# Integrating Multiple Commodities in a Model of Stochastic Price Dynamics

Raphael Paschke and Marcel Prokopczuk

University of Mannheim

23. October 2007

# Integrating Multiple Commodities in a Model of Stochastic Price Dynamics

Raphael Paschke and Marcel Prokopczuk*

October 2007

**Abstract**

In this paper we develop a multi-factor model for the joint dynamics of related commodity spot prices in continuous time. We contribute to the existing literature by simultaneously considering various commodity markets in a single, consistent model. In an application we show the economic significance of our approach. We assume that the spot price processes can be characterized by the weighted sum of latent factors. Employing an essentially-affine model structure allows for rich dependencies among the latent factors and thus, the commodity prices. The co-integrated behavior between the different spot price dynamics is explicitly taken into account. Within this framework we derive closed-form solutions of futures prices. The Kalman Filter methodology is applied to estimate the model for crude oil, heating oil and gasoline futures contracts traded on the NYMEX. Empirically, we are able to identify a common non-stationary equilibrium factor driving the long-term price behavior and stationary factors affecting all three markets in a common way. Additionally, we identify factors which only impact subsets of the commodities considered. To demonstrate the economic consequences of our integrated approach, we evaluate the investment into a refinery from a financial management perspective and compare the results with an approach neglecting the co-movement of prices. This negligence leads to radical changes in the project's assessment.

**JEL classification:** Q40, G13, C50

**Keywords:** Commodities, Integrated Model, Crude Oil, Heating Oil, Gasoline, Futures, Kalman Filter

---

*We thank Wolfgang Bühler for valuable comments. Raphael Paschke, Marcel Prokopczuk: Chair of Finance, University of Mannheim, D-68131 Mannheim, Germany. Contact e-Mail: paschke@uni-mannheim.de, prokopczuk@uni-mannheim.de.

# I Introduction

When making investment and risk management decisions it is necessary to consider the joint distribution of underlying factors. In this paper we demonstrate that when evaluating a project related to multiple commodities, it is of crucial importance to take the manifold dependence structure between these commodities into account. We develop a single multi-factor model and show in a real world example that it is not sufficient to model the dependencies via correlated returns, but it is necessary to allow for interdependencies in the price levels. These kind of relationships will alter the project's evaluation substantially, and thus, must be considered. To the best of our knowledge, this paper is the first considering more than one commodity in a single, consistent continuous time model.

It is standard to model commodity prices stochastically. However, in contrast to the stochastic behavior of stock prices, a pure random walk assumption does not seem to be justified, as supply and demand will directly respond to price changes and thus enforcing a mean reverting behavior (see Gibson and Schwartz (1990) and the references therein). On the other hand, Schwartz and Smith (2000) point out that uncertainty about the long term equilibrium price to which the process reverts exists. Thus, they propose modeling the stochastic behavior by two factors, a pure Brownian motion capturing the equilibrium level uncertainty, and an Ornstein-Uhlenbeck process, characterizing short-term deviations from this equilibrium.

In the literature various commodity markets are considered,[1] however, none of the articles take potential dependencies between the different markets into account. Furthermore, estimation of the risk processes' parameters is conducted separately for each market. For example, Cassasus and Collin-Dufresne (2005) propose a model which explicitly relates interest rates and commodity prices. However, their modeling approach leads empirically to the fact, that for each commodity market a different interest rate process is estimated, which is, as noted by the authors themselves, not consistent.

[1] Gibson and Schwartz (1990) and Schwartz and Smith (2000) consider crude oil, Schwartz (1997) crude oil, copper, and gold and Cassasus and Collin-Dufresne (2005) crude oil, copper, gold, and silver. Work applying the same methodology to other markets include Sorensen (2002) for agricultural commodities and Manoliu and Tompaidis (2002) for the natural gas market.

Our contribution is a model able to capture the stochastic behavior of multiple, related commodities simultaneously in a consistent way. As a typical example for the importance of considering the co-movement of commodity prices, we consider crude oil, heating oil, and gasoline, which are obviously related. However, our approach can easily be adjusted to other interrelated commodity markets and the explicit inclusion of an interest rate process. In a preliminary analysis we first confirm in Section II, that these commodity prices are not only correlated, but also co-integrated, i.e. follow a common long-term equilibrium process.[2] Our framework accounts for these kind of dependencies without assuming it ex-ante and also captures most stylized facts of commodity prices, namely backwardation, mean reversion, declining volatilities with contract horizon, and seasonality.

To illustrate the relevance of our model we consider the financial management of a long-term natural resource project, specifically an oil refinery. First, we show that considerable errors are made by neglecting the co-integrated behavior of crude oil, heating oil and gasoline. When computing the Value-at-Risk of an average refinery investment, these errors amount to more than 2 billion USD. Second, we demonstrate how to hedge a long-horizon exposure to all three commodities with short-term futures and compare the optimal hedge ratios with a model that allows only dependencies in returns. In contrast to the latter, where a substantial hedging demand sustains, the required hedge positions almost vanish for long-horizon exposures.

While Schwartz and Smith (2000) model the risk sources as latent factors, Schwartz (1997) and Cassasus and Collin-Dufresne (2005) explicitly incorporate the risk sources stochastic convenience yields and interest rates. Both approaches are equivalent, which was shown first by Schwartz and Smith (2000). We follow Schwartz and Smith (2000) and model the risk sources as latent state variables which are not directly observable. This implicit representation has several advantages. First, it keeps the model analytically tractable, second, it allows us to identify the maximal number of parameters for a given number of risk factors since we can adapt the general affine framework of Dai and Singleton

[2]Empirical evidence for the dependencies among various commodity markets can be found for instance in Pindyck and Rotemberg (1990). The co-integration of crude oil at its refined products has been also shown by Gjolberg and Johnson (1999).

(2000). Third, as it enables us to derive closed form solutions for the log-prices of futures written on the respective commodities which are linear in the state variables, the unknown parameters can be estimated by employing standard Kalman filtering techniques and maximum likelihood.

16 years of weekly sampled futures price data from the New York Mercantile Exchange is used to estimate our model. Given the estimated parameters, we analyze the joint behavior of the three spot price processes. We identify one non-stationary equilibrium process and find endogenously a common sensitivity of all commodities towards this factor. Furthermore, we find two factors causing deviations on all three markets with different degrees of persistence. A fourth factor mainly captures shocks which influence both refined products markets in a similar fashion, but not the crude oil market. In contrast, the fifth factor mainly affects the crude oil spot price, but has only weak impact on the heating oil and gasoline prices. Finally, the sixth factor captures shocks that influence the two derivative markets in a distinctive way.

The remainder of this article is structured as follows. We begin by conducting a preliminary analysis of the data considered. In Section III we derive the co-integrated factor model and provide closed from solutions for futures prices. Estimation using the Kalman filter is described in Section IV. In Section V we provide and discuss the results. Section VI demonstrates the implications for the evaluation of a long-term natural resource project. Section VII summarizes and concludes.

# II Preliminary Data Analysis

The data used in our study are prices of energy futures traded on the New York Mercantile Exchange (NYMEX). We consider three closely related commodities: (i) crude oil, (ii) heating oil, and (iii) gasoline, for which we sample generic futures prices from January 1990 to December 2005 on a weekly basis. Crude oil is the raw material for various products, including gasoline, heating oil, diesel, jet fuel etc. The two most important products refined from crude oil are heating oil and gasoline. Almost half of a barrel crude

Table 1: **Descriptive Summary Statistics of Crude Oil Futures**

*This table reports summary statistics for the crude oil futures data. The first column reports the futures Bloomberg Ticker wherein the number corresponds to the maturity in months. Prices are in US Dollars per barrel. DF stands for the Dickey-Fuller test of a unit root. * indicates significance at the 1% level.*

| | | Price | | | Return | | | Maturity | |
|---|---|---|---|---|---|---|---|---|---|
| Contract | NOBS | Mean | St.Dev. | DF | Mean | St.Dev. | DF | Mean | St.Dev. |
| CL1 | 835 | 25.55 | 10.87 | -0.8730 | 0.0606 | 0.3586 | -43.65* | 0.0435 | 0.0241 |
| CL2 | 835 | 25.42 | 10.90 | -0.4162 | 0.0644 | 0.3241 | -42.55* | 0.1269 | 0.0245 |
| CL3 | 835 | 25.25 | 10.89 | -0.0474 | 0.0665 | 0.2962 | -41.87* | 0.2100 | 0.0243 |
| CL4 | 835 | 25.08 | 10.85 | 0.2770 | 0.0682 | 0.2732 | -41.44* | 0.2937 | 0.0242 |
| CL5 | 835 | 24.90 | 10.80 | 0.5524 | 0.0698 | 0.2548 | -41.30* | 0.3770 | 0.0244 |
| CL6 | 835 | 24.74 | 10.73 | 0.7983 | 0.0712 | 0.2398 | -41.38* | 0.4603 | 0.0244 |
| CL7 | 835 | 24.58 | 10.66 | 1.0116 | 0.0724 | 0.2281 | -41.54* | 0.5439 | 0.0244 |
| CL8 | 835 | 24.43 | 10.59 | 1.2276 | 0.0734 | 0.2170 | -41.78* | 0.6272 | 0.0245 |
| CL9 | 835 | 24.30 | 10.52 | 1.3990 | 0.0741 | 0.2097 | -41.98* | 0.7105 | 0.0243 |
| CL10 | 835 | 24.17 | 10.45 | 1.5955 | 0.0746 | 0.2017 | -42.13* | 0.7941 | 0.0243 |
| CL11 | 835 | 24.06 | 10.38 | 1.7850 | 0.0749 | 0.1943 | -42.35* | 0.8773 | 0.0245 |
| CL12 | 835 | 23.96 | 10.32 | 1.9256 | 0.0752 | 0.1895 | -42.66* | 0.9606 | 0.0242 |
| CL13 | 835 | 23.86 | 10.25 | 2.0804 | 0.0754 | 0.1844 | -42.88* | 1.0443 | 0.0242 |
| CL14 | 835 | 23.78 | 10.19 | 2.2195 | 0.0756 | 0.1800 | -43.14* | 1.1275 | 0.0245 |
| CL15 | 834 | 23.70 | 10.14 | 2.3309 | 0.0714 | 0.1760 | -43.00* | 1.2108 | 0.0243 |
| CL16 | 835 | 23.62 | 10.07 | 2.4725 | 0.0760 | 0.1731 | -43.49* | 1.2945 | 0.0244 |
| CL17 | 835 | 23.56 | 10.01 | 2.5884 | 0.0761 | 0.1701 | -43.65* | 1.3776 | 0.0244 |
| CL18 | 814 | 23.59 | 10.06 | 2.6995 | 0.0696 | 0.1648 | -41.84* | 1.4578 | 0.0254 |

is used for gasoline, another quarter for heating oil.[3] Futures on these products are highly liquid.

