June 2003

The Informational Value of Crude Oil Futures Prices

Delphine LAUTIER*

Cahier de recherche n° **2003-06**

____

*CEREG, Université Paris Dauphine

THE INFORMATIONAL VALUE OF CRUDE OIL FUTURES PRICES

Delphine LAUTIER

ABSTRACT: In order to enhance the understanding of the term structure of commodity prices, this article examines the temporal integration of the American crude oil futures market. The study relies on a database including futures prices for very long maturities (as far as seven years) and compares their informational value for various delivery dates. Segmentation is defined as a situation where different parts of the prices curve are disconnected from each other's. Consequently, the information conveyed by some futures prices is useless to reconstitute the rest of the curve. Empirical tests are carried out with a term structure model, whose performances depend on the informational value of the futures prices retained for its estimation. The results lead to the conclusion that the crude oil futures market is segmented into three parts. The first corresponds to maturities below 28 months, the second is situated between the 29^{th} and the 47^{th} months, and the third consists of maturities ranging from the 4^{th} to the 7^{th} year.

KEY WORDS: Information – futures prices – term structure – segmentation – crude oil.

Professional titles and affiliations: Assistant professor, Cereg, University Paris IX
Associate Research Fellow, Cerna, ENSMP

Mailing Address: Delphine Lautier
Cereg, Université Paris IX
Place du Maréchal de Lattre de Tassigny
75 775 Paris Cedex 16

Telephone number: 33 1 44 05 46 42
Fax number: 33 1 44 05 40 23
Email: lautier@cerna.ensmp.fr

THE INFORMATIONAL VALUE OF CRUDE OIL FUTURES PRICES[1]

ABSTRACT: In order to enhance the understanding of the term structure of commodity prices, this article examines the temporal integration of the American crude oil futures market. The study relies on a database including futures prices for very long maturities (as far as seven years) and compares their informational value for various delivery dates. Segmentation is defined as a situation where different parts of the prices curve are disconnected from each other's. Consequently, the information conveyed by some futures prices is useless to reconstitute the rest of the curve. Empirical tests are carried out with a term structure model, whose performances depend on the informational value of the futures prices retained for its estimation. The results lead to the conclusion that the crude oil futures market is segmented into three parts. The first corresponds to maturities below 28 months, the second is situated between the 29th and the 47th months, and the third consists of maturities ranging from the 4th to the 7th year.

KEY WORDS: Information – futures prices – term structure – segmentation – crude oil.

INTRODUCTION

The term structure of commodity futures prices is the relationship linking the spot price and futures prices for different delivery dates. It synthesizes the information available in the market and the operators' expectations concerning the future. This information is very useful for management purposes: it can be used to hedge exposures on the physical market, to adjust the stocks level or the production rate. It can also be used to undertake arbitrage transactions, to evaluate derivatives instruments based on futures contracts, etc.

In the American crude oil market, this information is especially abundant because since 1999, there are futures contracts for maturities as far as seven years. Thus, this market provides publicly

[1]The author is grateful for the very helpful comments and precious suggestions of Professor Franklin Edwards (Columbia University). This study has also benefited from the support of the French Institute of Energy (Institut Français de l'Energie) and from TotalFinaElf that provided the data.

available prices – namely potentially informative and costless signals – whereas in most commodities markets, the only information for far maturities is private and given by forward prices. The introduction of these long term futures contracts authorizes empirical studies on the crude oil prices' curves that were only possible before with forward prices, whose informational content is not necessarily reliable or workable (the forward contracts are not standardized, the prices reporting mechanism does not force the operators to disclose their transactions prices, etc.).

This article aims to take advantage of the newly available information provided by the long term crude oil futures contracts in order to improve the understanding of the futures prices' behavior of storable commodities and in order to facilitate the use of the term structure as a management tool. More precisely, its objective is to compare the informational content of futures prices for different maturities. It reaches the following questions: do some maturities provide more information than others do? How much information gives one specific futures price on the rest of the prices curve? Is that information sufficient to reconstitute the whole term structure of futures prices? Are some maturities more important for certain parts of the curve, and irrelevant for the rest of the curve?

The answers given to these questions have crucial implications for financial decisions, particularly for all the hedging and valuation operations relying on the relationship between different futures prices. It is the case, for example, of the "stack and roll" hedging strategies that rely on short term futures contracts to protect long term positions on the physical market (Metallgesellschaft, in 1994, tried to build such a strategy in order to exploit the higher liquidity of the nearest contracts). The efficiency of these strategies can be affected by differences in the informational content of futures prices. It is also the case of investment decision, when the latter is based on the extrapolation from observed prices curves to value cash flows for maturities that are not available in the market[2]. All these operations rely on term structure models of commodity prices. Such a tool aims to reproduce the futures prices observed in the

---

[2] For real option analysis, see for example Brennan and Schwartz (1985), Schwartz (1997, 1998), Cortazar and Schwartz (1998), Schwartz and Smith (2000), Cortazar *et al* (2001).

market as accurately as possible and to extend the curve for very long maturities. Its use requires however, the estimation of its parameters, and the latter may depend on the informational content of futures prices.

In this empirical study, it is supposed that, all things being equal, the performances of a term structure model (*i.e.* its ability to replicate the prices curve) depend on the informational value of the futures prices retained for its parameters' estimation. The two-factor model proposed by Schwartz in 1997, whose performances were firmly established – on different commodity markets and different periods – authorizes the study of the futures prices' informational content.

The article proceeds as follows. First, a brief analysis of the term structure of commodity futures prices is presented and the central assumption of the paper is exposed. Second, the methodology of the empirical study is exposed, namely the term structure model retained, the method used to evaluate the informational content of futures prices, and the data. Third, the empirical results are exposed.

ANALYSIS OF THE TERM STRUCTURE OF COMMODITY FUTURES PRICES

The normal backwardation and the storage theories are traditionally used to explain the relationship between spot and futures prices in commodity markets. Keynes introduced the normal backwardation theory in 1930. In that case, the difference between spot and futures prices is due to an unbalance between short and long hedging positions in the futures market. To compensate for this lack of balance, there is a need for speculators. A premium remunerates the latter for the risk they undertake in their activity. Until now however, the theory of normal backwardation was never truly validated nor rejected, probably because the level and the direction of the risk premium are not constant.

