COMMODITY MODELS.

A STEP FORWARD

FRANCISCO JAVIER POBLACIÓN GARCÍA(*)

DOCTORAL THESIS

PhD IN QUANTITATIVE FINANCE AND BANKING

UNIVERSIDAD DE CASTILLA-LA MANCHA

DEPARTAMENTO DE ANÁLISIS ECONÓMICO Y FINANZAS

ADVISORS:

GREGORIO SERNA AND ÁNGEL LEÓN

DECEMBER 2008

(*) This doctoral thesis is sole responsibility of its authors.

To my parents, grandparents, brothers and sister.

"Now as touching things offered unto idols, we know that we all have knowledge. Knowledge puffeth up, but charity edifieth. If any man thinketh that he knoweth anything, he knoweth not yet as he ought to know; but if any man loveth God, the same is known by him", 1Co 8:1-3

PREFACE

The present doctoral thesis has been carried out during several years in which the author has had the pleasure to deal with some people and institutions. These lines try to be an acknowledgement to them.

First of all, the author wishes to grateful his family for the support and patience provided by them.

To his advisors who have given him the opportunity to work in theses exciting issues and for their help, availability and patience provided which has been essential to combine the thesis with his work.

To Andrés García Mirantes, one of the smartest person that the author has ever met. Without Andrés help this thesis has been never performed.

He also thanks his formers partners in Repsol YPF (Andrés Ubierna, Gabriel Paulik, Alicia Mateo and many others more) for their help in discussing many issues and in data collecting. The author especially thanks Juan Manuel Martín Prieto for his help. JuanMa is the person who more teach the author in this thesis subjects and many of the ideas included in this thesis come from never-ending discussions with JuanMa. The author also thanks Repsol YPF for its support.

His recognition to participants in many conferences and meeting for their helpful comments.

The author also acknowledges the generous financial support provided by Junta de Castilla-La Mancha, grant PAI05-074.

Any errors that remain are, however, entirely the author' own.

INDEX

INTRODUCTION

In this section we are going to exhibit some fundamentals about futures contract valuation in commodity markets which are going to be used in the whole doctoral thesis. These fundamentals consist on a general methodology which can be used in all kind of problems, is much simpler than the ad-hoc solutions presented in the literature that can only be used in the concrete problem for which they were developed, and avoids approximations.

1. Futures Contract Valuation

Most of the models proposed in the literature assume that the dynamics of a commodity price (or its log) is given by a linear stochastic differential system:

$$\begin{cases} dX_t = (b + AX_t)dt + RdW_t \\ Y_t \quad = cX_t \end{cases}$$

where Y_t is the commodity price (or its log), *b*, *A*, *R* and *c* are deterministic parameters independent of *t* ($b \in \Re^n, A, R \in \Re^{n\,x\,n},\ c \in \Re^n$) and W_t is a n-dimensional canonical Brownian motion (i.e. all components uncorrelated and its variance equal to unity).

It is easy to prove that the solution of that problem is:

$$X_t = e^{At}\left[X_0 + \int_0^t e^{-As} bds + \int_0^t e^{-As} RdW_s\right] \qquad \text{(I1)}$$

and this solution is unique (Oksendal, 1992). Moreover, even in the case that *b*, *A* and *R* were functions of *t*, if A_t and $\int_0^t A_s ds$ commute, the solution of that problem is (I1).

Accordingly, X_t is normally distributed with mean and variance:

$$E[X_t] = e^{At}\left[X_0 + \int_0^t e^{-As}\, bds\right] \qquad Var[X_t] = e^{At}\left[\int_0^t e^{-As} RR' e^{-As'} ds\right] e^{At'}$$

With this result it is easy to prove that the price of a futures contract traded at time t with maturity at time $t+T$, $F_{t,T}$, can be computed as:

$$F_{t,T} = \exp\left[ce^{AT}X_t + g(T)\right] \tag{I2}$$

where g is a, probably complicated, deterministic function.

2. Volatility of Futures Returns

The squared volatility of a futures contract traded at time t with maturity at time $t+T$ can be defined as: $\lim_{h\to 0}\frac{Var\left[\log F_{t+h,T} - \log F_{t,T}\right]}{h}$ (it is also possible to define it as $\lim_{h\to 0}\frac{Var\left[\log F_{t+h,T-h} - \log F_{t,T}\right]}{h}$). It can be proved that it is the expected value of the square of the coefficient of the Brownian motion (σ_t) in the expansion $d\log\left(F_{t,T}\right) = \mu_t ds + \sigma_t dW_t^F$, where W_t^F is a scalar canonical Brownian motion.

Hence, taking logarithms and differentials on both sides of Equation (I2), it follows that: $d\left(\log F_{t,T}\right) = ce^{AT}dX_t = ce^{AT}\left[b + AX_t\right]dt + ce^{AT}R\,dW_t$

Therefore, the squared volatility is: $ce^{AT}RR'e^{AT}{}'c'$. (I3)

Note that R does not need to be computed as RR' is the noise covariance matrix.

3. Empirical Models

In the general model presented above, it is easy to prove that knowing X_{t-1}, X_t can be written as:

$$X_t = c_t + M_t X_{t-1} + \psi_t \tag{I4}$$

where $c_t = e^A \left[\int_{t-1}^{t} e^{-As} b^{\#} ds \right]$, $M_t = e^A$ and ψ_t is a n-vector of serially uncorrelated Gaussian disturbances with zero mean and covariance matrix: $e^A \left[\int_{t-1}^{t} e^{-As} RR^{'} e^{-As^{'}} ds \right] e^{A^{'}}$.

4. Uses and remarks

This general methodology can be used in all kind of problems, is much simpler than the ad-hoc solutions presented in the literature that can only be used in the concrete problem for which they were developed, needing complex procedures like limit steps (Schwartz and Smith, 2000) or partial differential equations (Schwartz, 1997), and avoids approximations like in Schwartz (1997).

5. The doctoral thesis

As it is said above, due to the complexity of commodity prices dynamics, valuation of commodity contingent claims is carried out in the extant literature via ad-hoc solutions, which are very complex and sometimes include approximations. That is, owing to the cost of carry, the commodity prices dynamics is very complex. Therefore, it is needed to continue deepening in its study. This doctoral thesis tries to go an step forward in this sense.

This doctoral thesis is organized as follows. Chapter 1 contains an study about the extant commodity models and their implications in investment under uncertainty. Investment projects involving commodities typically require a large amount of capital, last many years and include clauses which can be interpreted as call and put options. Therefore, the price dynamics behaviour assumed for the commodity price is essential in valuing these investment projects. In this chapter it is analysed the optimal contract determination assuming several models proposed in the literature for the commodity price dynamics.

In chapter 2 it is proved that seasonality in some commodities (natural gas, gasoline and heating oil) is an stochastic factor instead of a deterministic one. This chapter proposes a general (n+2m)-factor model for the stochastic behavior of commodity prices, considering seasonality as an stochastic factor, with n non-seasonal factors, described in the literature, and m seasonal factors. The model, particularized for n = 1, 2, 3 and m = 1, has been applied to Henry Hub natural gas futures contracts traded at NYMEX. Similar results are obtained with other commodities traded at NYMEX and with commodities traded in other markets (ICE Futures Europe).

Chapter 3 provides evidence that crude oil and the main refining products are not only cointegrated but also have a common long-term trend. In this chapter there is definitive evidence of a common long-term trend for crude oil prices and the most important refining products prices, i.e. gasoline and heating oil, traded at NYMEX. These three commodities are not only cointegrated, but they have also a common long-term dynamics. We present definitive evidence of this fact by proposing different factor models to explain the dynamics of commodity prices jointly. These results are used to value the crack-spread options quoted at NYMEX, given that the most suitable way to value these options is assuming a common long-term dynamics for crude oil and refined products prices.

REFERENCES

- Oksendal, B. (1992) ‘Stochastic Differential Equations. An Introduction with Applications’, 3rd ed. Springer-Verlag, Berlin Heidelberg.
- Schwartz, E.S. (1997). The stochastic behavior of commodity prices: Implication for valuation and hedging. The 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.

CHAPTER 1: COMMODITY MODELS AND INVESTMENT UNDER UNCERTAINTY. THE OPTIMAL CONTRACT DETERMINATION

1. Introduction

One of the classical financial theory assumptions is the fact that the log-spot price (p_t) of a financial asset follows a random walk with drift: $p_t = r + \lambda + p_{t-1} + \varepsilon_t$, where r is the interest rate, λ is the risk premium and ε_t is an independent and identically distributed in t (iid) random noise, which is distributed following a Gaussian distribution with zero mean and variance $\sigma^2 \Delta t$ for each time "t" (σ is the annualized volatility). That is, except for error terms, future prices growth with a constant rate which is the sum of the interest rate and the risk-premium.

Due to the cost-of-carry, this assumption is not reasonable in the case of commodities. These assets present strong mean reversion, as can be appreciated in Figure 1. The natural way to extend the random walk to incorporate the mean reverting effect is through an AR(1) model. In recent years several authors have proposed more sophisticated models, in which the commodity price is assumed to be the sum of several factors. Specifically, we review the two-factor model by Schwartz and Smith (2000) and Schwartz (1997). In this model the commodity price is the sum of a long-term factor, which evolves according to a geometric Brownian motion, and a short-term factor, which evolves according to an Ornstein-Uhlenbeck process. More recently Cortazar and Schwartz (2003) have proposed an extension of the two factors model, allowing the spot price long term return to be stochastic. A generalization of this kind of models has been proposed by Cortazar and Naranjo (2006).

An alternative procedure has been proposed by Clewlow and Stricland (2000). These authors try to model the futures price dynamics directly, by assuming that the futures prices are a martingale, where the price of a futures contract at time t is its price at time t-1 plus an innovation which is given by the sum of several deterministic volatility functions. One possible way to choose these deterministic volatility functions is through a principal component analysis.

Several authors have tried to incorporate in the model the seasonal effects commonly observed in many commodity prices. Sorensen (2002) has tried to incorporate these effects in a deterministic way, whereas Chapter 2 includes seasonality as an additional stochastic factor in the model.

This chapter presents these models in an unified context, analyzing the relationships among them, and pointing out their advantages and limitations. Moreover, it is analyzed their relative performance with a common data set. Specifically, the data set is composed of three commodity prices series: WTI crude oil and natural gas futures contracts traded at NYMEX and Brent crude oil futures contracts traded at ICE Futures in London.

This comparative study between models has critical importance in investment under uncertainty. Concretely, when a company is planning to develop a crude oil or natural gas field, the investment is huge (it usually reaches thousands million dollars) and usually lasts many years (in fact they usually last between twenty and thirty years), however the main investment has to be made at the beginning, before getting any return. Consequently, the company needs a sell contract, which should last at least twenty years, to guarantee the investment recovery. Typically these contracts contain clauses with a minimum price to guarantee the seller's investment recovery, and a maximum price to protect the buyer from unexpected and steep price increases. It is easy to demonstrate that these clauses can be seen as put and call options and, therefore, the stochastic behaviour of commodity prices plays a crucial role in option valuation problems. As the volatility is a decisive parameter in the option valuation methodology, it is crucial to choose an appropriate model, with certain volatility assumptions, to characterize the commodity price dynamics.

This chapter is organized as follows. The comparative study between models is contained in section 2. Section 3 deals with investment under uncertainty and finally section 4 concludes with a summary and discussion.

2. Commodity Models

2.1. The AR(1) Model

Taking into account the arguments in the previous section, the natural extension of the random walk to incorporate in the model the mean reversion effects is the AR(1) model. The AR(1) model assumes that the commodity log-spot price (p_t) follows the following process:

$$\mathrm{p}_\mathrm{t} - c = \rho(\mathrm{p}_\mathrm{t-1} - c) + \varepsilon_t$$

where c is the long-term mean (the log-price converges to c in the long-tem), ρ is the reversion speed to this long-term mean and, as before, ε_t is an independent and identically distributed in t (iid) random noise, which is distributed following a Gaussian distribution with zero mean and variance $\sigma^2\Delta t$ for each time "t".

Therefore, under this model, at time "t", p_{t+1} is a random variable which is distributed N($c + \rho p_t - c$), $\sigma^2\Delta t$). Explicitly, the best prediction of the log-spot price in t+1 is $c + \rho(p_t - c)$ with precision given by $\sigma\sqrt{\Delta t}$ (short-term volatility). It is easy to see that p_{t+j}, with a high "j", is a random variable which is distributed N(c, $\sigma^2/(1-\rho^2)\Delta t$). Explicitly, the best prediction of the log-spot price in $t+j$ is c with precision given by $\sqrt{\sigma^2\Delta t/(1-\rho^2)}$ (long-term volatility). Consequently this model assumes that the volatility is bounded.

The annualized volatility of futures returns is: $\sigma(\ln F_{T,t}) = \rho^{T-t}\sigma$.

This model will be estimated with three commodity prices series: WTI and Brent crude oil and Henry Hub natural gas. Currently, however, there are not spot prices for these three commodities. Consequently in this work we use one month futures contracts quoted at NYMEX in the case of WTI crude oil and Henry Hub natural gas, and one month futures contracts quoted at ICE in the case of Brent crude oil. Hence, the data set to calculate these parameters consists on weekly observations of one month futures contracts from 4/2/1990 to 3/24/2008 for Henry

Hub natural gas, from 6/27/1988 to 3/24/2008 for Brent crude oil and from 1/1/1985 to 3/24/2008 for WTI crude oil. In the case of Henry Hub natural gas the deterministic seasonal component is removed. Table 1 contains some descriptive statistics of the data. The model parameter estimates for the three commodities: Brent and WTI crude oil and Henry Hub natural gas for the whole sample period are contained in Table 2.

As can be appreciated in Table 2, the reversion speed to the long-term mean (ρ) is equal to one in all cases and the long-term mean (c) is statistically non-different from zero, which means that these prices do not follow a pure mean reverting process. In Figure 1 it is possible to appreciate that there is a mean reversion effect in commodity prices until 1999, afterwards commodity prices exhibit a random walk behaviour, which is the dominating effect in the previous estimates. These results are coherent with the standard deviations observed in Table 1, which are too high for a mean reverting process, and they are also coherent with Cortazar and Naranjo (2006) findings.

Therefore, taking into account these results, we will assume that there is a structural break in 1999[1]. Specifically, the mean reverting behaviour can be appreciated if we select a data set which consists on weekly observations of one month futures contracts from 4/2/1990 to 12/27/1999 for Henry Hub natural gas, from 6/27/1988 to 12/27/1999 for Brent crude oil and from 1/1/1985 to 12/27/1999 for WTI crude oil. Some descriptive statistics for this data set are presented in Table 1. The model parameter estimates for the three commodities with this new data set are contained in Table 2.

In this case the reversion speed to the long-term mean (ρ) is not equal to one (although it is quite close to one), and the long-term mean (c) is statistically different from zero, which means that in this sample there are mean reverting effects.

The lack of economical transportation and the limited storability of natural gas make its supply unable to change in view of variations of demand. This is the reason why natural gas prices are more volatile (i.e. they have higher volatility, σ) and less mean reverting (mean reversion is

[1] Similar results are obtained if we choose 1998 or 2000 as breaking point.

higher because ρ is lower) than crude oil ones and it also explains why natural gas prices are strongly seasonal. Therefore, as we do not take into account seasonality in he model, the goodness of fit (R^2) is worse for natural gas prices than for crude oil ones, in spite of the deterministic seasonal component has been removed.

However, this model is too much simple and, as can be appreciated in Figures 2, 3 and 4, it does not estimate the volatility of futures returns properly.

Therefore, we can conclude that these one-factor mean-reverting models, like the AR(1), are not very realistic since they generate a volatility of futures returns which goes to zero as the time to maturity of the futures contract approaches infinity. As can be appreciated in the previous charts, the empirical volatility of futures returns does not go to zero when time to maturity goes to infinity. Even more, models considering a single source of uncertainty are not very realistic since they imply that futures prices for different maturities should be perfectly correlated, which defies existing evidence. In the following sections of this work we are going to present more complex models in which volatility of futures returns is better characterized.

2.2. The Two-Factor Model

The model that we are going to present now is the two-factor model introduced by Schwartz and Smith (2000), which is equivalent to the one proposed by Schwartz (1997). In this model it is assumed that the log-spot price is the sum of two components (or factors): one short term factor (χ_t) which follows an Ornstein-Uhlenbeck process and one long term factor (ε_t) which follows a standard Brownian motion. Writing the model in its discrete-time version we have that:

$$X_t = c + QX_{t-1} + \eta_t \qquad \text{where:}$$

$$X'_t = [\chi_t, \quad \varepsilon_t]; \;\; c' = [0, \quad \mu_\varepsilon \Delta t]; \quad Q = \begin{bmatrix} \exp(-k\Delta t) & 0 \\ 0 & 1 \end{bmatrix};$$

$$Var[\eta_t] = \begin{bmatrix} (1-\exp(-2k\Delta t))\dfrac{\sigma_\chi{}^2}{2k} & (1-\exp(-k\Delta t))\dfrac{\sigma_\chi \sigma_\varepsilon \rho_{\chi\varepsilon}}{k} \\ (1-\exp(-k\Delta t))\dfrac{\sigma_\chi \sigma_\varepsilon \rho_{\chi\varepsilon}}{k} & \sigma_\varepsilon{}^2 \Delta t \end{bmatrix};$$

and η_t independent of X_{t-1}.

In the appendix it is proved that the AR(1) model is a particular case of the two-factor one when the long-term factor is considered deterministic.

In the same way as before, it is possible to demonstrate that, under this model assumptions, at time t, p_{t+1} is a random variable and the best prediction of the log-spot price in t+1 is $\mu_\varepsilon\Delta t$ + exp(-$k\Delta t$)χ_t, with precision given by the following expression (short-term volatility):

$$\sqrt{((1-\exp(-2\mathrm{k}\Delta t))\sigma_\chi{}^2/(2k)+\sigma_\varepsilon{}^2\Delta t+2(1-\exp(-\mathrm{k}\Delta t))\sigma_\chi\sigma_\varepsilon\rho_{\chi\varepsilon}/k)}$$

In the same way, it is easy to see that the long-term volatility is $\sigma_\varepsilon\sqrt{t}$. Therefore, this model assumes that the volatility grows with time, consequently it is not bounded. This implies that the volatility of futures returns does not go to zero when time to maturity goes to infinity, which it is a desirable property.

The annualized volatility of futures returns is: $\sigma(\ln F_{T,t})=e^{-2kT}\sigma_\chi{}^2+\sigma_\varepsilon{}^2+2e^{-kT}\sigma_\chi\sigma_\varepsilon\rho_{\chi\varepsilon}$.

As there is not market quotation for the factors in which the spot price of the three commodities can be decomposed, the estimation has been performed using the Kalman filter methodology (see, for example Harvey, 1989)[2]. The data set employed in the estimation procedure consists on weekly observations of Henry Hub natural gas and WTI crude oil futures prices traded at NYMEX and Brent crude oil futures prices traded at ICE. The data set for Henry Hub natural gas is made of contracts F1, F5, F9, F13, F17, F21, F25, F29, F33, F37, F41, F44 and F48 where F1 is the contract closest to maturity, F2 is the second contract closest to maturity and so on. This data set contains 330 quotations of each contract from 12/03/2001 to 03/24/2008. The data set for WTI crude oil is made of contracts F1, F4, F7, F10, F13, F16, F19, F22, F25 and F28. This data set contains 654 quotations of each contract from 9/18/1995 to 03/24/2008. The data set for Brent crude oil is made of contracts F1, F4, F7, F10, F12, F16-18, F22-24 and F31-

[2] As explained in Schwartz and Smith (2000), the risk-neutral version of the model is necessary to estimate the parameters. This is the reason why there is a risk-premium for the short-term deviation (λ_χ) and also a risk-adjusted long-term drift (μ_ξ*).

36. This data set contains 537 quotations of each contract from 12/15/1997 to 03/24/2008. Table 3 contains some descriptive statistics of these data sets. The model parameter estimates for the three commodities are presented in Table 4.

The results are coherent with the ones obtained by Schwartz and Smith (2000). As in the previous section, comparing this estimation with the Schwartz and Smith (2000) one, which use data until 1995, we find that with our data (until 2008) the mean-reversion effect is lower than in the case of data until 1995. There is also less correlation between factors than in Schwartz and Smith (2000).

In this case, as the model accounts for long-term effects (i.e. random walk effects), there are not the problems described in the previous section. Even more, as can be appreciated in Figures 2, 3 and 4, this model estimates the volatility of futures returns properly, in a more accurate way than the previous one. As before, volatility and mean reversion (k) are higher in natural gas prices.

The in-sample predictive power ability of the two-factor model can be analyzed through the bias (real minus predicted prices) and the root mean squared error, which are shown in Table 5. As can be appreciated in the table, the model goodness of fit is worse in natural gas than in crude oil because we do not take into account seasonal effects in model specification.

2.3. The Three-Factor Model

This model was proposed by Cortazar and Schwartz (2003) and is an extension of the two-factor model presented above. The only difference is the introduction of a new risk factor: the drift, μ (i.e. the spot price long term return). Although the model formulation by the authors is slightly different from the one of the two-factor model presented above, it is proved that both formulations are equivalent (see Chapter 2).

In this work we use the same data set and estimation procedure (Kalman filter)[3] as in the two factor model case, just to compare the results between models. Even more, we are going to present the model with the same formulation than the two factor model one to make comparison easier. Therefore, it is assumed that the log-spot price is the sum of three components (or factors): two short term factors (χ_{1t} and χ_{2t}) which follow an Ornstein-Uhlenbeck process and one long term factor (ε_t) which follows a standard Brownian motion. Writing the model in its discrete-time version we have that:

$$X_t = c + QX_{t-1} + \eta_t \text{ where:}$$

$$X'_t = [\varepsilon_t, \quad \chi_{1t}, \quad \chi_{2t}]; \; c' = [\mu_\varepsilon \Delta t \quad 0 \quad 0]; \; Q = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \exp(-k_1 \Delta t) & 0 \\ 0 & 0 & \exp(-k_2 \Delta t) \end{bmatrix};$$

$$Var[\eta_t] = \begin{bmatrix} \sigma_\varepsilon^2 \Delta t & (1-\exp(-k_1\Delta t))\frac{\sigma_{\chi 1}\sigma_\varepsilon\rho_{\chi 1\varepsilon}}{k_1} & (1-\exp(-k_2\Delta t))\frac{\sigma_{\chi 2}\sigma_\varepsilon\rho_{\chi 2\varepsilon}}{k_2} \\ (1-\exp(-k_1\Delta t))\frac{\sigma_{\chi 1}\sigma_\varepsilon\rho_{\chi 1\varepsilon}}{k_1} & (1-\exp(-2k_1\Delta t))\frac{\sigma_{\chi 1}^2}{2k_1} & (1-\exp(-(k_1+k_2)\Delta t))\frac{\sigma_{\chi 1}\sigma_{\chi 2}\rho_{\chi 1\chi 2}}{k_1+k_2} \\ (1-\exp(-k_2\Delta t))\frac{\sigma_{\chi 2}\sigma_\varepsilon\rho_{\chi 2\varepsilon}}{k_2} & (1-\exp(-(k_1+k_2)\Delta t))\frac{\sigma_{\chi 1}\sigma_{\chi 2}\rho_{\chi 1\chi 2}}{k_1+k_2} & (1-\exp(-2k_2\Delta t))\frac{\sigma_{\chi 2}^2}{2k_2} \end{bmatrix}$$

and η_t independent of X_{t-1}.

The annualized volatility of futures returns is:

$$\begin{aligned}\sigma(\ln F_{T,t}) = \sigma_\varepsilon^2 + \sigma_{\chi 1}^2 e^{-2k_1 T} + \sigma_{\chi 2}^2 e^{-2k_2 T} + 2e^{-k_1 T}\sigma_\varepsilon\sigma_{\chi 1}\rho_{\xi\chi 1} + 2e^{-k_2 T}\sigma_\varepsilon\sigma_{\chi 2}\rho_{\xi\chi 2} + \\ + 2e^{-(k_1+k_2)T}\sigma_{\chi 1}\sigma_{\chi 2}\rho_{\chi 1\chi 2}\end{aligned}.$$

The model parameter estimates for the three commodities are contained in Table 6[4].

The results are consistent with those obtained in the previous section. In this model there are two short-term factors whose parameters are highly significant, and in all cases one of these factors has significantly higher speed of adjustment than the other. This means that there are two types of stochastic short-term effects, one (the one with higher *k*) with stronger mean

---

[3] Cortazar and Schwartz (2003) propose a very simple estimation procedure and apply it to an incomplete panel of oil futures prices. The methodology, however, does not make an optimal use of prices in the estimation of state variables (as opposed to the Kalman filter methodology), and is unable to obtain parameter estimation errors.

[4] As before, the risk-neutral version of the model is necessary to estimate the parameters.

reversion than the other (the one with lower *k*), and both of them significant. We can see in Table 6 that the less mean-reverting short-term factor has positive risk-premium whereas the more mean-reverting one has a negative risk-premium.

There is also high and negative correlation between short-term factors, which indicates that these short term-factors (one of them more mean-reverting than the order) usually work in opposite directions. It is also interesting to note that the long-term factor has negative correlation with the less reverting short-term factor, and positive correlation with the more mean-reverting one.

The in-sample predictive power ability of the three-factor model can be analyzed through the bias (real minus predicted prices) and the root mean squared error, which are shown in Table 7. It is also possible to compare these results with those obtained with the two-factor model (Table 5). As expected, the root mean squared errors obtained with the three-factor model are lower than those obtained with the two-factors one.

As before, volatility and mean reversion are higher in natural gas prices whereas goodness of fit is worse.

The relative performance of the two and three-factor models can also be analyzed through the Schwartz and Akaike Information Criteria (SIC and AIC respectively). If we define the Schwartz Information Criterion (SIC) as $\ln(L_{ML}) - q\ln(T)$, where q is the number of estimated parameters, T is the number of observations and L_{ML} is the value of the likelihood function, using the q estimated parameters, then the higher the SIC the better the fit. The same conclusions are obtained with the Akaike Information Criterion (AIC), which is defined as $\ln(L_{ML}) - 2q$. As expected, the values of both measures are higher with the three-factor model (Table 6) than with the two-factor one (Table 4).

The volatilities of futures returns for the three commodities are depicted in Figures 2, 3 and 4. It is easy to appreciate in the charts that with the three-factor model the volatility of futures returns fits better than with the two-factor one.

Finally, we can conclude that since the three-factor model has more structure (more factors and parameters), the goodness of fit is better than in the previous case. Therefore, in each case we have to decide between both models (the two-factor and the three factor model) taking into account that although the three-factor model fits better to the data, the two-factor one is simpler and therefore it is easier to estimate, the significance of each stochastic factor is more clear, needs less data to be estimated… Consequently in each case, depending on its characteristics, we will choose between the simplicity (the two factor model) and the goodness of fit (the three factor model).

Cortazar and Naranjo (2006) proved that, for WTI crude oil data, the three factor model is well enough and there is little improvement in adding more factors and defining a four factor model.

2.4. The Principal Components Model

In previous subsections we have presented models in which the assumptions were made on the spot price dynamics. Known the spot price dynamics, we derived the futures price dynamics. The following model, introduced by Clewlow and Stricland (2000), models the futures price dynamics directly.

This model departs from the fact that the best prediction in t for the price of a futures contract maturing in T is the futures price in t. (i.e., the futures price is a martingale). Therefore, the model assumes that, today (at time t), the futures price maturing in T follows the following process: $F_{T,t+1} = F_{T,t} + F_{T,t}\left[\sum_{i=1}^{N}\sigma_i(t,T)\varepsilon_{it}\right]$, where $\sigma_i(t,T)$ are deterministic volatility functions, ε_{it} are orthogonal iid random noises which are distributed following a Gaussian distribution with zero mean and variance Δt for each time "t"[5].

One possible way to choose N and the $\sigma_i(t,T)$ functions is through a principal components analysis. The "j" principal component is a vector which is defined as the squared root of the j-th

[5] If we choose in a proper way N and $\sigma_i\,(t,T)$, we get the factors models (AR(1) model, two-factor model and three-factor model) presented above. Consequently this model can be understood as a generalization of the previous ones.

highest eigenvalue times its eigenvector. It is easy to prove that the matrix which has the principal components in their columns times itself transposed is the covariance matrix Σ.

The data set to implement the principal components model is the same as the one used to estimate the two and three-factor models, which is described in Table 3. Specifically, for Henry Hub natural gas the data set is made of weekly observations of contracts F1 to F48 from 12/03/2001 to 03/24/2008, for WTI crude oil it is made of weekly observations of contracts F1 to F28 from 9/18/1995 to 03/24/2008, and for Brent crude oil it is made of weekly observations of contracts F1 to F31-36 from 12/15/1997 to 03/24/2008.

The first three principal components for the three commodities are depicted in Figures 5, 6 and 7. The first component is considered a long-term one as it has the same sign for all maturities and does not go to zero as maturity goes to infinity, which means that a random shock which follows this principal component has the same direction for all maturities and does not vanish with time. The second and the third ones are considered short-term ones as they change their sign depending on the maturity considered, which means that a random shock which follows these principal components has some direction in some periods time and the opposite in the others, and therefore, in the long-time its effects tend to vanish.

As can be seen in the Table 8, in the case of WTI and Brent crude oil, the first component explains more than the 90% of the volatility, and the first three ones explain almost the 100% of the volatility. Therefore, in the case of WTI and Brent crude oil, the other principal components have little importance. In the case of natural gas the first three principal components explain the 80% of the volatility, consequently, the other ones are important. The reason for this difference between crude oil and natural gas lies again on the fact that natural gas is a strongly seasonal commodity whereas the crude oil is not, thus, in natural gas the other principal components (the fourth, the fifth and so on) explain part of the volatility because they include seasonal effects.

As before, depending on the characteristic (the type of the data set, the precision needed,…) of our problem we will decide how many $\sigma_i(t,T)$ functions we will choose and the number of parameters that define each of them. The decision should be based on simplicity vs. goodness of fit.

2.5. Seasonality

As said above, the natural gas is a strongly seasonal commodity. One of the clearest ways to visualize this seasonality is through the forward curve. Figure 8 depicts the forward curve for Henry Hub natural gas futures contracts traded at NYMEX on 03/17/2008. In this figure it is possible to appreciate that Henry Hub natural gas prices are expected to be higher during winter months and lower during summer months, therefore, this commodity price is seasonal.

In the factor models presented above only the non-seasonal part of the price is modelled. Thus, a reasonable question to answer is how the results change if we take into account seasonality in the models. In Sorensen (2002) seasonality is captured through a deterministic factor and it is demonstrated that if seasonality is considered in the model, it fits better to the data for seasonal commodities like corn, soybean, and wheat. Chapter 2 shows that seasonality is a stochastic factor in seasonal commodity prices like natural gas, gasoline and heating oil.

As were pointed out by Blanco, Soronow and Stefiszyn (2002), in the principal components model it is possible to incorporate seasonality if we assume that: $F_{T,t+1} = F_{T,t} + F_{T,t}\left[\sum_{i=1}^{N}\sigma_i(t,T)\varepsilon_{it}\right]$ where $\sigma_i(t,T)$ are deterministic volatility functions whose forms are different depending on the month of the date "*t*".

Using a data set which is made of weekly observations of Henry Hub natural gas futures contracts (F1 to F17) from 04/22/1992 to 03/24/2008, we divide the sample in twelve sub-samples (one for each month) and carry out a principal components analysis for each month. As we need enough data each month, we need a data base different from the previous sections one.

The results show that the first component is similar in all months (see Figure 9) whereas the second and the third exhibit a clear seasonal pattern (see Figures 10 and 11). To be precise, the second and the third principal components exhibit the classical "tilt" and "bending" behaviour, however, this behaviour is mixed with a seasonal one.

As can be appreciated in Figure 7, in the case of the second principal component the "tilt" behaviour takes over the seasonal one, while for the third one we find the opposite, i.e. the seasonal behaviour dominates the "bending" one.

3. Investment under uncertainty. The optimal contract determination

Investment projects related with crude oil or natural gas are highly intensive in investment, therefore the stochastic behaviour of commodity prices has important implications for the valuation of projects related to the prices of those commodities.

Concretely, when a company is planning to develop a crude oil or natural gas field, the investment is huge (it usually reaches thousands million dollars) and it usually lasts many years (they usually last between twenty and thirty years), however the main investment has to be faced up before getting any return. Consequently, the company, which faces up the investment, needs a sell contract, which should last at least twenty years, to guaranty the investment recovery.

The most reasonable way to define the sell price is through the price of a contract which quotes in a liquid international market (like NYMEX or ICE). To choose the proper quotation variables, geographical location and product quality have to be taken into account. However, if the sell price is linked with the international one in a linear way, it is not possible to recover the investment when the commodity price goes down. For that reason these buy-sell contracts use to be designed with clauses which enforce the buyer to pay a minimum amount independently of international quotations. This guarantees the seller investment recovery. To compensate the buyer for this clause, and to hedge him against unexpected and steeper price increases, it is also common to introduce other clause to allow the buyer to pay at most a maximum amount independently of international quotations.

As a result, the typical buy-sell commodity contract is a long term contract (it lasts between twenty and thirty years) and is designed in the following way: in each period the exchange price is linked with an international quotation in a linear way with two main clauses, one to guaranty the seller investment recovery through a minimum price and other to protect the buyer from price increases through a maximum price. It is easy to see that, in each liquidation period, these clauses are a put option bought by the seller and a call option bought by the buyer respectively.

As these contracts contain many call and put options, some of them with maturity in a long period time, the stochastic behaviour of commodity prices plays a central roll in valuing these sort of commodity contingent contracts and the chosen model to carry-out the valuation, specially its volatility assumptions, is essential in the final result.

There are mainly two issues which depends on the chosen model, the first one is determining the put and call options value and the second one is determining the maximum price which, given a minimum price, makes the value of the put options, on average, equal to the call options value to get, on average, that the contract value is equal to the one without clauses.

Both issues are highly related. The first one is a general problem, which is determining the options value for calculating the whole contract value. The second one is a particular problem when it is decided that the buyer options (the call ones) have to be valued the same as the seller ones (the put ones). This second issue is essential in negotiating the contract by the seller and buyer company managers. As said above, the way to solve these problems is through the stochastic behaviour of commodity prices and the chosen model is essential. The valuation of the contract using two different models, specially two different models with highly different assumptions about volatility, could differ in hundred million dollars.

To illustrate this fact we propose three fictitious contracts, one for a WTI crude oil field, other for a Brent crude oil field and other for a natural gas field located in Henry Hub[6] which are defining in the following way: The contracts last twenty years from 1/1/2009 to 12/31/2028, the crude oil or natural gas amount exchanged is the same in each period, the liquidation is monthly

[6] Henry Hub is a hub located in Louisiana (EE.UU) near to the Texas border.

and the contract prices are the monthly average of the first month WTI, Brent and Henry Hub futures price traded at NYMEX and ICE respectively.

If the seller wants to include a clause to guarantee a minimum price to recover their initial investment there are two questions to deal with: the first one is the value of this clauses and the second one is what should be the maximum price that the buyer should include to do not lose money. Even more, if the buyer wants to introduce other clause to guarantee a maximum price there are an additional question: what should be its value.

We are going to answer these questions using three different models: the AR(1)[7] model, the two factor model and the three factor model. In all cases the valuation date is 3/24/2008 and the assumed risk free interest rate for the whole period is 5%[8]. As there are no forward curves for any commodity which cover the whole contracts period (from 1/1/2009 to 12/31/2028), it has been estimated a forward curve for the three commodities based on the observed forward curve at the validation date (3/24/2008). The estimatation has been carried out assuming in all cases that there is a long-term forward price and the observed futures prices converge to the long-term one through an exponential way. In the case of natural gas it has been assumed that seasonality is a deterministic factor that has been calculated based on the observed forward curve at the validation date and has been incorporated in the estimated one for the whole contract period.

Figures 12, 13 and 14 show the minimum price clause value as a function of their strike prices (the minimum price) for the three commodities presented above. Figures 15, 16 and 17 contain the maximum price clause valuations whereas Figures 18, 19 and 20 show the maximum price that the buyer should include to do not lose money depending on the minimum price calculated. As can be appreciated in the Figures, the results are presented in \$/bbl or \$/MMBtu to be compared with the commodity average price in the contract period, which is 95.97 \$/bbl in the case of WTI crude oil, 96.14 \$/bbl in the case of Brent crude oil and 8.56 \$/MMBtu in the case of natural gas in Henry Hub. Comparing the clauses values in \$/bbl or in \$/MMBtu with the

[7] The parameters used in the estimate are the ones calculated with data until 1999 because the ones calculated with the whole sample period have no sense as explained above.

[8] The same result has been obtained using different risk free interest rates for each period time.

average prices is equivalent to compare the clauses value with the whole contract value as the amount exchange in each period time is the same.

Table 9 present the differences in valuation using the three factor model vs. the other two models (the AR(1) and the two factor model) as percentage of the average price, which is equal to the differences in valuation as percentage of the whole contract value. Table 10 shows the differences in calculating the maximum price that the buyer should include to do not lose money if the seller includes a minimum price using the three factor model vs. the other two models as percentage of the result obtained with the three factor model.

The first issue to highlight is the fact that the results obtained are basically the same for all commodities, especially in the case of the natural gas, if we use the two factor and the three factor model. That brings again what is said above in relation to both models: as the two factor model is a particular case of the three factor one, the three factor models gets more accurate estimates, however, the two factor one is simpler an easy to deal with. Therefore, depending on the problem the optimal choice should be the accuracy or the simplicity. In this case, as thousand million dollars are involved, it sounds more reasonable to choose accuracy instead of simplicity. As said above, Cortazar and Naranjo (2006) realized that adding one more factor and getting the four factor model is useless and the three factor model is well enough to characterize this sort of commodities dynamics.

Opposite conclusions arise when we compare the valuation results obtained with the two and three factor models with those obtained with the AR(1) one. Just to get an idea about this differences amount we can see for example that the call options value if the maximum price is 100 \$/bbl or the put options value if the minimum price is 95 \$/bbl differs in almost 8 \$/bbl in the case of WTI crude oil and in almost 9 \$/bbl in the case of Brent crude oil, using the AR(1) model instead of the two or three factor models. In both cases it represents almost 10% of the average price. In the case of natural gas these differences round 1.15 \$/MMBtu, which represents almost 15% of the average price. Therefore, as these projects involve enormous amounts of commodity (as said above, its value usually reaches thousand million dollars), a

10% or a 15% of the whole project could represent hundreds million dollars. For that reason the chosen model has a crucial role in negotiating this kind of contracts. Comparing the differences between the two and the three factor model we get, at maximum, 1.3 $/bbl in the case of the crude oil and 0.07 $/MMBtu in the case of natural gas. These figures represent less than 1.5% and 1% in the crude oil case and natural gas case respectively. As the amount of money included in this kind of projects is so big, these differences are also important, however it extend is around ten times less than the previous ones.

