Principal Components Analysis for Correlated Curves and Seasonal Commodities: The Case of the Petroleum Market

CARLOS TOLMASKY
DMITRY HINDANOV*

This article presents a family of term structure models that can be applied to value contingent claims in multicommodity and seasonal markets. We apply the framework to the futures contracts on crude and heating oils trading on NYMEX. We show how to deal with the problem of having to value products depending on the "whole" market, such as spread options on contracts on a single commodity maturing at different times (time-spreads) or spread options on the added value of the products derived from the raw commodity (crack spreads). Also, we show how to build term structure models for a commodity that experiences seasonality, such as heating oil. © 2002 Wiley Periodicals, Inc. Jrl Fut Mark 22:1019–1035, 2002

*Correspondence author, Cargill Global Capital Markets, 12700 Whitewater Drive, Minnetonka, Minnesota 55343-9439; e-mail: dmitry_hindanov@cargill.com

Received September 2001; Accepted April 2002

- *Carlos Tolmasky is a Research Analyst at Cargill Energy in Minnetonka, Minnesota, and an Assistant Professor at the University of Minnesota in Minneapolis, Minnesota.*
- *Dmitry Hindanov is a Research Analyst at Cargill Global Capital Markets in Minnetonka, Minnesota.*

The Journal of Futures Markets, Vol. 22, No. 11, 1019–1035 (2002) © 2002 Wiley Periodicals, Inc.
Published online in Wiley InterScience (www.interscience.wiley.com). DOI: 10.1002/fut.10046

INTRODUCTION

Derivative securities have been trading in the financial markets for more than a couple of decades. The interest rate derivatives world has received most of the attention. The first generation of interest rate models assumed that the dynamics of the curve was driven by the short rate. They were one-factor models in the spirit of Black and Scholes (1973). It was not very long before researchers and practitioners noticed the shortcomings of trying to model a multidimensional object in a one-factor setting. Due to this, the modeling of derivatives instruments linked to the interest rate forward or yield curves has been extensively studied. One of the most celebrated models is the one due to Heath, Jarrow, and Morton (1990, 1992). They developed an arbitrage-free approach in which the forward rates are the state variables to model (instead of being derived from the short rate).

In the case of commodities, the developments were similar. The role of the short rate was played by the spot price of the commodity based on which the whole forward (or futures) curve was derived using arbitrage arguments. One of the first attempts to get rid of the one factor settings was done by Gibson and Schwartz (1990). They introduced a second factor corresponding to the convenience yield, which is defined as the difference between the interest rate and the cost of carry. As an alternative, and following the ideas from the interest rates world, we could try to model the forward prices directly and create a multifactor model. This was first done in the world of commodities by Reisman (1992) and later on by Cortazar and Schwartz (1994).

An interesting question that arises when using this type of models is how to choose the volatility functions. Litterman and Sheinkman (1991) found that the U.S. government bonds are almost entirely driven by three factors that they called *level, steepness*, and *curvature*. This result was "rediscovered" in the case of commodities by Cortazar and Schwartz when studying the stochastic movement of the copper futures prices.

Even though the similarities with the interest rate world are somewhat striking, it is clear that some differences should appear if we try to model futures prices for commodities that may be nonstationary.

Our aim is to extend the empirical work done by Cortazar and Schwartz to the petroleum market. Much of the trading in the petroleum market is generated by the economic relationships between the raw commodity (crude) and its products (heating oil, gasoline, jet fuel, etc.). One example of this is the spread between the crude and a product ("crack spread"), which is a proxy for the margin obtained by a refinery running crude.

A special feature of the petroleum market is the importance of seasonality, and the varying degrees of seasonal volatility across different petroleum products. We attempt to find a parsimonious way to parameterize the seasonality of volatilities and correlations, and study their effects on VAR calculations and option pricing.

A lot of research has previously been done on seasonal commodities. In the case of petroleum spreads, Girma and Paulson (1998) explore statistical relationships and seasonalities with the purpose of testing trading rules. According to them, "the relationship concerning intercommodity spreads can be complex and difficult to understand and there is a lack of concrete theoretical foundation that can explain the behavior of these spreads." To a certain extent, our article deals with the statistical aspects of this theoretical foundation (when it comes to derivative pricing).

Previous to Girma and Paulson's work, Milonas (1991) studied crop-related seasonalities in agricultural commodity prices. In the spirit of the abundant literature on the stock market, Gay and Kim (1987) test for day-of-the-week and month-of-the-year effects in the futures markets. They found that "the results are surprisingly similar to those found in studies of seasonality in stock returns, despite the existence of significant institutional differences between the two markets."

However, to the best of our knowledge, there has not yet been an attempt to integrate the theory of term structure modeling in the context of markets presenting seasonal behavior.

The starting point in our analysis is the term structure of NYMEX futures on crude oil (CL) and heating oil (HO).