The storage theory relies mainly on the storage costs and on the convenience yield to understand the situations of contango and backwardation. The concept of convenience yield was introduced by Kaldor in 1939. It can be briefly defined as the implicit gain associated with the holding of inventories. The stocks availability indeed prevents from production disruptions, and it enables to take advantage of unexpected rises in the demand (Brennan, 1958). The storage theory was validated on numerous commodity markets. It is the main theoretical basis for the elaboration of term structure models. However, when the analysis is centred on long term horizon, one may ask if the explanatory factors of the storage theory are still of use.

Gabillon[3] (1995) was the first to give a negative answer to that question. He proposed a theoretical analysis of the term structure of crude oil prices, where the curve is separated into two distinct segments. Each part of the curve reflects a specific economic behaviour of the operators. The first segment corresponds to the shorter maturities and it is mostly used for hedging purposes. As a result, production, consumption, stocks level and the fear of inventories disruptions are the most important explanatory factors of the prices relationship. For longer maturities however, the explanatory factors change: interest rates, anticipated inflation and the prices for competing energies determine the futures prices. In that case, the information provided by the prices is used for investment purposes. In 2000, Schwartz and Smith also propose to distinguish different explanatory factors in step with the maturity of futures prices.

If different market forces determine the short-term and the long-term futures prices, one may ask whether there is a temporal segmentation of the term structure. According to the market segmentation or preferred habitat approach to the term structure of interest rates (Modigliani and Sutch, 1966), segmentation can arise when a market participant has preferences for the holding of certain subsets of maturities, which might be termed the "preferred habitat" of that agent. In that situation, there exists a separate supply and demand for assets at each habitat, which traders are reluctant to quit, unless they earn a sufficiently high return.

Many reasons were invoked to account for segmentation. The latter can be the result of institutional barriers such as taxes or restrictions on short sales. It is also explained by ownership restrictions, constraints on capital, information costs, management risks, etc. In the case of crude oil

[3] Gabillon (1995) also proposed a term structure model of commodity prices, where the state variables are the spot price and the long term price. However, this model was never tested.

markets, according to Gabillon (1995), the existence of preferred habitat is due to the presence of several categories of participants in the futures market, each of them acting on different parts of the prices curve. Producers and consumers operate indeed primarily at the shorter end of the curve, whereas other investors (banks for example) are located at the longer maturities.

Whereas the spatial integration has already been examined in the case of commodity markets – see for example Kleit (2001) or Milonas and Henker (2001) – empirical tests on temporal integration were never carried out. This paper is an attempt to complement the existing literature on term structures of commodity prices by investigating this temporal integration. Segmentation is defined as a situation where different parts of the prices curve are disconnected from each other's. As a result of segmentation, the crude oil futures prices curve should be separated into several peaces – two if Gabillon is right, but maybe more – and the information conveyed by the prices should be of different nature when the contract's maturity changes. The validation of this assumption should be valuable for the comprehension of prices behavior and for the use of term structure models as management tools.

METHODOLOGY

This section exposes the method used to evaluate the informational content of futures prices. This method relies on the use of a term structure model. It is supposed that the model's performances depend on the informational value of the futures prices retained for the parameters' estimation. The term structure model retained for the tests is first presented. Then, the data and the information sets chosen for the parameters estimation are exposed.

The term structure model

The first term structure model of commodity prices was proposed by Brennan and Schwartz in 1985. In this model, the behavior of the futures price is explained by one single variable: the spot price, which is assumed to follow geometric Brownian motion. This model, which is very simple, is also not suited for long term maturities. Indeed, the spot price dynamics ignores that the operators in the physical market adjust their stocks to the evolution of the spot price and to the modifications in supply and demand. Moreover, this representation ignores that a futures price depends not only on the spot price but also on the convenience yield. In this model, the convenience yield is a constant parameter. However, in 1989, Gibson and Schwartz showed that such an analysis is limited. They proposed to introduce the convenience yield as a second state variable. Schwartz (1997) retained that proposition.

This author compares three term structure models of commodity prices. The first relies on a single state variable and it is unable to price correctly the term structure of futures prices. The second is the two-factor model retained in this study. In the third model, because he thought that for long term analysis, the interest rates should no longer be considered as a constant, Schwartz also introduced a stochastic interest rate. However, using forward prices provided by Enron, he showed that the two- and the three-factor models are empirically very similar. This result was confirmed by Lautier and Galli (2002, b), which applied a principal component analysis to term structure of futures prices and showed that only two factors (parallel and relative shifts) are sufficient to explain the movements of the prices curves[4]. Thus, the three-factor model is of little interest. In addition, the two-factor model has proven its ability to reproduce the term structure of futures prices on the crude oil and on the cooper markets, on the shorter as well as on the longest maturities, and on several periods. Moreover, it is of relatively easy use because it is linear and has an analytical solution. Finally, it was used as a reference to develop more sophisticated models[5]. All these reasons lead to retain it for the empirical study conducted in this article.

[4] More precisely, empirical tests carried out on futures prices with maturities ranging from 1 month to seven years showed that during 1999-2002 these two factors account for 99% of the total variance of crude oil prices' curve.

[5] In 1998, Hilliard and Reis proposed to introduce jumps in the spot price process, in order to take into account the large and abrupt changes, due to supply and demand shocks, that affect certain commodity markets (especially the energy commodities used for heating). Lautier (2002) introduced a mean reverting convenience yield with an asymmetrical behavior. This representation traduces the fact that in commodity markets, the basis is more volatile in backwardation than in contango. Yan (2002) introduced a stochastic volatility.

The two-factor model supposes that the spot price S and the convenience yield C can explain the behavior of the futures price F. The dynamics of these state variables is:

$$\begin{cases} dS = (\mu - C)Sdt + \sigma_S S dz_S \\ dC = [k(\alpha - C)]dt + \sigma_C dz_C \end{cases} \quad [1]$$

with: κ, σ_S, σ_C >0

where: - μ is the drift of the spot price,

- σ_S is the spot price volatility,

- dz_S is an increment to a standard Brownian motion associated with S,

- α is the long run mean of the convenience yield,

- κ is the speed of adjustment of the convenience yield,

- σ_C is the volatility of the convenience yield,

- dz_C is an increment to a standard Brownian motion associated with C.