Even more, comparing the results obtained with the two and three factor models in valuing the put and call options, we get a bias which is not too big and operates in the same direction in both cases. Subsequently, when we calculate the maximum price that the buyer should include to do not lose money if the seller includes a minimum price in the contract, the differences between models round 3%. If we compare the AR(1) model with the other two models, these differences reach more or less 15% in the case of the crude oil and 22% in the case of natural gas. As before, we can conclude that, as the amount of money involved in this type of projects is huge, the differences between models are important in all cases, however in comparing the AR(1) model with the other two, the differences get critical importance in contract determination.

These huge differences in the valuation results obtained with these models (the AR(1) model vs. the two and three factor models) comes from the fact that in the AR(1) volatility is bounded, whereas in the two and three factor models it is not bounded. As an option contract is not symmetric, the volatility plays a central role in its valuation (see for example Hull, 2006) and, as the contract lasts so much time, the fact the volatility is or not bounded is crucial in determining the value, as we have seen.

Other issue to highlight is the fact that, as can be appreciated in Figures 18, 19 and 20, there is not symmetry between the minimum and the maximum price. For example, with an average price of 95.97 $/bbl for the WTI crude oil, one naive manager could think that if the minimum price is settled at 50 $/bbl, the maximum price should be fixed at 95.97 + (95.97 – 50) = 141.97

$/bbl to get the same amount upwards than downwards. Nevertheless, as can be appreciated in the Figures this reasoning is false. With the AR(1), the two and three factor models the maximum price should be 192, 217 and 223 $/bbl respectively. The reason behind this result is again the options contracts asymmetry, and the reason of getting a higher maximum price than the one defined by the symmetric axis is the fact that the put option revenue is bounded (the maximum revenue is the strike price) whereas the call option revenue is not (for more details see, for example, Hull, 2006), therefore higher strike prices are needed in call options.

As the put option revenue is bounded and the call option one is not, there is not symmetry between the maximum and the minimum price independently of the model used in valuation. However, as we can see in Figures 18, 19 and 20 and in Table 10, and as we have analysed above, the degree of asymmetry depends on the chosen model, specially depends on the chosen model volatility assumptions.

As said above, the big differences found come from the fact that the AR(1) model assumes that the volatility is bounded in the long-term and the factor models do not. If we compare two models which assume that the volatility is bounded or two models which assume that the volatility is not bounded, in the long-term the differences are much smaller. However, as these sorts of contracts involves thousands million dollars, small differences in valuation involves much money.

4. Conclusions

This chapter presents an empirical study about the different models which can be used to characterize the commodity price dynamics, with special emphasis on the volatility of futures returns estimates. These models are presented in a unified context, analyzing the relationships among them, and pointing out their advantages and limitations. It is analyzed their relative performance with a common data set, which is composed of three commodity price series: WTI crude oil and natural gas in Henry Hub futures contracts traded at NYMEX and Brent crude oil futures contracts traded at ICE Futures in London. Finally we discuss the importance of these models in investment under uncertainty, concretely in valuing long-term contracts highly intensive in investment with clauses which are put and call options.

The first model analyzed is the basic AR(1) model. However, these one-factor mean-reverting models, like the AR(1), are not very realistic since they generate a volatility of futures returns which goes to zero as the time to maturity of the futures contract approaches infinity, and imply that futures prices for different maturities should be perfectly correlated, which defies existing evidence.

Next we analyze the empirical performance of the factor models proposed in the literature: the two-factor model introduced by Schwartz and Smith (2000) and the three-factor model proposed by Cortazar and Schwartz (2003) which is an extension of the two-factor one. Our results indicate that with these factor models the estimated volatility fits better to the empirical one than with the basic AR(1) model. Moreover, the three-factor model outperforms the two-factor one in terms of the in sample predictive ability. Finally, we can conclude that since the three-factor model has more structure (more factors and parameters), the goodness of fit is better than in the previous case, however, as it has more structure it is more difficult to estimate, the significance of each stochastic factor is less clear, needs more data to be estimated… Consequently in each case, depending on its characteristics, we will choose between the simplicity (the two factor model) and the goodness of fit (the three factor model).

In all cases, the factors models goodness of fit is worse with natural gas data than with crude oil data because we do not take into account seasonal effects in model specification. Other interesting evidence that we find in this study is the fact that the mean-reversion in the three commodities is less steep when using data until 2008 than in previous studies which did not use recent data. That is caused by the fact that there is mean reversion in commodity prices until 1999, afterwards commodity prices exhibit a random walk behaviour.

The principal components analysis proposed by Clewlow and Stricland (2000) is investigated next. These authors try to model the futures price dynamics directly, by assuming that the futures prices are a martingale, where the price of a futures contract at time t is its price at time t-1 plus an innovation which is given by the sum of several deterministic volatility functions. It is possible to define these volatility functions through a principal component analysis. In the other hand, given that many commodity prices are strongly seasonal, it is discussed the importance of explicitly incorporating seasonality effects in the models.

One we have compared the relative performance between models, we apply the conclusions to investment under uncertainty, especially in determining the best contract. The projects related with the commodities presented in this work are highly intensive in investment, last many years and have clauses which include put and call options. Therefore, their price dynamics characterization is essential in valuing these clauses.

We have seen a hypothetical project for each commodity to illustrate the big differences in contract valuation using different models. Concretely, if we use an AR(1) model instead of a two or three factor one, the clauses valuation can differ between the 10%-15% of the whole contract value, which should represent hundred million dollars. These big differences come from the fact that the AR(1) model assumes that the volatility is bounded in the long term and the two or three factor models do not. However, in options contract valuation volatility assumptions are crucial.

If we compare two models with the same assumption about volatility (bounded or not bounded) the differences are much smaller. However, as these sorts of contracts involve thousands million dollars, small differences in valuation entails much money.

In calculating the maximum price that the buyer should include in the contract to do not lose money if the seller includes a minimum price in the contract, there are also big differences if we use the AR(1) model instead of the two and the three factor ones. The reason is the same, in one case the volatility is assumed bounded and in the other it is assumed not bounded.

Other interesting issue to take into account is the fact that there is not symmetry between the maximum and the minimum price. The reason behind it is the fact that put option revenue is bounded whereas call options revenue is not. Therefore, higher strike prices are needed in call options. Even though it happens independently of the model used, once again, the model volatility assumptions are crucial in determining the degree of asymmetry between the maximum and the minimum price.

APPENDIX

Nested Models

As said above, the AR(1) model, the two factor model and the three factor model are nested models, that is, the two factor model is a particular case of the three factor model when one of the short-term factors is assumed deterministic. The AR(1) model is a particular case of the two factor model when the long-term factor is assumed deterministic.

To prove it, we are going to present the models in their formal way, through their stochastic differential equations (SDE). In this work, just for simplicity, we have presented the models in their discrete-time version. However, as can be seen in the standard literature (see for example Cortazar and Naranjo, 2006), the proper way to define a factor model is through its SDE. Afterwards, if it is necessary in the parameters estimate procedure, the discrete-time version of the model is presented.

We start with the three factor model SDE:

$$d\varepsilon_t = \mu_\varepsilon dt + \sigma_\varepsilon dW_{\varepsilon t}$$

$$d\chi_{1t} = -\kappa_1 \chi_{1t} dt + \sigma_{\chi 1} dW_{\chi 1t}$$

$$d\chi_{2t} = -\kappa_2 \chi_{2t} dt + \sigma_{\chi 2} dW_{\chi 2t}$$

In Chapter 2 it is proved that the discrete-time version of the model is the one presented in section 2.3.

The two factor model SDE are the ones presented above, except for the last one, which does not appear in this case (there is only one stochastic short-term factor):

$$d\varepsilon_t = \mu_\varepsilon dt + \sigma_\varepsilon dW_{\varepsilon t}$$

$$d\chi_t = -\kappa\chi_t dt + \sigma_\chi dW_{\chi t}$$

As before, it is easy to prove that the discrete-time version of the model is the one presented in section 2.2.

The AR(1) model SDE is the one presented for the short-term factor in the two factor model:

$$d\chi_t = -\kappa\chi_t dt + \sigma_\chi dW_{\chi t}$$

The discrete-time version of the model is the one presented in section 2.1.

Therefore, as it is said above, we can conclude that the two factor model is a particular case of the three factor one and the AR(1) model is a particular case of the two factor one.

REFERENCES

- Blanco, C., Soronow, D. and Stefiszyn, P. (2002) 'One step forward', *GARP Risk Review*, Vol. 9, pp. 39–43.
- Clewlow L. & Strickland, C. (2000). Energy Derivatives, Pricing and Risk Management. London (U.K.): Lamina Publications.
- Cortazar G. & Naranjo L. (2006). An N-Factor gaussian model of oil futures prices. *The Journal of Futures Markets*, 26, 209-313.

- Cortazar, G. and Schwartz, E.S. (2003) ‘Implementing a stochastic model for oil futures prices’. *Energy Economics*, Vol 25, pp. 215-218.
- Harvey, A.C. (1989). Forecasting Structural Time Series Models and the Kalman Filter. Cambridge (U.K.): Cambridge University Press.
- Hull J. C. (2006). Options, Futures and Other Derivatives. Sixth Edition. New Jersey, Upper Saddle River (USA): Prentice Hall.
- Schwartz, E. S. and Smith, J E. (2000) ‘Short-term variations and long-term dynamics in commodity prices’, *Management Science*, Vol. 46, pp. 893–911.
- Schwartz, E. S. (1997) ‘The stochastic behaviour of commodity prices: Implication for valuation and hedging’, *The Journal of Finance*, Vol. 52, pp. 923–973.
- Sorensen, C. (2002). ‘Modeling seasonality in agricultural commodity futures’. *The Journal of futures markets*, Vol. 22, pp. 393-426.

