have employed PCA to calculate the Value-at-Risk of fixed income portfolios (see also Alexander, 2001b, for an extensive description of the applications of PCA in Finance).

three factors govern the dynamics of the term structure of commodity futures. Following the terminology introduced by Litterman and Scheinkman (1991), the first three factors are interpreted as level, steepness, and curvature, respectively; the second and third factor change the term structure from backwardation to contango and vice versa (regime changes). The authors have suggested that in principle the PCA results can be used in option pricing and risk management applications. On the other hand, little attention has been paid to whether the proposed PCA models can be used to forecast the next day's futures term structure. To the best of our knowledge, the study by Cabibbo and Fiorenzani (2004) is the only one that has explored whether a PCA model can forecast the evolution of the Brent futures term structure in the International Petroleum Exchange (IPE) over the period 15/04/94 to 04/08/03. They approximated the term structure in terms of its level, steepness and curvature factors and they checked whether these characteristics can be forecasted by the retained factors. They found that the dynamics of the IPE futures term structure cannot be forecasted. However, they accept that their approach studies only "the macromovements (regime changes) without considering all those micro movements that can affect the string in the short run without inducing necessarily a regime switch".

This paper extends the study of Cabibo and Fiorenzani (2004) by investigating the predictability of the dynamics of the petroleum term structure of futures prices per se rather than forecasting its driving factors. To this end, four major futures markets are examined over the period 1993-2003: the New York Mercantile Exchange (NYMEX) futures traded on the WTI crude oil, heating oil and unleaded gasoline, and the IPE Brent crude oil futures. First, we apply PCA to each one of the four commodities separately, as well as jointly. The joint application of PCA allows incorporating any additional information stemming from any interactions between the four markets (see Tolamsky and Hindanov, 2002, for a similar approach). Then, for each commodity, forecasting regressions are performed. The term structure of futures prices is regressed on the retained from each commodity principal components that are used as explanatory variables; the factors are measured on the previous day and they can be regarded as the shocks that move the term structure of petroleum futures prices over time. In addition, the way that the forecasting regressions are set up allows detecting any possible spillover effects between the four petroleum markets. Tamvakis and Lin (2001) had examined the presence of spillover effects between the NYMEX and IPE crude oil futures markets over the period 1994-1997. Finally, the PCA results are compared with those obtained from standard vector and univariate autoregressions.

The paper is structured as follows. Section 2 describes the data set. Section 3 describes the PCA and discusses the results from the separate and joint PCA. A check of the robustness of the results is also carried out. Section 4 examines the forecasting power of principal components across commodities. Section 5 reports indicative results on the forecasting power of univariate and vector autoregressions. Section 6 concludes and presents the implications of this study.

2. The Data Set

We have obtained daily settlement futures prices on the West Texas Intermediate (WTI) crude oil, heating oil, and gasoline futures trading on NYMEX and the Brent crude oil futures trading on the IPE from Bloomberg (ticker names CL, HO, HU, and CO, respectively).

The NYMEX light sweet (low sulfur) crude oil futures contract is the world's most heavily traded commodity futures contract. It has been trading since 1983. Each futures contract is written on 1,000 barrels of crude oil. On any given day, there are contracts trading for the next 30 consecutive months as well as contracts for delivery in 36, 48, 60, 72, and 84 months (35 futures contracts in total). The delivery period is a full month, meaning that deliveries must be initiated on or after the first calendar day and completed on or before the last calendar day of the delivery month. Trading terminates at the close of the third business day prior to the 25th calendar day of the month preceding the delivery month. Settlement is done with physical delivery, even though most of the contracts are closed before expiration. The underlying asset can be thought to be the WTI that serves as the reference for most crude oil transactions. However, a number of other grades of crude are also deliverable[3]. The delivery point is Cushing, Oklahoma.

[3] Deliverable US crudes are crudes with a sulfur content of 0.42% by weight (or less) and an American Petroleum Institute (API) gravity between 37^{o} and 42^{o}. Deliverable streams are the WTI, Low Sweet Mix, New Mexico Sweet, North Texas Sweet, Oklahoma Sweet, and South Texas Sweet. Deliverable non-US crudes are crudes with an API gravity between 34^{o} and 42^{o}. Deliverable streams are the UK's Brent and Forties and Norway's Oseberg Blend at a discount of \$0.30 per barrel, Nigeria's Bonny Light and Colombia's Cusiana at a premium of \$0.15 per barrel, and Nigerian Qua Iboe at a premium of \$0.05 per barrel.

The IPE in London is the second most liquid crude oil market in the world. The Brent Crude futures contract has been trading on the IPE since 1988. It is part of the Brent blend complex (that also consists of the physical and forward Brent) that is used as a basis for pricing the two thirds of the world's traded crude oil. Each futures contract is 1,000 barrels of Brent crude oil. There are contracts trading for the next twelve consecutive months, then quarterly out to a maximum 24 months, and then half-yearly out to a maximum 36 months (eighteen futures contracts in total). Trading terminates at the close of the business day immediately preceding the 15th day prior to the first day of the delivery month. Settlement is done with physical delivery or alternatively there is the option to settle in cash against the IPE Brent Index price of the day following the last trading day of the futures contract[4]. The underlying asset is the pipeline-exported Brent blend supplied at the Sullom Voe terminal in the North Sea. The prices of the NYMEX and IPE contracts are quoted in US dollars and cents per barrel and are used as benchmarks for pricing crude oil and its refined products on an international basis.

Gasoline and heating oil (also known as No. 2 fuel oil) are two most important refined products, accounting for approximately 40% and 25% of the yield of a crude oil barrel respectively. Both heating oil and gasoline futures trade in NYMEX in contracts of 42,000 US gallons (equivalent to 1,000 barrels). Prices are quoted in US dollars and cents per gallon. There exist contracts for the next 18 consecutive months for heating oil and the next 12 consecutive months for gasoline. The delivery period begins on the day after the fifth business day of the delivery month and ends on the last business day of the delivery month. Trading terminates at the close of business on the last business day of the month preceding the delivery month. Settlement is done with physical delivery. The grade and quality of the deliverable heating oil and gasoline conform to industry standards for fungible No. 2 heating oil, and for Phase II Complex Model Reformulated Gasoline in accordance with Colonial Pipeline Co. specifications for fungible A grade, 87 octane index gasoline, respectively. The three NYMEX petroleum futures contracts are traded by open outcry from 10:05am until 2:30pm New York time. The IPE contract is traded by open outcry from 10:02am until 7:30pm London time (5:02am until 2:30pm New York time).

---

[4] The IPE Brent Index 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. The IPE Index is

Bloomberg provides daily data on the above petroleum futures contracts for any maturity. It also rolls over contracts to construct generic series that contain the contracts that fall within a certain range of days-to-maturity. For example, the first generic *CL1* is the shortest maturity futures contract traded on NYMEX at any point in time, the second generic *CL2* is the second shortest maturity futures contract traded on NYMEX at any point in time, etc. In particular, there are 35 generics for crude oil futures traded on NYMEX (labeled *CL1*-*CL35*), 18 generics for crude oil futures traded on the IPE (labeled *CO1*-*CO18*), 18 generics for heating oil futures traded on NYMEX (labeled *HO1*-*HO18*), and twelve generics for gasoline futures traded on NYMEX (labeled *HU1*-*HU12*). For the purposes of this study, we have used the Bloomberg's generic contracts. We have chosen the generics to roll to the next contract month seven days prior to expiration so as to avoid noise in prices due to increased trading activity. Trading in petroleum futures increases significantly a few days prior to maturity; this results in increased volatility and price spikes.