All data is obtained from Bloomberg. Figure 1 displays price paths of a short term (1 month) and a long term (12 months) future for each respective commodity. Clearly, these prices are dependent on each other. Our data set includes also the recent strong price increase since 2003. Descriptive summary statistics for prices and log-returns of these contracts are reported in Tables 1, 2 and 3, respectively. For crude oil and heating oil we consider 18 different maturities, for gasoline only 12 different contracts are available. The relatively large standard deviations compared to former studies are caused by the sharp price increase in the final sample period. The average maturities between the different contracts with same maturities (in months) do deviate slightly due to different trading

[3]According to the Energy Information Administration (EIA) of the U.S. Department of Energy the precise numbers for 2005 are 44% and 24%.

Table 2: **Descriptive Summary Statistics of Heating Oil Futures**

*This table reports summary statistics for the heating oil futures data. The first column reports the futures Bloomberg Ticker wherein the number corresponds to the maturity in months. Prices are in US Dollar Cents per gallon. DF stands for the Dickey-Fuller test of a unit root. * indicates significance at the 1% level.*

| | | Price | | | Return | | | Maturity | |
|---|---|---|---|---|---|---|---|---|---|
| Contract | NOBS | Mean | St.Dev. | DF | Mean | St.Dev. | DF | Mean | St.Dev. |
| HO1 | 835 | 70.65 | 31.56 | -0.9929 | 0.0583 | 0.3760 | -43.42* | 0.0391 | 0.0242 |
| HO2 | 835 | 70.46 | 31.80 | -0.5472 | 0.0652 | 0.3333 | -43.30* | 0.1242 | 0.0240 |
| HO3 | 835 | 70.21 | 31.75 | -0.2180 | 0.0704 | 0.3081 | -42.28* | 0.2068 | 0.0243 |
| HO4 | 835 | 69.85 | 31.47 | -0.0353 | 0.0729 | 0.2872 | -41.66* | 0.2903 | 0.0243 |
| HO5 | 835 | 69.44 | 31.08 | 0.1274 | 0.0737 | 0.2673 | -41.53* | 0.3741 | 0.0243 |
| HO6 | 835 | 69.01 | 30.61 | 0.3211 | 0.0741 | 0.2503 | -41.69* | 0.4568 | 0.0244 |
| HO7 | 835 | 68.61 | 30.15 | 0.6147 | 0.0739 | 0.2353 | -42.28* | 0.5410 | 0.0244 |
| HO8 | 835 | 68.23 | 29.73 | 0.9022 | 0.0735 | 0.2246 | -42.69* | 0.6236 | 0.0244 |
| HO9 | 835 | 67.89 | 29.39 | 1.2296 | 0.0737 | 0.2164 | -42.86* | 0.7076 | 0.0244 |
| HO10 | 835 | 67.60 | 29.18 | 1.5494 | 0.0741 | 0.2105 | -42.98* | 0.7908 | 0.0244 |
| HO11 | 835 | 67.33 | 29.07 | 1.8728 | 0.0744 | 0.2058 | -43.12* | 0.8742 | 0.0242 |
| HO12 | 832 | 67.13 | 29.12 | 2.1527 | 0.0767 | 0.2014 | -43.57* | 0.9580 | 0.0244 |
| HO13 | 826 | 66.96 | 29.27 | 1.7698 | 0.0848 | 0.1980 | -43.05* | 1.0397 | 0.0243 |
| HO14 | 803 | 67.01 | 29.68 | 1.7209 | 0.0721 | 0.1840 | -42.60* | 1.1248 | 0.0242 |
| HO15 | 776 | 67.13 | 30.12 | 1.5085 | 0.0990 | 0.1852 | -41.32* | 1.2075 | 0.0244 |
| HO16 | 737 | 67.41 | 30.65 | 2.0374 | 0.1132 | 0.1857 | -40.40* | 1.2911 | 0.0243 |
| HO17 | 662 | 68.81 | 31.69 | 1.7305 | 0.1089 | 0.1892 | -37.75* | 1.3750 | 0.0244 |
| HO18 | 569 | 71.22 | 33.14 | 1.5685 | 0.0979 | 0.1946 | -33.98* | 1.4574 | 0.0244 |

rules regarding the last trading day.[4]

For both, the price as well as the return data, the Dickey-Fuller test of an unit root shows clear evidence that all time series are integrated of order 1.[5] Table 4 presents the results of the Johansen (1991) Trace test as well as the Maximum Eigenvalue test for the three different futures price series with identical maturity. The results yield clear evidence of a co-integration relationship among these commodity markets. The Trace test of no co-integration relationship is rejected at the 1% level for all maturities. The Maximum Eigenvalue test is significant at the 5% level for four maturities and at the 1%

[4]According to to the NYMEX (www.nymex.com) the following rules apply: (i) Crude oil: Trading terminates at the close of business on the third business day prior to the 25th calendar day of the month preceding the delivery month. If the 25th calendar day of the month is a non-business day, trading shall cease on the third business day prior to the business day preceding the 25th calendar day. (ii) Heating oil and gasoline: Trading terminates at the close of business on the last business day of the month preceding the delivery month.

[5]We have conducted Augmented-Dickey-Fuller Tests for various lag-lengths as well, yielding identical conclusions.

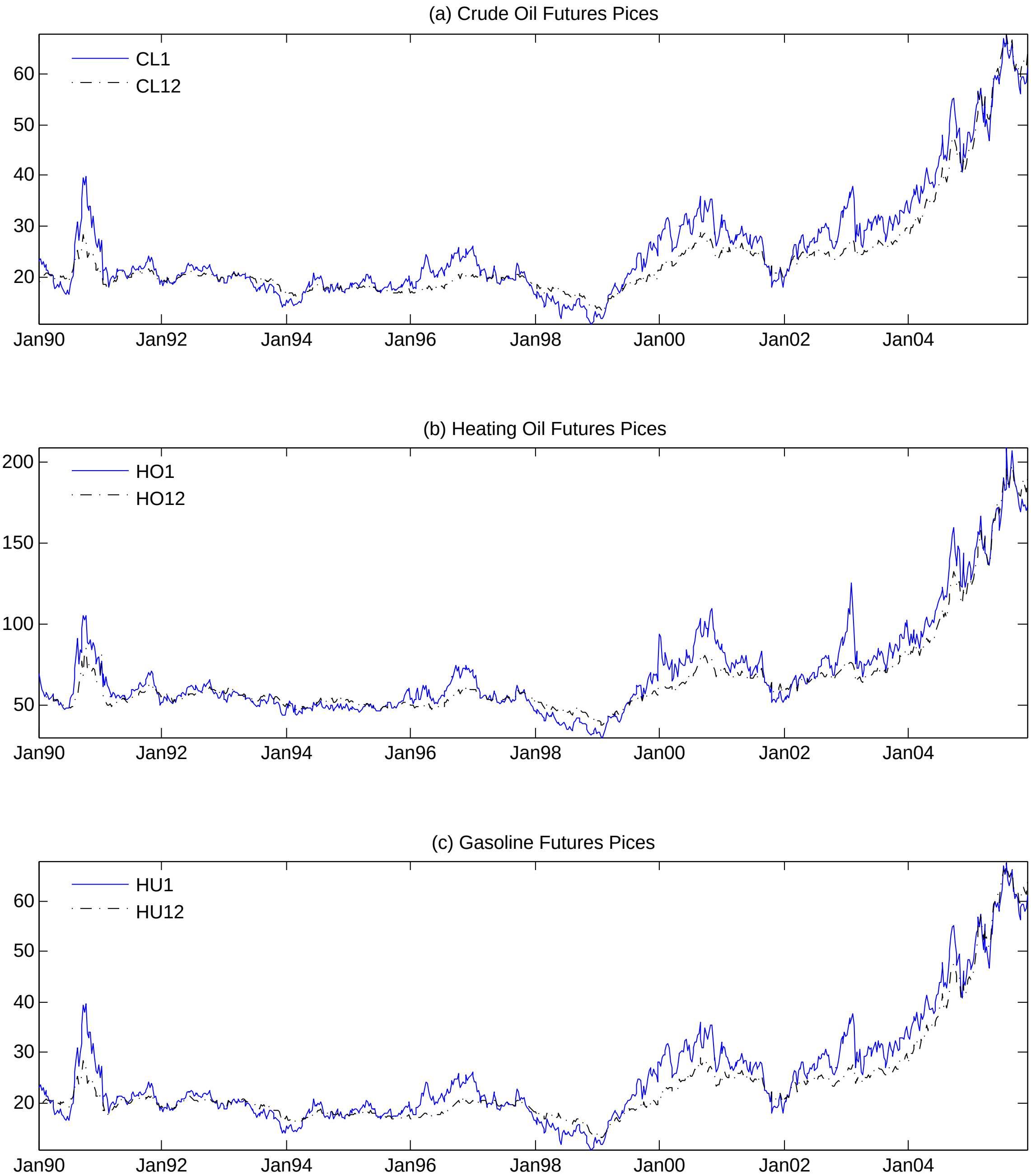


Figure 1: **Futures Prices**

This figure shows weekly future prices of one month and twelve months futures for (a) crude oil (CL1 and CL12), (b) heating oil (HO1 and HO12) and (c) gasoline (HU1 and HU12) from January 1990 to December 2005. Prices for crude oil are in US Dollars per barrel, prices for heating oil and gasoline are in Cents (USD) per gallon.