First, we determine empirically what factors drive the stochastic movement of future prices for the crude and heating oil separately. In this case, our results are very similar to the ones already reported by Cortazar and Schwartz for the case of copper and previously by Litterman and Scheinkman for the case of Treasury bonds. The resulting model can then be used to price derivatives depending on different points on the futures curve such as options on timespreads or swaptions.

Holding factor loadings fixed across seasons, we find that in the case of heating oil the relative importances of the factors vary, depending on the season. We show that the correlations among contracts are higher when the commodity is off-season and, therefore, the first factor, the level, gains in importance.

Next, we study the joint dynamics of the crude oil and heating oil. As we have mentioned, the relations among the different commodities and different maturities are at the core of the economics of the petroleum market. To be able to price options on crack spreads, timespreads, swaptions, or combinations of those we would need to build a model involving multiple curves. To achieve this, we study the dynamics of the "complete" (futures) market by extending the analysis done for individual curves to an analysis that shows the factors driving both curves. We find again that the most important one is the parallel move that now applies evenly across maturities and across commodities. Next in importance are the *steepness* factors that are parallel from commodity to commodity.

THE GENERAL APPROACH

We assume that the futures market is frictionless and arbitrage-free. Therefore (Harrison & Kreps, 1979; Harrison & Pliska, 1981) there exists a unique probability measure under which futures prices are martingales. The general framework that will be used is the one due to Heath, Jarrow, and Morton. Our starting point is a set of correlated futures curves F_j ($j = 1, \ldots, m$), corresponding to m different commodities to be modeled. Each futures contract will be denoted by $F_{jk}(t, T_k)$ where k will run from 1 to k_j, the index of the contract maturing furthest in the future among those considered for the jth commodity. So, if n is the total number of contracts we have that

$$n = \sum_{j=1}^{m} k_j$$

In what follows and with the purpose of easing the notation, we will drop the index denoting the commodity (j). In this model, the dynamics of the futures curve is given (in the risk neutral world) by

$$\frac{dF_k(t, T_k)}{F_k(t, T_k)} = \sum_{i=1}^{l} \sigma_{i,k}(t)\, dW^i(t) \tag{1}$$

where $W^1, \ldots, W^l$ are l independent Brownian motions under the equivalent martingale measure, and $\sigma_{i,k}(t)$ are volatility functions corresponding to the futures prices that will be determined empirically. Their dependence on the current time t will be found crucial when applying the model to commodities that experience seasonality. We stress the fact that we are looking for a model that accounts for the correlations, and therefore explains the joint dynamics existing in all m curves. To find the volatility functions we will analyze the principal components of the correlation matrix of the historical returns. This methodology will ultimately allow us to capture the variance of the multiple-curve market with the minimum number of factors (which will lead to a less computationally intensive model). With this objective in mind, we impose the following factor–analytic structure on the covariance in (1):

$$\sigma_{i,k}(t) = \sigma_k(t)\sqrt{\lambda_i(\xi)V_i(T_k - t)}$$

Here, $\sigma_k(t)$ is the same across the l factors, the eigenvalues $\lambda_1(\xi),\ldots,\lambda_l(\xi)$ depend only on the season of the year $\xi(t)$, and the eigenvectors $V_1,\ldots,V_l$ depend only on commodity and time to maturity. By setting $l < n$, we achieve dimensionality reduction and, as a side effect, constrain the possible correlation matrices of the commodities. The factor communalities by commodity, season and time to expiration are

$$\nu(\xi, T_k - t) = \sum_{i=1}^{l} V_i^2(T_k - t)\lambda_i(\xi(t)) \tag{2}$$

We assume that the factor communalities are always equal to 1. For the standard principal component analysis the estimated communalities are equal to 1 when $l = n$, but vary with commodity, season, and time to expiration when $l < n$. This formulation characterizes the seasonality of the correlations among different commodities and curve points. As for the seasonality of volatility, we maintain the assumption that it can be recovered from the term structures of the implied volatilities of the futures.

We rewrite (1) as

$$\frac{dF_k(t,T_k)}{F_k(t,T_k)} = \sigma_k(t)\sum_{i=1}^{l}\sqrt{\lambda_i(\xi)}\,V_i(T_k - t)dW^i(t) \tag{3}$$

which leaves F_k's lognormally distributed over any time step Δt, with volatility

$$\sqrt{\int_t^{t+\Delta t} \sigma_k(s)\nu(\xi(s), T_k - s)\,ds} \tag{4}$$

All the usual caveats apply when calibrating (4) to the market. In the case of a flat term structure of volatilities of the kth contract, we need only one implied volatility per contract to calibrate the model. For nonflat, including seasonal, term structures of volatilities we need to either observe the term structure for each individual contract, or come up with reasonable assumptions about the volatility surface.

