# Drift Matters: An Analysis of Commodity Derivatives

**OLAF KORN**

This article presents a reduced-form, two-factor model to price commodity derivatives, which generalizes the model by Schwartz and Smith (2000). The model allows for two mean-reverting stochastic factors and therefore implies that spot and futures prices can be stationary. An empirical study for the crude oil market tests the new model. Out-of-sample pricing and hedging results for futures and forwards show that the new model dominates the nonstationary model by Schwartz and Smith in the following sense: It works equally well for short-term contracts but leads to major improvements for long-term contracts. This finding is particularly relevant for typical applications like the valuation of commodity-linked real assets with long maturities. © 2005 Wiley Periodicals, Inc. Jrl Fut Mark 25:211–241, 2005

I am especially grateful to B. Philipp Kellerhals, Abraham Lioui, and an anonymous referee for their valuable comments and suggestions, and to Wolfgang Bühler for many helpful discussions. I also thank Jens Göbel, Stinson Gibner, Vince Kaminski, and John Nolan for providing the forward data.
For correspondence, University of Mannheim, D-68131 Mannheim, Germany; e-mail: o.korn@uni-mannheim.de

*Received October 2003; Accepted May 2004*

■ *Olaf Korn is an Assistant Professor of Finance at the University of Mannheim in Mannheim, Germany.*

The Journal of Futures Markets, Vol. 25, No. 3, 211–241 (2005)

Published online in Wiley InterScience (www.interscience.wiley.com). DOI: 10.1002/fut.20139

## INTRODUCTION

In recent years the trading volume of many commodity contingent claims has grown rapidly, new derivatives products have been developed, and options concepts have gained acceptance as important tools to value commodity-linked real assets. Therefore, it is not surprising that the issues of valuation and hedging of commodity contingent claims have received much attention from practitioners and researchers.

Reduced-form Gaussian models from the affine class are undoubtedly the workhorse valuation models for commodity derivatives.[1] The major advantage of these models is that closed-form solutions for futures and European-style options are available. Moreover, affine Gaussian models lend themselves to maximum likelihood estimation of the unknown model parameters by means of the Kalman filter algorithm. Many specific models have been presented in the literature, following basically two approaches. The first approach uses the notion of a convenience yield (e.g., Brennan, 1991; Brennan & Schwartz, 1985; Casassus & Collin-Dufresne, 2002; Gibson & Schwartz, 1990; Hilliard & Reiss, 1998; Schwartz, 1997, Model 3). In the second one, the commodity spot price or its logarithm is specified directly as the sum of stochastic factors (e.g., Cortazar & Naranjo, 2003; Ross, 1997; Schwartz, 1997 Model 1; Schwartz & Smith, 2000; Sørensen, 2002). However, a model presented in terms of a certain factor representation can in general be equivalently presented in terms of another factor representation (see e.g., Schwartz & Smith, 2000, p. 899 ff.).

A first important aspect of the concrete model choice is how many stochastic factors should be used. Empirical studies by Brennan (1991) and Schwartz (1997) suggest that a single-factor model is often too restrictive to explain observed derivatives prices and their dynamics. However, multi-factor models with three or more factors (e.g., Casassus & Collin-Dufresne, 2002; Cortazar & Naranjo, 2003; Schwartz, 1997, Model 3) have other disadvantages. Because of the increased number of parameters, estimation becomes less accurate, parameters are more likely to be unstable over time, and there is generally a higher model risk. Moreover, many model applications that involve the valuation of real options (see, e.g., Brennan & Schwartz, 1985) can only be solved numerically. For such problems, models with more than two factors usually cannot be applied because of restrictions in the numerical solutions. Thus, in many cases a two-factor model is a good middle way.

[1]Relatively few papers, like Richard and Sundaresan (1981), Routledge, Seppi, and Spatt (2000), and Bühler, Korn, and Schöbel (in press), price commodity derivatives in equilibrium.

A second important aspect of the concrete model choice is the drift specification of the stochastic factors. This second aspect is the main focus of this article, which extends the two-factor model by Schwartz and Smith (2000) to a model with a more general drift specification. In contrast to existing two-factor models that contain at least one nonstationary factor and lead to nonstationary spot and futures prices, the new model variant allows for two mean-reverting factors. Therefore, futures prices can be stationary. It is shown that the seemingly minor change in the drift of the second factor, which leads to only one extra model parameter, has major effects on the pricing and hedging of long-term derivatives contracts.

The new model is tested in an extensive empirical study and compared to a stationary one-factor model and a nonstationary two-factor model. Three aspects of out-of-sample model performance are considered: (1) pricing, (2) short-horizon risk measurement and hedging, and (3) the ability to create synthetic long-maturity contracts from existing short-maturity contracts. The empirical study uses two data sets of crude oil futures and forward contracts that supplement each other. The futures data set (New York Mercantile Exchange crude oil futures prices) covers a relatively long historical time period, which improves the statistical power of the analysis. The forward data set (over-the-counter [OTC] forward prices) covers a wide range of maturities, which allows for an analysis of maturity effects. The empirical results support a clear recommendation for the stationary two-factor model. It dominates the nonstationary model in the sense that it performs equally well for short-term contracts but leads to major improvements for long-term contracts.

## MODELS OF THE SPOT PRICE DYNAMICS

The analysis begins with an introduction of different specifications of the exogenous commodity price process. It is assumed that prices consist of several, potentially unobservable components. Following Schwartz and Smith (2000), a first component can be interpreted as the long-term or equilibrium price level, determined by long-term supply and demand conditions. Deviations from this equilibrium level as a result of temporal shocks on the supply or demand side build a second price component. Finally, a deterministic time trend or seasonal patterns can be superimposed if appropriate.[2] However, note that the reduced-form valuation

[2]A seasonal component is the main focus of Sørensen's (2002) paper and can be added to the models presented here. However, the crude oil futures and forward prices analyzed in the empirical part of this paper do not show seasonal patterns.

models presented here also allow for other interpretations of the stochastic factors. For example, Schwartz and Smith (2000) have shown that their "short-term–long-term" model is equivalent to the stochastic convenience yield model by Gibson and Schwartz (1990).

The zero mean Ornstein-Uhlenbeck (O-U) process $\chi_t$ of Equation (1) below describes the transitory deviations from the long-term level, where $\kappa > 0$ is the mean-reversion parameter, $\sigma_\chi$ the volatility parameter, and $z_\chi$ a standard Wiener process:

$$d\chi_t = -\kappa \chi_t \, dt + \sigma_\chi \, dz_\chi \tag{1}$$

The transitory component given in Equation (1) is a common part of all models of the spot price dynamics.

The specification of the long-term or equilibrium price level distinguishes different models. In the simplest case, the long-term price level is an intertemporal constant, $\Theta$. A combination of this long-term level with the short-term deviations of Equation (1) leads to the following definition of the log spot price,[3] $\ln S_t$:

$$\ln S_t \equiv \chi_t + \Theta \tag{2}$$

The specification in Equations (1) and (2) is equivalent to the Ornstein-Uhlenbeck process of Equation (3), with stationary mean $\Theta$, $\sigma_S = \sigma_\chi$, and $dz_S = dz_\chi$, the model proposed by Schwartz (1997) for the log spot price dynamics.

$$d \ln S_t = \kappa(\Theta - \ln S_t) \, dt + \sigma_S \, dz_S \tag{3}$$

Schwartz and Smith (2000) suggest an extension of this model. These authors assume a stochastic long-term level $\xi_t$, which follows the Brownian motion process (4), with drift rate $\mu$, volatility $\sigma_\xi$, and a standard Wiener process $z_\xi$, such that $dz_\chi \, dz_\xi = \rho_{\chi\xi} \, dt$.

$$d\xi_t = \mu \, dt + \sigma_\xi \, dz_\xi \tag{4}$$

The following definition of the log spot price in Equation (5),

$$\ln S_t \equiv \chi_t + \xi_t - \frac{\mu}{\kappa} \tag{5}$$

results in the stochastic differential Equation (6) for the price dynamics:

$$d \ln S_t = \kappa(\xi_t - \ln S_t) \, dt + \sigma_S \, dz_S \tag{6}$$

[3]Modeling the log spot price instead of the spot price guarantees positive prices.

with $\sigma_S = \sqrt{\sigma_\chi^2 + \sigma_\xi^2 + 2\rho_{\chi\xi}\sigma_\chi\sigma_\xi}$ and $dz_S = (\sigma_\chi dz_\chi + \sigma_\xi dz_\xi)/\sigma_S$. Equation (6) highlights the interpretation of $\xi_t$ as a long-term mean level.

Because of the introduction of a second stochastic factor, Equation (6) offers a greater flexibility in the specification of the commodity price dynamics compared to Equation (3). But there is another crucial difference between the two processes. The price process (3) is stationary (mean-reverting), but the stochastic, nonstationary mean level of Equation (4) renders the price process (6) nonstationary. To allow for a (potentially) stationary two-factor model, the analysis further extends model (6).