Table 3: **Descriptive Summary Statistics of Gasoline Futures**

*This table reports summary statistics for the gasoline futures data. The first column reports the futures Bloomberg Ticker wherein the number corresponds to the maturity in months. Prices are in US Dollar Cents per gallon. DF stands for the Dickey-Fuller test of a unit root. * indicates significance at the 1% level.*

| | | Price | | | Return | | | Maturity | |
|---|---|---|---|---|---|---|---|---|---|
| Contract | NOBS | Mean | St.Dev. | DF | Mean | St.Dev. | DF | Mean | St.Dev. |
| HU1 | 835 | 75.05 | 30.69 | -1.9374 | 0.0630 | 0.3900 | -43.35* | 0.0391 | 0.0242 |
| HU2 | 835 | 74.33 | 30.20 | -1.2397 | 0.0633 | 0.3339 | -42.65* | 0.1225 | 0.0242 |
| HU3 | 835 | 73.67 | 29.63 | -0.7187 | 0.0634 | 0.3024 | -42.37* | 0.2060 | 0.0243 |
| HU4 | 835 | 73.13 | 29.34 | -0.0583 | 0.0644 | 0.2712 | -43.33* | 0.2892 | 0.0244 |
| HU5 | 835 | 72.60 | 29.22 | 0.2041 | 0.0666 | 0.2611 | -43.09* | 0.3727 | 0.0243 |
| HU6 | 833 | 72.19 | 29.24 | 0.6238 | 0.0691 | 0.2516 | -42.51* | 0.4561 | 0.0244 |
| HU7 | 825 | 71.95 | 29.43 | 0.5605 | 0.0762 | 0.2446 | -43.51* | 0.5394 | 0.0244 |
| HU8 | 810 | 71.76 | 29.69 | 0.7093 | 0.0631 | 0.2431 | -42.59* | 0.6229 | 0.0243 |
| HU9 | 780 | 71.83 | 30.08 | 0.7784 | 0.0732 | 0.2350 | -41.69* | 0.7063 | 0.0244 |
| HU10 | 711 | 72.54 | 30.84 | 0.3192 | 0.0555 | 0.2355 | -41.05* | 0.7895 | 0.0243 |
| HU11 | 652 | 73.07 | 31.59 | 0.4610 | 0.0630 | 0.2325 | -39.24* | 0.8731 | 0.0242 |
| HU12 | 575 | 74.672 | 32.71 | 0.0309 | 0.0580 | 0.2416 | -36.86* | 0.9565 | 0.0244 |

level for the remaining maturities. Both tests also indicate the existence of more than one co-integrating vector at high significance levels.[6]

# III Integrated Commodity Model

In this section, we develop a Gaussian $n$-factor model for related commodity futures prices. Our model can be viewed as an extension of Schwartz and Smith (2000)[7]. We generalize their model by allowing for more complex factor dependencies and we show how multiple related commodities can be modeled simultaneously.

Uncertainty in the economy is represented by a filtered probability space $(\Omega, \mathcal{F}, \mathbb{P})$, on which an independent $n$-dimensional standard Brownian motion $Z_t^P$ is defined. All stochastic processes are assumed to be adapted to the filtration $\mathcal{F}_t$ and to be well

[6]This is, of course, only true for maturities up to 12 months as there cannot be two co-integrating vectors when considering only two series.

[7]We are not modeling convenience yields in an explicit form. As the authors showed, the canonical form we will use is equivalent to an explicit economic specification with convenience yields.

Table 4: **Co-Integration Analysis**

*This table reports the test statistics of the Johansen Trace and Maximum Eigenvalue tests. The column headlines state the null hypothesis were $r$ is the number of co-integrating vectors. The alternative hypothesis for the trace test is a greater value of $r$, whereas $r+1$ is the alternative hypothesis at the maximum eigenvalue test. * indicates significance at the 1%, $^\circ$ at the 5% level.*

| | Trace | | Maximum Eigenvalue | |
|---|---|---|---|---|
| Maturity | $r=0$ | $r \leq 1$ | $r=0$ | $r=1$ |
| 1 | 141.26* | 42.25* | 99.01* | 42.14* |
| 2 | 114.88* | 29.75* | 85.14* | 29.67* |
| 3 | 101.27* | 23.95* | 77.33* | 23.94* |
| 4 | 95.61* | 19.06° | 76.55° | 18.69* |
| 5 | 101.92* | 19.01° | 82.91° | 18.39* |
| 6 | 124.10* | 24.07* | 100.04* | 23.54* |
| 7 | 108.40* | 25.63* | 82.77* | 25.02* |
| 8 | 97.44* | 22.59* | 74.85* | 21.32* |
| 9 | 91.21* | 20.85* | 70.37° | 18.65* |
| 10 | 83.10* | 19.89° | 63.21° | 17.78* |
| 11 | 78.66* | 22.04* | 56.62* | 21.30* |
| 12 | 74.69* | 25.86* | 48.82* | 25.75* |
| 13 | 69.17* | NA | 68.59* | NA |
| 14 | 80.51* | NA | 78.57* | NA |
| 15 | 99.31* | NA | 93.15* | NA |
| 16 | 51.54* | NA | 48.94* | NA |
| 17 | 109.03* | NA | 100.69* | NA |
| 18 | 38.19* | NA | 33.12* | NA |

defined satisfying the usual regularity conditions. For notational convenience conditional expectations and variances at time $t$ with respect to the probability measure $\mathbb{P}$ of a random vector $x_T$ are denoted by $\mathbf{E}_t^P[x_T]$ and $\mathbf{V}_t^P[x_T]$, respectively.

We assume that the log-spot price of commodity $k$ can be represented by

$$\ln S_{k,t} = \delta_k \, x_t + \delta_k^0 + s_k(t), \tag{1}$$

where $\delta_k^0$ is the constant log-price level, $\delta_k$ is a $(1 \times n)$ vector of factor loadings, and $s_k(t)$ deterministically adjusts for seasonality effects in each commodity $k$. Prices are driven by the $n$-dimensional vector $x_t$ of latent state variables with Gaussian diffusion

$$dx_t = (a^P - K^P x_t)dt + dZ_t^P \tag{2}$$

where $a^P$ is a $(n \times 1)$ vector and $K^P$ a $(n \times n)$ positive semi-definite matrix. In the spirit of Schwartz and Smith (2000) we wish to decompose the price dynamics into a common non-stationary long-term component and short-term deviations from this equilibrium. Therefore, we assume the first state variable to follow a standard arithmetic Brownian motion superimposed by the cross-effects of stationary Ornstein-Uhlenbeck processes. Thus, changes in the first variable are persistent and represent fundamental changes in the economic environment. The dependence on the weighted non-stationary factor incorporates the empirical fact that commodity prices are not only correlated but also co-integrated. Note that zero weights result in no co-integration. All other variables capture deviations from the equilibrium process. As $K^P$ needs not be diagonal, we allow for interdependencies between the different factors.[8] The model of Schwartz and Smith (2000) is a special case of our model with only one commodity and and a restricted matrix $K^P$.

As derived in the Appendix, the state variable $(x_T|\mathcal{F}_t)$ with $t < T$, is normally distributed with mean

$$\mathbf{E}_t^P[x_T] = \Psi^P(t,T)x_t + \Phi^P(t,T)a^P \tag{3}$$

and variance

$$\mathbf{V}_t^P[x_T] = \Omega^P(t,T), \tag{4}$$

where the matrix-valued functions $\Psi^P(t,T)$, $\Phi^P(t,T)$, and $\Omega^P(t,T)$ are provided in equations (19), (20), and (21).

Following the well known interest rate literature (see e.g. Duffee (2002) and Dai and Singleton (2002)) we allow for affine-linear market prices of risk. Consequently, the change of measure is of the form

$$dZ_t^Q = dZ_t^P + (\lambda + \Lambda x_t)dt$$

where $Z_t^Q$ is an orthonormal Brownian motion under the new pricing measure $\mathbb{Q}$ equivalent to $\mathbb{P}$. $\lambda$, resp. $\Lambda$, are a $(n \times 1)$, resp.$(n \times n)$ matrix of market prices of risk. Existence

[8] We choose to use orthonormal Brownian motion as diffusion part, since the correlations and volatilities will be captured by other parameters, as will become obvious in the risk neutral valuation framework.

of a risk neutral pricing measure ensures arbitrage-free prices.[9] Under the risk-neutral measure the dynamics of the latent factors follow

$$dx_t = (a^Q - K^Q x_t)dt + dW_t^Q. \tag{5}$$

Since the factors are latent and we are interested in the model with the maximal number of identifiable parameters we first apply the techniques presented by Dai and Singleton (2000) and Cassasus and Collin-Dufresne (2005) to reduce equation (5) to its canonical form. As noted by these authors we can identify in the Gaussian case as many parameters as observing them separately. We specify the model in its canonical form under $\mathbb{Q}$ merely to facilitate fast estimation. Therefore $K^Q$ reduces to a $(n \times n)$ upper triangular matrix with elements $k_{i,j}^Q (i, j = 1, ..., n)$. The assumption of a non-stationary process in the first state variable under the risk neutral measure implies that only the first element of $a^Q$ is different from zero and $k_{1,1}^Q$ equals zero.

Futures $F_k(t, T)$ with different maturities $T$ for each commodity $k$ are traded in the market. Standard theory within affine frameworks implies futures prices to be equal to the risk neutral expectation of the spot price at maturity[10], i.e.

$$\begin{aligned}
\ln F_k(x_t, t; T) &= \mathbf{E}_t^Q[\ln S_{k,T}] + \tfrac{1}{2}\mathbf{V}_t^Q[\ln S_{k,T}] \\
&= \delta_k \, \mathbf{E}_t^Q[x_T] + \delta_k^0 + s_k(T) + \tfrac{1}{2}\mathbf{V}_t^Q[\delta_{x,k}\, x_T] \\
&= \left[\, \delta_k \, \Psi^Q(t,T) \right] x_t + \left[\, \delta_k \Phi^Q(t,T) a^Q + \tfrac{1}{2}\delta_k \Omega^Q(t,T)\delta_k^T + \delta_k^0 + s_k(T) \,\right] \\
&\equiv \mathcal{A}_k^Q(t,T) x_t + \mathcal{B}_k^Q(t,T),
\end{aligned} \tag{6}$$

where closed form solutions for the functions $\Psi^Q(t, T)$, $\Phi^Q(t, T)$, and $\Omega^Q(t, T)$ are provided in the Appendix (equations (19), (20), and (21)).

Applying Ito's lemma on $\ln F_k(x_t, t; T)$ one can easily recover the dynamics of Black (1976)

[9] See Harrison and Pliska (1981).

[10] Strictly speaking, this is true for forward prices only. We are aware of the fact that futures and forwards may have different values in certain economic environments. For a clear cut exposition of the differences in a similar framework, see e.g. Miltersen and Schwartz (1998). In what follows we abstract from these differences and treat the two instruments as equal.

and thus, the instantaneous volatility of returns

$$\frac{\mathbf{V}_t^Q[d\ln F_k(x_t,t;T)]}{dt} = \delta_k\,\Psi^Q(t,T)\Psi^Q(t,T)'\delta_k' \tag{7}$$

of commodity $k$ and the covariance between the two commodities $k$ and $l$

$$\frac{\mathbf{C}_t^Q[d\ln F_k(x_t,t;T_k);d\ln F_l(x_t,t;T_l)]}{dt} = \delta_k\,\Psi^Q(t,T_k)\Psi^Q(t,T_l)'\delta_l'.$$

The spot price volatility and the long-term covariation of two commodity futures is respectively

$$\frac{\mathbf{V}_t^Q[d\ln S_t]}{dt} = \delta_k{\delta'}_k \tag{8}$$

$$\frac{\mathbf{C}_t^Q[d\ln F_k(x_t,t;T_k);d\ln F_l(x_t,t;T_l)]}{dt} \overset{T\to\infty}{\longrightarrow} \delta_{k,1}\delta_{l,1}. \tag{9}$$

The variance structure of futures only depends on the parameters in the vector $\delta_k$ at the short end and in the long term behavior. The matrix $K^Q$ describes the strength of decay of volatility between spot prices and the long end of the term structure.