However, liquidity considerations make possible the use of only a subset of the original data set in terms of the number of generic contracts and the time period. Therefore, we have only used *CL1*-*CL9*, *CO1*-*CO7*, *HO1*-*HO9*, and *HU1*-*HU7* that have satisfactory liquidity and hence their prices are likely to reflect the market dynamics; long-dated contracts are relatively illiquid. Furthermore, despite the fact that crude oil futures have been trading on NYMEX since 1983, data are limited; there is no open interest or volume data from May 30, 1983 to June 30, 1986 and from January 1, 1987 to July 31, 1989. In addition, trading in longer maturity futures did not become available until several years later. Similarly, the data were scarce in the case of Brent contracts on the IPE as well as for heating oil and gasoline contracts on NYMEX until the early '90s. Therefore, we have decided to use data from 1/1/1993 to 31/12/2003. To eliminate further problems arising from thin trading, we have excluded quotes for contracts that have daily volume less than ten contracts.

Table 1 shows the summary statistics of the daily changes of futures prices for each maturity; the results are reported for each one of the four commodity futures under scrutiny. Excluded data correspond either to days where data was unavailable (e.g., public holidays) or to days that were omitted because of the ten-contract volume constraint.

---

issued by the IPE on a daily basis at 12:00 noon London time.

Notice that excluded data account for only about 7-10% of total for the nearest contracts but as much as 14% for *ΔCL9*, 32% for *ΔCO7*, 27% for *ΔHO9*, and 42% for *ΔHU7*. Application of the Jarque-Bera test showed that the series are not normally distributed. We can see that for each commodity, the volatility of the daily changes of futures prices decreases as we move to longer maturities; this has been termed "Samuelson effect" (Samuelson, 1965).

(INSERT TABLE 1 HERE)

3. Principal Components Analysis

In this Section, first we describe the Principal Components Analysis (PCA). Then, we apply PCA to the daily change of the term structure of futures prices for each commodity separately (separate PCA). Next, PCA is applied to the daily change of the term structure of futures prices by grouping all four commodities (joint PCA).

3.1 Description

PCA is used to explain the systematic behavior of observed variables, by means of a smaller set of unobserved latent random variables. Its purpose is to transform p correlated variables to an orthogonal set which reproduces the original variance-covariance structure (or correlation matrix). In this paper, we apply PCA to decompose the correlation structure of the first differences of petroleum futures prices. To achieve this, for any given underlying commodity, we measure the daily differences of petroleum futures prices across different times-to-maturity. For example, within the IPE crude oil contract, *ΔCO1* provides a time series of the first differences of the futures prices that correspond to the nearest maturity contract.

In general, denote time by $t=1,\ldots,T$ and let p be the number of variables. Such a variable is a $(T\times 1)$ vector $\boldsymbol{x}$. The purpose of the PCA is to construct p artificial variables (Principal Components - PCs hereafter) as linear combinations of the $\mathbf{x}$ vectors orthogonal to each other, which reproduce the original variance-covariance structure. The first PC is constructed to explain as much of the variance of the original p variables, as possible (maximization problem). The second PC is constructed to explain as much of the remaining variance as possible, under the additional condition that it is uncorrelated with the first one, and so on. The coefficients with which these linear combinations are formed are called the loadings. In matrix notation

$$Z = XA \tag{1}$$

where $\boldsymbol{X}$ is a ($T \times p$) matrix, $\boldsymbol{Z}$ is a ($T \times p$) matrix of PCs, and $\boldsymbol{A}$ is a ($p \times p$) matrix of loadings. The first order condition of this maximization problem results to

$$(X'X - lI)A = 0 \tag{2}$$

where l_i are the Lagrange multipliers and I is a ($p \times p$) identity matrix. Equation (2) shows that the PCA is simply the calculation of the eigenvalues l_i, and the eigenvectors A of the variance-covariance matrix $\boldsymbol{S=X'X}$. Furthermore, the variance of the *ith* PC is given by the *ith* eigenvalue, and the sum of the variances of the PCs equals the sum of the variances of the $\boldsymbol{X}$ variables.

In the case that the p variables are measured in different units, or they have unequal variances, PCA should be performed on standardized variables. This is equivalent to using the correlation matrix (instead of the variance-covariance matrix). When both variables and components are standardized to unit length, the elements of A' are correlations between the variables and PCs; they are called correlation loadings (Basilevsky, 1994).

It is often the case that a few principal components account for a large part of the total variance of the original variables. In such a case one may omit the remaining components. The result is a substantial reduction of the dimension of the problem. If we retain $r<p$ PCs then

$$X = Z_{(r)}A'_{(r)} + \varepsilon_{(r)} \tag{3}$$

where $\varepsilon_{(r)}$ is a ($T \times p$) matrix of residuals and the other matrices are defined as before having r rather than p columns. The percentage of variance of $\boldsymbol{x}$ that is explained by the retained PCs (communality of $\boldsymbol{x}$) is calculated from the correlation loadings. The concept of "communality" is analogous to that of determination coefficient in a linear regression set-up. After retaining $r<p$ components, we use equation (3) to examine the size of the communalities, and the meaning of the retained components. The interpretation of the PCs is revealed by the correlation loadings that show how each component affects ("loads on") each variable.

There is not a unanimous way on deciding on the number of components to retain. It is common practice to use a variety of rules of thumb, e.g. keep the components that explain 90% of the total variance. However, these are ad-hoc rules with no statistical theory underlying them. There are statistical tests to determine the number of PCs to be retained. Their limitation is that they are parametric being based on the assumption that the original *x* variables follow a multivariate normal distribution. However, multivariate normality does not hold in our case. Some non-parametric criteria have also been suggested (e.g. Velicer's criterion, bootstrapping), but their accuracy is questionable (see e.g., Jackson, 1991, Basilevsky, 1994). The final decision for the number of components to retain is a result of considering the employed formal/informal rule, the interpretation of the components, and the explained communalities.

3.2 Separate PCA: Results and Discussion

We perform PCA on the block of futures series for each of the four commodities under examination. Frachot et *al*. (1992) have shown that PCA yields reliable results in the case where it is applied to stationary series. Hence, we tested for stationarity by applying the Augmented Dickey-Fuller (ADF) test (four lagged terms were employed) to the daily settlement prices of the generic series *CL1-CL9*, *CO1-CO7*, *HO1-HO9*, and *HU1-HU7*. We found that the series were non-stationary while their first differences were stationary. Therefore, PCA will be applied to the first differences *ΔCL1-ΔCL9*, *ΔCO1-ΔCO7*, *ΔHO1-ΔHO9*, and *ΔHU1-ΔHU7* of the original series. In the case where there were missing values for any one variable at any one date, the data were excluded listwise.

Table 2 (Panel A) shows the cumulative percentage of variance explained by all PCs for each one of the four commodities. We can see that the first three PCs explain 96%-99% of the variance of the changes in futures prices across the four commodities; the percentage of variance explained by the first three PCs is smallest in the case of gasoline since the pairwise correlations (not reported) between the futures expiries is slightly smaller compared to those for the other three commodities. The fourth PC increases the amount of explained variance marginally.

(INSERT TABLE 2 HERE)

Table 3 (Panel A) shows the descriptive statistics of the first three standardized PCs for each one of the four commodities. Application of the Jarque-Bera test shows that they are non-normally distributed; this implies that the stochastic process that drives the term structure of commodity futures is not normally distributed. The number of observations is sufficient in order to obtain reliable results from the PCA; it ranges from 1,451 - 2,353 depending on the commodity.

