VOLATILITY DYNAMICS AND SEASONALITY IN THE NYMEX ENERGY COMPLEX

Hiroaki Suenaga
School of Economics and Finance, Curtin University of Technology

Aaron Smith
Department of Agricultural and Resource Economics, University of California, Davis

ABSTRACT

We model the volatility dynamics of the three major petroleum commodities traded on the NYMEX: crude oil, unleaded gasoline, and heating oil. Using the partially overlapping time-series (POTS) framework of Smith (2005), we model jointly all contracts with maturity dates up to a year into the future. In so doing, we capture the nonlinearities induced by seasonality and storage. Our model produces volatility dynamics that are consistent with the observed seasonality in demand and storage of the three commodities. The results show that the short-horizon price variance in the crack spread is twice as large late in the peak-demand season as in the low-demand season. Long-horizon crack-spread price risk is much lower than short-horizon risk in all seasons, but it is still substantial in magnitude. The variance of long-horizon crack-spread shocks is about 0.5 percent per day, compared to about 3 percent for long-horizon shocks to the price of crude oil and about 8 percent for short-horizon shocks to the crack spread. Crack spread hedgers ignore high-season nearby price risk at their peril, but they would also be remiss to ignore the long horizon.

1. INTRODUCTION

Demand for motor gasoline in the United States peaks in the summer driving season, whereas demand for heating oil peaks in winter. Because refined petroleum products are imperfect substitutes in the production process, this mismatched seasonal pattern leads oil refiners to produce more gasoline than is needed in the winter and more heating oil than is needed in the summer. The resulting seasonal storage cycle induces complex dynamic relationships between the prices of crude oil and the two refined products. For example, during the inventory accumulation period, current and future prices for a particular commodity are linked by temporal arbitrage. However, late in the peak-demand season, a tight supply-demand balance means that short-term shocks often cannot be mitigated through storage. In this event, the temporal arbitrage link is broken and futures prices for delivery after the peak season will be unaffected by current price shocks. These price patterns spill over to the other two commodities.

Numerous recent studies have examined the co-variability of crude oil and refined petroleum product prices (Asche et al., 2003; Girma and Paulson, 1999; Haigh and Holt, 2002; Lanza et al., 2005; Ng and Pirrong, 1996; Serletis, 1994). These studies have commonly employed the cointegration approach and found that each petroleum commodity price is non-stationary or I(1) while sharing common stochastic trends. Although this method is useful for depicting the long-run equilibrium relationship of cointegrated time series and how each series responds to the deviation from the equilibrium, most of previous studies presume unrealistic distributional structures of petroleum prices. In particular, an application of the standard cointegration approach presumes that the cointegration vector is time-invariant even though demand seasonality differences across commodities imply that the equilibrium price relationship of these commodities exhibits seasonal variation. Furthermore, the standard error correction model presumes that stochastic dynamics of the short-term deviation to be time-invariant. These assumptions cannot capture the price dynamics induced by seasonality and storage.

Aside from modeling deficiencies, most of these studies have used only a subset of the available price data. Multiple contracts with different maturity dates trade simultaneously on futures exchanges, but a common practice is to construct a single time series by splicing the nearby futures prices and stacking them together. Such practice discards much information available from the markets about the persistence of price shocks. It also distorts the temporal

dynamics of futures prices by switching from one contract to another at the end of every calendar month (Smith 2005).

In this paper, we apply the partially overlapping time-series (POTS) model of Smith (2005), which decomposes the daily returns into common factors that affect the prices of all simultaneously traded contracts. The model allows the factor loadings and the variance of the idiosyncratic errors to be flexible non-parametric functions so as to capture time-to-maturity effects, seasonal and cross-sectional variations, and other non-linear volatility dynamics resulting from the peculiarities of the commodities. We extend Smith's original model into a three-commodity, six-factor setting so that each of three commodities is subject to two common factors: short- and long-term.

The rest of the paper is organized as follows. Section 2 reviews the peculiarities of three petroleum commodity markets in the US and identifies key features expected for their price volatility. Section 3 develops the model that is capable of capturing these features. Section 4 describes the data and presents the estimation results with emphasis on their implications for the volatility dynamics of three commodities and three widely considered crack spread positions. Section 5 concludes the paper.

2. SEASONALITY IN PETROLEUM PRODUCT DEMAND AND ITS IMPLICATIONS FOR VOLATILITY DYNAMICS

Three market features cause petroleum commodities to exhibit complex stochastic dynamics. First, demand for motor gasoline and heating oil in the US exhibits strong seasonality with the former peaking in the summer driving season and the latter in winter for space heating. Second, finished petroleum products, manufactured from crude oil and other hydrocarbons, are imperfect substitutes in the refining process. A conventional refining process heats crude oil in a distillation column and extracts motor gasoline, distillate fuel oil, and other refined petroleum products at different temperatures. In the early 1980's, refineries developed the capacity for secondary refining to reprocess the heavier fractions of refined products into more valuable, lighter products. However, such alteration of the production mix incurs additional cost. Besides, the capacity of coker or downstream processing cannot be expanded in short run, providing refiners with only a limited flexibility in the proportion of the finished products extracted from a unit of crude oil.

Third, environmental regulations require the specifications of motor gasoline to be altered across seasons (EIA, 1997). In particular, NYMEX defines three different commodity specifications; the Summer grade (delivered from April to mid September), the Winter grade (November to the end of February), and the intermediate specification (delivered in the two transition periods). The shift in product specification requires refiners to clear inventory and adjust their production schedules at or prior to the change in the product specification. Together with the limited substitutability in refining process, this regulatory requirement affects refiners' production and storage decisions of other refined petroleum products.

Figure 1 shows that the supply of heating oil and motor gasoline exhibits seasonal variation that is similar in pattern but much smaller than the seasonal variation observed for their demand.[1] For both commodities, monthly consumption generally exceeds net production in their peak demand seasons, the difference of which is supplemented by releasing inventory. For heating oil, this is seen in a rapid decline of inventory from December to March. Inventory reaches its lowest point around the end of winter. It then gradually increases from March to August and stays around the same level until it is depleted during the subsequent winter.

For motor gasoline, inventory is the lowest in August right after the summer driving season. It is also low in October in transition of the commodity specification from the summer to winter grade. Inventory starts building up in November and increases until it reaches an annual high typically in January. During this post-summer season, gasoline production is kept high because a large amount of heating oil needs to be co-produced and stored to meet high winter demand. High gasoline inventory accumulated in this period is released during the subsequent winter before the specification switches to the summer grade in mid March. Since the demand is low throughout winter, contemporaneous supply is kept low and many refineries perform regular maintenance in this period. The production level starts increasing in April and peaks in July. Over the same period, the inventory increases initially and then quickly depletes to meet with high demand over summer driving season.

While the seasonal demand patterns observed for the two refined products themselves imply higher average prices in summer for unleaded gasoline and in winter for heating oil, inter-temporal price differentials are smoothed out, albeit imperfectly, through storage. According to

[1] **Figure 1** illustrates monthly averages of demand (product supplied for end use and for refinery and blender plus exports), supply (US domestic production plus imports), and inventory for each of the three commodities, over the period between 1981 and 2008. For heating oil, the figure plots the monthly averages of demand, supply, and storage of distillate fuel oil.

the theory of storage (e.g. Williams and Wright, 1991), the equilibrium constellation of spot and futures prices represents the prices at which the marginal benefit of current consumption is equal to or above the expected marginal benefit of future consumption plus the cost of carry. The weak inequality stems from the constraint that firms cannot borrow from the future when the market experiences a shortage of physical commodity. This constraint creates a discontinuity in the inter-temporal price linkage; two prices are linked by temporal arbitrage when inventory is plentiful whereas they are disconnected when discretionary inventory is scarce.

The data in Figure 1 indicate that average inventory ranges from 20-24 days of supply for gasoline and 31-37 days of supply for heating oil, depending on the season. Average crude oil inventories range from 23-24 days of supply. The span of these ranges reflects the seasonality in demand for each commodity. Overall, the number of days supply in inventory is low for the energy complex compared to many other commodities. For example, US corn inventories average about three months of supply and gold inventories average much more than a year of supply. Low average inventory levels suggest a somewhat limited ability of above-ground storage to smooth price shocks, even in seasons with relatively high inventory. Thus, shocks to current supply and demand will not transmit equally to futures prices with distant maturity dates. These shocks will affect prices with nearby delivery dates more than prices with distant delivery dates, thereby inducing some mean reversion in spot prices and implying greater volatility in nearby prices than distant prices. This feature of commodity markets is known as a Samuelson effect (Samuelson, 1965), or a time-to-delivery effect.

The observed seasonality in demand and storage also creates nonlinear volatility dynamics of three commodity prices. Heating oil prices tend to be more volatile in December through March because the high marginal cost of production together with price inelastic demand means that demand and supply shocks of even a small magnitude can cause a large price swing. High inventory in early winter allows short-term demand shocks to be absorbed though releasing inventory. Such flexibility is inevitably low toward the end of winter and early spring when inventory is low. For unleaded gasoline, the price tends to be more volatile in July and August when demand peaks and in October and March when inventory is low due to the switching of product specifications. For both commodities, the inter-temporal price linkage breaks at the end of demand season when the inventory is effectively zero and, for motor gasoline, when the commodity specification changes. Thus, short-term market shocks prior to the stock out should not affect prices of futures contracts for post season delivery.

In the next section, we specify a model of futures prices that captures these dynamic features of energy prices. Specifically, it captures four effects: (1) seasonal and temporal variation in the link between futures prices with nearby and distant delivery dates, (2) a time-to-delivery effect, (3) seasonal variation in the level of volatility, and (4) time variation in cross-commodity correlation. In addition, the model permits volatility clustering, which has been documented for many commodity markets.

3. ECONOMETRIC MODEL

In commodity futures markets, multiple contracts trade simultaneously. These contracts differ by time to delivery. Each month, some contracts reach delivery and cease to exist, while others are born and begin trading. From an econometric perspective, a set of futures prices presents a potentially large number of partially overlapping time series. The partially overlapping time-series (POTS) model of Smith (2005) models jointly all contracts trading on a given day. The POTS model is a factor model for partially overlapping time series that incorporates time varying conditional heteroskedasticity and time and cross-sectional variation in the factor loadings and innovation variances. It seeks to explain the price changes on all traded contracts with a small number of components, while accounting for time-to-delivery effects, storability, and seasonality.