Our integrated model derived above nests many other models such as Gibson and Schwartz (1990), Schwartz (1997), Ross (1997), and Schwartz and Smith (2000). Note that if one wants to go along the line of Cassasus and Collin-Dufresne (2005) and incorporate a term structure of interest rates as well, one can treat zero bonds as yet another "commodity" $\overline{k}$ with adapted boundary condition $F_{\overline{k}}(T,T) = 1$.

# IV Estimation

In this section the integrated model is implemented for crude oil, heating oil and gasoline as a six-factor model. The choice of six latent factors is, to some extend, arbitrary. However, as it is well known in the empirical commodity literature, two-factor models seem to do the best job in describing the dynamics of one commodity (see e.g. Schwartz (1997)). As we consider three related commodities we suspect that six factors should be sufficient.

Our choice is motivated for expositional reasons as well. When applying the integrated model there will be a natural benchmark to compare our results with (see Section VI). The entire analysis was conducted for the general case with an affine-linear market price of risk and a restricted model, imposing $\Lambda = 0$, i.e. constant market price of risk. As the results did not improve significantly using the unrestricted model we present the estimation for the latter case only.[11] Note, that this restriction implies $K^P = K^Q \equiv K$.[12]

To estimate the model parameters we fit observed futures prices employing standard Kalman filtering[13]. Thus, we are able to explore time series as well as cross-sectional properties of the data at the same time. It is well known that in linear and Gaussian models, the Kalman filter is the optimal filter.

In what follows, we present the state space form of our model first in the general notation following Harvey (1989) and afterwards using the functions derived in the previous section. The state space transition equation can be deduced from equations (3) and (4),

$$\begin{aligned} x_{t+\Delta t} &= Tx_t + c + \eta_{\Delta t} \\ &= \underset{(6\times6)}{\Psi(t,t+\Delta t)} \underset{(6\times1)}{x_t} + \underset{(6\times6)}{\Phi(t,t+\Delta t)} \underset{(6\times1)}{a^P} + \underset{(6\times1)}{\eta_{\Delta t}} \end{aligned} \tag{10}$$

for time step $\Delta t$ and $\eta_{\Delta t}$ serially uncorrelated, normally distributed disturbances with zero mean and constant variance

$$\mathbf{V}_t^P[\eta_{\Delta t}] = \underset{(6\times6)}{\Omega(t,t+\Delta t)} \; .$$

The measurement equations for one commodity $k$ at time $t$ is given by adding measurement errors $\varepsilon$ to equation (6), hence

[11] Allowing for affine market prices of risk adds another 36 parameters. As a consequence, already noted by Duffee (2002), the maximum likelihood function has a large number of local maxima in this framework.

[12] Details on the estimation equations for the general case are available upon request.

[13] For a rigorous treatment of Kalman filtering see e.g. Harvey (1989) and the references therein.

$$\begin{aligned}
\underset{(n(k,t)\times 1)}{y_{k,t}} &= \ln F_k(x_t, t; T) + \varepsilon_{k,t} \\
&= Z_k x_t + d_k + \varepsilon_{k,t} \\
&= \underset{(n(k,t)\times 6)}{\mathcal{A}_k(t,T)}\; x_t + \underset{(n(k,t)\times 1)}{\mathcal{B}_k(t,T)} \;+\; \underset{(n(k,t)\times 1)}{\varepsilon_{k,t}}
\end{aligned} \tag{11}$$

where $y_{k,t}$ is the vector of futures log-prices at time $t$ of commodity $k$ for all $n(k,t)$ available maturities. The time-varying size of the matrices in (11) is due to missing observations. All commodity prices are stacked into a vector $y_t$ of length $n(t) = n(1,t)+n(2,t)+n(3,t)$. The vector $\varepsilon_t$ of serially uncorrelated, normally distributed disturbances captures the differences between observed and theoretical prices.

Lastly, we assume the trigonometrical functional form

$$s_k(t) = \sum_{m_k=1}^{M_k} \gamma_{k,m_k} \cos(2\pi\, m_k t) + \bar{\gamma}_{k,m_k} \sin(2\pi\, m_k t) \tag{12}$$

for the seasonality component in $\mathcal{B}_k(t,T)$ as proposed by Hannan et al. (1970).[14] It is well known that the crude oil market does not exhibit significant seasonality and thus, can be modelled without a seasonal component[15] which implies $M_1 = 0$. Heating oil and gasoline, however, do exhibit seasonal behavior. The origin of this can be directly linked to the demand side of heating oil, which is, obviously, changing during the year. Since the crack ratio, i.e. the ratio of heating oil and gasoline (and other minor derivatives) refined from one barrel of crude oil cannot be changed discretionary, an increase in heating oil production will also yield an increased production of gasoline. Consequently the price of gasoline will decrease, as demand stays relatively stable throughout the year. As we wish to keep our model parsimonious and there is no seasonality detectable for periods of less than a year we choose $M_2 = M_3 = 1$.

[14] This specification of the seasonality adjustment is frequently used when modelling commodity prices, for instance Sorensen (2002).

[15] See e.g. Schwartz (1997) or Schwartz and Smith (2000).

Table 5: **Estimation Results**

*This table reports the maximum-likelihood estimates based on the Kalman filter. The sample period is 01/01/1990 through 12/31/2005 with weekly sampling frequency.*

| Parameter | Estimate | SE | Parameter | Estimate | SE | Parameter | Estimate | SE |
|---|---|---|---|---|---|---|---|---|
| $k_{12}$ | 0.0369 | 0.1285 | $x_{10}$ | 0.8909 | 0.9920 | $\delta_{11}$ | -0.1294 | 0.0011 |
| $k_{13}$ | 0.5989 | 0.0196 | $x_{20}$ | 1.5032 | 0.7873 | $\delta_{12}$ | 0.1827 | 0.0080 |
| $k_{14}$ | 0.0359 | 0.0075 | $x_{30}$ | 0.5641 | 0.3761 | $\delta_{13}$ | -0.0206 | 0.0027 |
| $k_{15}$ | -0.4966 | 0.0668 | $x_{40}$ | -0.0187 | 0.3271 | $\delta_{14}$ | -0.3127 | 0.0018 |
| $k_{16}$ | -0.4409 | 0.0485 | $x_{50}$ | -0.4366 | 0.2136 | $\delta_{15}$ | -0.1566 | 0.0031 |
| $k_{22}$ | 2.2079 | 0.0114 | $x_{60}$ | 0.6141 | 0.1879 | $\delta_{16}$ | 0.0054 | 0.0068 |
| $k_{23}$ | -0.1615 | 0.0127 | $a_1^Q$ | 0.0175 | 0.0201 | $\delta_{21}$ | -0.1331 | 0.0012 |
| $k_{24}$ | 0.7519 | 0.0096 | $a_1^P$ | -0.4222 | 0.0604 | $\delta_{22}$ | 0.2007 | 0.0061 |
| $k_{25}$ | 0.9768 | 0.0514 | $a_2^P$ | 0.1752 | 0.3205 | $\delta_{23}$ | -0.1065 | 0.0010 |
| $k_{26}$ | 0.4500 | 0.0286 | $a_3^P$ | -0.6340 | 0.7029 | $\delta_{24}$ | -0.3053 | 0.0028 |
| $k_{33}$ | 0.6613 | 0.0075 | $a_4^P$ | -0.1823 | 0.3869 | $\delta_{25}$ | -0.0369 | 0.0033 |
| $k_{34}$ | 0.1410 | 0.0075 | $a_5^P$ | -0.4276 | 0.2105 | $\delta_{26}$ | 0.1245 | 0.0041 |
| $k_{35}$ | -0.4111 | 0.0435 | $a_6^P$ | 0.2522 | 0.3108 | $\delta_{31}$ | -0.1260 | 0.0011 |
| $k_{36}$ | 0.4370 | 0.0149 | $\gamma_2$ | 0.0408 | 0.0001 | $\delta_{32}$ | 0.2936 | 0.0072 |
| $k_{44}$ | 0.8071 | 0.0037 | $\bar{\gamma}_2$ | -0.0072 | 0.0002 | $\delta_{33}$ | -0.0817 | 0.0018 |
| $k_{45}$ | -0.2082 | 0.0251 | $\gamma_3$ | -0.0440 | 0.0002 | $\delta_{34}$ | -0.2482 | 0.0021 |
| $k_{46}$ | 0.3569 | 0.0121 | $\bar{\gamma}_3$ | 0.0151 | 0.0001 | $\delta_{35}$ | -0.0612 | 0.0042 |
| $k_{55}$ | 2.7888 | 0.0240 | $\delta_1^0$ | 2.9710 | 0.0111 | $\delta_{36}$ | -0.0803 | 0.0027 |
| $k_{56}$ | 0.3462 | 0.0288 | $\delta_2^0$ | 4.0149 | 0.0108 | | | |
| $k_{66}$ | 1.9105 | 0.0211 | $\delta_2^0$ | 4.0496 | 0.0109 | | Log-likelihood: | 159,950 |

# V Results

Table 5 shows the maximum likelihood parameter estimates for the data set described in Section II. The left part reports the elements of the matrix $K$, which is upper diagonal. On the right, the $(3 \times 6)$ matrix of factor loadings is reported. The filtered starting values $x_{i0}$, the drift vectors under the real and risk neutral measure $a^P$ and $a^Q$, the seasonality parameters $\gamma_i$ and $\bar{\gamma}_i$, $i = 2, 3$, as well as the level vector $\delta_k^0$ are reported in the middle column of the table. Note that $k_{11}$ and $a_i^Q$, $i = 2, ..., 6$ are zero by construction and thus, not reported.

All elements, except for $k_{12}$, of the matrix $K$ are statistically significant. The main diagonal elements $k_{ii}$ can be interpreted as mean-reversion parameters as they represent the eigenvalues of $K$. Their size varies from 0.66 ($k_{33}$) through 2.79 ($k_{55}$). Converting these into "half-lifes" of deviations from equilibrium, we get values between 3 and 13 months, indicating different degrees of persistence of the price shocks due to the different factors.[16] The significance of almost all off-diagonal elements indicates their value in explaining the dependence structure among the three commodity markets.