In this model, the convenience yield is mean reverting. This formulation relies on the hypothesis that there is a level of stocks, which satisfies the needs of industry under normal conditions. The behaviour of the operators in the physical market guarantees the existence of this normal level. When the convenience yield is low, the stocks are abundant and the operators sustain a high storage cost compared with the benefits related to holding the raw materials. So if they are rational, they try to reduce these surplus stocks. Conversely, when the stocks are rare the operators tend to reconstitute them.

Moreover, as the storage theory showed it, the two state variables are correlated. Both the spot price and the convenience yield are indeed an inverse function of the inventories level. Nevertheless, as Gibson and Schwartz (1990) demonstrated it, the correlation between these two variables is not perfect. Therefore, the increments to standard Brownian motions are correlated, with:

$$E[dz_S \times dz_C] = \rho dt$$

where ρ is the correlation coefficient.

An arbitrage reasoning and the construction of a hedge portfolio lead to the solution of the model. It expresses the relationship at t between an observable futures price F for delivery at T and the state variables S and C. This solution is:

$$F(S,C,t,T) = S(t) \times \exp\left[-C(t)\frac{1-e^{-\kappa\tau}}{\kappa} + B(\tau) \right] \qquad [2]$$

with: - $B(\tau) = \left[\left(r - \hat{\alpha} + \frac{\sigma_C^2}{2\kappa^2} - \frac{\sigma_S \sigma_C \rho}{\kappa} \right) \times \tau \right] + \left[\frac{\sigma_C^2}{4} \times \frac{1-e^{-2\kappa\tau}}{\kappa^3} \right] + \left[\left(\hat{\alpha}\kappa + \sigma_S \sigma_C \rho - \frac{\sigma_C^2}{\kappa} \right) \times \left(\frac{1-e^{-\kappa\tau}}{\kappa^2} \right) \right]$

- $\hat{\alpha} = \alpha - (\lambda / \kappa)$

where: - r is the risk free interest rate, assumed constant,

- λ is the market price of convenience yield risk,

- $\tau = T - t$ is the maturity of the futures contract.

The model's performances measure its ability to reproduce the term structure of commodity prices. To assess these performances, criteria and parameters values are needed. Two criteria are retained: the mean pricing error and the root mean squared error[6]. The parameters values are also necessary to compute the estimated futures prices and to compare them with empirical data. The parameters estimation is not obvious though, because as many term structure models, Schwartz's one relies on non-observable state variables. Both the spot price and the convenience yield are indeed regarded as non-observable because most of the time there are no reliable time series for the spot price and the convenience yield is not a traded asset. In order to cope with this difficulty, the method that was proposed by Schwartz in 1997, namely a Kalman filter, is used[7].

Relying on these criteria and parameters values, the sensitivity of the model performances to the information used for the estimation can be tested. This information is constituted of temporal series of futures prices corresponding to a few selected maturities.

---

[6] Appendix 1 presents these criteria.

[7] Appendix 1 presents this method.

Information used

In this study, the informational content of futures prices is supposed to be captured by the parameters of the model. In order to measure this informational content, several sets of futures prices having various maturities are selected and used for the estimation of the parameters. Each set gathers four maturities[8]. The latter are then used to determine, at each observation date, the corresponding futures prices (first for the four maturities selected, afterwards for the remaining ones). These prices constitute a theoretical curve based upon a specific information set. Lastly, relying on the performance criteria previously defined, this curve is compared with the empirical one observed at the same date. Such a comparison, on the whole study period, makes it possible to appreciate whether the maturities set retained contains enough information to correctly reconstitute the curve. The repetition of this test with different sets of maturities makes it possible to investigate whether some sets lead to better performances than others do, namely whether they provide more information than others do.

The empirical tests are carried out with seven sets of maturities. In order to examine if the assumption – according to which the short- and the long-term futures prices are influenced by different factors – is valuable, the two extremities of the prices curve were first of all retained: the nearby delivery months (1st, 2nd, 3rd and 4th months), and the longer maturities (48th, 60th, 72nd, and 84th months). Then, keeping the information on the shorter expiration date (1st month), the other maturities were progressively moved away. So the parameters corresponding to the 1st, 3rd, 6th, and 9th months were estimated. Then the 1st, 5th, 10th, and 15th months were used. The fifth set corresponds to the 1st, 6th, 12th, and 18th months, the sixth set to the 1st, 12th, 24th and 48th months. Finally, a set with maturities regularly distributed along the prices curve was retained: it gathers the 1st, 24th, 48th, and the 84th months.

Data

The database is an important element of the study. Considering the volume and the maturity of the transactions, the American crude oil futures market is today the most developed commodity futures market. Thus, working with crude oil prices makes it possible to study maturities as far as seven years[9]. The database covers the period from 06/01/1999 to 01/14/02. The crude oil prices are daily settlement prices for the West Texas Intermediate futures contract of the New York Mercantile Exchange. They have been operated such as the first futures price maturity corresponds to the one month maturity, such as the second futures price corresponds to the two months maturity, and so forth. Lastly, weekly data were retained for the estimations (139 weeks).

The choice of the temporal series of futures prices is guided by liquidity concerns. On the crude oil market as on other commodity markets, the transaction volume is always higher on the nearest maturities. However, the crude oil market is also characterized by a lack of liquidity on intermediate maturities: there are no regularly available data, on the study period, for the futures contracts corresponding to maturities situated between the 29th and the 47th months. Thus, they were removed of the database. Finally, thirty-two different maturities were selected. From the 1st to the 28th months, each temporal series corresponds to the maturity of a futures contract. On a longer horizon, the maturities for the 48th, the 60th, the 72nd, and the 84th delivery months were also retained.

The interest rates are daily T-bill rates for a three months maturity and were extracted from Datastream. Because interest rates are supposed to be constant in the model, the mean of the observations between June 1999 and January 2002 was retained.