FIGURE 1

FIGURE 2

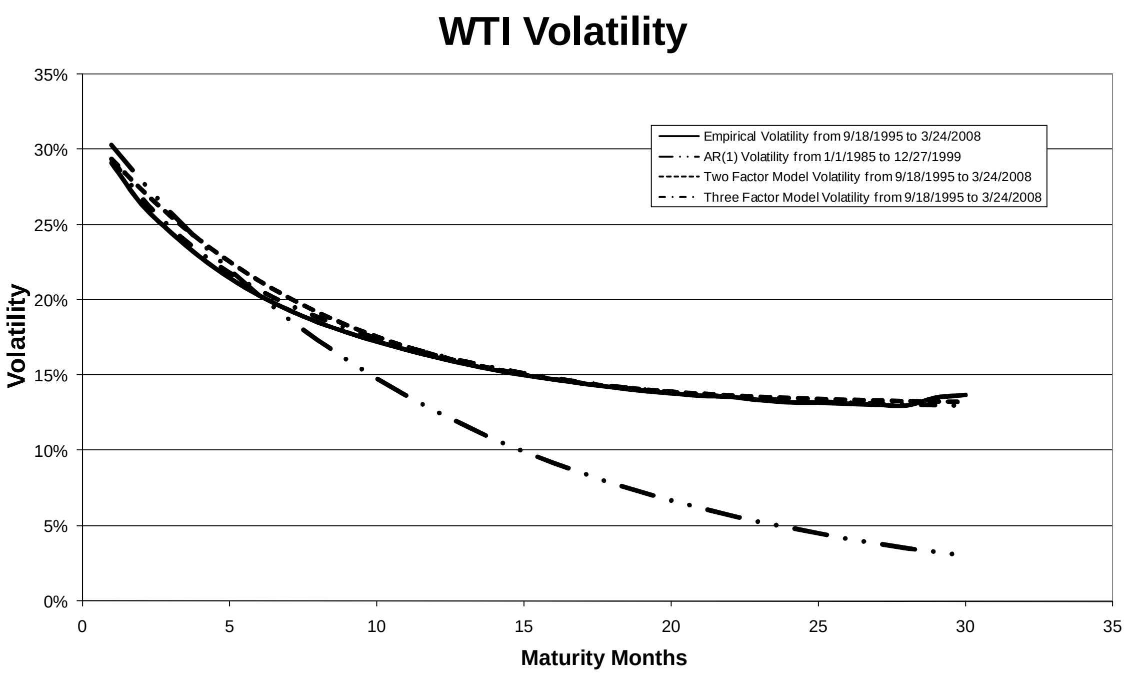


FIGURE 3

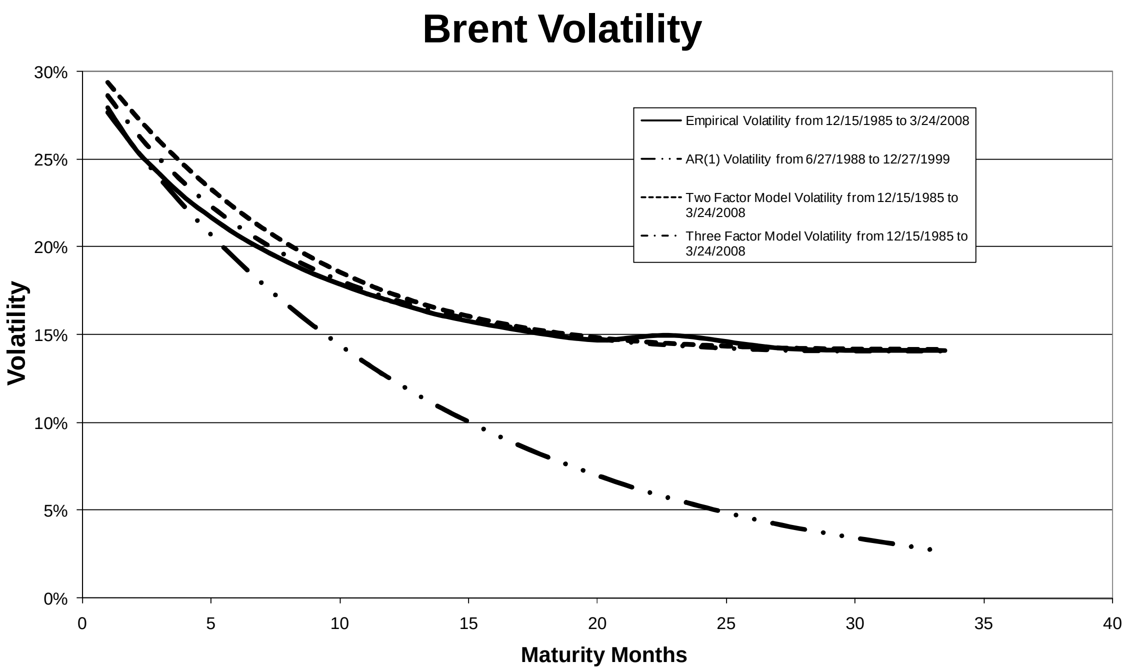


FIGURE 4

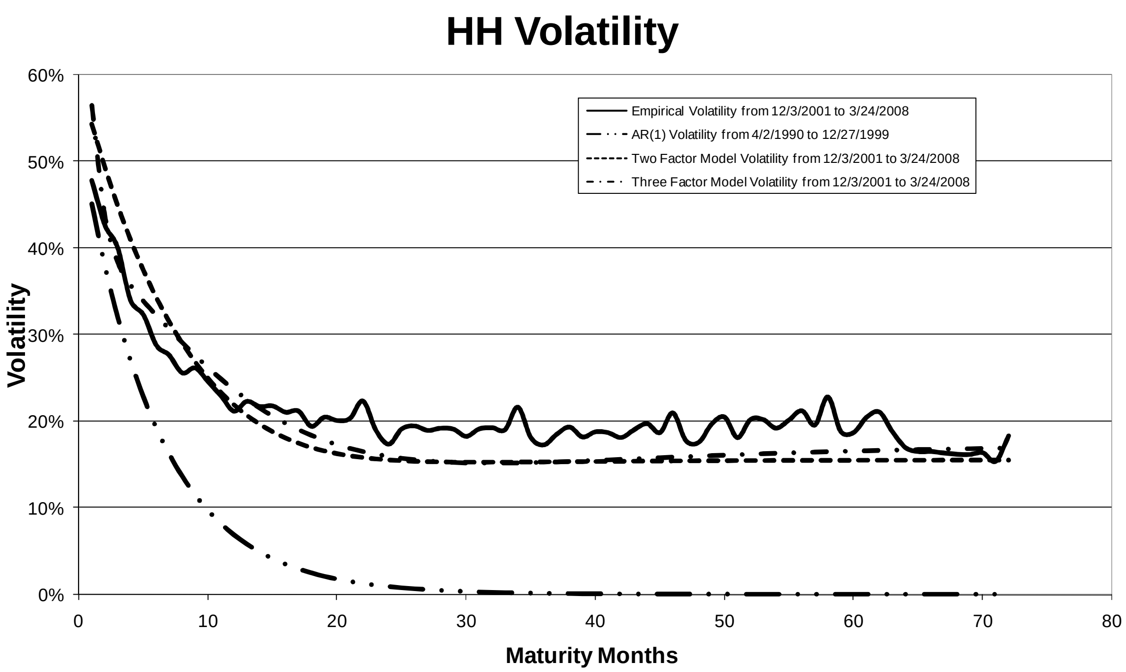


FIGURE 5

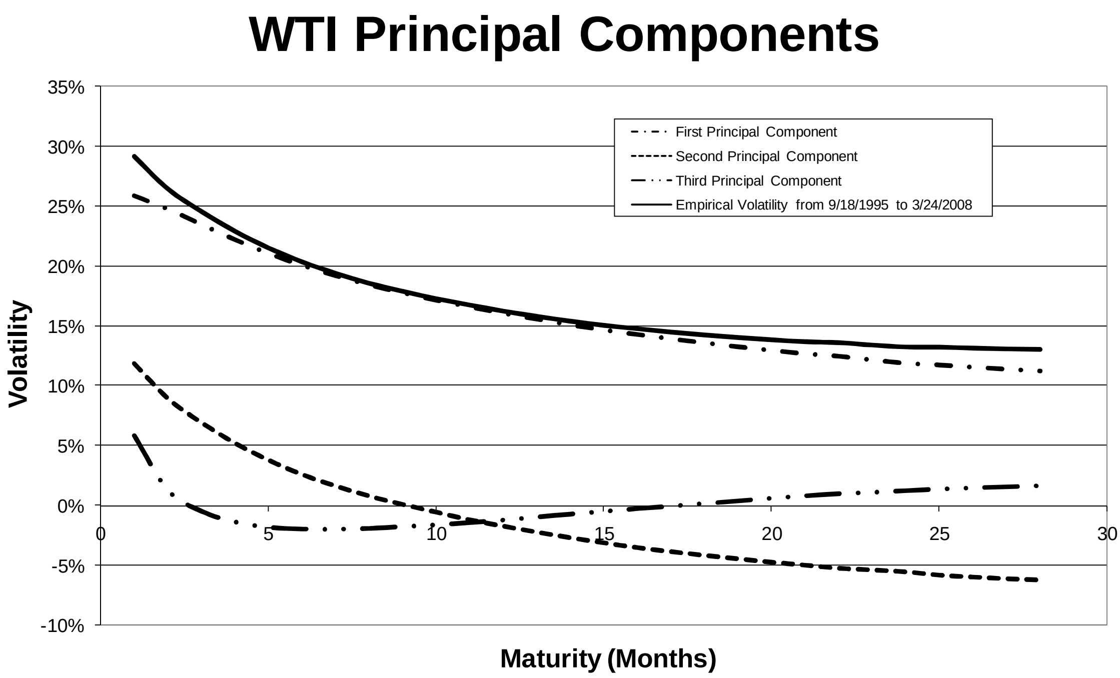


FIGURE 6

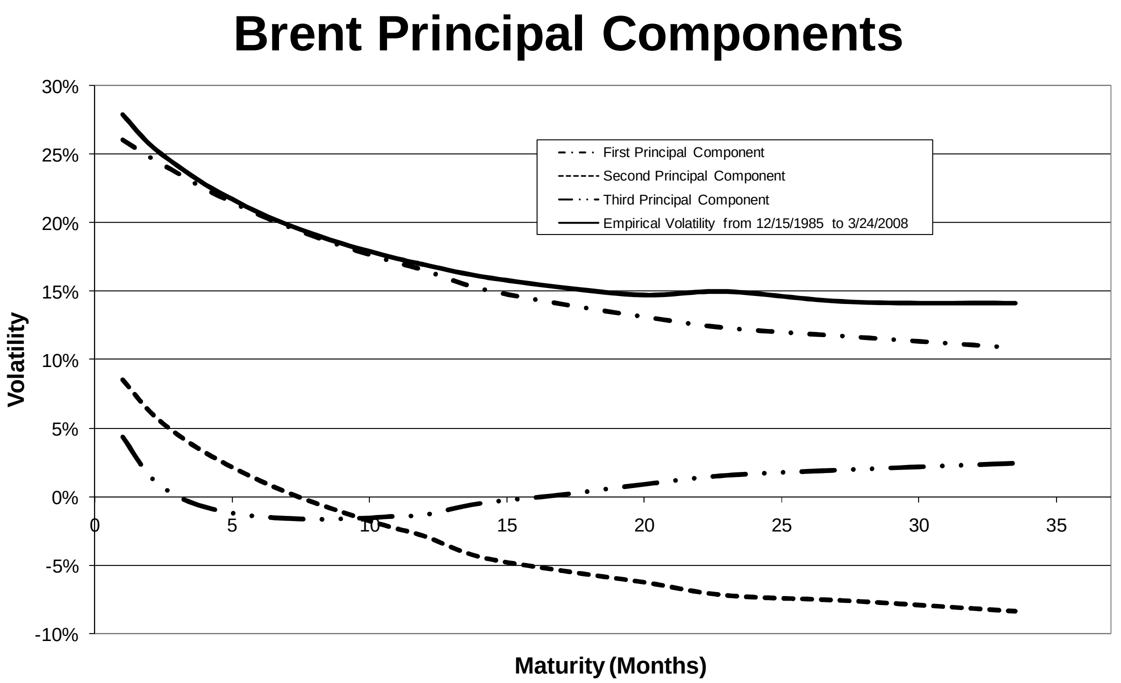


FIGURE 7

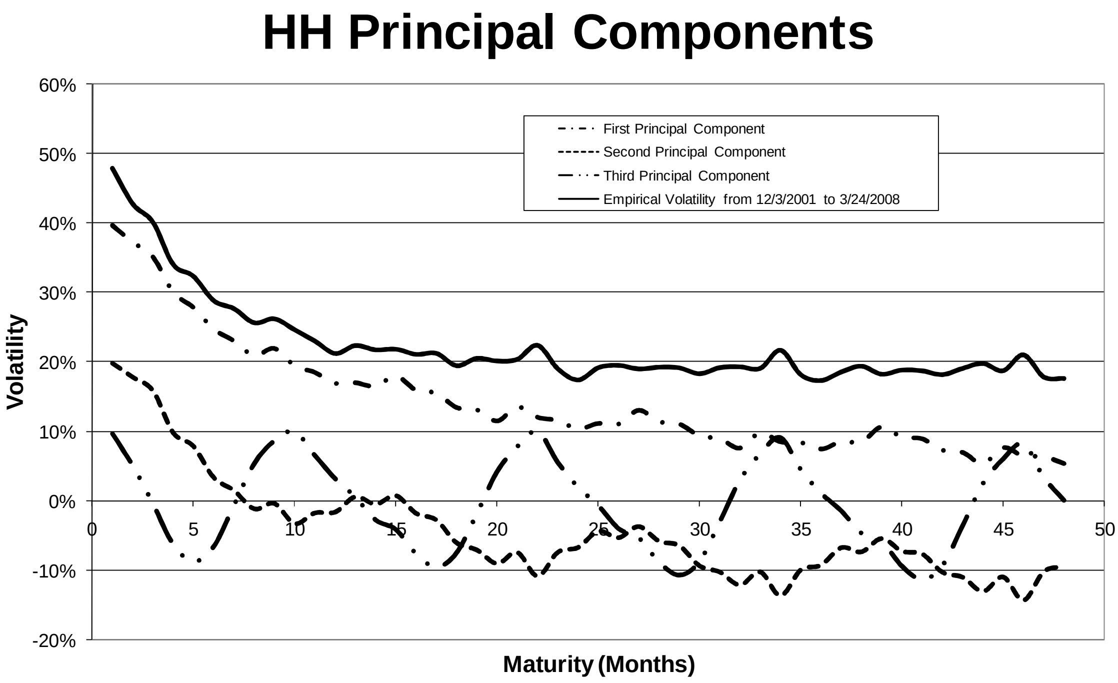


FIGURE 8

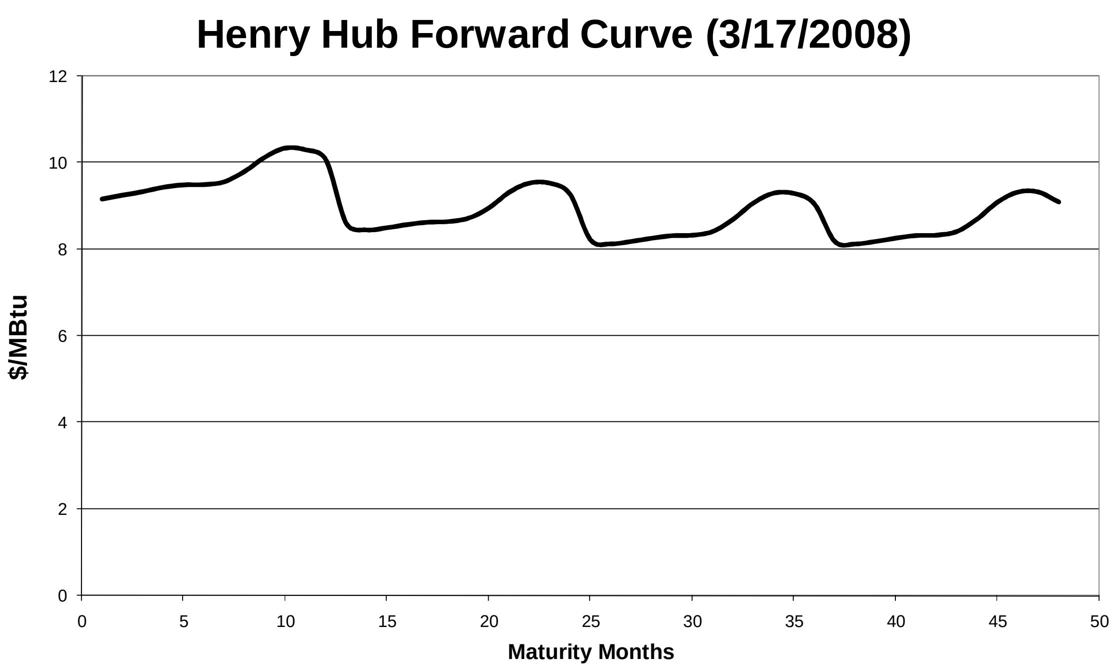


FIGURE 9

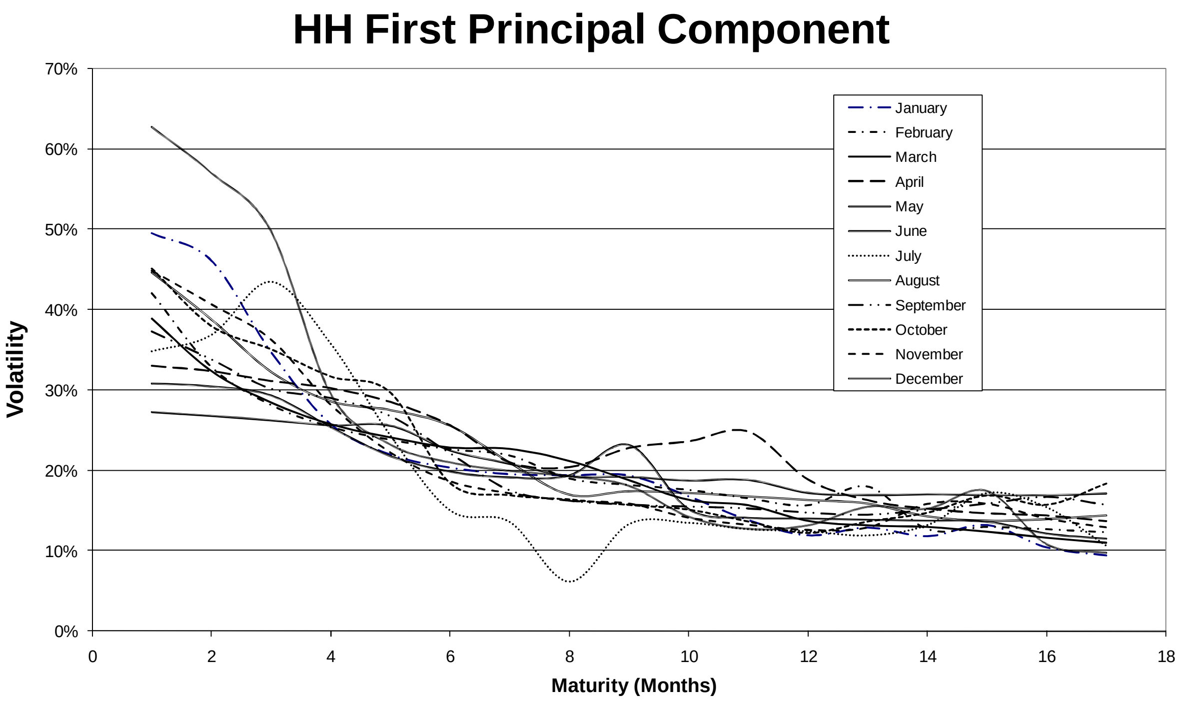


FIGURE 10

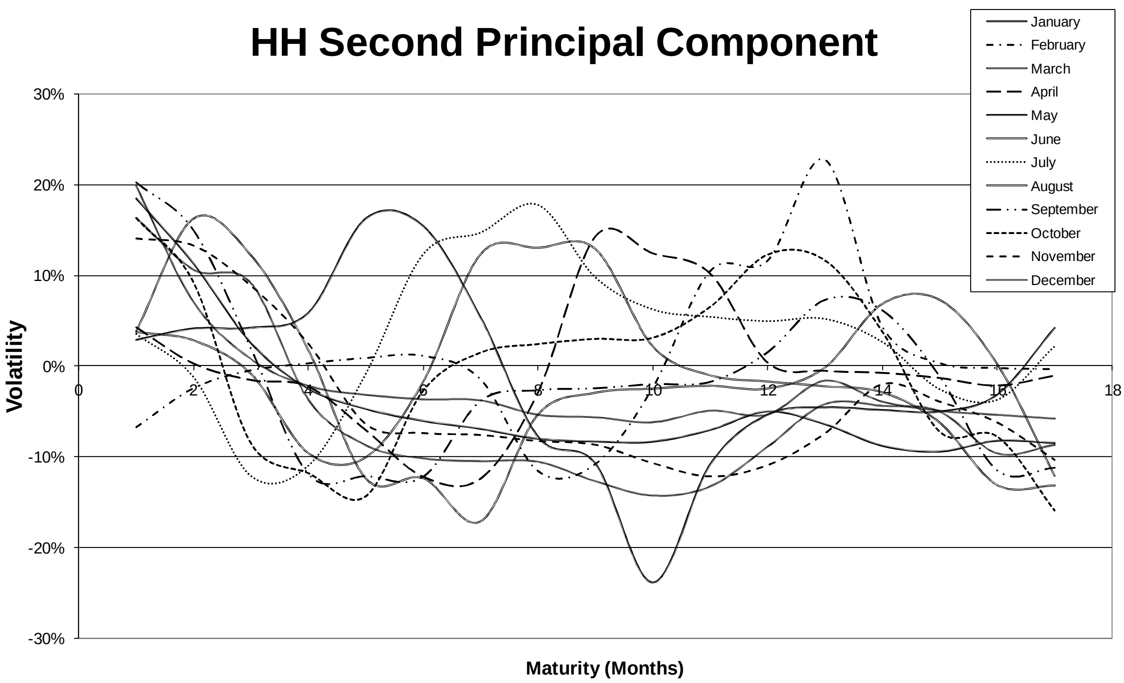


FIGURE 11

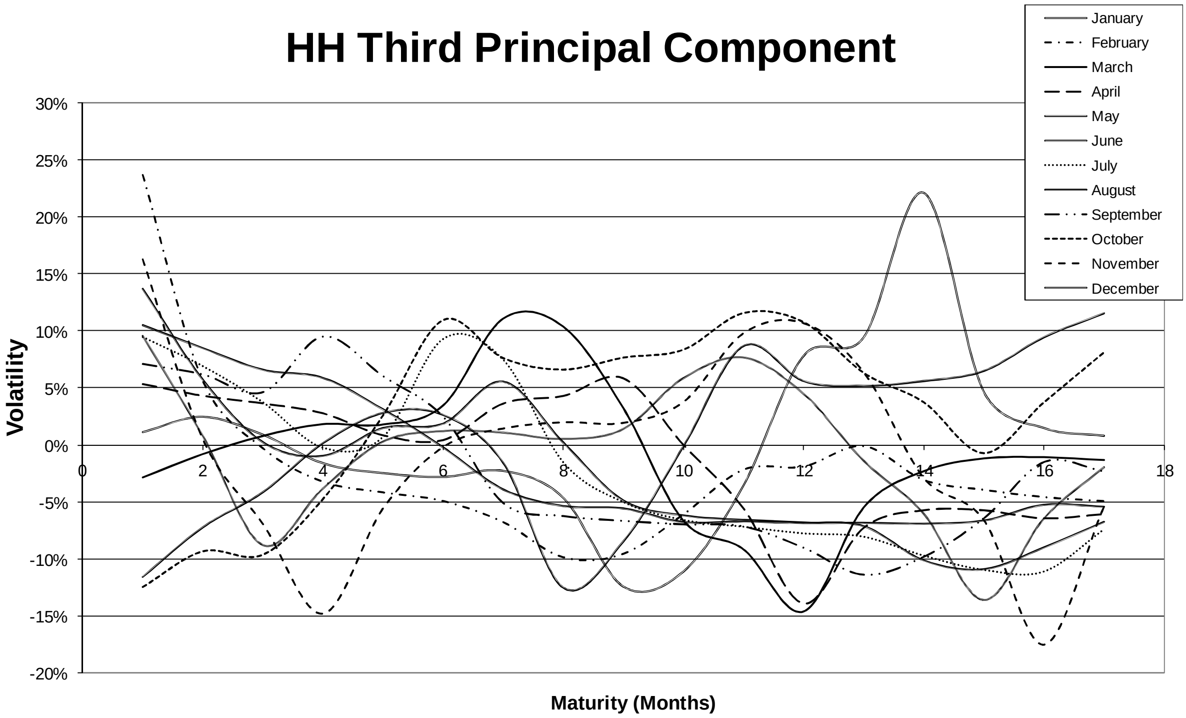


FIGURE 12

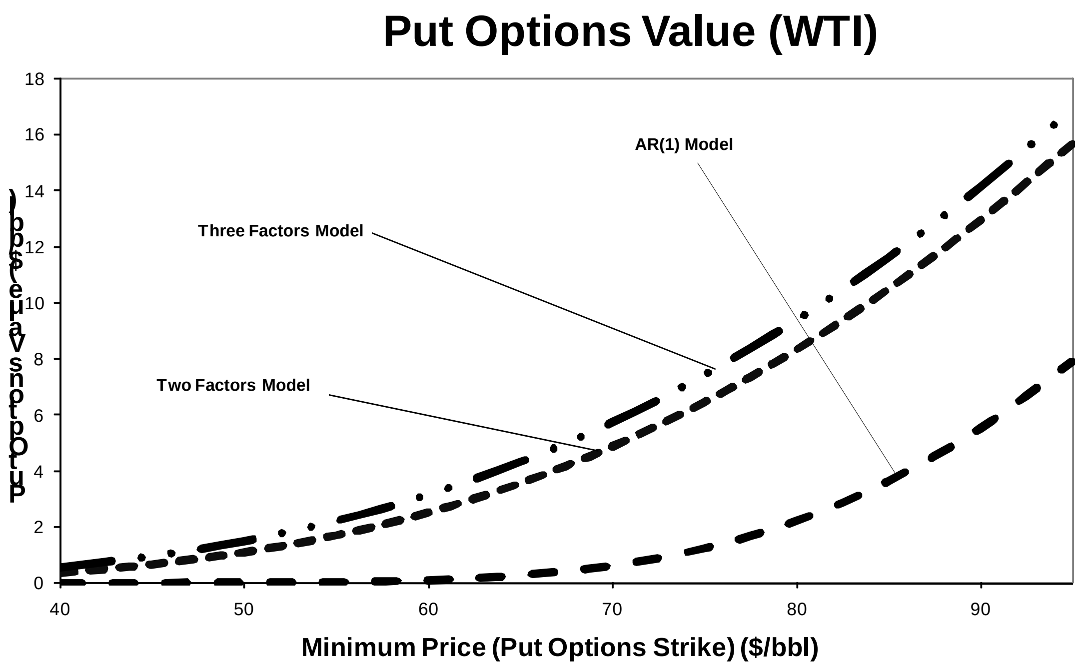


FIGURE 13

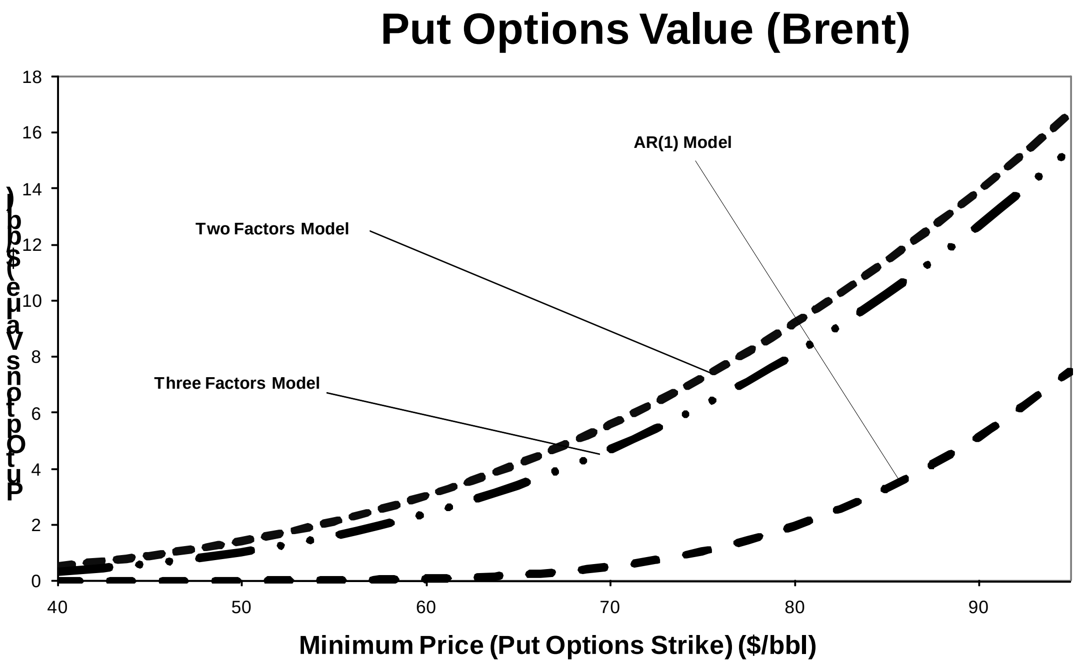


FIGURE 14

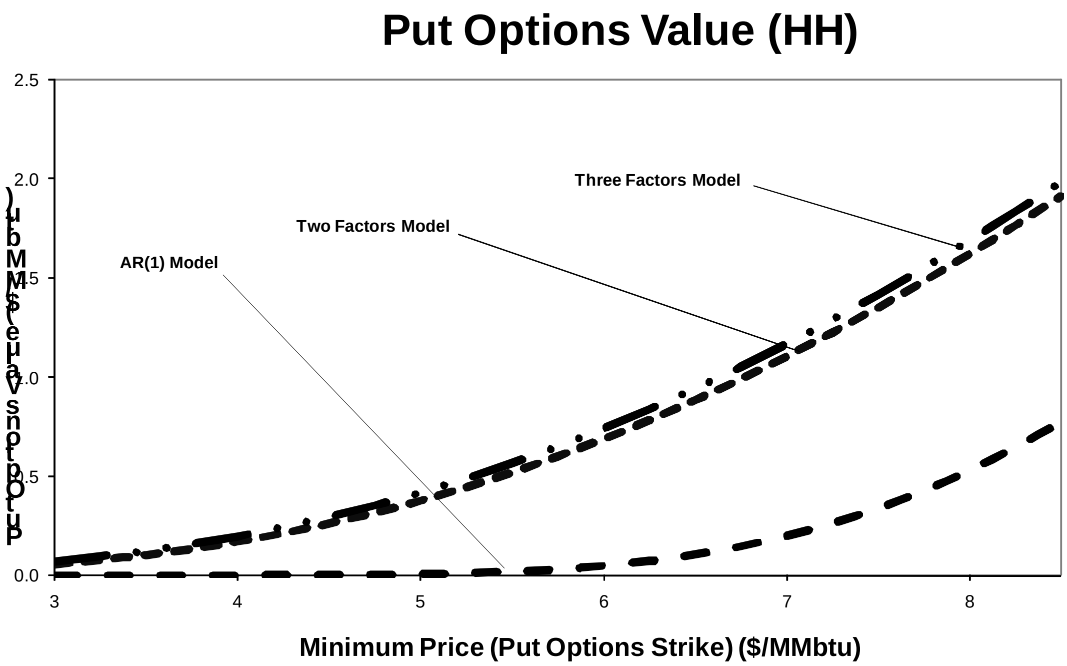


FIGURE 15

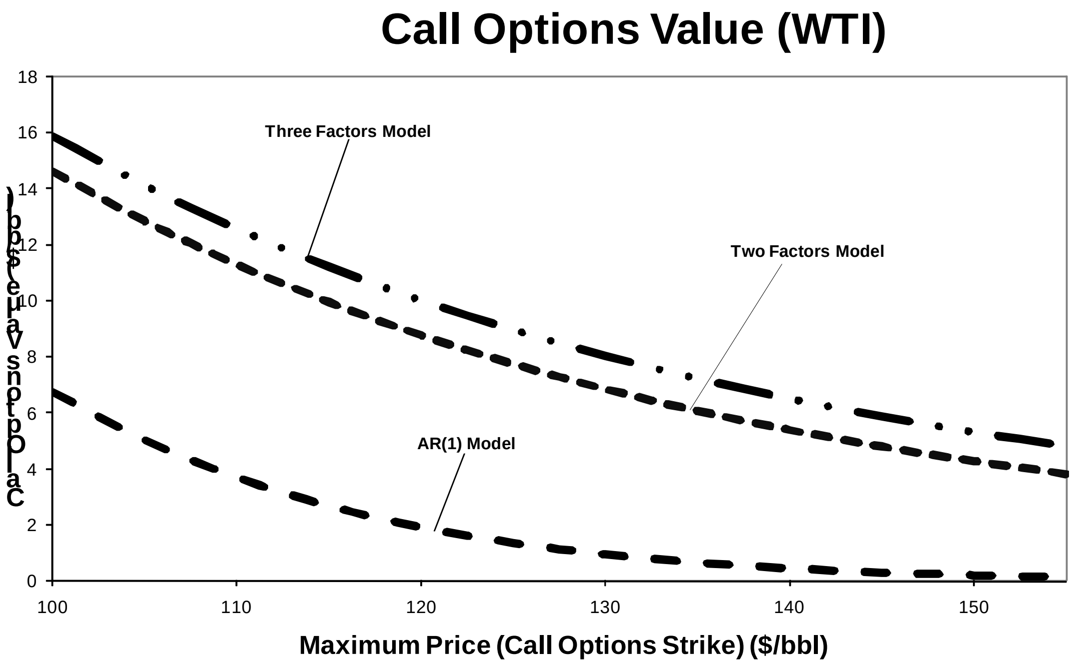


FIGURE 16

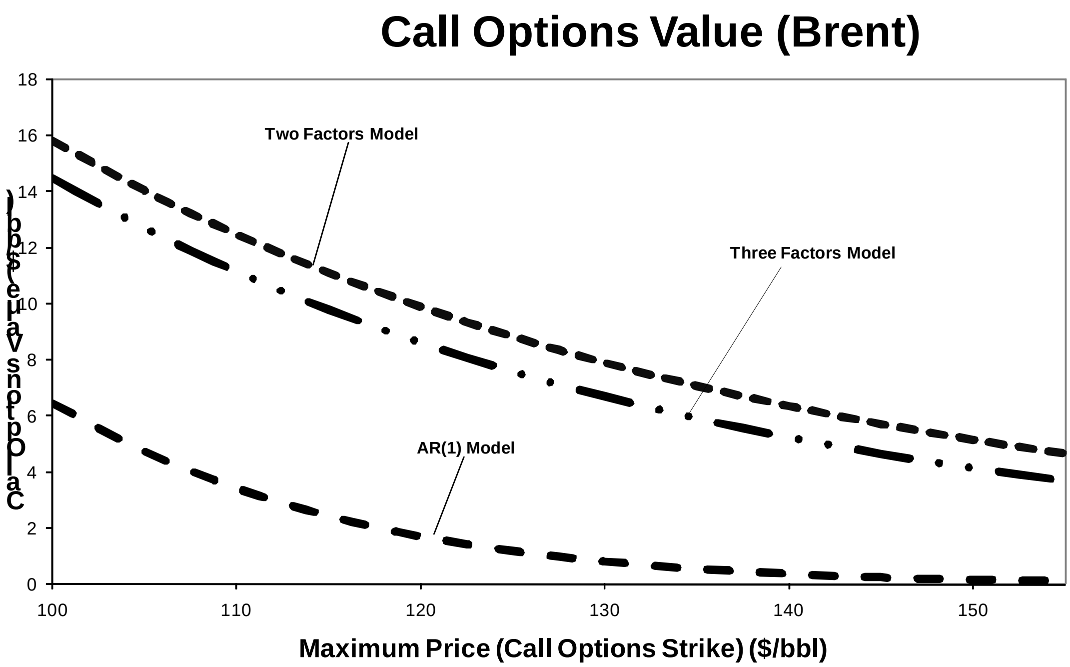


FIGURE 17

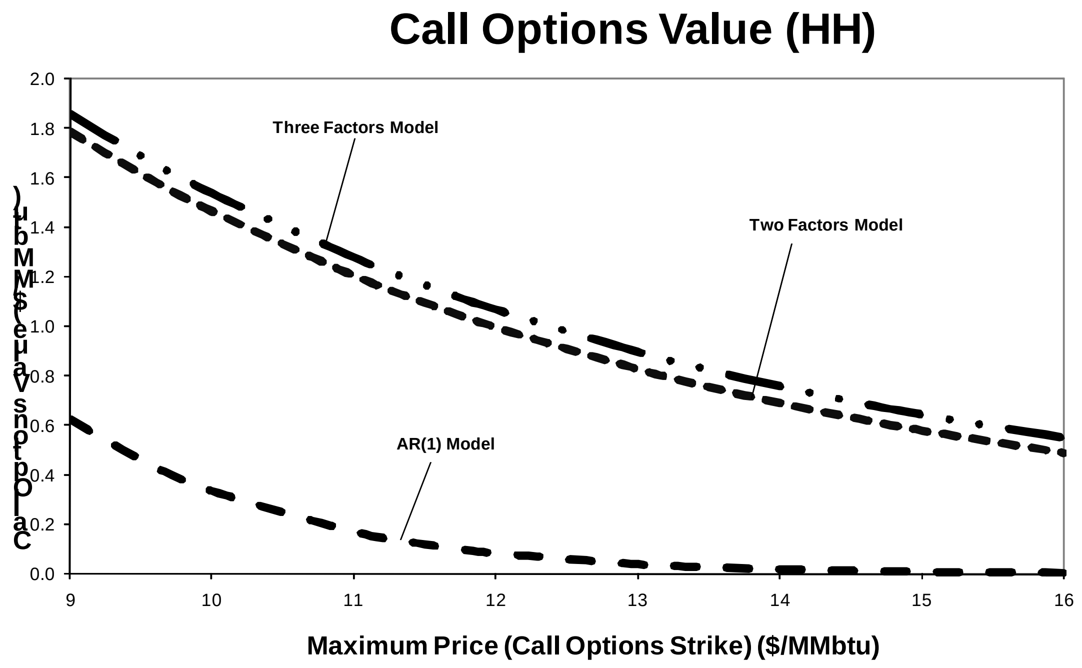


FIGURE 18

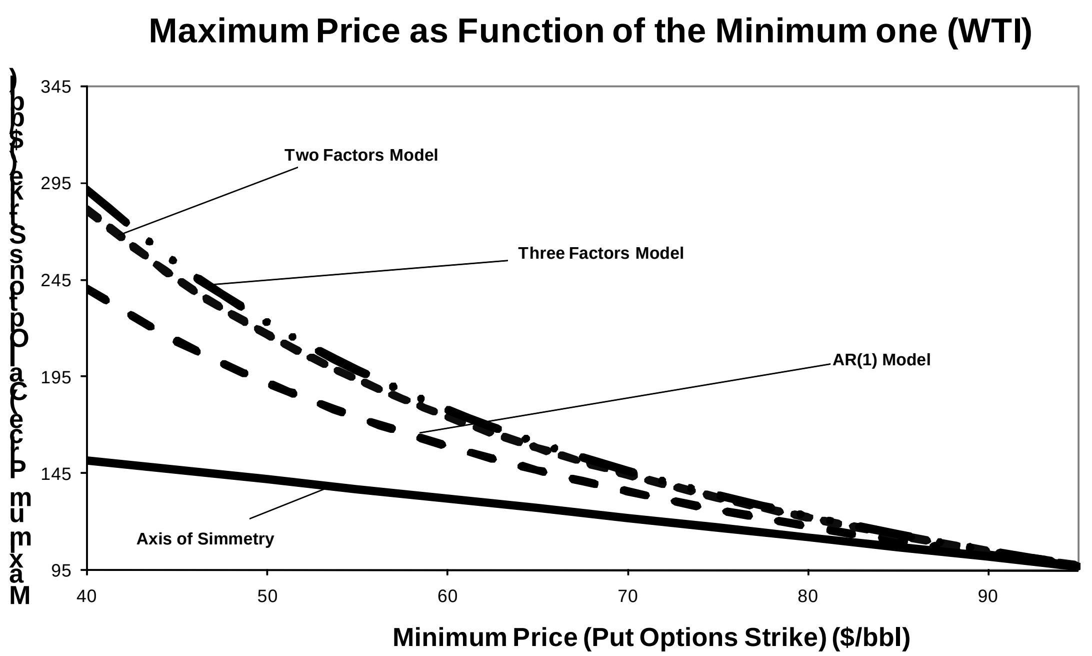


FIGURE 19

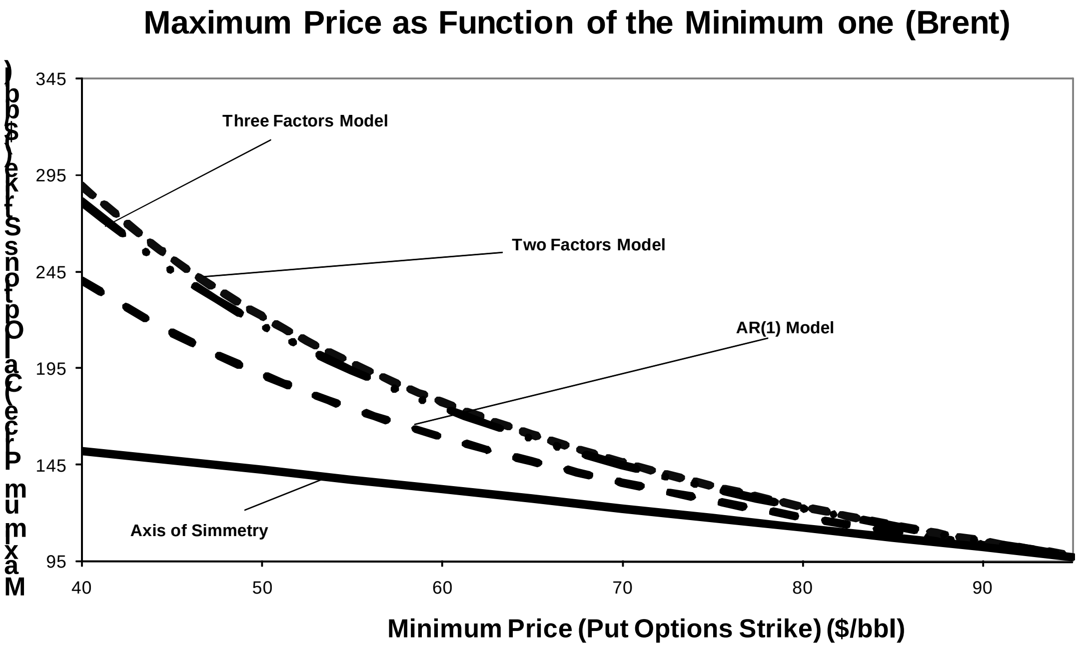


FIGURE 20

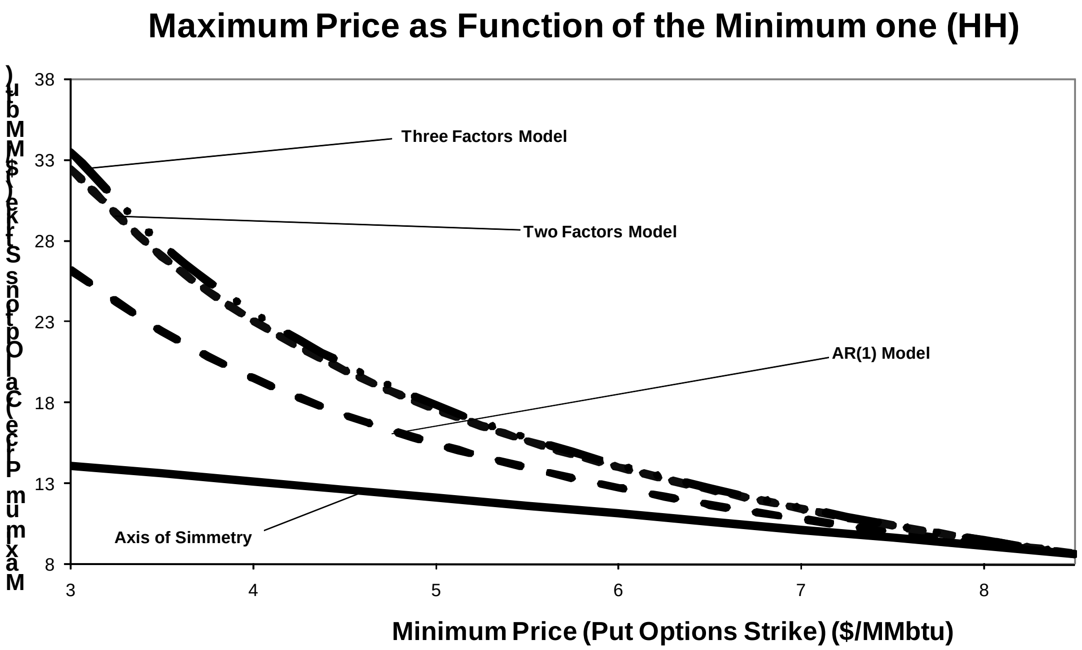


TABLE 1

DESCRIPTIVE STATISTICS FOR THE DATA SET USED IN THE AR(1) ESTIMATION.

The table shows the main descriptive statistics of the three commodity prices series employed in the estimation. Weekly Observations of one month futures contracts are used in all cases. For the whole sample period: NYMEX WTI crude oil from 1/1/1985 to 3/24/2008, NYMEX Henry Hub natural gas from 4/2/1990 to 3/24/2008 and ICE Brent crude oil from 6/27/1988 to 3/24/2008. Until 1999: NYMEX WTI crude oil from 1/1/1985 to 12/27/1999, NYMEX Henry Hub natural gas from 4/2/1990 to 12/27/1999 and ICE Brent crude oil from 6/27/1988 to 12/27/1999.

One Month Futures Contract	Whole Sample Period		Until 1999	
	Mean	Standard Deviation	Mean	Standard Deviation
WTI	28.75 $/bbl	17.68 $/bbl	19.64 $/bbl	4.27 $/bbl
Brent	28.88 $/bbl	18.82 $/bbl	18.11 $/bbl	3.85 $/bbl
Henry Hub	3.8 $/MMBtu	2.5 $/MMBtu	2.03 $/MMBtu	0.48 $/MMBtu

TABLE 2

AR(1) PARAMETER ESTIMATES

The Table shows the parameter estimates of the AR(1) model, which assumes that the commodity log-spot price (p_t) follows the process: $\mathrm{p}_t - c = \rho(\mathrm{p}_{t\text{-}1} - c) + \varepsilon_t$. There are two data sets. For the whole sample period: NYMEX WTI crude oil from 1/1/1985 to 3/24/2008, NYMEX Henry Hub natural gas from 4/2/1990 to 3/24/2008 and ICE Brent crude oil from 6/27/1988 to 3/24/2008. Until 1999: NYMEX WTI crude oil from 1/1/1985 to 12/27/1999, NYMEX Henry Hub natural gas from 4/2/1990 to 12/27/1999 and ICE Brent crude oil from 6/27/1988 to 12/27/1999. Weekly observations of one month futures contracts are used in all cases. Standard deviations in parenthesis.

Param.	Whole Sample Period			Until 1999		
	Brent Crude Oil	WTI Crude Oil	Henry Hub Natural Gas	Brent Crude Oil	WTI Crude Oil	Henry Hub Natural Gas
ρ	1.00 (0.002)	1.00 (0.003)	0.99 (0.003)	0.98 (0.008)	0.98 (0.007)	0.96 (0.01)
c	-1.97 (20.69)	14.03 (76.10)	1.51 (0.93)	2.93 (1.35)	2.9 (1.18)	0.69 (0.23)
σ	28%	30%	48%	28%	30%	45%
R^2	99%	99%	99%	96%	96%	92%
Number Obs.	1031	*1213*	*939*	601	783	509

TABLE 3

DESCRIPTIVE STATISTICS FOR THE DATA SET USED IN THE TWO AND THREE-FACTOR MODELS ESTIMATION

The Table shows the main descriptive statistics for the data set used in the two and three factor model estimation: NYMEX WTI crude oil futures contracts from 9/18/1995 to 3/24/2008, NYMEX Henry Hub natural gas futures contracts from 12/3/2001 to 3/24/2008, and ICE Brent crude oil futures contracts from 12/15/1997 to 3/24/2008. Weekly Observations in all cases.

Henry Hub ($/MMBtu)			WTI ($/bbl)			Brent ($/bbl)		
Contract	Mean	Standard Deviation	Contract	Mean	Standard Deviation	Contract	Mean	Standard Deviation
F1	6.40	2.22	F1	36.33	20.94	F1	38.32	21.98
F5	6.86	2.25	F4	36.07	21.42	F4	38.23	22.40
F9	6.93	2.26	F7	35.62	21.64	F7	37.89	22.67
F13	6.85	2.26	F10	35.20	21.74	F10	37.50	22.81
F17	6.77	2.26	F13	34.82	21.76	F12	37.24	22.86
F21	6.66	2.19	F16	34.48	21.73	F16-18	36.69	22.84
F25	6.53	2.10	F19	34.20	21.66	F22-24	36.14	22.69
F29	6.44	2.09	F22	33.96	21.57	F31-36	35.42	22.33
F33	6.33	2.03	F25	33.76	21.46			
F37	6.24	1.94	F28	33.59	21.36			
F41	6.18	1.93						
F45	6.10	1.87						
F48	6.05	1.80						

TABLE 4

TWO FACTOR MODEL PARAMETER ESTIMATES

The Table shows the parameter estimates in the two-factor model. The data set is composed of weekly observations of NYMEX WTI crude oil futures contracts from 9/18/1995 to 3/24/2008, NYMEX Henry Hub natural gas futures contracts from 12/3/2001 to 3/24/2008, and ICE Brent crude oil futures contracts from 12/15/1997 to 3/24/2008. Standard errors in parentheses. The estimated values are reported with * denoting significance at 10%, ** denoting significance at 5%, and *** denoting significance at 1%.

Contract	Brent	Henry Hub	WTI
μ_ξ	0.1641^{***} (0.0293)	0.1524^{***} (0.0286)	0.1376^{***} (0.0247)
k	0.8854^{***} (0.0071)	1.1879^{***} (0.0425)	1.0598^{***} (0.0083)
σ_ξ	0.1441^{***} (0.0033)	0.1553^{***} (0.0059)	0.1315^{***} (0.0027)
σ_χ	0.3038^{***} (0.0081)	0.6057^{***} (0.0240)	0.2905^{***} (0.0066)
$\rho_{\xi\chi}$	-0.1736^{***} (0.0342)	-0.1791^{***} (0.0585)	-0.0240^{***} (0.0310)
μ_ξ^*	-0.0269^{***} (0.0008)	-0.0601^{***} (0.0016)	-0.0219^{***} (0.0007)
λ_χ	0.0792 (0.0630)	-0.0225 (0.1327)	0.1120^{**} (0.0547)
σ_η	0.0122^{***} (0.0001)	0.0735^{***} (0.0006)	0.0127^{***} (0.0001)
Log-likelihood	30847	17635	47232
AIC	30831	17619	47215
SIC	30797	17588	47180
Number Obs.	537	330	654

TABLE 5

IN-SAMPLE PREDICTIVE ABILITY OF THE TWO-FACTOR MODEL

The Table shows the mean error (real minus predicted) and the root mean squared error (RMSE) obtained with the two-factor model, for the three commodity prices series, using the parameter estimates in Table 4.

BRENT			HENRY HUB			WTI		
Contract	Mean	RMSE	Contract	Mean	RMSE	Contract	Mean	RMSE
F1	0.0035	0.0428	F1	0.0175	0.0945	F1	0.0025	0.0457
F4	-0.0031	0.0323	F5	-0.0159	0.0870	F4	-0.0026	0.0333
F7	-0.0027	0.0282	F9	-0.0161	0.0838	F7	-0.0019	0.0291
F10	-0.0008	0.0261	F13	-0.0066	0.0676	F10	-0.0006	0.0265
F12	0.0006	0.0249	F17	0.0012	0.0772	F13	0.0006	0.0240
F17	0.0021	0.0231	F21	0.0044	0.0749	F16	0.0014	0.0220
F23	0.0021	0.0219	F25	0.0087	0.0717	F19	0.0016	0.0203
F34	-0.0017	0.0256	F29	0.0104	0.0720	F22	0.0009	0.0197
-	-	-	F33	0.0085	0.0706	F25	-0.0003	0.0207
-	-	-	F37	0.0057	0.0725	F28	-0.0017	0.0232
-	-	-	F41	0.0007	0.0709	-	-	-
-	-	-	F45	-0.0054	0.0749	-	-	-
-	-	-	F48	-0.0127	0.0844	-	-	-

TABLE 6

THREE FACTOR MODEL PARAMETER ESTIMATES

The table shows the parameter estimates of the three-factor model. The data set is composed of weekly observations of NYMEX WTI crude oil futures contracts from 9/18/1995 to 3/24/2008, NYMEX Henry Hub natural gas futures contracts from 12/3/2001 to 3/24/2008, and ICE Brent crude oil futures contracts from 12/15/1997 to 3/24/2008. Standard errors in parentheses. The estimated values are reported with * denoting significance at 10%, ** denoting significance at 5%, and *** denoting significance at 1%.

Contract	Brent	Henry Hub	WTI
μ_ξ	0.1346*** (0.0252)	0.1598*** (0.0308)	0.1766*** (0.0317)
k_1	0.8477*** (0.0119)	0.7001*** (0.0424)	0.8348*** (0.0288)
k_2	2.2362*** (0.0286)	6.2955*** (0.0000)	1.0800*** (0.0328)
σ_ξ	0.1300*** (0.0026)	0.1726*** (0.0086)	0.1515*** (0.0037)
$\sigma_{\chi 1}$	0.2963*** (0.0069)	0.5333*** (0.0216)	0.8987*** (0.1946)
$\sigma_{\chi 2}$	0.2826*** (0.0074)	0.8514*** (0.0000)	0.9187*** (0.1956)
$\rho_{\xi\chi 1}$	-0.1497*** (0.0289)	-0.4790*** (0.0607)	-0.3921*** (0.0314)
$\rho_{\xi\chi 2}$	0.1951*** (0.0288)	0.3197*** (0.0531)	0.3763*** (0.0313)
$\rho_{\chi 1\,\chi 2}$	-0.4867*** (0.0276)	-0.3741*** (0.0487)	-0.9553*** (0.0195)
μ_ξ^*	0.0032*** (0.0009)	-0.0458*** (0.0018)	0.0041*** (0.0015)
$\lambda_{\chi 1}$	0.1961*** (0.0548)	0.3781*** (0.1061)	0.1424 (0.1686)
$\lambda_{\chi 2}$	-0.1042** (0.0521)	-0.4271** (0.1858)	-0.0849 (0.1778)
σ_η	0.0045*** (0.0000)	0.0684*** (0.0005)	0.0055*** (0.0000)
Log-likelihood	35575	18106	57240
AIC	35549	18080	57214
SIC	35493	18031	57156
Number Obs.	537	330	654

TABLE 7

IN-SAMPLE PREDICTIVE ABILITY OF THE THREE-FACTOR MODEL

The Table shows the mean error (real minus predicted) and the root mean squared error (RMSE) obtained with the three-factor model, for the three commodity prices series, using the parameter estimates in Table 6.

BRENT			HENRY HUB			WTI		
Contract	Mean	RMSE	Contract	Mean	RMSE	Contract	Mean	RMSE
F1	0.0007	0.0395	F1	0.0163	0.0816	F1	0.0002	0.0406
F4	-0.0020	0.0321	F5	0.0120	0.0768	F4	-0.0007	0.0328
F7	-0.0009	0.0276	F9	-0.0004	0.0823	F7	0.0002	0.0270
F10	0.0001	0.0249	F13	-0.0010	0.0710	F10	0.0003	0.0241
F12	0.0006	0.0235	F17	0.0006	0.0695	F13	0.0001	0.0222
F17	-0.0003	0.0222	F21	0.0005	0.0740	F16	0.0000	0.0210
F23	-0.0014	0.0215	F25	0.0033	0.0697	F19	-0.0001	0.0199
F34	0.0003	0.0198	F29	0.0052	0.0746	F22	-0.0003	0.0190
-	-	-	F33	0.0045	0.0690	F25	-0.0002	0.0185
-	-	-	F37	0.0037	0.0682	F28	0.0003	0.0185
-	-	-	F41	0.0016	0.0755	-	-	-
-	-	-	F45	-0.0010	0.0684	-	-	-
-	-	-	F48	-0.0055	0.0718	-	-	-

TABLE 8

PRINCIPAL COMPONENTS ANALYSIS

The Table shows the percentage of the empirical volatility explained by each principal component for each of the three commodities. The data set is composed of weekly observations of NYMEX WTI crude oil futures contracts from 9/18/1995 to 3/24/2008, NYMEX Henry Hub natural gas futures contracts from 12/3/2001 to 3/24/2008, and ICE Brent crude oil futures contracts from 12/15/1997 to 3/24/2008.

	Henry Hub	WTI	Brent
First Principal Component	61%	95%	94%
First Two Principal Components	75%	99%	99%
First Three Principal Components	80%	100%	99%

TABLE 9

CLASUSES VALUATION

The Table shows the differences in options value using the three factor model vs. the other two models (the AR(1) and the two factors ones) as percentage of the average price which is equal to the differences in valuation as percentage of the whole contract value.

WTI					
	Put Value			**Call Value**	
Minimum Price	Three Factor vs. AR(1)	Three Factor vs. Two Factors	Maximum Price	Three Factor vs. AR(1)	Three Factor vs. Two Factor
40	0.6%	0.2%	100	9.5%	1.3%
45	1.0%	0.3%	105	9.5%	1.3%
50	1.6%	0.4%	110	9.2%	1.3%
55	2.3%	0.5%	115	8.9%	1.3%
60	3.2%	0.7%	120	8.4%	1.3%
65	4.2%	0.8%	125	7.9%	1.3%
70	5.3%	0.9%	130	7.4%	1.2%
75	6.4%	1.0%	135	6.8%	1.2%
80	7.5%	1.1%	140	6.3%	1.2%
85	8.3%	1.2%	145	5.8%	1.1%
90	8.9%	1.2%	150	5.3%	1.1%
95	9.4%	1.3%	155	4.9%	1.0%

TABLE 9 (CONTINUATION)

Brent					
	Put Value			**Call Value**	
Minimum Price	Three Factors vs AR(1)	Three Factors vs Two Factors	Maximum Price	Three Factors vs AR(1)	Three Factors vs Two Factors
40	0.4%	-0.2%	100	8.4%	-1.4%
45	0.6%	-0.3%	105	8.3%	-1.4%
50	1.1%	-0.4%	110	8.0%	-1.4%
55	1.7%	-0.5%	115	7.6%	-1.4%
60	2.4%	-0.6%	120	7.2%	-1.3%
65	3.3%	-0.8%	125	6.7%	-1.3%
70	4.4%	-0.9%	130	6.1%	-1.3%
75	5.4%	-1.0%	135	5.6%	-1.2%
80	6.4%	-1.1%	140	5.1%	-1.2%
85	7.2%	-1.2%	145	4.6%	-1.1%
90	7.8%	-1.3%	150	4.1%	-1.1%
95	8.2%	-1.3%	155	3.7%	-1.0%

TABLE 9 (CONTINUATION)

HH					
	Put Value			**Call Value**	
Minimum Price	Three Factors vs AR(1)	Three Factors vs Two Factors	Maximum Price	Three Factors vs AR(1)	Three Factors vs Two Factors
40	0.8%	0.2%	100	14.4%	0.9%
45	1.4%	0.2%	105	14.4%	0.9%
50	2.3%	0.3%	110	14.1%	0.9%
55	3.4%	0.4%	115	12.9%	0.9%
60	4.8%	0.5%	120	12.2%	0.9%
65	6.4%	0.5%	125	11.5%	0.9%
70	8.1%	0.6%	130	10.0%	0.8%
75	9.8%	0.7%	135	9.3%	0.8%
80	11.3%	0.7%	140	8.6%	0.8%
85	12.6%	0.8%	145	8.0%	0.8%
90	13.6%	0.8%	150	7.4%	0.8%
95	14.2%	0.8%	155	6.4%	0.7%

TABLE 10

MINIMUM PRICES vs. MAXIMUM PRICE

The Table shows the differences in calculating the maximum price that the buyer should include to do not lose money if the seller includes a minimum price using the three factor model vs. the other two models (the AR(1) and the two factors ones) as percentage of the result obtained with the three factors model.

WTI		
Minimum Price	Three Factors vs. AR(1)	Three Factors vs. Two Factors
40	17.5%	3.5%
45	15.7%	3.1%
50	13.9%	2.7%
55	12.1%	2.3%
60	10.4%	2.0%
65	8.7%	1.6%
70	7.1%	1.3%
75	5.6%	1.0%
80	4.2%	0.7%
85	2.8%	0.5%
90	1.5%	0.2%
95	0.2%	0.0%

TABLE 10 (CONTINUATION)

Brent		
Minimum Price	Three Factors vs AR(1)	Three Factors vs Two Factors
40	14.4%	-3.0%
45	13.0%	-2.6%
50	11.5%	-2.3%
55	10.0%	-2.0%
60	8.6%	-1.7%
65	7.2%	-1.4%
70	5.9%	-1.2%
75	4.7%	-0.9%
80	3.5%	-0.7%
85	2.4%	-0.5%
90	1.3%	-0.3%
95	0.2%	0.0%

TABLE 10 (CONTINUATION)

HH		
Minimum Price	Three Factors vs AR(1)	Three Factors vs Two Factors
40	21.9%	3.1%
45	19.4%	2.6%
50	17.3%	2.2%
55	15.4%	1.8%
60	13.6%	1.4%
65	11.8%	1.2%
70	10.0%	0.9%
75	8.2%	0.7%
80	6.3%	0.5%
85	4.4%	0.3%
90	2.4%	0.2%
95	0.3%	0.0%

CHAPTER 2: THE STOCHASTIC SEASONAL BEHAVIOR OF THE NATURAL GAS PRICE

1. Introduction

In recent times, both academics and practitioners have been paying attention to the valuation and hedging of commodity contingent claims and to the procedures for evaluating natural resources investment projects, especially to the rule for determining when it is optimal to invest. The stochastic behavior of commodity prices plays a central role in this area.

Early studies on the stochastic behavior of commodity prices assumed that spot prices follow a geometric Brownian motion (see for example Brennan and Schwartz, 1985; Paddock et al., 1988, among others). However, the geometric Brownian motion hypothesis implies a constant rate of growth in the commodity price and a constant volatility of futures price returns, which are not realistic assumptions. In practice it is found that commodity prices show mean-reversion and the volatility of futures price returns is a decreasing function of time.

Consequently, in recent years several authors, such as Laughton and Jacobi (1993) and (1995), Ross (1997) or Schwartz (1997), have considered that a mean-reverting process is more appropriate to model the stochastic behavior of commodity prices. Unfortunately, these one-factor mean-reverting models are not very realistic since they generate a volatility of futures returns which goes to zero as the time to maturity of the futures contract approaches infinity. Even more, models considering a single source of uncertainty are not very realistic since they imply that futures prices for different maturities should be perfectly correlated, which defies existing evidence.

Looking for more realistic results, multi-factor models have been developed (Schwartz, 1997; Schwartz and Smith, 2000; Cortazar and Schwartz, 2003; Cortazar and Naranjo, 2006, among others). All these multi-factor models assume that the spot price is the sum of short-term and long-term components. Long-term factors account for the long-term dynamics of commodity prices, which is assumed to be a random walk, whereas short-term factors account for the mean-reversion components in the commodity price.

Most of these articles are focused on oil prices. In recent years, however, there have been many papers addressing the study of natural gas prices. See for example the papers by Clewlow and Strickland (2000), Wei and Zhu (2006) and Mu (2006) among others.

Natural gas represents almost the fourth part of the world energy consumption, with similar figures to coal and only behind oil. World natural gas consumption is about 45 millions barrels of oil equivalent per day while world oil consumption is about 80 millions barrels per day. World proved reserves are more or less the same for natural gas and oil which are roughly one trillion barrels of oil equivalent for each one.

Almost the third part of natural gas world consumption is located in the United States. In this country gas natural represents the 25% of the consumed energy and this percentage is growing very fast. This is the reason why the most developed gas natural markets are located in the United Sates. The lack of economical transportation makes the natural gas price substantially different along the country. The most liquid and famous natural gas market is located in Louisiana, near to the Texas border, which is named Henry Hub.

This lack of economical transportation and the limited storability of natural gas make its supply unable to change in view of seasonal variations of demand. Therefore, natural gas prices are strongly seasonal. One of the clearest ways to visualize this seasonality is through the forward curve. In Figure 1 it is possible to appreciate that Henry Hub natural gas prices are expected to be higher during winter months and lower during summer months.

It is also possible to notice that in spot price historical series the highest prices have been reached in winter while the lowest prices appear in summer.

There are studies taking into account the seasonal behavior of some commodity prices, such as Lucia and Schwartz (2002), Sorensen (2002), Tolmasky and Hindanov (2002) and Borovkova and Geman (2006) among others, but, to the best of our knowledge, seasonality has never been considered as an stochastic factor.

As pointed out by Schwartz (1997), the stochastic process assumed for the commodity price is important not only for derivatives valuation purposes, but also for the valuation of natural resource investment projects, specially for the rule for determining when it is optimal to invest.

In this chapter, it has been developed a general (n+2m)-factor model considering seasonality as an stochastic factor. This general (n+2m)-factor model assumes that the log-spot price is the sum of n and m stochastic factors (n non-seasonal and m seasonal). The non-seasonal factors are the factors of the models mentioned above. The seasonal factors are trigonometric components generated by stochastic processes. Then, this general model has been particularized for $m = 1$ and $n = 1, 2, 3$, thus, three, four and five-factor models have been obtained to explain the stochastic behavior of Henry Hub natural gas prices. The Kalman filter methodology has been applied to estimate the parameters of the models based on Henry Hub natural gas futures contracts traded at the New York Mercantile Exchange (NYMEX). Finally, using the estimated parameters, it is analyzed the models goodness of fit to the spot price dynamics and the term structure of futures prices and volatilities. Interestingly, it is found that models allowing for stochastic seasonality outperform standard models with deterministic seasonality.

This chapter is organized as follows. Section 2 deals with seasonality in Henry Hub natural gas prices. The model allowing for stochastic seasonality is developed in Section 3. The estimation methodology is discussed in Section 4. Sections 5 and 6 present the data and the empirical results regarding the estimation of the models. The goodness of fit of the models regarding the spot price, forward curve and volatility of futures returns estimations is contained in section 7.

Section 8 presents the results obtained for other commodities (RBOB gasoline and heating oil traded at NYMEX) and other markets (natural gas and gas oil traded at ICE Futures Europe, London). Finally, section 9 concludes with a summary and discussion.

2. Seasonality in Henry Hub Natural Gas Prices

Looking at Figure 1 it seems clear that Henry Hub natural gas prices are seasonal with a one year period. A very simple algorithm can be developed to understand it more clearly. Let S_t be the spot price and Y_t the centered moving average in a year of S_t defined as follows. If $\{S_t\}_{t=1,2,3,\dots}$ is the spot price time series with monthly frequency, then $Y_t = (0.5S_{t+6} + S_{t+5} + S_{t+4} + \dots + S_t + \dots + S_{t-5} + 0.5S_{t-6})/12$.

Let us define $\tau_t = S_t/Y_t$, which is a measure of how big is the spot price in month "*t*" with respect to the prices in one-year period centered in this month "*t*". If the price in this month "*t*" is higher than the price in the previous and following months, then τ_t will be grater than one, if not, τ_t will be less than one. Let us also define i_m as the average of τ_t for month "*m*" (*m* = January, February,…, December) and r_m, the scaling factor for month "*m*", as $r_m = i_m / \sqrt[12]{i_1 i_2 \dots i_{12}}$. It is obvious that $r_1 r_2 \dots r_{12} = 1$, and it is also easy to show that if r_m is greater than one, then the prices in month "*m*" are higher than the average price and if r_m is less than one, then the prices in month "*m*" are less than the average price.