The factor loadings are, except for $\delta_{16}$, also all significantly different from zero. The fact that $\delta_{16}$ is close to zero, indicates that the sixth stochastic factor does not drive the price of crude oil. However, $\delta_{26}$ as well as $\delta_{36}$, are non-zero, which shows that $x_6$ captures deviations from equilibrium on the refined product markets.

In Figure 2 the filtered state variables, weighted with their respective factor loadings are plotted for each commodity spot price. The sum of each of the six components equals the filtered log-spot price. Thus, we can see how each factor attributes to the three different markets. The upper left graph displays the filtered equilibrium process. The factor loading of the non-stationary equilibrium process, modelled by state variable $x_1$, is non-zero for each market. Note, that in contrast to Schwartz and Smith (2000) we do not assume the existence of such a process ex-ante. The case of no common equilibrium factor is the special case of our model where all or some of the $\delta_{k1}$ equal zero. Moreover, the sensitivities towards this equilibrium process are of equal size, namely $\delta_{11} \cong \delta_{21} \cong \delta_{31} \cong -0.13$, revealing that indeed not only a common non-stationary process can be identified, but also that all prices react by the same degree regarding the long-term equilibrium.

The weighted state variables $x_2$ and $x_4$ demonstrate very similar behavior for all three markets. Thus, they can be interpreted as shocks resulting in deviations from equilibrium affecting all energy commodities in a similar fashion. These can be best explained by the supply side of the market since shortages of crude oil will propagate to its refined products markets. However, the persistence of shocks in the two state variables is different. Whereas $x_2$ has a half-life of 0.32 years, i.e. about 4 months, deviations due to shocks

[16] The "half-life" can be computed as $\ln(2)/k$.

from $x_4$ take more than 10 months (half-life of 0.86 years) to halve instead. Thus, we find two factors with different persistence levels causing deviations from equilibrium in the energy markets considered.

The state variable $x_3$ demonstrates different behavior. Almost no contribution to the crude oil price process, but concurrent deviations in both derivatives processes can be observed. Thus, this state variable captures shocks mainly affecting the heating oil and gasoline markets, but only weakly the crude oil market ($\delta_{13}$ is small, but statistically significant). Analogously, the weighted factor $x_5$ mainly captures deviation from the equilibrium in the crude oil market, the two refined product markets are less involved. As discussed above, $\delta_{16}$ is close to zero, thus the crude oil market is not affected by $x_6$ which is also clearly visible in the bottom right graph of Figure 2. This factor represents shocks on the refined product markets which, in contrast to shocks due to $x_3$, have different impacts on both markets.

Summarizing, we are able to identify a common equilibrium process ($x_1$), two factors causing deviations from this equilibrium with different degrees of persistence for all products, ($x_2$ and $x_4$), a factor mainly affecting the refined products in a similar way ($x_3$), a factor mainly capturing deviations on the crude oil market ($x_5$), as well as one factor only having an influence on the heating oil and gasoline markets, however, in a distinctive fashion.

The level variables $\delta_k^0$ represent the different price levels of the three commodities. These different levels are due to different trading units in the three markets as well as the costs of the refining process, which are assumed to be constant. All three level variables are highly significant.

The drift parameters $a_i^Q$ and $a_i^P$ weighted with the factor loadings $\delta_{ij}$, characterize the risk premiums of the commodity spot prices with respect to the various risk factors. More precisely, the risk premium of commodity $k$ related to factor $i$ can be computed as $\delta_{ki}(a_i^P - a_i^Q) \equiv \lambda_{ki}$. Summing over all $i$ one gets the entire risk premium for each spot commodity contract $\lambda_k = \sum_{i=1}^{6} \lambda_{ki}$. These quantities are reported in Table 6. The

Table 6: **Risk premia**

*This table reports the risk premia for the three commodity spot prices related to the six risk factors as well as the entire risk premium of each contract.*

| $\lambda_{k1}$ | $\lambda_{k2}$ | $\lambda_{k3}$ | $\lambda_{k4}$ | $\lambda_{k5}$ | $\lambda_{k6}$ | $\lambda_k$ |
|---|---|---|---|---|---|---|
| 0.0569 | 0.0320 | 0.0130 | 0.0570 | 0.0669 | 0.0014 | 0.2273 |
| 0.0585 | 0.0352 | 0.0675 | 0.0557 | 0.0158 | 0.0314 | 0.2640 |
| 0.0554 | 0.0514 | 0.0518 | 0.0452 | 0.0262 | -0.0202 | 0.2098 |

equilibrium factor risk premium is slightly above 0.055 for all three contracts. The overall risk premium is 0.23 for the crude oil, 0.26 for the heating oil, and 0.21 for the gasoline spot market.

To investigate the power of the six estimated factors in explaining the variation in spot price volatility, we calculate the ratios of factor volatilities to the overall spot price volatility. Each factor $i$ contributes $\delta_{ki}^2 \equiv \omega_{ki}\delta_k\delta_k'$ to the entire spot price variance $\delta_k\delta_k'$ of commodity $k$, which can be directly seen from (7) for $T \to t$. Table 7 displays these ratios. The equilibrium factor explains around 10% of the variation in the three spot price dynamics. The major part is captured by the two factors influencing all three markets, $x_2$ and $x_4$. The third and the sixth factor do not contribute in explaining the variation of crude oil prices, the fifth factor mainly drives the crude oil volatility.

When modelling prices of financial or real assets we also wish to be able to fit the term structure of volatilities. Equipped with the estimated parameters we are able to compare the empirical volatilities reported in Section II with the model implied ones using formula (7). The empirical and theoretical term structures are presented in Figure 3. The estimated term structures are slightly biased upwards, compared to the empirical ones, but the overall shape as well as the fit at the long and short ends seem to be satisfactory and are well in line with the Samuelson effect.[17]

The seasonality parameters $\gamma_i$ and $\bar{\gamma}_i$ are highly significant. To visualize the effects of the adjustments, Figure 4 shows the trigonometric seasonality function $\exp(s(t))$. As

[17]See Samuelson (1965).

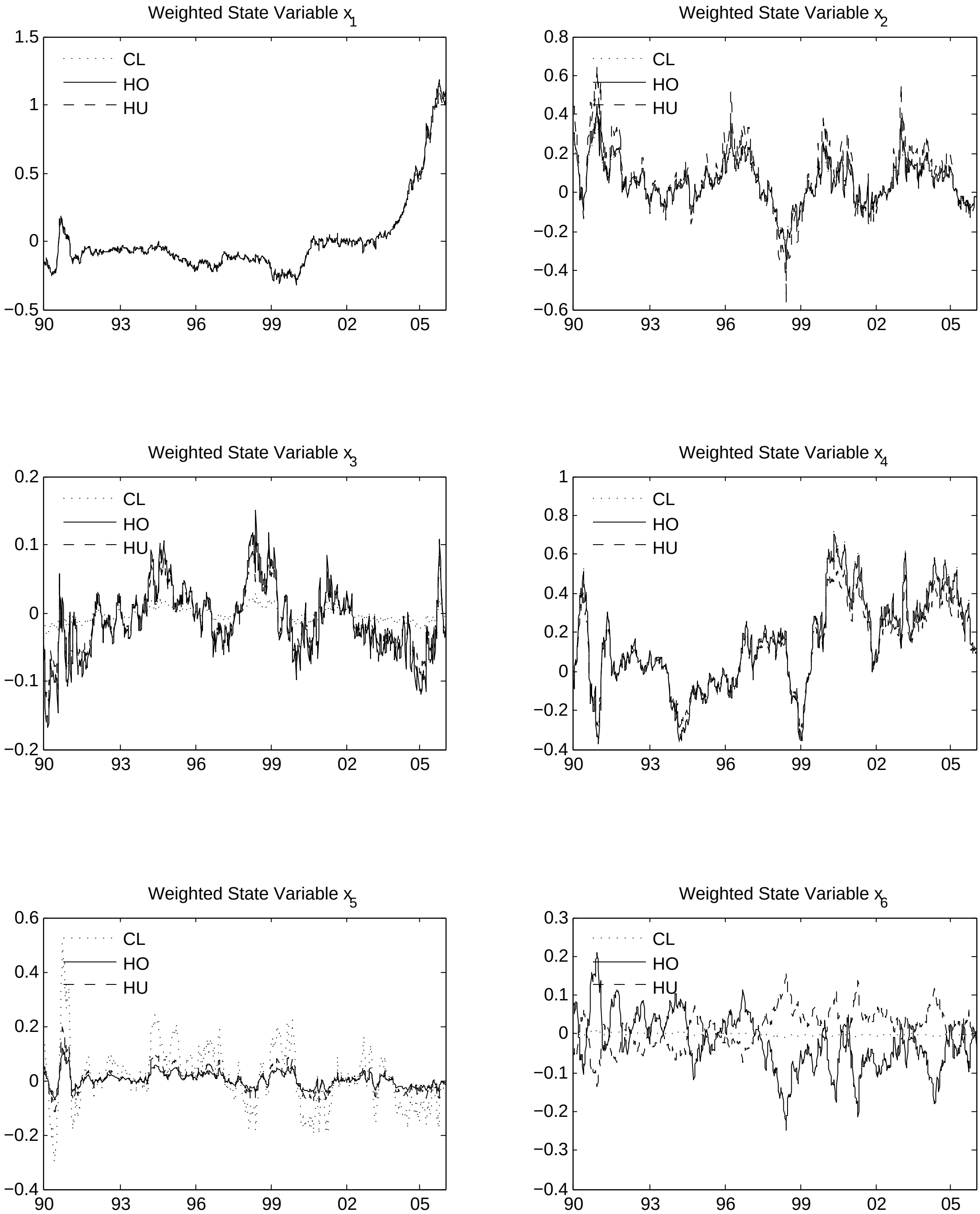


Figure 2: **Weighted State Variables**

This figure shows the weighted state variables, i.e. $\delta_i x_i$. The dotted line shows the contribution of the respective state variable to the crude oil price, the solid line corresponds to heating oil, and the dashed line to gasoline.

Table 7: **Explained variation**

*This table reports the ratios of explained variation due to the respective factor. $\omega_{ki}$ is computed as $\frac{\delta_{ki}^2}{\sum_{i=1}^{6}\delta_{ki}^2}$.*

| $\omega_{k1}$ | $\omega_{k2}$ | $\omega_{k3}$ | $\omega_{k4}$ | $\omega_{k5}$ | $\omega_{k6}$ |
|---|---|---|---|---|---|
| 0.0968 | 0.1931 | 0.0024 | 0.5657 | 0.1418 | 0.0002 |
| 0.0987 | 0.2246 | 0.0632 | 0.5196 | 0.0076 | 0.0864 |
| 0.0880 | 0.4775 | 0.0370 | 0.3411 | 0.0208 | 0.0357 |

discussed in the previous section, we can observe the expected pattern of price increases of heating oil and decrease of gasoline during the cold season and converse behavior in the summer. The price deviations captured with this deterministic part are of the size +/- 4% around the annual mean.