We extend Smith's two-factor, single commodity model to a six-factor model of three petroleum commodity prices. We index the price of a futures contract, F, with two subscripts: d represents the number of trading days until the first day of the delivery month and t represents the date on which a particular price observation occurs. We use the superscript $j \in \{\text{CO, HO, UG}\}$ to denote the commodity. This (d, t, j) triple is sufficient to identify any price observation in the sample. The model is

$$\Delta \ln F^j_{d,t} = \theta^j_{1,d,t}\varepsilon^j_{1,t} + \theta^j_{2,d,t}\varepsilon^j_{2,t} + \theta^j_{3,d,t}u^j_{d,t}$$

where $\Delta \ln F^j_{d,t} = \ln F^j_{d,t} - \ln F^j_{d+1,t-1}$ denotes the log price change on date t of the futures contract of commodity j for delivery at date t+d. The functions $\theta^j_{1,d,t}$ and $\theta^j_{2,d,t}$ represent the factor loadings, and $\theta^j_{3,d,t}$ represents the innovation standard deviation, each of which are deterministic functions of d and t. The factors $\varepsilon^j_{1,t}$ and $\varepsilon^j_{2,t}$ capture the common components across contract maturities,

and the idiosyncratic term $u^j_{d,t}$ captures the component of $\Delta \ln F^j_{d,t}$ that is uncorrelated with the factors. If the two factors were to explain all of the variation in observed prices, then $u^j_{d,t}$ would equal zero for all *t*. In essence, the POTS model uses two unobserved random factors to explain as much of the observed variation in prices as possible. It gains its richness by permitting the factor loadings, which connect the factors to observed prices, to be flexible functions of the season and time-to-delivery. It also gains richness by allowing time varying conditional correlations between the factors and time-varying conditional volatility of the factors.

We define the first factor for each commodity to be a long-run factor and the second factor to be a short-run factor. We generate this interpretation by setting $\theta^j_{2,d,t} = 0$ for contracts that are far from delivery, i.e., a short-run shock today cannot affect contracts that mature on a distant date in the future. The condition is equivalent to the one used by Schwartz and Smith (2000) to identify their two-factor model of commodity price dynamics.

Stacking all observed prices into a single vector, the model can be written as

(1) $\mathbf{\Delta ln F}_t = \mathbf{\Theta}_{1,t}\, \boldsymbol{\varepsilon}_{1,t} + \mathbf{\Theta}_{2,t}\, \boldsymbol{\varepsilon}_{2,t} + \mathbf{\Theta}_{3,t}\, \mathbf{u}_t$

or, in long form as

$$\begin{bmatrix} \Delta \ln \mathbf{F}^{CO}_t \\ \Delta \ln \mathbf{F}^{HO}_t \\ \Delta \ln \mathbf{F}^{UG}_t \end{bmatrix} = \begin{bmatrix} \Theta^{CO}_{1,t} & 0 & 0 \\ 0 & \Theta^{HO}_{1,t} & 0 \\ 0 & 0 & \Theta^{UG}_{1,t} \end{bmatrix} \begin{bmatrix} \varepsilon^{CO}_{1,t} \\ \varepsilon^{HO}_{1,t} \\ \varepsilon^{UG}_{1,t} \end{bmatrix} + \begin{bmatrix} \Theta^{CO}_{2,t} & 0 & 0 \\ 0 & \Theta^{HO}_{2,t} & 0 \\ 0 & 0 & \Theta^{UG}_{2,t} \end{bmatrix} \begin{bmatrix} \varepsilon^{CO}_{2,t} \\ \varepsilon^{HO}_{2,t} \\ \varepsilon^{UG}_{2,t} \end{bmatrix} + \begin{bmatrix} \Theta^{CO}_{3,t} & 0 & 0 \\ 0 & \Theta^{HO}_{3,t} & 0 \\ 0 & 0 & \Theta^{UG}_{3,t} \end{bmatrix} \begin{bmatrix} \mathbf{u}^{CO}_t \\ \mathbf{u}^{HO}_t \\ \mathbf{u}^{UG}_t \end{bmatrix}$$

In (1), $\mathbf{\Delta ln F}_t$ is an $n_t \times 1$ vector of daily futures log price changes observed on date *t* where $n_t = n^{CO}_t + n^{HO}_t + n^{UG}_t$ with n^j_t representing the number of futures contracts of commodity *j*. The $n^j_t \times 1$ sub-vector, $\Delta \ln \mathbf{F}^j_t$, is comprised of $\Delta \ln F^j_{d,t} = \ln F^j_{d,t} - \ln F^j_{d+1,t-1}$. We denote the factors by $\varepsilon_t = \left[\varepsilon^{CO}_{1,t} \quad \varepsilon^{HO}_{1,t} \quad \varepsilon^{UG}_{1,t} \quad \varepsilon^{CO}_{2,t} \quad \varepsilon^{HO}_{2,t} \quad \varepsilon^{UG}_{2,t} \right]'$, which is a 6 × 1 vector of latent factors with E[$\boldsymbol{\varepsilon}_t$] = 0 and E[$\boldsymbol{\varepsilon}_t \boldsymbol{\varepsilon}_t'$] = $\mathbf{\Omega}$ where $\mathbf{\Omega}$ is the unconditional correlation matrix with diagonal elements of unity and off-diagonal elements s_{jk} ($j \neq k$). The idiosyncratic term $\mathbf{u}_t$ is an $n_t \times 1$ vector with $\mathbf{u}_t \sim N(0, I_{n_t})$ where I_{n_t} is an identity matrix of dimension n_t. For identification, we assume E[$\varepsilon^j_{i,t} u^k_{d,t}$] = 0 for all *i*, *j*, *k*, *d*, and *t*. These assumptions together imply that E[$\mathbf{\Delta ln F}_t \,|\, \Im^{t-1}$] = 0 where $\Im^{t-1}$ denotes the

information set available at $t - 1$. That is, a series of daily futures price changes follows a martingale difference sequence, which implies a zero risk premium.[2]

For the two factors, $\Theta_{1,t}$ and $\Theta_{2,t}$ are $n_t \times 3$ block diagonal loading matrices. For the idiosyncratic term, $\Theta_{3,t}$ is an $n_t \times n_t$ diagonal matrix that determines the idiosyncratic standard deviation. We specify the components *of* $\Theta_{i,t}$ as deterministic functions of the delivery month of contract (m = 1, …, 12) and time to maturity (d). This setup allows a particular shock to affect prices differently depending on time to maturity and season. For example, short-run shocks such as hurricanes have a greater effect on nearby prices than those with more distant delivery dates. We consider the following three specifications,

$$\text{(2a)} \quad \theta^j_{i,d,t} = \exp\left[\lambda^j_i(d)\right] = \exp\left[a^j_{i,0} + a^j_{i,1}d + \sum_{k=1}^{K}\left(a^j_{i,2k}\sin\left(\frac{2\pi kd}{d_{\max}}\right) + a^j_{i,2k+1}\cos\left(\frac{2\pi kd}{d_{\max}}\right)\right)\right] \quad \text{… Spec 1}$$

$$\text{(2b)} \quad \theta^j_{i,d,t} = \exp\left[\lambda^j_i(m,d)\right] = \exp\left[a^j_{i,m,0} + a^j_{i,m,1}d + \sum_{k=1}^{K}\left(a^j_{i,m,2k}\sin\left(\frac{2\pi kd}{d_{\max}}\right) + a^j_{i,m,2k+1}\cos\left(\frac{2\pi kd}{d_{\max}}\right)\right)\right] \text{… Spec 2}$$

$$\text{(2c)} \quad \theta^j_{i,d,t} = \exp\left(\lambda^j_{1,i}(d)\right)\exp\left(\lambda^j_{2,i}(t)\right)\left(1+\exp\left(\lambda^j_{2,i}(t)\right)\right)^{-1}$$

$$\lambda^j_{1,i}(d) = a^j_{i,0} + a^j_{i,1}d + \sum_{k=1}^{K}\left(a^j_{i,2k}\sin\left(\frac{2\pi kd}{d_{\max}}\right) + a^j_{i,2k+1}\cos\left(\frac{2\pi kd}{d_{\max}}\right)\right)$$

$$\lambda^j_{2,i}(t) = \sum_{k=1}^{K}\left(b^j_{i,2k}\sin\left(\frac{2\pi ks(t)}{365}\right) + b^j_{i,2k+1}\cos\left(\frac{2\pi ks(t)}{365}\right)\right) \quad \text{… Spec 3}$$

where d_{max} is the maximum days to maturity for which the model is estimated, and $s(t)$ is day in year (calendar day) of t.

The three specifications allow different levels of flexibility in modelling the volatility dynamics of futures prices of three commodities. Spec 1 is the most restrictive of three specifications. It specifies the factor loadings and the variance of idiosyncratic error as functions of time-to-delivery only and assumes each of three functions (i = 1, 2, and 3) to be common to all contract delivery months (m). Spec 1 does not allow for seasonality. Spec 2 is the most flexible of three specifications. Like Spec 1, it specifies the factor loadings and the variance of idiosyncratic error as functions of time-to-maturity, yet it specifies a different set of three functions for each delivery month. Spec 3 is an intermediate specification. It specifies the factor loadings and the variance of idiosyncratic error as products of the two components. The first component,

[2] We also estimated a model in which we allowed the factors and the idiosyncratic errors to have non-zero mean. The estimated mean parameters were very small for all three commodities and had no effect on the results we present in the

$\exp(\lambda^j_{1,i}(d))$, captures the time-to-maturity effect whereas the second component, $\exp\left(\lambda^j_{2,i}(t)\right)\left(1+\exp(\lambda^j_{2,i}(t))\right)^{-1}$, captures seasonal variation. To identify the second factor as the short-run factor, we set $\theta^j_{2,t}(d_{\max}) = 0$, i.e., the loading on the second factor loading is zero at d = d_{max}. We impose this constraint as $a^j_{2,0} = \sum_{k=1}^{K} a^j_{0,2k+1} - \lambda_{\min}$, $a^j_{2,m,0} = \sum_{k=1}^{K} a^j_{0,m,2k+1} - \lambda_{\min}$, and $a^j_{2,0} = \sum_{k=1}^{K} a^j_{2,2k+1} - \lambda_{\min}$, respectively, for three specifications. We set λ_{min} = 10 so that, for all three specifications, $\theta^j_{2,d_{\max},t} = \exp[\lambda^j_2(d_{\max})] = \exp(-10) \approx 0$.