As can be appreciated in Figure 2, the spot price and the forward curve scaling factors present the same pattern: In winter months they are higher than one and in summer months they are less than one. This is clear evidence of seasonality in the price of natural gas in Henry Hub.

A more sophisticated analysis can be implemented to complete this result. The spectrum of an stationary process is defined as the Fourier transformation of its autocovariance function. It can be proved (Wei, 2005) that deterministic seasonality appears in the spectrums as sharp peaks (actually Dirac delta functions) in several frequencies whereas stochastic seasonality shows a softer pattern.

This implies that in data analysis a sharp spike in the sample spectrum may indicate a possible deterministic cyclical component, while broad peaks often imply a nondeterministic seasonal component. Of course, the results should be taken with some care as estimation errors and aliasing effects can take place, confusing deterministic and stochastic "ideal" patterns.

It can be proved (Wei, 2005) that in a general stationary ARMA(p,q) model, $\psi_p(\mathrm{L})S_t = \theta_q(\mathrm{L})e_t$, where L is the lag function and e_t is white noise with variance σ_e^2, the spectrum is given by: $f(w) = \dfrac{\sigma_e^2}{2\pi}\left|\dfrac{\theta_q(e^{-iw})}{\psi_p(e^{-iw})}\right|^2$.

If it is assumed that Henry Hub spot natural gas prices follow an AR(1) with yearly stochastic seasonality, that is, $(1-\varphi\mathrm{L})(1-\Phi\mathrm{L}^{12})S_t = e_t$, its spectrum should be:

$$f(w) = \frac{\sigma_e^2}{2\pi}\frac{1}{(1+\Phi^2-2\Phi\cos(12w))(1+\varphi^2-2\varphi\cos(w))}$$

Thus, for $\Phi>0$, other than a peak at $w = 0$, the spectrum also exhibits peaks and troughs at seasonal harmonic frequencies $w = 2\pi k/12$ and $w = \pi(2k-1)/12$ respectively (k = 1, 2, 3, 4, 5 and 6). The Henry Hub natural gas price spectrum is depicted in Figure 3. It seems that, more or less, the spectrum exhibits peaks and troughs at those frequencies.

Therefore, this analysis suggests that natural gas in Henry Hub price has a yearly stochastic seasonal component.

3. A model for stochastic seasonality

As mentioned above, in this section it is presented the general n+2m-factor model. This model assumes that the log-spot price (X_t) is the sum of $n+m$ stochastic factors (n non-seasonal and m seasonal):

$$X_t = \xi_t + \sum_{i=1}^{n-1} \chi_{it} + \sum_{j=1}^{m} \alpha_{jt} \qquad (1)$$

The non-seasonal factors (ξ_t and χ_{it}) are the same that Cortazar and Naranjo (2006) use in their paper. Their stochastic differential equations (SDE) are:

$$d\xi_t = \mu_\xi dt + \sigma_\xi dW_{\xi_t} \qquad (2)$$

$$d\chi_{it} = -\kappa_i \chi_{it} dt + \sigma_{\chi_i} dW_{\chi_{it}} \qquad i = 1, 2, 3,..., n-1 \qquad (3)$$

where μ_ξ, κ_i, σ_ξ and $\sigma_{\chi i}$ are constants and $dW_{\xi t}$ and $dW_{\chi it}$ are correlated Brownian motions increments.

Each seasonal factor is modeled through a trigonometric component. The trigonometric SDE is complex:

$$da_{jt} = -i2\pi\varphi_j a_{jt} dt + Q_{\alpha j} dW_{ajt}$$

where a_{jt} is a complex factor ($a_{jt} = \alpha_{jt} + i\alpha_{jt}^*$), $Q_{\alpha j}$ is a complex number ($Q_{\alpha j} = Q_{\alpha j1} + iQ_{\alpha j2}$) and W_{ajt} a complex Brownian motion ($W_{ajt} = W_{\alpha jt} + iW_{\alpha^* jt}$), provided that $d\alpha_{jt}$ and $d\alpha_{jt}^*$ are uncorrelated and with the same variance (Oksendal, 1992).

To get it, a necessary and sufficient condition is to assume that $dW_{\alpha jt}$ and $dW_{\alpha^* jt}$ are uncorrelated. A proof of this fact can be found in appendix A. In appendix A it is also proved that the argument of $Q_{\alpha j}$ ($Q_{\alpha j}$ expressed in polars is $Q_{\alpha j} = e^{i\theta_j}\sigma_{\alpha j}$) has no effect in the model once expressed using only real numbers, this is to say equalling components in the previous equation. This is the reason why θ_j is chosen equal to zero and, consequently, $Q_{\alpha j} = \sigma_{\alpha j}$. Therefore the last SDE can be written as:

$$da_{jt} = -i2\pi\varphi_j a_{jt} dt + \sigma_{aj} dW_{ajt}$$

Equalling components in the previous equation yields two real SDEs for each seasonal factor:

$$d\alpha_{jt} = 2\pi\varphi_j \alpha_{jt}^{*} dt + \sigma_{\alpha j} dW_{\alpha jt} \tag{4}$$

$$d\alpha_{jt}^{*} = -2\pi\varphi_j \alpha_{jt} dt + \sigma_{\alpha j} dW_{\alpha^* jt} \qquad j = 1, 2, 3, ..., m \tag{5}$$

where $W_{\alpha jt}$ and $W_{\alpha * jt}$ are uncorrelated.

To assess derivatives contracts the "risk-neutral" version of the model has to be used. The SDEs for the factors under the equivalent martingale measure can be expressed as:

$$d\xi_t = (\mu_\xi - \lambda_\xi) dt + \sigma_\xi dW_{\xi t}^{\lozenge} \tag{6}$$

$$d\chi_{it} = (-\kappa_i \chi_{it} - \lambda_{\chi i}) dt + \sigma_{\chi i} dW_{\chi it}^{\lozenge} \qquad i = 1, 2, 3, ..., n\text{-}1 \tag{7}$$

$$d\alpha_{jt} = (2\pi\varphi_j \alpha_{jt}^{*} - \lambda_{\alpha j}) dt + \sigma_{\alpha j} dW_{\alpha jt}^{\lozenge} \tag{8}$$

$$d\alpha_{jt}^{*} = (-2\pi\varphi_j \alpha_{jt} - \lambda_{\alpha^* j}) dt + \sigma_{\alpha j} dW_{\alpha^* jt}^{\lozenge} \qquad j = 1, 2, 3, ..., m \tag{9}$$

where λ_ξ, $\lambda_{\chi i}$, $\lambda_{\alpha j}$ y $\lambda_{\alpha * j}$ are each factor "risk-premia" and $W_{\xi t}^{\lozenge}$, $W_{\chi t}^{\lozenge}$, $W_{\alpha jt}^{\lozenge}$ and $W_{\alpha * jt}^{\lozenge}$ are each factor Brownian motions under the equivalent martingale measure. It is admitted any correlation structure among Brownian motions with the restriction explained above ($W_{\alpha jt}$ and $W_{\alpha * jt}$ are uncorrelated). For each seasonal factor, both components (corresponding to real and imaginary parts in a complex process) should have equal variance and be uncorrelated.

General expressions for the price of a futures contract and for the volatility of futures returns can be found in Appendix B.

Next, the general model presented above will be particularized for n = 1, 2, 3 and m = 1. This is to say, we present three, four and five-factor models in order to explain the stochastic behavior of Henry Hub natural gas prices. Due to the analysis developed in Section 2, these particular models will have only one seasonal factor and it is expected that the estimated phase is one year (φ =1).

The non-seasonal part will be the same as in Schwartz (1997) for the three-factor model, Schwartz and Smith (2000) for the four-factor model and Schwartz and Cortazar (2003) for the five-factor model. Due to the seasonal factor, the Schwartz (1997) one-factor model becomes a three-factor model, the Schwartz and Smith (2000) two-factor model becomes a four-factor model and the Schwartz and Cortazar (2003) three-factor model becomes a five-factor model.

3.1. The Three-Factor Model

In this model it is assumed that the log-spot price (X_t) is the sum of two stochastic factors: a short-term component (χ_t) and a seasonal component (α_t), and a deterministic factor: the long-term component (ξ_t).

$$X_t = \xi_t + \chi_t + \alpha_t \tag{10}$$

The third stochastic factor is the other seasonal factor (α_t^*) which complements α_t.

The SDE of these factors are:

$$d\xi_t = \mu_\xi dt \tag{11}$$

$$d\chi_t = -\kappa\chi_t dt + \sigma_\chi dW_{\chi t} \tag{12}$$

$$d\alpha_t = 2\pi\varphi\alpha_t^* dt + \sigma_\alpha dW_{\alpha t} \tag{13}$$

$$d\alpha_t^* = -2\pi\varphi\alpha_t dt + \sigma_\alpha dW_{\alpha^* t} \tag{14}$$

Equation (12) is identical to equations (2) in Schwartz (1997).

The “risk-neutral” SDE are:

$$d\xi_t = \mu_\xi dt \tag{15}$$

$$d\chi_t = \left(-\kappa\chi_t - \lambda_\chi\right)dt + \sigma_\chi dW_{\chi t}^{\pm} \tag{16}$$

$$d\alpha_t = \left(2\pi\varphi\alpha_t^* - \lambda_\alpha\right)dt + \sigma_\alpha dW_{\alpha t}^{\pm} \tag{17}$$

$$d\alpha_t^* = \left(-2\pi\varphi\alpha_t - \lambda_{\alpha^*}\right)dt + \sigma_\alpha dW_{\alpha^* t}^{\pm} \tag{18}$$

Applying the result in Appendix B, expression (B4), the log-price of a futures contract with maturity at time “*T*+*t*” traded at time *t* is:

$$\ln\left[F(X_t, T+t)\right] = \xi_0 + e^{-kT}\chi_0 + \cos(2\pi\varphi T)\alpha_0 + sen(2\pi\varphi T)\alpha_0^* + A_3(T) \tag{19}$$

where:

$$\begin{aligned} A_3(T) = {} & (\mu_\xi + 0.5\sigma_\alpha^2)T - (1-e^{-kT})\lambda_\chi / k - (\lambda_{\alpha^*} + \lambda_\alpha sen(2\pi\varphi T) - \\ & -\lambda_{\alpha^*}\cos(2\pi\varphi T))/(2\pi\varphi) + \sigma_\chi^2(1-e^{-2kT})/(4k) + \\ & + \frac{\sigma_\chi\sigma_\alpha\rho_{\chi\alpha}\left\{k + e^{-kT}\left(2\pi\varphi sen(2\pi\varphi T) - k\cos(2\pi\varphi T)\right)\right\}}{k^2 + (2\pi\varphi)^2} + \\ & + \frac{\sigma_\chi\sigma_{\alpha^*}\rho_{\chi\alpha^*}\left\{2\pi\varphi - e^{-kT}\left(2\pi\varphi\cos(2\pi\varphi T) + k sen(2\pi\varphi T)\right)\right\}}{k^2 + (2\pi\varphi)^2} \end{aligned} \tag{20}$$

It is important to note that the trigonometric terms in the expressions above try to capture the seasonality in the forward curve.

Particularizing equation (B5) in Appendix B, it is determined the squared volatility of futures returns implied by this model:

$$\sigma_{F3}^2(T) = \sigma_\chi^2 e^{-2kT} + \sigma_\alpha^2 + 2e^{-kT}\sigma_\chi\sigma_\alpha\rho_{\chi\alpha}\cos(2\pi\varphi T) + 2e^{-kT}\sigma_\chi\sigma_\alpha\rho_{\chi\alpha^*} sen(2\pi\varphi T) \tag{21}$$

As can be appreciated in the previous equation, an interesting fact of this model with respect to the one factor model in Schwartz (1997) is the fact that in this case the volatility of futures returns does not go to zero as the time to maturity of futures contract approaches infinity, which is an undesirable property. It happens because seasonal factors are long-term factors.

Other interesting fact to take into account is the presence of seasonality in the volatility of futures returns. In this model seasonality disappears when the time to maturity of futures contracts approaches infinity. It happens because seasonality comes from the correlation between seasonal and non-seasonal factors. In this model, however, there is only one non-seasonal stochastic factor, which is a short-term factor. In following sections this facts are going to be discussed again.

3.2. The Four-Factor Model

In this model the log-spot price (X_t) is the sum of three stochastic factors: a long-term component (ξ_t), a short-term component (χ_t) and a seasonal component (α_t).

$$X_t = \xi_t + \chi_t + \alpha_t \tag{22}$$

The fourth stochastic factor is the other seasonal factor (α_t^*) which complements α_t.

The SDE of these factors are:

$$d\xi_t = \mu_\xi dt + \sigma_\xi dW_{\xi t} \tag{23}$$

$$d\chi_t = -\kappa\chi_t dt + \sigma_\chi dW_{\chi t} \tag{24}$$

$$d\alpha_t = 2\pi\varphi\alpha_t^* dt + \sigma_\alpha dW_{\alpha t} \tag{25}$$

$$d\alpha_t^* = -2\pi\varphi\alpha_t dt + \sigma_\alpha dW_{\alpha^* t} \tag{26}$$

Equations (23) and (24) are identical to equations (2) and (1) respectively in Schwartz and Smith (2000).

The "risk-neutral" SDE are:

$$d\xi_t = \mu_\xi' dt + \sigma_\xi dW_{\xi t} \tag{27}$$

$$d\chi_t = \left(-\kappa\chi_t - \lambda_\chi\right)dt + \sigma_\chi dW_{\chi t} \tag{28}$$

$$d\alpha_t = \left(2\pi\varphi\alpha_t^* - \lambda_\alpha\right)dt + \sigma_\alpha dW_{\alpha t} \tag{29}$$

$$d\alpha_t^* = \left(-2\pi\varphi\alpha_t - \lambda_{\alpha^*}\right)dt + \sigma_\alpha dW_{\alpha^* t} \tag{30}$$

where $\mu_\xi' = \mu_\xi - \lambda_\xi$ is the "risk-neutral" drift.

As before, the log-price of a futures contract with maturity at time "*T+t*" traded at time *t* can be calculated applying the result in Appendix B, expression (B4):

$$\ln\left[F(X_t, T+t)\right] = \xi_0 + e^{-kT}\chi_0 + \cos(2\pi\varphi T)\alpha_0 + sen(2\pi\varphi T)\alpha_0^* + A_4(T) \tag{31}$$

where:

$$\begin{aligned}
A_4(T) &= (\mu_\xi' + 0.5\sigma_\xi^2 + 0.5\sigma_\alpha^2)T - (1-e^{-kT})\lambda_\chi / k - \\
&- (\lambda_{\alpha^*} + \lambda_\alpha sen(2\pi\varphi T) - \lambda_{\alpha^*}\cos(2\pi\varphi T))/(2\pi\varphi) + \\
&+ \sigma_\xi\sigma_\chi\rho_{\xi\chi}(1-e^{-kT})/k + \frac{\sigma_\xi\sigma_\alpha\rho_{\xi\alpha}\{1-\cos(2\pi\varphi T)\} + \sigma_\xi\sigma_{\alpha*}\rho_{\xi\alpha*}sen(2\pi\varphi T)}{2\pi\varphi} + \\
&+ 0.25\sigma_\chi^2(1-e^{-2kT})/k + \frac{\sigma_\chi\sigma_\alpha\rho_{\chi\alpha}\left\{k - e^{-kT}\left(k\cos(2\pi\varphi T) + 2\pi\varphi sen(2\pi\varphi T)\right)\right\}}{k^2 + (2\pi\varphi)^2} + \\
&\frac{\sigma_\chi\sigma_{\alpha*}\rho_{\chi\alpha*}\left\{2\pi\varphi + e^{-kT}\left(ksen(2\pi\varphi T) - 2\pi\varphi\cos(2\pi\varphi T)\right)\right\}}{k^2 + (2\pi\varphi)^2}
\end{aligned} \tag{32}$$

And the squared volatility of futures returns is:

$$\begin{aligned}
\sigma_{F4}^2(T) &= \sigma_\xi^2 + \sigma_\chi^2 e^{-2kT} + \sigma_\alpha^2 + 2e^{-kT}\sigma_\xi\sigma_\chi\rho_{\xi\chi} + 2\sigma_\xi\sigma_\alpha\rho_{\xi\alpha}\cos(2\pi\varphi T) + \\
&+ 2\sigma_\xi\sigma_\alpha\rho_{\xi\alpha*}sen(2\pi\varphi T) + 2e^{-kT}\sigma_\chi\sigma_\alpha\rho_{\chi\alpha}\cos(2\pi\varphi T) + 2e^{-kT}\sigma_\chi\sigma_\alpha\rho_{\chi\alpha*}sen(2\pi\varphi T)
\end{aligned} \tag{33}$$

Given that in this model there is a long-term stochastic factor, seasonality in the volatility of futures returns does not disappear when the time to maturity of futures contracts approaches infinity.

3.3. The Five-Factor Model

In this model the log-spot price (X_t) is the sum of four stochastic factors: a long-term component (ξ_t), two short-term components (χ_{1t} and χ_{2t}) and a seasonal component (α_t).

$$X_t = \xi_t + \chi_{1t} + \chi_{2t} + \alpha_t \tag{34}$$

The fifth stochastic factor is the other seasonal factor (α_t^*) which complements α_t.

The SDE of these factors are:

$$d\xi_t = \mu_\xi dt + \sigma_\xi dW_{\xi t} \tag{35}$$

$$d\chi_{1t} = -\kappa_1 \chi_{1t} dt + \sigma_{\chi 1} dW_{\chi 1t} \tag{36}$$

$$d\chi_{2t} = -\kappa_2 \chi_{2t} dt + \sigma_{\chi 2} dW_{\chi 2t} \tag{37}$$

$$d\alpha_t = 2\pi\varphi\alpha_t^* dt + \sigma_\alpha dW_{\alpha t} \tag{38}$$

$$d\alpha_t^* = -2\pi\varphi\alpha_t dt + \sigma_\alpha dW_{\alpha^* t} \tag{39}$$

In Appendix C.1 it can be seen that the non-seasonal part of this model is equivalent to the three factor model proposed in Cortazar and Schwartz (2003).

The “risk-neutral” SDE are:

$$d\xi_t = \mu_\xi ' dt + \sigma_\xi dW_{\xi t} \tag{40}$$

$$d\chi_{1t} = \left(-\kappa \ \ \chi_{1t} - \lambda_{\chi 1}\right) dt + \sigma_{\chi 1} dW_{\chi 1t} \tag{41}$$

$$d\chi_{2t} = \left(-\kappa_2 \chi_{2t} - \lambda_{\chi 2}\right) dt + \sigma_{\chi 2} dW_{\chi 2t} \tag{42}$$

$$d\alpha_t = \left(2\pi\varphi\alpha_t^* - \lambda_\alpha\right)dt + \sigma_\alpha dW_{\alpha t} \tag{43}$$

$$d\alpha_t^* = \left(-2\pi\varphi\alpha_t - \lambda_{\alpha^*}\right)dt + \sigma_\alpha dW_{\alpha^* t} \tag{44}$$

where $\mu_\xi' = \mu_\xi - \lambda_\xi$ is the "risk-neutral" drift.

The log-price of a futures contract with maturity at time "*T*+*t*" traded at time *t* and the squared volatility of futures returns can be calculated in the same way as in the three and four-factor models. Their complex expressions are given in Appendix C.2.

4. Estimation methodology

As stated in previous studies, one of the main difficulties in estimating the parameters of the model is the fact that the factors (or state variables) are not directly observable and must be estimated from spot and/or futures prices. Intuitively, the non-seasonal factors (long term and short-term factors) are going to be estimated based on the relationship between long-maturity futures and short-maturity futures (or spot prices) and the seasonal factors are going to be estimated through the relationship between futures contracts maturing in different months.

The formal way to do this is through the Kalman filtering methodology. This methodology enables the calculation of the likelihood of a data series given a particular set of model parameters and a prior distribution of the variables which permits the estimation of the parameters using maximum likelihood techniques. Detailed accounts of Kalman filtering are given in Harvey (1989).

The Kalman filter methodology is a recursive methodology that estimates the unobservable time series, the state variables or factors (Z_t), based on an observable time series (Y_t) which depends on these state variables.

In the traditional version of this methodology, two conditions need to be fulfilled. First, no missing points in the data set. Second, the length of vector Y_t must be independent of "*t*". An improved version of this methodology has been developed in Cortazar and Naranjo (2006) to handle with incomplete data sets and vectors Y_t whose length depends on "*t*". The problem using this methodology with a data set with a lot of missing points in the futures contracts with higher maturities is the unbalance in the relationship between long and short effects and, in the case handled in this work, the unbalance in the relationship between futures contracts whose maturity occurs in different months, which accounts for the seasonal effects. As explained below, taking into account these considerations and other issues related to data liquidity, the data set used in this study has not missing points and all vectors have the same length. Therefore in this chapter the traditional version of the Kalman filter methodology is used.

In Cortazar and Schwartz (2003) an alternative to the Kalman filter methodology has been developed to estimate the model parameters and the state variables. This technique is a simple one which only needs a spreadsheet to be implemented.

To estimate the parameters of the models through the Kalman filter methodology, or through the Cortazar and Schwartz (2003) technique, we need a discrete-time version of the models.

As stated in Appendix B, the solution to the general problem of this chapter is (B1) and Z_t is Gaussian with mean and variance given by expressions (B2) and (B3) respectively. Thus, if the difference between the current period and the initial period is one period time, Z_t follows the discrete process:

$$Z_t = c_t + TZ_{t-1} + \psi_t \qquad\qquad t = 1, \ldots, N_t \qquad (45)$$

where $c_t = e^{At}\int_{t-1}^{t} e^{-As} b\,ds \in \Re^h$, $T = e^{A} \in \Re^{hxh}$ and $\psi_t \in \Re^h$ is a vector of serially uncorrelated Gaussian disturbances with zero mean and covariance matrix $Q = \left(e^{A}\right)\left[\int_{t-1}^{t} e^{-As} R \left(e^{-As}\right)^T ds\right]\left(e^{A}\right)^T$. This equation will be called, following standard conventions in the literature, the *transition equation* of each model.

The *measurement equation* is just the expression of the log-futures prices (Y_t) in terms of the factors (Z_t) by adding serially uncorrelated disturbances with zero mean (η_t) to take into account measurement errors derived from bid-ask spreads, price limits, non-simultaneity of observations, errors in the data, etc. To avoid dealing with a great amount of parameters, the covariance matrix H_t will be assumed diagonal with all its diagonal elements being equal. This simple structure for the measurement errors is imposed so that the serial correlation and cross correlation in the log-prices is attributed to the variation of the unobservable state variables. The measurement equation will be expressed as:

$$Y_t = d_t + M_t Z_t + \eta_t \qquad \text{t = 1, …, } N_t \qquad (46)$$

where $Y_t, d_t \in \Re^n, M_t \in \Re^{n x n}, \ Z_t \in \Re^h$ (h is the number of state variables, or factors, in the model) and $\eta_t \in \Re^n$ is a vector of serially uncorrelated Gaussian disturbances with zero mean and covariance matrix H_t.

The specific transition and measurement equations for the particular models considered in this study (three, four and five factor models) are derived in Appendix D.

Let $Y_{t|t-1}$ be the conditional expectation of Y_t and let Ξ_t be the covariance matrix of Y_t conditional on all information available at time t – 1. Then the log-likelihood function can be expressed as (after omitting unessential constants)[9]

$$l = -\sum_t \ln|\Xi_t| - \sum_t (Y_t - Y_{t|t-1})'\Xi_t^{-1}(Y_t - Y_{t|t-1}) \qquad (47)$$

5. Data

The data set employed in the estimation procedure consists of weekly observations of Henry Hub natural gas futures prices traded at NYMEX[10]. Four different data sets are used in the

[9] See Harvey (1989) for a detailed discussion of the estimation of a model in state space form by maximum likelihood.

estimation procedure. In all cases nine futures contracts (i.e. n = 9) have been used. The three first ones data sets (data sets 1, 2 and 3) are made of contracts F1, F5, F9, F14, F18, F22, F27, F31 and F35 where F1 is the contract closest to maturity, F2 is the second contract closest to maturity and so on. Data set 1 contains 438 quotations of each contract from 09/01/1997 to 01/16/2006, data set 2 contains 222 quotations from 09/01/1997 to 11/26/2001, and data set 3 contains 216 quotations from 12/03/2001 to 01/16/2006. The last data set (data set 4) entails contracts F1, F8, F15, F22, F29, F36, F43, F50 and F57 from 12/03/2001 to 01/16/2006 (216 quotations).

As explained in Schwartz (1997), since futures contracts have a fixed maturity date, the time to maturity changes as time progresses, but remaining in a narrow time interval. This is the reason why, as in Schwartz (1997), it is assumed that the time to maturity does not change with time and it is equal to one month for F1, to two months for F2 and so on.

There are currently 72 contracts traded for different maturities ranging from 1 to 72 months. However, until December 2001 the maximum maturity traded at NYMEX was only 36 months. In this date new contracts were introduced to include maturities up to 72 months, but the liquidity of these new contracts is relatively low and in recent times there is not quotation for the contracts with the latest maturities in some dates. Before September 1997 there were also liquidity problems with the contracts with the latest maturities and, to be precise, there is not quotation for the latest maturities ones (from F20 to F36) in many dates.

As stated above, it is expected that the seasonal factor in the particular models has one year period. Thus, the total number of futures contracts, as long as the number of contracts with different maturities needed in the present study, is considerable higher than in previous studies. Specifically, we need futures contracts with more than one year to maturity, contracts with different maturities and as many contracts as possible to account for it. Futures contracts with long-term and short-term maturities are also necessary to estimate properly the parameters of

[10] The NYMEX is the biggest market for natural gas.

the non-seasonal factors (long-term and short-term factors). Moreover, the higher the number of contracts used in the estimation, the more precise the estimates of the parameters.

Nevertheless, there are also other considerations to take into account. As explained above, in order to estimate properly the relationship between long and short-term effects and the parameters in the seasonal process, it is desirable that the data set includes more or less the same quotations for all futures contracts.

Moreover, to account for structural changes in the natural gas price dynamics it is desirable to consider different data sets with different sample periods.

Taking into account all these arguments, it has been defined the sets of futures contracts specified above. Nonetheless, the estimation has been repeated using different data sets (data sets with more futures contracts and with future contracts with other maturities) and the results are more or less the same than those presented in this chapter.

NYMEX quotations are market prices, that is, market prices at NYMEX arise as a result of matching bid and ask orders. However, there are not Henry Hub natural gas spot prices. Some institutions, such as Bloomberg, provide natural gas in Henry Hub spot prices asking the participants in the market for their best estimation and applying an internal procedure. Therefore, spot prices are not properly market prices.

Table I contains the main descriptive statistics of all NYMEX data series employed in this study. In all cases, units are \$/MMBtu[11].

---

[11] "MMBtu" means "millions of British thermal units", which is an energy measure. US \$ per MMBtu is the accepted way to represent the natural gas price (in MYMEX, natural gas quotes in US \$ per MMBtu). The energy contained in a crude oil barrel depends on the kind of the oil, but broadly speaking an oil barrel contains something less than six MMBtu of energy.

6. Empirical Results

6.1. The Three-Factor Model

Assuming that the variance-covariance matrix of η_t is diagonal and all diagonal elements are the same, there are eleven parameters to be estimated in the three-factor model: μ_ξ, k, φ, σ_χ, σ_α, $\rho_{\chi\alpha}$, $\rho_{\chi\alpha^*}$, λ_χ, λ_α, λ_{α^*} and σ_η.

Table II presents the results for the three-factor model applied to the four data sets described above.

One interesting issue from these results is the fact that in all cases the seasonal period is more or less one year and the standard deviation of the seasonal factor (σ_α) is significantly different from zero. This implies that seasonality in Henry Hub natural gas prices is stochastic with one year period, which is consistent with the results in Section 2.

The speed of adjustment (*k*) is highly significant which implies, as in the case of oil (Schwartz, 1997), mean reversion in the natural gas price. The market prices of risk, the λ´s, are not significantly different from zero in most of the cases. The long-term trend (μ_ξ) is positive and significantly different from zero in all cases, implying long-term growth in the natural gas price.

As can be appreciated in the table, it has been obtained more or less the same results with the first three data sets. The results with the last data set are slightly different. Specifically, there is significantly less mean reversion, the volatility of future returns, which is calculated by substituting the parameters is equation (21), is also significantly lower, the price of risk for the sort-term factor (χ) is highly significant and the long-term trend (μ_ξ) is also different. This fact is also a first sign that more structure is needed. Finally, the values of the Akaike and Schwartz Information Criteria (AIC and SIC respectively) are shown at the bottom of the table. These values will be useful for comparing the models (section 7).

As stated above, the seasonal factor is a long-term factor, but, as it will be discussed below, it is unable to capture all stochastic long-term effects present in the natural gas price. Therefore, at least one stochastic long-term factor is needed.

6.2. The Four-Factor Model

Assuming that the variance-covariance matrix of η_t is diagonal and all diagonal elements are the same, there are sixteen parameters to be estimated in the four-factor model: μ_ξ, k, φ, σ_ξ,, σ_χ, σ_α, $\rho_{\xi\chi}$, $\rho_{\xi\alpha}$, $\rho_{\xi\alpha^*}$, $\rho_{\chi\alpha}$, $\rho_{\chi\alpha^*}$, μ_ξ', λ_χ, λ_α, λ_{α^*} and σ_η.

Table III presents the results for the four-factor model applied to the data sets described above.

As in the previous model, in all cases the seasonal period is one year and the standard deviation of the seasonal factor (σ_α) is significantly different from zero although its magnitude is lower than in the three-factor model. This is due to the fact that in the three-factor model the seasonal factor captures stochastic long-term effects which are captured by the stochastic long-term factor in the four-factor model.

The speed of adjustment (k) is higher than in the previous model, implying more mean reversion, probably due to the fact that the short-term factor in the three-factor model was capturing long-term stochastic effects, whereas the parameters of the new stochastic factor, the long-term one, are also highly significant, which confirms the previous perception that a stochastic long-term factor is needed.

Moreover, the market prices of risk, the λ´s, are not significantly different from zero in most of the cases, the long-term trend is positive and the long-term trend adjusted by risk (μ_ξ') is negative and in both cases significantly different from zero.

In this case the results are more or less the same in all data sets. The only difference which remains is that in the fourth data set, which uses futures with higher maturities, there is less mean reversion. This difference is consistent with the results obtained by Schwartz (1997), since the fourth data set contains futures contracts with higher maturities than those in the first three data sets. Like in the oil case, Schwartz (1997), it has important implications in valuation and hedging natural gas contingent claims or in investment decisions, since it is necessary to account for the payments term structure when choosing the futures contracts to estimate the model parameters.

6.3. The Five-Factor Model

Assuming that the variance-covariance matrix of η_t is diagonal and all diagonal elements are the same, there are twenty three parameters to be estimated in the five-factor model: μ_ξ, k_1, k_2, φ, σ_ξ, $\sigma_{\chi 1}$, $\sigma_{\chi 2}$, σ_α, $\rho_{\xi\chi 1}$, $\rho_{\xi\chi 2}$, $\rho_{\xi\alpha}$, $\rho_{\xi\alpha*}$, $\rho_{\chi 1\,\chi 2}$, $\rho_{\chi 1\alpha}$, $\rho_{\chi 1\alpha*}$, $\rho_{\chi 2\alpha}$, $\rho_{\chi 2\alpha*}$, μ_ξ', $\lambda_{\chi 1}$, $\lambda_{\chi 2}$, λ_α, $\lambda_{\alpha*}$ and σ_η.

Table IV presents the results for the five-factor model applied to the data sets described above.

The results are quite close to the four-factor model ones. In this model there are two short-term factors whose parameters are highly significant, and in all cases one of these factors has higher speed of adjustment than the other. This means that there are two types of stochastic short-term effects, one (the one with higher k) with stronger mean reversion than the other (the one with smaller k), and both of them significant. It is interesting to note that, in all cases, the price of risk for the short-term factor with stronger mean reversion is highly significant.

As the non-seasonal part of this model is equivalent to the Schwartz and Cortzar (2003) one (appendix C.1), the fact that both sort-term factors are significant implies that Henry Hub natural gas prices long-term drift is stochastic.

7. Comparing the models

In this section the relative performance of the models will be compared. Specifically, we will analyze the in-sample and out-of-sample performance, and the goodness of fit for the spot price, the forward curve and the volatility of futures returns. The analysis of the goodness of fit for the spot price and the forward curve will be developed using the first data set because it is the biggest one, whereas the goodness of fit for the volatility will be developed using the fourth one, to account for volatilities of futures with maturities higher than 36 months. Nevertheless, using whatever other data set the results are broadly the same.

The results will be also compared with those obtained with the Schwartz and Smith (2000) two-factor model with deterministic seasonality, which is Sorensen (2002) proposal.

7.1 Deterministic versus stochastic seasonality

Before comparing the relative performance of the three models with stochastic seasonality presented in this chapter, it could be useful to compare the results obtained with the models with stochastic seasonality with those obtained with the standard models with deterministic seasonality as in Sorensen (2002).

Table V presents the results of the estimation of the Schwartz and Smith (2000) two-factor model with deterministic seasonality, which will be compared with those obtained with the four-factor model with stochastic seasonality (Table III). In order to save space only the comparison of the results obtained with the Schwartz and Smith (2000) two-factor model with deterministic seasonality and the four-factor model with stochastic seasonality are presented, although similar conclusions are obtained with other models. As it is going to be pointed out below, the four-factor model is much more accurate to capture Henry Hub natural gas prices dynamics than the three-factor one and simpler than the five-factor one. This is the reason why the four-factor model is the one chosen for the comparison.

First of all, it is interesting to compare the value of the Schwartz Information Criterion (SIC) obtained with both models. If we define the SIC as $\ln(L_{ML}) - q\ln(T)$, where q is the number of estimated parameters, T is the number of observations and L_{ML} is the value of the likelihood function, defined in (47), using the q estimated parameters, then the preferred model is the one with the highest SIC. It is found that the value of the SIC for the four-factor model with stochastic seasonality, shown at the bottom of Table III, is higher than the corresponding value obtained with the standard model with deterministic seasonality, shown at the bottom of Table

V. The same conclusions are obtained with the Akaike Information Criterion (AIC), which is defined as $\ln(L_{ML}) - 2q$.

A second way of comparing the models is though their predictive ability. The in-sample predictive ability of the Schwartz and Smith (2000) two-factor model with deterministic seasonality and the four-factor model with stochastic seasonality is presented in Table VI, using the whole data set, i. e. data set 1. It is found that, in general, the model accounting for stochastic seasonality outperforms the standard model with deterministic seasonality. The advantages of the stochastic seasonality model over the deterministic seasonality one are even clearer when we analyze the out-of-sample predictive ability (Table VII). The out-of-sample results are obtained valuing the contracts in data set 3 with the parameters obtained estimating the models with data set 2. As expected, out of sample pricing errors are slightly higher than the corresponding in-sample values.

7.2 Spot Price

Following a procedure close to the Kalman filter technique it has been obtained factors estimations in each time based on the information available until this time, for each model and for the parameters estimated with each data set. The actual spot price estimation, provided by Bloomberg, and the spot price estimated by each model for the first data set, together with the deviation from the actual spot price of the spot prices estimated by each model are depicted in Figure 4.

In Figure 4 it is possible to notice that the five-factor model is the model with the best performance, the four-factor model the second one and the three-factor model is the model with the worst performance of the models presented in this work, in estimating the actual spot price. The standard deviation of the deviation from actual spot price is 9% for the five-factor model, 9.6% for the four-factor model and 13.1% for the three-factor model.

One interesting fact, which has been observed in the previous section, is the difference in the estimations obtained with the three-factor model and with the four and five-factor models. The estimations of the last two models are much better than the three-factor model ones. It confirms that a long-term stochastic factor is needed to understand the natural gas price dynamics. The five-factor model estimations are better than the four-factor model ones, but both of them are close and highly accurate.

7.3 Forward Curve

Proceeding in a similar manner as in the previous sub-section and using the first data set as well, it is also possible to obtain estimations of the forward curve. The actual forward curve and the one estimated by the models for a randomly chosen date are depicted in Figure 5.

As previously obtained in the case of the goodness of fit for the spot price, the model with the best fit to actual data is the five-factor model, the second one is the four-factor model and so on. Moreover, the estimations for the five and four-factor models are close and are much better than the three-factor model ones.

7.4 Volatility of Futures Returns

The volatility of futures returns can be calculated by substituting the estimated parameters presented section 6 in the corresponding formulas developed in section 3. The volatility of futures returns estimated by each model compared with the actual one are depicted in Figure 6. In this case it has been used the fourth data set because with the first one it is not possible to calculate the actual volatility for maturities higher than 36 months. Anyway, the conclusion of the analysis will be the same if the first data set were used.

The comments in the previous sub-sections also apply here. One interesting issue to note is that the actual natural gas price volatility is seasonal, oppositely to oil. As stated in section 3, in the models proposed in this chapter, natural gas price volatility is also seasonal because the seasonal factors are stochastic. With deterministic seasonal factors, as in Sorensen (2002), it is not possible to get seasonal volatility. Therefore, the seasonal volatility in actual data is a new evidence in favor of using stochastic seasonal factors instead of deterministic ones, in explaining natural gas prices dynamics.

In the case of the three-factor model, seasonality decreases when futures maturity grows, going to zero when maturity goes to infinity, whereas it does not in the case of the four and five-factor models. Looking at Figure 6 it seems that the seasonality in the actual volatility does not decrease, which is a new evidence in favor of the four and five-factor models and against the three-factor one. Moreover, in the case of the three-factor model it is not possible to appreciate the seasonality even for short maturity contracts, because in this case σ_α is much smaller than σ_χ.

8. Other commodities and other markets

In this section we also apply our model with stochastic seasonality to other commodities. Specifically, we have investigated heating oil and RBOB gasoline futures contracts traded at NYMEX[12] and gas oil and natural gas futures contracts traded at ICE Futures Europe (London).

Tables VIII and IX present the results for heating oil futures contracts traded at NYMEX. Table VIII contains the results for the Schwartz and Smith (2000) two-factor model with deterministic seasonality and Table IX contains the results for the four-factor model with stochastic seasonality. Both models have been estimated with weekly observations, using contracts F1, F3, F5, F7, F10, F12, F14, F16 and F18. The whole sample period (data set 1) consists of 524 weekly observations from 09/09/1996 to 09/18/2006. Data sets 2 and 3 consists of weekly observations from 09/09/1996 to 09/10/2001 and from 09/17/2001 to 09/19/2006 respectively (262 observations each one). The results are similar to those obtained for the natural gas futures prices. Specifically, the long-term trend (μ_ξ) is positive and significant (except for data set 2), implying long-term growth in heating oil prices. It is worth noting that the long-term trend is higher in the second sub-period (data set 3), implying more long-term grown in the last years of the sample, which is consistent with the high growth experimented by oil products during the last years.