Table 8 presents summary statistics about the model fit. For each commodity and maturity, the mean, standard deviation and maximum of the pricing errors are reported. The crude oil futures prices are fitted best by the model. Only the one and two months errors are slightly higher, but all middle and long-term prices are fitted very well. Although the results are not comparable directly with the results of Schwartz and Smith (2000), since our dataset ranges to 2005, we can observe a better fit of the model. For instance, the one months futures standard deviation of pricing errors is 0.0226 for our model, opposed to 0.0414 in the study of Schwartz and Smith (2000), indicating that using information from the refined products markets can help explain the crude oil futures prices. One could argue that the better fit is a direct result of increasing the number of parameters. However, we do not only increase the number of stochastic factors, but also the number of modelled commodities. Thus, comparing our results of a six factor model for three commodity markets with a two factor model for one market seems reasonable.

The heating oil and gasoline futures prices are fitted with less precision. Again, the short end of the futures curve does exhibit the largest errors, the middle and long end shows a

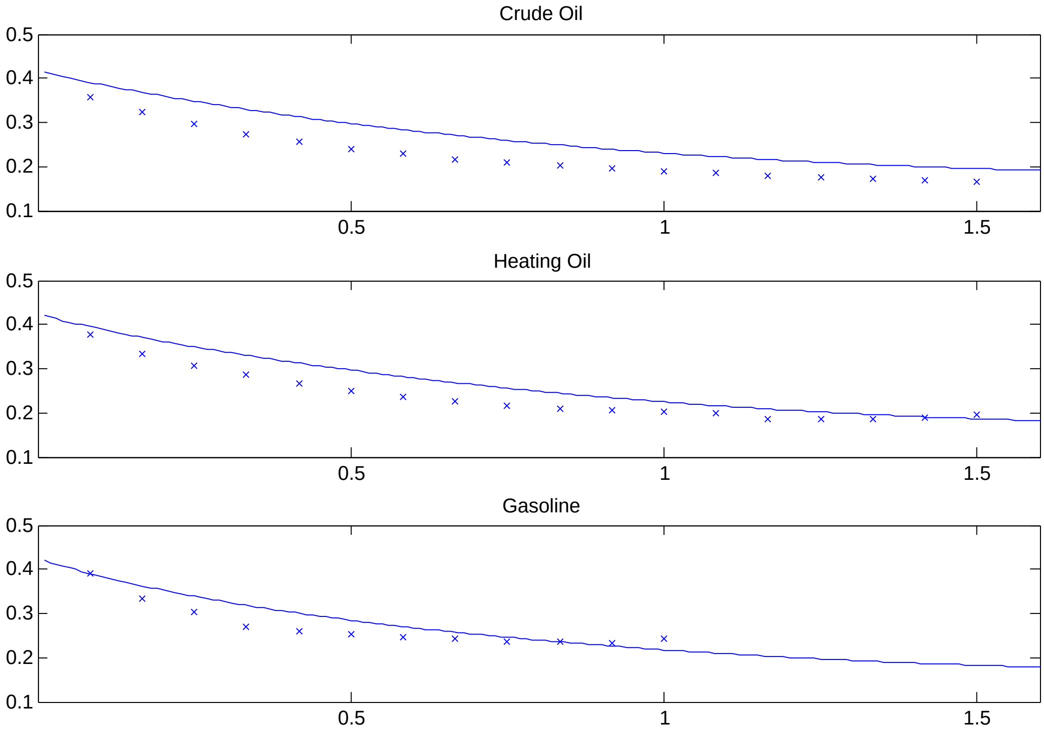


Figure 3: **Term structure of volatilities**

This figure shows the term structure of volatilities. The empirical volatilities are plot by $\times$, the model volatilities are given by the solid line. The maturities on the abscissa are given in years.

stable behavior.[18]

# VI Applications

To illustrate the economic significance of modelling multiple related commodities in an integrated framework we consider an investment project into a refinery. In analyzing a long-term horizon project, the differences between our approach and the ad-hoc approach of modelling the dependencies of the considered commodities only via a correlation structure in returns will be most noticeable.

[18]We observe one large maximum error of 0.26 for the heating oil contract with 13 months maturity, which is due to a sharp price increase from \$69 to \$82 in the second week of 1991. This may be a data error, however, as we also observe a positive trading volume, we decided not to exclude this data point from the analysis.

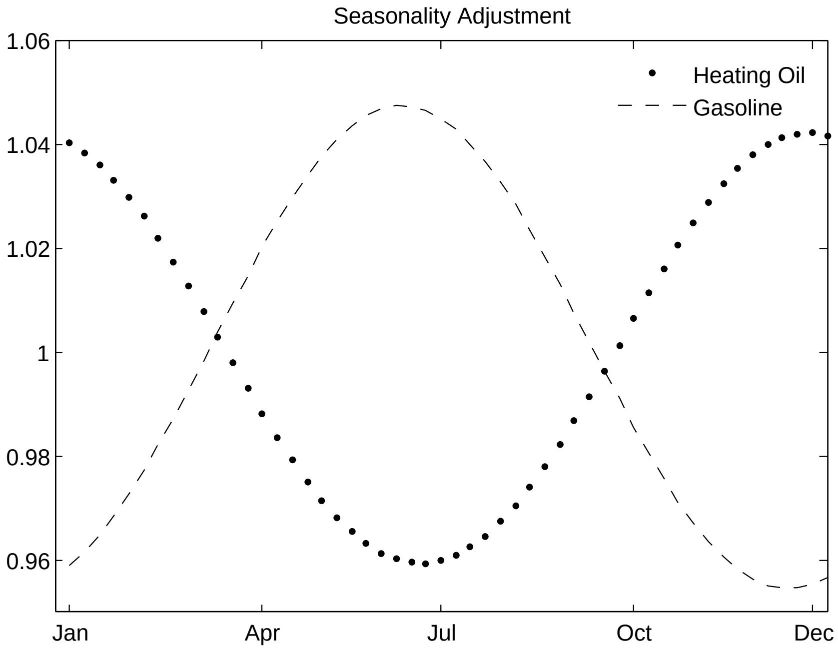


Figure 4: **Seasonality adjustments**

This figure shows the trigonometric seasonality adjustment function for one year.

We assume that the refinery can be described, from a financial point of view, as the weighted sum of three cash flows. First, one has to purchase one unit of crude oil. This is refined into $\alpha$ units of heating oil and $1-\alpha$ units of gasoline which are sold in the spot market.[19] Thus, the present value of the refinery can be represented by the following sum of cash flows

$$PV(T) = \sum_{t=1}^{T} D(0,t)[-S_{1,t} + \alpha S_{2,t} + (1-\alpha)S_{3,t}], \tag{13}$$

where $D(0,t)$ denotes the discount function, $S_1$ the crude oil spot price, $S_2$ the heating oil spot price and $S_3$ the gasoline spot price. To compute the distribution of $PV(T)$ we

[19]Notice that we make some simplifying assumptions. First, the refining process is assumed to be immediate, i.e. no time lag between purchase and selling date exists. Second, the two considered derivatives are assumed to be the only refined products of crude oil. As described in Section II, these actually represent around 75% of the refining output. These assumptions are made for tractability and keep things as simple as possible to make the main point clear. Furthermore, we neglect any costs, i.e. we calculate gross cash flows. Assuming that the process costs are deterministic, or at least independent of the spot prices, the inclusion will only shift the distribution of cash flows, not altering the shape.

Table 8: **Statistics of Pricing Errors**

*This table reports statistics of the pricing errors of the fitted six factor model. For each commodity and maturity mean pricing errors, standard deviation (s.d.) of pricing errors, and the maximal (max.) errors are reported. Pricing errors are computed as* $e_t = \ln(F(t,T)) - \ln(\hat{F}(t,T))$ *where* $F(t,T)$ *denotes the observed, and* $\hat{F}(t,T)$ *the fitted futures price with maturity* $T$ *and time* $t$.

| | Crude Oil | | | Heating Oil | | | Gasoline | | |
|---|---|---|---|---|---|---|---|---|---|
| Contract Maturity | Mean Error | S.d. of Error | Max. Error | Mean Error | S.d. of Error | Max. Error | Mean Error | S.d. of Error | Max. Error |
| 1 | -0.0034 | 0.0226 | 0.1390 | 0.0004 | 0.0281 | 0.1940 | -0.0028 | 0.0366 | 0.1408 |
| 2 | -0.0012 | 0.0087 | 0.0554 | 0.0003 | 0.0138 | 0.0598 | -0.0027 | 0.0217 | 0.0758 |
| 3 | -0.0003 | 0.0027 | 0.0149 | 0.0009 | 0.0088 | 0.0647 | -0.0016 | 0.0170 | 0.0839 |
| 4 | 0.0001 | 0.0013 | 0.0067 | 0.0010 | 0.0085 | 0.0481 | -0.0001 | 0.0146 | 0.0619 |
| 5 | 0.0002 | 0.0018 | 0.0084 | 0.0005 | 0.0108 | 0.0425 | 0.0007 | 0.0145 | 0.0698 |
| 6 | 0.0002 | 0.0018 | 0.0086 | 0.0001 | 0.0120 | 0.0501 | 0.0013 | 0.0160 | 0.0718 |
| 7 | 0.0001 | 0.0014 | 0.0095 | -0.0001 | 0.0124 | 0.0357 | 0.0015 | 0.0172 | 0.0747 |
| 8 | -0.0001 | 0.0016 | 0.0036 | -0.0003 | 0.0116 | 0.0510 | 0.0018 | 0.0179 | 0.0891 |
| 9 | -0.0002 | 0.0014 | 0.0104 | -0.0006 | 0.0104 | 0.0396 | 0.0020 | 0.0184 | 0.0754 |
| 10 | -0.0002 | 0.0016 | 0.0084 | -0.0007 | 0.0101 | 0.0424 | 0.0022 | 0.0184 | 0.0739 |
| 11 | -0.0002 | 0.0018 | 0.0062 | -0.0011 | 0.0102 | 0.0399 | 0.0025 | 0.0191 | 0.0706 |
| 12 | -0.0001 | 0.0014 | 0.0057 | -0.0017 | 0.0110 | 0.0429 | 0.0040 | 0.0197 | 0.0692 |
| 13 | 0.0000 | 0.0011 | 0.0050 | -0.0016 | 0.0150 | 0.2567 | | | |
| 14 | 0.0000 | 0.0008 | 0.0030 | -0.0019 | 0.0116 | 0.0414 | | | |
| 15 | 0.0001 | 0.0011 | 0.0041 | -0.0017 | 0.0125 | 0.0859 | | | |
| 16 | 0.0002 | 0.0016 | 0.0065 | -0.0015 | 0.0124 | 0.0382 | | | |
| 17 | 0.0003 | 0.0023 | 0.0092 | -0.0005 | 0.0124 | 0.0606 | | | |
| 18 | 0.0003 | 0.0033 | 0.0197 | -0.0001 | 0.0119 | 0.0653 | | | |

rely on simulation as the sum of three log-normal distributed random variables is not analytically feasible.