Solving the SDE for F_k gives:

$$\log \frac{F_k(t+\Delta t, T_k)}{F_k(t, T_k)} = -\frac{1}{2}\int_t^{t+\Delta t} \sigma_k(s)^2 \nu(\xi(s), T_k - s)\, ds + \sum_{i=1}^{l} \int_t^{t+\Delta t} \sigma_k(s)\sqrt{\lambda_i(\xi(s))}\, V_i(T_k - s)\, dW_i(s) \quad (5)$$

which, for the purpose of simulating, could be approximated by

$$\log \frac{F_k(t+\Delta t, T_k)}{F_k(t, T_k)} = -\frac{\sigma_k^2(t)}{2}\nu(\xi, T_k - t)\Delta t + \sigma_k(t)\sum_{i=1}^{l} \phi_i \sqrt{\lambda_i(\xi) V_i^2(T_k - t)\Delta t} \quad (6)$$

where $\phi_1, \ldots, \phi_l$ are standard normal i.i.d. random variables.

Let us pause to contemplate what we can expect from the above model. First, note that for a futures contract taken in isolation it is irrelevant how the communalities in (2) are split among the different factors. That is, the multifactor aspect of the model is only relevant for pricing derivatives that depend on the correlations among different futures maturities and commodities. On the other hand, these correlations have direct bearing on computing VARs of portfolios of futures. The same is true for the seasonality of correlations.

Second, it is reasonable to expect that the principal components of any single market behave similarly to what was shown by Cortazar and Schwartz (1994). That is, one would look for a parallel shift first, then for changes in slope and curvature and expect these to explain a large proportion of the futures' volatilities. This is because the futures contracts are positively correlated, and the correlation declines with the difference in maturity. Thus, a joint move will tend to be more important than a separating move of the same frequency, and a low-frequency move will tend to be more important than a higher frequency one. The mathematics is worked out in Forzani and Tolmasky (2002).

The factor pattern may be less obvious in the case of multiple commodity curves, and may even be unstable. This is because in this case a factor's importance depends not only on the correlations within the same commodity, but also on the correlations among the commodities. To find out the behavior of the factor structure, we can extend the analysis from Forzani and Tolmasky (2002). If we assume that the correlation between two nearby contracts on the same commodity is a fixed number ρ and the correlation between contracts from different commodities (but same tenors) is a fixed number μ, we obtain, in the case of two commodities, the correlation matrix:

$$\begin{pmatrix} C_\rho & \mu C_\rho \\ \mu C_\rho & C_\rho \end{pmatrix}$$

where

$$C_\rho = \begin{pmatrix} 1 & \rho & \rho^2 & \dots & \dots & \rho^n \\ \rho & 1 & \rho & \dots & \dots & \rho^{n-1} \\ \dots & \dots & \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots & \dots & \dots \\ \rho^{n-1} & \rho^{n-2} & \rho^{n-3} & \dots & 1 & \rho \\ \rho^n & \rho^{n-1} & \rho^{n-2} & \dots & \rho & 1 \end{pmatrix}$$

We can check that the correlation matrices found in reality are similar to this idealized one. It is easy to check that if V_n is the nth eigenvector for C_ρ with eigenvalue λ_n, then (V_n, V_n) and $(V_n, -V_n)$ are eigenvectors of $C_{\rho,\mu}$ with eigenvalues $\lambda_n(1 + \mu)$ and $\lambda_n(1 - \mu)$, respectively. Therefore, depending on the correlation between the curves μ, the eigenvectors for the whole system will be a joint frequency move or a separating one. For example, if the curves were to be very correlated (μ very close to 1) then all the most important eigenvectors would turn out to be of the form (V, V) and the ones of the form $(V, -V)$ would have very small eigenvalues. On the contrary, if the correlation between the curves is very low (μ close to 0) the eigenvectors (V, V) and $(V, -V)$ would have similar importance.

CRUDE OIL

In this case, the model we present follows Cortazar and Schwartz almost verbatim. As was said before, we use principal components analysis to find the orthogonal factors and corresponding importances. We work with the history of NYMEX crude oil futures from 1983 through 2000, and utilize the term structure of 10 contracts nearest expiration. The data is ordered according to the remaining time to maturity: first contract, second contract, and so on. We then form the term structure of the weekly log-returns (rolling the contracts when necessary). To remove the effects of seasonal volatility, we divide the returns by the estimated volatility for the relevant season. Then we compute the correlation matrix of returns.