[12] It is worth noting that although crude oil prices do not exhibit seasonality, the main oil products are strongly seasonal. This is because seasonality in these products goes in opposite direction. Specifically, heating oil prices are higher in winter months, whereas gasoline prices are higher in summer months. The sum of these effects results in non-seasonality for oil prices.

The speed of adjustment (k) is significantly different from zero in all cases, implying mean reversion, which is consistent with our previous results for the natural gas and also with the results obtained by Schwartz (1997) in the case of oil. Moreover, as in the case of natural gas, the volatility of the seasonal factor (σ_α) is significantly different from zero (Table IX), implying that seasonality in heating oil prices is stochastic with one year period (Φ close to one). It is worth noting that the short-term volatility (σ_χ) is higher than the long-term volatility (σ_ξ). Also, it can be appreciated from Table IX that all three volatilities (σ_ξ,σ_χ and σ_α) are lower than those obtained in the natural gas case for the same four-factor model (Table III), and (σ_ξ and σ_χ) are similar to those presented in Schwartz-Smith (2000) for the case of oil. As before, the market prices of risk are not significantly different from zero in most of the cases.

Finally, it is found that the value of the SIC and AIC for our four-factor model (Table IX) are higher than the corresponding values obtained with the standard model with deterministic seasonality (Table VIII).

Table X presents the out of sample pricing errors results for heating oil futures prices (NYMEX). As in the natural gas case, these results are obtained valuing the contracts in data set 3 with parameters obtained estimating the models with data set 2. The results are quite similar to those obtained with the natural gas series. This is to say, the model accounting for stochastic seasonality outperforms the standard model with deterministic seasonality. Moreover, the (root mean squared) errors obtained with the heating oil series are lower than those obtained with the natural gas series.

The estimation results for RBOB gasoline (NYMEX) and natural gas and gas oil (ICE Futures Europe) are contained in Table XI. The data set for RBOB gasoline (NYMEX) is composed of weekly observations from 12/08/2003 to 01/30/2006, contracts F1, F3, F5, F7, F9 and F12 (113 observations). Due to liquidity constrains, the data set for natural gas (ICE futures) is composed of weekly observations from 03/30/1998 to 09/25/2000, contracts F1, F3, F5, F7, F9, F11, F13 and F15 (131 observations). Finally, the data set for gas oil (ICE futures) is composed of weekly observations from 10/04/1999 to 09/18/2006 (364 observations). The results are quite similar to those obtained for natural gas and heating oil (NYMEX). From the point of view of this chapter goal, the most important fact is that all three commodities in Table XI show stochastic seasonality (σ_α is significantly different from zero), with one year period (Φ close to one). This confirms that stochastic seasonality is commonly observed in commodity futures markets. Consequently, stochastic seasonality is a fact that should be considered in a commodity futures valuation model.

9. Conclusions

Most studies on the stochastic behavior of commodity prices are focused on oil prices. The number of papers addressing the study of natural gas prices is still scarce. However natural gas represents almost the fourth part of the world energy consumption. The lack of economical transportation and the limited storability of natural gas make its supply unable to change in view of seasonal variations of demand. Therefore, natural gas price is strongly seasonal.

Analyzing the natural gas price spectrum, it seems highly probable that seasonality in natural gas price is an stochastic factor and not a deterministic one. However, to the best of the authors´ knowledge, seasonality has never been considered as an stochastic factor in previous studies. Therefore, in this chapter it has been developed a general n+2m-factor model of the stochastic behavior of commodity prices, considering seasonality as an stochastic factor. Then, this general model has been particularized for m = 1 and n = 1,2,3, thus, three, four and five-factor models have been obtained to explain the stochastic behavior of Henry Hub natural gas prices. The parameters of the models have been estimated through the Kalman filter methodology and using NYMEX data.

One of the main conclusions of this study is the confirmation of the fact that seasonality in the natural gas price is stochastic and not deterministic, as many studies assume. It is also found that the natural gas prices seasonal period is one year. Similar results are obtained with other commodities and with commodities traded in other markets. Specifically we have confirmed the presence of stochastic seasonality in the case of heating oil and RBOB gasoline futures contracts traded at NYMEX and natural gas and gas oil futures contracts traded at ICE Futures Europe (London).

The seasonal factors considered in this chapter have a long-term component, but it is demonstrated that this long-term component is unable to capture properly the natural gas long-term dynamics. Consequently, it is necessary to use a random walk as a long-term factor. Therefore the four and five-factor models are much better than the three-factor one in explaining the natural gas price behavior. The classical two-factor model fails in this task because it does not account for seasonality.

Consequently, the four and five-factor models presented in this work seem appropriate to value all natural gas contingent claims or investment projects. The five-factor model is better than the four-factor one although it needs more structure. Hence, the use of the four or the five-factor models will depend on the precision needed.

APPENDICES

Appendix A: Seasonal Factors

As stated above, the stochastic differential equation (SDE) for each seasonal factor is:

$$da_{jt} = -i\varphi_j a_{jt} dt + Q_{\alpha j} dW_{ajt}$$

where a_{jt} is a complex factor ($a_{jt} = \alpha_{jt} + i\alpha_{jt}^*$), $Q_{\alpha j}$ a complex number ($Q_{\alpha j} = Q_{\alpha j1} + iQ_{\alpha j2}$) and W_{ajt} a complex Brownian motion ($W_{ajt} = W_{\alpha jt} + iW_{\alpha * jt}$), provided that $d\alpha_{jt}$ and $d\alpha_{jt}^*$ are uncorrelated and with the same variance.

Equalling components in the previous equation yields two real SDEs:

$$d\alpha_{jt} = \varphi_j \alpha_{jt}{}^* dt + Q_{\alpha j1} dW_{\alpha jt} - Q_{\alpha j2} dW_{\alpha^* jt}$$

$$d\alpha_{jt}{}^* = -\varphi_j \alpha_{jt} dt + Q_{\alpha j2} dW_{\alpha jt} + Q_{\alpha j1} dW_{\alpha^* jt}$$

If $W_{\alpha jt}$ and $W_{\alpha^* jt}$ are uncorrelated, then $d\alpha_{jt}$ and $d\alpha^*_{jt}$ are also uncorrelated and with the same variance, which is $\sigma^2_{\alpha j} = Q^2_{\alpha j1} + Q^2_{\alpha j2}$, being $\sigma_{\alpha j}$ the complex number $Q_{\alpha j}$ module (the sufficiency condition).

On the other hand:

$$\operatorname{cov}(d\alpha_{jt}, d\alpha^*_{jt}) = \left(Q^2_{\alpha j1} - Q^2_{\alpha j2}\right)\operatorname{cov}(dW_{\alpha jt}, dW_{\alpha^* jt}),$$

$$Var(d\alpha_{jt}) = Q^2_{\alpha j1} dt + Q^2_{\alpha j2} dt - 2Q_{\alpha j1} Q_{\alpha j2} \operatorname{cov}(dW_{\alpha jt}, dW_{\alpha^* jt})$$

$$Var(d\alpha^*_{jt}) = Q^2_{\alpha j1} dt + Q^2_{\alpha j2} dt + 2Q_{\alpha j1} Q_{\alpha j2} \operatorname{cov}(dW_{\alpha jt}, dW_{\alpha^* jt})$$

To get $\operatorname{cov}(d\alpha_{jt}, d\alpha^*_{jt}) = 0$ and $Var(d\alpha_{jt}) = Var(d\alpha^*_{jt})$ it is necessary that $\operatorname{cov}(dW_{\alpha jt}, dW_{\alpha^* jt}) = 0$ (the necessary condition).

Let now $Q_{\alpha j}$ be expressed in polar coordinates $Q_{\alpha j} = e^{i\theta_j}\sigma_{\alpha j}$ where $\sigma_{\alpha j}$ is the module and θ_j the phase.

Defining $dW_{ajt}{}^{\times} = e^{i\theta_j} dW_{ajt}$, it is clear that $W_{ajt}{}^{\times} = \int_0^t e^{i\theta_j} dW_{ajs}$, or equivalently: $W_{ajt}{}^{\times} = e^{i\theta_j} W_{ajt}$. Equalling components in the last equation yields:

$$\begin{pmatrix} W_{\alpha jt}{}^{\times} \\ W_{\alpha^* jt}{}^{\times} \end{pmatrix} = \begin{pmatrix} \cos\theta_j & -sen\theta_j \\ sen\theta_j & \cos\theta_j \end{pmatrix} \begin{pmatrix} W_{\alpha jt} \\ W_{\alpha^* jt} \end{pmatrix}$$

From the previous equation, taking into account that $W_{\alpha jt}$ and $W_{\alpha^* jt}$ are uncorrelated, it follows that:

$$Var\begin{pmatrix} W_{1t}^{\times} \\ W_{2t}^{\times} \end{pmatrix} = Var\begin{pmatrix} W_{\alpha jt} \\ W_{\alpha^* jt} \end{pmatrix} = I$$

Where I the 2x2 identity matrix.

Therefore, the phase θ_j is indistinguishable.

Appendix B: Futures Contract Valuation

Let $Z_t = \begin{pmatrix} \xi_t & \chi_{1t} & \cdots & \chi_{n-1t} & \alpha_{1t} & \alpha_{1t}^{*} & \cdots & \alpha_{mt} & \alpha_{mt}^{*} \end{pmatrix}'$ be the vector of all factors. The "risk-neutral" SDE of Z_t can be expressed as:

$$dZ_t = \left(b^{\Diamond} + AZ_t\right)dt + \Omega dW_{Zt}^{\Diamond}$$

where $dW_{Zt}^{\Diamond}$ is a vector of independent Brownian motions, and therefore Var(dZ_t) = $R = \Omega\Omega^{T}$ (Ω^{T} is the transpose matrix of Ω) with the restriction explained above, $b^{\Diamond} = \begin{pmatrix} \mu_{\xi} - \lambda_{\xi} & -\lambda_{\chi 1} & \cdots & -\lambda_{\chi n-1} & -\lambda_{\alpha 1} & -\lambda_{\alpha^* 1} & \cdots & -\lambda_{\alpha m} & -\lambda_{\alpha^* m} \end{pmatrix}$ and:

$$A = \begin{pmatrix}
0 & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 & 0 \\
0 & -k_1 & \cdots & 0 & 0 & 0 & \cdots & 0 & 0 \\
\cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\
0 & 0 & \cdots & -k_{n-1} & 0 & 0 & \cdots & 0 & 0 \\
0 & 0 & \cdots & 0 & 0 & \varphi_1 & \cdots & 0 & 0 \\
0 & 0 & \cdots & 0 & -\varphi_1 & 0 & \cdots & 0 & 0 \\
\cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\
0 & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 & \varphi_m \\
0 & 0 & \cdots & 0 & 0 & 0 & \cdots & -\varphi_m & 0
\end{pmatrix}$$

Under this notation $X_t = cZ_t$, where $c = \begin{pmatrix} 1 & \overbrace{1 \quad \cdots \quad 1}^{n-1} & \overbrace{1 \quad 0 \quad \cdots \quad 1 \quad 0}^{m} \end{pmatrix}$.

It is easy to prove that the (unique) solution of that problem is (Oksendal, 1992) [13]:

$$Z_t = e^{At}\left[Z_0 + \int_0^t e^{-As} b^{\Diamond} ds + \int_0^t e^{-As} \Omega dW_{Zs}^{\Diamond} \right] \tag{B1}$$

It is clear that, under the risk-neutral measure, given Z_0, Z_t is Gaussian, with mean and variance[14 15]

$$E^*[Z_t] = e^{At}\left[Z_0 + \int_0^t e^{-As}\, b^{\Diamond} ds \right] \tag{B2}$$

$$Var^*[Z_t] = e^{At}\left[\int_0^t e^{-As} R (e^{-As})^T ds \right] (e^{At})^T \tag{B3}$$

As $X_t = c\,Z_t = \xi_t + \sum_{i=1}^{n-1} \chi_{it} + \sum_{j=1}^{m} \alpha_{jt}$, then under the risk-neutral measure, X_t is also Gaussian with mean and variance:

$$E^*(X_t) = cE^*(Z_t)$$

$$Var^*(X_t) = cVar^*(Z_t)c^T$$

This provides a valuation scheme for all sorts of commodity contingent claims as financial derivatives on commodity prices, real options, investment decisions and other more.

In particular, the price of a futures contract traded at time "*t*" with maturity at time "*t+T*" is: $F_{t,T} = E^*\left[S_{t+T} \| I_t\right] = \exp\left\{ E^*\left[X_{t+T} \| I_t\right] + \frac{1}{2} Var^*\left[X_{t+T} \| I_t\right] \right\}$, where I_t is the information available at time "*t*". It can be expressed as:

$$F_{t,T} = \exp\left[ce^{AT} Z_t + g(T) \right] \tag{B4}$$

[13] This methodology is general and can be used in all kind of problems. It does not matter which is b, A and R. Even in the case that b, A and R were function of t, if At and commute, the solution of that problem is (B1).

[14] E*[] and Var*[] are the mean and variance under the risk neutral measure.

[15] Superscript T indicate transpose matrix.

where $g(T) = ce^{AT}\int_t^{t+T} e^{-As}\, b^{\Diamond} ds + \frac{1}{2} ce^{AT}\left[\int_t^{t+T} e^{-As} R(e^{-AT})^T\, ds\right](e^{AT})^T c^T$, which is a deterministic function.

The squared volatility of a futures contract traded at time "*t*" with maturity at time "*t+T*" is defined as[16] $\lim_{h\to 0}\frac{Var\left[\log F_{t+h,T} - \log F_{t,T}\right]}{h}$. It is easy to prove that it is the expected value of the square of the coefficient of the Brownian motion (σ_t) in the expansion: $d\log(F_{t,T}) = \mu_t ds + \sigma_t dW_t^F$, where W_t^F is a scalar canonical Brownian motion. Therefore, taking logarithms and differentials on both sides of Equation (B4), it follows that:

$$d\left(\log F_{t,T}\right) = ce^{AT} dZ_t = ce^{AT}\left[b + AZ_t\right]dt + ce^{AT}\Omega dW_t$$

So, the futures squared volatility is:

$$ce^{AT} R(e^{AT})^T c^T \qquad \text{(B5)}$$

Appendix C: The Five-Factor Model

C.1 Five-factor model reformulation

In the five factors model the log-spot price (X_t) is given by:

$$X_t = \xi_t + \chi_{1t} + \chi_{2t} + \alpha_t$$

The SDEs of the factors can be expressed as:

[16] The same results are going to be obtained if the volatility is defined as: $\lim_{h\to 0}\frac{Var\left[\log F_{t+h,T-h} - \log F_{t,T}\right]}{h}$.

$$\begin{pmatrix} d\xi_t \\ d\chi_{1t} \\ d\chi_{2t} \\ d\alpha_t \\ d\alpha_t^* \end{pmatrix} = \begin{pmatrix} \mu_\xi \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix} dt + \begin{pmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & -k_1 & 0 & 0 & 0 \\ 0 & 0 & -k_2 & 0 & 0 \\ 0 & 0 & 0 & 0 & \varphi \\ 0 & 0 & 0 & -\varphi & 0 \end{pmatrix} \begin{pmatrix} \xi_t \\ \chi_{1t} \\ \chi_{2t} \\ \alpha_t \\ \alpha_t^* \end{pmatrix} dt + \begin{pmatrix} \sigma_\xi dW_{\xi t} \\ \sigma_{\chi 1} dW_{\chi 1t} \\ \sigma_{\chi 2} dW_{\chi 2t} \\ \sigma_\alpha dW_{\alpha t} \\ \sigma_\alpha dW_{\alpha^* t} \end{pmatrix}$$

Defining $y_t = \kappa_1 \chi_{1t}$ and $v_t^{\otimes} = -k_2 \chi_{2t}$ it is clear that:

$$dy_t = \kappa_1 d\chi_{1t} = \kappa_1 \left(-\kappa_1 \chi_{1t} dt\right) + \kappa_1 \sigma_{\chi 1} dW_{\chi 1t} = -\kappa_1^2 \left(\frac{y_t}{\kappa_1}\right) + \kappa_1 \sigma_{\chi 1} dW_{\chi 1t} =$$
$$= -\kappa_1 y_t dt + \kappa_1 \sigma_{\chi 1} dW_{\chi 1t}$$

$$dv_t^{\otimes} = -k_2 d\chi_{2t} = -k_2(-k_2 \chi_{2t} dt) - k_2 \sigma_{\chi 2} dW_{\chi 2t} = -k_2 v_t^{\otimes} d - k_2 \sigma_{\chi 2} dW_{\chi 2t}$$

Let $v_t = v_t^{\otimes} + \mu_\xi$ and $\xi_t^{\otimes} = \xi_t - \frac{\mu_\xi}{k_2}$, in this case:

$$dv_t = dv_t^{\otimes} = -k_2 v_t^{\otimes} dt - k_2 \sigma_{\chi 2} dW_{\chi 2t} = k_2(\mu_\xi - v_t) dt - k_2 \sigma_{\chi 2} dW_{\chi 2t}$$

$$d\xi_t^{\otimes} = d\xi_t = \mu_\xi dt + \sigma_\xi dW_{\xi t}$$

It is easy to see that:

$$X_t = \xi_t + \chi_{1t} + \chi_{2t} + \alpha_t = \xi_t^{\otimes} + \frac{y_t}{k_1} + \frac{v_t}{k_2} + \alpha_t$$

Therefore:

$$dX_t = d\xi_t + d\chi_{1t} + d\chi_{2t} + d\alpha_t = d\xi_t^{\otimes} + \frac{dy_t}{k_1} - \frac{dv_t}{k_2} + d\alpha_t = \mu_\xi dt + \sigma_\xi dW_{\xi t} - y_t dt + \sigma_{\chi 1} dW_{\chi 1t} +$$
$$-(\mu_\xi - v_t) dt + \sigma_{\chi 2} dW_{\chi 2t} + \varphi \alpha_t^* dt + \sigma_\alpha dW_{\alpha^* t} = \left(v_t - y_t + \varphi \alpha_t^*\right) dt + \sigma_X dW_{Xt}$$

where $\sigma_X dW_{Xt} = \sigma_\xi dW_{\xi t} + \sigma_{\chi 1} dW_{\chi 1t} + \sigma_{\chi 2} dW_{\chi 2t} + \sigma_\alpha dW_{\alpha t}$.

Thus, the model can be expressed in terms of the new variables. The new factors are: X_t, y_t, v_t, α_t and α_t^*. The SDE of these new factors are:

$$dX_t = \left(v_t - y_t + \varphi\alpha_t^*\right)dt + \sigma_X dW_{Xt}$$

$$dy_t = -k_1 y_t dt + k_1 \sigma_{\chi 1} dW_{\chi 1t}$$

$$dv_t = k_2(\mu_\xi - v_t)dt - k_2 \sigma_{\chi 2} dW_{\chi 2t}$$

$$d\alpha_t = \varphi\alpha_t^* dt + \sigma_\alpha dW_{\alpha t}$$

$$d\alpha_t^* = -\varphi\alpha_t dt + \sigma_\alpha dW_{\alpha^* t}$$

Therefore, it is obvious that this model is a generalization of the three-factor model developed in Cortazar and Schwartz (2003).

C.2 The log-price of futures contracts and the squared volatility of futures returns

In the context of the five-factor model, the log-price of a futures contract with maturity at time "*T*+*t*" traded at time *t* can be calculated applying the result in equation (B4):

$$\ln\left[F(X_t, T+t)\right] = \xi_0 + e^{-k_1 T}\chi_{10} + e^{-k_2 T}\chi_{20} + \cos(2\pi\varphi T)\alpha_0 + sen(2\pi\varphi T)\alpha_0^* + A_5(T)$$

where:

$$A_5(T) = (\mu_\xi' + 0.5\sigma_\xi^2 + 0.5\sigma_\alpha^2)T - (1-e^{-k_1 T})\lambda_{\chi 1}/k_1 - (1-e^{-k_2 T})\lambda_{\chi 2}/k_2 - (\lambda_{\alpha^*} + \lambda_\alpha sen(2\pi\varphi T) -$$

$$- \lambda_{\alpha^*}\cos(2\pi\varphi T))/(2\pi\varphi) + \sigma_\xi \sigma_{\chi 1}\rho_{\xi\chi 1}(1-e^{-k_1 T})/k_1 + \sigma_\xi \sigma_{\chi 2}\rho_{\xi\chi 2}(1-e^{-k_2 T})/k_2 +$$

$$+ \frac{\sigma_\xi \sigma_\alpha \rho_{\xi\alpha}\{1-\cos(2\pi\varphi T)\} + \sigma_\xi \sigma_{\alpha^*}\rho_{\xi\alpha^*} sen(2\pi\varphi T)}{(2\pi\varphi)} + 0.25\sigma_{\chi 1}^2(1-e^{-2k_1 T})/k_1 +$$

$$+ \sigma_{\chi 1}\sigma_{\chi 2}\rho_{\chi 1\chi 2}(1-e^{-(k_1+k_2)T})/(k_1+k_2) + 0.25\sigma_{\chi 2}^2(1-e^{-2k_2 T})/k_2 +$$

$$+ \frac{\sigma_{\chi 1}\sigma_\alpha \rho_{\chi 1\alpha}\{k_1 - e^{-k_1 T}(k_1\cos(2\pi\varphi T) + \varphi sen(2\pi\varphi T))\}}{k_1^2 + (2\pi\varphi)^2} +$$

$$\frac{\sigma_{\chi 1}\sigma_{\alpha^*}\rho_{\chi 1\alpha^*}\{2\pi\varphi + e^{-k_1 T}(k_1 sen(2\pi\varphi T) - 2\pi\varphi\cos(2\pi\varphi T))\}}{k_1^2 + (2\pi\varphi)^2}$$

$$+ \frac{\sigma_{\chi 2}\sigma_\alpha \rho_{\chi 2\alpha}\{k_2 - e^{-k_2 T}(k_2\cos(2\pi\varphi T) + \varphi sen(2\pi\varphi T))\}}{k_2^2 + (2\pi\varphi)^2} +$$

$$+ \frac{\sigma_{\chi 2}\sigma_{\alpha^*}\rho_{\chi 2\alpha^*}\{2\pi\varphi + e^{-k_2 T}(k_2 sen(2\pi\varphi T) - \varphi\cos(2\pi\varphi T))\}}{k_2^2 + (2\pi\varphi)^2}$$

And, from equation (B5), the squared volatility of futures returns is:

$$\sigma_{F5}^2(T) = \sigma_\xi^2 + \sigma_{\chi 1}^2 e^{-2k_1 T} + \sigma_{\chi 2}^2 e^{-2k_2 T} + \sigma_\alpha^2 + 2e^{-k_1 T}\sigma_\xi \sigma_{\chi 1}\rho_{\xi\chi 1} + 2e^{-k_2 T}\sigma_\xi \sigma_{\chi 2}\rho_{\xi\chi 2}$$

$$+ 2\sigma_\xi \sigma_\alpha \rho_{\xi\alpha}\cos(2\pi\varphi T) + 2\sigma_\xi \sigma_\alpha \rho_{\xi\alpha^*} sen(2\pi\varphi T) + 2e^{-(k_1+k_2)T}\sigma_{\chi 1}\sigma_{\chi 2}\rho_{\chi 1\chi 2} +$$

$$+ 2e^{-k_1 T}\sigma_{\chi 1}\sigma_\alpha \rho_{\chi 1\alpha}\cos(2\pi\varphi T) + 2e^{-k_1 T}\sigma_{\chi 1}\sigma_\alpha \rho_{\chi 1\alpha^*} sen(2\pi\varphi T) +$$

$$+ 2e^{-k_2 T}\sigma_{\chi 2}\sigma_\alpha \rho_{\chi 2\alpha}\cos(2\pi\varphi T) + 2e^{-k_2 T}\sigma_{\chi 2}\sigma_\alpha \rho_{\chi 2\alpha^*} sen(2\pi\varphi T)$$

Appendix D: Transition and measurement equations

The Three-Factor Model

Transition equation:

$$Z_t = c_t + T_t Z_{t-1} + \psi_t \qquad \mathbf{t} = 1, \ldots, N_t$$

where :

$$Z_t = \begin{pmatrix} \xi_t \\ \chi_t \\ \alpha_t \\ \alpha_t^* \end{pmatrix}, \quad c_t = \begin{pmatrix} \mu_\xi \Delta t \\ 0 \\ 0 \\ 0 \end{pmatrix}, \quad T_t = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & e^{-k\Delta t} & 0 & 0 \\ 0 & 0 & \cos(\varphi\Delta t) & \sin(\varphi\Delta t) \\ 0 & 0 & -\sin(\varphi\Delta t) & \cos(\varphi\Delta t) \end{pmatrix}$$

and

$$Var(\psi_t) = \begin{pmatrix} 0 & - & - & - \\ 0 & \sigma_\chi^2 (1 - e^{-2k\Delta t})/(2k) & - & - \\ 0 & -\sigma_\chi \sigma_\alpha \rho_{\chi\alpha} \dfrac{k\{-1 + e^{-k\Delta t}(\cos(\varphi\Delta t) + sen(\varphi\Delta t))\} + \varphi\{-1 + e^{-k\Delta t}(\cos(\varphi\Delta t) - sen(\varphi\Delta t))\}}{k^2 + \varphi^2} & \sigma_\alpha^2 \Delta t & - \\ 0 & \sigma_\chi \sigma_\alpha \rho_{\chi\alpha} \dfrac{k\{1 + e^{-k\Delta t}(-\cos(\varphi\Delta t) + sen(\varphi\Delta t))\} + \varphi\{-1 + e^{-k\Delta t}(\cos(\varphi\Delta t) + sen(\varphi\Delta t))\}}{k^2 + \varphi^2} & 0 & \sigma_\alpha^2 \Delta t \end{pmatrix}$$

Measurement equation:

$$Y_t = d_t + M_t Z_t + \eta_t \qquad t = 1, \ldots, N_t$$

where:

$$Y_t = \begin{pmatrix} \ln F_{T1} \\ \vdots \\ \ln F_{Tn} \end{pmatrix}, \; d_t = \begin{pmatrix} A_3(T_1) \\ \vdots \\ A_3(T_n) \end{pmatrix}, \; M_t = \begin{pmatrix} 1 & e^{-kT_1} & \cos(\varphi T_1) & sen(\varphi T_1) \\ \vdots & \vdots & \vdots & \vdots \\ 1 & e^{-kT_n} & \cos(\varphi T_n) & sen(\varphi T_n) \end{pmatrix}$$

The Four-Factor Model

Transition equation:

$$Z_t = c_t + T_t Z_{t-1} + \psi_t \qquad t = 1, \ldots, N_t$$

where:

$$Z_t = \begin{pmatrix} \xi_t \\ \chi_t \\ \alpha_t \\ \alpha_t^* \end{pmatrix}, \; c_t = \begin{pmatrix} \mu_\xi \Delta t \\ 0 \\ 0 \\ 0 \end{pmatrix}, \; T_t = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & e^{-k\Delta t} & 0 & 0 \\ 0 & 0 & \cos(\varphi\Delta t) & \sin(\varphi\Delta t) \\ 0 & 0 & -\sin(\varphi\Delta t) & \cos(\varphi\Delta t) \end{pmatrix}$$

and

$$Var(\psi_t)=\begin{pmatrix} \sigma_\xi^2\Delta t & - & - & - \\ \sigma_\xi\sigma_\chi\rho_{\chi\xi}(1-e^{-k\Delta t})/k & \sigma_\chi^2(1-e^{-2k\Delta t})/(2k) & - & - \\ \sigma_\xi\sigma_\alpha\rho_{\xi\alpha}(1+sen(\varphi\Delta t)-\cos(\varphi\Delta t))/\varphi & -\sigma_\chi\sigma_\alpha\rho_{\chi\alpha}\dfrac{k\left\{-1+e^{-k\Delta t}\left(\cos(\varphi\Delta t)+sen(\varphi\Delta t)\right)\right\}+\varphi\left\{-1+e^{-k\Delta t}\left(\cos(\varphi\Delta t)-sen(\varphi\Delta t)\right)\right\}}{k^2+\varphi^2} & \sigma_\alpha^2\Delta t & - \\ \sigma_\xi\sigma_\alpha\rho_{\xi\alpha}(-1+sen(\varphi\Delta t)+\cos(\varphi\Delta t))/\varphi & \sigma_\chi\sigma_\alpha\rho_{\chi\alpha}\dfrac{k\left\{1+e^{-k\Delta t}\left(-\cos(\varphi\Delta t)+sen(\varphi\Delta t)\right)\right\}+\varphi\left\{-1+e^{-k\Delta t}\left(\cos(\varphi\Delta t)+sen(\varphi\Delta t)\right)\right\}}{k^2+\varphi^2} & 0 & \sigma_\alpha^2\Delta t \end{pmatrix}$$

Measurement equation:

$$Y_t = d_t + M_t Z_t + \eta_t \qquad t = 1, \ldots, N_t$$

where:

$$Y_t = \begin{pmatrix} \ln F_{T1} \\ \vdots \\ \ln F_{Tn} \end{pmatrix},\ d_t = \begin{pmatrix} A_4(T_1) \\ \vdots \\ A_4(T_n) \end{pmatrix},\ M_t = \begin{pmatrix} 1 & e^{-kT_1} & \cos(\varphi T_1) & sen(\varphi T_1) \\ \vdots & \vdots & \vdots & \vdots \\ 1 & e^{-kT_n} & \cos(\varphi T_n) & sen(\varphi T_n) \end{pmatrix}$$

The Five-Factor Model

Transition equation:

$$Z_t = c_t + T_t Z_{t-1} + \psi_t \qquad t = 1, \ldots, N_t$$

where:

$$Z_t = \begin{pmatrix} \xi_t \\ \chi_{1t} \\ \chi_{2t} \\ \alpha_t \\ \alpha_t^* \end{pmatrix},\ c_t = \begin{pmatrix} \mu_\xi \Delta t \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix},\ T_t = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & e^{-k_1\Delta t} & 0 & 0 & 0 \\ 0 & 0 & e^{-k_2\Delta t} & 0 & 0 \\ 0 & 0 & 0 & \cos(\varphi\Delta t) & \sin(\varphi\Delta t) \\ 0 & 0 & 0 & -\sin(\varphi\Delta t) & \cos(\varphi\Delta t) \end{pmatrix}$$

and $Var(\psi_t)$ is the same as in the other models but with a new column:

$$Var(\psi_t)=\begin{pmatrix} \vdots & \vdots & \sigma_{\xi}\sigma_{\chi 2}\rho_{\chi 2\xi}(1-e^{-k_2\Delta t})/k_2 & \vdots & \vdots \\ \vdots & \vdots & \sigma_{\chi 1}\sigma_{\chi 2}\rho_{\chi 1\chi 2}(1-e^{-(k_1+k_2)\Delta t})/(k_1+k_2) & \vdots & \vdots \\ \vdots & \vdots & \sigma_{\chi 2}^2(1-e^{-2k_2\Delta t})/(2k_2) & \vdots & \vdots \\ \vdots & \vdots & -\sigma_{\chi 2}\sigma_{\alpha}\rho_{\chi 2\alpha}\dfrac{k_2\left\{-1+e^{-k_2\Delta t}\left(\cos(\varphi\Delta t)+sen(\varphi\Delta t)\right)\right\}+\varphi\left\{-1+e^{-k_2\Delta t}\left(\cos(\varphi\Delta t)-sen(\varphi\Delta t)\right)\right\}}{k_2^2+\varphi^2} & \vdots & \vdots \\ \vdots & \vdots & \sigma_{\chi 2}\sigma_{\alpha}\rho_{\chi 2\alpha}\dfrac{k_2\left\{1+e^{-k_2\Delta t}\left(-\cos(\varphi\Delta t)+sen(\varphi\Delta t)\right)\right\}+\varphi\left\{-1+e^{-k_2\Delta t}\left(\cos(\varphi\Delta t)+sen(\varphi\Delta t)\right)\right\}}{k_2^2+\varphi^2} & \vdots & \vdots \end{pmatrix}$$

Measurement equation:

$$Y_t = d_t + M_t Z_t + \eta_t \qquad\qquad t = 1, \ldots, N_t$$

where :

$$Y_t = \begin{pmatrix} \ln F_{T1} \\ \vdots \\ \ln F_{Tn} \end{pmatrix},\ d_t = \begin{pmatrix} A_5(T_1) \\ \vdots \\ A_5(T_n) \end{pmatrix},\ M_t = \begin{pmatrix} 1 & e^{-k_1T_1} & e^{-k_2T_1} & \cos(\varphi T_1) & sen(\varphi T_1) \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & e^{-k_1T_n} & e^{-k_2T_n} & \cos(\varphi T_n) & sen(\varphi T_n) \end{pmatrix}$$

REFERENCES

- Borovkova, S. and Geman, H. (2006). Seasonal and stochastic effects in commodity forward curves. Review of Derivatives Research, 9, 167-186.

- Brennan, M.J. and Schwartz, E.S. (1985). Evaluating natural resource investments. Journal of Business, 58, 133-155.

- Clewlow L. and Strickland, C. (2000). Energy Derivatives, Pricing and Risk Management. London (U.K.): Lamina Publications.

- Cortazar, G. and Schwartz, E.S. (2003). Implementing a stochastic model for oil futures prices. Energy Economics, 25, 215-218.

- Cortazar G. and Naranjo L. (2006). An N-Factor gaussian model of oil futures prices. The Journal of Futures Markets, 26, 209-313.
- Harvey, A.C. (1989). Forecasting Structural Time Series Models and the Kalman Filter. Cambridge (U.K.): Cambridge University Press.
- Laugthton, D.G. and Jacoby, H.D. (1993). Reversion, timing options, and long-term decision making. Financial Management, 33, 225-240.
- Laugthton, D.G. and Jacoby, H.D. (1995). The effects of reversion on commodity projects of different length (pp. 185-205). In L. Trigeorgis, (Ed.), Real Options in Capital Investment: Models, Strategies and Applications. Praeger, Westport, CT.
- Lucia, J. and Schwartz, E.S. (2002). Electricity Prices and Power derivatives: Evidence from the Nordic Power Exchange. Review of derivative Research, 5, 5-50.
- Mu, X. (2007). Weather, storage, and natural gas price dynamics: Fundamentals and volatility. Energy Economics, 29, 46-63.
- Oksendal, B. (1992). Stochastic Differential Equations. An Introduction with Applications, (3rd ed). Springer-Verlag, Berlin Heidelberg.
- Ross, S.A. (1997) Hedging long run commitments: Exercises in incomplete market pricing. Banca Monte Econom. Notes, 26, 99-132.
- Schwartz, E.S. (1997). The stochastic behavior of commodity prices: Implication for valuation and hedging. The Journal of Finance, 52, 923-973.
- Schwartz, E.S. and Smith, J.E. (2000). Short-term variations and long-term dynamics in commodity prices, Management Science, 46, 893-911.
- Sorensen, C. (2002). Modeling seasonality in agricultural commodity futures. The Journal of futures markets, 22, 393-426.

- Tolmasky, C. and Hindanov, D. (2002). Principal Components Analysis for Correlated Curves and Seasonal Commodities: The Case of the Petroleum Market. The Journal of Futures Markets, 22, 1019-1035.

- Wei, S.Z.C. and Zhu, Z. (2006). Commodity convenience yield and risk premium determination: The case of US natural gas market. Energy Economics, 28, 523-534.

- Wei, W.S. (2005). The time series analysis. Univariate and multivariate methods (2th ed.). Addison Wesley.