In our analysis we assume a risk free rate of 2%. The crack ratio $\alpha$ is set to 1/3, i.e. we assume that one barrel of crude oil is refined into 0.33 barrels of heating oil and 0.67 barrels of gasoline. This ratio reflects the real ratio (0.25 and 0.5) proportionally adjusted for neglecting the other refined products.

To put the results into perspective, we compare our integrated model with a natural extension of Schwartz and Smith (2000) to the multiple commodity case. The latter model was originally proposed for modelling univariate spot prices. The ad-hoc extension of this

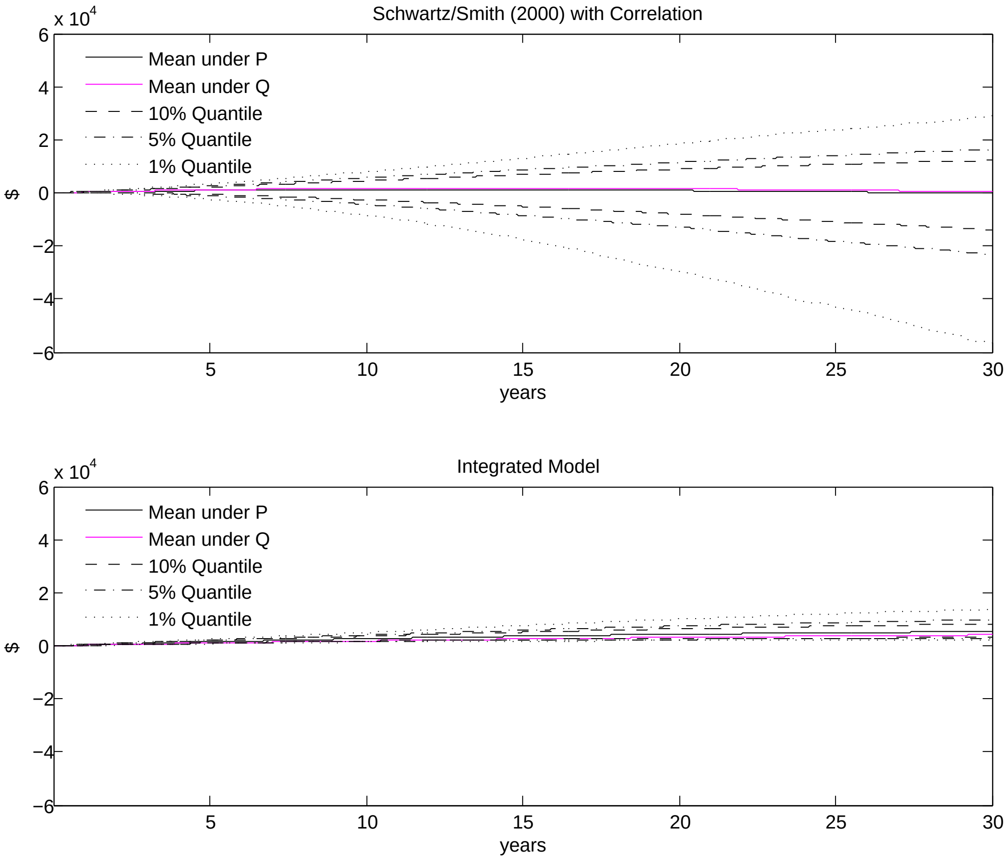


Figure 5: **Present value distribution of the refinery**

This figure shows the distribution of the present value of the refinery project for the two different modelling approaches. The upper graph provides the results employing the model of Schwartz and Smith (2000) with correlated returns and the bottom graph shows the results using the integrated model. Furthermore, the present value under $\mathbb{Q}$ is given for each model.

model for multiple commodities is to allow for non-zero correlations among the factors driving the three different commodity price dynamics and simultaneously estimating the model.

We evaluate the refinery from a risk management and pricing perspective, i.e. under the real and the risk neutral measure. The mean present value as defined in (13) is computed under $\mathbb{Q}$ for all $T$ up to 30 years. Under $\mathbb{P}$ we simulate the entire distribution of $PV(T)$. The results are provided in Figure 5. The upper graph shows the present value with respect to the maturity of the project using the extended model of Schwartz and Smith (2000), the lower graph displays the results using the integrated model for one unit per week. One can clearly observe a significant difference in the two graphs under $\mathbb{P}$, whereas the difference under $\mathbb{Q}$ is small. Neglecting the co-integrated behavior of the three price processes leads to a distribution which is much wider. Considering an average US refinery, which processes around 850,000 barrels of crude oil per week[20] at an assumed profit margin of 10%, the difference of the Value-at-Risk at the 5% level after 10 years yields approximately 400 million USD and 2.1 billion USD after 30 years.

Lastly, we show how hedging of a long-term exposure changes due to the co-movements of prices.[21] More precisely, we demonstrate how a single uncertain monthly cash flow

$$CF_T = -S_{1,T} + \alpha S_{2,T} + (1-\alpha) S_{3,T} \tag{14}$$

can be hedged efficiently. As no futures contracts for long-term horizons exist, this cannot be done by simply going long and short in the respective contracts. However, assuming that at least six futures contracts with distinctive underlyings or maturities are available for continuous trading, one can build up a riskless position at each point in time $t$. These hedge portfolios $h$ are obtained by solving the following system of linear equations:

$$\sum_{j=1}^{6} h_{jt} \frac{\partial F_j(t, T_j)}{\partial x_i} = \frac{\partial \mathbf{E}_t^Q [CF_T]}{\partial x_i} \qquad \forall i = 1, ..., 6. \tag{15}$$

To exemplify this hedging strategy we consider two futures on each commodity with maturities of three and 11 months. Notice that the use of all three commodity futures is not necessary to apply a hedging strategy within the integrated model framework. One could also rely on only one or two of the three contracts. As the factor loading $\delta_{16}$ is not significant it will be advisable in practice to use at least one crude oil and one refined product future. To keep things comparable we use two contracts for each commodity since the extended model of Schwartz and Smith (2000) requires exactly this composition of the hedging portfolio.

The hedge ratios for different maturities up to 10 years are provided in Figure 6. The

[20] See the webpage of the EIA (www.eia.doe.gov).

[21] The hedging of long-term exposures with short-term futures was discussed in the one commodity case in Schwartz (1997) and Korn (2005).

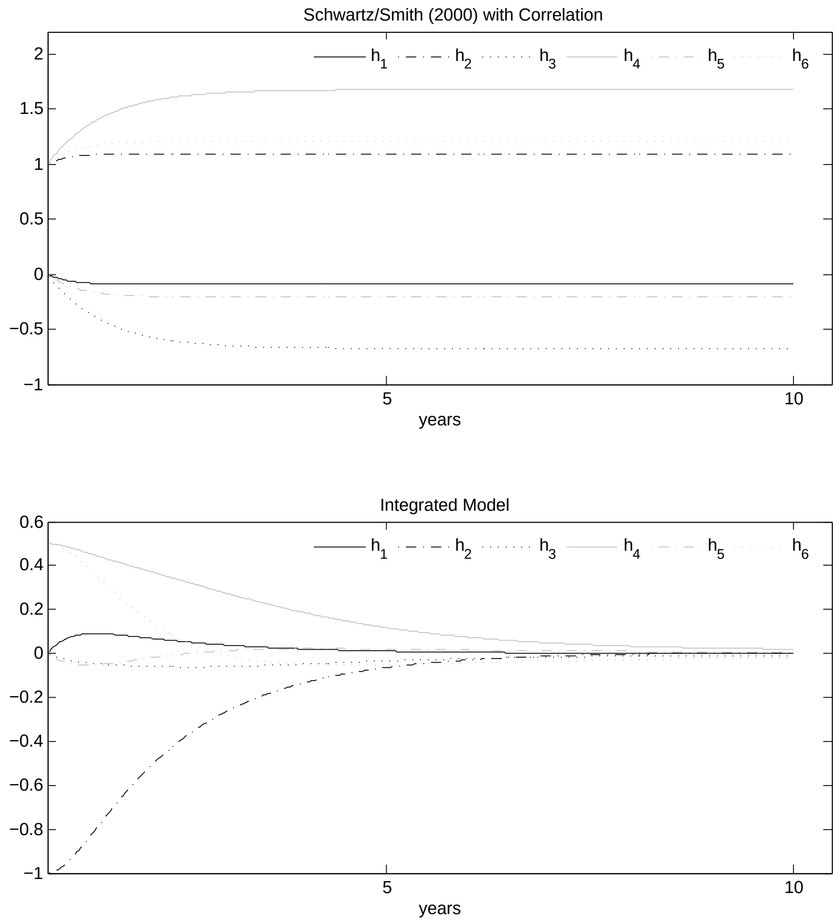


Figure 6: **Hedge ratios**

This figure shows the hedge positions in the six contracts considered to hedge one monthly cash flow of the refinery at $T$, where $T$ is on the abscissa. The upper graph gives these positions using the model of Schwartz and Smith (2000) incorporating correlations in returns, the lower graph provides the hedge positions implied by the integrated model.

upper graph shows the hedging demand employing the benchmark model, whereas the lower graph shows the hedge positions in the integrated model. It is clearly visible that both strategies differ massively. In the first case, the position in futures needed increases up to a maturity of one year and afterwards remains significantly different from zero between 0.7 and 1.6 contracts. On the contrary, the hedging positions in the integrated model decrease fast, yielding a hedging demand for a five years maturity which is already close to zero. For a long time horizon, the required exposures are between $-0.02$ and

0.02.[22] Thus, using the benchmark model for hedging a position in crude oil, heating oil and gasoline leads to a huge overhedge causing transaction costs to negatively affect the project's financial success.