(INSERT TABLE 3 HERE)

Figure 1 plots the correlation loadings of the first three PCs for each one of the four commodities. The interpretation of the PCs is the same across commodities. We can see that the first PC affects the term structure of futures prices by the same amount. Hence, it can be interpreted as a parallel shift. The second PC moves the shortest expiries to a different direction from the longer expiries and hence it can be interpreted as a slope factor. The third PC can be interpreted as a curvature factor: it causes prices of short-maturity and long-maturity futures to move in the same direction and prices of mid-maturity futures to move in the opposite direction. The third PC is steeper for the short expiries than for the long ones (see also Tolmasky and Hindanov, 2002, for a similar finding). The second and third PCs change the term structure from contango to backwardation and vice versa. The communalities of the first three PCs range from 93%-99% depending on the commodity and the futures series. The fourth PC does not have a clear interpretation and it can be regarded as noise; hence it is not shown here. The correlation loadings of the first three PCs have similar values across commodities. Our results on the number of retained PCs, the amount of the variance that they explain, and their interpretation is in general in line with the previous related literature on the dynamics of the term structures of commodity futures (see e.g., Cortazar and Schwartz, 1994, Clewlow and Strickland, 1999b, and Tolmasky and Hindanov, 2002)[5].

(INSERT FIGURE 1 HERE)

[5] The study by Järvinen (2003) provides different results in the PCA of the term structure of the IPE crude oil futures contracts. He used Brent crude oil swap quotes from 1997 to 2002 to derive the futures curve. He concluded that the first three principal components explain 89% of total variance. The interpretation of the first two PCs was also different. The first factor sloped upwards for maturities of up to 21 months before flattening out and even had an opposite sign for three-month and six-month maturities. The second factor showed a more complex behavior, representing shocks that move contracts with maturities of up to 21 months in one direction and then contracts with longer maturities in the other direction, albeit with a curvature in the middle.

3.3 Joint PCA: Results and Discussion

We perform PCA on the changes of futures prices across maturities for all four commodities simultaneously (joint PCA, see also Tolmasky and Hindanov, 2002, for a similar approach). Hence, the derived PCs explain the joint evolution of the term structure of all four commodities ($\Delta CL1$-$\Delta CL9$, $\Delta CO1$-$\Delta CO7$, $\Delta HO1$-$\Delta HO9$, and $\Delta HU1$-$\Delta HU7$). Table 2 (Panel B) shows the cumulative percentage of variance explained by the first three joint principal components. The first three joint PCs explain a slightly smaller amount of the total variance compared to the one explained by the PCs obtained from the separate PCA (93% compared to 96%-99%). Table 3 (Panel B) shows the summary statistics of the first three joint standardized PCs. We can see that they are non-normally distributed, as was the case with the PCs obtained from the separate PCA.

Figure 2 plots the loadings of the first three components. The first PC can be interpreted as a parallel shift (this is not that clear in the case of gasoline though) as in the case of the first PC obtained from the separate PCA. However, the interpretation of the second and third PC has changed now. The second PC cannot be interpreted as slope any longer. Instead, it has a level characterization for the NYMEX and IPE contracts. This is less evident for the heating oil contracts while it is downward sloping for the gasoline ones. Interestingly, the second joint PC moves the term structure of the crude oil (NYMEX and IPE) contracts to different direction from the heating oil and gasoline ones. The third PC does not have the curvature interpretation any longer that was attributed to it in the separate PCA. It moves the crude oil contracts to the opposite direction of the heating oil contracts while it slopes upwards in the case of gasoline. The fourth PC had a noisy behavior and hence it is not reported here. Tolmasky and Hindanov (2002) had also found that the joint PCA might yield PCs that do not have the same interpretation with the ones obtained from the PCA applied to each commodity, separately.

(INSERT FIGURE 2 HERE)

3.4 Stability of the PCA results

For the purposes of our subsequent analysis we need to check whether the PCA results are stable over time. Figure 3 shows the evolution of the WTI crude oil price over the period 1/1/1993 to 31/12/2003. We can see that the crude oil prices (and hence the futures prices) have fluctuated widely over the period under examination depending on supply and demand conditions as well as on global political and economic events (similar graphs are obtained for the other three commodities). Therefore, we performed PCA on the four commodities by breaking up the period in two sub-periods: 1/1/1993 to 13/5/1997 and 14/5/1997 to 31/12/2003, the cutoff point being the day we can identify as the beginning of the Asian crisis.[6] The Asian crisis led to stagnant oil demand. This in conjunction with the increased production in the Middle East caused oil prices to plummet. It may be the case that the Asian crisis has created a structural break in the data generating process. To check this, PCA was performed on each commodity separately, as well as jointly on the four commodities, within each sub-period.

(INSERT FIGURE 3 HERE)

We find that the results obtained from PCA for each of the two sub-periods are not significantly different than the results we obtained from the analysis on the full sample. This makes us confident about the stability of the PCA results. Therefore, in the remaining of the paper we will use the PCA results obtained in the full period from 1/1/1993 to 31/12/2003.

4. PCA and Forecasting Power

In this Section, we use the PCA results to examine whether the movements of the term structure of futures prices can be forecasted. To this end, a multiple regression setup is employed. For any given commodity and maturity, two alternative approaches are taken. First, the changes of the futures prices are regressed on the twelve retained PCs (three for each commodity) obtained from the separate PCA in Section 3.1. Next, the changes of the futures prices are regressed on the three retained PCs obtained from the joint PCA in Section 3.2.

4.1 Separate PCA: The regression setup & Results

Let ΔF_t^j be the daily changes of futures prices measured at time *t* for any generic contract (maturity) j = *CL1*,…, *CL9*, *CO1*,…, *CO7*, *HO1*,…, *HO9*, *HU1*,…, *HU7*. We regress

[6] On May 14, 1997, the Thai bhat depreciated dramatically. This was due to the fact that the country's economic slowdown and political instability urged speculators to proceed to massive sell orders.

ΔF_t^j on the three retained principal components PC$_k$ (k=1,2,3) measured at time t-1[7]. To fix ideas, the following regression is estimated

$$\Delta F_t^j = c + \sum_{k=1}^{3} a_k CLPC_{k,t-1} + \sum_{k=1}^{3} b_k COPC_{k,t-1} + \sum_{k=1}^{3} c_k HOPC_{k,t-1} + \sum_{k=1}^{3} d_k HUPC_{k,t-1} + u_t \quad (4)$$

where $CLPC_{k,t-1}, COPC_{k,t-1}, HOPC_{k,t-1}, HUPC_{k,t-1}$ are the time series of the k retained PCs extracted from the PCA on the NYMEX crude oil, IPE crude oil, heating oil, and gasoline futures contracts, respectively.

There are two advantages of using the PCs rather than alternative ad hoc variables for the purposes of forecasting. The first is that the PCs summarize the dynamics of the term structure of futures prices. Hence, the forecasting information in any alternative variables would be a subset of the information contained in the PCs. The second advantage is that the use of the PCs allows checking for spillover effects across commodities. A general-to-specific approach is used. We start off with all 12 principal components (three per commodity) as regressors and we drop the ones that are not statistically significant at the 5% significance level.

Table 4 shows the results of the regressions for each one of the four commodities. The first column shows the dependent variable in equation (4). The next 13 columns show the estimated constant term and the estimated coefficients of the regressors along with their t-statistics in parentheses. The t-statistics are calculated by the Newey-West standard errors so as to correct for the detected heteroscedasticity and autocorrelation. The following column shows the R^2 statistic. Finally, the last column shows the F-statistic that tests the null hypothesis that all coefficients (excluding the constant term) are zero. The F-statistic's p-values are shown in parentheses.