For the conditional variance of $\boldsymbol{\varepsilon}_t$, we specify a multivariate GARCH(1,1) model with dynamic conditional correlation (DCC), as in Engle (2002) and Engle and Sheppard (2001),

(3) $\quad \mathrm{E}[\boldsymbol{\varepsilon}_t \boldsymbol{\varepsilon}_t' \mid \Im^{t-1}] = \mathbf{H}_t = \mathbf{D}_t \mathbf{R}_t \mathbf{D}_t$

where

$$\mathbf{D}_t = diag(\sqrt{h^{CO}_{1,t}}, \sqrt{h^{CO}_{2,t}}, \sqrt{h^{HO}_{1,t}}, \sqrt{h^{HO}_{2,t}}, \sqrt{h^{UG}_{1,t}}, \sqrt{h^{UG}_{2,t}})$$

$$h^j_{i,t} = \varpi^j_i + \beta^j_i h^j_{i,t-1} + \alpha^j_i \mathrm{E}[(\varepsilon^j_{i,t-1})^2 \mid \Im^{t-1}]$$

$$\mathbf{R}_t = (1 - \gamma_1 - \gamma_2)\, \boldsymbol{\Omega} + \gamma_1\, \boldsymbol{\eta}_{t-1}\boldsymbol{\eta}_{t-1}' + \gamma_2\, \mathbf{R}_{t-1}$$

$$\boldsymbol{\eta}_t = \mathbf{D}_t^{-1}\, \boldsymbol{\varepsilon}_t$$

The matrix $\mathbf{R}_t = (\rho_{jk,t})$ is symmetric and positive definite with $\rho_{jk,t} = 1$ for $j = k$ and $\rho_{jk,t} = \rho_{kj,t}$ for $j \neq k$. Because the unconditional variance of $\varepsilon^j_{i,t}$ equals unity, $\varpi^j_i = 1 - \alpha^j_i - \beta^j_i$, $\forall$ i and j. The conditional expectation in the expression for $h^j_{i,t}$ is

(4) $\quad \mathrm{E}[(\varepsilon^j_{i,t-1})^2 \mid \Im^{t-1}] = (\varepsilon^j_{i,t-1|t-1})^2 + P^j_{i,t-1|t-1}$

where $\varepsilon^j_{i,t-1|t-1} = \mathrm{E}[\varepsilon^j_{i,t-1} \mid \Im^{t-1}]$ and $P^j_{i,t-1|t-1} = \mathrm{E}[(\varepsilon^j_{i,t-1} - \varepsilon^j_{i,t-1|t-1})^2 \mid \Im^{t-1}]$, which are obtained through the Kalman filter (Hamilton, 1994).

The POTS model in (1)-(4) is econometrically similar to some other factor models of commodity price dynamics. See Lautier (2005) for a comprehensive review of the application of these models to commodities and Gibson and Schwartz (1990), Cortazar and Naranjo (2006), Schwartz (1997), Todorova (2004), and Schwartz and Smith (2000) for applications to crude oil. The POTS model differs from the models in this literature in two important ways. First, the POTS model specifies the dynamics of daily price *changes* whereas conventional factor models specify

---

remainder of the paper.

the dynamics of daily price *levels*. By modeling daily price *changes*, we avoid specifying seasonal or any other deterministic variations in the underlying spot price level and temporal dynamics of the deviation from such deterministic price level. The model is, hence, free from the bias in approximating such a deterministic variation.

Second, our model employs flexible functional forms in specifying the factor loadings ($\Theta_{1,t}$, $\Theta_{2,t}$) and the variance of the idiosyncratic term ($\Theta_{3,t}$). In contrast, conventional factor models specify stochastic processes of latent factors with a small number of parameters, thereby imposing rather restrictive structure on the factor loadings in the futures pricing equation. In particular, the two-factor model of Schwartz and Smith (2002) specifies the long- and short-term factor to follow the geometric Brownian motion and mean-reversion process, respectively. Such a specification imposes that, for all contracts, the factor loading is unity for the first factor and decays exponentially at identical rate for all contracts with the time-to-maturity of contract for the second factor. Such specification is incapable of capturing seasonal, cross-sectional, and/or any other deterministic variations in the factor loadings among simultaneously traded contracts with different maturity dates.

4. DATA AND ESTIMATION

4.1 Data

We estimate the model in (1)-(4) using daily settlement price data from the NYMEX crude oil (CO), unleaded gasoline (UG), and heating oil (HO) futures markets. All prices are converted into natural log. Contracts traded in these markets are defined in monthly blocks and prices are quoted per barrel of crude oil and per gallon of unleaded gasoline and heating oil. Crude oil contracts trade for delivery as far as seven years whereas heating oil and unleaded gasoline contracts trade as far as 18 and 12 months, respectively. These contracts trade until the third business day prior to the 25th calendar day of the month prior to the delivery month for crude oil and until the last business day of the month prior to the delivery month for two refined commodities. We use data from December 1, 1984, the date that crude oil and unleaded gasoline markets were opened, to December 31, 2006.[3] Because distant delivery contracts are not always

[3] The Heating Oil futures contract was introduced in 1978.

actively traded, we exclude contracts of more than 12 months to maturity from our sample. Excluding these observations leaves a sample of 185,023 prices among 792 contracts.

Estimating the vector of the daily log-price changes of three commodities requires intensive computing due to a large sample size and a large number of parameters defining the model. For example, when $k = 3$ the model includes 857 parameters for specification 2. Therefore, following Engle (2002), we estimate the model in two steps. First, we construct three sub-models, each comprised of the daily log-price changes of a single commodity, ignoring the correlation coefficients, ρ's. Setting the number of trigonometric terms (k) equal to 3 and $d_{max} = 252$, the number of trade days per calendar year, we estimate each of three sub-models by the method of Maximum Likelihood. Second, we estimate the DCC process (3), using the predicted values of six latent factors and their conditional variances from the first step. This procedure yields less efficient but consistent estimators of model parameters (Engle, 2002).

4.2 Estimation Results

Model Selection

Table 1 shows the Schwartz Information Criterion calculated for each specification of the factor loading functions in (2). For all three commodities, Spec 2 with the maximum flexibility of the factor loadings and the variance of idiosyncratic error is supported over the other two specifications. Of the two restrictive specifications, Spec 3 is supported over Spec 1. The results indicate that not only does volatility exhibit significant seasonal variation and time-to-maturity effects for three petroleum commodities, but also that the time-to-maturity effect varies across the twelve delivery months. The three specifications in (2) become more flexible as k increases. Although this extra flexibility allows the model to fit the observed data better, it also makes the coefficient estimates more sensitive to extreme observations. By using only six trigonometric terms, we allow sufficient flexibility to capture seasonality and time-to-delivery effects while avoiding excess sensitivity to outliers. Given these results, we focus on the results obtained for Spec 2 and k=3 in our subsequent discussions on the volatility dynamics of the NYMEX petroleum futures.

Seasonality in Total Variance

Figure 2 plots the unconditional variance of daily returns as implied by the estimated model. They are computed as $\sum_{i=1}^{3} \hat{\theta}^{j}_{i,m,m-d}{}^{2}$ for each of three commodities (j = CO, HO, UG), each of twelve delivery months (m = 1, …, 12), and d ranging from 0 to 252 days before maturity. The figure reveals four features about the volatility of daily returns of three commodities. First, in general, the unconditional variance of daily returns increases rapidly as the contract approaches maturity. This feature, often called the Samuelson effect (Samuelson, 1965) or the time-to-delivery effect, implies that shocks to the underlying spot prices are mean-reverting, that is, they are expected to dissipate over time.

Second, volatility in the last two months of trading exhibits substantial seasonality for the two refined commodities. For heating oil, volatility in the last two months of trading is higher for the contracts maturing in late winter to early spring than the other contracts. It is particularly high for the December through March contracts because the demand for heating oil peaks in winter months and demand and supply shocks of even a small magnitude can cause a large price swing. Volatility is also high for the April contract because low inventory after the peak demand season requires that, in years with extended cold weather, higher than usual demand be met mostly by contemporaneous supply. For unleaded gasoline, the unconditional variance in the last two months of trading is substantially higher for the September and October contracts because gasoline inventory is depleted by the end of the summer driving season, and the commodity specification switches from the summer to intermediate grade in mid-September and then to the winter grade at the end of October.

Third, although the two refined petroleum commodities exhibit different seasonal patterns in their volatility, high volatility of one commodity appears partially transmitted into the other commodity market. The September and October contracts exhibit high volatility not only for unleaded gasoline but also for heating oil, even though demand for heating oil is only moderate in these two months. Similarly, despite low winter demand, volatility increases for the January through March gasoline contracts from mid-November to mid-January. This pattern likely reflects the limited flexibility of the production mix. For example, higher than usual September demand for gasoline can be met by increasing production, which is associated with high co-production of heating oil, which in turn affects the heating oil price.

Fourth, seasonal variation in the volatility of heating oil and unleaded gasoline futures in the last few months of trading is not reflected in crude oil markets. Thus, price shocks during this period are mostly demand driven. They result from short-term shocks to refinery sector and/or to the demand of these refined products that are mitigated by above ground storage.

Decomposition of Petroleum Price Variance

Table 2 presents, for each of three commodities, the proportion of the total variance explained by the two factors, $\varepsilon_{1,t}^{j}$ and $\varepsilon_{2,t}^{j}$. For each delivery month m = 1, …, 12, we calculate this proportion by $\left(\sum_{t}\left(\hat{\theta}_{1,m,m-d}^{j}\hat{\varepsilon}_{1,t}+\hat{\theta}_{2,m,m-d}^{j}\hat{\varepsilon}_{2,t}\right)\right)\left(\sum_{t}\Delta\ln F_{m,t}^{j}\right)^{-1}$, and we use $\left(\sum_{m=1}^{12}\sum_{t}\left(\hat{\theta}_{1,m,m-d}^{j}\hat{\varepsilon}_{1,t}+\hat{\theta}_{2,m,m-d}^{j}\hat{\varepsilon}_{2,t}\right)\right)\left(\sum_{m=1}^{12}\sum_{t}\Delta\ln F_{m,t}^{j}\right)^{-1}$ for the overall share. The table shows that the model explains a large share, on average 97.8%, 97.3%, and 95.4%, of variation in the daily returns of crude oil, heating oil, and unleaded gasoline futures, respectively. The share of the total variance accounted for by the two common factors is almost identical for all twelve crude oil contracts, whereas it is lower than average for the high demand season in the two refined commodities. Specifically, there is more idiosyncratic variation in the January through April heating oil contracts and in the September through November unleaded gasoline contracts.

Figure 3 illustrates, for each of three commodities and for each of twelve contracts, how the proportion of total variance accounted for by the two common factors changes over the one year trading horizon. We measure this proportion by $\sum_{i=1,2}\hat{\theta}_{i,m,m-d}^{j\ 2}/\sum_{i=1}^{3}\hat{\theta}_{i,m,m-d}^{j\ 2}$ for commodity j, m = 1, …, 12, and d = 0, …, 252. For all three commodities and for all twelve contracts, this share drops substantially in the last few months of trading. That is, much of the high volatility in this period is unrelated to the price movements of more distant contracts that are concurrently traded. The share explained by the factors is particularly low in the last few months of trading for the January through April heating oil contracts and the September through November unleaded gasoline contracts. These are the same contracts with the greatest increase in volatility in the last two months of trading, which reveals that this volatility increase comes from sources uncorrelated with the shocks that affect longer-dated contracts.

For heating oil, this period corresponds to the end of winter peak-demand season when inventory is inherently low. At this time, demand, supply and other market shocks affect only the

prices of contracts maturing before the end of the peak demand season and not those maturing after the demand season. Similarly, gasoline stock is naturally low in September, which comes at the end of the summer driving season and just before the specification switches from the summer to winter grade. Low inventory during this period breaks the inter-temporal price linkage between contracts near maturity and more distant contracts, resulting high volatility of the September through November contracts that is unrelated to the two factors. For both refined petroleum commodities, this result shows that moderate inventory of physical commodity toward the end of peak-demand season provides only limited price buffering.