1.0000	0.9742	0.9488	0.9264	0.9063	0.8873	0.8679	0.8468	0.8265	0.8056
0.9742	1.0000	0.9917	0.9789	0.9649	0.9503	0.9344	0.9164	0.8992	0.8811
0.9488	0.9917	1.0000	0.9962	0.9884	0.9781	0.9658	0.9513	0.9370	0.9213
0.9264	0.9789	0.9962	1.0000	0.9973	0.9910	0.9822	0.9709	0.9593	0.9459
0.9063	0.9649	0.9884	0.9973	1.0000	0.9979	0.9924	0.9843	0.9752	0.9642
0.8873	0.9503	0.9781	0.9910	0.9979	1.0000	0.9980	0.9931	0.9865	0.9778
0.8679	0.9344	0.9658	0.9822	0.9924	0.9980	1.0000	0.9982	0.9941	0.9876
0.8468	0.9164	0.9513	0.9709	0.9843	0.9931	0.9982	1.0000	0.9984	0.9942
0.8265	0.8992	0.9370	0.9593	0.9752	0.9865	0.9941	0.9984	1.0000	0.9982
0.8056	0.8811	0.9213	0.9459	0.9642	0.9778	0.9876	0.9942	0.9982	1.0000

FIGURE 1
The correlation matrix of the changes of the first 10 crude oil futures prices.

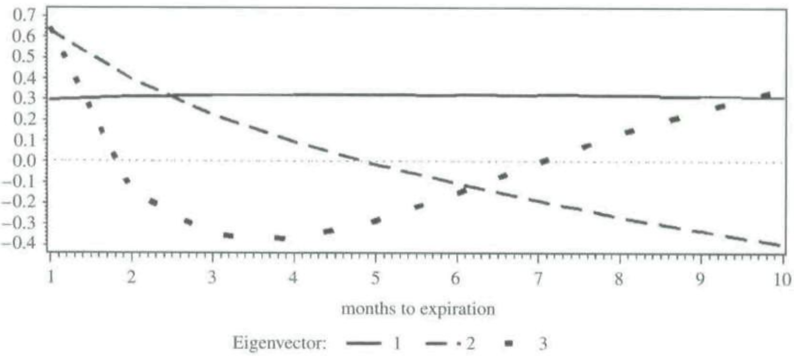

FIGURE 2
Crude oil: first three eigenvectors of the correlation matrix.

Figure 1 shows the computed correlation matrix for the period in study. The pattern of correlations is typical of a futures curve: the correlations in general increase with time to expiration and decline with time distance between two expirations. The tendency of the correlations to increase with time to expiration is the major difference between this correlation matrix and the stylized one analyzed in Forzani and Tolmasky (2002).

We now diagonalize the correlation matrix to find the factor variances λs and factor loadings V^is. The factor loadings for the first three factors are shown in Figure 2. We find that most of the variance (95.92%) is explained by a parallel move, the "level" factor. Next in importance comes a move affecting long and short maturity contracts in opposite ways, the "steepness" factor. These two factors together account for 99.53% of the variance. Adding one more, the "curvature" factor, gets the model to 99.89% of variance explanation. Therefore, we conclude that the dynamics of the futures curve, in the case of crude oil, resembles the dynamics found previously by Litterman and Scheinkman

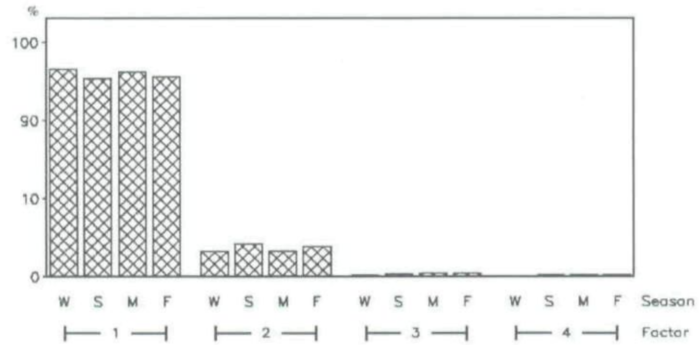

FIGURE 3
Crude oil: relative importance of the first four factors by season.

in the case of treasury bonds, and by Cortazar and Schwartz in the case of copper. The third eigenvector is steeper for short expirations than for long ones, a result of the higher correlations among longer expirations.

To see how the correlations among the contracts vary by season, we compute the fitted factor scores and compare their variances by season. Figure 3 shows how much each of the first four factors contributes to their total variance each season. Each vertical bar on the chart is the ratio of a fitted factor's variance to the sum of variances of the first four factors, by season. The proportion of the variance explained by the first factor is stable across the seasons.

SEASONALITY IN CORRELATIONS: A MODEL FOR HEATING OIL

The main question we try to answer in this section is how the results of the PCA differ when we build a model for a commodity that experiences seasonality. If we analyze the explanatory power of each of the principal components in the case of crude we find that it is fairly stable across trading periods. Due to seasonality effects, we can guess that this will not be the case in the case of heating oil.

First, note that, as one would expect, the factor pattern for the heating oil is remarkably similar to that of the crude oil (Fig. 4). Similarly, 95.80% of the total variance is due to changes in the "level," 99.02% is explained by the "level" and "slope," and 99.63% by the first three factors. This is to be expected given that the heating oil's correlation matrix is stereotypical of many commodity markets. The factor pattern is remarkably similar to that of the crude oil, as Figure 4 demonstrates.