(INSERT TABLE 4 HERE)

We can see that in the case of the NYMEX crude oil contract, the changes of the futures prices can be forecasted only for the three intermediate maturities (*CL3*, *CL4* and *CL5*) by the third PC of the IPE crude oil futures; the PCs of the NYMEX crude oil have no forecasting power themselves. The estimated parameters and the R^2 value are very small though (0.004). In the case of the IPE crude oil, the pattern is different. The first PC

[7] Application of the Augmented Dickey-Fuller to each one of the three retained PCs (individual and common PCs) revealed that these are stationary.

of the NYMEX crude oil and the IPE crude oil can forecast the changes of futures prices; this holds for all maturities. The sign of the estimated coefficients of the NYMEX PC is positive while that of the IPE PC is negative. This implies that good news in the NYMEX (IPE) market would increase (decrease) the daily change in the IPE futures prices; this holds across the whole spectrum of maturities. This is important for speculators who form spreads with different underlying assets. Moreover, our findings imply that the NYMEX crude oil market leads the IPE market given that the latter opens before the former. This is in accordance with the results of Lin and Tamvakis (2001). The third PC of the IPE crude oil can also forecast two maturities (*CO2* and *CO3*). The estimated parameters as well as the R^2 values are greater now.

In the case of the heating oil, the third PC of the IPE crude oil can forecast the changes of the futures prices for all maturities. Finally, in the case of the gasoline contracts, there is not a clear pattern since only a few maturities can be forecasted (shortest and the two longest) by the PCs of different commodities; the first PC of the NYMEX crude oil and the gasoline contracts forecast the changes of the longest gasoline series.

In general, the R^2 values are small for all regressions despite the fact that certain PCs are statistically significant; the greatest values are obtained in the case of the IPE contract (1%-3%)[8]. The magnitude of the estimated regression coefficients is also small. These results suggest that the obtained PCs have limited power in order to forecast the subsequent changes in the futures prices. Interestingly, for any given commodity with the exception of IPE, the variables that can be used for forecasting purposes are not the PCs of the same commodity; the IPE and NYMEX PCs can forecast the changes of the futures prices of the other commodities. This indicates that there is a spillover effect between the various markets.

4.2 Joint PCA: The Regression Setup and Results

We test whether the PCs that were derived from the joint PCA on all four commodities can be used to predict the futures prices. The same multiple regression setup is employed

[8] One could argue that the small R^2 is expected in the cases where only the second and third PCs are found to be significant since these explain a small amount of the variance of the changes of the term structure. However, the small R^2 appears also in the cases where the first PC (explaining more than 90% of the total variance) is also significant.

as in Section 4.1. ΔF_t^j is regressed on the three joint PCs $PC_{i,t-1}$ (i=1,2,3) measured at time t-1. Hence, the regression equations are formed as follows

$$\Delta F_t^j = c + a_1 PC_{1,t-1} + a_2 PC_{2,t-1} + a_3 PC_{3,t-1} + u_t \tag{5}$$

j = *CL1*,…, *CL9*, *CO1*,…, *CO7*, *HO1*,…, *HO9*, *HU1*,…, *HU7*. Again, a general to specific approach has been used to estimate equation (5).

Table 5 shows the results of the regressions per commodity. The first column shows the dependent variable of Equation (5). The next four columns show the constant term and the coefficient values of the regressors along with their t-statistics in parentheses (corrected for autocorrelation and heteroscedasticity). The following column shows the R^2 statistic. Finally, the last column shows the F-statistic that tests the null hypothesis that all coefficients (excluding the constant term) are zero. The F-statistic's p-values are shown in parentheses.

(INSERT TABLE 5 HERE)

We can see that in the case of NYMEX and IPE crude oil futures, the joint principal components have no predictive power. On the other hand, the second joint PC can forecast the changes of the heating oil and gasoline futures prices of all maturities but the shortest. The coefficients of the second PC are consistently negative of a relatively high magnitude. However, the R^2 statistics are again small (1%-2.2%) as in the case of the regressions with the PCs obtained from the separate PCA. The small R^2 suggests that the joint PCs cannot forecast the changes of the prices of petroleum futures, just as in the case of the PCs obtained from the separate PCA.

5. Univariate and Vector Autoregressions

In this Section, we check whether the dynamics of the term structure of petroleum futures can be forecasted by running univariate and vector autoregressions as alternative models to the PCA approach. The univariate autoregressions are of the form

$$\Delta F_t^j = c + a_1 \Delta F_{t-1}{}^j + u_t \tag{6}$$

j = *CL1*,…, *CL9*, *CO1*,…, *CO7*, *HO1*,…, *HO9*, *HU1*,…, *HU7*.

The vector autoregressions (VAR) are of the form

$$\Delta F_t^{\,l} = c^l + \Phi^l \Delta F_{t\text{-}1}^{\ \ l} + u_t^{\,l} \tag{7}$$

where ΔF_t^l is the $(J\times 1)$ vector that consists of the changes of the $j=1,\ldots,J$ maturity futures prices for each commodity $l=CL, CO, HO, HU,$, Φ^1 is the $(J\times J)$ matrix of coefficients of the l commodity to be estimated and c^l, u_t^l are the l-commodity $(J\times 1)$ vectors of constants and error terms respectively; the error terms of the j maturities may be correlated. Equations (6) and (7) can be viewed as a test of the weak form of market efficiency (Fama, 1970, 1991).

The results obtained from the regressions given by equations (6) and (7) are evaluated on the grounds of the R^2. Table 6 shows the results from the univariate and VAR autoregressions (Panel A and B, respectively). We can see that almost all the regression coefficients are statistically insignificant and the R^2's are zero for all commodities. The results are similar for the other three commodities, as well, and hence they are not reported due to space limitations. Overall, they are in accordance with the PCA results and they confirm that they dynamics of the petroleum term structures cannot be forecasted.

(INSERT TABLE 6 HERE)

6. Conclusions

The prediction of the evolution of the term structure of petroleum futures is of paramount importance for the participants in the energy derivatives markets. In this paper, we have investigated whether the dynamics of the petroleum futures prices can be forecasted in four major petroleum markets: the NYMEX crude oil, heating oil, gasoline and the IPE crude oil. Following Stock and Watson (2002), we have used the Principal Components Analysis (PCA) to let the data decide on the variables to be used as predictors rather than assuming ourselves ad hoc forecasting variables. PCA was first applied to the time series of daily changes of petroleum futures prices across the whole spectrum of maturities. This enabled us to model the dynamics of the term structure of petroleum futures prices parsimoniously by means of a few retained principal components (PCs). PCA was performed on each commodity market separately (separate PCA), as well as on the four markets jointly (joint PCA). Both the separate and the joint PCA have shown that three components drive the dynamics of the term structure of futures prices. The retained components summarise the dynamics of the whole term structure of futures prices. Their number (three) and interpretation (level, slope, curvature, in the separate PCA) are in line with the results from the previous literature. Then, the retained PCs of all commodities were used in a multiple regression setup to forecast the subsequent daily changes of futures prices.

The forecasting regressions for all commodities under scrutiny yield low R^2's despite that fact that some PCs are statistically significant. In particular, some of the NYMEX and IPE crude oil factors affect the next days' dynamics of all four commodities. Interestingly, the joint PCA that takes into account the interactions in the dynamics of the four markets does not increase the forecasting power of the retained components. Low R^2's also occur in the case where we run standard univariate and vector autoregression models as alternatives to the PCA approach to forecasting.

This study has at least three implications. First, the evidence on the R^2 suggests that the dynamics of the term structure of petroleum futures cannot be forecasted. This is in accordance with the results in Cabibo and Fiorenzani (2004) who had used a different methodology and a single petroleum market to study whether the dynamics of the IPE Brent futures term structure can be forecasted. Second, the dynamics of the term structure of petroleum futures are stable over time in terms of the number and interpretation of factors that drive them; the PCA results obtained from our updated and rich data set are in line with those reported in the previous related literature. Finally, spillover effects are detected between the four markets. This complements the study by Lin and Tamvakis (2001) who had found that there are substantial spillover effects between the NYMEX and IPE crude oil futures prices. Future research should investigate whether non-linear models can forecast the evolution of petroleum futures prices; the linear autoregressive models that have been employed in the current study can be considered as a first-order approximation of the true model.