It is now assumed that the long-term price level follows the O-U process of Equation (7), with stationary mean $\Theta$, mean-reversion parameter $\gamma$, volatility $\sigma_\xi$, and $dz_\chi\, dz_\xi = \rho_{\chi\xi}\, dt$.

$$d\xi_t = \gamma(\Theta - \xi_t)\, dt + \sigma_\xi\, dz_\xi \tag{7}$$

With the alternative definition of $\ln S_t$ in Equation (8),

$$\ln S_t \equiv \chi_t + \frac{\kappa}{\kappa - \gamma}\xi_t - \frac{\gamma\Theta}{\kappa - \gamma} \tag{8}$$

the log spot price process has the same form as in Equation (6), but with $\xi_t$ from Equation (7) instead of Equation (4). Moreover, $\sigma_S = \sqrt{\sigma_\chi^2 + \frac{\kappa^2}{(\kappa-\gamma)^2}\sigma_\xi^2 + 2\rho_{\chi\xi}\frac{\kappa}{(\kappa-\gamma)}\sigma_\chi\sigma_\xi}$ and $dz_S \equiv (\sigma_\chi\, dz_\chi + \frac{\kappa}{\kappa-\gamma}\sigma_\xi\, dz_\xi)/\sigma_S$. Equations (7) and (8) specify the most general of the three models of the commodity price dynamics. The first model is a special case with $\xi_t \equiv \Theta$, and the second one results in the limit for $\gamma \to 0$ and $\gamma\Theta \to \mu$.

## PRICING FUTURES CONTRACTS

Because $\chi_t$ and $\xi_t$ are unobservable state variables and not prices of traded assets, one can not uniquely deduce the drift rate of the commodity price process under the risk-neutral measure from no-arbitrage arguments. Instead, the analysis introduces market prices of risk $\lambda_\chi$ and $\lambda_\xi$, referring to the state variables $\chi_t$ and $\xi_t$, as additional parameters. For simplicity, these market prices of risk are assumed to be constant over time. With constant market prices of risk, under the most general specification of the commodity price dynamics of the previous section the risk-neutral processes of the state variables become

$$d\chi_t = -\kappa\left(\chi_t + \frac{\lambda_\chi}{\kappa}\right)dt + \sigma_\chi\, dz_\chi^* \tag{9}$$

$$d\xi_t = -\gamma\left(\xi_t - \Theta + \frac{\lambda_\xi}{\gamma}\right)dt + \sigma_\xi\, dz_\xi^* \tag{10}$$

Given the dynamics of the state variables under the risk-neutral measure, risk-neutral valuation leads to the prices of derivatives. Duffie and Stanton (1992) show that under the risk-neutral measure the futures price equals the expected spot price at maturity of the contract. In the Appendix, this expectation is calculated under the dynamics in Equations (9) and (10), and the definition of the spot price in Equation (8). The resulting futures price $F(0, T)$ at time zero of a contract maturing at time $T$ equals

$$F(0, T) = \exp\left\{e^{-\kappa T}\chi_0 + \frac{\kappa}{\kappa - \gamma}e^{-\gamma T}\xi_0 + A(T)\right\} \tag{11}$$

with

$$A(T) \equiv \left[-(1 - e^{-\kappa T})\frac{\lambda_\chi}{\kappa} + \frac{\kappa}{\kappa - \gamma}\left[(1 - e^{-\gamma T})\left(\Theta - \frac{\lambda_\xi}{\gamma}\right)\right] - \frac{\gamma\Theta}{\kappa - \gamma} + (1 - e^{-2\kappa T})\frac{\sigma_\chi^2}{4\kappa} \right.$$
$$\left. + \frac{\kappa^2}{(\kappa - \gamma)^2}(1 - e^{-2\gamma T})\frac{\sigma_\xi^2}{4\gamma} + \frac{\kappa}{(\kappa - \gamma)}(1 - e^{-(\kappa + \gamma)T})\,\frac{\rho_{\chi\xi}\sigma_\chi\sigma_\xi}{(\kappa + \gamma)}\right]$$

Futures prices according to the Schwartz (1997) and Schwartz and Smith (2000) models result from Equation (11) when taking the appropriate limits of the relevant parameters. Notably, if the long-term price level follows a mean-reverting process ($\gamma > 0$) and $\kappa > \gamma$, the sensitivity of the futures price for changes in $\xi_0$ decreases with the time to maturity of the contract. However, if the process of the long-term level is not mean-reverting ($\gamma = 0$), the sensitivity is the same for all maturities. Thus, Equation (11) suggests that the stationarity or nonstationarity of the spot price process is particularly important for the pricing and hedging of contracts with a long time to maturity.

Closed-form solutions for European options on futures can also be derived for the presented models (details are available from the author upon request). Because the stationarity or nonstationarity of the spot price process affects the distribution of the futures price on the option's maturity date—in particular, the volatility of the futures price—different drift specifications lead to different option prices.

## HEDGING LONG-TERM COMMODITY PRICE EXPOSURE

An interesting example for the importance of the drift specification is the problem of hedging long-term commodity price exposure. For many commodities, traded futures and options contracts extend only a couple of years. In such cases, one way to hedge long-term commodity price exposure is to devise dynamic strategies based on short-maturity futures, which synthetically create long-maturity futures, forwards, or options. The question of how to design such strategies has recently received much attention.[4]

In theory, if at least two short-term futures contracts with different maturities are available and priced according to one of the model variants presented here, there is a dynamic strategy that delivers the pay-off of a model-consistent long-term derivative. For concreteness, assume that one wants to create a synthetic long-position of one futures contract $F_3$ with expiration date $T_3$, or equivalently hedge a short-position of one such contract. Further assume that two traded futures contracts $F_1$ and $F_2$ are always available, with times to maturity $\tau_1$ and $\tau_2$, respectively. All contracts refer to one unit of the commodity. The replication strategy is characterized by the hedge positions $h_{1t}$ and $h_{2t}$, $t \in [0, T_3]$, that is, the number of contracts held in the two short-term futures. These hedge positions are the solutions to the following system of equations:

$$\begin{aligned} h_{1t}\frac{\partial F_1(t, \tau_1 + t)}{\partial \chi} + h_{2t}\frac{\partial F_2(t, \tau_2 + t)}{\partial \chi} &= \frac{\partial F_3(t, T_3)}{\partial \chi}, \\ h_{1t}\frac{\partial F_1(t, \tau_1 + t)}{\partial \xi} + h_{2t}\frac{\partial F_2(t, \tau_2 + t)}{\partial \xi} &= \frac{\partial F_3(t, T_3)}{\partial \xi} \end{aligned} \tag{12}$$

For futures prices according to the most general drift specification given in Equation (11), System (12) has the following solution:

$$\begin{aligned} h_{1t} &= e^{-\gamma(T_3 - t - \tau_1)}\left(1 - \frac{1 - e^{(\gamma-\kappa)(T_3 - t - \tau_1)}}{1 - e^{(\gamma-\kappa)(\tau_2 - \tau_1)}}\right)\frac{F_3}{F_1}, \\ h_{2t} &= e^{-\gamma(T_3 - t - \tau_2)}\left(\frac{1 - e^{(\gamma-\kappa)(T_3 - t - \tau_1)}}{1 - e^{(\gamma-\kappa)(\tau_2 - \tau_1)}}\right)\frac{F_3}{F_2} \end{aligned} \tag{13}$$

Equations (14) give the hedge positions for the Schwartz and Smith (2000) special-case model:

$$h_{1t} = \left(1 - \frac{1 - e^{-\kappa(T_3 - t - \tau_1)}}{1 - e^{-\kappa(\tau_2 - \tau_1)}}\right)\frac{F_3}{F_1}, \qquad h_{2t} = \left(\frac{1 - e^{-\kappa(T_3 - t - \tau_1)}}{1 - e^{-\kappa(\tau_2 - \tau_1)}}\right)\frac{F_3}{F_2} \tag{14}$$

[4]This attention stems mainly from the problems of the German firm Metallgesellschaft in hedging their long-term delivery commitments in oil. Culp and Miller (1999) provide a collection of important contributions to the controversy surrounding the case and the related hedging strategy.

Equations (15) give the hedge positions for the Schwartz (1997) one-factor model:

$$h_{1t} = 0, \qquad h_{2t} = e^{-\kappa(T_3 - t - \tau_2)} \frac{F_3}{F_2} \tag{15}$$

For the one-factor model, one hedging instrument suffices; that is, there can always be a zero position in one of the short-term contracts. Because of the linear relation between the log futures prices and the risk factors, hedge positions are functions of the futures prices and the parameters $\gamma$ and $\kappa$ only.

Figure 1 gives a visual impression of the replication strategies that lead to a synthetic 10-year futures contract. The figure shows how the hedge positions $h_{1t}$ and $h_{2t}$ that result from the three model variants with different drift specifications vary with the time to maturity $T_3 - t$, $t \in [0, T_3]$, of the long-term futures. Short-term futures have times to maturity of three months ($\tau_1$) and 7 months ($\tau_2$). The parameter values $\gamma$ and $\kappa$ are estimates from oil futures data, as given in Table III, and a flat futures curve is used.

Figure 1 shows that a negligible hedge position follows the one-factor model (Model 1) for long maturities $T_3 - t$. Only when $T_3 - t$ is less than two years, substantial positions of more than 0.5 contracts appear.