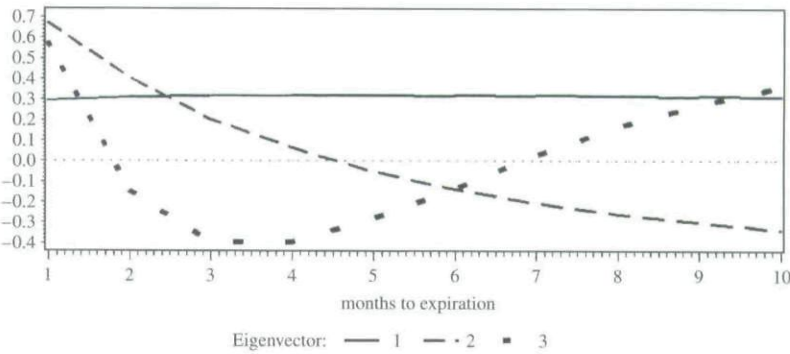

FIGURE 4
Heating oil: first three eigenvectors of the correlation matrix.

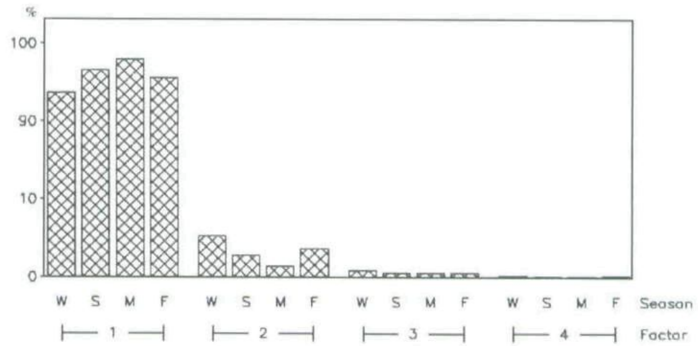

FIGURE 5
Heating oil: relative importance of the first four factors by season.

The similarities end when we take a look at the seasonal pattern of the factor scores. Figure 5 shows how the relative importances of the heating oil's fitted factor scores visibly change by season. The first factor, the "level," appears to be most important during the summer and spring, which are the quarters in which the heating oil is off-season. In other words, we see that when the heating oil is off-season, the futures curve behaves more like a one-factor driven curve. Changes in slope and curvature are relatively more important in the winter and also in the fall, when the heating oil stocks are built up and the market is trying to forecast the interplay of the spot supply and demand during the heating season.

Are these seasonal differences statistically significant? Although some results on hypothesis testing in PCA models are available in the literature (e.g., Kshirsagar, 1972), we are not aware of any work on the sampling distribution of the ratio of the first eigenvalue of a correlation matrix to the sum of the n largest. Bootstrapping 10,000 times the difference between the summer and winter contributions of the first factor gives us the standard errors of 1.43% for the crude and 1.17% for the heating oil. Taking further liberties with distributional assumptions, the p-values for rejecting the null that the first factor is more important in the winter than in the summer are: 0.54 for the crude, 0.00 for the heating oil. But these numbers should not be viewed as more than a rough measure of the ratio of signal to noise present in the data.

A MODEL FOR PETROLEUM PRODUCTS

In this section we generalize the empirical approach to a set of correlated curves. Heating oil and gasoline are among the most actively traded petroleum products. Both also have futures contracts traded on NYMEX.[1] Together with crude, they give a fairly complete picture of the state of the U.S. petroleum market. The contracts are not only traded to express a view or to hedge exposure on each commodity but also to arbitrage their relationships. For example, as a means of quantifying the economic value of keeping oil in its crude form versus the value of refining it speculators and hedgers look at the "cracks." A heating oil crack is the value resulting from subtracting the price of the crude oil from the price of the heating oil. It is used as a proxy for the added value of running the crude to obtain a product. As in the case of the crude, timespreads are also traded on the derived products and on the cracks. It is thus important to be able to uncover the dynamics governing the "whole" market. Knowing it will allow us to develop a model in which we can price derivatives on any combination across time and commodity. The methodology we use here is the natural extension of the previous two sections. We compute the covariance matrix of the first 10 crude and the first 10 heating oil contracts and factorize it. The principal components and their relative importances are shown in Figures 6 and 7.

We see that the general flavor of the one-commodity analysis stays with us. The most important factor, explaining 93.46% of the variance, is a parallel move in the whole market. Next we find three factors that separate the commodities. The second and the third factors separate the long end of one curve from the short end of the other, bringing the share of the total variance explained in 98.30%. The fourth factor separates equal maturities of the two curves, and at the same time is responsible

[1]Heating oil (HO) and gasoline (HU) started trading on NYMEX in November 1978 and December 1984, respectively.