References

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

Alexander, C. O. (2001b). *Market Models: A Guide to Financial Data Analysis*, John Wiley & Sons Ltd.

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.

Basilevsky, A., (1994). *Statistical Factor Analysis and Related Methods, Theory and Applications*, Wiley Series in Probability and Mathematical Statistics.

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

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.

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

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.

Clewlow, L. and Strickland, C. (1999b). A Multi-Factor Model for Energy Derivatives, Working Paper, School of Finance and Economics, University of Technology, Sydney.

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.

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

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

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.

Jackson, E. (1991). *A User's Guide to Principal Components*, Wiley Series in Probability and Mathematical Statistics.

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

Kamal, M., and Derman, E. (1997). The Patterns of Change in Implied Index Volatilities. Quantitative Strategies Research Notes, Goldman Sachs, New York.

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

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.

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.

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

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.

Samuelson, P.A. (1965). Proof that Properly Anticipated Prices fluctuate Randomly, *Industrial Management Review*, 6, 41-49.

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.

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.

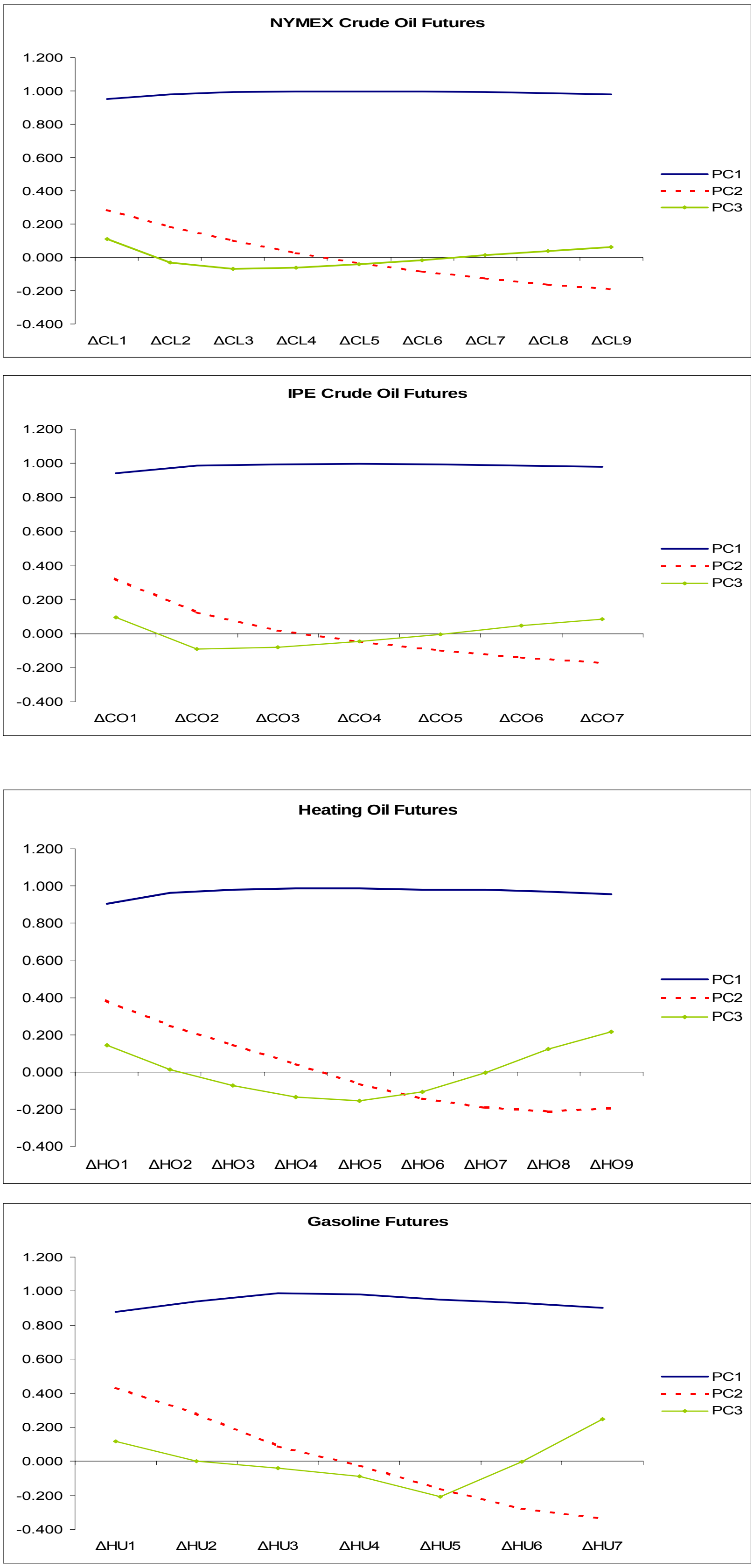

Figure 1: NYMEX IPE crude oil, Heating Oil and Gasoline futures: Correlation Loadings of the first three principal components.

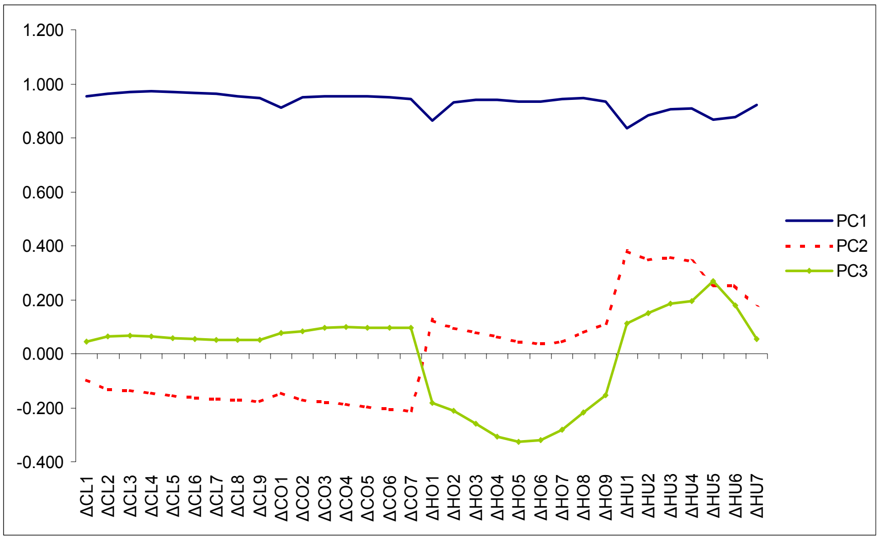

Figure 2: Correlation loadings of the first three joint principal components. Principal Components Analysis has been applied to all four commodities (NYMEX IPE crude oil, Heating Oil and Gasoline futures) jointly for the period from 1/1/1993 to 31/12/2003.