# VII Summary and Conclusion

In this paper we investigate the problem of modelling multiple commodities within a single stochastic framework. This is of high relevance for any market participant being exposed to risks related to more than one commodity. As a concrete example we consider a refinery, the financial success of which is highly dependent on the prices of its main resource, crude oil and its selling products which are mainly heating oil and gasoline.

In the first step, we argue and show that the considered commodity prices are not only correlated but also co-integrated. This fact will alter the results of any economic evaluation, and therefore, must be considered when developing and applying a model describing the joint stochastic price dynamics.

Applying essential-affine modelling technique we develop an integrated latent factor model allowing for a common stochastic trend as well as any number of stationary processes which represent deviations from the long-term equilibrium. This model captures well known properties of commodity prices[23], namely backwardation, mean reversion, declining volatilities with contract horizon, seasonality as well as the co-integrated behavior described above. Furthermore, it nests many well known models such as Gibson and Schwartz (1990), Schwartz (1997), Schwartz and Smith (2000), and Cassasus and Collin-Dufresne (2005).[24]

[22] The fact of decreasing hedging demand with increasing maturity is an implication of the equal sensitivities towards the common equilibrium process. Considering exemplary two commodities, the variance of a futures contract which exchanges commodity $k$ for $l$ at time $T$, $-S_{k,T} + S_{l,T}$, will have a variance of approximately $(\delta_{k,1}^2 - 2\delta_{k,1}\delta_{l,1} + \delta_{l,1}^2) \cong 0$, if $\delta_{k,1} \cong \delta_{l,1}$.

[23] See e.g. Routledge et al. (2000).

[24] Cassasus and Collin-Dufresne (2005) also consider a jump diffusion specification which is not nested in our framework, however, it is noted by the authors that the jump component does not contribute significantly to explaining the price process.

Using NYMEX futures data, a six factor model for crude oil, heating oil and gasoline is estimated using standard Kalman filtering and maximum likelihood. We find that the core features of our model are highly significant. We are not only able to identify a non-stationary common long-term component driving all commodity prices but also find that all commodities exhibit the same factor loading. Two of the stationary factors impact the markets in a common fashion, however with different levels of persistence. In contrast, the other three components influence the three commodities in a distinctive way.

To emphasize the economic significance of our results we apply the integrated model in the context of a refinery. From the viewpoint of risk management, the dispersion of the present value distribution decreases severely when compared to an approach neglecting the co-movements of commodity prices. Finally, we show when hedging a single cash flow with short-term futures contracts, the optimal hedging changes considerably for short, and even more for long horizons. For the latter case, the hedging demand decreases significantly.

# A Appendix

Consider the dynamics

$$dx_t = (a^M - K^M x_t)dt + dZ_t^M$$

under any $P$-equivalent measure $M$, where $Z_t^M$ is a standard Brownian motion under $M$, $a^M$ is a constant vector and $K_M$ is a positive semi-definite constant matrix. Decomposing the matrix $K^M \equiv UVU^{-1}$ where $V$ is the diagonal matrix of distinctive eigenvalues $\{v_i \geq 0\}$ of $K^M$ and $U$ the matrix of associated eigenvectors. Defining the functions

$$\psi(v; t, T) = \exp(-v(T-t)) \overset{v \to 0}{\longrightarrow} 1 \tag{16}$$

and

$$\phi(v; t, T) = \int_t^T \psi(v; s, T)ds = \frac{(1 - \exp(-v(T-t))}{v} \overset{v \to 0}{\longrightarrow} (T-t) \tag{17}$$

and the matrix

$$\mathcal{L}_\psi(K^M; t, T) = \begin{pmatrix} \psi(v_1; t, T) & 0 & \cdots & 0 \\ 0 & \psi(v_2; t, T) & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \psi(v_n; t, T) \end{pmatrix} \tag{18}$$

and $\mathcal{L}_\phi$ analogously, where the function $\psi$ is replaced by $\phi$. Integrating the matrix $\mathcal{L}_\psi$ from $t$ to $T$ results in the matrix $\mathcal{L}_\phi$. The matrix $U\mathcal{L}_\psi(K^M; t, T)U^{-1}$ can be seen as matrix-valued equivalent to the function $\exp(-k(T-t))$.

First, applying Ito's lemma to the the function $U\mathcal{L}_\psi(K^M; t, T)U^{-1}x_t$, with $\frac{\partial(U\mathcal{L}_\psi(K^M; t, T)U^{-1})}{\partial t} = U\mathcal{L}_\psi(K^M; t, T)U^{-1}K$, results in

$$d(U\mathcal{L}_\psi(K^M; s, T)U^{-1}x_t) = U\mathcal{L}_\psi(K^M; s, T)U^{-1}a^M dt + U\mathcal{L}_\psi(K^M; t, T)U^{-1}dZ_t^M,$$

then integration of the LHS results in

$$\begin{aligned}\int_t^T d(U\mathcal{L}_\psi(K^M;s,T)U^{-1}x_s) &= U\mathcal{L}_\psi(K^M;T,T)U^{-1}x_T - U\mathcal{L}_\psi(K^M;t,T)U^{-1}x_t \\ &= x_T - U\mathcal{L}_\psi(K^M;t,T)U^{-1}x_t.\end{aligned}$$

Rearranging yields

$$\begin{aligned}x_T = \quad & U\mathcal{L}_\psi(K^M;t,T)U^{-1}x_t + \int_t^T U\mathcal{L}_\psi(K^M;s,T)U^{-1}a^M dt \\ & + \int_t^T U\mathcal{L}_\psi(K^M;s,T)U^{-1}dZ_s^M\end{aligned}$$

as the solution.

Thus, $(x_T|\mathcal{F}_t)$ is Gaussian and the conditional expected value can be calculated as

$$\begin{aligned}\mathbf{E}_t^M[x_T] &= U\mathcal{L}_\psi(K^M;t,T)U^{-1}x_t + \int_t^T U\mathcal{L}_\psi(K^M;s,T)U^{-1}a^M ds \\ &= U\mathcal{L}_\psi(K^M;t,T)U^{-1}x_t + U\int_t^T \left(\mathcal{L}_\psi(K^M;s,T)ds\right) U^{-1}a^M \\ &= \Psi^M(t,T)x_t + \Phi^M(t,T)a^M\end{aligned}$$

with

$$\Psi^M(t,T) = U\mathcal{L}_\psi(K^M;t,T)U^{-1} \tag{19}$$

$$\Phi^M(t,T) = U\mathcal{L}_\phi(K^M;t,T)U^{-1} \tag{20}$$

as well as the conditional variance as

$$
\begin{aligned}
\mathbf{V}_t^M[x_T] &= \mathbf{E}_t^M[\int_t^T (U\mathcal{L}_\psi(K^M;s,T)U^{-1}dZ_t^M)(U\mathcal{L}_\psi(K^M;s,T)U^{-1}dZ_t^M)'] \\
&= \int_t^T U\mathcal{L}_\psi(K^M;s,T)(U^{-1}U'^{-1})\mathcal{L}'_\psi(K^M;s,T)U'ds \\
&= U\int_t^T \mathcal{L}_\psi(K^M;s,T)H\mathcal{L}'_\psi(K^M;s,T)ds\,U' \\
&= U\mathcal{H}(K^M;t,T)U' \\
&= \Omega^M(t,T) \qquad (21)
\end{aligned}
$$

where $H \equiv U^{-1}U'^{-1}$ and the matrix $\mathcal{H}$ can be easily derived by considering one element $\mathcal{H}_{ij}$ of the matrix $\mathcal{H}$

$$
\begin{aligned}
\int_t^T \psi(v_i;s,T)H_{ij}\psi(v_j;s,T)ds &= \int_t^T H_{ij}\psi(v_i+v_j;s,T)ds \\
&= H_{ij}\phi(v_i+v_j;s,T) \\
&\equiv \mathcal{H}_{ij}(K^M;t,T). \qquad (22)
\end{aligned}
$$

# References

F. Black. The pricing of commodity contracts. *Journal of Financial Ecomomics*, 3: 167–179, 1976.

J. Cassasus and P. Collin-Dufresne. Stochastic convenience yield implied from commodity futures and interest rates. *Journal of Finance*, 60(5):2283–2331, 2005.

Q. Dai and K.J. Singleton. Specification analysis of affine term structure models. *Journal of Finance*, 55(5):1943–1978, 2000.

Q. Dai and K.J. Singleton. Expectation puzzles, time-varying risk premia, and affine models of the term structure. *Journal of Financial Economics*, 63:415–441, 2002.

G. R. Duffee. Term premia and interest rate forecasts in affine models. *Journal of Finance*, 57(1):405–443, 2002.

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

O. Gjolberg and T. Johnson. Risk management in the oil industry: can information on long-run equilibrium prices be utilized? *Energy Economics*, 21:517–527, 1999.

E. J. Hannan, R. D. Terrell, and N. E. Tuckwell. The seasonal adjustment of economic time series. *International Economic Review*, 11(1):24–52, 1970.

J. M. Harrison and S. R. Pliska. Martingales and stochastic integrals in the theory of continuous trading. *Stochastic Process. Appl.*, 11(3):215–260, 1981.

A. C. Harvey. *Forecasting, structural time series models and the Kalman filter*. Cambridge University Press, 1989.

S. Johansen. Estimation and hypothesis testing of cointegration vectors in Gaussian vector autoregressive models. *Econometrica*, 59(6):1551–1580, 1991.

O. Korn. Drift matters: An analysis of commodity derivatives. *Journal of Futures Markets*, 25(3):211–241, 2005.

M. Manoliu and S. Tompaidis. Energy futures prices: term structure models with Kalman filter estimation. *Applied Mathematical Finance*, 9:21–43, 2002.

K. R. Miltersen and E. S. Schwartz. Pricing of options on commodity futures with stochastic term structures of convenience yields and interest rates. *Journal of Financial and Quantitative Analysis*, pages 77–105, 1998.

R.S. Pindyck and J.J. Rotemberg. The excess co-movement of commodity prices. *The Economic Journal*, 100:1173–1189, 1990.

S. A. Ross. Hedging long run commitments: Exercises in incomplete market pricing. *Banca Monte Economic Notes*, 26:99–132, 1997.

B R. Routledge, D J. Seppi, and C S. Spatt. Equilibrium forward curves for commodities. *Journal of Finance*, 55(3):1297–1338, 2000.

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

E. S. Schwartz. The stochastic behavior of commodity prices: Implication for valuation and hedging. *Journal of Finance*, 52(3):923–973, 1997.

E. S. Schwartz and J. E. Smith. Short-term variations and long-term dynamics in commodity prices. *Management Science*, 46(7):893–911, 2000.

C. Sorensen. Modeling seasonality in agricultural commodity futures. *Journal of Futures Markets*, 22(5):393–426, 2002.