FIGURE 1

HENRY HUB NATURAL GAS PRICES. WEEKLY OBSERVATIONS FROM NYMEX

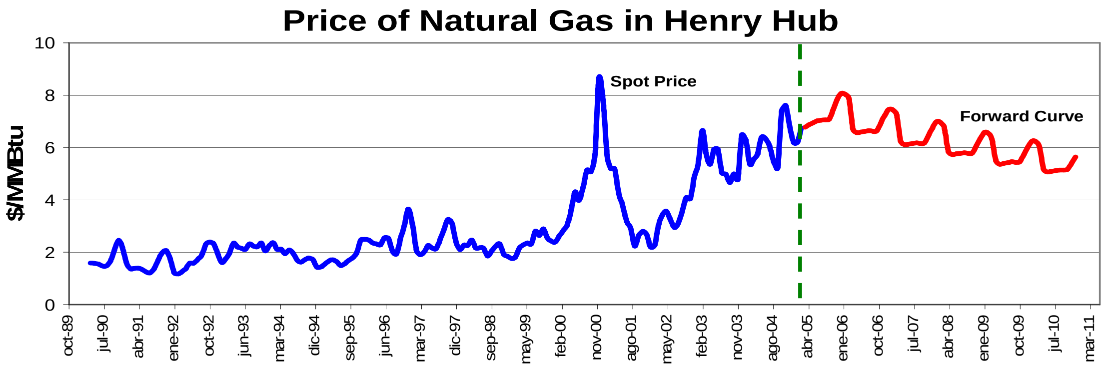
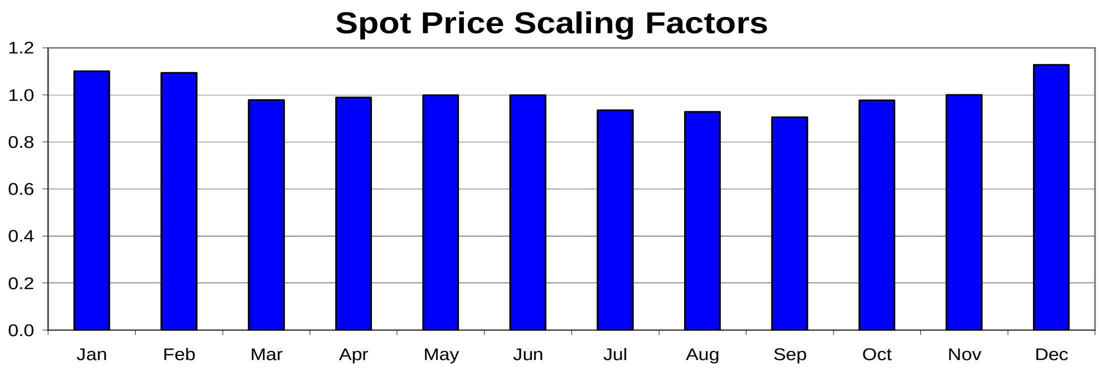


FIGURE 2

SCALING FACTORS (WEEKLY OBSERVATIONS FROM NYMEX)



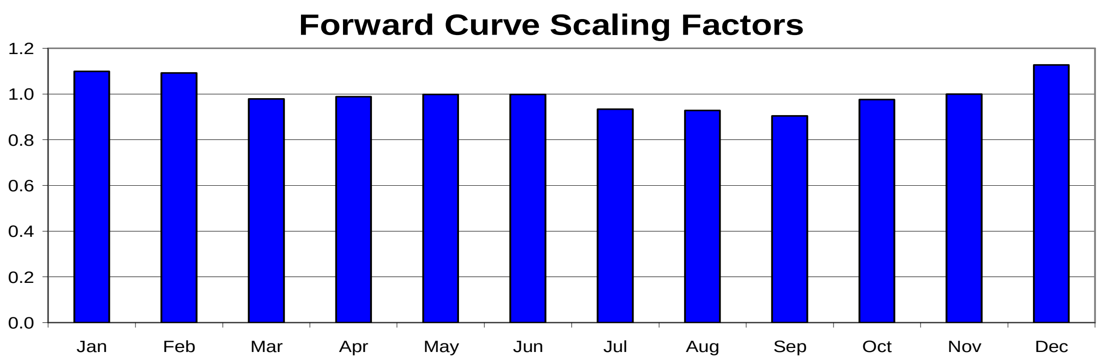


FIGURE 3: SPECTRUM

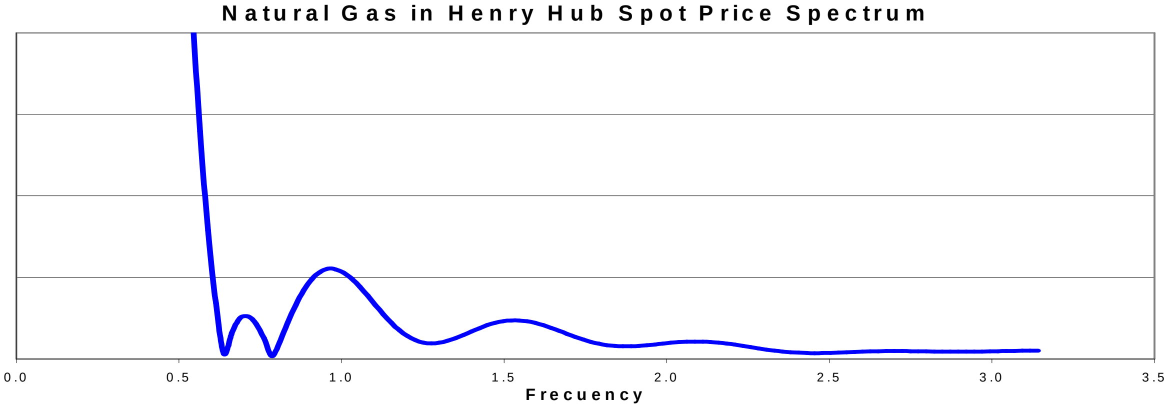
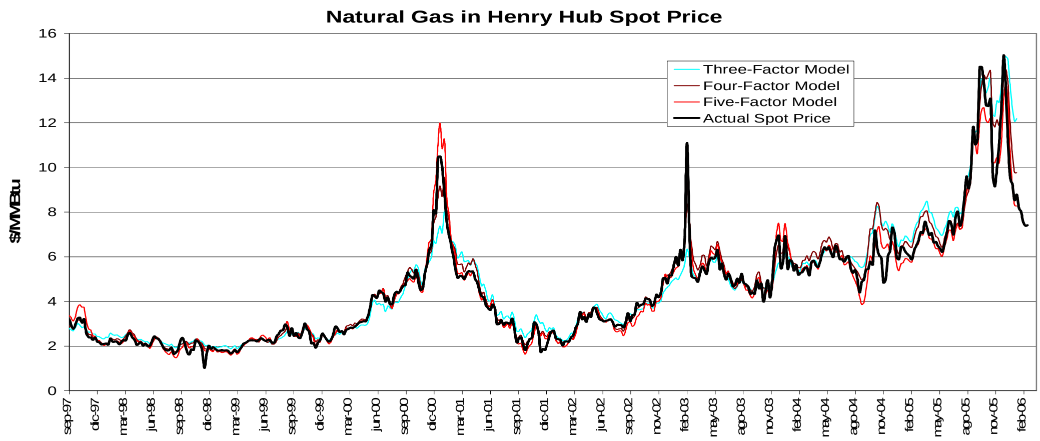


FIGURE 4: SPOT PRICE ESTIMATION



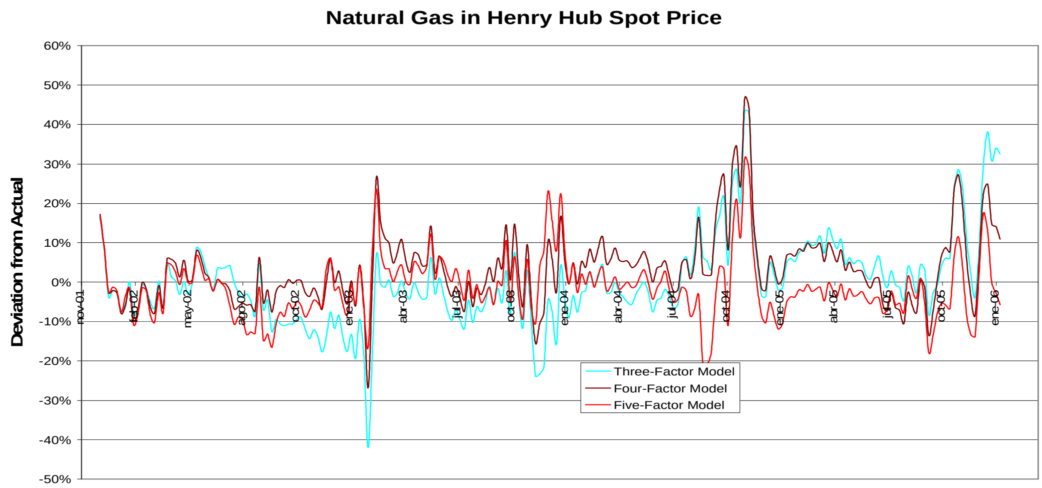


FIGURE 5

FORWARD CURVE 05/27/2002

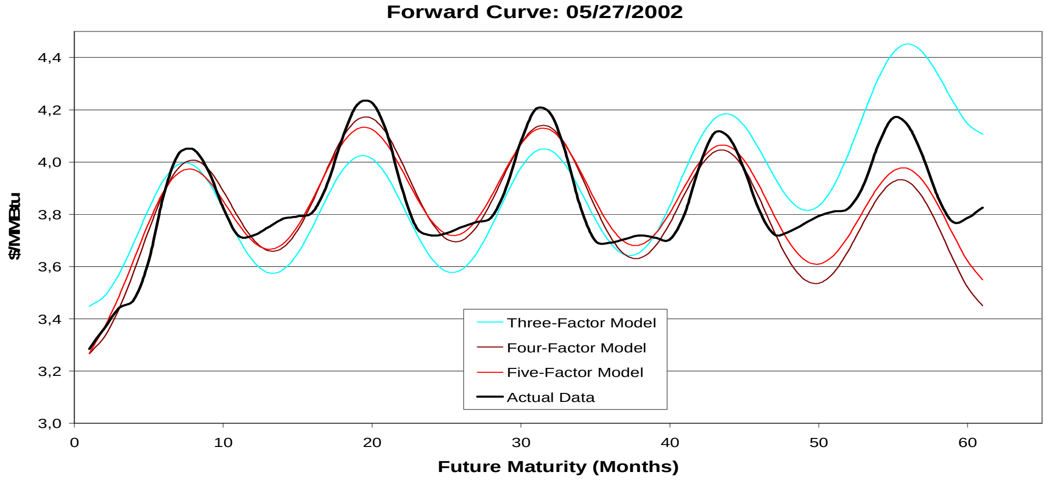


FIGURE 6

VOLATILITY OF FUTURES RETURNS

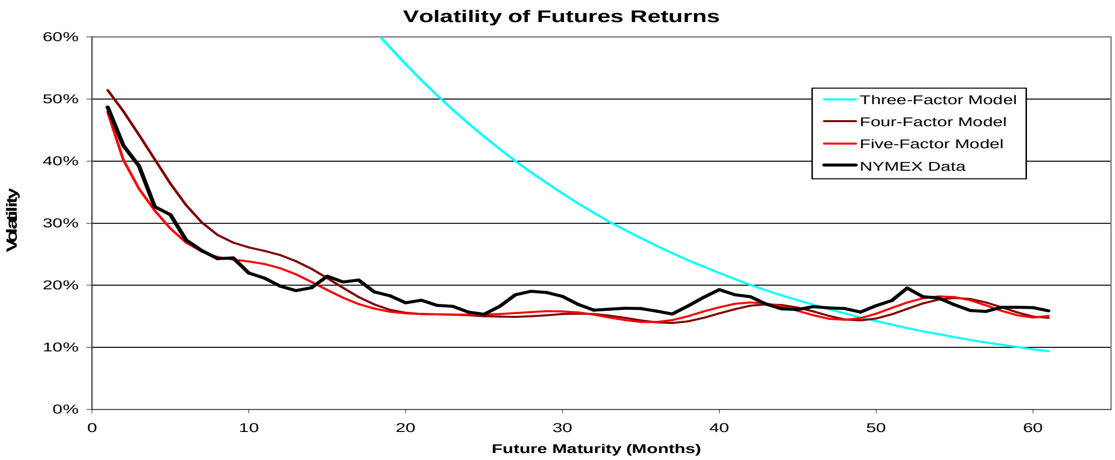


TABLE I

HENRY HUB NATURAL GAS FUTURES PRICES. DESCRIPTIVE STATS ($/MMBtu)

The table shows the main descriptive stats of Henry Hub natural gas futures prices traded at NYMEX for four different data sets.

	Data-Set 1		Data-Set 2		Data-Set 3		Data-Set 4		
Num. Obs.	438		222		216		216		
Contract	Mean	Stand. Dev.	Mean	Stand. Dev.	Mean	Stand. Dev.	Contract	Mean	Stand. Dev.
F1	4.57	2.53	3.19	1.51	5.99	2.57	F1	5.99	2.57
F5	4.61	2.38	3.16	1.10	6.10	2.57	F8	5.96	2.07
F9	4.50	2.15	3.10	1.05	5.94	2.57	F15	5.81	2.19
F14	4.42	2.20	3.03	0.89	5.86	2.56	F22	5.51	1.71
F18	4.28	1.96	2.97	0.82	5.63	2.56	F29	5.34	1.70
F22	4.20	1.85	2.93	0.78	5.51	2.55	F36	5.22	1.55
F27	4.17	1.89	2.92	0.72	5.45	2.54	F43	5.07	1.35
F31	4.06	1.70	2.88	0.70	5.27	2.54	F50	5.08	1.43
F35	4.03	1.65	2.88	0.68	5.21	2.53	F57	4.92	1.09

TABLE II

HENRY HUB NATURAL GAS (NYMEX). THREE-FACTOR MODEL

The table presents the results for the three-factor model applied to the four data sets described in the chapter. Standard errors in parentheses. The estimated values are reported with * denoting significance at 10%, ** denoting significance at 5%, and *** denoting significance at 1%.

	Data-Set 1	Data-Set 2	Data-Set 3	Data-Set 4
Number obs.	438	222	216	216
μ_ξ	0.1429^{***} (0.0004)	0.1446^{***} (0.0009)	0.1655^{***} (0.0025)	0.0155^{**} (0.0062)
k	0.4025^{***} (0.0046)	0.4723^{***} (0.0062)	0.5714^{***} (0.0143)	0.1886^{***} (0.0058)
Φ	0.9941^{***} (0.0023)	0.9885^{***} (0.0030)	1.010^{***} (0.0034)	1.0025^{***} (0.0027)
σ_χ	1.0056^{***} (0.0196)	1.0047^{***} (0.0245)	1.4358^{***} (0.0313)	0.3014^{***} (0.0174)
σ_α	0.0846^{***} (0.0041)	0.0824^{***} (0.0054)	0.0510^{***} (0.0083)	0.0670^{***} (0.0061)
$\rho_{\chi\alpha}$	0.2010^{***} (0.0310)	0.1216^{***} (0.0359)	0.6616^{***} (0.0390)	0.1549^{***} (0.0316)
$\rho_{\chi\alpha^*}$	0.5762^{***} (0.0275)	0.6709^{***} (0.0312)	0.3140^{***} (0.0406)	0.0960^{***} (0.0250)
λ_χ	0.6596^{**} (0.2904)	0.4807 (0.4299)	0.6497 (0.5719)	0.4470^{***} (0.1027)
λ_α	0.0025 (0.0112)	-0.0135 (0.0151)	-0.0351^{***} (0.0124)	0.0008 (0.0144)
λ_{α^*}	-0.0084 (0.0137)	-0.0042 (0.0206)	-0.0121 (0.0114)	-0.0317^{**} (0.0137)
σ_η	0.0544^{***} (0.0005)	0.0497^{***} (0.0006)	0.0592^{***} (0.0007)	0.0680^{***} (0.0008)
Log-likelihood	17849	9387.03	8471.95	8279.9
AIC	17827	9365.03	8449.95	8257.9
SIC	17782.1	9327.6	84128.2	82207.7

TABLE III

HENRY HUB NATURAL GAS (NYMEX). FOUR-FACTOR MODEL

The table presents the results for the four-factor model applied to the four data sets described in the chapter. Standard errors in parentheses. The estimated values are reported with * denoting significance at 10%, ** denoting significance at 5%, and *** denoting significance at 1%.

	Data-Set 1	Data-Set 2	Data-Set 3	Data-Set 4
Number obs.	438	222	216	216
μ_{ξ}	0.1791^{***} (0.0436)	0.1182 (0.1220)	0.2476^{***} (0.0674)	0.2292 (0.1433)
k	0.9135^{***} (0.0038)	1.0968^{***} (0.0126)	0.7817^{***} (0.0037)	0.5912^{***} (0.0061)
Φ	0.9973^{***} (0.0015)	0.9937^{***} (0.0020)	0.9999^{***} (0.0028)	1.0008^{***} (0.0033)
σ_{ξ}	0.1704^{***} (0.0015)	0.1765^{***} (0.0038)	0.1611^{***} (0.0020)	0.1995^{***} (0.0020)
σ_{χ}	0.5592^{***} (0.0076)	0.5509^{***} (0.0149)	0.5922^{***} (0.0014)	0.7542^{***} (0.0025)
σ_{α}	0.0561^{***} (0.0012)	0.0515^{***} (0.0023)	0.0576^{***} (0.0004)	0.0529^{***} (0.0008)
$\rho_{\xi\chi}$	-0.3908^{***} (0.0039)	-0.2560^{***} (0.0063)	-0.5398^{***} (0.0047)	-0.7236^{***} (0.0022)
$\rho_{\xi\alpha}$	-0.2617^{***} (0.0037)	-0.2193^{***} (0.0048)	-0.2946^{***} (0.0047)	-0.0729^{***} (0.0022)
$\rho_{\xi\alpha*}$	-0.1096^{***} (0.0041)	-0.0911^{***} (0.0059)	-0.1760^{***} (0.0053)	-0.1626^{***} (0.0023)
$\rho_{\chi\alpha}$	0.3942^{***} (0.0084)	0.3956^{***} (0.0393)	0.3735^{***} (0.0010)	0.0540^{***} (0.0051)
$\rho_{\chi\alpha*}$	0.2716^{***} (0.0067)	0.3006^{***} (0.0246)	0.3072^{***} (0.0071)	0.2618^{***} (0.0053)
μ_{ξ}’	-0.0479^{***} (0.0020)	-0.0312^{***} (0.0016)	-0.0661^{***} (0.0042)	-0.0264^{***} (0.0024)
λ_{χ}	0.0029 (0.0894)	-0.0726 (0.0476)	0.0791^{**} (0.0317)	0.1926^{*} (0.1014)
λ_{α}	0.0019 (0.0036)	0.0003 (0.0102)	0.0059 (0.0056)	0.0081 (0.0050)
$\lambda_{\alpha*}$	-0.0117^{**} (0.0057)	-0.0149 (0.0121)	-0.0075 (0.0072)	-0.0038 (0.0201)
σ_{η}	0.0324^{***} (0.0002)	0.0300^{***} (0.0004)	0.0324^{***} (0.0004)	0.0373^{***} (0.0004)
Log-likelihood	21555.9	11150.2	10661.5	10193.6
AIC	21523.8	11118.2	10629.5	10161.6
SIC	21458.5	11063.8	10575.5	10107.6

TABLE IV
HENRY HUB NATURAL GAS (NYMEX). FIVE-FACTOR MODEL

The table presents the results for the five-factor model applied to the four data sets described in the chapter. Standard errors in parentheses. The estimated values are reported with * denoting significance at 10%, ** denoting significance at 5%, and *** denoting significance at 1%.

	Data-Set 1	Data-Set 2	Data-Set 3	Data-Set 4
Number obs.	438	222	216	216
μ_ξ	0.1678***	0.1248	0.2340***	0.0930
	(0.0262)	(0.1270)	(0.0539)	(0.0816)
k_1	0.5660***	7.5352***	6.1064***	3.3543***
	(0.0009)	(0.1789)	(0.2765)	(0.0235)
k_2	5.3633***	0.6259***	0.7001***	0.4411***
	(0.0500)	(0.0032)	(0.0102)	(0.0012)
Φ	0.9916***	0.9923***	0.9897***	0.9948***
	(0.0019)	(0.0034)	(0.0019)	(0.0019)
σ_ξ	0.1995***	0.2020***	0.1613***	0.1933***
	(0.0044)	(0.0060)	(0.0013)	(0.0022)
$\sigma_{\chi 1}$	0.4550***	0.6390***	0.5568***	0.5195***
	(0.0058)	(0.0126)	(0.0133)	(0.0022)
$\sigma_{\chi 2}$	0.5441***	0.4070***	0.4808***	0.4754***
	(0.0222)	(0.0070)	(0.0345)	(0.0084)
σ_α	0.0600***	0.0525***	0.0674***	0.0459***
	(0.0099)	(0.0016)	(0.0021)	(0.0018)
$\rho_{\xi\chi 1}$	-0.6173***	0.1029***	0.2626***	0.3012***
	(0.0043)	(0.0022)	(0.0064)	(0.0027)
$\rho_{\xi\chi 2}$	0.1573***	-0.5064***	-0.5517***	-0.7455***
	(0.0043)	(0.0056)	(0.0049)	(0.0028)
$\rho_{\xi\alpha}$	-0.2705***	-0.1820***	-0.3167***	-0.1479***
	(0.0068)	(0.0059)	(0.0043)	(0.0028)
$\rho_{\xi\alpha*}$	0.0728***	0.0743***	0.0670***	0.1572***
	(0.0049)	(0.0057)	(0.0044)	(0.0028)
$\rho_{\chi 1\,\chi 2}$	-0.0985***	0.0612*	-0.2374***	-0.3005***
	(0.0057)	(0.0332)	(0.0688)	(0.0186)
$\rho_{\chi 1\alpha}$	0.3786***	-0.3751***	-0.3277***	-0.0534***
	(0.0437)	(0.0270)	(0.0567)	(0.0153)
$\rho_{\chi 1\alpha*}$	-0.0245***	0.0902***	0.0540***	-0.0208
	(0.0040)	(0.0329)	(0.0122)	(0.0206)
$\rho_{\chi 2\alpha}$	-0.3071***	0.3125***	0.4265***	0.1141***
	(0.1160)	(0.0085)	(0.0028)	(0.0082)
$\rho_{\chi 2\alpha*}$	0.0628	-0.0828***	0.0084	-0.0963***
	(0.0657)	(0.0091)	(0.0171)	(0.0100)
μ_ξ’	-0.0137***	-0.0033	-0.0420***	-0.0056**
	(0.0048)	(0.0039)	(0.0029)	(0.0023)
$\lambda_{\chi 1}$	0.1258	-0.6780***	-0.6910***	-0.6639***
	(0.0901)	(0.0936)	(0.1455)	(0.1110)
$\lambda_{\chi 2}$	-0.5237***	0.0165	0.1992***	0.6836***
	(0.0605)	(0.0487)	(0.0738)	(0.0602)
λ_α	-0.0051	-0.0046	-0.0004	-0.0085
	(0.0233)	(0.0070)	(0.0030)	(0.0077)
$\lambda_{\alpha*}$	0.0124	0.0070	0.0235***	0.0233*
	(0.0327)	(0.0051)	(0.0067)	(0.0139)
σ_η	0.0251***	0.0224***	0.0267***	0.0289***
	(0.0003)	(0.0004)	(0.0002)	(0.0002)
Log-likelihood	22858.2	11917.9	11113.5	10950.1
AIC	22812.2	11871.9	11167.5	10904.1
SIC	22718.3	11793.6	10989.8	10826.4

TABLE V

HENRY HUB NATURAL GAS (NYMEX). SCHWARTZ AND SMITH (2000) TWO-FACTOR MODEL WITH DETERMINISTIC SEASONALITY

The table presents the results for the Schwartz and Smith (2000) two-factor model with deterministic seasonality, applied to the four data sets described in the chapter. Standard errors in parentheses. The estimated values are reported with * denoting significance at 10%, ** denoting significance at 5%, and *** denoting significance at 1%.

	Data-Set 1	Data-Set 2	Data-Set 3	Data-Set 4
Number obs.	438	222	216	216
μ_ξ	0.1781*** (0.0326)	0.1160** (0.0513)	0.2512*** (0.0376)	0.2283*** (0.0436)
k	1.0292*** (0.0201)	1.2208*** (0.0273)	0.9726*** (0.0280)	0.6358*** (0.0148)
Φ	0.9843*** (0.0003)	0.9822*** (0.0009)	0.9929*** (0.0010)	0.987***0 (0.0007)
σ_ξ	0.1688*** (0.0051)	0.1769*** (0.0071)	0.1476*** (0.0070)	0.1864*** (0.0040)
σ_χ	0.5554*** (0.0175)	0.5505*** (0.0236)	0.5621*** (0.0265)	0.6966*** (0.0152)
$\rho_{\xi\chi}$	-0.3296*** (0.0409)	-0.2267*** (0.0570)	-0.3483*** (0.0626)	-0.6583*** (0.0125)
μ_ξ’	-0.0484*** (0.0017)	-0.0324*** (0.0021)	-0.0665*** (0.0024)	0.0425*** (0.0005)
λ_χ	-0.0059 (0.1210)	-0.0885 (0.1704)	0.0773 (0.1755)	-0.0306*** (0.0013)
λ_α	-0.0023 (0.0027)	-0.0080** (0.0035)	0.0470*** (0.0050)	0.1868 (0.1892)
$\lambda_{\alpha*}$	0.0260*** (0.0034)	0.0271*** (0.0044)	-0.0575*** (0.0048)	-0.0484*** (0.0007)
σ_η	0.0362*** (0.0003)	0.0336*** (0.0004)	0.0368*** (0.0005)	0.0526*** (0.0041)
Log-likelihood	20988,8	10861,5	10334.1	9831.87
AIC	20966.7	10839.6	10312	9809.9
SIC	20921.8	10802.1	10274.9	9727.7

TABLE VI

HENRY HUB NATURAL GAS (NYMEX).

IN-SAMPLE PREDICTIVE ABILITY (DATA SET 1)

The table presents several metrics trying to compare the in-sample predictive power ability of the Schwartz and Smith (2000) two-factor model with deterministic seasonality and the four-factor model with stochastic seasonality. The results are based on the first data set.

	Mean error (Real-Predicted)	Std. Dev. Error	Std. Dev. error (% mean)	Root Mean Squared Error	Root Mean Squared Error (% mean)
PANEL A: SCHWARTZ AND SMITH (2000) TWO-FACTOR MODEL WITH DETERMINISTIC SEASONALITY					
F1	0.011873	0.073741	5.317492	0.074691	0.05386
F5	-0.01563	0.058013	4.111995	0.06008	0.042585
F9	-0.00594	0.048836	3.494517	0.049196	0.035203
F14	0.001658	0.042283	3.05686	0.042315	0.030592
F18	0.002208	0.037473	2.755082	0.037538	0.027599
F22	0.007588	0.034945	2.593316	0.035759	0.026537
F27	0.002152	0.036535	2.727677	0.036599	0.027324
F31	-0.00175	0.038182	2.887859	0.038222	0.028909
F35	-0.00205	0.036748	2.7867	0.036805	0.02791
PANEL B: FOUR-FACTOR MODEL WITH STOCHASTIC SEASONALITY					
F1	0.007864	0.071977	5.190253	0.072405	0.052211
F5	-0.0092	0.052742	3.738416	0.053538	0.037948
F9	-0.0056	0.045167	3.231979	0.045513	0.032567
F14	-0.00239	0.037048	2.678386	0.037125	0.02684
F18	0.007655	0.035478	2.608369	0.036294	0.026684
F22	0.005313	0.034116	2.531833	0.034528	0.025623
F27	-0.00042	0.035255	2.63206	0.035257	0.026322
F31	0.003122	0.039431	2.982327	0.039554	0.029917
F35	-0.00615	0.036058	2.734364	0.036579	0.027739

TABLE VII

HENRY HUB NATURAL GAS (NYMEX)

OUT-OF-SAMPLE PREDICTIVE ABILITY

The table presents several metrics trying to compare the out-of-sample predictive power ability of the Schwartz and Smith (2000) two-factor model with deterministic seasonality and the four-factor model with stochastic seasonality. The results are obtained valuing the contracts in data set 3 with the parameters obtained estimating the models with data set 2.

	Mean error (Real-Predicted)	Std. Dev. Error	Std. Dev. error (% mean)	Root Mean Squared Error	Root Mean Squared Error (% mean)
PANEL A: SCHWARTZ AND SMITH (2000) TWO-FACTOR MODEL					
F1	0.02061	0.070913	6.58469	0.073847	0.068572
F5	-0.02638	0.062268	5.679274	0.067626	0.061679
F9	-0.02824	0.050649	4.685677	0.057991	0.053648
F14	-0.02244	0.047966	4.48428	0.052958	0.049509
F18	-0.01421	0.041866	3.97686	0.044213	0.041998
F22	-0.00423	0.037435	3.589902	0.037674	0.036128
F27	0.003337	0.043051	4.128993	0.04318	0.041414
F31	0.010411	0.045073	4.373193	0.04626	0.044883
F35	0.018478	0.04249	4.120222	0.046334	0.04493
PANEL B: FOUR-FACTOR MODEL WITH STOCHASTIC SEASONALITY					
F1	0.015791	0.070325	6.530147	0.072076	0.066927
F5	-0.01559	0.054871	5.004566	0.057042	0.052026
F9	-0.0296	0.044777	4.142435	0.053675	0.049656
F14	-0.02492	0.0342	3.197334	0.042315	0.03956
F18	-0.00598	0.034203	3.248993	0.034722	0.032983
F22	-0.01016	0.031116	2.983895	0.032732	0.031389
F27	0.003033	0.035621	3.416454	0.03575	0.034288
F31	0.016003	0.037406	3.629328	0.040686	0.039475
F35	0.010345	0.037622	3.648209	0.039018	0.037836

TABLE VIII

HEATING OIL (NYMEX). SCHWARTZ AND SMITH (2000) TWO-FACTOR MODEL WITH DETERMINISTIC SEASONALITY

The table presents the results for the Schwartz and Smith (2000) two-factor model with deterministic seasonality, applied to the three data sets described in the chapter. Standard errors in parentheses. The estimated values are reported with * denoting significance at 10%, ** denoting significance at 5%, and *** denoting significance at 1%.

	Data-Set 1	Data-Set 2	Data-Set 3
Number obs.	524	262	262
μ_ξ	0.1361*** (0.0309)	0.0576 (0.0373)	0.2097*** (0.0479)
k	1.5156*** (0.0185)	1.3826*** (0.0257)	1.4036*** (0.0263)
Φ	0.9914*** (0.0002)	0.9889*** (0.0007)	0.9832*** (0.0005)
σ_ξ	0.1506*** (0.0036)	0.1378*** (0.0054)	0.1583*** (0.0051)
σ_χ	0.2978*** (0.0073)	0.3223*** (0.0116)	0.2710*** (0.0091)
$\rho_{\xi\chi}$	0.0552 (0.0360)	-0.1707*** (0.0546)	0.2426*** (0.0446)
μ_ξ’	-0.0582*** (0.0016)	-0.0460*** (0.0029)	-0.0697*** (0.0018)
λ_χ	-0.0396 (0.0603)	-0.0114 (0.0909)	-0.0537 (0.0807)
λ_α	-0.0537*** (0.0016)	-0.0538*** (0.0026)	0.0007 (0.0015)
$\lambda_{\alpha*}$	-0.0091*** (0.0013)	-0.0025 (0.0020)	-0.0061*** (0.0013)
σ_η	0.0164*** (0.0001)	0.0194*** (0.0002)	0.0111*** (0.0001)
Log-likelihood	31741,6	15259,5	17407,2
AIC	31719.6	15237.5	17385.2
SIC	31672.7	15198.2	17345.9

TABLE IX

HEATING OIL (NYMEX). FOUR-FACTOR MODEL

The table presents the results for the four-factor model applied to the three data sets described in the chapter. Standard errors in parentheses. The estimated values are reported with * denoting significance at 10%, ** denoting significance at 5%, and *** denoting significance at 1%.

	Data-Set 1	Data-Set 2	Data-Set 3
Number obs.	524	262	262
μ_ξ	0.1435*** (0.0323)	0.0651 (0.0456)	0.2097*** (0.0480)
k	1.1372*** (0.0170)	1.0354*** (0.0250)	1.3996*** (0.0223)
Φ	1.0031*** (0.0021)	0.9949*** (0.0037)	1.0091*** (0.0020)
σ_ξ	0.1462*** (0.0036)	0.1254*** (0.0048)	0.1493*** (0.0049)
σ_χ	0.3078*** (0.0076)	0.3255*** (0.0120)	0.2701*** (0.0092)
σ_α	0.0271*** (0.0010)	0.0333*** (0.0018)	0.0218*** (0.0009)
$\rho_{\xi\chi}$	-0.0167*** (0.0056)	0.0437*** (0.0084)	0.0433*** (0.0071)
$\rho_{\xi\alpha}$	0.0454*** (0.0059)	-0.0859*** (0.0090)	-0.0056 (0.0084)
$\rho_{\xi\alpha*}$	0.0269*** (0.0064)	-0.0368*** (0.0112)	-0.0180** (0.0081)
$\rho_{\chi\alpha}$	-0.0614*** (0.0121)	0.1148*** (0.0191)	0.0016 (0.0152)
$\rho_{\chi\alpha*}$	-0.0448*** (0.0138)	0.0834*** (0.0240)	-0.0172 (0.0147)
μ_ξ’	-0.0568*** (0.0017)	-0.0450*** (0.0031)	-0.0689*** (0.0015)
λ_χ	-0.0385 (0.0664)	-0.0109 (0.1040)	-0.0532 (0.0822)
λ_α	-0.0067* (0.0040)	-0.0079 (0.0076)	-0.0072 (0.0045)
$\lambda_{\alpha*}$	-0.0137*** (0.0032)	-0.0072 (0.0057)	-0.0185*** (0.0039)
σ_η	0.0128*** (0.0001)	0.0156*** (0.0002)	0.0087*** (0.0001)
Log-likelihood	32972,4	15695,6	18133,6
AIC	32940.4	15663.6	18101.6
SIC	32872.2	15606.5	18044.5

TABLE X

HEATING OIL (NYMEX). OUT-OF-SAMPLE PREDICTIVE ABILITY

The table presents several metrics trying to compare the out-of-sample predictive power ability of the Schwartz and Smith (2000) two-factor model with deterministic seasonality and the four-factor model with stochastic seasonality. The results are obtained valuing the contracts in data set 3 with the parameters obtained estimating the models with data set 2.

PANEL A: SCHWARTZ AND SMITH (2000) TWO-FACTOR MODEL

	Mean error (Real-Predicted)	Std. Dev. Error	Std. Dev. error (% mean)	Root Mean Squared Error	Root Mean Squared Error (% mean)
F1	0,002912	0,045427	1,421227	0,04552	0,014241
F3	-0,00269	0,039435	1,234471	0,039527	0,012373
F5	-0,00541	0,037057	1,161716	0,037451	0,01174
F7	-0,01228	0,03582	1,125849	0,037866	0,011902
F10	-0,01366	0,031593	0,99601	0,034421	0,010852
F12	-0,00553	0,029557	0,933671	0,030069	0,009498
F14	0,002676	0,026995	0,854329	0,027127	0,008585
F16	0,004152	0,024821	0,786907	0,025166	0,007978
F18	-0,00438	0,025607	0,813414	0,02598	0,008252

PANEL B: FOUR-FACTOR MODEL WITH STOCHASTIC SEASONALITY

F1	0,000999	0,04439	1,388795	0,044401	0,013891
F3	-0,00543	0,038672	1,210584	0,039051	0,012225
F5	-0,00201	0,034761	1,089737	0,034819	0,010916
F7	-0,00136	0,032437	1,019496	0,032465	0,010204
F10	-0,00336	0,027752	0,874921	0,027954	0,008813
F12	-0,00336	0,027192	0,858962	0,027398	0,008655
F14	-0,00166	0,025963	0,821682	0,026016	0,008234
F16	0,000415	0,024686	0,782607	0,024689	0,007827
F18	-0,00206	0,025379	0,806161	0,025462	0,008088

TABLE XI

GAS OIL AND NATURAL GAS (ICE FUTURES, LONDON) AND RBOB GASOLINE (NYMEX). FOUR-FACTOR MODEL AND SCHWARTZ AND SMITH (2000) TWO-FACTOR MODEL

The table presents the results for the Schwartz and Smith (2000) two-factor model with deterministic seasonality and for the four-factor model with stochastic seasonality. Standard errors in parentheses. The estimated values are reported with * denoting significance at 10%, ** denoting significance at 5%, and *** denoting significance at 1%.

	GAS OIL IPE FOUR-F.	GAS OIL IPE DETERM.	GASOLINE NYMEX FOUR-F.	GASOLINE NYMEX DETERM.	NATURAL GAS IPE FOUR-F.	NATURAL GAS IPE DETERM.
Number obs.	364	364	113	113	131	131
μ_ξ	0.1960***	0.1962***	0.4282***	0.5193***	0.1600**	0.1918
	(0.0389)	(0.0387)	(0.0782)	(0.1093)	(0.0626)	(0.0641)
k	1.3650***	1.4259***	1.5407***	0.5998***	2.0348***	4.2371***
	(0.0236)	(0.0201)	(0.3065)	(0.0904)	(0.4678)	(0.2891)
Φ	1.0217***	0.9950***	0.9580***	0.9733***	1.0142***	0.9907***
	(0.0038)	(0.0007)	(0.0124)	(0.0020)	(0.0057)	(0.0015)
σ_ξ	0.1522***	0.1528***	0.1768***	0.2889***	0.1942***	0.1918***
	(0.0042)	(0.0042)	(0.0095)	(0.0475)	(0.0148)	(0.0097)
σ_χ	0.2685***	0.2663***	0.3587***	0.5243***	0.4275***	0.5148***
	(0.0079)	(0.0078)	(0.0295)	(0.0606)	(0.0284)	(0.0376)
σ_α	0.0257***		0.0487***		0.1021***	
	(0.0009)		(0.0037)		(0.0074)	
$\rho_{\xi\chi}$	-0.0117	0.0004	-0.0673	-0.7898***	-0.2348	0.2465**
	(0.0409)	(0.0407)	(0.1960)	(0.0785)	(0.1861)	(0.1012)
$\rho_{\xi\alpha}$	0.1127***		-0.0053		-0.1011	
	(0.0427)		(0.1297)		(0.1452)	
$\rho_{\xi\alpha*}$	0.1176***		0.1791*		0.2293***	
	(0.0428)		(0.1041)		(0.0885)	
$\rho_{\chi\alpha}$	-0.1630***		0.5269***		0.3204**	
	(0.0447)		(0.0924)		(0.1465)	
$\rho_{\chi\alpha*}$	-0.1842***		-0.3494***		-0.0712	
	(0.0460)		(0.0790)		(0.1016)	
μ_ξ’	-0.0357***	-0.0366***	-0.1470***	-0.2682***	-0.0928***	-0.0309**
	(0.0012)	(0.0012)	(0.0256)	(0.0814)	(0.0296)	(0.0128)
λ_χ	0.0974	0.0970	-0.2805*	-0.4626*	-0.1419	-0.1075
	(0.0677)	(0.0660)	(0.1670)	(0.2507)	(0.1430)	(0.1885)
λ_α	0.0107***	-0.0022	-0.0085	-0.0226***	0.0211	0.1909***
	(0.0041)	(0.0015)	(0.0195)	(0.0042)	(0.0280)	(0.0135)
$\lambda_{\alpha*}$	-0.0192***	-0.0044**	0.0377**	0.0191	-0.0430	-0.0263
	(0.0043)	(0.0018)	(0.0184)	(0.0160)	(0.0270)	(0.0203)
σ_η	0.0115***	0.0134***	0.0104***	0.0144***	0.0561***	0.0686***
	(0.0001)	(0.0001)	(0.0003)	(0.0003)	(0.0009)	(0.0011)
Log-likelihood	23744.84	23190.87	4757.064	4571.973	4732.799	4400.313
AIC	23712.84	23168.87	4725.064	4549.973	4700.799	4378.313
SIC	23650.49	23126	4681.426	4519.971	4654.796	4346.686

CHAPTER 3: CRUDE OIL AND REFINED PRODUCTS. A COMMON LONG-TERM TREND

1. Introduction

In recent years the study of the stochastic behaviour of commodity prices has grown in importance among academics and practitioners, since it plays a central role in the valuation and hedging of commodity contingent claims and in defining procedures for evaluating natural resources investment projects, especially in determining the optimal investment rules.

In spite of the existing empirical evidence, which suggests that commodity prices show mean-reversion and the volatility of futures returns is a decreasing function of time, the first works on the stochastic behaviour of commodity prices assumed, as in the equity assets context, that commodity prices follow a geometric Brownian motion, which implies a constant rate of growth in the commodity price and a constant volatility of futures returns (see for example Brennan and Schwartz, 1985; Paddock et al., 1988, among others).

More recently, several authors, such as Laughton and Jacobi (1993) and (1995), Ross (1997) or Schwartz (1997), have considered that a mean-reverting process is more appropriate to model the stochastic behaviour of commodity prices. Unfortunately, these one-factor mean-reverting models are not very suitable, since they generate a volatility of futures returns which goes to zero as the time to maturity of the futures contract approaches infinity, which is not a realistic assumption.

In addition, these models, that consider only one source of uncertainty, are not very reasonable since they entail that futures prices for different maturities should be perfectly correlated, which defies existing evidence. Looking for more realistic results, Schwartz (1997), Schwartz and Smith (2000), Cortazar and Schwartz (2003) and Cortazar and Naranjo (2006), among others, have developed multi-factor models. All these multi-factor models assume that the spot price is the sum of short-term and long-term components. Long-term factors account for the long-term dynamics of commodity prices, which is assumed to be a random walk, whereas the short-term factors account for the mean-reversion components in the commodity prices.

Most of these articles are focused on crude oil prices and not on the refining products derived from crude oil. Briefly speaking[17], due to the refining process 47% of a crude oil barrel is transformed into gasoline, 24% into diesel fuel and heating oil, 13% into jet fuel oil, 4% into heavy fuel oil, 4% into liquefied Petroleum Gas (LPG) and 8% into other products like asphalts and others more. Each of these products has a market price which quotes in an organized market, as crude oil. Therefore, there is a relation between product prices and crude oil prices and the difference between refining product and crude oil prices is known as the refining margin.