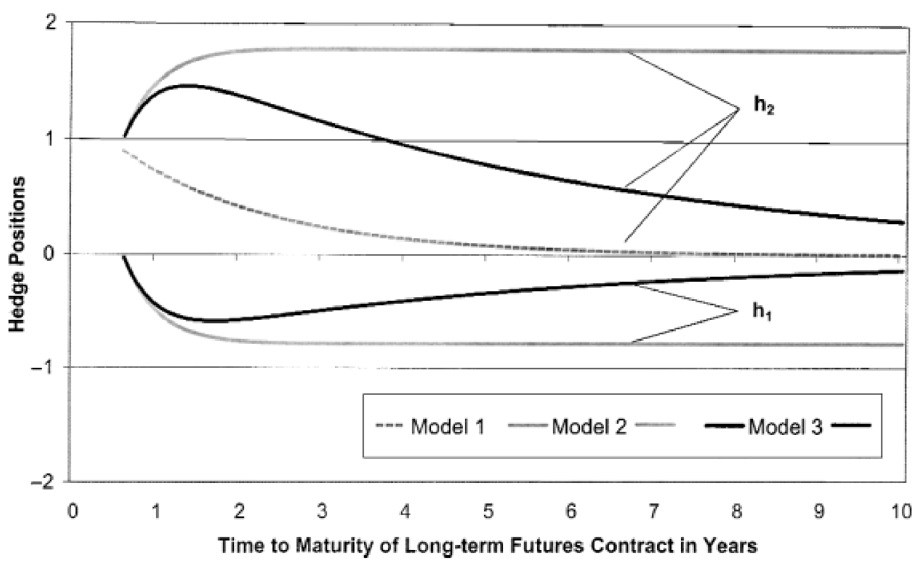


**FIGURE 1**
Hedge positions resulting from different model variants (drift specifications). The figure shows the numbers of short-term contracts needed to hedge one long-term contract.

In contrast, the Schwartz and Smith (2000) model (Model 2) leads to larger absolute hedge positions. The net position $h_{1t} + h_{2t}$, which is the natural quantity to compare with $h_{2t}$ of the one-factor model, equals one for all $T_3 - t$, $t \in [0, T_3]$. Thus, on a net basis, Model 2 leads to a one-to-one hedging strategy.

Model 3, the more general two-factor model with a stationary long-term level, shares with Model 2 the feature that long positions in the seven-months futures and short positions in the three-months futures always appear. However, the sizes of hedge positions derived from the two models are very different. Model 3 has a net hedge position $h_{1t} + h_{2t}$ closer to $h_{2t}$ of the one-factor model, starting from a very low value of about 0.15 for long maturities and gradually increasing to one as $T_3 - t$ decreases.

According to Figure 1, the stationarity or nonstationarity of the long-term price level has a strong impact on the (net) hedge positions. The intuition behind this finding is as follows: If $\xi_t$ follows a Brownian motion (Model 2), then current changes have a permanent effect on the level of the process and thus influence futures prices for all maturities in the same way. As Equation (11) shows, when the futures curve is flat and $\gamma \to 0$ the sensitivity of the futures price for changes in $\xi_t$ does not depend on the time to maturity of the futures. When $\xi_t$ follows an O-U process (Model 3), current changes have only a temporary effect on the spot price level. Prices of futures with short maturities respond more strongly to changes in $\xi_t$ than do futures with long maturities. Thus, a hedging strategy based on a stationary model needs fewer short-term contracts to replicate a long-term contract.

One can create synthetic long-term options in a way similar to that used for long-term futures. The corresponding replication strategies for European-style options are easily derived from Equations (13) to (15) and the closed-form solutions for the option prices. The replication strategies crucially depend on the options' deltas. Because the deltas are functions of the volatility, which itself depends on the drift specification, the drift specification enters via yet another channel if options on long-term futures are replicated.

## EMPIRICAL STUDY

Depending on the kind of application, the empirical performance of a valuation model has several important aspects. First, an investor requests a price quote for a new derivative product, for example. This price quote could be obtained by fitting a model to the prices of commodity derivatives that depend on the same risk factors as the new product and that already trade in the market. In such an application, a model should lead to a high pricing accuracy.

Second, to measure and manage the risk of a trading portfolio of commodity derivatives on a short-term basis, a valuation model could be used to calculate the portfolio's sensitivities with respect to the relevant risk factors. Here, the model should have good short-term hedging properties.

Third, one could derive from a valuation model a dynamic trading strategy leading to the pay-off of a synthetic long-term derivative contract. For this purpose, a model with good long-term hedging properties is needed.

The empirical study compares the performance of different factor specifications (Models 1, 2, and 3) for pricing accuracy, and short- and long-term hedging quality.

## Preliminary Data Analysis

The empirical study uses two data sets that complement each other. The first data set contains weekly prices of New York Mercantile Exchange (Nymex) crude oil futures contracts over the period from July 2, 1986, to December 31, 2002. Prices are settlement prices on the last trading day of a week. The holder of a short position of one Nymex crude oil futures contract is committed to deliver 1,000 barrels of West Texas Intermediate (WTI) crude in Cushing, Oklahoma. Up until 1989, the longest maturity contracts were for 12 months. Currently, contracts with initial maturity of up to 84 months are traded. However, liquidity is very low for the longer-term contracts. Early in the data period, even prices for contracts with maturities of more than nine months were not always available. The analysis concentrates on those maturities for which prices exist over the whole 1986 to 2002 period.

The second data set consists of OTC forward prices for the period from January 22, 1993, to August 30, 1996, and was made available by Enron. Again, prices refer to the last trading day of a week. The main advantage of the forward data set lies in its complete coverage of times to maturity from one month to nine years. The main disadvantage compared to the futures data set is the relatively short time period covered.

A preliminary data analysis indicates the appropriateness of certain factor specifications. Table I shows the means and standard deviations of the time series of futures and forward prices for different maturities. On average, the futures curve is downward sloping, that is, the Nymex

**TABLE I**
Means and Standard Deviations of Futures and Forward Prices

| | *Futures data* | | *Forward data* | |
|---|---|---|---|---|
| *Contract maturity* | *Mean* | *Standard deviation* | *Mean* | *Standard deviation* |
| 1 month | 20.68 | 4.90 | 18.61 | 2.01 |
| 2 months | 20.52 | 4.68 | 18.38 | 1.71 |
| 3 months | 20.36 | 4.47 | 18.23 | 1.53 |
| 4 months | 20.22 | 4.26 | 18.14 | 1.42 |
| 5 months | 20.09 | 4.08 | 18.09 | 1.34 |
| 6 months | 19.97 | 3.91 | 18.06 | 1.29 |
| 7 months | 19.86 | 3.75 | 18.04 | 1.25 |
| 8 months | 19.76 | 3.61 | 18.02 | 1.21 |
| 9 months | 19.68 | 3.48 | 18.02 | 1.18 |
| 10 months | — | — | 18.02 | 1.16 |
| 12 months | — | — | 18.05 | 1.12 |
| 18 months | — | — | 18.15 | 1.07 |
| 24 months | — | — | 18.32 | 1.02 |
| 36 months | — | — | 18.73 | 0.95 |
| 48 months | — | — | 19.15 | 0.92 |
| 60 months | — | — | 19.56 | 0.90 |
| 72 months | — | — | 19.91 | 0.88 |
| 84 months | — | — | 20.21 | 0.85 |
| 96 months | — | — | 20.51 | 0.81 |
| 108 months | — | — | 20.81 | 0.78 |

*Note.* Weekly data are for the period 7/2/1986 to 12/20/2002 (futures, 861 observations) and the period 1/22/1993 to 8/30/1996 (forwards, 186 observations).

futures market is in backwardation. The forward data confirm this finding for maturities of up to 10 months. However, for longer maturities, the forward curve is upward sloping on average. A forward curve that is downward sloping at the short end and upward sloping at the long end is not compatible with Model 1, the one-factor model by Schwartz (1997). The standard deviation of prices generally decreases with the time to maturity. This result confirms the prediction made by all presented valuation models, because each model contains at least one stationary component.[5]

An important difference between Models 1 and 3, on the one hand, and Model 2, on the other hand, is the stationarity or nonstationarity of the log spot price process as well as the derived futures price processes. Hence, it is interesting to see if there is an empirical indication for stationarity.

[5]Absolute numbers should not be compared between the futures and forward data presented in Table I, because the means and standard deviations refer to different time periods.

**TABLE II**
Results of Augmented Dickey-Fuller-Tests and KPSS-Tests for Logarithmic Futures and Forward Prices

| | *Futures data* | | | *Forward data* | | |
|---|---|---|---|---|---|---|
| | *Dickey-Fuller-test* | | *KPSS-test* | *Dickey-Fuller-test* | | *KPSS-test* |
| *Contract maturity* | *Test statistic* | *Lags* | *Test statistic* | *Test statistic* | *Lags* | *Test statistic* |
| 1 month | −3.23* | 4 | 0.350 | −1.67 | 4 | 0.259 |
| 2 months | −3.28* | 4 | 0.352 | −1.93 | 4 | 0.208 |
| 3 months | −3.24* | 4 | 0.354 | −1.95 | 4 | 0.168 |
| 4 months | −3.19* | 4 | 0.355 | −1.97 | 4 | 0.153 |
| 5 months | −3.16* | 4 | 0.357 | −2.00 | 4 | 0.160 |
| 6 months | −2.87* | 0 | 0.359 | −2.04 | 4 | 0.178 |
| 7 months | −2.89* | 0 | 0.360 | −2.08 | 4 | 0.203 |
| 8 months | −2.91* | 0 | 0.360 | −2.11 | 4 | 0.229 |
| 9 months | −2.94* | 0 | 0.361 | −2.13 | 4 | 0.257 |
| 10 months | — | — | — | −2.14 | 4 | 0.283 |
| 12 months | — | — | — | −2.01 | 4 | 0.328 |
| 18 months | — | — | — | −1.93 | 4 | 0.394 |
| 24 months | — | — | — | −2.19 | 7 | 0.416 |
| 36 months | — | — | — | −2.13 | 8 | 0.430 |
| 48 months | — | — | — | −1.88 | 8 | 0.423 |
| 60 months | — | — | — | −1.68 | 8 | 0.400 |
| 72 months | — | — | — | −1.26 | 3 | 0.375 |
| 84 months | — | — | — | −1.31 | 3 | 0.354 |
| 96 months | — | — | — | −1.38 | 3 | 0.333 |
| 108 months | — | — | — | −1.47 | 3 | 0.312 |

*Note.* Weekly data are for the period 7/2/1986 to 12/20/2002 (futures, 861 observations) and the period 1/22/1993 to 8/30/1996 (forwards, 186 observations).
*Significant on a 5%-level.