---

[8] This number of maturities is retained because it increases the stability of the estimation procedure.

[9] However, the transaction volume is not very high for the longer maturities. Therefore, the quality of the information conveyed by those prices may not be excellent.

EMPIRICAL RESULTS

This section presents the empirical results. The model's performances as well as their sensitivity to the informational content of the parameters are analyzed. Then the parameters obtained with the seven maturities' sets are exposed.

The model's performances and their sensitivity to the information used

The model's performances are first of all presented for all the maturities retained in the study. However, the nature of the results leads to separate the shorter from the longer maturities.

Reconstitution of the prices curve

Table I illustrates the MPE and RMSE obtained on the study period, for all the maturities. Each line represents the averages MPE and RMSE related to a specific information set (*i.e.* one set of parameters). Appendix 2 provides the entire results, maturity by maturity.

Table I.
Average MPE and RMSE for all the maturities

Set of maturities	*MPE*	*RMSE*
1st, 2nd, 3rd, 4th months	6.2710	7.2579
1st, 3rd, 6th, 9th months	2.4462	3.5619
1st, 5th, 10th, 15th months	1.0169	2.0829
1st, 6th, 12th, 18th, months	-0.6909	1.4162
1st, 12th, 24th, 48th months	-0.1779	1.1467
1st, 24th, 48th, 84th months	-0.1189	1.0518
48th, 60th, 72nd, 84th months	1.2777	2.2264

Unit : USD/b

The empirical tests carried out on the study period arouse two statements. First, the informational content of futures prices changes with the contract's expiration date. Indeed, the choice of a specific maturities' set alters significantly the model's performances. Its ability to reproduce the prices curve is weak with the shorter maturities, and very good with the set gathering the 1^{st}, the 24^{th}, the 48^{th}, and the 84^{th} months. Thus, even if Schwartz's model supposes that the interest rate is constant and ignores political risk or eventual shocks on demand and supply, it is suited for the replication of the shorter as well as the longer part of the prices curve (MPE lower than 12 cents, and average RMSE around USD 1.05), provided the proper information set is retained. Second, each extremity of the prices curve has a specific and useful informational content. Indeed, the best performances are achieved with the set including the two extremities. Thus, both of them provide some useful information for the reconstitution of the prices curve, and their informational content is different, as the observation of the performances associated with each extremity (lines 2 and 8 of Table I) corroborates it. This incites to examine separately the shorter and the longer maturities.

The shorter part of the curve

Table II exposes the average MPE and RMSE obtained with a shortened prices curve: from the 1^{st} to the 28^{th} months.

Table II.
Average MPE and RMSE for the 1^{st} to the 28^{th} months

Set of maturities	MPE	RMSE
1^{st}, 2^{nd}, 3^{rd}, 4^{th} months	0.7262	1.7463
1^{st}, 3^{rd}, 6^{th}, 9^{th} months	0.1696	1.3924
1^{st}, 5^{th}, 10^{th}, 15^{th} months	0.0312	1.2165
1^{st}, 6^{th}, 12^{th}, 18^{th}, months	-0.4948	1.2787
1^{st}, 12^{th}, 24^{th}, 48^{th} months	-0.0565	1.1154
1^{st}, 24^{th}, 48^{th}, 84^{th} months	-0.1474	1.0767
48^{th}, 60^{th}, 72^{nd}, 84^{th} months	1.4884	2.4225

Unit : USD/b

The elimination of the end of the curve makes it possible to reach a third statement: the segmentation witnessed by Gabillon in 1995 has disappeared. The latter supposed that the two segments of the prices curve were separated at the 18th month. However, the observation of the MPE and the RMSE month per month (see Appendix 2) provide no evidence of sizeable differences between the performances obtained with the maturities ranging from the 1st to the 18th months and those corresponding to the 18th to the 28th months. On the contrary, focusing on the shorter part of the curve brings together the performances of the different maturity sets, and shows that the shorter maturities can be reasonably retained in order to reconstitute the curve up to 28 months. Moreover, the performances of the first set of parameters are dramatically improved: compared with Table I, the RMSE falls from USD 7.2579 to USD 1.7463.

The longer part of the curve

The focus on the longer part of the curve (4th to 7th years), illustrated by Table III, arouses a fourth statement: the information concentrated on the shorter maturities is useless to reconstitute the end of the prices curve. Indeed, when estimated on the nearest maturities, Schwartz's model leads to long term futures prices with no economical sense: the average MPE are around USD 45 per barrel!

Table III.

Average MPE and RMSE for the 4th to the 7th years

Set of maturities	MPE	RMSE
1st, 2nd, 3rd, 4th months	45.0846	45.8397
1st, 3rd, 6th, 9th months	18.3828	18.7487
1st, 5th, 10th, 15th months	7.9163	8.1481
1st, 6th, 12th, 18th, months	-2.0639	2.3785
1st, 12th, 24th, 48th months	-1.0277	1.3657
1st, 24th, 48th, 84th months	0.0808	0.8773
48th, 60th, 72nd, 84th months	-0.1970	0.8532

Unit : USD/b

Parameters

The last empirical results presented are the optimal parameters corresponding to each set of maturities. Schwartz's model includes seven parameters, all of them supposed to be constant. Nevertheless, in 1997, relying on forward prices for long term expiration dates, the author showed that they change with the maturities retained for the estimation. Table IV illustrates that the same result is reached with futures prices.