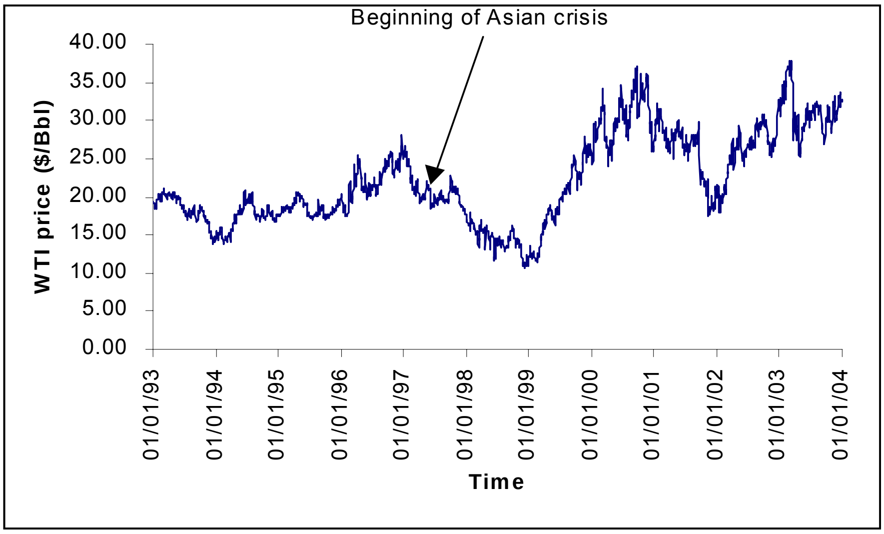

Figure 3: Spot WTI crude oil prices for the period from 1/1/1993 to 31/12/2003.

Panel A: NYMEX Crude Oil generic contracts									
	ΔCL1	*ΔCL2*	*ΔCL3*	*ΔCL4*	*ΔCL5*	*ΔCL6*	*ΔCL7*	*ΔCL8*	*ΔCL9*
Retained observ.	2645	2641	2637	2633	2634	2615	2606	2576	2476
Excluded observ.	224	228	232	236	235	254	263	293	393
Mean	0	0	0	0	0	0	0	0	0
Std. Deviation	0.50	0.45	0.41	0.37	0.35	0.33	0.32	0.30	0.30
Skewness	-0.57	-0.58	-0.46	-0.48	-0.52	-0.46	-0.43	-0.43	-0.42
Kurtosis	5.14	5.51	4.26	4.54	4.94	4.94	4.57	4.33	4.14
Panel B: IPE Crude Oil generic contracts									
	ΔCO1	*ΔCO2*	*ΔCO3*	*ΔCO4*	*ΔCO5*	*ΔCO6*	*ΔCO7*		
Retained observ.	2590	2667	2666	2649	2589	2325	1954		
Excluded observ.	279	202	203	220	280	544	915		
Mean	0	0	0	0	0	0	0		
Std. Deviation	0.47	0.42	0.39	0.36	0.34	0.33	0.33		
Skewness	-0.56	-0.52	-0.44	-0.47	-0.49	-0.44	-0.50		
Kurtosis	5.33	4.99	4.71	4.68	4.72	4.71	4.89		
Panel C: Heating Oil generic contracts									
	ΔHO1	*ΔHO2*	*ΔHO3*	*ΔHO4*	*ΔHO5*	*ΔHO6*	*ΔHO7*	*ΔHO8*	*ΔHO9*
Retained observ.	2606	2580	2581	2576	2551	2547	2477	2272	2100
Excluded observ.	263	289	288	293	318	322	392	597	769
Mean	0	0	0	0	0	0	0	0	0
Std. Deviation	1.47	1.27	1.16	1.07	1.02	0.98	0.94	0.91	0.88
Skewness	-0.48	-0.24	-0.20	-0.25	-0.41	-0.53	-0.46	-0.41	-0.41
Kurtosis	5.23	3.35	3.13	3.21	4.07	4.54	4.26	3.48	3.35
Panel D: Gasoline generic contracts									
	ΔHU1	*ΔHU2*	*ΔHU3*	*ΔHU4*	*ΔHU5*	*ΔHU6*	*ΔHU7*		
Retained observ.	2645	2633	2619	2597	2507	2188	1656		
Excluded observ.	224	236	250	272	362	681	1213		
Mean	0	0	0	0	0	0	0		
Std. Deviation	1.64	1.38	1.21	1.12	1.07	1.04	1.03		
Skewness	-0.85	-0.36	-0.46	-0.26	-0.23	0.05	-0.13		
Kurtosis	10.03	4.56	4.69	4.43	3.77	5.34	4.45		

Table 1: Summary Statistics of the first differences of the futures prices. The results are reported for each expiry (generic contract, i.e., shortest, second shortest, etc), and for each one of the four underlying commodities (NYMEX & IPE crude oil, Heating Oil and Gasoline). The sample corresponds to the period from 1/1/1993-31/12/2003.

Principal component	NYMEX crude oil	IPE crude oil	Heating oil	Gasoline
		Panel A: Separate PCA		
1	97.21	96.66	93.56	88.11
2	99.58	99.23	97.74	95.08
3	99.90	99.73	99.31	96.90
4	99.96	99.88	99.81	98.18
		Panel B: Joint PCA		
1		87.12		
2		90.79		
3		93.60		
4		95.23		

Table 2: Cumulative percentage of variance explained by the principal components (up to four components) obtained from the separate and joint PCA. Results are reported for each one of the four underlying commodities (NYMEX & IPE Crude Oil, Heating Oil and Gasoline). The sample corresponds to the period from 1/1/1993-31/12/2003.

Panel A: Separate PCA - Standardised PCs			
	PC1	***PC2***	***PC3***
	NYMEX crude oil		
Retained observations	2353	2353	2353
Missing observations	516	516	516
Skewness	-0.44	-0.73	0.15
Kurtosis	4.54	10.77	13.79
	IPE crude oil		
Retained observations	1651	1651	1651
Missing observations	1218	1218	1218
Skewness	-0.45	-0.14	0.44
Kurtosis	4.51	6.26	6.96
	Heating oil		
Retained observations	1624	1624	1624
Missing observations	1245	1245	1245
Skewness	-0.37	0.75	0.19
Kurtosis	2.96	18.57	33.24
	Gasoline		
Retained observations	1451	1451	1451
Missing observations	1418	1418	1418
Skewness	-0.34	-3.24	-3.32
Kurtosis	2.97	54.58	100.31

Panel B: Joint PCA - Standardised PCs			
	PC1	***PC2***	***PC3***
Retained observations	563	563	563
Excluded observations	2306	2306	2306
Skewness	-0.52	-1.30	-0.08
Kurtosis	3.26	11.49	4.60

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

Table 4: Results from regressing ΔF_t^j (j = *CL1*,…, *CL9*, *CO1*,…, *CO7*, *HO1*,…, *HO9*, *HU1*,…, *HU7*) on the twelve retained principal components obtained from the separate PCA on the four commodities.

j	c	a_1	a_2	a_3	b_1	b_2	b_3	c_1	c_2	c_3	d_1	d_2	d_3	R^2	F-stat (prob)
	(t-stat)	(t-stat)	(t-stat)	(t-stat)	(t-stat)	(t-stat)	(t-stat)	(t-stat)	(t-stat)	(t-stat)	(t-stat)	(t-stat)	(t-stat)		
Panel A: Dependent variables are the NYMEX Crude Oil generic futures															
CL1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
	-	-	-	-	-	-	-	-	-	-	-	-	-		-
CL2	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
	-	-	-	-	-	-	-	-	-	-	-	-	-		-
CL3	-	-	-	-	-	-	0.029	-	-	-	-	-	-	0.004	6.629
	-	-	-	-	-	-	(2.2)	-	-	-	-	-	-		(0.01)
CL4	-	-	-	-	-	-	0.026	-	-	-	-	-	-	0.004	6.285
	-	-	-	-	-	-	(2.2)	-	-	-	-	-	-		(0.01)
CL5	-	-	-	-	-	-	0.024	-	-	-	-	-	-	0.004	6.052
	-	-	-	-	-	-	(2.2)	-	-	-	-	-	-		(0.01)
CL6	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
	-	-	-	-	-	-	-	-	-	-	-	-	-		-
CL7	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
	-	-	-	-	-	-	-	-	-	-	-	-	-		-
CL8	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
	-	-	-	-	-	-	-	-	-	-	-	-	-		-
CL9	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
	-	-	-	-	-	-	-	-	-	-	-	-	-		-