Table II provides the results of two different unit root tests. The first one is the augmented test by Dickey and Fuller (1979), which tests the null hypothesis of a unit root, the second one is the test by Kwiatkowski, Phillips, Schmidt, and Shin (KPSS, 1992), which states stationarity as the null hypothesis. Because there is no time trend in the data, results show the test variants with a constant term but without a deterministic time trend. The procedure by Hall (1994) is applied to select the number of augmentation terms used in the Dickey-Fuller tests for each series. The KPSS tests are adjusted for autocorrelation by applying the Newey and West (1987) variance estimate, with a number of lags equal to one-third of the sample size.

For the futures data set, the unit root hypothesis is rejected for all maturities on a 5% significance level. This result supports valuation models with stationary price processes.[6] For the forward data neither stationarity nor nonstationarity is rejected on a 5% level.[7] Such a finding might be explained by the relatively short data period of less than four years. As the unit roots literature shows, it is not the number of observations, but the length of the data period that is primarily important to detect mean-reversion (see, e.g., Campbell & Perron, 1991, for an overview and further references).

## Model Estimation

To estimate the unknown parameters of the continuous time models, a maximum likelihood approach is used that combines time series and cross-sectional information.[8] Assume that at times $t = 1, \ldots, m$ the $n$-vectors $y_t \equiv [\ln(F_{T_1}), \ln(F_{T_2}), \ldots, \ln(F_{T_n})]$ of log futures prices with times to maturity $T_1, \ldots, T_n$ are observed. Further denote the length of the sampling interval by $h$, which is one week in the relevant case of weekly data. Then from the conditional distribution of the bivariate random variable $x_t \equiv [\chi_t, \xi_t]$ and the valuation formula (11) the following state space model, (16) and (17), results:

$$x_t = C + Gx_{t-1} + \omega_t, \qquad t = 1, \ldots, m \tag{16}$$

with $C \equiv [0, (1 - e^{-\gamma h})\Theta]$ and $G \equiv \begin{bmatrix} e^{-\kappa h} & 0 \\ 0 & e^{-\gamma h} \end{bmatrix}$.

The error terms $\omega_t$ are bivariate i.i.d. normal random variables with zero expectation and variance-covariance matrix

$$\Omega \equiv \begin{bmatrix} (1 - e^{-2\kappa h})\dfrac{\sigma_\chi^2}{2\kappa} & (1 - e^{-(\kappa+\gamma)h})\dfrac{\rho_{\chi\xi}\sigma_\chi\sigma_\xi}{(\kappa + \gamma)} \\ (1 - e^{-(\kappa+\gamma)h})\dfrac{\rho_{\chi\xi}\sigma_\chi\sigma_\xi}{(\kappa + \gamma)} & (1 - e^{-2\gamma h})\dfrac{\sigma_\gamma^2}{2\gamma} \end{bmatrix}$$

[6]The stationarity of the crude oil spot price is also supported empirically by Pilipovic (1998). Bessembinder, Coughenour, Seguin, and Smoller (1995) find evidence that investors expect crude oil spot prices to be mean reverting.
[7]The null hypothesis of the Dickey-Fuller test is rejected on a 5% significance level for test statistics below: −2.86, the null hypothesis of the KPSS test is rejected on a 5% significance level for test statistics above 0.463.
[8]Schwartz (1997), Schwartz and Smith (2000), Sørensen (2002), Casassus and Collin-Dufresne (2002), and Cortazar and Naranjo (2003) have applied this approach to models of commodity derivatives. Similar applications to term structure models are, for example, Chen and Scott (1993), Pearson and Sun (1994), and Duffie and Singleton (1997). Kellerhals (2001) contains a detailed treatment of the estimation of continuous-time models via Kalman filter techniques and provides many additional references.

Equation (16) is the transition equation of the state space model. The measurement equation is given as:

$$y_t = D_t + H_t x_t + \nu_t, \qquad t = 1,\ldots, m \tag{17}$$

with $n$-vector $D_t \equiv [A(T_1), A(T_2), \ldots, A(T_n)]'$

and $n \times 2$ matrix $H_t \equiv \begin{bmatrix} e^{-\kappa T_1} & e^{-\kappa T_2} & \cdots & e^{-\kappa T_n} \\ \dfrac{\kappa}{\kappa - \gamma} e^{-\gamma T_1} & \dfrac{\kappa}{\kappa - \gamma} e^{-\gamma T_2} & \cdots & \dfrac{\kappa}{\kappa - \gamma} e^{-\gamma T_n} \end{bmatrix}'.$

The $n$-vectors $\nu_t$ can be interpreted as pricing errors. If the model explains futures prices perfectly, all $\nu_t$'s equal zero. The $\nu_t$'s are the only component of the state space model that is not derived from the valuation model. To account for possible serial correlation of the pricing errors, the following AR(1) model is assumed for the $\nu_t$'s:

$$\nu_t = \delta \nu_{t-1} + \varepsilon_t, \qquad t = 1, \ldots, m \tag{18}$$

where the $\varepsilon_t$'s are zero mean i.i.d. normal random vectors and $\delta$ is a scalar parameter. The assumption of a single $\delta$ coefficient, irrespective of the time to maturity of the futures contract, has been made to limit the number of free model parameters. For the same reason, a diagonal variance-covariance matrix $V$ of the $\varepsilon_t$'s has been assumed, whose elements are denoted by $s^2_{T_1}, \ldots, s^2_{T_n}$.

With constant model parameters, it will usually be impossible to fit all $m \times n$ futures prices exactly. For the two-factor models (Models 2 and 3), it is possible to set at most two elements of $V$ equal to zero, that is, chose two contract maturities for which model prices equal observed prices for all $t = 1,\ldots, m$. For the one-factor model (Model 1), at most one zero element of $V$ will generally be possible.

The likelihood function of the above state space model is computed recursively by means of the Kalman filter algorithm, and is maximized with respect to the model parameters. In a second step, estimates of the state variables $\chi_t, \xi_t, t = 1, \ldots, m,$ can be obtained (for details see, e.g., Harvey, 1989; Gourieroux & Monfort, 1997, Ch. 15). Notably, the estimation of Models 1 and 2 proceeds in the same way as the estimation of Model 3, but the matrices $C$, $G$, $\Omega$, $D_t$, and $H_t$ must be modified appropriately. It is also notable that all empirical results obtained for Model 2, that is, concerning parameter estimation, pricing accuracy, and hedging performance, also hold for the convenience yield model by Gibson and

**TABLE III**
Estimation Results for the Futures Data

| | *Model 1 (One-factor-model: stationary)* | | *Model 2 (Two-factor-model: nonstationary)* | | *Model 3 (Two-factor-model: stationary)* | |
|---|---|---|---|---|---|---|
| *Parameter* | *Estimate* | *Standard error* | *Estimate* | *Standard error* | *Estimate* | *Standard error* |
| $\kappa$ | 0.552 | 0.006 | 2.459 | 0.030 | 2.566 | 0.033 |
| $\sigma_\chi$ | 0.311 | 0.004 | 0.280 | 0.005 | 0.270 | 0.005 |
| $\lambda_\chi$ | 0.301 | 0.142 | 0.128 | 0.074 | 0.113 | 0.071 |
| $\gamma$ | — | — | — | — | 0.189 | 0.015 |
| $\Theta$ | 3.114 | 0.140 | — | — | 3.260 | 0.319 |
| $\mu$ | — | — | 0.040 | 0.052 | — | — |
| $\sigma_\xi$ | — | — | 0.200 | 0.003 | 0.217 | 0.003 |
| $\lambda_\xi$ | — | — | 0.087 | 0.051 | 0.095 | 0.054 |
| $\rho_{\chi\xi}$ | — | — | 0.251 | 0.029 | 0.130 | 0.030 |
| $\delta$ | 0.947 | 0.003 | 0.912 | 0.004 | 0.877 | 0.005 |
| $S_1$ | 0.022 | 0.000 | 0.015 | 0.000 | 0.014 | 0.000 |
| $S_3$ | 0.006 | 0.000 | 0 | — | 0 | — |
| $S_5$ | 0 | — | 0.002 | 0.000 | 0.002 | 0.000 |
| $S_7$ | 0.004 | 0.000 | 0 | — | 0 | — |
| $S_9$ | 0.007 | 0.000 | 0.002 | 0.000 | 0.002 | 0.000 |
| AIC | −34,499.54 | | −36,650.48 | | −36,694.36 | |
| SIC | −34,456.72 | | −36,598.16 | | −36,637.28 | |

*Note.* Weekly data are for the period 7/2/1986 to 12/20/2002.

Schwartz (1990). The reason is the formal equivalence of the two models, as shown by Schwartz and Smith (2000).

Table III shows the estimation results for the futures data set. Estimates are obtained from 861 weekly observations of futures with maturities closest to one, three, five, seven, and nine months. Because the five-months futures serve as the hedging instrument for strategies based on Model 1, these contracts should be correctly priced, thus $s_5$ is set equal to zero. The two-factor models impose the restrictions $s_3 = 0$ and $s_7 = 0$, because the three-months and seven-months futures serve as hedging instruments. All parameters are significantly different from zero, except for the market prices of risk $\lambda_\chi$ and $\lambda_\xi$, and the drift parameter $\mu$ in Model 2. In particular the $\gamma$-parameter in Model 3, which provides evidence on whether the second stochastic factor is mean-reverting or not, has a $t$-statistic of 12.58.