Figure 4 decomposes the unconditional variance of futures prices into three sources; short- and long-term factors, and idiosyncratic shocks, and presents how these three components change over the one-year trading horizon. For illustration, the figure presents this decomposition only for the February and September contracts, which exhibit the highest volatility near maturity for heating oil and unleaded gasoline, respectively. Our identification condition specifies that short-term shocks have zero effect on the price for delivery one year hence. It follows that the short-run factor captures intra-year shocks, those that affect prices for delivery within a year, but are uncorrelated with long-run shocks. The idiosyncratic term captures any price movements uncorrelated with the two factors and, as Figures 3 and 4 show, it appears mostly in the last two months of trading.

Figure 4 shows that, for all three commodities, virtually all futures price volatility around six months before maturity and earlier emanates from the long-term factor. The long-term shock exhibits weak, if any, seasonality for all twelve delivery months including those not presented in **Figure 4**. This weak seasonality is evident in **Figure 4** through the fact that the long-run factor contributes a similar amount of price variance to the February contracts as to the September contracts. It is also notable that the variance contribution of the long-run factor is of a similar magnitude across the three commodities. These patterns imply that, on average, the long-run factor reflects non-seasonal supply and demand shocks that are transmitted across the commodities.

For all three commodities, much of volatility increase in the last four months of trading emanates from the short-term factor and the idiosyncratic term. For crude oil, these two components exhibit little seasonal variation; for all twelve delivery months, the short-term shock starts increasing around three months before maturity whereas the idiosyncratic error increases more rapidly in the last two months of trading. In contrast, the two components exhibit

substantial seasonal variation for heating oil and unleaded gasoline. For the February heating oil contract, the short-term shock starts increasing in September, around five months before maturity. The same seasonal pattern is also observed for the other winter contracts (December through March) not shown in **Figure 4**, for which the short-tem shock starts increasing as early as July but more rapidly over September through January. For the two post-winter contracts (April and May), the short-term shock becomes significant in mid-December. For the June through November contracts, the short-term shock starts increasing around three to four months before maturity, before which it has almost no impact on the futures price of these contracts. The short-term and the idiosyncratic shocks are also much more volatile for the last two months of trading in the winter and spring contracts than for the last two months of trading in the other contracts.

The observed seasonality is consistent with the seasonal pattern in demand and storage of heating oil. Weather in fall through early winter determines the amount of inventory carried into the winter peak-demand season. Since the inventory accumulated in this period is not carried over to the post peak-demand season in a normal year, volatility starts increasing for the winter contracts early fall and is captured by the short-term shocks. In the same period, the contracts maturing after the subsequent winter exhibit no significant movement. The January through March contracts exhibit high volatility at the end of winter peak-demand season due to tight demand-supply balance and low inventory. This high volatility is captured by the idiosyncratic error, which we can think of as capturing very-short-run shocks. Demand, supply, and other market shocks in late winter also affect the contracts for post-winter delivery because they determine the amount of inventory carried over to early spring in the case of mild winter. These impacts are rather small and are captured by the short-term factor, which starts increasing as early as mid December for the April and May contract.

The short-term and idiosyncratic shocks also exhibit a strong seasonal pattern for unleaded gasoline. For the August and September contracts, the short-term shock starts increasing in importance around mid March when the commodity specification switches to the summer grade. For the other ten contracts, the short-term shocks do not start increasing until around four months before maturity. The idiosyncratic shocks are also greater for the September contract than for the February contract, mainly because demand needs to be met only by cotemporaneous supply in September due to low inventory. Low inventory also breaks the inter-temporal linkage between the September and subsequent contracts, resulting in high volatility of the September contract to be captured by the short-term and contract specific shock.

4.3 Volatility Persistence and Cross-Commodity Correlation (GARCH + DCC)

Tables 3 and 4 summarize the estimates of GARCH parameters. In **Table 3**, the sum of the coefficient estimates of α_j^i and β_j^i is close to one for each of two factors and for each of three commodities. The sum of the two estimated coefficients of DCC process, γ_1 and γ_2, is also close to one. These estimates indicate that both the conditional variance and conditional correlation of the underlying factors are highly persistent yet stationary.

In **Table 4**, the estimated unconditional correlation, obtained as a sample correlation of the standardized factors, is high for the long-term factors; 0.86 for CO-HO, 0.84 for CO-UG and HO-UG pair, whereas it is substantially smaller for the short-term factors; ranging from 0.33 for HO-UG pair to 0.43 for CO-HO pair. Lower cross-commodity correlations for the short-term factors than for the long-term factors imply that the short-term factors capture demand, supply, and other market shocks that are specific to each market. Examples of such market shocks include outages of refining facilities and, for the case of heating oil, unusually cold weather in the winter peak-demand season. These shocks will naturally generate a sharp increase in the refined product prices, which revert gradually to the normal level as demand and/or supply respond and the inventory adjusts to the normal level. Comparing between the two refined commodities, the correlation with crude oil price is higher for heating oil than for unleaded gasoline in both the short- and long-term factors. However, the cross-factor correlations between crude oil and the refined commodities (e.g., correlation between CO Factor 1 and HO Factor 2) are higher for unleaded gasoline than heating oil. Although these differences are small, they suggest that heating oil dynamics match crude oil more closely than unleaded gasoline.

Figure 5 shows the conditional correlation of the latent factors as predicted by the estimated DCC process. To enhance clarity, the figure presents monthly averages of the daily conditional correlations. In **panel (a)** of **Figure 5**, the correlations between the long-term factors exhibit only moderate fluctuations around the corresponding unconditional correlation value, although they increased gradually from an average of 0.8 in the early part of the sample to an average of 0.9 in the latter part of the sample. In contrast, the correlations between the short-run factors vary widely, ranging from -0.17 to 0.80 for CO-HO pair, -0.07 to 0.76 for CO-UG pair, and -0.18 to 0.73 for HO-UG pair (**panel (b)**). High variation in the conditional correlations of the short-term factors is consistent with the fact that the identified short-term factors sometimes represent

commodity-specific shocks, such as a change in seasonal demand, and other times represent a common shock such as a refinery outage.

The correlations between the short-term factors show a gradual trend upwards for most of the sample, rising from about 0.2 in 1989 to a high of about 0.6 in 2002. In the ensuing four years, these correlations dropped back to around 0.3. Nominal oil prices fluctuated around $20 per barrel throughout the 1985-2001 period, which implies a declining real price of oil during this interval. When oil prices began to increase in 2002, the correlation between the short-run factors began declining. The increase from 1985-2001 may reflect the reduction in the refining margin that occurred in the 1990s. Narrower margins may reflect crude oil prices moving more closely with the prices of the refined products, so that both short-run and long-run shocks are more likely to be passed through.

Between 2003 and 2006, the conditional variance of the short-run factor decreased by almost 50 percent. This decrease reveals that the prolonged increase in oil prices from 2003 through the end of the sample was driven by long-run shocks. Most of the price variation came from long-run shocks and there were few common short-run shocks. This decline of common short-run shocks is also reflected in the declining cross-commodity correlation between the short-run factors.

4.4 Crack Spread Dynamics

Decomposition of Crack Spread Variance—Description

In this section, we explore the volatility dynamics of crack spread, the margin received from producing either or both of the two refined petroleum products from crude oil. We consider three crack spreads widely considered in the industry (NYMEX, 1999, 2000); -3:2:1, -1:1:0, and -1:0:1, where three numbers represent the short (long if negative) position in the CO, UG, and HO markets, respectively. Based on the estimated POTS model, we decompose the variance of each of these positions into three components; long-term, short-term, and idiosyncratic shocks, and examine how the magnitudes of these components change by delivery month and time-to-maturity.

Let **A** denote a 1 by 3 vector of positions in the three commodity markets. The variance of the crack spread is

$$(5) \qquad V[\mathbf{A}\Delta\ln\mathbf{F}_t] = \mathbf{A}V[(\Theta_t\boldsymbol{\varepsilon}_t + \Theta_{3,t}\mathbf{u}_t)]\mathbf{A}' = \mathbf{A}\Theta_t\Omega\Theta_t'\mathbf{A}' + \mathbf{A}\Theta_{3,t}\Psi\Theta_{3,t}'\mathbf{A}'$$

where $\boldsymbol{\Theta}_t = \{\boldsymbol{\Theta}_{1,t}', \boldsymbol{\Theta}_{2,t}'\}'$ and $\boldsymbol{\Psi}_t = E[\mathbf{u}_t\mathbf{u}_t']$ is the covariance matrix of idiosyncratic errors, $\mathbf{u}_t$. The two terms in the right-hand side of (5) represent the components of the crack-spread variance that originate in the six common factors and in the idiosyncratic errors, respectively. Although the model assumes $E[\mathbf{u}_t\mathbf{u}_t'] = I_{n_t}$, some of the estimated idiosyncratic error terms are correlated. In particular, the idiosyncratic errors for nearby contracts are correlated with the errors for the second and third positions contracts. The idiosyncratic errors for the same maturity are also correlated across commodities. In the calculations below, we estimate $\boldsymbol{\Psi}_t$ using the sample covariances. Specifically, we use their estimates obtained as the sample correlation of the predicted values of idiosyncratic errors, $\hat{u}_{d,t}^{j}$, calculated for each trade month, for each combinations of the commodities, and for each of 144 combinations of days-to-maturity in monthly block.

Since the six latent factors are contemporaneously correlated, most of the off-diagonal elements of $\boldsymbol{\Theta}_t\boldsymbol{\Omega}\boldsymbol{\Theta}_t'$ are non zero (see **Table 4**). For a six factor model, the matrix $\boldsymbol{\Theta}_t\boldsymbol{\Omega}\boldsymbol{\Theta}_t'$ contains 12 non-zero covariance terms and three zero covariance terms. The three zero terms each represent the covariance between the two factors for a given commodity. The presence of 12 non-zero covariance terms makes $\boldsymbol{\Theta}_t\boldsymbol{\Omega}\boldsymbol{\Theta}_t'$ difficult to interpret. Therefore, we orthogonalize the covariance matrix of latent factors through the Cholesky decomposition,

(6) $\omega\omega' = \boldsymbol{\Omega}$

where the lower triangular matrix, ω, is the Cholesky factor of $\boldsymbol{\Omega}$. Substituting (6) into (5) yields,

(7) $V[\mathbf{A}\Delta\mathbf{lnF}_t] = \boldsymbol{\Gamma}_t\boldsymbol{\Gamma}_t' + \mathbf{A}\boldsymbol{\Theta}_{3,t}\boldsymbol{\Psi}_t\boldsymbol{\Theta}_{3,t}'\mathbf{A}'$

where $\boldsymbol{\Gamma}_t = \mathbf{A}\boldsymbol{\Theta}_t\omega$.

The values of Cholesky factors, ω, and hence, $\boldsymbol{\Gamma}_t$ depend on the order of latent factors in the vector $\boldsymbol{\varepsilon}_t$. However, of all the possible orderings, only those that place the three long-term factors first preserve our identification condition that three of the factors correspond to short-term shocks, i.e., they have zero loading at $d = d_{max}$. Furthermore, for those orderings that place the three long-term factors first, the values of the six components individually are sensitive to the ordering of the three long- and short-term factors. However, the sum of the first three components of $\boldsymbol{\Theta}_t\omega$ and the sum of the last three components are insensitive to the ordering. These two sums effectively represent the aggregate long- and short-term shocks to the crack spread variance. Thus, to understand the components of the crack spread, we order long-term factors before short-term factors and report the aggregate of each.