Table IV

Optimal parameters corresponding to each set of maturities

Parameters	1^{st}, 2^{nd}, 3^{rd}, 4^{th} months	1^{st}, 3^{rd}, 6^{th}, 9^{th} months	1^{st}, 5^{th}, 10^{th}, 15^{th} months	1^{st}, 6^{th}, 12^{th}, 18^{th}, months	1^{st}, 12^{th}, 24^{th}, 48^{th} months	1^{st}, 24^{th}, 48^{th}, 84^{th} months	48^{th}, 60^{th}, 72^{nd}, 84^{th} months
κ	2.8078	1.4307	1.0373	0.8365	0.5085	0.8512	0.0006
	(0.4887)	(0.3278)	(0.3731)	(1.4217)	(0.0913)	(0.0283)	(0.003)
μ	0.2228	0.3588	0.3818	0.3332	0.3631	0.2983	0.1700
	(0.0966)	(0.1053)	(0.0990)	(0.5118)	(0.0574)	(0.0793)	(0.0237)
σ_S	0.2927	0.2899	0.3005	0.3984	0.2147	0.2333	0.0936
	(0.0230)	(0.0316)	(0.0358)	(0.2778)	(0.0341)	(0.0231)	(0.0026)
α	0.1212	0.2171	0.2455	0.1861	0.2976	0.2437	0.0932
	(0.0632)	(0.0797)	(0.0780)	(0.6206)	(0.0565)	(0.0694)	(0.0060)
σ_C	0.4692	0.3052	0.2842	0.3113	0.0998	0.1956	0.0114
	(0.0830)	(0.0577)	(0.0892)	(0.6985)	(0.0331)	(0.0239)	(0.0002)
ρ	0.9272	0.9725	0.9563	0.9699	0.8979	0.9577	0.5553
	(9.0629)	(31.09)	(8.1609)	(0.0573)	(2.7469)	(5.7346)	(0.0577)
λ	-0.301	0.1485	0.1821	0.1519	0.1466	0.1782	0.0058
	(0.2068)	(0.1092)	(0.0916)	(0.3173)	(0.033)	(0.0625)	(0.0577)

The parameters values retained to initiate the optimization are: $\kappa = 0.5$; $\mu = 0.1$; $\sigma_S = 0.3$; $\alpha = 0.1$; $\sigma_C = 0.4$; $\rho = 0.5$; $\lambda = 0.1$. Standard deviations are in parentheses. For two sets of maturities (the 4^{th} and the 7^{th}) the parameters are obtained with a precision of 1^e-2 on the gradients, instead of 1^e-5.

The most important changes concern the speed of adjustment of the convenience yield, the volatilities of the two state variables and their correlation coefficient. All of them tend to decrease with the maturity. Indeed, in the model mean reversion concerns the stocks, which are of little importance for long term maturities. As a result, a low speed of adjustment characterises the longer part of the curve. The same kind of explanation can be evoked for the volatilities. Their level decreases when the maturity rises because then, the shocks on supply and demand have a lowest impact on the futures prices. Considering these changes, the parameters of the term structure models should ideally be maturity dependent: the speed of adjustment and the volatilities of the state variables should be considered as decreasing functions of the expiration date. However, the parameters being also time dependent, such a modification will improve strongly the model's complexity.

CONCLUSION AND POLICY IMPLICATIONS

This article is centered on the informational value of futures prices. The interest of the study lies in a better understanding of the behavior of the term structure of commodity prices and in an enhanced appreciation of the way to use the term structure models for management purposes. Relying on the performances of Schwartz's model to appreciate the informational content of futures prices, this empirical study shows that the information conveyed by the prices changes with the contract's maturity. More precisely, two important statements are reached.

First, each extremity of the prices curve has a specific and useful informational content. The differences between the two extremities are so high that the information concentrated on the nearest delivery dates is totally useless to reconstitute the long term futures prices. This first result leads to conclude that the crude oil futures prices' curve is segmented. Second, the study shows that on an informational point of view there are three coherent groups of futures prices: the first corresponds to maturities ranging from the 1^{st} to the 28^{th} months, the second is situated between the 29^{th} and the 47^{th} months, and the last consists of maturities from the 4^{th} to the 7^{th} years. Among these groups, two only have a real informational value: the first and the third. Thus, the prices curve is segmented into three parts, and the segmentation evolves with time: in 1995, according to Gabillon, it was situated around the 18^{th} month and there were only two segments.

This evolution can be explained by the repining process of the crude oil futures market. Since 1995, the Nymex experienced a growth in its transactions volume, pushing away the boundary of actively traded contracts. This phenomenon is reported in most derivatives markets, and one can expect that segmentation will move to longer maturities in the future. However, what is specific to the crude oil market is the introduction, since 1999, of long term futures contracts. This created a new segment in the prices curve, which is separated from the shorter segment by intermediate maturities with no informational content. This long term extremity of the curve is poorly connected with the nearest maturities. Thus, despite significant steps toward temporal integration, the latter is far from achieved in the case of the crude oil futures market.

The reasons explaining this segmentation of the crude oil futures markets can probably be found in the existence of different categories of participants located at the two extremities of the curve: the hedgers, acting on the short term maturities, and the investors situated on far distant prices. Insufficient liquidity prevents presumably these operators to leave their preferred habitat and to undertake arbitrage operations between maturities. The difficulty to initiate reverse cash and carry operations, due to the lack of physical stocks in markets characterized by backwardation is another explanatory factor, especially important for the crude oil market. Moreover, there are ownership restrictions on the Nymex that may prevent temporal integration. For example, the rulebook of the exchange states that the position on crude oil futures contracts shall not exceed 20 000 contracts.

The segmentation of the crude oil market has policy implications. Indeed, term structure models are used for hedging and investment purposes. This article shows however, than one must be very cautious when using such a model to extend the prices curve. The model must not only be well suited for long term applications. The information used for the estimation is also crucial because it alters significantly the model's performances. When the proper set of maturities is chosen – namely the two extremities of the curve – the ability of Schwartz's model to reproduce the futures prices for very long maturities is excellent. However, the same model can lead to prices with no economical sense if the nearest futures prices are used to reconstitute the long term ones.

Further work could be undertaken. Another investigation relying on the private information given by forward prices for maturities longer than seven years could conduct to the discovery of new segments in the crude oil market. The same methodology could also be reproduced on other periods, in order to examine if the segmentation really evolves with time. Such studies should lead to a more accurate use of the information and the hedging instruments provided by the futures markets.

BIBLIOGRAPHY

Brennan, M.J. (1958). The supply of storage. American Economic Review, 47, 50-72.

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

Cortazar, G., & Schwartz, E.S. (1998). Monte-carlo evaluation model of an undeveloped oil field. Journal of Energy Finance and Development, 3, 73-84.