The last two rows of Table III show the values of the information criteria AIC (Akaike, 1973) and SIC (Schwarz, 1978), which are useful for a statistical model comparison. Both criteria attain the best (lowest) values for Model 3, followed by Model 2. Model 1 is the worst specification.

**TABLE IV**
Estimation Results for the Forward Data

| Parameter | Model 1 (One-factor-model: stationary) Estimate | Model 1 Standard error | Model 2 (Two-factor-model: nonstationary) Estimate | Model 2 Standard error | Model 3 (Two-factor-model: stationary) Estimate | Model 3 Standard error |
|---|---|---|---|---|---|---|
| $\kappa$ | 0.542 | 0.017 | 2.600 | 0.077 | 3.277 | 0.112 |
| $\sigma_\chi$ | 0.237 | 0.015 | 0.238 | 0.019 | 0.231 | 0.017 |
| $\lambda_\chi$ | 0.036 | 0.287 | 0.203 | 0.145 | 0.277 | 0.143 |
| $\gamma$ | — | — | — | — | 0.161 | 0.010 |
| $\Theta$ | 2.928 | 0.283 | — | — | 2.905 | 0.670 |
| $\mu$ | — | — | −0.011 | 0.107 | — | — |
| $\sigma_\xi$ | — | — | 0.147 | 0.011 | 0.163 | 0.011 |
| $\lambda_\xi$ | — | — | −0.019 | 0.106 | −0.026 | 0.101 |
| $\rho_{\chi\xi}$ | — | — | 0.334 | 0.107 | 0.339 | 0.099 |
| $\delta$ | 0.980 | 0.003 | 0.934 | 0.004 | 0.932 | 0.005 |
| $S_1$ | 0.023 | 0.001 | 0.021 | 0.001 | 0.022 | 0.001 |
| $S_3$ | 0.006 | 0.000 | 0 | — | 0 | — |
| $S_5$ | 0 | — | 0.002 | 0.000 | 0.001 | 0.000 |
| $S_7$ | 0.003 | 0.000 | 0 | — | 0 | — |
| $S_9$ | 0.004 | 0.001 | 0.002 | 0.000 | 0.002 | 0.000 |
| $S_{12}$ | 0.014 | 0.001 | 0.015 | 0.001 | 0.014 | 0.001 |
| $S_{36}$ | 0.013 | 0.002 | 0.015 | 0.003 | 0.013 | 0.002 |
| $S_{60}$ | 0.016 | 0.005 | 0.017 | 0.007 | 0.014 | 0.004 |
| $S_{84}$ | 0.017 | 0.007 | 0.018 | 0.013 | 0.015 | 0.006 |
| $S_{108}$ | 0.017 | 0.006 | 0.018 | 0.011 | 0.015 | 0.005 |
| AIC | −14,619.84 | | −14,620.62 | | −14,955.54 | |
| SIC | −14,574.76 | | −14,569.10 | | −14,900.80 | |

*Note.* Weekly data are for the period 1/22/1993 to 8/30/1996.

Table IV shows the estimation results for the forward data set. Estimates are obtained from 186 weekly observations for 10 different futures contracts. In addition to the contracts with one, three, five, seven, and nine months to maturity, the analysis uses 12, 36, 60, 84, and 108-month contracts. All mean-reversion parameters and variance parameters of the risk factors are statistically significant, and the $\gamma$ parameter in Model 3 has a $t$-statistic of 16.35. A model comparison based on the AIC leads to the same ordering as for the futures data set. According to the SIC, which penalizes large models more heavily than the AIC, a drift specification with two stationary factors (Model 3) is again the best (lowest value). However the ordering of the other two models has changed. The one-factor drift specification (Model 1) is second best, followed by the nonstationary two-factor model (Model 2).

## Results: Pricing Accuracy

The first important aspect of a model comparison based on economic criteria is pricing derivative products conditional on price information contained in other derivatives. Therefore, estimation does not use all futures and forward contracts but keeps certain contracts for model evaluation using cross-sectional predictions. These latter contracts refer to maturities of two, four, six, and eight months for the futures data set.

Cross-sectional predictions are obtained from the different valuation models. To avoid the usage of in-sample information, a rolling data window is used to estimate model parameters. The length of the data window is 60 weeks for the futures data set. Because futures contracts with five different times to maturity (one, three, five, seven, and nine months) are available for estimation, the data window contains 300 data points.

The first set of parameter values is estimated from the data of the first 60 weeks. Using the estimated model parameters and estimates of the state variables $\chi_t$ and $\xi_t$, theoretical futures prices for maturities of two, four, six, and eight months are calculated for each of the following 10 weeks. Subtraction of the observed market prices from the theoretical prices delivers 10 pricing errors for each of the four maturities and each of the three model variants. In the next step, parameters are estimated using the data from week 11 to 70. The resulting parameter values are then used in the following 10 weeks to calculate price predictions and pricing errors. The whole procedure is repeated until the end of the data period. As a final result, a time series of 801 pricing errors for each of the four maturities and each of the three drift specifications is obtained.

Table V shows the mean absolute pricing errors and the mean pricing errors per barrel of oil for the different model variants and contract maturities. Model 1 leads to the largest mean absolute errors for all maturities, which reach up to \$0.42. Compared to the two-factor specifications, pricing errors are significantly larger both economically and statistically for the two-months contracts and the eight-months contracts. This result is highlighted in the last part of the table, which gives the mean differences of the absolute errors together with their standard deviations.[9] The one-factor model does not seem to be flexible enough to price futures of all maturities equally well as the two-factor models.

[9]Because the time series of absolute errors show positive autocorrelation, the analysis uses standard deviations according to the Newey and West (1987) estimator with a lag length equal to one third of the sample size.

**TABLE V**
Results of Pricing Accuracy for Futures Data (801 observations)

| | *Model 1* | *Model 2* | *Model 3* |
|---|---|---|---|
| *Contract maturity* | *Mean absolute error (in $)* | | |
| 2 months | 0.420 | 0.203 | 0.203 |
| 4 months | 0.105 | 0.097 | 0.097 |
| 6 months | 0.085 | 0.076 | 0.076 |
| 8 months | 0.216 | 0.096 | 0.096 |
| *Contract maturity* | *Mean error (in $)* | | |
| 2 months | 0.012 | 0.069 | 0.070 |
| 4 months | 0.000 | −0.043 | −0.043 |
| 6 months | 0.003 | −0.038 | −0.038 |
| 8 months | 0.021 | 0.052 | 0.052 |
| | *Mean difference of absolute errors (in $) (Standard deviation in parenthesis)* | | |
| *Contract maturity* | *Model 1 – Model 2* | *Model 2 – Model 3* | *Model 1 – Model 3* |
| 2 months | 0.216** | 0.000 | 0.217** |
| | (0.080) | (0.001) | (0.080) |
| 4 months | 0.008 | 0.000 | 0.009 |
| | (0.039) | (0.000) | (0.039) |
| 6 months | 0.009 | 0.000 | 0.009 |
| | (0.030) | (0.001) | (0.030) |
| 8 months | 0.120** | 0.000 | 0.120** |
| | (0.029) | (0.001) | (0.029) |

*Significantly different from zero: 5%-level.
**Significantly different from zero: 1%-level.

Mean absolute errors for Models 2 and 3 are almost identical. It seems to make no difference which of the models is used to price the missing contracts.

The mean errors show that the relatively bad performance of Model 1 (high mean absolute errors) is not because of a prediction bias but to a high variance of the prediction errors. The prediction bias is even lower in absolute terms than for Models 2 and 3.

Table VI shows the corresponding pricing results for the forward data set. The rolling window that is employed for parameter estimation has a length of 30 weeks here. Because futures contracts with 10 different maturities are used in the cross section (1, 3, 5, 7, 12, 36, 60, 84, and 108 months), each estimation round is based on 300 data points, the same number as for futures. For the forward data, the analysis makes

**TABLE VI**
Results of Pricing Accuracy for Forward Data (156 observations)

| | *Model 1* | *Model 2* | *Model 3* |
|---|---|---|---|
| *Contract maturity* | | *Mean absolute error (in \$)* | |
| 2 months | 0.451 | 0.090 | 0.086 |
| 4 months | 0.106 | 0.032 | 0.028 |
| 6 months | 0.091 | 0.019 | 0.018 |
| 8 months | 0.255 | 0.022 | 0.022 |
| 10 months | 0.388 | 0.070 | 0.073 |
| 18 months | 0.682 | 0.250 | 0.285 |
| 24 months | 0.798 | 0.304 | 0.338 |
| 48 months | 1.092 | 0.503 | 0.426 |
| 72 months | 1.524 | 0.676 | 0.431 |
| 96 months | 1.969 | 0.866 | 0.493 |
| *Contract maturity* | | *Mean error (in \$)* | |
| 2 months | 0.409 | 0.051 | 0.025 |
| 4 months | 0.082 | −0.021 | −0.011 |
| 6 months | −0.052 | −0.008 | −0.000 |
| 8 months | −0.123 | 0.004 | −0.007 |
| 10 months | −0.165 | 0.005 | −0.030 |
| 18 months | −0.161 | −0.000 | −0.132 |
| 24 months | −0.071 | 0.001 | −0.185 |
| 48 months | 0.573 | 0.057 | −0.152 |
| 72 months | 1.291 | −0.048 | 0.008 |
| 96 months | 1.889 | −0.423 | 0.133 |

| | | *Mean difference of absolute errors (in \$) (Standard deviation in parenthesis)* | |
|---|---|---|---|
| *Contract maturity* | *Model 1 – Model 2* | *Model 2 – Model 3* | *Model 1 – Model 3* |
| 2 months | 0.361** | 0.004 | 0.365** |
| | (0.064) | (0.006) | (0.061) |
| 4 months | 0.075** | 0.004 | 0.079** |
| | (0.018) | (0.002) | (0.016) |
| 6 months | 0.072** | 0.001 | 0.073** |
| | (0.012) | (0.002) | (0.011) |
| 8 months | 0.233** | 0.000 | 0.234** |
| | (0.037) | (0.002) | (0.036) |
| 10 months | 0.318** | −0.004 | 0.314** |
| | (0.054) | (0.007) | (0.050) |
| 18 months | 0.423** | −0.035 | 0.396** |
| | (0.104) | (0.029) | (0.084) |
| 24 months | 0.495** | −0.035 | 0.460** |
| | (0.156) | (0.043) | (0.124) |
| 48 months | 0.589 | 0.077 | 0.666* |
| | (0.372) | (0.086) | (0.305) |
| 72 months | 0.848 | 0.246 | 1.094* |
| | (0.559) | (0.131) | (0.475) |
| 96 months | 1.103 | 0.373 | 1.476* |
| | (0.797) | (0.262) | (0.591) |