j	***c*** (***t*-stat**)	a_1 (***t*-stat**)	a_2 (***t*-stat**)	a_3 (***t*-stat**)	b_1 (***t*-stat**)	b_2 (***t*-stat**)	b_3 (***t*-stat**)	c_1 (***t*-stat**)	c_2 (***t*-stat**)	c_3 (***t*-stat**)	d_1 (***t*-stat**)	d_2 (***t*-stat**)	d_3 (***t*-stat**)	R^2	***F*-stat (prob)**
						Panel B: Dependent variables are the IPE Crude Oil generic futures									
CO1	-	0.176	-	-	-0.203	-	-	-	-	-	-	-	-	0.019	13.293
	-	(3.7)	-	-	(-3.7)	-	-	-	-	-	-	-	-		(0.00)
CO2	-	0.145	-	-	-0.178	-	0.041	-	-	-	-	-	-	0.025	12.145
	-	(3.5)	-	-	(-3.8)	-	(2.7)	-	-	-	-	-	-		(0.00)
CO3	-	0.143	-	-	-0.169	-	0.036	-	-	-	-	-	-	0.026	12.721
	-	(3.9)	-	-	(-4.2)	-	(2.8)	-	-	-	-	-	-		(0.00)
CO4	-	0.133	-	-	-0.164	-	-	-	-	-	-	-	-	0.022	15.666
	-	(4.1)	-	-	(-4.6)	-	-	-	-	-	-	-	-		(0.00)
CO5	-	0.125	-	-	-0.162	-	-	-	-	-	-	-	-	0.025	18.108
	-	(4.2)	-	-	(-4.9)	-	-	-	-	-	-	-	-		(0.00)
CO6	-	0.114	-	-	-0.155	-	-	-	-	-	-	-	-	0.027	18.859
	-	(4.0)	-	-	(-4.8)	-	-	-	-	-	-	-	-		(0.00)
CO7	-	0.105	-	-	-0.149	-	-	-	-	-	-	-	-	0.030	19.581
	-	(3.7)	-	-	(-4.7)	-	-	-	-	-	-	-	-		(0.00)

j	c	a_1	a_2	a_3	b_1	b_2	b_3	c_1	c_2	c_3	d_1	d_2	d_3	R^2	F-stat (prob)
	(t-stat)	(t-stat)	(t-stat)	(t-stat)	(t-stat)	(t-stat)	(t-stat)	(t-stat)	(t-stat)	(t-stat)	(t-stat)	(t-stat)	(t-stat)		

Panel C: Dependent variables are the Heating Oil generic futures

j	c	a_1	a_2	a_3	b_1	b_2	b_3	c_1	c_2	c_3	d_1	d_2	d_3	R^2	F-stat (prob)
HO1	-	-	-	-	-	-	0.091	-	-	-	-	-	-	0.003	4.796
	-	-	-	-	-	-	(2.3)	-	-	-	-	-	-		(0.03)
HO2	-	-	-	-	-	-	0.095	-	-	-	-	-	-	0.005	7.039
	-	-	-	-	-	-	(2.5)	-	-	-	-	-	-		(0.01)
HO3	-	-	-	-	-	-	0.096	-	-	-	-	-	-	0.006	8.697
	-	-	-	-	-	-	(2.8)	-	-	-	-	-	-		(0.00)
HO4	-	-	-	-	-	-	0.088	-	-	-	-	-	-	0.006	8.563
	-	-	-	-	-	-	(2.5)	-	-	-	-	-	-		(0.00)
HO5	-	-	-	-	-	-	0.087	-	-	-	-	-	-	0.006	9.315
	-	-	-	-	-	-	(2.7)	-	-	-	-	-	-		(0.00)
HO6	-	-	-	-	-	-	0.094	-	-	-	-	-	-	0.008	11.852
	-	-	-	-	-	-	(3.0)	-	-	-	-	-	-		(0.00)
HO7	-	-	-	-	-	-	0.086	-	-	-	-	-	-	0.007	10.821
	-	-	-	-	-	-	(2.8)	-	-	-	-	-	-		(0.00)
HO8	-	-	-	-	-	-	0.075	-	-	-	-	-	-	0.006	8.843
	-	-	-	-	-	-	(2.5)	-	-	-	-	-	-		(0.00)
HO9		-	-	-	-	-	-	-	-	-	-	-0.084	-	0.008	8.941
	-	-	-	-	-	-	-	-	-	-	-	(-3.0)	-		(0.00)

j	*c* (*t*-stat)	a_1 (*t*-stat)	a_2 (*t*-stat)	a_3 (*t*-stat)	b_1 (*t*-stat)	b_2 (*t*-stat)	b_3 (*t*-stat)	c_1 (*t*-stat)	c_2 (*t*-stat)	c_3 (*t*-stat)	d_1 (*t*-stat)	d_2 (*t*-stat)	d_3 (*t*-stat)	R^2	*F*-stat (prob)
Panel D: Dependent variables are the Gasoline generic futures															
HU1	-	-	-	-	-	-	-	-	-0.087	-	-	-	-	0.002	3.817
	-	-	-	-	-	-	-	-	(-2.0)	-	-	-	-		(0.05)
HU2	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
	-	-	-	-	-	-	-	-	-	-	-	-	-		-
HU3	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
	-	-	-	-	-	-	-	-	-	-	-	-	-		-
HU4	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
	-	-	-	-	-	-	-	-	-	-	-	-	-		-
HU5	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
	-	-	-	-	-	-	-	-	-	-	-	-	-		-
HU6	-	-	-	-	-	-	0.083	-	-	-	-0.131	-	-	0.021	8.852
	-	-	-	-	-	-	(2.1)	-	-	-	(-2.6)	-	-		(0.00)
HU7	-	0.283	-	-	-	-	-	-	-	-	-0.353	-	-	0.025	14.033
	-	(3.9)	-	-	-	-	-	-	-	-	(-5.1)	-	-		(0.00)

Table 5: Results from regressing ΔF_t^j (where $(j = CL1,\ldots, CL9, CO1,\ldots, CO7, HO1,\ldots, HO9, HU1,\ldots, HU7)$) on the three retained common principal components obtained from the joint PCA on the four commodities.

j	***c*** (***t*-stat**)	***a_1*** (***t*-stat**)	***a_2*** (***t*-stat**)	***a_3*** (***t*-stat**)	***R^2***	***F*-stat (prob)**
Panel A: Dependent variables are the NYMEX crude oil generic futures						
No significant results found for any maturity.						
Panel B: Dependent variables are the IPE crude oil generic futures						
No significant results found for any maturity.						

j	c (t-stat)	a_1 (t-stat)	a_2 (t-stat)	a_3 (t-stat)	R^2	F-stat (prob)
			Panel C: Dependent variables are the Heating Oil generic futures			
HO1	-	-	-	-	-	-
	-	-	-	-		-
HO2	-	-	-0.172	-	0.012	6.621
	-	-	(-2.4)	-		(0.01)
HO3	-	-	-0.177	-	0.015	8.596
	-	-	(-2.7)	-		(0.00)
HO4	-	-	-0.179	-	0.018	9.845
	-	-	(-3.0)	-		(0.00)
HO5	-	-	-0.167	-	0.016	9.101
	-	-	(-2.9)	-		(0.00)
HO6	-	-	-0.164	-	0.016	9.003
	-	-	(-2.9)	-		(0.00)
HO7	-	-	-0.148	-	0.015	8.102
	-	-	(-2.6)	-		(0.00)
HO8	-	-	-0.133	-	0.014	7.276
	-	-	(-2.5)	-		(0.01)
HO9	-	-	-0.122	-	0.012	6.331
	-	-	(-2.2)	-		(0.01)
			Panel D: Dependent variables are the Gasoline generic futures			
HU1	-	-	-	-	-	-
	-	-	-	-		-
HU2	-	-	-	-	-	-
	-	-	-	-		-
HU3	-	-	-0.150	-	0.009	5.160
	-	-	(-2.3)	-		(0.02)
HU4	-	-	-0.172	-	0.013	7.004
	-	-	(-2.7)	-		(0.01)
HU5	-	-	-0.167	-	0.013	7.336
	-	-	(-2.7)	-		(0.01)
HU6	-	-	-0.236	-	0.030	15.924
	-	-	(-3.8)	-		(0.00)
HU7	-	-	-0.214	-	0.027	13.046
	-	-	(-3.6)	-		(0.00)