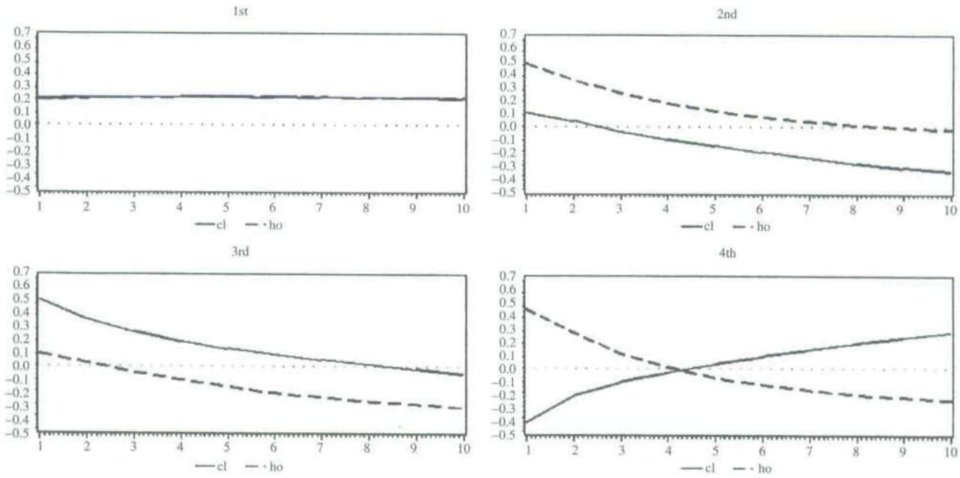

FIGURE 6
Petroleum markets: first four eigenvectors by commodity.

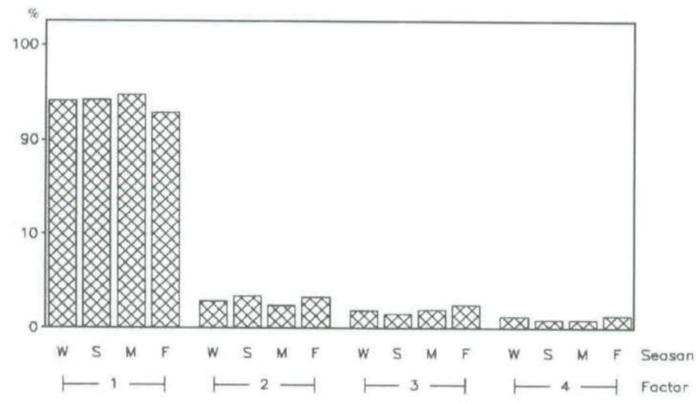

FIGURE 7
Petroleum markets: relative importance of the first four factors by season.

for moving the long and the short contracts of each commodity in opposite directions. Together, the first four factors account for 99.36% of the two petroleum products' covariance. It is interesting to note that factors 2 and 3 do not differ substantially in their contributions to the covariance, or in their roles. The seasonality is not as pronounced as in the case of heating oil alone, and is more difficult to interpret, although the relative importance of a parallel shift of both curves is still higher in the summer. Overall, the complexity of the PCA results has increased tremendously in making a small step from a one-commodity to a two-commodity setup.

APPLICATIONS

VAR Calculation

What, if any, is the practical relevance of the above findings? One way to answer this question is to see if they matter for such a well established procedure as value-at-risk calculation. Let us consider a position long a (heat-crude) crack spread 6 months out on the curve and short the same spread 2 months out. The parameters of the spread are summarized in Table I. We assume that the volatilities of each contract have been recovered from option prices, and that the term structure of volatilities is flat.

In the notation from the General Approach Section, we hold a portfolio $w = (w_1, \ldots, w_n)$ of futures, with each w_i being our position in the ith futures contract times the point value of that contract. We want to calculate the VAR of this position based on the assumptions of the previous section. That is, we want to determine a tail (say, first percentile) of the distribution of profit-and-loss (PNL), given by $w'(F_{t+\Delta t} - F_t)$. In our example and using the setup of the previous section, i ranges from 1 to 10 for crude oil, and from 11 to 20 for heating oil. Our w_is are zero except for i in (2,6,12,16). With lognormal F_{jk}s the distribution of $w'(F_{t+\Delta t} - F_t)$ is not known, so we simulate it using four factors in (6). In that equation, we use the estimated factor loadings V_i from the previous section, and use the percentage weights of each factor for λ_is for the pertinent season (the numbers from Fig. 7). The σs are assumed to be at-the-money implied volatilities from the option market.

Table II shows the risk calculations for the trade in Table I. Column 3 is the one to note. It tells us that our hypothetical position can be 10% bigger in the summer than in the winter before it has the same VAR.