Cortazar, G., Schwartz, E.S., & Casassus, J. (2001). Optimal exploration investments under price and geological-technical uncertainty: a real options model. R & D Management, 31, 181-189.

Gabillon, J. (1995). Analysing the forward curve, in Managing Energy Price risk (1st ed.), London: Risk Books.

Gibson, R., & Schwartz, E.S. (1989). Valuation of long term oil-linked assets. Working Paper, Anderson Graduate School of Management, UCLA.

Gibson, R., & Schwartz, E.S. (1990). Stochastic convenience yield and the pricing of oil contingent claims. Journal of Finance, 45, 959-975.

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

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

Javaheri, A., Lautier, D., & Galli, A. (2003). Filtering in finance. Willmott Magazine, 5, 67-83.

Kaldor, N. (1939). A note on the theory of the forward market. Review of Economic Studies, 8, 196-201.

Keynes, J.M. (1930). A Treatise on Money : The applied Theory of Money. Londres, England: Macmillan, vol. 2.

Kleit, A.N. (2001). Are regional oil markets growing closer together? An arbitrage cost approach. Energy Journal, 22, 1-15.

Lautier, D. (2002). Trois modèles de structure par terme des prix du pétrole: une comparaison. Banque et Marchés, 57, 46-57.

Lautier, D., & Galli, A. (2002,a). The Kalman filters in finance: an application to term structure models of commodity prices. Working Paper, Cereg, University Paris IX.

Lautier, D., & Galli, A. (2002, b). Report on term structure models of commodity prices: elaboration and improvement. French Institute of Energy Research (IFE), 70 p.

Milonas, T.N., & Henker, T. (2001). Price spread and convenience yield behaviour in the international oil market. Applied financial economics, 11, 23-36.

Modigliani, F., & Sutch, R. (1966). Innovations in interest rate policy. American Economic Review, 56, 178-197.

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

Schwartz, E.S. (1998). Valuing long term commodity assets. Journal of Energy Finance and Development, 3, 85-99.

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

Yan X., (2002). Valuation of commodity derivatives in a new multi-factor model. Review of Derivatives Research, 5, 251-271.

APPENDIX 1. KALMAN FILTER, PARAMETERS' ESTIMATION, AND PERFORMANCE CRITERIA

The Kalman filters are powerful tools, which can be used for models estimation in many areas in finance. A Kalman filter is an interesting method when a large volume of information must be taken into account, because it is very fast. When associated with an optimization procedure, it can also be used for the estimation of the parameters, if the model relies on non-observable data.

The simple Kalman filter is the most common version of the Kalman filter. It can be used when the model is linear, as is the case for Schwartz's model. First, the main principles of the method are presented. Second, the parameters estimation is exposed.

1. Presentation[10]

The main principle of the Kalman filter is to use temporal series of observable variables in order to reconstitute the value of non-observable variables. The model has to be expressed in a state-space form characterized by a transition equation and a measurement equation. Once this has been made, a three step iteration process can begin.

The state-space form model, in the simple filter, is characterized by the following equations:

- Transition equation: $$\alpha_{t/t-1} = T\alpha_{t-1} + c + R\eta_t \qquad [4]$$

where α_t is the m-dimensional vector of non-observable variables at t, also called state vector, T is a matrix (m × m), c is an m-dimensional vector, and R is (m × m)

- Measurement equation: $$y_{t/t-1} = Z\alpha_{t/t-1} + d + \varepsilon_t \qquad [5]$$

where $y_{t/t-1}$ is an N-dimensional temporal series, Z is a (N×m) matrix, and d is an N-dimensional vector.

[10] Harvey (1989) inspires this presentation. For a more detailed presentation, see for example Lautier and Galli (2002,a), or Javaheri, Lautier and Galli (2003).

η_t and ε_t are white noises whose dimensions are respectively m and N. They are supposed to be normally distributed, with zero mean and with Q and H as covariance matrices:

$$E[\eta_t] = 0 \text{ and } Var[\eta_t] = Q$$

$$E[\varepsilon_t] = 0 \text{ and } Var[\varepsilon_t] = H$$

The initial value of the system is supposed to be normal, with mean and variance:

$$E[\alpha_0] = \widetilde{\alpha}_0 ,\ Var[\alpha_0] = P_0$$

If $\widetilde{\alpha}_t$ is a non biased estimator of α_t, conditionally to the information available in t, then:

$$E_t[\alpha_t - \widetilde{\alpha}_t] = 0$$

Consequently, the following expression[11] defines the covariance matrix P_t :

$$P_t = E_t[(\widetilde{\alpha}_t - \alpha_t)(\widetilde{\alpha}_t - \alpha_t)']$$

During the iteration, three steps are successively tackled: prediction, innovation and updating.

• Prediction:

$$\begin{cases} \widetilde{\alpha}_{t/t-1} = T\widetilde{\alpha}_{t-1} + c \\ P_{t/t-1} = T\ P_{t-1}\ T' + R\ Q\ R' \end{cases}$$

where $\widetilde{\alpha}_{t/t-1}$ and $P_{t/t-1}$ are the best estimators of $\alpha_{t/t-1}$ and $P_{t/t-1}$, conditionally to the information available at *(t-1)*.

• Innovation:

$$\begin{cases} \widetilde{y}_{t/t-1} = Z\widetilde{\alpha}_{t/t-1} + d \\ v_t = y_t - \widetilde{y}_{t/t-1} \\ F_t = ZP_{t/t-1}Z' + H \end{cases}$$

where $\widetilde{y}_{t/t-1}$ is the estimator of the observation y_t conditionally to the information available at *(t-1)*, and v_t is the innovation process, with F_t as a covariance matrix.

• Updating:

$$\begin{cases} \widetilde{\alpha}_t = \widetilde{\alpha}_{t/t-1} + P_{t/t-1}Z' F_t^{-1} v_t \\ P_t = (I - P_{t/t-1}Z' F_t^{-1} Z) P_{t/t-1} \end{cases}$$

[11] $(\widetilde{\alpha}_t - \alpha_t)'$ is the transposed matrix of $(\widetilde{\alpha}_t - \alpha_t)$.