Decomposition—Results

Figure 6 shows, for each of twelve maturity dates, how the variance of three crack spread portfolios changes as the contracts approach maturity. Three key observations are of particular interest in the figure. First, for all three spreads, volatility at long trading horizons is low until around two months before delivery, at which point it increases rapidly. Second, the two pairwise crack spreads exhibit seasonal volatility patterns similar to that observed for the relevant refined commodity, with volatility in the last three months of trading substantially greater for the September delivery of the CO-UG spread and the February delivery of the CO-HO spread than the other deliveries. The -3:2:1 spread reflects the seasonal patterns of the two pairwise spreads but more of the CO-UG spread than CO-HO spread. Third, volatility is generally higher for the CO-UG spread than for the CO-HO spread.

Figure 7 decomposes the variance of three crack spreads into three components; the short-term, long-term, and idiosyncratic shocks, and illustrates, for each of February and September delivery, how these components change over one year of trading horizon. Although not shown in **Figure 7**, the variance of crack spreads for other ten delivery months exhibits the dynamics similar to the pattern observed for February and September delivery. The figure shows that the low variance of three spreads in the first several months of trading originates predominantly in the long-term factor. This pattern reflects the result shown in **Figure 4** that essentially all of the price variation at horizons greater than six months emanates from the long-term factors. Long-term shocks tend to be passed through and have little effect on the crack spread, which is reflected in the high correlation of the long-term shocks across commodities and the relatively low long-horizon variance of the crack-spread portfolios.

The volatility of the crack spread indicates the potential daily effectiveness of hedging price risk. Specifically, suppose a refiner plans to buy crude oil and sell heating oil in December of a particular year. In January of that year, the December futures prices represent the expected spot prices 11 months later. The change in futures prices from day to day therefore equals the change in the expected December price. If the percentage change in the crude oil futures price exactly offsets the percentage change in the heating oil price, then the crack-spread volatility equals zero and there is no gain to hedging today rather than waiting until tomorrow. On the other hand, positive crack-spread volatility represents the daily exposure of a firm with an unhedged crack-spread position. It measures the potential change in the value of the firm's position from one day to the next. For all twelve delivery dates and for all three spreads, the crack spread variance at

one year before maturity is around 20 percent of the variance of the portfolio without an opposite position in crude oil. Thus, although the crack-spread variance is much lower than in an individual commodity, substantial benefits to long-horizon hedging remain.

As delivery approaches, all three commodities become increasingly subject to the short-term shocks. Although these short-term shocks exhibit only moderate correlation across three commodities, they are less volatile for crude oil than for the two refined commodities. Consequently, until around about two months before maturity, the crack-spread portfolio variance decreases gradually relative to refined-product portfolio without an opposite position in crude oil. The crack-spread variance is lower for the CO-HO than for the CO-UG spread because the correlation in the short-term factors is higher for CO-HO pair than for CO-UG pair.

In the last two months before delivery, all three commodities become subject to the idiosyncratic shocks, which are only weakly correlated between crude oil and the two refined commodities. Consequently, the correlation between crude oil and the two refined commodity prices decreases rapidly in this period, and the volatility of the crack-spread portfolios is almost as large as for the individual refined commodities. The benefits of crack-spread hedging are thus magnified during this period. The magnitudes of the idiosyncratic shocks differ substantially across the twelve delivery dates, resulting in large seasonal variation in the benefits of crack-spread hedging in the last one month of trading period.

The -3:2:1 spread exhibits much greater variance than two pairwise spreads because it includes three times the number of contracts as the pairwise portfolios. However, once scaled to one ninth to make it comparable to the pairwise spreads, the variance of the -3:2:1 crack spread is substantially smaller than the pairwise portfolios, particularly near the maturity date. This is because, in addition to the correlation between crude oil and two refined commodities, the prices of the two refined commodities are correlated each other in all three components. The -3:2:1 spread also reflects the seasonal pattern of unleaded gasoline more than that of heating oil, because the portfolio contains twice as more units of the former than the latter. In addition, as shown in **Figure 4**, the February contract of heating oil and the September contract of unleaded gasoline are subject to high idiosyncratic shocks in the last one month of trading period. High volatility of these contracts is mitigated in the -3:2:1 spread through moderate correlation of idiosyncratic errors between the two refined commodities. The variance is reduced by a greater extent for the February than for the September delivery, because the February contract for

unleaded gasoline is subject to greater idiosyncratic shocks than the September contract for heating oil.

5. CONCLUSION

We examine the volatility dynamics of the futures prices of three petroleum commodities traded at NYMEX via the partially overlapping time-series model of Smith (2005). We extend Smith's original specification to a six-factor, three-commodity setting, which we use to decompose the daily futures returns for each commodity into two common factors and a contract specific component. The estimated model reveals that the volatility of the three commodity prices exhibits both time-to-maturity effects and substantial seasonal variations in patterns consistent with the observed seasonality in demand and storage of the three commodities.

For heating oil, the volatility in the last few months of trading is higher for the contracts maturing in late fall through the end of winter, during which the demand peaks for space heating. It is particularly high for the January through March contracts due to low inventory at the end peak-demand season. For unleaded gasoline, the volatility is high for contracts maturing in late summer through fall when gasoline inventory is low after the summer peak-demand season and before the commodity specification shifts from the summer to winter grade. Relatively weak seasonal variation is depicted for crude oil in its long-term factor, with volatility slightly higher for all twelve contracts during early fall through winter when demand peaks for spacing heating in the US and global market. The same seasonal pattern is also observed for the two refined commodities, implying that the long-term shocks to the crude oil market are transmitted to the markets for two refined commodities.

For all three commodities, a large share of price variation emanates from the short-term factor in the last several months of trading and from idiosyncratic errors in the last two months of trading. Idiosyncratic shocks are particularly large for heating oil and unleaded gasoline contracts maturing in their respective peak demand seasons. These contract-specific shocks are virtually uncorrelated, resulting in low price correlation across concurrently traded contracts towards the end of the peak-demand season. This finding is consistent with the theory of storage, which suggests that low inventory at the end of peak-demand season breaks the inter-temporal price linkage so that contracts maturing in post peak-demand season do not exhibit much movement before the end of the demand season. The correlation is also low earlier in the peak

demand season, implying that high inventory at the beginning of the peak demand season provides only limited buffering of price shocks. The short-term and idiosyncratic shocks also exhibit much lower correlation than the long-term shocks across three commodities. Consequently, the crack spread is much more volatile at short horizons than long horizons. Nonetheless, significant long-horizon price risk remains, with long-horizon (greater than six months) volatility still 20 percent of its level for the individual commodities.

REFERENCES

Asche, F. Gjolberg, O, and Volker, T. 2003. "Price relationships in the petroleum market: an analysis of crude oil and refined product prices," Energy Economics, 25: 289-301.

Bollerslev, T. 1990. "Modeling the coherence in short-run nominal exchange rates: A multivariate generalized ARCH model, Review of Economics and Statistics, 72: 498-505.

Cortazar, G., and Naranjo, L. 2006. "An N-factor Gaussian model of oil futures prices," Journal of Futures Markets, 26: 243-268.

EIA. 1997. Petroleum 1996: Issues and Trends. Energy Information Administration, US Department of Energy, Washington DC.

Engle, R., and Sheppard, K. 2001. "Theoretical and empirical properties of dynamic conditional correlation multivariate GARCH," NBER working paper #8554.

Engle, R. 2002. "Dynamic conditional correlation: A simple class of multivariate generalized autoregressive conditional heteroskedasticity models," Journal of Business and Economic Statistics, 20: 339-350.

Gibson, R., and Schwartz, E. S. 1990. "Stochastic convenience yield and the pricing of oil contingent claims," Journal of Finance, 45:959-976.

Girma, P. B., and Paulson, A. S. 1999. "Risk arbitrage opportunities in petroleum futures spreads," Journal of Futures Markets, 19: 931-955.

Haigh, M. S., and Holt, M. T. 2002. "Crack spread hedging: Accounting for time-varying volatility spillovers in the energy futures markets," Journal of Applied Econometrics, 17: 269-289.

Hamilton, J. D. 1994. "State-Space Models," in Engle, R. F., and McFadden, D. L. eds. Handbook of Econometrics, Vol. 4. Elsevier, Amsterdam.

Lanza, A., Manera, M., and Giovannini, M. 2005. "Modeling and forecasting cointegrated relationships among heavy oil and product prices," Energy Economics, 27: 831-848.

Lautier, D. 2005. "Term Structure Models of Commodity Prices: A Review," Journal of Alternative Investments, 8(1): 42-64.

New York Mercantile Exchange. 2000. A Guide to Energy Hedging.

New York Mercantile Exchange. 1999. Crack Spread Handbook.

Ng, V. K., and Pirrong, S. C. 1996. "Price dynamics in refined petroleum spot and futures markets," Journal of Empirical Finance, 2: 359-388.

Samuelson, P. A. 1965. "Proof that properly anticipated prices fluctuate randomly," Industrial Management Review, 6: 41–49.

Schwartz, E. S. 1997. "The stochastic behavior of commodity prices: Implications for valuation and hedging," Journal of Finance, 52: 923-973.

Schwartz, E., and Smith, JE. 2000. "Short-term variations and long-term dynamics in commodity prices," Management Science, 46:893-911.

Serletis, A. 1994. "A cointegration analysis of petroleum futures prices," Energy Economics, 16: 93-97.

Smith, A. 2005. "Partially overlapping time-series model: A new model for volatility dynamics in commodity futures," Journal of Applied Econometrics, 20:405-422.

Todorova, M. I. 2004. "Modeling energy commodity futures: Is seasonality part of it?" Journal of Alternative Investment, Fall 2004: 10-31.

Williams, J. C., and Wright, B. D. 1991. Storage and Commodity Markets. Cambridge University Press, New York.

Table 1. Results of specification test

	Model Specification		
	1	2	3
Crude oil (CO)	-9.749	-9.820	-9.763
Heating oil (HO)	-9.132	-9.369	-9.151
Unleaded gasoline (UG)	-8.356	-8.557	-8.403

Note: Numbers presented are the Schwartz Information Criterion calculated for each of the three specifications of the factor loadings and innovation standard deviation in equation (2) and for each of three commodities.