*Significant on a 5%-level.
**Significant on a 1%-level.

cross-sectional predictions for contracts with 2, 4, 6, 8, 10, 18, 24, 48, 72, and 96 months to maturity.

The forward data confirm the results of Table V for short- and medium-term maturities of up to 24 months. The one-factor model always leads to the largest mean absolute errors, whereas Model 2 and 3 are very similar in their higher pricing accuracy. For longer-maturity forwards, mean absolute errors increase considerably for all models. The strongest effect is in Model 1, followed by Model 2. A comparison between Model 2 and Model 3 shows important deviations. For the two longest maturities (72 and 96 months), the differences in mean absolute errors of about $0.25 and $0.37 are clearly economically significant. However, as a result of the high autocorrelation and the relatively small sample size of 156 observations, the differences are not statistically significant on a 5% significance level. For the 72-month contract, there is a significance on the 10% level.

In sum, the results on the pricing accuracy show that a two-factor model should be preferred to a one-factor model. The choice of a stationarity or nonstationarity second factor seems to be irrelevant for short- and medium-term contracts. However, for long-term contracts, a stationary factor seems to bring some improvement.

## Results: Short-Term Delta Hedging

A second important application refers to short-term risk measurement and hedging. Banks and other firms usually calculate risk measures (e.g., value-at-risk) for their commodity positions on a short-term basis, both for internal use and for external capital requirements. It is common practice for "linear" derivatives positions like forwards and futures to use a delta approximation, based on a valuation model. But which model should be used to calculate deltas? The following analysis provides some empirical evidence on the accuracy of delta-based risk measures derived from different models. Seen from another perspective, the analysis shows how well short-term delta hedging works for different models.

Three different futures portfolios are built at the end of each week of the data period. The portfolios are initially delta neural and therefore instantaneously risk free according to the three model variants with alternative drift specifications. These portfolios contain two futures with different maturities for the one-factor model and three futures with different maturities for the two-factor models. Portfolios referring to Model 1 build on Equations (15): A short position of one contract with maturity closest to nine months to deliver one barrel of crude oil, and $h_{2t}$

contracts with maturity closest to five months. Portfolio positions referring to Models 2 and 3 build on Equations (14) and (13), respectively, with one contract short in the nine-months futures, $h_{1t}$ contracts in the three-months futures, and $h_{2t}$ contracts in the seven-months futures. After a holding period of one week, during which no rebalancing of the portfolio or marking-to-market is considered, the new portfolio value is calculated. Because the portfolio has an initial value of zero and is delta neutral according to the respective model, it would be treated as a zero position for risk measurement purposes. Thus it is natural to ask how far the realized change in value of the portfolio (after one week) deviates from a value of zero, that is, to which extent the risk of the portfolio has been misjudged. The deviations of the realized changes in value from zero can also be interpreted as the weekly hedging errors of a model-based delta hedging strategy that hedges a short position in nine-months futures with long positions in futures with shorter times to maturity.

To assess the out-of-sample performance of the different model variants, model parameters entering Equations (13) to (15) are again estimated using a rolling data window. The size of the data window is 60 weeks for the futures data set and 30 weeks for the forward data set. After each estimation round, the resulting parameter values have been used to set up the delta-neural portfolios for the following 10 weeks.

Table VII provides mean absolute hedging errors and mean hedging errors for the futures data set. As for the pricing results, Model 1 clearly shows the highest mean absolute error (more than \$0.11), and the errors of Models 2 and 3 are very close (a difference of less than \$0.001). The differences between the errors of Model 1, on the one hand, and the errors of Models 2 and 3, on the other hand, are statistically significant.

The forward data set allows the investigation of how far the deltas derived from different models capture the comovement between short- and long-term contracts. Thus, different portfolios are constructed with short positions of one futures to deliver one barrel in 9, 24, 48, 72, and 108 months, respectively. Again, Equations (13) to (15) provide the positions in the three-, five-, and seven-months futures, using the appropriate adjustments for the use of forwards instead of futures.

Table VIII provides some evidence on weekly errors of delta hedging the longer-term futures with the shorter-term contracts. For all models, the mean absolute error increases considerably when moving from a nine-months contract to a 24-months contract. However, a further increase of the time to maturity of the long-term contract up to 108 months does not have such a significant effect on the errors. Except for the portfolio that includes the 24-months contract, Model 3 leads to

**TABLE VII**
Results of Short-Term Delta Hedging for Futures Data (800 observations)

| | *Futures data* | | |
|---|---|---|---|
| *Maturity of long-term contract* | *Model 1* | *Model 2* | *Model 3* |
| | *Mean absolute error (in \$)* | | |
| 9 months | 0.114 | 0.047 | 0.048 |
| | *Mean error (in \$)* | | |
| 9 months | 0.013 | 0.013 | 0.012 |
| | *Mean difference of absolute errors (in \$) (Standard deviation in parenthesis)* | | |
| *Maturity of long-term contract* | *Model 1 – Model 2* | *Model 2 – Model 3* | *Model 1 – Model 3* |
| 9 months | 0.067** (0.010) | −0.000 (0.001) | 0.067** (0.012) |

*Significantly different from zero: 5%-level.
**Significantly different from zero: 1%-level.

the lowest mean absolute errors. Model 2 shows the highest errors. Differences between models are often statistically significant at least at a 5% significance level. This is true in particular for the two-factor models. For maturities of the long-term contract of 24 months and longer, the errors resulting from the stationary model (Model 3) are significantly smaller than the errors resulting form the nonstationary model (Model 2). Even though a difference between mean absolute errors of about three cents per barrel or about \$30 per 1,000 barrel contract does not seem very large, it can well be economically significant. If the portfolio consisted, for example, of 1,000 48-month futures contracts and appropriate positions in the shorter maturity futures, one would still predict the same risk of zero. However, the difference between Models 2 and 3 in the mean absolute realized changes in portfolio value would then be \$300,000.

In sum, we see from the results of this second application that it never harms to use a stationary two-factor model instead of a nonstationary two-factor model. However, it brings some benefits if long-term contracts are contained in the portfolios. In this case, it might even be better to use a stationary one-factor model instead of a nonstationary two-factor model.

**TABLE VIII**

Results of Short-term Delta Hedging for Forward Data (155 observations)

| | *Forward data* | | |
|---|---|---|---|
| *Maturity of long-term contract* | *Model 1* | *Model 2* | *Model 3* |
| | *Mean absolute error (in \$)* | | |
| 9 months | 0.056 | 0.020 | 0.019 |
| 24 months | 0.147 | 0.169 | 0.150 |
| 48 months | 0.169 | 0.185 | 0.154 |
| 72 months | 0.181 | 0.190 | 0.158 |
| 108 months | 0.170 | 0.176 | 0.149 |
| | *Mean error (in \$)* | | |
| 9 months | 0.015 | −0.003 | −0.002 |
| 24 months | 0.025 | −0.008 | 0.003 |
| 48 months | 0.018 | −0.011 | 0.004 |
| 72 months | 0.013 | −0.014 | 0.004 |
| 108 months | 0.009 | −0.015 | 0.003 |
| | *Mean difference of absolute errors (in \$) (Standard deviation in parenthesis)* | | |
| *Maturity of long-term contract* | *Model 1 – Model 2* | *Model 2 – Model 3* | *Model 1 – Model 3* |
| 9 months | 0.036** | 0.001 | 0.037** |
| | (0.006) | (0.001) | (0.005) |
| 24 months | −0.021* | 0.019** | −0.003 |
| | (0.009) | (0.005) | (0.005) |
| 48 months | −0.016 | 0.031** | 0.016** |
| | (0.015) | (0.011) | (0.005) |
| 72 months | −0.009 | 0.033* | 0.023** |
| | (0.014) | (0.013) | (0.003) |
| 108 months | −0.006 | 0.027* | 0.021** |
| | (0.010) | (0.012) | (0.003) |

*Significantly different from zero: 5%-level.
**Significantly different from zero: 1%-level.

## Results: Synthetic Long-Term Forwards

A third application is the creation of synthetic long-term commodity derivatives by using dynamic trading strategies with short-term derivatives and a risk-free asset. This subsection compares the replication errors of such strategies between valuation models with different drift specification.