An interesting question that will not be addressed in this chapter (and has not been taken into account in the literature until recently) is the fact that both gasoline and heating oil are seasonal whereas crude oil is not. This is because seasonality in these products goes in opposite direction. Specifically, heating oil prices are higher in winter months, whereas gasoline prices are higher in summer months. The sum of these effects results in non-seasonality for oil prices (see, for example, Tolmasky and Hyndanov, 2002, or Chapter 2).

In order to keep our models parsimonious, we shall use no seasonal models, as they would only improve the refined product part. The reader should keep in mind, however, that a seasonal model is more realistic and convenient when only gasoline or heating oil are considered (see Chapter 2 for a more detailed analysis of seasonal continuous time models).

This work aims to understand the relationship between crude oil and refining product prices. The first issue to take into account is the fact that the refining margin is an stationary series, whereas crude oil and refined products prices series are not. Therefore, crude oil and refining

[17] See the Oil Market Report (2006) elaborated by the International Energy Agency for more information about these issues.

products prices should be cointegrated. In this work we present an empirical analysis of the relation among crude oil (WTI) prices and the main refining products prices (gasoline and heating oil), traded at NYMEX, using unit root and cointegration tests. In previous studies like Serletis (1994), Gjolberg and Johnsen (1999), Asche, Gjolberg and Völker (2003) and Lanza, Manera and Giovanninni (2005) we can find evidence of unit root and cointegration in crude oil and refined product prices, but no evidence of stacionarity in the refining margin is found. Nevertheless, these works carry out their studies with more refining products. However, in this chapter we demonstrate that these three commodities are not only cointegrated, but they have also a common long-term dynamics. The first evidence of it is achieved through a principal component analysis. Following the studies carried out by Clewlow and Strickland (2000) or Tolmasky and Hyndanov (2002), it can be considered that the first principal component is a long-term factor and the second and third ones are short-term factors. In this work we show that when we calculate the principal components of two commodities jointly, the sign of the first principal component does not change depending on the maturity of the futures contracts, but the second and the third do, which is a new evidence of a common long-term dynamics.

A definitive evidence of this fact is achieved by proposing different factor models to explain the dynamics of commodity prices jointly. It is found that the most suitable model in terms of simplicity and fit is the one which assumes a common long-term trend for all three commodities.

This fact will have straightforward applications in the valuation and hedging of commodity contingent claims and in defining procedures for evaluating investment projects in natural resources, especially when determining optimal investment rules. Specifically, we use these results to value the so-called crack-spread options quoted at NYMEX, and we find that, assuming a common long-term trend for crude oil and refined product prices, option valuation is as accurate as using models with more factors and parameters.

This chapter is organized as follows. Section 2 presents some preliminary results about the existence of a common long-term trend for crude oil prices and the most important refining products prices, i.e. the results of the unit root and cointegration tests and the principal component analysis. The results of the estimation of the factor models with a common long-term dynamics are contained in section 3. Section 4 shows the valuation results of the so-called crack-spread options quoted at NYMEX, with a model which assumes a common long-term trend for crude oil and the main refining product prices, with a model which allows for a long-term trend for each commodity and with a model which postulates uncorrelated models for each commodity. Finally, section 5 concludes with a summary and discussion.

2. Unit Root Tests, Cointegration tests and Principal Component Analysis

In this section we find evidence of unit root in crude oil, gasoline and heating oil spot prices, but not in the refining margin. Taking into account that, as it is said above, the main part of a crude oil barrel is transformed into gasoline and heating oil, this fact suggests that these three commodities should be cointegrated. In order to check it, cointegration tests have been implemented. It is found that these three commodities are not only cointegrated, but they also share a common long-term dynamics. The first evidence of it is achieved through a principal component analysis. More evidence on these issues will be presented in the next section.

2.1. Unit Root Tests

Next we present three unit root tests to show that the refining margin is stationary, whereas crude oil, gasoline and heating oil prices are not. The first one is the Augmented Dickey-Fuller (ADF) test and the second one is the Phillips-Perron test. Both of them are based on the Dickey-Fuller test, which is only valid if the series is an AR(1) process and tests if the AR(1) coefficient minus one is statistically different from cero (Dickey and Fuller, 1979).

As the basic Dickey-Fuller test needs a rather concrete specification, several alternatives have been proposed. The ADF test allows for higher-order correlation by adding lagged difference terms of the dependent variable. Said and Dickey (1984) demonstrate that the ADF test remains valid even when the series has a moving average (MA) component, provided that enough lagged difference terms have been added to the regression. Phillips and Perron (1988) propose a nonparametric method to control for higher-order serial autocorrelation.

The third unit root test is based on the works by Boswijik (2001) and Boswijik and Doornik (2005). These authors point out that standard Dickey-Fuller tests based on LS estimators are often sensitive in the presence of GARCH errors, which is a typical phenomenon in financial high-frequency data. This problem becomes serious when the volatility process is near integrated. Therefore, given that our commodity prices series are serious candidates to show GARCH errors, the test proposed by Boswijk and Doornik should be implemented. This test is based on a likelihood ratio statistic, which substantially improves the asymptotic local power of the standard Dickey-Fuller tests. The likelihood ratio test will be based on the following model:

$$
\begin{aligned}
\Delta X_t &= (\phi - 1)(X_{t-1} - \mu) + \varepsilon_t \\
\varepsilon_t &= h_t^{1/2}\eta_t ; \eta_t \; i.i.d.\, N(0,1) \\
h_t^{1/2} &= \beta_0 + \beta_1 \varepsilon_{t-1}^2 + \beta_2 h_{t-1}
\end{aligned}
$$

Where X_t is a commodity price series and the parameter ϕ describes the degree of mean-reversion. The null hypothesis will be H_0: $(\phi - 1) = 0$, which is tested against the alternative H_1: $(\phi - 1) < 0$. The distribution for the likelihood ratio statistic (under the null) is approximated by a gamma distribution (see Boswijk and Doornik, 2005).

To the best of the author knowledge there are not spot prices for gasoline, heating oil or crude oil associated with the futures traded at NYMEX. Therefore, weekly observations for WTI (light sweet) crude oil prices, RBOB gasoline and heating oil from 9/9/1996 to 9/4/2006 (522 observations) of one month futures prices are going to be used to test cointegration relations.

The refining margin used in the tests below is calculated by subtracting the cost of WTI crude oil and the freight cost from the value of oil products produced by a refinery in the US Gulf Coast (catalytic cracking refinery). The value of oil products is calculated by adding together a fixed percentage of each refined oil product produced by a refinery. Weekly observations for this refining margin from 04/20/2001 to 09/15/2006 (283 observations) are used in the tests.

The main descriptive statistics of the series are contained in Table 1. Table 2 presents the results for the unit root tests for gasoline, heating oil and WTI crude oil prices, together with the refining margin described above. Our results indicate significant presence of GARCH errors in all the commodity series[18], which confirms that the Boswijk and Doornik test should be implemented. It is found that at 1% significance level the unit root test hypothesis is rejected for the refining margin with the ADF and the Phillips-Perron tests and at 5% significance level it is also rejected with the Boskijk-Doornik test[19], but it is not possible to reject it for gasoline, heating oil and crude oil prices, even at 10% significance level. Thus, the empirical evidence already found in previous studies of a unit root in crude oil, heating oil and gasoline prices is confirmed in the present work even in the presence of GARCH errors. Therefore, given that the refining margin does not show evidence of a unit root, we can conclude that there should be a cointegration relation among crude oil prices and the main refining product prices.

2.2. Cointegration Tests

Cointegration is the phenomenon that occurs when each component, $Y_{i,t}$ (i = 1,...,k), of a vector time series process Y_t is a unit root process, possibly with drift, but certain linear combinations of theses components are stationary. In the previous section it was found that gasoline, heating oil and crude oil prices have a unit root, but the refining margin has not. As stated above, the refining margin is a linear combination of these commodity prices and other more like jet fuel oil, heavy fuel oil, liquefied petroleum gas (LPG), asphalts, etc. These two facts together suggest a cointegration relation among all the refining products and the crude oil. However, it is not possible to conclude that there is a cointegration relation among crude oil, heating oil and gasoline. In this section we show that this cointegration relation indeed exists, even more, there

[18] For short, the results of the estimation of the GARCH models are not presented in the paper, although they are available from the authors upon request.

[19] It should be noted that at 1% significance level the unit root test hypothesis is almost rejected with the Boswijk-Doornik test, for the refining margin.

are also cointegration relations between crude oil and heating oil, crude oil and gasoline, heating oil and gasoline and also among heating oil, gasoline and crude oil. This is a first evidence of a common long-term dynamics among the three commodity prices.

In order to demonstrate that these cointegration relations exist, the Johansen test (Johansen, 1988) will be used. The approach by Johansen consists in estimating the Vector Error Correction Model (VECM), which is a model for the vector time series in first differences, including the cointegration relation. The model is estimated by maximum likelihood, under various assumptions about the trend or intercept parameters and the number "r" of cointegrating vectors. Once the model has been estimated, we can conduct likelihood ratio tests. Assuming that the VECM errors are independently distributed, and given the cointegrating restrictions on the trend or intercept parameters, the maximum likelihood is a function of the cointegration rank r. Johansen proposes two types of tests for r: The lambda-max test and the trace test.

The first one is based on the log-likelihood ratio $\ln[\mathrm{Lmax}(r)/\mathrm{Lmax}(r+1)]$, and is conducted sequentially for $r = 0,1,..,k-1$. In this test the null hypothesis is that the cointegration rank is equal to r and the alternative is that the cointegration rank is equal to $r + 1$. The second one is based on the log-likelihood ratio $\ln[\mathrm{Lmax}(r)/\mathrm{Lmax}(k)]$, and is conducted sequentially for $r = k-1,...,1,0$. In this test the null hypothesis is that the cointegration rank is equal to r and the alternative is that the cointegration rank is k. In this study we will present the results from the first one, however it has been checked that the same results are obtained using the second one. Johansen´s test also dispatches the cointegration relations if it exists.

As in the previous section, weekly observations from 9/9/1996 to 9/4/2006 (522 observations) of one month futures prices for crude oil, gasoline and heating oil, traded at NYMEX, are going to be used to test cointegration relations.

Tables 3, 4, 5 and 6 present the results for the cointegration test and the normalized cointegrating coefficients for gasoline and crude oil, heating oil and crude oil, gasoline and crude oil and, finally, for gasoline, heating oil and crude oil. All the tests have been performed assuming that it is possible to find a deterministic trend in the cointegration equation. However, similar results have been obtained when we repeated the tests assuming that this kind of trend is not feasible.

As can be seen in the tables, the results suggest that crude oil, gasoline and heating oil prices have a common trend. At 1% significance level, there are cointegration relations between gasoline and heating oil prices, between gasoline and crude oil prices and between heating oil and crude oil prices, and the (normalized) cointegrating coefficients are, more or less, 1 and -1 in all cases. There is also a cointegration relation among gasoline, heating oil and crude oil prices at the same significance level.

2.3. Principal Component Analysis

In this section we carry out a principal component analysis with the three commodities presented above trying to find more evidence of a common trend. As in the previous sub-sections, we use weekly observations for WTI (light sweet) crude oil prices, RBOB gasoline and heating oil from 9/9/1996 to 9/4/2006, traded at NYMEX. However, given that we need to analyze the term structure of futures contracts prices, the twelve futures contracts closest to expiration will be employed (522 weekly observations). The principal component analysis will be based on the eigenvalue decomposition of the covariance matrix of the weekly log-returns.

Figure 1 shows the results of the principal component analysis for each commodity separately. The first three principal components are depicted in the charts. The first component and, therefore, the primary dynamics of the forward curve, the shift, is a roughly parallel shift of the whole curve either up or down depending on the direction of the random shock. The fact that the function is not flat indicates that the volatility of the short end of the forward curve is greater than the long end one. The second and third components represent tilts and bendings of the forward curve respectively. The tilt is a movement of the short end of the forward curve either up or down depending on the direction of the random shock and a movement of the long end of the curve in the opposite direction. The bending is a movement of the short end and the long end of the forward curve either up or down depending on the direction of the random shock and a movement of the middle end of the curve in the opposite direction.

Moreover, it should be noted that the first component has the same sign for all maturities and does not go to zero as maturity goes to infinity, implying that a random shock following this component has the same direction for all maturities and does not vanish with time. However, the second and third components have a certain sign for some maturities and the opposite sign for the other ones, implying that a random shock following these components has a certain direction in some periods and the opposite direction in other periods and, therefore, in the long-term its effect tends to vanish. For these reasons we can consider that the first principal component is a long-term component, whereas the second and third components are short-term ones.

The same conclusion is derived from Table 7. These results are similar to those obtained by Clewlow and Stricland (2000) and Tolmasky and Hyndanov (2002). It can be appreciated from Table 7 that the first component explains more than 90% of the volatility (more than 98% in the case of the crude oil). This evidence is consistent with the extant literature about modelling the stochastic behaviour of commodity prices, where it is always considered one long-term factor and one or more short-term factors (see for example Cortazar and Schwartz, 2003; Cortazar and Naranjo, 2006). As in Clewlow and Stricland (2000) or Tolmasky and Hyndanov (2002), the first three components explain almost 100% of the volatility.

In spite of the fact that in all cases the first principal component explains more than the 90% of the volatility, as can be appreciated in the first panel of Table 7, this explanatory power is greater for crude oil than for the other two commodities. This means that short-term shocks have more importance in refining product prices than in crude oil prices. This is coherent with previous papers, see for example Radchenko (2005a) and (2005b), who state that in the case of refining products, particularly in the case of gasoline, there are short-term imperfections in the markets like lags and asymmetries in response to movements in crude oil prices. Other cause could be the fact that, as explained in the Introduction Section and as can be seen in Tolmasky and Hyndanov (2002) or in Chapter 2, gasoline and heating oil prices are seasonal whereas crude oil prices are not, and therefore as the principal components are calculated without considering this fact, seasonality could be taken as a short-term shock. Moreover, the lack of liquidity of the refining products long maturity futures prices quotation, especially in the case of gasoline, could be other explanation. Other possible explanation could be related with the fact that gasoline is more volatile than the other two commodities, as can be appreciated in Figure 1 and also in Table 1.

Figure 2 shows the results of the principal component analysis with the three commodities considered in this study joined in pairs. In this case, given that the estimation is performed with more than one commodity, the same number of futures contracts have been used for each commodity, in order not to upset the balance of each commodity. Similar maturities were also taken for each commodity futures, so that short-term and long-term relationships were not decompensated[20]. Thus, we take the first, third, fifth, seventh, ninth and eleventh maturating futures contracts for the first commodity and the second, fourth, sixth, eighth, tenth and twelfth maturating futures contracts for the second one, and we carry out the principal component analysis in the same way as before.

Looking at Figure 2 we can appreciate that the first component follows the same pattern as in Figure 1, but the second and the third principal components do not. This is new evidence in favour of a common long-term trend for these three commodities, despite of their differences in the short-term. Moreover, from the second panel of Table 7 it can be appreciated that the percentage of the volatility explained by each component, with the three commodities joined in pairs, is broadly the same than in the case of each commodity separately, which confirms the evidence of only one long-term trend for all commodities, since the first principal component explains more than the 90% of the volatility.

The same conclusion is obtained when we perform the principal component analysis with the three commodities jointly (Figure 3). Specifically, we take the first, fourth, seventh and tenth

[20] Similar results have been obtained with futures with different maturities.

maturating futures contracts for heating oil, the second, fifth, eighth and eleventh for WTI crude oil and the third, sixth, ninth, and twelfth maturating futures contracts for unleaded gasoline, and carried out the principal component analysis. In this case the first component explains 92.93% of the volatility, the first two components 94.90% and the first three 96.25%.

Given these first signs of a common long-term for the three commodities considered in this study, the next section tries to shed more light on these issues by proposing different factor models for the stochastic behaviour of commodity prices, including models assuming a common long-term trend for all three commodities.

As it is said above, the previous studies about the relation among crude oil and refining product prices were focused on the cointegration relations among them. However, as we will see in the following sections, in this chapter we also find evidence of a common long-term trend among crude oil and refining product prices.

3. Factor Models

In order to find definitive evidence of a common long-term trend among crude oil, gasoline and heating oil prices, we are going to fit the data to different factor models. It seems clear that modelling each commodity separately is the way to get the best fit to data. However, if we get a similar goodness of fit when modelling the three commodities jointly with a common long-term trend, the conclusion is a definitive evidence of a common long-term trend. Of course, it is also possible to compare the results of modelling the commodities jointly with and without a common long-term trend. If there is a common long-term trend the results must be comparable.

3.1. Theoretical Models

In order to model each commodity separately, we shall use the two factor model proposed by Schwartz and Smith (2000). Given the existing empirical evidence (see for example Schwartz, 1997), this is a reasonable approach for this kind of commodities. In this model the log-spot price (X_t) is assumed to be the sum of two stochastic factors: a short-term deviation (χ_t) and a long-term equilibrium price level (ξ_t):

$$X_t = \xi_t + \chi_t \tag{1}$$

The stochastic differential equations (SDE) of these factors are:

$$\begin{cases} d\xi_t = \mu_\xi dt + \sigma_\xi dW_{\xi t} \\ d\chi_t = -\kappa \chi_t dt + \sigma_\chi dW_{\xi t} \end{cases} \tag{2}$$

where $dW\xi_t$ and $dW\chi_t$ can be correlated ($dW\xi_t dW\chi_t = \rho\xi\chi dt$).

In order to test the existence of a common long-term trend for each pair of commodities, we will compare the model goodness of fit for each commodity separately in the context of the two factor model presented above, and the goodness of fit for the commodities in pairs in a joint model in which the log-spot price (X_{it}) of the commodity "*i*" is assumed to be the sum of two stochastic factors: a short-term deviation (χ_{it}), different for each commodity, and a common long-term equilibrium price level (ξ_t) for both commodities: $X_{it} = \xi_t + \chi_{it}$ (i = 1,2). We shall also compare this joint model, which possesses a common long-term trend for the pair of commodities, with another joint model without a common long-term trend in which the log-spot price (X_{it}) of each commodity is taken as the sum of two stochastic factors: a short-term deviation (χ_{it}) and a different long-term equilibrium price level (ξ_{it}) for each commodity: $X_{it} = \xi_{it} + \chi_{it}$ (i = 1,2).

The SDE of the factors are the same as the ones presented above. For the model with common long-term trend they are:

$$\begin{cases} d\xi_t = \mu_\xi dt + \sigma_\xi dW_{\xi t} \\ d\chi_{1t} = -\kappa_1 \chi_{1t} dt + \sigma_{\chi 1} dW_{\chi 1t} \\ d\chi_{2t} = -\kappa_2 \chi_{2t} dt + \sigma_{\chi 2} dW_{\chi 2t} \end{cases} \tag{3}$$

where any correlation structure can exits among $dW_{\xi t}$, $dW_{\chi 1t}$ and $dW_{\chi 2t}$ ($dW_{\xi t}dW_{\chi 1t} = \rho_{\xi\chi 1}dt$, $dW_{\xi t}dW_{\chi}2t = \rho\xi\chi 2dt$ and $dW\chi 1tdW\chi 2t = \rho\chi 1\chi 2dt$).

For the model without common long-term trend the SDE are:

$$\begin{cases} d\xi_{1t} &= \mu_{1\xi}dt + \sigma_{1\xi}dW_{\xi 1t} \\ d\xi_{2t} &= \mu_{2\xi}dt + \sigma_{2\xi}dW_{\xi 2t} \\ d\chi_{1t} &= -\kappa_1\chi_{1t}dt + \sigma_{\chi 1}dW_{\chi 1t} \\ d\chi_{2t} &= -\kappa_2\chi_{2t}dt + \sigma_{\chi 2}dW_{\chi 2t} \end{cases} \tag{4}$$

where again $dW_{\xi 1t}$, $dW_{\xi 1t,}$ $dW_{\chi 1t}$ and $dW_{\chi 2t}$ can show any correlation structure ($dW_{\xi 1t}dW_{\xi 2t} = \rho_{\xi 1\xi 2}dt$, $dW_{\xi 1t}dW_{\chi 1t} = \rho_{\xi 1\chi 1}dt$, $dW_{\xi 1t}dW_{\chi 2t} = \rho_{\xi 1\chi 2}dt$, $dW_{\xi 2t}dW_{\chi 1t} = \rho_{\xi 2\chi 1}dt$, $dW_{\xi 2t}dW_{\chi 2t} = \rho_{\xi 2\chi 2}dt$ and $dW_{\chi 1t}dW_{\chi 2t} = \rho_{\chi 1\chi 2}dt$).

As can be appreciated, the model for each commodity separately and the model for pairs of commodities jointly have the same SDE. However in the first case, there is no correlation between factors of different commodities, which is, clearly, an undesirable property in valuing commodity contingent claims, as shall be discussed below.

If there is a common long-term trend for pairs of commodities, the log-likelihood for these three methodologies (one model for each commodity separately, the joint model with common long-term trend and the joint model without common long-term trend) must be similar. In this case, the most suitable model for pairs of commodities is the joint model with common long-term trend. This model is preferable to the joint model without common long-term trend as it contains less parameters (being, therefore, simpler) and because it has only one long-term factor, which is an advantage in valuing long-term commodity contingent claims. From our point of view, the economic interpretation of both commodities to be common long term trended is also quite appealing in qualitative terms. The point is that, all things equal, a model whose parameters are interpretable is better. Let us briefly compare our models from a qualitative point of view.

In a standard two factor model framework each commodity factors are, by design, uncorrelated to the other commodity ones. So nothing can be said about the relationship between both series. Should we impose or estimate a correlation structure among different commodities factors we would end up with 4 more parameters and therefore the model would be equivalent to a four factor model. Four factor models, such as the model in (4), i.e. the joint model without common long term trend for commodities in pairs, do account for the relationship between both series but they do so in a rather ambiguous way. We have 6 correlations to look at, and none of them is negligible, in general. We have two correlated long term trends but the relationship between both series does not stop here. We can not take this correlation as the only measure, as the long term trend for crude oil is also correlated with the short term trend for refined products. Moreover, one cannot answer to questions like, what is the market general trend? This question is meaningless in our model without common long term trend unless we assume some combination of both long term trends as “representative”.

In contrast, with only one long term trend this question can be fully answered, as the general trend is the common long term trend. We can even see the relationship between the general trend and each of the series, just looking at its long run-short run correlation coefficient. Moreover, the (isolated) influence of a series on the other can be directly seen just looking at the short term-short term correlation coefficient.

Finally, the joint model with a common long-term trend is also more suitable than a model for each commodity separately (without correlation among different commodities factors) as the first one has less parameters and factors and because, as stated above, it takes into account the correlation between different commodities factors, which is essential in valuing commodity contingent claims, a fact that will be discussed below.

To test the existence of a common long-term trend for the three commodities jointly we shall compare the goodness of fit if we estimate each commodity separately (using a two factor model as presented above) and the goodness of fit for a joint model (for the three commodities together) in which the log-spot price (X_{it}) of the commodity "*i*" is assumed to be the sum of two stochastic factors: a short-term deviation **(χ_{it}),** which is different for each commodity, and a common long-term equilibrium price level (ξ_t), which is equal for all commodities[21]: $X_{it} = \xi_t + \chi_{it}$ (i = 1,2,3).

The SDE of the factors for the joint model with common long term trend for all three commodities are the same as before:

$$\begin{cases} d\xi_t = \mu_\xi dt + \sigma_\xi dW_{\xi t} \\ d\chi_{1t} = -\kappa_1 \chi_{1t} dt + \sigma_{\chi 1} dW_{\chi 1t} \\ d\chi_{2t} = -\kappa_2 \chi_{2t} dt + \sigma_{\chi 2} dW_{\chi 2t} \\ d\chi_{3t} = -\kappa_3 \chi_{3t} dt + \sigma_{\chi 3} dW_{\chi 3t} \end{cases} \qquad (5)$$

where $dW_{\xi t}$, $dW_{\chi 1t}$, $dW_{\chi 2t}$ and $dW_{\chi 3t}$ can show any correlation structure ($dW_{\xi t}dW_{\chi 1t} = \rho_{\xi\chi 1}dt$, $dW_{\xi t}dW_{\chi 2t} = \rho_{\xi\chi 2}dt$, $dW_{\xi t}dW_{\chi 3t} = \rho_{\xi\chi 3}dt$, $dW_{\chi 1t}dW_{\chi 2t} = \rho_{\chi 1\chi 2}dt$, $dW_{\chi 1t}dW_{\chi 3t} = \rho_{\chi 1\chi 3}dt$ and $dW_{\chi 1t}dW_{\chi 3t} = \rho_{\chi 1\chi 3}dt$).

As in the previous case, if there is a common long-term trend for the three commodities, the log-likelihood for both methodologies (one model for each commodity separately and the joint model with common long-term trend) must be similar. Should this be the case, the most suitable model for the three commodities will be the joint model with common long-term trend, because it is simpler (it has less structure) and because it takes into account the correlation between different commodities factors, which is essential in valuing commodities contingent claims, as we will discuss below.

In all models presented in this sub-section, the log-spot price of each commodity is the sum of two factors: a short-term deviation **(χ_t)** and a long-term equilibrium price level (ξ_t) ($X_t = \xi_t + \chi_t$). Therefore, the expression for the log-price of a futures contract for each

[21] Of course, we could also estimate a joint model without common long term trend for all three commodities. However, apart from the fact that this model would have a high number of parameters, its interpretation would be ambiguous, as said above for the case of the joint model without common long term trend for commodities in pairs.

commodity with maturity at time “T+t” traded at time t is the one obtained by Schwartz and Smith (2000)[22]:

$$\ln[F(X_t, T+t)] = \xi_0 + e^{-kT}\chi_0 + A(T) \qquad (6)$$

where:

$$A(T) = (\mu_\xi' + 0.5\sigma_\xi^2)T - (1-e^{-kT})\lambda_\chi / k + \sigma_\xi\sigma_\chi\rho_{\xi\chi}(1-e^{-kT})/k + 0.25\sigma_\chi^2(1-e^{-2kT})/k$$

As explained in Schwartz and Smith (2000), the risk-neutral version of the model is necessary to calculate this expression. This is the reason why there is a risk-premium for the short-term deviation (λ_χ) and also a long-term drift corrected by risk (μ_ξ’) instead of the original one for the long-term equilibrium price level.

Finally, it should be noted that the differences between the models presented above lay on the number of factors and in the correlation assumed among them.

3.2. Estimation Methodology

Given that the factors (or state variables) are not directly observable, the model’s parameters must be estimated using the Kalman Filter. This methodology enables the estimation of the likelihood of a data series, given a particular set of model parameters and a prior distribution of the variables, which allows the estimation of the parameters by maximum likelihood techniques[23].

The Kalman filter technique is a recursive methodology that estimates the unobservable time series, the state variables or factors (Z_t), based on an observable time series (Y_t) which depends on these state variables. The *measurement equation* accounts for the relationship between the observable time series and the state variables:

[22] For the same reason the volatility of the futures returns has the same expression as in Schwartz and Smith (2000).
[23] Detailed accounts of Kalman filtering are given in Harvey (1989).

$$Y_t = d_t + M_t Z_t + \eta_t \qquad t = 1, \ldots, N_t \tag{7}$$

where $Y_t, d_t \in \Re^n, M_t \in \Re^{n\,x\,n},\ Z_t \in \Re^h$, h is the number of state variables, or factors, in the model, and $\eta_t \in \Re^n$ is a vector of serially uncorrelated Gaussian disturbances with zero mean and covariance matrix H_t.

The *transition equation* accounts for the evolution of the state variables:

$$Z_t = c_t + T_t Z_{t-1} + \psi_t \qquad t = 1, \ldots, N_t \tag{8}$$

where $c_t \in \Re^h, T_t \in \Re^{h\,x\,h}\mathit{and}\ \psi_t \in \Re^h$ is a vector of serially uncorrelated Gaussian disturbances with zero mean and covariance matrix Q_t.

Let $Y_{t|t-1}$ be the conditional expectation of Y_t and let Ξ_t be the covariance matrix of Y_t conditional on all information available at time t – 1. Then, after omitting unessential constants, the log-likelihood function can be expressed as:

$$l = -\sum_t \ln|\Xi_t| - \sum_t (Y_t - Y_{t|t-1})'\Xi_t^{-1}(Y_t - Y_{t|t-1}) \tag{9}$$

Two conditions need to be fulfilled in the original version of this methodology. Firstly, the data must have no missing points and, secondly, the length of vector Y_t must be independent of "*t*". Cortazar and Naranjo (2006) introduce a new version of this methodology to handle with incomplete data sets and vectors Y_t whose length depends on "*t*". However, with a data set with a lot of missing points in the futures contracts with longer maturities, we face the problem of the unbalance in the relationship between long and short effects. As explained below, the data set employed in this chapter has not missing points and all vectors have the same length. Therefore in this work the traditional version of the Kalman filter methodology can be used.

A discrete-time version of the models is needed in order to estimate the parameters of the models through the Kalman filter methodology. Given that the expression for the log-price of a futures contract for each commodity is independent of the model, the discrete time version of each model will be similar to the ones obtained by Schwartz and Smith (2000). The differences are due to the correlation among factors. The discrete time versions of the models presented above are developed in the Appendix.

3.3. Data

The data set employed in the estimation procedure consists on weekly observations of unleaded gasoline, heating oil and WTI crude oil futures prices traded at NYMEX[24].

There are currently futures traded at NYMEX for WTI crude oil from one month to seven years, for heating oil from one to eighteen months and for gasoline from one to twelve months. However, in the case of the gasoline there is not enough liquidity for the futures of longer maturities. In the estimations presented in this work, for models with more than one commodity, we have chosen to use futures contracts with the same maturities for all commodities, in order not to decompensate the short-term-long-term relations[25]. Therefore, to estimate the parameters for the different models presented above the following data sets have been set up:

- The data set for the joint model with common long-term trend for the three commodities is made of contracts F1, F3, F5, F7 and F9 from 06/30/1997 to 04/24/2006, which implies 461 quotations of each contract, where F1 is the contract for the month closest to maturity, F2 is the contract for the second month closest to maturity and so on.
- The data set for the joint model for WTI crude oil and heating oil (with and without common long-term trend) is made of contracts F1, F4, F7, F11, F15 and F18 from 09/09/1996 to 09/18/2006, which implies 522 quotations of each contract.
- The data set for the joint model for WTI crude oil and gasoline (with and without common long-term trend) is made of contracts F1, F3, F5, F7 and F9 from 06/30/1997 to 04/24/2006, which implies 461 quotations of each contract.

[24] Details about the contracts can be found in the NYMEX homepage.
[25] Similar results have been obtained with futures with different maturities.

- The data set for the joint model for heating oil and gasoline (with and without common long-term trend) is made of contracts F1, F3, F5, F7 and F9 from 06/30/1997 to 04/24/2006, which implies 461 quotations of each contract.
- The data set for the two factor model for gasoline is made of contracts F1, F3, F5, F7 and F9 from 06/30/1997 to 04/24/2006, which implies 461 quotations of each contract.
- There are two data sets for the two factor model for WTI crude oil. The first one is made of contracts F1, F3, F5, F7 and F9 from 06/30/1997 to 04/24/2006, which implies 461 quotations of each contract. The second one is made of F1, F4, F7, F11, F15 and F18 from 09/09/1996 to 09/18/2006, which implies 522 quotations of each contract. Depending of the case we will employ one or the other. The first data set is employed when crude oil is used jointly with gasoline, whereas the second one will be used in all other cases.
- There are also two data sets for the two factor model for WTI heating oil. The first one is made of contracts F1, F3, F5, F7 and F9 from 06/30/1997 to 04/24/2006, which implies 461 quotations of each contract. The second one is made of F1, F4, F7, F11, F15 and F18 from 09/09/1996 to 09/18/2006, which implies 522 quotations of each contract. As in the previous case we shall employ one or the other depending on the situation, the first data set when heating oil is used jointly with gasoline, the second one in all other situations.

The use of different data sets for gasoline and the other two commodities is due to liquidity constrains. Specifically, as stated above, in the case of gasoline the available futures contracts are less liquid and their maturities are shorter than the other two commodities contracts. Therefore, given that in the case of heating oil there are futures contracts with enough liquidity until eighteen months of maturity, we have decided to use a data set with more futures contracts and with futures contracts with longer maturities than in the case of gasoline, in which there are not enough liquidity for futures with maturities longer than nine months. In the case of crude oil there are available futures contracts until seven years of maturity. However, as it is said above, it has been decided to use crude oil futures contracts with the same maturity as the ones used for the other commodities, in order to not decompensate the short-term-long-term effects. Schwartz

(1997) realized that mean reversion effects tend to be lower for contracts with longer maturities. The same evidence in the case of the natural gas has been presented in chapter 2. Therefore, to avoid undesirable effects, for models with more than one commodity, it has been decided to use futures contracts with the same maturities for all commodities.

As explained in Schwartz (1997), since futures contracts have a fixed maturity date, the time to maturity changes as time progresses, but remains in a narrow time interval. This is the reason why, as in Schwartz (1997), it is assumed that the time to maturity does not change with time and it is equal to one month for F1, two months for F2 and so on.

3.4. Results

Table 8 presents the results for the two factor model applied to each commodity (WTI crude oil, heating oil and gasoline) separately and for all data sets described above. Tables 9 and 10 present the results for the joint model for pairs of commodities with and without common long-term trend respectively. Finally, Table 11 presents the results for the joint model with common long-term trend for all three commodities. As stated above, for models with more than one commodity, we have chosen to use futures contracts with the same maturities for each commodity.

The first notable issue is the fact that the gasoline price is more volatile than the other two commodity prices, since in Tables 8, 9, 10 and 11 the volatility coefficients are higher when the gasoline appears in the model. It can be also appreciated in Table 1 and in Figure 1.

If we define the Schwartz Information Criterion (SIC) as $\ln(L_{ML}) - q\ln(T)$, where q is the number of estimated parameters, T is the number of observations and L_{ML} is the value of the likelihood function, defined in (9), using the q estimated parameters, then the higher the SIC the better the fit. The same conclusions are obtained with the Akaike Information Criterion (AIC), which is defined as $\ln(L_{ML}) - 2q$.

The main result of this part of the work is the fact that there is a common long-term trend for these three commodities. To reach this conclusion, first of all we can appreciate that the SIC and the AIC values in the joint model with common long-term trend for pairs of commodities (Table 9) are more or less the same as the ones obtained with the joint model without common long-term trend (Table 10), and are also more or less the same as the sum of the SICs or the AICs values obtained with the two factor model for each commodity separately (Table 8). Specifically, in the case of heating oil and WTI crude oil, in the in the joint model with common long-term trend, for the period time from 09/09/1996 to 09/18/2006 and using the contracts F1, F4, F7, F11, F15 and F18, the SIC is 39195.4 (Table 9), whereas in the joint model without common long-term trend the SIC is 39456.3 (Table 10). The sum of the SICs estimated in the two factor model for each commodity separately is 23444.8 + 17761.8 = 41206.6 (Table 8). Therefore, we can conclude that all the models fit more or less the same and, consequently, as the joint model with common long-term trend is the simplest one, we don’t need a second long-term factor in modelling heating oil and WTI crude oil jointly. They have the same long-term trend.

Similar results are obtained for gasoline and WTI crude oil and for gasoline and heating oil for the period 06/30/1997 to 04/24/2006 and using contracts F1, F3, F5, F7 and F9 in both cases. In the case of gasoline and WTI crude oil the SIC in the joint model with common long-term trend is 26436.2 (Table 9) whereas in the joint model without common long-term trend and adding the SICs of the two factor models separately we obtain 27672.7 (Table 10) and 30113.9 (Table 8) respectively. In the case of gasoline and heating oil these figures are: 24306.2 (Table 9), 26313 (Table 10) and 25912.4 (Table 8) respectively.

Using the previous evidence we can conclude that the three commodities have the same long-term trend when we analyse them in pairs. Using the transitive property we could conclude that all three commodities have the same long-term trend. However, in order to reach a definitive evidence of this fact, we can just compare the SIC for the joint model with common long-term trend for the three commodities, which is 39033.7 (Table 11), and the sum of the SICs of the two factor model for each of these three commodities separately, which is 17812.6 + 13611.1 + 12301.3 = 43725 (Table 8). The period used goes from 06/30/1997 to 04/24/2006 and the contracts used are F1, F3, F5, F7 and F9. As the figures are similar, we can conclude definitively that the three commodities have the same long-term trend.

Moreover, it is worth noting that in Table 8 the values of the SIC and the AIC are higher for crude oil than for the other two commodities, using the same time period (06/30/1997 to 04/24/2006) and the same contracts (F1, F3, F5, F7 and F9)[26]. As in the case of the principal component analysis, the better fit to crude oil data could be related with the short term imperfections in the refining product markets, with the fact that the two factor models do not account for seasonality, whereas gasoline and heating oil prices are seasonal and, especially in the case of gasoline, or the lack of liquidity in longer maturity futures prices quotations. Nevertheless, in all cases the fit is well enough for the purpose of this study.

The relative fit of the models to our three commodity prices series can be assed more formally looking at their in-sample predictive ability. The in-sample predictive ability of the Schwartz and Smith (2000) two-factor model is presented for each commodity separately (Table 12), for pairs of commodities with common long-term trend (Table 13), and for pairs of commodities without common long-term trend (Table 14). It is found that, whenever the results are comparable (i.e. contracts F1, F3, F5, F7 and F9), although the model for each commodity separately performs slightly better (Table 12) in terms of the root mean squared error than the other two models, the differences, in general, are low. Therefore, given that the predictive ability of all three models is very similar, as before, we can conclude that all three commodities have the same long-term trend. Moreover, looking at the results in Table 12, it can be appreciated that the root mean squared error values for crude oil are lower than the corresponding values obtained with the other two commodities, which confirms our previous guess.

[26] It should be pointed out that, although in principle the values of the SIC and the AIC are not directly comparable whenever the series are different, our three commodity prices series have the same order of magnitude, the time period (06/30/1997 to 04/24/2006) and the futures contracts used in the estimation procedure (F1, F3, F5, F7 and F9) are the same.

Finally, it is also noticeable that previous studies have already found that crude oil and the main refined product prices are cointegrated but in the present work this conclusion is extended, as we find out that they have also a common long-term trend. Moreover, to the best of our knowledge this is the first time that a factor model with a common long-term trend for crude oil and its main refined product prices is proposed and estimated.

In the next section we shall use these results to value the so-called crack-spread options quoted at NYMEX, assuming a common long-term trend for crude oil and the main refining product prices.

4. Crack Spread Option Valuation

It is well known that commodity markets have been growing fast during the last years. According to the Hedge Funds Review (2008), we can distinguish three types of investors in commodity markets. Firstly, investors looking for a portfolio diversification tool, especially pension funds. Secondly, investors looking for a source of alpha, especially hedge funds. And thirdly, European banks, which use commodity derivatives to structure products that they retail to their customers. Therefore, in this context it is important to have techniques which can be implemented easily to value commodity derivatives, such as the so called crack spread options. In this section we present a technique to value these crack spread options assuming a common long-term trend, for WTI crude oil and gasoline and for WTI and heating oil, which can be implemented relatively easy and is simpler than the standard techniques.