The matrices *T, c, R, Z, d, Q,* and *H* are the system matrices associated with the state-space model. They are not time dependent in the case considered in this article.

2. Parameters' estimation

When the Kalman filter is applied to term structure models of commodity prices, the aim is the estimation of the parameters of the measurement equation, in order to obtain estimated futures prices for different maturities $\widetilde{F}(\tau_i)$, and to compare them with empirical futures prices $F(\tau_i)$. The closest the firsts are with the seconds, the best is the model.

Suppose that the non-observable variables and the errors are normally distributed. Then the Kalman filter can be used to estimate the model parameters, which are supposed to be constant. On that purpose, the logarithm of the likelihood function is computed for the innovation v_t, for given iteration and parameters vector:

$$\log l(t) = -\left(\frac{n}{2}\right) \times \ln(2\Pi) - \frac{1}{2}\ln(dF_t) - \frac{1}{2} v_t' \times F_t^{-1} \times v_t$$

where F_t is the covariance matrix associated with the innovation v_t, and dF_t its determinant[12].

Relying on the hypothesis that the model measurement equation admits continuous partial derivatives of first and second order on the parameters, another recursive procedure is used to estimate the parameters[13]. An initial M-dimensional vector of parameters is first used to compute the innovations and the logarithms of the likelihood function. Then the iterative procedure researches the parameters vector that maximizes the likelihood function and minimizes the innovations. Once this optimal vector is obtained, the Kalman filter is used, for the last time, to reconstitute the non-observable variables and the measure $\widetilde{y}$ associated with these parameters.

[12] The value of *logl(t)* is corrected when dF_t is equal to zero.

[13] The optimization algorithm used is the one proposed by Broyden, Fletcher, Goldfarb and Shanno (BFGS).

3. Application to Schwartz's model

The simple Kalman filter can be applied to Schwartz's model because the later can be easily expressed on a linear form, as follows:

$$\ln\left(F(S,C,t,T)\right) = \ln\left(S(t)\right) - C(t) \times \frac{1 - e^{-\kappa\tau}}{\kappa} + B(\tau)$$

Letting $G = ln(S)$, we also have:

$$\begin{cases} dG = (\mu - C - \frac{1}{2}\sigma_S^2)dt + \sigma_S dz_S \\ dC = \left[k(\alpha - C)\right]dt + \sigma_C dz_C \end{cases}$$

The transition equation is:

$$\begin{bmatrix} \tilde{G}_{t/t-1} \\ \tilde{C}_{t/t-1} \end{bmatrix} = c + T \times \begin{bmatrix} \tilde{G}_{t-1} \\ \tilde{C}_{t-1} \end{bmatrix} + R\eta_t\,, \quad \text{t = 1, ... NT}$$

where:

- $c = \begin{bmatrix} \left(\mu - \frac{1}{2}\sigma_S^2\right)\Delta t \\ \kappa\alpha\Delta t \end{bmatrix}$ is a (2×1) vector, and Δt is the period separating 2 observation dates
- $T = \begin{bmatrix} 1 & -\Delta t \\ 0 & 1 - \kappa\Delta t \end{bmatrix}$ is a (2 × 2) matrix,
- R is the identity matrix, (2 × 2),
- η_t are uncorrelated errors, with :

$$E[\eta_t] = 0, \text{ and } \qquad Q = Var[\eta_t] = \begin{bmatrix} \sigma_S^2\Delta t & \rho\sigma_S\sigma_C\Delta t \\ \rho\sigma_S\sigma_C\Delta t & \sigma_C^2\Delta t \end{bmatrix}$$

The measurement equation is:

$$\tilde{y}_{t/t-1} = d + Z \times \begin{bmatrix} \tilde{G}_{t/t-1} \\ \tilde{C}_{t/t-1} \end{bmatrix} + \varepsilon_t\,, \quad \text{t = 1, ... NT}$$

where:

- the i[th] line of the N dimensional vector of the observable variables $\tilde{y}_{t/t-1}$ is $\left[\ln\left(\tilde{F}(\tau_i)\right)\right]$, with i = 1,..,N, where N is the number of maturities retained for the estimation.
- $d = [B(\tau_i)]$ is the i[th] line of the d vector, with i = 1,..., N
- $Z = \left[1 \quad , \quad -\dfrac{1-e^{-\kappa\tau_i}}{\kappa}\right]$ is the i[th] line of the Z matrix, which is (N×2), with i = 1,...,N
- ε_t is a white noise's vector, (N×1), with no serial correlation : $E[\varepsilon_t] = 0$ and $H = Var[\varepsilon_t]$. H is (N × N)

4. Performance criteria

To assess the model's performances, two criteria were retained: the mean pricing error and the root mean squared error.

The mean pricing error (MPE) is defined as follows:

$$MPE = \frac{1}{N}\sum_{n=1}^{N}\left(\tilde{F}(n,\tau) - F(n,\tau)\right)$$

where N is the number of observations, $\tilde{F}(n,\tau)$ is the estimated futures price for a maturity τ at date n and $F(n,\tau)$ is the observed futures price. The MPE is expressed in US dollar. It measures the estimation bias for one given maturity. When the estimation is good, the MPE is close to zero.

Retaining the same notations, the root mean squared error (RMSE) is also expressed in US dollar and is defined in the following way, for one given maturity τ:

$$RMSE = \sqrt{\frac{1}{N}\sum_{n=1}^{N}\left(\tilde{F}(n,\tau) - F(n,\tau)\right)^2}$$

When there is no bias, the RMSE can be considered as an empirical variance. It measures the estimations stability. This second criterion is considered as more representative because prices errors can offset themselves and the MPE can be low even if there are strong deviations.