Option Pricing

Another application of our results is pricing correlation-dependent options on petroleum products. The PCA is helpful if, first, the option's

TABLE I

A Hypothetical Heat–Crude Crack Spread Position

Commodity	*Months to Expiration*	*Position*	*Price*	*Point Value*	*1-Month Implied Volatility*
CL	2	100	26.42	1000	28.58
HO	2	−100	93.60	420	27.42
CL	6	−100	25.01	1000	22.88
HO	6	100	66.90	420	22.38

TABLE II
Risks of the Trade in Table I by Season

Season	*Std. Deviation, '000 USD*	*Std. Deviation Relative to Winter*	*1% VAR '000 USD*
Winter	122.31	100.00	−305.12
Spring	112.96	92.35	−284.03
Summer	109.20	89.28	−273.89
Fall	123.11	100.65	−307.74

payoff depends on correlations between many different curve points and/or curves and, second, the option will be priced by Monte Carlo simulation. Under these circumstances, the PCA provides a valuable dimensionality reduction for the Monte Carlo.

Two most typical examples are options on cracks and timespreads (spread options), and options on weighted averages of several contracts (basket options). Asian options and average strike options are widely used contracts whose payoff depends on a linear combination of prices, but they do not depend on crosscommodity or crossexpiration correlations, and are hedged with a single futures expiration. Hence, they fall outside the present framework.

A generic put on a spread pays the difference between the strike and the difference between two prices, if in the money (symmetrically for a call). With zero strike, a spread option is an exchange option studied by Magrabe (1978). One of the earliest treatments of the general case is Wilcox (1990), who assumes that the spread itself follows an arithmetic Brownian motion. A similar approach is taken by Poitras (1998), who starts out with both assets being arithmetic Brownian motions. The resulting option price is similar to that of Wilcox but with explicit no-arbitrage restrictions on the parameters. An excellent reference, in our view, is Shimko (1994), who covers the no-arbitrage conditions, the pricing methodology, as well as some hedging issues, although many details of the presentation are specific to a structural model of futures as derivatives on the cash commodity with stochastic convenience yield. Shimko shows an expression for the price which is suitable for numeric integration, and a closed-form four-moment Edgeworth expansion based on the work of Jarrow and Rudd (1982). If the third and fourth moments of the latter expansion are omitted, the resulting formula is based on the normal distribution and is similar in form (but not in the hedges) to those of Poitras and Wilcox. Pearson (1995) starts with the same numerical integral as Shimko, and shows how the quadrature relates to an approximate hedge with vanilla options. Of these approaches, the numerical integral is the one consistent with lognormal underlying prices, and is quite amenable to evaluation to desired precision. The four-moment Edgeworth expansion can give negative values even for reasonable volatilities and expirations. A normal approximation is quite good at the money, but can give pricing errors in excess of 100% for realistic out/in the money strikes. American spread options can be handled by bivariate trees or PDEs, but the speed of convergence drops noticeably relative to the univariate case.

Numerical methods based on trees and PDEs become all but infeasible once we turn to the more general case of options on linear combinations of several assets. Approximation and Monte Carlo simulation are the alternatives not tailored to the structure of a specific problem. The latter is easy to set up and adopt to a given situation, and is the method we choose for the example below. Among the approximation methods, replacing the arithmetic average with the geometric one was suggested early on. Later work is based on matching the moments of the linear combination, which can be readily computed, to another distribution. Milevsky and Posner (1998) suggested using the inverse gamma distribution for the case when the weights of the linear combination are all positive. Other straightforward possibilities are lognormal (for positive combinations), normal, and mixtures of them. Going one step further, the matching distribution can be expanded in Edgeworth series, as in Jarrow and Rudd (1982), or in Gram–Charlier series. Expansion-based approximations, however, may be very poor for some parameter ranges.

As an example of a basket of spreads, consider a product that at time T caps at K the total of the futures cracks expiring at times $T_1, \ldots, T_i$. The cracks are between commodities 2 (HO) and 1 (CL). At T the option pays

$$\max(0, wF_{2,T_1} - F_{1,T_1} + wF_{2,T_2} - F_{1,T_2} + \cdots + wF_{2,T_i} - F_{1,T_i} - K)$$

where w is a factor converting cents per gallon into dollars per barrel. This option depends on both crack spreads and timespreads. The no-arbitrage price of this option is the discounted risk-neutral expectation of its payoff. Under the risk-neutral measure the drifts of the futures prices are zero, exactly as in (1) and (3), so the terminal prices that enter the option's payoff can be simulated via (6).

Let us narrow the example further to an at-the-money European put on the average of two cracks (long heat, short crude) with times to maturity, prices and volatilities as described in Table I, and with one month to expiration. Table III displays the prices of this option by season. The price of the option drops in the summer, when the contribution of the first factor rises. The price is higher in the spring than in the winter

TABLE III
Option Prices by Season

Season	*Premium '000 USD*	*Premium Relative to Winter*
Winter	0.355	100.00
Spring	0.356	100.33
Summer	0.345	97.16
Fall	0.378	106.53

because the second factor adds more intercommodity volatility. The first factor dominates in the summer, reducing the volatilities of both the timespreads and the cracks. In the fall, the contribution of the first factor drops sharply, leading to reduced correlations between both the commodities and between the 2-month and 6-month maturities.