Table 2. Proportion of the variance explained by common factors

	By contract												Overall
	1	2	3	4	5	6	7	8	9	10	11	12	
CO	97.7%	97.8%	98.7%	97.7%	97.7%	98.3%	97.2%	97.3%	97.2%	98.0%	97.5%	98.5%	97.8%
HO	96.6%	94.8%	95.1%	95.5%	97.2%	98.8%	99.1%	99.2%	96.8%	98.7%	97.7%	98.7%	97.3%
UG	96.1%	95.5%	95.8%	96.2%	97.6%	96.9%	97.8%	97.1%	90.7%	91.8%	93.2%	96.6%	95.4%

Note: These numbers represent the share of variations in log-returns that are accounted for by the two common factors. They are calculated as $\left(\sum_t \left(\hat{\theta}^j_{1,m,m-d}\hat{\varepsilon}_{1,t} + \hat{\theta}^j_{2,m,m-d}\hat{\varepsilon}_{2,t}\right)\right)\left(\sum_t \Delta \ln F^j_{m,t}\right)^{-1}$ for each of 12 delivery month (m = 1, …, 12) and $\left(\sum_{m=1}^{12}\sum_t \left(\hat{\theta}^j_{1,m,m-d}\hat{\varepsilon}_{1,t} + \hat{\theta}^j_{2,m,m-d}\hat{\varepsilon}_{2,t}\right)\right)\left(\sum_{m=1}^{12}\sum_t \Delta \ln F^j_{m,t}\right)^{-1}$ for overall share.

Table 3. Estimates of MGARCH coefficients

	Crude Oil		Heating Oil		Unleaded Gasoline		DCC	
	Coefficient	Std. Err.	Coefficient	Std. Err.	Coefficient	Std. Err.	Coefficient	Std. Err.
α_1^j	0.056	0.022	0.066	0.041	0.052	0.011		
α_2^j	0.074	0.066	0.071	0.030	0.068	0.039		
β_1^j	0.944	0.024	0.934	0.041	0.946	0.010		
β_2^j	0.926	0.066	0.929	0.030	0.932	0.040		
γ_1							0.025	0.002
$\gamma_1 + \gamma_2$							0.993	0.001

Note: These coefficients are estimated sequentially. First, the estimates of coefficients α_i^j and β_i^j (i = 1, 2) are obtained by estimating bivariate GARCH model for each of three commodities (j = CO, HO, and UG) separately. Second, the coefficients in DCC process, γ_1 and γ_2, are estimated on the predicted values of six latent factors obtained from the first step.

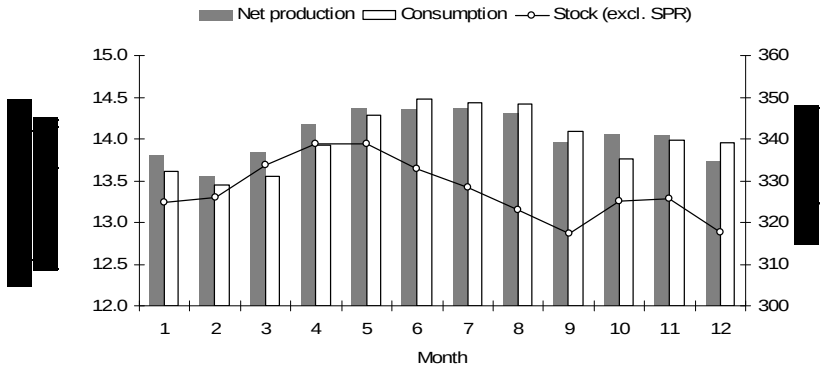
Table 4. Estimates of unconditional correlation of latent factors

		Commodity					
		CO		HO		UG	
Commodity		Factor 1	Factor 2	Factor 1	Factor 2	Factor 1	Factor 2
CO	Factor 1		0.005	0.870	0.095	0.841	0.135
	Factor 2			0.237	0.431	0.276	0.356
HO	Factor 1				0.009	0.857	0.188
	Factor 2					0.175	0.308
UG	Factor 1						0.017
	Factor 2						

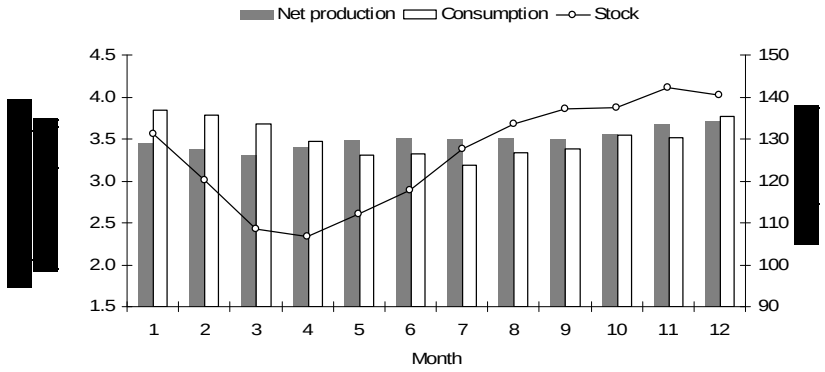
Note: Numbers presented are obtained as the sample correlation of the predicted factor values.

Figure 1. Seasonality in demand, supply, and storage of US crude oil, unleaded gasoline, and high sulfur distillate fuel oil

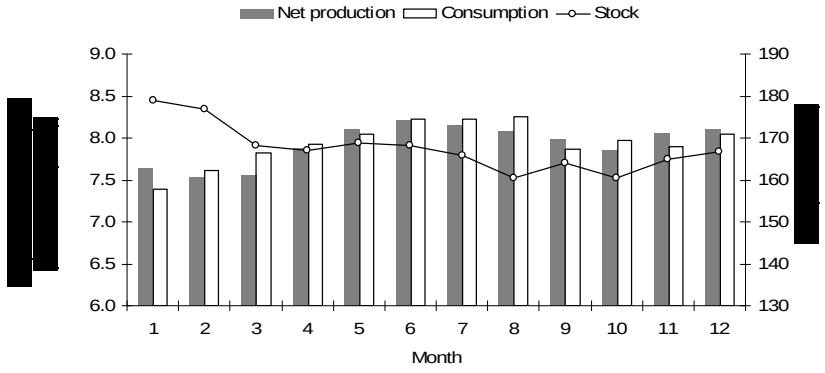
(a) Monthly averages of consumption, net production, and storage of crude oil



(b) Monthly averages of consumption, net production, and stock of distillate fuel oil



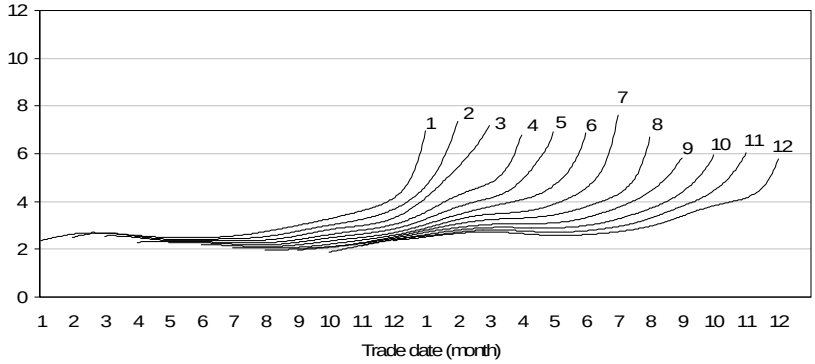
(c) Monthly averages of consumption, net production, and stock of finished motor gasoline



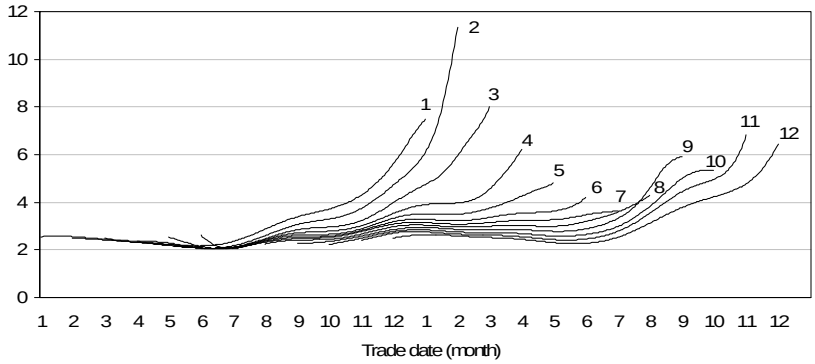
Note: Three figures are created based on the EIA's data on the US supply and disposition of crude oil and finished petroleum products (obtained from http://tonto.eia.doe.gov/dnav/pet/pet_sum_snd_d_nus_mbblpd_m_cur.htm) for the period between Jan. 1981 and Dec. 2008. For CO, net production represents the sum of the US field production, import, and supply adjustment whereas consumption represents the sum of the US refinery and blender net input, export, and product supplied of crude oil. For HO and UG, net production represents the sum of the US refinery and blender net production, import, and supply adjustment whereas consumption represents the sum of the export and product supplied of each petroleum commodity.

Figure 2. Unconditional variance of daily log returns as implied by the estimated POTS model

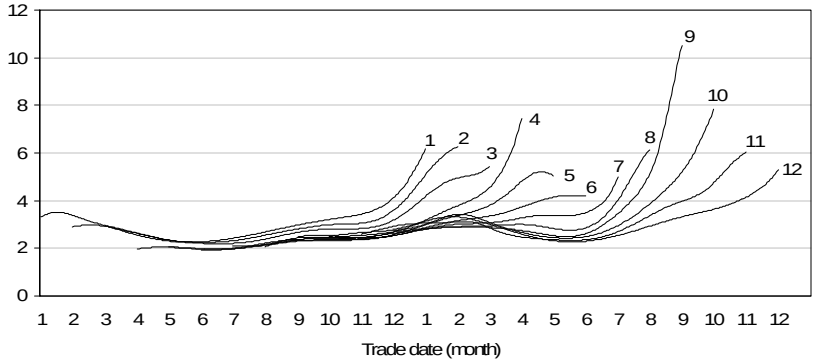
(a) Crude Oil



(b) Heating oil



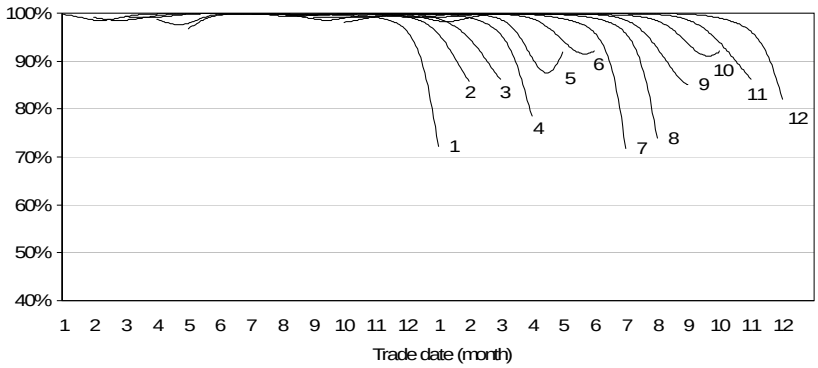
(c) Unleaded gasoline



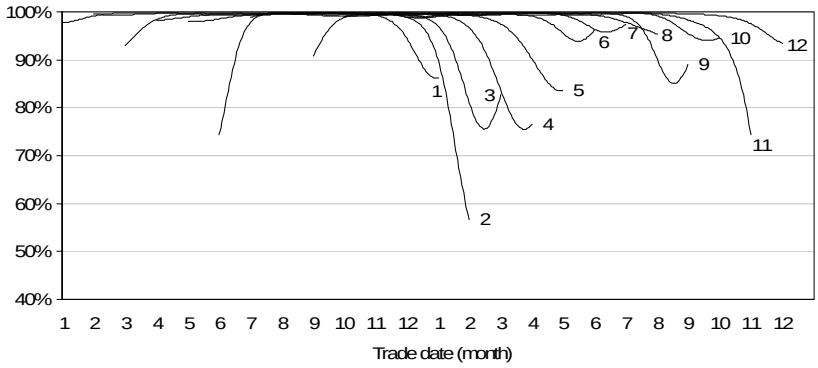
Note: The unconditional variance of daily log returns, computed as $\sum_{i=1}^{3} \hat{\theta}^{j}_{i,m,m-d}{}^{2}$ for each of the three commodities (j = CO, HO, UG) and for each of twelve delivery months (m = 1, …, 12), is plotted for time-to-delivery (d) ranging from 0 to 252.