The evaluation of a strategy to replicate, for example, the pay-off of a 10-year forward contract with short-term futures, would require hundreds of years of historical data if the analysis were to perform a historical simulation such as that undertaken in the previous subsection. As an alternative approach, available Nymex futures prices are used to estimate a multivariate time series model and to simulate 20,000 sample paths from this data model. For each of these sample paths, model-consistent replication strategies are implemented for the three different drift specifications. These strategies lead to 20,000 replication errors for each model variant.[10]

The first equation of the data model (19) refers to the log oil price. Because monthly observations on the last trading day of an expiring contract are used to estimate the data model, the price of a one-month futures (the expiring contract) should be a good proxy for the spot price. Therefore, one-month futures prices are used in the first equation. The second to seventh equation of the data model refer to the two- to seven-months relative futures bases, that is, the difference between the spot price and the futures price divided by the spot price. Lagged values referring to the same contract explain the relative basis $BAS_t^k$ of the $k$-months futures at time $t$. Modeling of the relative basis instead of the (log) futures prices avoids unrealistic deviations between spot prices and futures prices in the simulated price paths which the analysis obtains from the data model. The number of lags is specified separately for each equation of the data model using Schwarz's (1978) information criterion. Equations (19) give the final result of this specification procedure.

$$
\begin{aligned}
\ln S_t &= a_1 + b_1 \ln S_{t-1} + u_{1t} \\
BAS_t^2 &= a_2 + b_2\, BAS_{t-1}^3 + u_{2t} \\
BAS_t^3 &= a_3 + b_3\, BAS_{t-1}^4 + u_{3t} \\
BAS_t^4 &= a_4 + b_4\, BAS_{t-1}^5 + u_{4t} \\
BAS_t^5 &= a_5 + b_5\, BAS_{t-1}^6 + u_{5t} \\
BAS_t^6 &= a_6 + b_6\, BAS_{t-1}^7 + u_{6t}, \\
BAS_t^7 &= a_7 + b_7\, BAS_{t-1}^8 + u_{7t}, \qquad \text{with } t = 2, \ldots, 198
\end{aligned} \tag{19}
$$

In System (19), $a_1, b_1, \ldots, b_7$ denote constant model parameters and $u_{1t}, \ldots, u_{7t}$ are error terms with zero expectation.

[10]The same or similar bootstraps methods have been applied by Ross (1997), Bollen and Whaley (1998), and Bühler, Korn, and Schöbel (in press), to evaluate the hedging strategy of the German firm Metallgesellschaft.

**TABLE IX**
Estimation Results for the Data Model

| | *Parameter* | | | *Ljung-Box-test (12 lags)* | *LR-test for ARCH (12 lags)* |
|---|---|---|---|---|---|
| *Contract* | $\hat{a}$ *(St. dev.)* | $\hat{b}$ *(St. dev.)* | $R^2$ | *Test statistic (p-value)* | *Test statistic (p-value)* |
| Spot | 0.300<br>(0.089) | 0.902<br>(0.029) | 0.83 | 17.31<br>(0.136) | 24.81<br>(0.016) |
| 2 months | 0.002<br>(0.002) | 0.587<br>(0.040) | 0.53 | 11.00<br>(0.529) | 4.72<br>(0.960) |
| 3 months | 0.003<br>(0.003) | 0.669<br>(0.039) | 0.60 | 12.45<br>(0.410) | 5.35<br>(0.945) |
| 4 months | 0.003<br>(0.003) | 0.716<br>(0.038) | 0.64 | 12.76<br>(0.387) | 6.79<br>(0.871) |
| 5 months | 0.004<br>(0.003) | 0.747<br>(0.038) | 0.66 | 12.97<br>(0.371) | 8.92<br>(0.710) |
| 6 months | 0.005<br>(0.004) | 0.769<br>(0.038) | 0.68 | 13.51<br>(0.333) | 11.13<br>(0.518) |
| 7 months | 0.005<br>(0.004) | 0.785<br>(0.037) | 0.69 | 13.90<br>(0.307) | 12.39<br>(0.415) |

Table IX shows single equation OLS estimates of System (19) and some results from diagnostic tests. The Ljung-Box tests give no indication of autocorrelation in the residuals of the data model. Moreover, there is little evidence for ARCH effects in the residuals. ARCH effects are always insignificant on a 1% level, and they are significant on a 5% level in only one case. Overall, if there is any time-series dependence in the residuals, it is weak. Based on this result, the analysis can apply a standard bootstrap to sample the residuals of Model (19). In general, whole residual vectors $u_{1t}, \ldots, u_{7t}$ are drawn to maintain any cross-sectional correlation.

The simulation of price paths from the data model starts with the values of the log spot price and relative bases on December 20, 2002, the last trading day of the January 2003 contract. In a first step, a residual vector is drawn from the data model, and the simulated log spot price and bases are calculated for the next month according to System (19). Starting from these new values, another residual vector is drawn that allows the calculation of new prices again. This procedure is continued until log spot prices and relative bases for a time horizon of up to nine years are generated. From these data, price paths for spot oil and futures

**TABLE X**
Results of Creation of Synthetic Forward Contracts (one barrel of crude oil)

| | *Model 1* | *Model 2* | *Model 3* |
|---|---|---|---|
| *Contract Maturity* | | *Theoretical forward price (in \$)* | |
| 48 months | 18.594 | 22.259 | 20.275 |
| 72 months | 17.793 | 21.088 | 18.786 |
| 108 months | 17.479 | 19.448 | 17.432 |
| | | *Mean absolute error (in \$)* | |
| 48 months | 1.815 | 2.102 | 1.181 |
| 72 months | 1.751 | 2.574 | 1.266 |
| 108 months | 1.750 | 3.122 | 1.315 |
| | | *Mean error (in \$)* | |
| 48 months | −0.811 | 1.445 | −0.188 |
| 72 months | −0.541 | 1.778 | −0.162 |
| 108 months | −0.397 | 2.222 | −0.101 |

*Note.* Model parameters and data model have been estimated using the whole data set.

are calculated. These simulated prices together with the parameter estimates from Table III are then used to implement the replication strategies according to Equations (13) to (15).[11] In Equations (13) to (15), the short-term contracts are the five-months futures (Model 1) and the three- and seven-months futures (Models 2 and 3). Initial maturities of the long-term contracts are either 48, 72, or 108 months. Long-term forward prices entering the replication strategies are always theoretical prices, obtained from the simulated prices of short-term futures and the model parameters. The replication portfolios are rebalanced (rolled over) at the end of a month and held constant in the meantime to avoid excessive rebalancing. No daily marking-to-market is considered. After a month, the (positive or negative) proceeds from the short-term futures are calculated and added to a savings account, which allows either investing or borrowing for a fixed rate of 5% per year. At the end of the hedge horizon, replication errors are calculated by subtracting the payments of the expiring long-term forward from the balance of the savings account.

Table X shows the mean absolute replication errors and mean replication errors of the strategies implied by different drift specifications.

[11]Because futures contracts are used to create a synthetic long-term forward contract, the price $F_3$ in Equations (13) to (15) must be replaced by the discounted price.

Replication refers to the creation of a synthetic forward that delivers one barrel of crude oil in either 48, 72, or 108 months. Means are taken over the 20,000 replication errors for each strategy. Mean absolute errors are generally quite high. Model 3 shows the smallest values of about \$1.3, followed by Models 1 and 2. Differences between model variants lie within the range of \$0.43 and \$1.8. Thus, there is a significant difference between alternative drift specifications in economic terms.[12] Notably, for the stationary models (Model 1 and Model 3) the mean absolute error barely changes with the time to maturity of the long-term contract. For the nonstationary model, Model 2, the mean absolute replication error is much larger for a 108-months contract than for a 48-months contract. Figure 1 shows a possible reason for this result. Because Model 2 takes a one-to-one net hedge position regardless of the time to maturity of the long-term contract, basis risk increases almost proportionally with the number of times the position is rolled over. In contrast, basis risk should not increase considerably for the stationary models.

A positive mean error, as is observed from Model 2, means that a strategy that is short one long-term forward, and then hedges this position by rolling over short-term futures leads to gains on average. However, a strategy that is long one long-term forward and hedges with short-term futures looses money on average. Thus, a nonzero mean error could be either a desirable or undesirable property in practice, depending on whether a long position or a short position is replicated. However, in comparison to a perfect model, which always delivers a replication error of zero, a nonzero mean error is certainly a model deficiency.

So far, both the parameters of the valuation models and the parameters of the data model used for simulation have been estimated from the whole data set. In this sense, Table X provides in-sample results that might be too optimistic. To check this conjecture, the following control study has been performed: The last 10 years of the data set are reserved for estimation of the data model. The remaining six and half years are used for estimation of the parameters of the valuation models, that is, there is no overlap in the two periods. Accordingly, the new starting date for the simulation is now January 20, 1993, the first date that has not been used to estimate parameters of the valuation models.

[12] Some indication about the statistical significance of the differences would also be useful. In principle, one could bootstrap the data model. Then, for every bootstrap sample, one would have to estimate the data model, simulate the paths, and evaluate all replication strategies. However, this procedure is hardly feasible for a reasonable number of bootstrap samples.