CONCLUSIONS

The theory of commodity term structure modeling has profited a lot from the developments in the interest rate literature. However, commodities present their own challenges, many of which remain open.

In this article we have extended the analysis done by Cortazar and Schwartz (1994) to the petroleum market. In doing so, we have shown how to use all the information present in the current two-curve term structure of futures prices. Analyzing the futures data for crude oil and heating oil that trade on NYMEX we have found that four factors are enough to explain 99% of the variance present in the multiple curve market. These factors turn out to be natural extensions of the ones found by Cortazar and Schwartz.

We have also studied how to extend the analysis to seasonal commodities. We have found that, in the case of heating oil, the most important factors are still level, steepness, and curvature, but their importances follow a seasonal pattern. On-season periods produce a gain in importance by the second and third factors. We have also noticed that this effect increases when we consider modeling the crack spread (used as a proxy for the added value of producing heating oil). We expect that similar effects would be found in other seasonal commodities such as natural gas.

BIBLIOGRAPHY

Black, F., & Scholes, M. (1973). The pricing of options and corporate liabilities. Journal of Political Economy, 81, 637–654.

Cortazar, G., & Schwartz, E. (1994). The valuation of commodity–contingent claims. The Journal of Derivatives, 1(4), 27–39.

Forzani, L., & Tolmasky, C. F. (2001, December). On the spectral decomposition of empirical correlation matrices. Journal of Knot Theory and Its Ramifications, 10(8), 1201–1214.

German, M. (1995). Spread the load. Risk, 8(5), 68.

Gay, G. D., & Kim, T. H. (1987). An investigation into seasonality in the futures market. Journal of Futures Markets, 7(2), 169–182.

Gibson, R., & Schwartz, E. (1990). Stochastic convenience yield and the pricing of oil contingent claims. Journal of Finance, 45, 959–976.

Girma, P. B., & Paulson, A. S. (1998). Seasonality in petroleum futures spreads. Journal of Futures Markets, 18(5), 581–598.

Harrison, J. M., & Kreps, D. M. (1979). Martingales and arbitrage in multiperiod security markets. Journal of Economic Theory, 20, 381–408.

Harrison, J. M., & Pliska, S. (1981). Martingales and stochastic integrals in continuous trading. Stochastic Processes and Their Applications, 11, 215–260.

Heath, D., Jarrow, R., & Morton, A. (1990). Bond pricing and the term structure of interest rates: A discrete time approximation. Journal of Financial and Quantitative Analysis, 25, 419–440.

Heath, D., Jarrow, R., & Morton, A. (1992). Bond pricing and the term structure of interest rates: A new methodology for contingent claims evaluation. Econometrica, 60(1), 77–105.

Jarrow, R., & Rudd, A. (1982). Approximate option valuation for arbitrary stochastic processes. Journal of Financial Economics, 10, 347–369.

Kshirsagar, A. M. (1972). Multivariate analysis. New York: Marcel Dekker.

Litterman, R., & Scheinkman, J. (1991). Common factors affecting bond returns. Journal of Fixed Income, 1, 54–61.

Magrabe, W. (1987). The value of an option to exchange one asset for another. Journal of Finance, 33, 117–186.

Milevsky, M. A., & Posner, S. E. (1998). A closed-form approximation for valuing basket options. The Journal of Derivatives, 5(4), 54–61.

Milonas, N. T. (1991). Measuring seasonalities in commodity markets and the half-month effect. Journal of Futures Markets, 11(3), 331–346.

Milonas, N. T. (1987). The effects of USDA crop announcements on commodity prices. Journal of Futures Markets, 7, 571–589.

Pearson, N. D. (1995). An efficient approach for pricing spread options. The Journal of Derivatives, 3(1), 76–91.

Poitras, G. (1998). Spread options, exchange options, and arithmetic Brownian motion. The Journal of Futures Markets, 18(5), 487–517.

Reismann, H. (1992). Movements of the term structure of commodity futures and pricing of commodity claims (Working Paper). Haifa, Israel: Faculty of I.E. and Management, Technion-Israel Institute of Technology.

Shimko, D. S. (1994). Options on futures spreads: Hedging, speculation, and valuation. The Journal of Futures Markets, 14, 183–213.

Wilcox, D. (1990). Energy futures and options: Spread options in energy markets. New York: Goldman Sachs.

Copyright of Journal of Futures Markets is the property of John Wiley & Sons, Inc. / Business and its content may not be copied or emailed to multiple sites or posted to a listserv without the copyright holder's express written permission. However, users may print, download, or email articles for individual use.