Figure 3. Share of price variation explained by the common factors

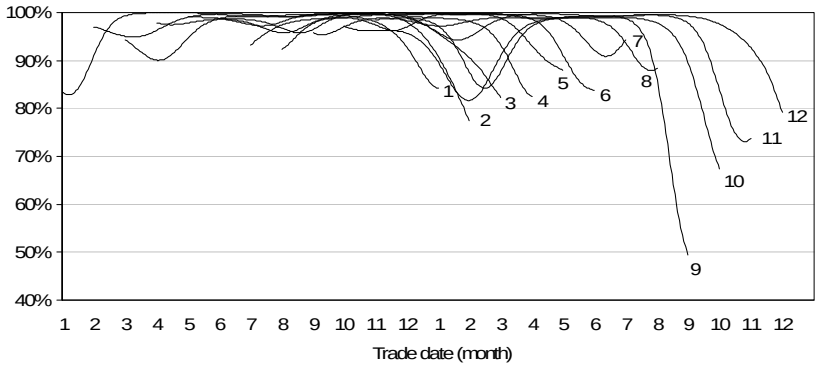
(a) Crude oil



(b) Heating oil



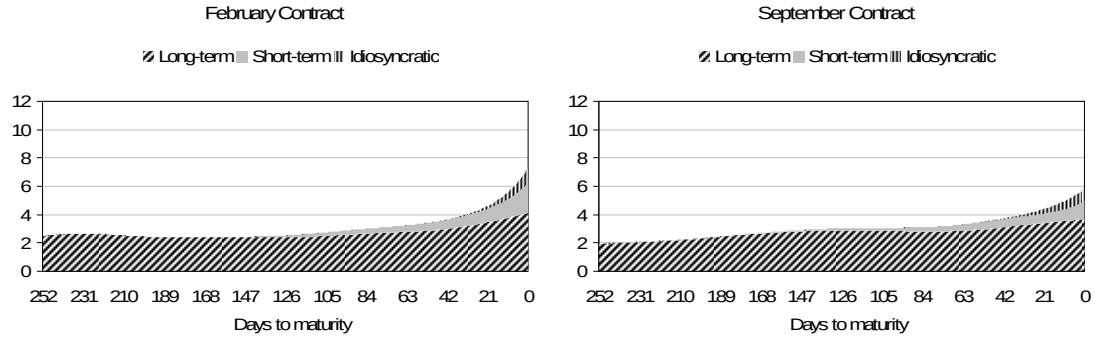
(c) Unleaded gasoline



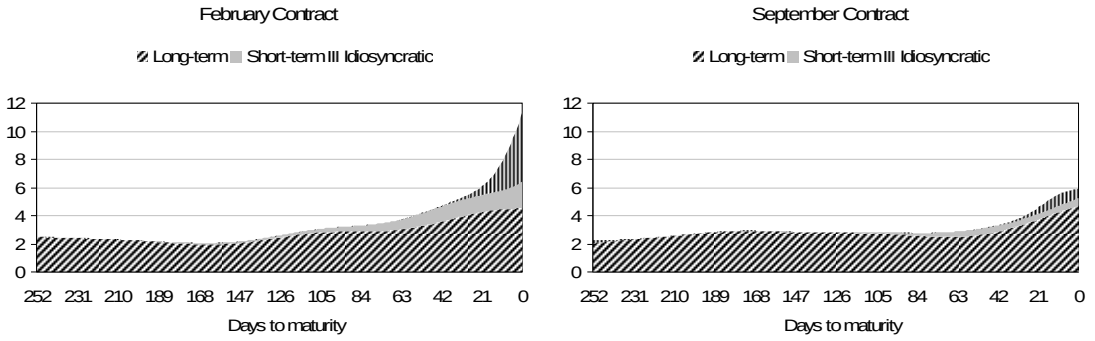
Note: The share of total variance accounted for by the two common factors, as measured by $\sum_{i=1,2} \hat{\theta}^{j}_{i,m,m-d}{}^{2} / \sum_{i=1}^{3} \hat{\theta}^{j}_{i,m,m-d}{}^{2}$ for each of the three commodities (j = CO, HO, UG) and for each of twelve delivery months (m = 1, …, 12), is plotted for time-to-delivery (d) ranging from 0 to 252.

Figure 4. Variance decomposed into three components

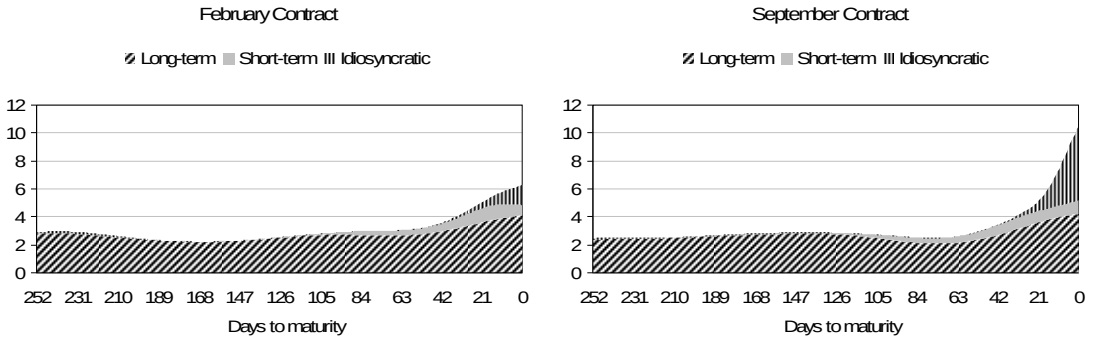
(a) Crude Oil



(b) Heating Oil



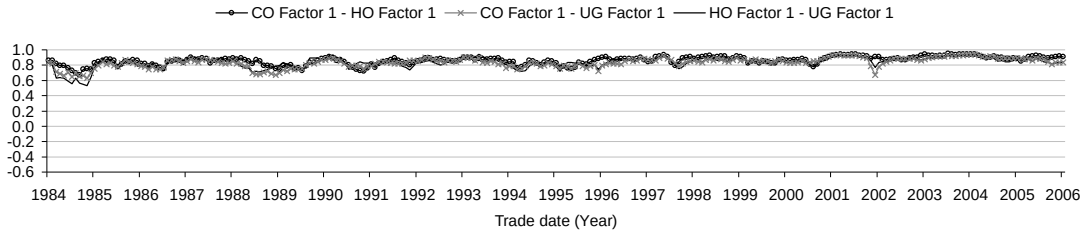
(c) Unleaded Gasoline



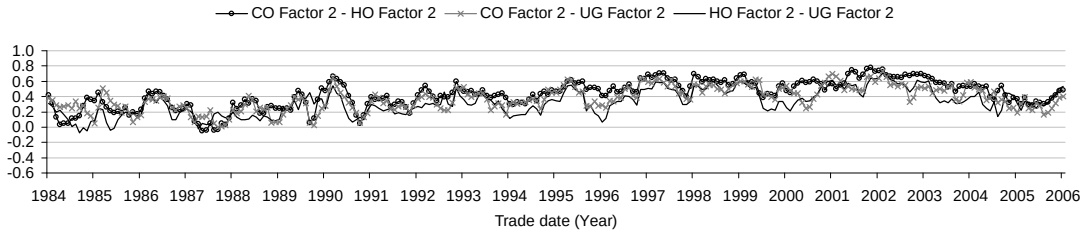
Note: Three sources of the unconditional variance are measured by $\hat{\theta}^{j}_{i,m,m-d}{}^{2}$ for each of the three commodities (j = CO, HO, UG), for each of twelve delivery months (m = 1, …, 12), and for the time-to-maturity (d) ranging from 0 to 252. The figure plots how each of the three sources changes over one year of trading horizon for the February (m = 2) and the September (m = 9) contract.

Figure 5. Monthly averages of conditional variances and conditional correlations predicted by the estimated DCC model

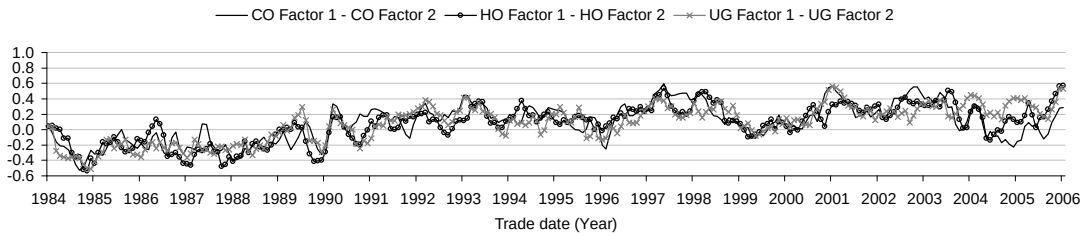
(a) Factor 1 across three commodities



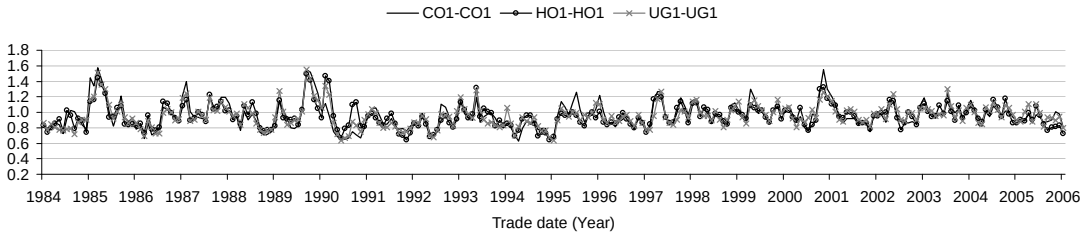
(b) Factor 2 across three commodities



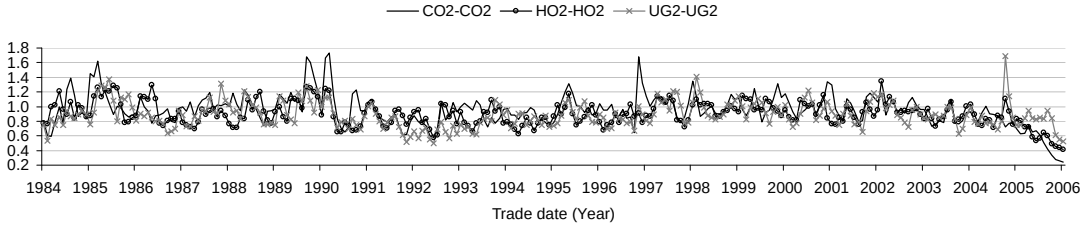
(c) Factors 1 and 2 for each of three commodities