**TABLE XI**
Results of Creation of Synthetic Forward Contracts (one barrel of crude oil)

| | *Model 1* | *Model 2* | *Model 3* |
|---|---|---|---|
| *Contract maturity* | | *Theoretical forward price (in \$)* | |
| 48 months | 16.812 | 18.691 | 18.140 |
| 72 months | 16.393 | 18.214 | 17.485 |
| 108 months | 16.163 | 17.522 | 16.774 |
| | | *Mean absolute error (in \$)* | |
| 48 months | 2.049 | 3.245 | 1.914 |
| 72 months | 2.210 | 5.326 | 2.436 |
| 108 months | 2.299 | 8.331 | 2.777 |
| | | *Mean error (in \$)* | |
| 48 months | −0.745 | 2.151 | 0.815 |
| 72 months | −0.975 | 4.349 | 1.382 |
| 108 months | −1.158 | 7.580 | 1.806 |

*Note.* Model parameters have been estimated using the first six and a half years of the data period. The data model has been estimated using the last ten years of the data period.

Table XI shows the results of the control study. As expected, the mean absolute errors increase for all models. The highest increase is for Model 2. Similar to the results of Table X, the errors of the stationary models are relatively stable if the time to maturity increases. But there is an enormous increase in errors with the time to maturity for the nonstationary model.

In sum, the results on the synthetic creation of forwards show again that a stationary two-factor models leads to considerable improvements compared to a nonstationary two-factor model if long-term contracts are involved. In this application, even a stationary one-factor model performs better than the nonstationary two-factor model.

## CONCLUSION

This article presents a reduced-form affine two-factor model for the pricing of commodity derivatives. The model generalizes the model by Schwartz and Smith (2000) with respect to the drift specification of the stochastic factors. It allows for two mean-reverting factors and therefore for stationary spot and futures prices.

An empirical study for the crude oil market compares the new model with the stationary one-factor model by Schwartz (1997) and the nonstationary two-factor model by Schwartz and Smith (2000). Based on statistical criteria and on the results for different pricing and hedging applications, the paper comes to two main conclusions: First, a one-factor model does not seem to be flexible enough to provide satisfactory results in many instances. Second, the use of the more general stationary two-factor model instead of the nonstationary two-factor model brings additional benefits with no additional costs. Both two-factor models are almost identical in terms of analytical tractability and ease of implementation, because the more general model has only one more parameter. However, this extra parameter could either help or hurt in out-of sample applications. As the results show, it never hurts but leads rather to considerably better results with respect to long maturity contracts. This is an important finding, because for many applications with real options, where the use of more general three- or four-factor models is not feasible, very long time horizons are relevant.

Of course, the analysis of this article is restricted to the crude oil market. However, because mean reversion is a well-known stylized fact for many commodities, it can be conjectured that the simple model extension suggested here is also useful for other markets. Tests of this conjecture are left for future research.

## APPENDIX

It follows from the properties of the Ornstein-Uhlenbeck processes in Equations (9) and (10) that conditional on $\chi_0$ and $\xi_0$, the distribution of the state variables at time $T$ under the risk-neutral measure is bivariate normal with the following mean vector and variance-covariance matrix:

$$E_0^*([\chi_T,\ \xi_T]) = \left[e^{-\kappa T}\chi_0 - (1-e^{-\kappa T})\frac{\lambda_\chi}{\kappa},\ e^{-\gamma T}\xi_0 + (1-e^{-\gamma T})\left(\Theta - \frac{\lambda_\xi}{\gamma}\right)\right] \quad (20)$$

$$Cov_0^*([\chi_T,\ \xi_T]) = \begin{bmatrix} (1-e^{-2\kappa T})\dfrac{\sigma_\chi^2}{2\kappa} & (1-e^{-(\kappa+\gamma)T})\dfrac{\rho_{\chi\xi}\sigma_\chi\ \sigma_\xi}{(\kappa+\gamma)} \\ (1-e^{-(\kappa+\gamma)T})\dfrac{\rho_{\chi\xi}\sigma_\chi\ \sigma_\xi}{(\kappa+\gamma)} & (1-e^{-2\gamma T})\dfrac{\sigma_\gamma^2}{2\gamma} \end{bmatrix} \quad (21)$$

Using the above moments, the definition of the log spot price in Equation (8) implies the following expectation and variance of $\ln S_T$:

$$E_0^*(\ln S_T) = \left[e^{-\kappa T}\chi_0 - (1-e^{-\kappa T})\frac{\lambda_\chi}{\kappa}\right.$$
$$\left. + \frac{\kappa}{\kappa-\gamma}\left[e^{-\gamma T}\xi_0 + (1-e^{-\gamma T})\left(\Theta - \frac{\lambda_\xi}{\gamma}\right)\right] - \frac{\gamma\Theta}{\kappa-\gamma}\right] \quad (22)$$

$$Var_0^*(\ln S_T) = \left[ (1 - e^{-2\kappa T}) \frac{\sigma_\chi^2}{2\kappa} + \frac{\kappa^2}{(\kappa - \gamma)^2}(1 - e^{-2\gamma T})\frac{\sigma_\xi^2}{2\gamma} + 2\frac{\kappa}{(\kappa - \gamma)}(1 - e^{-(\kappa+\gamma)T})\frac{\rho_{\chi\xi}\sigma_\chi \sigma_\xi}{(\kappa + \gamma)} \right] \quad (23)$$

Because $\ln S_T$ is normally distributed, it follows:

$$F(0, T) = E_0^*(S_T) = \exp\{E_0^*(\ln S_T) + 1/2(Var_0^*(\ln S_T))\} \quad (24)$$

Inserting the expressions from Equations (22) and (23) into Equation (24) delivers the pricing formula (11).

## BIBLIOGRAPHY

Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle. In B. N. Petrov & F. Csáki (Eds.), Second international symposium on information theory. Budapest, Akadémiai Kiadó.

Bessembinder, H., Coughenour, J. F., Seguin, P. J., & Smoller, M. A. (1995). Mean reversion in asset prices: Evidence from the futures term structure. Journal of Finance, 50, 361–375.

Bollen, N. P., & Whaley, R. E. (1998). Simulating supply. Risk, 11, 143–147.

Brennan, M. J. (1991). The price of convenience and the valuation of commodity contingent claims. In D. Lund & B. Øksendahl (Eds.), Stochastic models and option values, applications to resources, environment and investment problems. Amsterdam: North-Holland.

Brennan, M. J., & Schwartz, E. S. (1985). Evaluating natural resource investments. Journal of Business, 58, 135–157.

Bühler, W., Korn, O., & Schöbel, R. (in press). Hedging long-term forwards with short-term futures: A two-regime approach. Review of Derivatives Research.

Campbell, J. Y., & Perron, P. (1991). Pitfalls and opportunities: What macroeconomists should know about unit roots. In O. Blanchard & S. Fischer (Eds.), NBER macroeconomics annual. Cambridge, MA: MIT Press.

Casassus, J., & Collin-Dufresne, P. (2002). Maximal affine model of convenience yields implied from interest rates and commodity futures. Discussion paper, Carnegie Mellon University.

Chen, R. R., & Scott, L. (1993). Maximum likelihood estimation for a multifactor equilibrium model of the term structure of interest rates. Journal of Fixed Income, 3, 14–31.

Cortazar, G., & Naranjo, L. (2003). A multi-factor stochastic model and estimation procedure for the valuation and hedging of commodity contingent claims. Discussion paper, Pontificia Universidad Catolica de Chile.

Cox, J. C., Ingersoll, J. E., & Ross, S. A. (1985). An intertemporal general equilibrium model of asset prices. Econometrica, 53, 363–384.

Culp, C. L., & Miller, M. H. (1999). Corporate hedging in theory and practice. London: Risk Books.

Dickey, D. A., & Fuller, W. A. (1979). Distribution of the estimators for autoregressive time series with a unit root. Journal of the American Statistical Association, 74, 427–431.

Duffie, D., & Singleton, K. J. (1997). An econometric model of the term structure of interest-rate swap yields. Journal of Finance, 52, 1287–1321.

Duffie, D., & Stanton, R. (1992). Pricing continuously resettled contingent claims. Journal of Economic Dynamics and Control, 16, 561–573.

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

Gourieroux, C., & Monfort, A. (1997). Time series and dynamic models. Cambridge: Cambridge University Press.

Hall, A. (1994). Testing for a unit root in time series with pre-test data-based model selection. Journal of Business and Economics Statistics, 12, 461–470.

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

Hilliard, J. E., & Reis, J. (1998). Valuation of commodity futures and options under stochastic convenience yields, interest rates, and jump diffusions in the spot. Journal of Financial and Quantitative Analysis, 33, 61–86.

Kellerhals, B. P. (2001). Financial pricing models in continuous time and Kalman filtering. Berlin: Springer.

Kwiatkowski, D., Phillips, P. C. B., Schmidt, P., & Shin, Y. (1992). Testing the null hypothesis of stationarity against the alternative of a unit root: How sure are we that economic time series have a unit root? Journal of Econometrics, 54, 159–178.

Newey, W. K., & West, K. D. (1987). A simple positive semi-definite heteroskedasticity and autocorrelation-consistent covariance matrix. Econometrica, 55, 703–708.

Pearson, N. D., & Sun, T. S. (1994). Exploiting the conditional density in estimating the term structure: An application to the Cox, Ingersoll, and Ross model. Journal of Finance, 49, 1279–1304.

Pilipovic, D. (1998). Energy risk: Valuing and managing energy derivatives. New York: McGraw Hill.

Richard, S. F., & Sundaresan, M. (1981). A continuous-time equilibrium model of forward prices and futures prices in a multigood economy. Journal of Financial Economics, 9, 347–371.

Ross, S. A. (1997). Hedging long run commitments: Exercises in incomplete market pricing. Economic Notes by Banca Monte dei Paschi di Siena SpA, 26, 385–420.

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

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

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

Schwarz, G. (1978). Estimating the dimension of a model. Annals of Statistics, 6, 461–464.

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