APPENDIX 2. THE MODEL PERFORMANCES FOR ALL THE MATURITIES

	1-2-3-4		**1-3-6-9**		**1-5-10-15**		**1-6-12-18**		**1-12-24-48**		**1-24-48-84**		**48-60-72-84**	
Maturity	MPE	RMSE	MPE	RMSE	MPE	RMSE	MPE	RMSE	MPE	RMSE	MPE	RMSE	MPE	RMSE
1 month	-0,2160	2,2207	0,0961	2,1769	0,1048	1,9196	-0,5356	2,3705	0,2025	2,2266	0,1419	1,9311	0,0685	2,7829
2 months	-0,1676	2,0503	-0,0102	1,9847	-0,0107	1,7862	-0,5391	2,1169	0,1124	1,9223	0,0156	1,6889	-0,0133	2,4780
3 months	-0,1592	1,9154	-0,0951	1,8438	-0,1044	1,6646	-0,5336	1,9033	0,0378	1,6740	-0,0828	1,4987	-0,0982	2,2194
4 months	-0,1519	1,7992	-0,1316	1,7314	-0,1484	1,5476	-0,4892	1,7083	0,0079	1,4767	-0,1260	1,3544	-0,1554	2,0079
5 months	-0,0208	1,8062	-0,0216	1,7311	0,0332	1,5436	-0,6161	1,8228	-0,0428	1,5843	0,0007	1,4143	-0,3180	1,8908
6 months	-0,0169	1,6871	-0,0046	1,6283	0,0309	1,4307	-0,5247	1,6195	-0,0305	1,4085	-0,0489	1,2776	-0,2795	1,7063
7 months	-0,0504	1,5846	0,0077	1,5468	0,0284	1,3435	-0,4403	1,4476	-0,0201	1,2667	-0,0890	1,1776	-0,2573	1,5529
8 months	-0,1169	1,4979	0,0130	1,4807	0,0233	1,2763	-0,3645	1,3033	-0,0136	1,1542	-0,1231	1,1077	-0,2522	1,4261
9 months	0,6591	1,7346	0,0165	1,4975	-0,0105	1,3295	-0,6537	1,5074	-0,1095	1,2766	-0,1328	1,1759	0,8705	1,6755
10 months	0,5391	1,6043	0,0236	1,4304	-0,0123	1,2612	-0,5641	1,3577	-0,0865	1,1646	-0,1469	1,1007	0,8404	1,5698
11 months	0,4026	1,4804	0,0262	1,3702	-0,0134	1,2030	-0,4783	1,2239	-0,0633	1,0695	-0,1540	1,0399	0,8011	1,4706
12 months	0,2538	1,3699	0,0255	1,3167	-0,0132	1,1540	-0,3948	1,1063	-0,0387	0,9911	-0,1539	0,9917	0,7545	1,3787
13 months	1,1760	1,8505	0,1517	1,3510	0,0134	1,2136	-0,6442	1,3158	-0,1583	1,1011	-0,2152	1,0709	1,8015	2,2333
14 months	0,9826	1,6709	0,1557	1,2917	0,0284	1,1533	-0,5417	1,1806	-0,1154	1,0134	-0,1948	1,0068	1,7248	2,1317
15 months	0,7799	1,5012	0,1503	1,2376	0,0382	1,1005	-0,4458	1,0608	-0,0759	0,9387	-0,1744	0,9515	1,6393	2,0283
16 months	0,5674	1,3494	0,1339	1,1909	0,0413	1,0571	-0,3578	0,9596	-0,0415	0,8790	-0,1559	0,9075	1,5439	1,9222
17 months	1,5711	2,0636	0,3785	1,3113	0,0937	1,1466	-0,5842	1,1599	-0,1367	0,9918	-0,2327	1,0041	2,5152	2,8823
18 months	1,3219	1,8426	0,3471	1,2593	0,0996	1,1025	-0,4889	1,0546	-0,0968	0,9289	-0,2004	0,9521	2,3927	2,7494
19 months	1,0648	1,6280	0,3040	1,2058	0,0960	1,0591	-0,4027	0,9607	-0,0636	0,8717	-0,1736	0,9041	2,2608	2,6081
20 months	0,8041	1,4296	0,2531	1,1535	0,0868	1,0177	-0,3218	0,8768	-0,0332	0,8205	-0,1488	0,8600	2,1234	2,4629
21 months	1,7668	2,1678	0,2969	1,2135	-0,0147	1,0483	-0,6652	1,1911	-0,1317	0,8949	-0,2747	0,9112	3,0253	3,3939
22 months	1,4796	1,9127	0,2298	1,1586	-0,0341	1,0091	-0,5886	1,1149	-0,1022	0,8439	-0,2390	0,8664	2,8695	3,2293
23 months	1,1748	1,6550	0,1405	1,0992	-0,0744	0,9687	-0,5325	1,0568	-0,0911	0,8000	-0,2224	0,8289	2,6937	3,0408
24 months	0,9027	1,4503	0,0793	1,0651	-0,0855	0,9433	-0,4468	0,9893	-0,0486	0,7632	-0,1753	0,7943	2,5481	2,8956
25 months	1,9020	2,2914	0,6767	1,2781	0,2087	0,9953	-0,5405	0,9562	-0,1603	0,8478	-0,2566	0,8861	3,3981	3,7883
26 months	1,5959	2,0222	0,5900	1,2086	0,1832	0,9608	-0,4633	0,8808	-0,1284	0,8082	-0,2229	0,8487	3,2278	3,6101
27 months	1,2941	1,7700	0,5025	1,1429	0,1571	0,9282	-0,3859	0,8110	-0,0945	0,7724	-0,1883	0,8142	3,0590	3,4347
28 months	0,9954	1,5397	0,4131	1,0811	0,1293	0,8971	-0,3095	0,7485	-0,0598	0,7404	-0,1539	0,7827	2,8906	3,2610
48 months	66,1569	67,0793	23,7877	24,1712	10,1052	10,3425	-2,6051	2,9408	-1,4159	1,7768	-0,0196	1,0849	-0,2572	0,8673
60 months	50,2521	51,0392	20,0129	20,3778	8,6222	8,8483	-2,2986	2,6083	-1,2274	1,5405	0,0971	0,9218	-0,2246	0,8291
72 months	37,2891	37,9790	16,5183	16,8774	7,2011	7,4291	-1,8768	2,1826	-0,8957	1,2133	0,1595	0,7993	-0,1659	0,8280
84 months	26,6402	27,2612	13,2121	13,5685	5,7366	5,9723	-1,4749	1,7824	-0,5717	0,9324	0,0862	0,7031	-0,1400	0,8883