(d) Conditional variance of factor 1



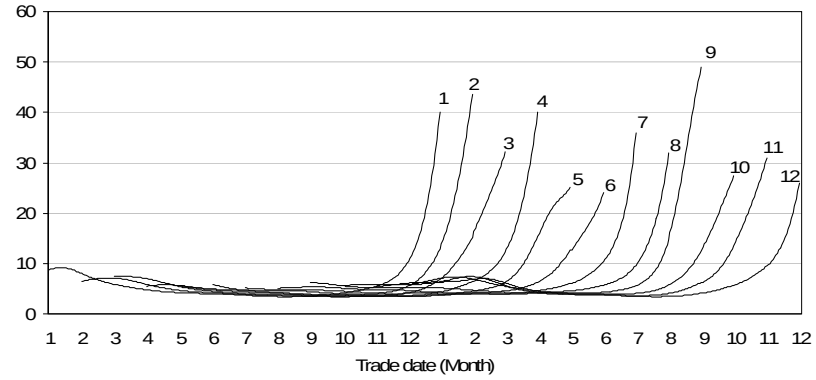
(e) Conditional variance of factor 2



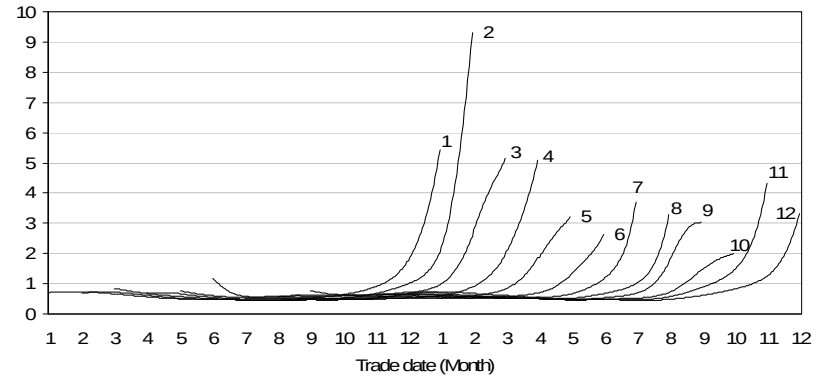
Note: Panels (a) and (b) plot the monthly averages of conditional correlation across three commodities whereas panel (c) plots the monthly averages of conditional correlation between factors 1 and 2 for each of three commodities, as estimated by the DCC model. Panels (d) and (e) plot the monthly averages of conditional variance of factors 1 and 2 for three commodities.

Figure 6. Variance of crack spread position implied by the estimated POTS model

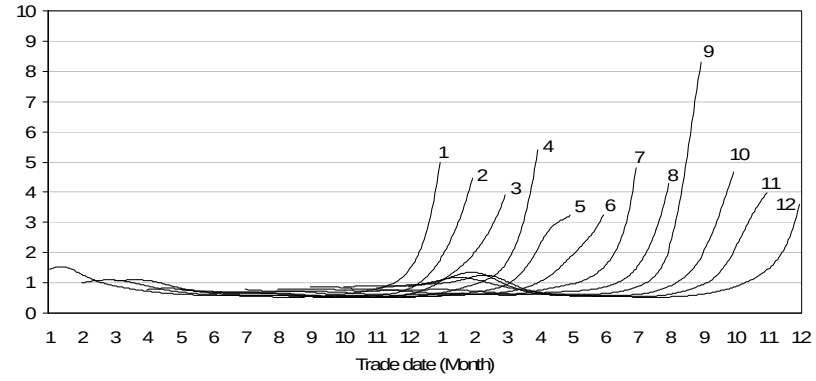
(a) -3:2:1 Crack Spread



(b) -1:0:1 Crack Spread



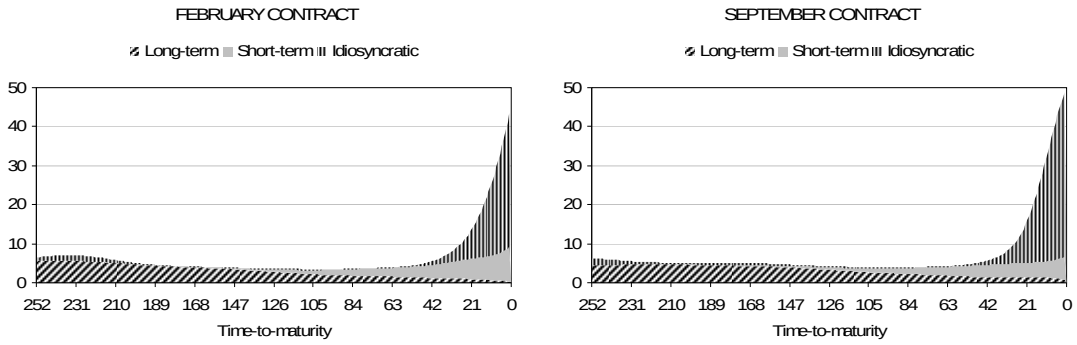
(c) -1:1:0 Crack Spread



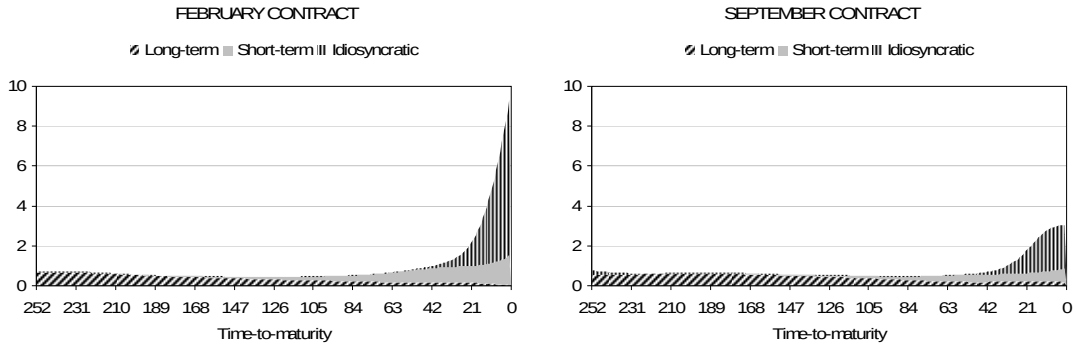
Note: The unconditional variance of daily log returns to a crack-spread position is calculated as $V[\mathbf{A}\Delta\ln F_{m,m-d}] = \mathbf{A}\boldsymbol{\Theta}_{m,m-d}\boldsymbol{\Omega}\boldsymbol{\Theta}_{m,m-d}'\mathbf{A}' + \mathbf{A}\boldsymbol{\Theta}_{3,m,m-d}\boldsymbol{\Psi}_{m,m-d}\boldsymbol{\Theta}_{3,m,m-d}'\mathbf{A}'$ where $\mathbf{A}$ is a 1 by 3 vector of futures positions in three commodity markets, $\boldsymbol{\Theta}_{m,m-d} = \{\boldsymbol{\Theta}_{1,m,m-d}', \boldsymbol{\Theta}_{2,m,m-d}'\}'$ with $\boldsymbol{\Theta}_{i,m,m-d}'$ representing a 3 by 3 diagonal matrix with its elements $\{\theta^{CO}{}_{i,m,m-d}, \theta^{UG}{}_{i,m,m-d}, \theta^{HO}{}_{i,m,m-d}\}$, $\boldsymbol{\Omega}$ is the unconditional variance of six latent factors, and $\boldsymbol{\Psi}_{m,m-d} = E[\mathbf{u}_{m,m-d}\,\mathbf{u}_{m,m-d}']$ is a 3 by 3 covariance matrix of idiosyncratic errors with $\mathbf{u}_{m,m-d} = (\mathbf{u}^{CO}{}_{m,m-d}, \mathbf{u}^{UG}{}_{m,m-d}, \mathbf{u}^{HO}{}_{m,m-d})'$. The figure plots the unconditional variance of three commonly considered crack-spread positions: -3:2:1, -1:0:1, and -1:1:0, where three numbers represents, respectively, the short (long if negative) position in the CO, UG, and HO markets.

Figure 7. Variance of crack spread position decomposed into three components

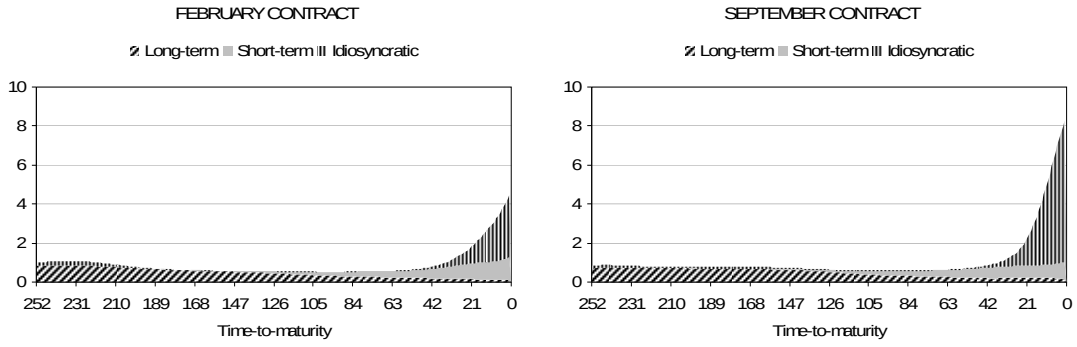
(a) -3:2:1 Crack Spread



(b) -1:0:1 Crack Spread



(c) -1:1:0 Crack Spread



Note: The unconditional variance of daily log returns to a crack-spread position is decomposed into three sources: long-term, short-term, and idiosyncratic shock. We implement this decomposition by: (i) reorder six latent factors so that three long-term factors are placed first and then three short-term factors, (ii) apply Cholesky decomposition to the covariance matrix of reordered latent factors, i.e., $\omega\omega' = \mathbf{\Omega}$, and (iii) calculate the aggregate long- and short-term shocks as the sum of the first three components and the second three components of $\mathbf{\Theta}_t\omega$, respectively. The unconditional variance of idiosyncratic shocks is calculated as $\mathbf{A}\mathbf{\Theta}_{3,m,m-d}\mathbf{\Psi}_{m,m-d}\mathbf{\Theta}_{3,m,m-d}'\mathbf{A}'$. We calculate such variance decomposition for three commonly considered crack-spread positions: -3:2:1, -1:0:1, and -1:1:0, where three numbers represents, respectively, the short (long if negative) position in the CO, UG, and HO markets. The figure illustrates how three components of unconditional variance change over one year of trading horizon for each of the three crack-spread positions and for each of February and September delivery.