Panel A: NYMEX Crude Oil Futures - Univariate Autoregressions

	ΔCL01	**ΔCL02**	**ΔCL03**	**ΔCL04**	**ΔCL05**	**ΔCL06**	**ΔCL07**	**ΔCL08**	**ΔCL09**
c	0.002	0.000	0.001	0.001	0.001	0.002	0.002	0.001	0.002
	(0.2)	(0.0)	(0.1)	(0.1)	(0.2)	(0.2)	(0.4)	(0.2)	(0.3)
ΔCL01(-1)	-0.113	-0.112	-0.166	-0.164	-0.161	-0.140	-0.132	-0.136	-0.118
	(-0.6)	(-0.7)	(-1.1)	(-1.2)	(-1.3)	(-1.2)	(-1.2)	(-1.3)	(-1.2)
ΔCL02(-1)	0.401	0.466	0.614	0.486	0.413	0.267	0.249	0.307	0.232
	(0.9)	(1.2)	(1.7)	(1.5)	(1.4)	(0.9)	(0.8)	(1.2)	(0.8)
ΔCL03(-1)	-0.272	-0.266	-0.455	-0.136	-0.077	0.166	0.183	0.055	0.236
	(-0.3)	(-0.4)	(-0.7)	(-0.2)	(-0.1)	(0.3)	(0.4)	(0.1)	(0.5)
ΔCL04(-1)	-0.133	-0.348	-0.428	-0.684	-0.500	-0.628	-0.640	-0.632	-0.801
	(-0.1)	(-0.3)	(-0.4)	(-0.7)	(-0.5)	(-0.7)	(-0.8)	(-0.8)	(-1.0)
ΔCL05(-1)	-0.324	-0.097	0.090	0.125	-0.068	-0.060	-0.156	-0.122	-0.057
	(-0.2)	(-0.1)	(0.1)	(0.1)	(-0.1)	(-0.1)	(-0.1)	(-0.1)	(-0.1)
ΔCL06(-1)	0.874	0.891	0.841	0.754	0.674	0.732	0.946	0.962	0.956
	(0.5)	(0.6)	(0.6)	(0.6)	(0.5)	(0.6)	(0.9)	(0.9)	(0.9)
ΔCL07(-1)	-0.743	-1.143	-0.827	-0.714	-0.698	-0.557	-0.703	-0.429	-0.617
	(-0.4)	(-0.8)	(-0.6)	(-0.6)	(-0.7)	(-0.6)	(-0.7)	(-0.4)	(-0.7)
ΔCL08(-1)	-1.271	-0.895	-0.865	-0.782	-0.687	-0.917	-0.895	-1.344	-1.063
	(-1.0)	(-0.7)	(-0.8)	(-0.8)	(-0.8)	(-1.1)	(-1.1)	(-1.2)	(-1.4)
ΔCL09(-1)	1.656	1.583	1.250	1.146	1.110	1.123	1.124	1.302	1.183
	(1.8)	(2.0)	(1.7)	(1.7)	(1.8)	(1.9)	(1.9)	(1.9)	(2.2)
R^2	**0.004**	**0.005**	**0.006**	**0.007**	**0.009**	**0.009**	**0.012**	**0.016**	**0.019**

Panel B: NYMEX Crude Oil Futures - Vector Autoregressions

	ΔCL01	**ΔCL02**	**ΔCL03**	**ΔCL04**	**ΔCL05**	**ΔCL06**	**ΔCL07**	**ΔCL08**	**ΔCL09**
c	0.000	0.000	0.001	0.001	0.001	0.001	0.001	0.001	0.001
	(0.04)	(0.00)	(0.09)	(0.12)	(0.16)	(0.16)	(0.17)	(0.15)	(0.18)
ΔCL01(-1)	-0.013	-0.039	-0.107	-0.112	-0.110	-0.110	-0.111	-0.111	-0.111
	(-0.1)	(-0.3)	(-1.0)	(-1.1)	(-1.2)	(-1.3)	(-1.3)	(-1.4)	(-1.5)
ΔCL02(-1)	0.111	0.254	0.482	0.343	0.258	0.224	0.224	0.229	0.228
	(0.3)	(0.8)	(1.6)	(1.2)	(1.0)	(0.9)	(1.0)	(1.0)	(1.1)
ΔCL03(-1)	-0.057	-0.051	-0.310	0.070	0.157	0.213	0.210	0.204	0.229
	(-0.1)	(-0.1)	(-0.5)	(0.1)	(0.3)	(0.4)	(0.4)	(0.4)	(0.5)
ΔCL04(-1)	-0.192	-0.562	-0.684	-0.979	-0.786	-0.781	-0.799	-0.786	-0.851
	(-0.1)	(-0.5)	(-0.6)	(-1.0)	(-0.8)	(-0.9)	(-0.9)	(-1.0)	(-1.1)
ΔCL05(-1)	-0.239	0.006	0.188	0.240	0.026	0.036	-0.017	-0.082	-0.039
	(-0.1)	(0.0)	(0.1)	(0.2)	(0.0)	(0.0)	(0.0)	(-0.1)	(0.0)
ΔCL06(-1)	0.954	1.203	1.262	1.108	1.006	0.848	1.060	1.062	1.039
	(0.5)	(0.8)	(0.9)	(0.8)	(0.8)	(0.7)	(0.9)	(1.0)	(1.0)
ΔCL07(-1)	-0.283	-0.834	-0.637	-0.556	-0.544	-0.435	-0.619	-0.465	-0.503
	(-0.2)	(-0.5)	(-0.5)	(-0.4)	(-0.5)	(-0.4)	(-0.6)	(-0.4)	(-0.5)
ΔCL08(-1)	-2.319	-1.822	-1.673	-1.456	-1.322	-1.337	-1.344	-1.512	-1.392
	(-1.2)	(-1.0)	(-1.1)	(-1.0)	(-1.0)	(-1.0)	(-1.1)	(-1.3)	(-1.2)
ΔCL09(-1)	2.128	1.927	1.538	1.372	1.324	1.334	1.375	1.430	1.358
	(1.9)	(1.9)	(1.7)	(1.6)	(1.7)	(1.8)	(1.9)	(2.1)	(2.1)
R^2	**0.003**	**0.004**	**0.004**	**0.004**	**0.006**	**0.008**	**0.012**	**0.015**	**0.019**

Table 6: Results from the univariate and vector autoregressions run for the NYMEX crude oil futures data set. The univariate autoregressions are of the form $\Delta F_t^{\,j} = c + a_1 \Delta F_{t-1}^{\,j} + u_t$ j = *CL1*,…, *CL9*, *CO1*,…, *CO7*, *HO1*,…, *HO9*, *HU1*,…, *HU7*. The vector autoregression is of the form $\Delta F_t = c + \Phi \Delta F_{t\text{-}1} + u_t$.