4.1. Data

The data set employed in the estimation procedure consists on two sets of daily observations of crack-spread put and call options quoted at NYMEX. The first one contains Heating Oil vs. WTI and the second one Gasoline vs. WTI.

In NYMEX there are only quotations for these two sorts of crack spread options (Heating Oil vs. WTI and Gasoline vs. WTI) and contracts maturity occurs each month for the following eighteen months for different strike prices. Specifically, the strike prices are the at-the-money one, five additional strikes both above and below the established "at-the-money" strike price at \$0.25 (25¢) increments, three additional out-of-the-money strike prices are added above and below those strikes at \$1.00 intervals, and two additional strikes will be added above and below at \$2.00 intervals. Options traded at NYMEX are American style, so the holder can exercise his right at any time.

The market for this kind of options is much less liquid than the one for futures. Due to the scarcity of data, we have chosen to use daily instead of weekly data. This fact has forced us to do same minor changes in the model, as we shall see bellow. Even more, these liquidity constrains are the cause of the lack of data for many dates in many contracts. As a result, our data set includes daily quotations from January 2004 to January 2007. However, there are many missing dates or contracts. Specifically, for the case of the Gasoline vs. WTI crack-spread options we have data corresponding to four maturating dates: March, April, August 2006 and January 2007, with twelve exercise prices, from 5 to 16 dollars. In the case of the Heating Oil vs. WTI crack spread option we have data corresponding to contracts maturating from January 2004 to January 2007, with only two exercise prices available, which are 5 and 8 dollars. A brief summary of these options is given in table 15.

Let us give a brief review of the option description. Crack spread options are used to protect the refining margin, while at the same time allowing market participants to take advantages of favourable changes in the spread. A crack-spread call option is a contract that gives the holder the right (not the obligation) to buy a refined product futures contract from the writer and sell him a crude oil futures contract, paying a previously agreed crack spread price. Note that these three actions are simultaneous and can not be split, i.e. the holder is unable to just buy the contract and leave the rest of the transaction for later. Conversely, a crack spread put allows to sell a refined product futures contract and to buy a crude oil futures contract, paying a crack spread price.

4.2. Option Valuation Methodology

In order to value crack spread options, we shall follow the methodology described in Barraquand and Martineau (1995).

First of all, let us keep in mind what the problem is. We have a set of crack-spread American options that we would like to replicate. Of course, there is no close analytic expression for their price so one has to resort to simulation. In addition to that, our models are three dimensional (common trended one) or four dimensional (when we consider one trend for each commodity) which severely narrows down the methods we can use.

There are several ways to value an American option. All of them involve discretization of the state space, so at each point of time and each value of the factors, the holder of the option must decide whether to exercise his right or not and that decision can only be based on future dynamics. However, when the number of factors increases, one can not help finding the "course of dimensionality" (a concept due to Bellman) as full discretization is almost infeasible due to computational complexity for more than three factors.

In order to reduce this problem, three are the main approaches (Bally, Caramellino and Zanette, 2005). One idea is to perform state aggregation in order to develop a "synthetic indicator" in fewer dimensions, so that the holder of the option decides whether to exercise or not just depending on it and not on the three or more factors. This is the idea in Barraquand and Martineau (1995), which we shall follow because its simplicity. The other ones use either base functions (Longstaff and Schwartz, 2001 or Tsitsiklis and VanRoy, 1999) or compute conditional expectations via Malliavin Calculus, see Fournié, et al. (1999) for numerical applications and Bouchard and Touzi (2004) for general theory and variance reduction.

The basic idea in the Barraquand and Martineu (BM) method is to construct a one-dimensional indicator summarizing all states. The synthetic indicator will be the difference between the values of the refined product and the crude oil futures. We shall sum up the algorithm in the following scheme.

1. Divide the time interval [0,T] into k subintervals $0 = t_0 < t_1 < ... < t_k = N$ (uniformly spaced). The value of an American option will be approximated by a Bermuda option that can only be exercised at these intervals. Let $\Delta t = t_{i+1} - t_i$.

2. Start at t = 0 and simulate N trajectories of the process starting at X_0. For each instant of time, one can measure the difference in futures pricess, $D(X_{t_i})$, which is one-dimensional. Discretize this quantity in r intervals $\{P_1(t_i),...,P_r(t_i)\}$, so that each interval contains $\frac{1}{r}$ of probability.

3. Using simulation, compute the transition probabilities given by the formula:

$$\pi_{ij}(t) = P\left(X_{t+\Delta t} \in P_j(t) / X_t \in P_i(t)\right) \tag{10}$$

4. Compute a mean for each interval:

$$f_i(t) = E\left[D(X_t) / X_t \in P_i(t)\right] \tag{11}$$

5. Solve the problem by backwards iteration, i.e. for the interval r – 1, the value of not exercising is $e^{-r\Delta t}\sum_{j=1}^{r}\pi_{ij} f_j(t_r)$, so the value of the option at that interval is given by the maximum:

$$C(i,t_{r-1}) = \max\left(f_i(t_{r-1}), e^{-r\Delta t}\sum_{j=1}^{r}\pi_{ij} f_j(t_r)\right) \tag{12}$$

6. Repeat 5 until reaching :

$$X_0\text{, i.e. } C(i,t_{k-1}) = \max\left(f_i(t_{k-1}), e^{-r\Delta t}\sum_{j=1}^{r}\pi_{ij} C(j,t_k)\right) \tag{13}$$

Note that for each valuation, one has to compute X_0, which means that we must have an estimation of the state for each instant in time. This is done via an aliasing algorithm, which computes the expectation of the state X_t conditional on the information of the futures at time *t* (see Harvey, 1989, for details). This is the most reasonable choice, as there is no reason why a practitioner would not employ the (already available) futures price at time *t*, but future prices at times *t*+*h*, *t*+2*h*… can never be observed prior to time *t* in a real application.

All parameters and dynamics were estimated using weekly data. However, changing the model to daily dynamics is not really an issue in continuous time. We have just resampled the model (we did not estimate parameters again) and used an aliasing algorithm to get daily states. The RMSE is a bit higher, but the difference is small.

Finally, once we have an estimation of the option price, we must keep in mind the possibility of direct exercise, this is, buying the option and exercising it right away, thus obtaining the differences in futures prices minus the strike price. Should this quantity be higher than our estimated price, the real price would be the former, as we would exercise the option immediately.

4.3. Commodity Price Dynamics

In order to value an option, we need a full description of the model. In matrix form, the state dynamics can be described as:

$$dX_t = (\mu + AX_t)dt + dW_t \tag{14}$$

In order to clarify matters, let us take U_t to be a unit Brownian motion (i.e. $dU_t dU_t^T = I\,dt$) and rewrite (14) as:

$$dX_t = (\mu + AX_t)dt + RdU_t \tag{15}$$

If we just consider the models in equations (3) and (4) the only thing that is left to do is to estimate the states via an aliasing algorithm. In the case of two models put together, we have two equations in matrix form:

$$dX_{it} = (\mu_i + A_i X_{it})dt + R_i dW_{it} \quad (i = 1,2) \tag{16}$$

Therefore, if we assume uncorrelation the global model is:

$$\begin{pmatrix} dX_{1t} \\ dX_{2t} \end{pmatrix} = \left[\begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix} + \begin{pmatrix} A_1 & 0 \\ 0 & A_2 \end{pmatrix}\begin{pmatrix} X_{1t} \\ X_{2t} \end{pmatrix}\right]dt + \begin{pmatrix} R_1 & 0 \\ 0 & R_2 \end{pmatrix}\begin{pmatrix} dU_{1t} \\ dU_{2t} \end{pmatrix} \tag{17}$$

Whereas if we allow a free correlation structure the model is:

$$\begin{pmatrix} dX_{1t} \\ dX_{2t} \end{pmatrix} = \left[\begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix} + \begin{pmatrix} A_1 & 0 \\ 0 & A_2 \end{pmatrix}\begin{pmatrix} X_{1t} \\ X_{2t} \end{pmatrix}\right]dt + R^* \begin{pmatrix} dU_{1t} \\ dU_{2t} \end{pmatrix} \tag{18}$$

4.4. Results

Table 16 presents several metrics in order to analyze the in-sample predictive power ability of the models, for both types of options: the heating oil vs. WTI and the gasoline vs. WTI crack-spread options. The models considered are the (three-factor) joint model with common long-term trend, the (four-factor) joint model without common long-term trend, and the (two-factor) model for both commodities separately.

What we see is that we get better results with the four and three factor models. Even in some cases our model with common long-term trend gives better results than the model without common long-term trend. The difference between the joint models with and without common long-term trend is rather small when compared with the difference between each one of them and the uncorrelated model, which gives the worst results. This confirms our hypothesis of common trend.

Finally, we can conclude that, given that our joint model with common long-trend is simpler and easier to implement for the purposes of option valuation, and given that the valuation errors obtained with the joint models with and without common trend are quite similar and lower than those obtained with the uncorrelated model, the preferred model is the one with common long-term trend. This is further evidence of the convenience of using a common trend for both series, as soundness is always a prerequisite for accuracy.

5. Conclusion

In this chapter we have found evidence of a common long-term trend for crude oil (WTI) prices and the prices of the most important refined products (gasoline and heating oil), traded at NYMEX. This evidence was obtained from three independent sources, which makes our findings even more convincing.

First of all, studies on stationarity and Johansen tests show a cointegration relation. This fact would suggest common non-stationary dynamics but we have been able to go further.

Secondly, a principal component analysis shows that these three commodities are not only cointegrated, but they have also a common long-term dynamics, which can be obtained from one of these components (the first one).

Our next step was to propose a model that takes this fact into account and to compare it with independent models. In order to do so, we have proposed a joint model with common long-term trend for two or three commodities, within the framework of the factor models proposed by Schwartz (1997) and Schwartz and Smith (2000). The results indicate that our joint model with common long-term trend gives similar results in terms of goodness of fit and Schwartz and Akaike information criteria than a joint model without common long-term trend and a standard model for the commodities separately, suggesting that there is a common long-term trend.

These three models have also been used to value crack-spread options traded at NYMEX. Specifically, we have used Heating Oil vs. WTI and Gasoline vs. WTI crack-spread options traded at NYMEX. The results indicate that the valuation errors obtained with our common long-term model are quite similar to the results obtained with the model without common trend, and lower than those obtained with the uncorrelated model. Therefore, we can conclude that, given that our joint model with common long-trend is simpler and easier to implement for the purposes of option valuation, the preferred model is the one with common long-term trend.

Finally, we can conclude that crude oil and its main refined products have a common long-term trend and therefore a model taking into account this fact is the most useful. This model can be used not only for valuation of claims, but also for qualitative analysis.

APPENDIX

The Two-Factor Model for Each Commodity Separately

Transition equation:

$$Z_t = c_t + T_t Z_{t-1} + \psi_t \qquad t = 1, \ldots, N_t$$

where $Z_t = \begin{pmatrix} \xi_t \\ \chi_t \end{pmatrix}$, $c_t = \begin{pmatrix} \mu_\xi \Delta t \\ 0 \end{pmatrix}$, $T_t = \begin{pmatrix} 1 & 0 \\ 0 & e^{-k\Delta t} \end{pmatrix}$ and

$$Var(\psi_t) = \begin{pmatrix} \sigma_\xi^2 \Delta t & \sigma_\xi \sigma_\chi \rho_{\chi\xi}(1-e^{-k\Delta t})/k \\ \sigma_\xi \sigma_\chi \rho_{\chi\xi}(1-e^{-k\Delta t})/k & \sigma_\chi^2 (1-e^{-2k\Delta t})/(2k) \end{pmatrix}$$

Measurement equation:

$$Y_t = d_t + M_t Z_t + \eta_t \qquad t = 1, \ldots, N_t$$

where $Y_t = \begin{pmatrix} \ln F_{T1} \\ \vdots \\ \ln F_{Tn} \end{pmatrix}$, $d_t = \begin{pmatrix} A(T_1) \\ \vdots \\ A(T_n) \end{pmatrix}$, $M_t = \begin{pmatrix} 1 & e^{-kT_1} \\ \vdots & \vdots \\ 1 & e^{-kT_n} \end{pmatrix}$ and F_{T1} is the price of a futures contract on the commodity with maturity at time "T_1+t" traded at time t.

This model is the two factor model presented in Schwartz and Smith (2000).

The Joint Model With Common Long-Term Trend for Pairs of Commodities

Transition equation:

$$Z_t = c_t + T_t Z_{t-1} + \psi_t \qquad t = 1, \ldots, N_t$$

where $Z_t = \begin{pmatrix} \xi_t \\ \chi_{1t} \\ \chi_{2t} \end{pmatrix}$, $c_t = \begin{pmatrix} \mu_\xi \Delta t \\ 0 \\ 0 \end{pmatrix}$, $T_t = \begin{pmatrix} 1 & 0 & 0 \\ 0 & e^{-k_1 \Delta t} & 0 \\ 0 & 0 & e^{-k_2 \Delta t} \end{pmatrix}$ and

$$Var(\psi_t) = \begin{pmatrix} \sigma_\xi^2 \Delta t & \sigma_\xi \sigma_{\chi_1} \rho_{\xi\chi_1} (1 - e^{-k_1 \Delta t})/k_1 & \sigma_\xi \sigma_{\chi_2} \rho_{\xi\chi_2} (1 - e^{-k_2 \Delta t})/k_2 \\ \sigma_\xi \sigma_{\chi_1} \rho_{\xi\chi_1} (1 - e^{-k_1 \Delta t})/k_1 & \sigma_{\chi_1}^2 (1 - e^{-2k_1 \Delta t})/(2k_1) & \sigma_{\chi_1} \sigma_{\chi_2} \rho_{\chi_1\chi_2} (1 - e^{-(k_1 \Delta t + k_2 \Delta t)})/(k_1 + k_2) \\ \sigma_\xi \sigma_{\chi_2} \rho_{\xi\chi_2} (1 - e^{-k_2 \Delta t})/k_2 & \sigma_{\chi_1} \sigma_{\chi_2} \rho_{\chi_1\chi_2} (1 - e^{-(k_1 \Delta t + k_2 \Delta t)})/(k_1 + k_2) & \sigma_{\chi_2}^2 (1 - e^{-2k_2 \Delta t})/(2k_2) \end{pmatrix}$$

Measurement equation:

$$Y_t = d_t + M_t Z_t + \eta_t \qquad t = 1, \ldots, N_t$$

where $Y_t = \begin{pmatrix} \ln F_{T1}^1 \\ \vdots \\ \ln F_{Tn}^1 \\ \ln F_{T1}^2 \\ \vdots \\ \ln F_{Tn}^2 \end{pmatrix}$, $d_t = \begin{pmatrix} A^1(T_1) \\ \vdots \\ A^1(T_n) \\ A^2(T_1) \\ \vdots \\ A^2(T_n) \end{pmatrix}$, $M_t = \begin{pmatrix} 1 & e^{-k_1 T_1} & 0 \\ \vdots & \vdots & \vdots \\ 1 & e^{-k_1 T_n} & 0 \\ 1 & 0 & e^{-k_2 T_1} \\ \vdots & \vdots & \vdots \\ 1 & 0 & e^{-k_2 T_n} \end{pmatrix}$ and F_{T1}^i is the price of a futures contract on the commodity "*i*" (*i*=1,2) with maturity at time "$T_1 + t$" traded at time *t*. In principle, it would be possible to use a different number of futures contracts for each commodity, but in this work we consider more suitable to use the same number ("*n*") of futures contracts for both commodities.

The Joint Model Without Common Long-Term Trend for Pairs of Commodities

Transition equation:

$$Z_t = c_t + T_t Z_{t-1} + \psi_t \qquad t = 1, \ldots, N_t$$

where $Z_t = \begin{pmatrix} \xi_{1t} \\ \chi_{1t} \\ \xi_{2t} \\ \chi_{2t} \end{pmatrix}$, $c_t = \begin{pmatrix} \mu_{\xi 1}\Delta t \\ 0 \\ \mu_{\xi 2}\Delta t \\ 0 \end{pmatrix}$, $T_t = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & e^{-k_1\Delta t} & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & e^{-k_2\Delta t} \end{pmatrix}$ and

$$Var(\psi_t) = \begin{pmatrix} \sigma_{\xi_1}^2 \Delta t & \sigma_{\xi_1}\sigma_{\chi_1}\rho_{\xi_1\chi_1}(1-e^{-k_1\Delta t})/k_1 & \sigma_{\xi_1}\sigma_{\xi_2}\rho_{\xi_1\xi_2}\Delta t & \sigma_{\xi_1}\sigma_{\chi_2}\rho_{\xi_1\chi_2}(1-e^{-k_1\Delta t})/k_2 \\ \sigma_{\xi_1}\sigma_{\chi_1}\rho_{\xi_1\chi_1}(1-e^{-k_1\Delta t})/k_1 & \sigma_{\chi_1}^2(1-e^{-2k_1\Delta t})/(2k_1) & \sigma_{\chi_1}\sigma_{\xi_2}\rho_{\chi_1\xi_2}(1-e^{-k_1\Delta t})/k_1 & \sigma_{\chi_1}\sigma_{\chi_2}\rho_{\chi_1\chi_2}(1-e^{-(k_1\Delta t+k_2\Delta t)})/(k_1+k_2) \\ \sigma_{\xi_1}\sigma_{\xi_2}\rho_{\xi_1\xi_2}\Delta t & \sigma_{\chi_1}\sigma_{\xi_2}\rho_{\chi_1\xi_2}(1-e^{-k_1\Delta t})/k_1 & \sigma_{\xi_2}^2\Delta t & \sigma_{\xi_2}\sigma_{\chi_2}\rho_{\xi_2\chi_2}(1-e^{-k_2\Delta t})/k_2 \\ \sigma_{\xi_1}\sigma_{\chi_2}\rho_{\xi_1\chi_2}(1-e^{-k_1\Delta t})/k_2 & \sigma_{\chi_1}\sigma_{\chi_2}\rho_{\chi_1\chi_2}(1-e^{-(k_1\Delta t+k_2\Delta t)})/(k_1+k_2) & \sigma_{\xi_2}\sigma_{\chi_2}\rho_{\xi_2\chi_2}(1-e^{-k_2\Delta t})/k_2 & \sigma_{\chi_2}^2(1-e^{-2k_2\Delta t})/(2k_2) \end{pmatrix}$$

Measurement equation:

$$Y_t = d_t + M_t Z_t + \eta_t \qquad t = 1, \ldots, N_t$$

where $Y_t = \begin{pmatrix} \ln F_{T1}^1 \\ \vdots \\ \ln F_{Tn}^1 \\ \ln F_{T1}^2 \\ \vdots \\ \ln F_{Tn}^2 \end{pmatrix}$, $d_t = \begin{pmatrix} A^1(T_1) \\ \vdots \\ A^1(T_n) \\ A^2(T_1) \\ \vdots \\ A^2(T_n) \end{pmatrix}$, $M_t = \begin{pmatrix} 1 & e^{-k_1T_1} & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 1 & e^{-k_1T_n} & 0 & 0 \\ 0 & 0 & 1 & e^{-k_2T_1} \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 1 & e^{-k_2T_n} \end{pmatrix}$ and F_{T1}^{i} is the price of a futures contract on the commodity "*i*" (i=1,2) with maturity at time "T_1+t" traded at time *t*. As in the previous case, in principle, it would be possible to use a different number of futures contracts for each commodity, but in this work we consider more suitable to use the same number ("*n*") of futures contracts for both commodities.

<u>The Joint Model With Common Long-Term Trend for the Three Commodities</u>

Transition equation:

$$Z_t = c_t + T_t Z_{t-1} + \psi_t \qquad t = 1, \ldots, N_t$$

where $Z_t = \begin{pmatrix} \xi_{1t} \\ \chi_{1t} \\ \chi_{2t} \\ \chi_{3t} \end{pmatrix}$, $c_t = \begin{pmatrix} \mu_{\xi 1}\Delta t \\ 0 \\ 0 \\ 0 \end{pmatrix}$, $T_t = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & e^{-k_1\Delta t} & 0 & 0 \\ 0 & 0 & e^{-k_2\Delta t} & 0 \\ 0 & 0 & 0 & e^{-k_3\Delta t} \end{pmatrix}$ and

$$Var(\psi_t)=\begin{pmatrix} \sigma_{\xi_1}^2\Delta t & \sigma_{\xi_1}\sigma_{\chi_1}\rho_{\xi_1\chi_1}(1-e^{-k_1\Delta t})/k_1 & \sigma_{\xi_1}\sigma_{\xi_2}\rho_{\xi_1\xi_2}\Delta t & \sigma_{\xi_1}\sigma_{\chi_2}\rho_{\xi_1\chi_2}(1-e^{-k_1\Delta t})/k_2 \\ \sigma_{\xi_1}\sigma_{\chi_1}\rho_{\xi_1\chi_1}(1-e^{-k_1\Delta t})/k_1 & \sigma_{\chi_1}^2(1-e^{-2k_1\Delta t})/(2k_1) & \sigma_{\chi_1}\sigma_{\chi_2}\rho_{\chi_1\chi_2}(1-e^{-(k_1\Delta t+k_2\Delta t)})/(k_1+k_2) & \sigma_{\chi_1}\sigma_{\chi_3}\rho_{\chi_1\chi_3}(1-e^{-(k_1\Delta t+k_3\Delta t)})/(k_1+k_3) \\ \sigma_{\xi_1}\sigma_{\chi_2}\rho_{\xi_1\chi_2}(1-e^{-k_1\Delta t})/k_2 & \sigma_{\chi_1}\sigma_{\chi_2}\rho_{\chi_1\chi_2}(1-e^{-(k_1\Delta t+k_2\Delta t)})/(k_1+k_2) & \sigma_{\chi_2}^2(1-e^{-2k_2\Delta t})/(2k_2) & \sigma_{\chi_2}\sigma_{\chi_3}\rho_{\chi_2\chi_3}(1-e^{-(k_2\Delta t+k_3\Delta t)})/(k_2+k_3) \\ \sigma_{\xi_1}\sigma_{\chi_3}\rho_{\xi_1\chi_3}(1-e^{-k_3\Delta t})/k_3 & \sigma_{\chi_1}\sigma_{\chi_3}\rho_{\chi_1\chi_3}(1-e^{-(k_1\Delta t+k_3\Delta t)})/(k_1+k_3) & \sigma_{\chi_2}\sigma_{\chi_3}\rho_{\chi_2\chi_3}(1-e^{-(k_2\Delta t+k_3\Delta t)})/(k_2+k_3) & \sigma_{\chi_3}^2(1-e^{-2k_3\Delta t})/(2k_3) \end{pmatrix}$$

Measurement equation:

$$Y_t = d_t + M_t Z_t + \eta_t \qquad\qquad t = 1, \ldots, N_t$$

where $Y_t = \begin{pmatrix} \ln F_{T1}^1 \\ \vdots \\ \ln F_{Tn}^1 \\ \ln F_{T1}^2 \\ \vdots \\ \ln F_{Tn}^2 \\ \ln F_{T1}^3 \\ \vdots \\ \ln F_{Tn}^3 \end{pmatrix}$, $d_t = \begin{pmatrix} A^1(T_1) \\ \vdots \\ A^1(T_n) \\ A^2(T_1) \\ \vdots \\ A^2(T_n) \\ A^3(T_1) \\ \vdots \\ A^3(T_n) \end{pmatrix}$, $M_t = \begin{pmatrix} 1 & e^{-k_1T_1} & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 1 & e^{-k_1T_n} & 0 & 0 \\ 1 & 0 & e^{-k_2T_1} & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 1 & 0 & e^{-k_2T_n} & 0 \\ 1 & 0 & 0 & e^{-k_3T_1} \\ \vdots & \vdots & \vdots & \vdots \\ 1 & 0 & 0 & e^{-k_3T_n} \end{pmatrix}$ and F_{T1}^i is the price of a futures contract on the commodity "*i*" (*i*=1,2,3) with maturity at time "T_1+t" traded at time *t*. As in the previous cases, in principle, it would be possible to use a different number of futures contracts for each commodity, but in this work we consider more suitable to use the same number ("*n*") of futures contracts for all the commodities.

REFERENCES

- Asche F., Gjolberg O., Völker T. Price relationships in the petroleum market: an analysis of crude oil and refined product prices. Energy Economics 2003; 289-301.
- Bally V., Caramellino L., Zanette A. Pricing American options by Monte Carlo methods using a Malliavin Calculus approach. Monte Carlo Methods and Applications 2005;11; 97-133.

- Barraquand J., Martineau D. Numerical Valuation of High Dimensional Multivariate American Securities. The Journal of Financial and Quantitative Analysis 1995; 30 (3); 383-405.
- Boswijk, P.H. Testing for a Unit Root with Near-Integrated Volatility. Tinbergen Institute Discussion Paper 2001; 01-077/4 (2001).
- Boskijk, H.P., Doornik, J.a. Distribution Approximations for Cointegration Tests with Stationary Exogenous Regressors. Journal of Applied Econometrics 2005; 20; 797-810.
- Bouchard B., Touzi N. On the Malliavin approach to Monte Carlo approximation of conditional expectations. Finance and Stochastics 2004; 8(1); 45–71.
- Brennan, M.J, E.S. Schwartz. Evaluating natural resource investments. Journal of Business 1985; 58; 133-155.
- Cortazar, G., E.S. Schwartz. Implementing a stochastic model for oil futures prices. Energy Economics 2003; 25; 215-218.
- Cortazar G., Naranjo L. An N-Factor gaussian model of oil futures prices. Journal of Futures Markets 2006; 26 (3); 209-313.
- Clewlow L., Strickland C. Energy Derivatives, Pricing and Risk Management. Lamina Publications, London, 2000.
- Dickey D.A., W.A. Fuller. Distribution of the Estimators for Autoregressive Time Series with a Unit Root. Journal of the American Statistical Association 1979; 74; 427–431.
- Fournié, E., Lasry, J.M., Lebouchoux, J., Lions, P.L., Touzi, N. Applications of Malliavin calculus to Monte Carlo methods in Finance. Finance & Stochastics 1999; 3; 391-412.
- Gjolberg O., Johnsen T., 1999. Risk management in the oil industry: can information on long-run equilibrium prices be utilized? Energy Economics 1999; 21(6); 517-527.
- Harvey A.C. Forecasting, Structural Time Series Models and the Kalman Filter. Cambridge University Press, Cambridge, U.K, 1999.

- Hedge Funds Review. Commodities are hot with investors. http://www.hedgefundsreview.com, 2008.
- Johansen S. Statistical Analysis of Cointegrating Vectors. Journal of Economic Dynamics and Control 1988; 12,; 231-254.
- Lanza A., Manera M., Giovanninni M. Modeling and forecasting cointegrated relationships among heavy oil and product prices. Energy Economics 2005; 831-848.
- Laugthton D.G., Jacoby, H.D. Reversion, timing options, and long-term decision making, Financial Management 1993; 33; 225-240.
- Laugthton D.G., Jacoby H.D. The effects of reversion on commodity projects of different length. In L. Trigeorgis (Ed.), Real Options in Capital Investment: Models, Strategies and Applications. Praeger, Westport, CT; 1995; 185-205.
- Longstaff F., Schwartz E. Valuing American options by simulations a simple least squares approach. The Review of Financial Studies 2001;14 (1); 113-147.
- Oil Market Report, Annual Statistical Supplement. International Energy Agency. http://www.omrpublic.iea.org; 2006.
- Phillips, P.C.B., Perron, P. Testing for a unit root in time series regression, Biometrika 1988; 75(2); 335-346.
- Radchenko S. Lags in the response of gasoline prices to changes in crude oil prices: The role of short-term and long-term shocks. Energy Economics 2005a; 27(4); 573-602.
- Radchenko S. Oil price volatility and the asymmetric response of gasoline prices to oil price increases and decreases. Energy Economics 2005b; 27(5); 708-730.
- Said, S.E., Dickey, D.A. Testing for unit roots in autoregressive-moving average models of unknown order. Biometrika, 1984; 71(3); 599-607.
- Schwartz, E.S. The stochastic behaviour of commodity prices: Implication for valuation and hedging. The Journal of Finance 1997; 52; 923-973.
- Schwartz, E.S., Smith, J.E. Short-term variations and long-term dynamics in commodity prices. Management Science 2000; 893-911.

- Serletis A. A cointegration analysis of petroleum futures prices, Energy Economics 1994; 16 (2); 93-97.
- Tolmaski C., Hindanov D. Principal component analysis for correlated curves and seasonal commodities: The case of petroleum market. The Journal of Futures Markets 2002; 22(11); 1019–1035.
- Tsitsiklis, J. N., VanRoy, B. Optimal stopping for Markov processes: Hilbert space theory, approximation algorithm and an application to pricing high-dimensional financial derivatives. IEEE Trans. Autom. Controll 1999; 44; 1840-1851.

TABLE 1

DESCRIPTIVE STATISTICS

The table shows the mean and volatility of the four commodity series prices. F1 is the futures contract closest to maturity, F2 is the second contract closest to maturity and so on.

	WTI cruce oil		Heating Oil		Gasoline		Ref. Margin	
	Mean	Volatility	Mean	Volatility	Mean	Volatility	Mean	Stand.Dev.
Spot	-	-	-	-	-	-	4.02	4.99
F1	32.07	30%	37.10	31%	39.26	35%	-	-
F2	32.05	27%	37.16	28%	38.98	30%	-	-
F3	31.94	26%	37.12	26%	38.63	27%	-	-
F4	31.78	24%	37.01	25%	38.28	24%	-	-
F5	31.60	22%	36.85	23%	37.90	22%	-	-
F6	31.42	21%	36.69	22%	37.31	22%	-	-
F7	31.23	20%	36.52	21%	36.62	21%	-	-
F8	31.05	19%	36.34	19%	36.08	21%	-	-
F9	30.88	18%	36.16	18%	35.50	20%	-	-
F10	30.71	18%	35.99	18%	35.15	20%	-	-
F11	30.55	17%	35.82	17%	34.89	19%	-	-
F12	30.40	17%	35.67	16%	34.73	19%	-	-

TABLE 2

UNIT ROOT TESTS

The Table shows the results of the Augmented Dickey-Fuller (ADF), Phillips-Perron and Boswijk-Doornik unit root tests. The reported critical values for the ADF and Phillips-Perron tests are the MacKinnon critical values for rejection of the null hypothesis of a unit root. In the case of the Boswijk-Doornik test the reported critical values are asymptotic p-values obtained by the gamma approximation proposed by Boswijk and Doornik (2005).

WTI CRUDE OIL	ADF	PHILLIPS-PERRON	BOSWIJK-DOORNIK
Test statistic	-0.0062	0.1460	3.2198
1% Critical value	-3.4427	3.4427	12.5284
5% Critical value	-2.8669	-2.8669	9.1422
10% Critical Value	-2.5697	-2.5697	7.5999
HEATING OIL	ADF	PHILLIPS-PERRON	BOSWIJK-DOORNIK
Test statistic	0.2975	0.0117	1.5572
1% Critical value	-3.4427	-3.4427	12.5642
5% Critical value	-2.8669	-2.8669	9.0843
10% Critical Value	-2.5697	-2.5697	7.5063
GASOLINE	ADF	PHILLIPS-PERRON	BOSWIJK-DOORNIK
Test statistic	-1.1137	-1.1564	0.9192
1% Critical value	-3.4427	-3.4427	12.5406
5% Critical value	-2.8669	-2.8669	9.1284
10% Critical Value	-2.5697	-2.5697	7.5761
REFINING MARGIN	ADF	PHILLIPS-PERRON	BOSWIJK-DOORNIK
Test statistic	-3.9437	-4.5629	11.4066
1% Critical value	-3.4537	-3.4535	11.8653
5% Critical value	-2.8717	-2.8716	8.0008
10% Critical Value	-2.5723	-2.5722	6.3071

TABLE 3

JOHANSEN COINTEGRATION TEST FOR GASOLINE AND HEATING OIL

Hypothesized Number of Cointegration Equations	Likelihood Ratio	5 Percent Critical Value	1 Percent Critical Value
None **	38.67195	15.41	20.04
At most 1	0.041342	3.76	6.65
Normalized Cointegrating Coefficients: 1 Cointegrating Equation			
Gasoline	Heating Oil	Trend Coefficient	Log likelihood
1.00000	-0.913957	-5.41611	-1983.337

*(**) denotes rejection of the hypothesis at 5%(1%) significance level

TABLE 4

JOHANSEN COINTEGRATION TEST FOR GASOLINE AND CRUDE OIL

Hypothesized Number of Cointegration Equations	Likelihood Ratio	5 Percent Critical Value	1 Percent Critical Value
None **	41.69961	15.41	20.04
At most 1	0.311977	3.76	6.65
Normalized Cointegrating Coefficients: 1 Cointegrating Equation			
Gasoline	Crude Oil	Trend Coefficient	Log likelihood
1.00000	-1.134830	-2.919235	-1813.967

*(**) denotes rejection of the hypothesis at 5%(1%) significance level

TABLE 5

JOHANSEN COINTEGRATION TEST FOR HEATING OIL AND CRUDE OIL

Hypothesized Number of Cointegration Equations	Likelihood Ratio	5 Percent Critical Value	1 Percent Critical Value
None **	42.03124	15.41	20.04
At most 1	0.040317	3.76	6.65
Normalized Cointegrating Coefficients: 1 Cointegrating Equation			
Gasoline	Crude Oil	Trend Coefficient	Log likelihood
1.00000	-1.22888	2.321953	-1598.295

*(**) denotes rejection of the hypothesis at 5%(1%) significance level

TABLE 6

JOHANSEN COINTEGRATION TEST FOR GASOLINE, HEATING OIL AND CRUDE OIL

Hypothesized Number of Cointegration Equations	Likelihood Ratio	5 Percent Critical Value	1 Percent Critical Value
None **	87.88729	29.68	35.65
At most 1 **	37.97700	15.41	20.04
At most 2	0.215044	3.76	6.65

Normalized Cointegrating Coefficients: 1 Cointegrating Equation

Gasoline	Heating Oil	Crude Oil	Trend Coefficient	Log likelihood
1.000000	1.42074	-2.881141	0.39226	-2527.315

Normalized Cointegrating Coefficients: 2 Cointegrating Equations

Gasoline	Heating Oil	Crude Oil	Trend Coefficient	Log likelihood
1.000000		-1.130430	-3.060284	-2508.434
	1.000000	-1.232253	2.430102	

*(**) denotes rejection of the hypothesis at 5%(1%) significance level

TABLE 7

PRINCIPAL COMPONENT ANALYSIS

The Table shows the percentage of the volatility explained by each principal component.

Panel 1: Gasoline, Heating Oil and Crude Oil separately

	Gasoline	Heating Oil	Crude Oil
First Component	91,25%	95,05%	98,01%
First Two Components	94,55%	97,70%	99,74%
First Three Components	96,99%	99,19%	99,95%

Panel 2: Gasoline, Heating Oil and Crude Oil in pairs

	Gasoline and Heating Oil	Gasoline and Crude Oil	Heating Oil and Crude Oil
First Component	92,27%	92,87%	94,74%
First Two Components	94,21%	95,67%	97,11%
First Three Components	96,41%	97,08%	98,26%

TABLE 8

THE TWO FACTOR MODEL FOR EACH COMMODITY SEPARATELY

The table presents the results for the Schwartz and Smith (2000) two-factor model for each commodity separately. Standard errors in parentheses.

	WTI Crude Oil	WTI Crude Oil	Heating Oil	Heating Oil	Gasoline
Contracts	F1, F3, F5, F7 and F9	F1, F4, F7, F11, F15 and F18	F1, F3, F5, F7 and F9	F1, F4, F7, F11, F15 and F18	F1, F3, F5, F7 and F9
Period	06/30/1997 to 04/24/2006	09/09/1996 to 09/18/2006	06/30/1997 to 04/24/2006	09/09/1996 to 09/18/2006	06/30/1997 to 04/24/2006
Number obs.	461	522	461	522	461
μ_ξ	0.1536 (0.0383)	0.1488 (0.0293)	0.1579 (0.0494)	0.1471 (0.0284)	0.1321 (0.0429)
k	1.4996 (0.0302)	1.1304 (0.0124)	1.0678 (0.0829)	1.3624 (0.0402)	1.5505 (0.1166)
σ_ξ	0.1712 (0.0042)	0.1403 (0.0033)	0.2957 (0.0194)	0.1652 (0.0042)	0.2707 (0.0164)
σ_χ	0.3037 (0.0081)	0.3089 (0.0078)	0.4929 (0.0262)	0.3337 (0.0091)	0.5267 (0.0226)
$\rho_{\xi\chi}$	-0.0640 (0.0398)	-0.0666 (0.0351)	-0.7163 (0.0443)	-0.1974 (0.0408)	-0.6450 (0.0515)
μ_ξ’	-0.1198 (0.0031)	-0.0531 (0.0013)	-0.1330 (0.0149)	-0.0522 (0.0030)	-0.0624 (0.0141)
λ_χ	-0.0543 (0.0689)	0.0142 (0.0660)	-0.1077 (0.0939)	-0.0241 (0.0586)	0.0955 (0.0986)
σ_η	0.0066 (0.0001)	0.0089 (0.0001)	0.0231 (0.0003)	0.0289 (0.0003)	0.0329 (0.0004)
Log-likelihood	17861.6	23494.9	13660.2	17811.9	12350.4
AIC	17845.6	23478.9	13644.2	17795.9	12334.4
SIC	17812.6	23444.8	13611.1	17761.8	12301.3

TABLE 9

THE JOINT MODEL WITH COMMON LONG-TERM TREND FOR PAIRS OF COMMODITIES

The table presents the results for the Schwartz and Smith (2000) two-factor model assuming a common long-term trend for pairs of commodities. Standard errors in parentheses.

	WTI Crude Oil and Gasoline	WTI Crude Oil and Heating Oil	Heating Oil and Gasoline
Contracts	F1, F3, F5, F7 and F9	F1, F4, F7, F11, F15 and F18	F1, F3, F5, F7 and F9
Period	06/30/1997 to 04/24/2006	09/09/1996 to 09/18/2006	06/30/1997 to 04/24/2006
Number obs.	461	522	461
μ_ξ	0.2334 (0.0399)	0.1771 (0.0296)	0.1553 (0.0358)
k_1	1.1533 (0.0535)	1.1349 (0.0181)	2.1553 (0.0764)
k_2	1.7312 (0.0761)	1.3854 (0.0242)	2.2657 (0.0749)
σ_ξ	0.1962 (0.0066)	0.1433 (0.0034)	0.1819 (0.0050)
$\sigma_{\chi 1}$	0.3233 (0.0127)	0.2768 (0.0072)	0.3097 (0.0104)
$\sigma_{\chi 2}$	0.4308 (0.0132)	0.3182 (0.0080)	0.3637 (0.0110)
$\rho_{\xi\chi 1}$	-0.3035 (0.0580)	0.0043 (0.0368)	0.0726 (0.0474)
$\rho_{\xi\chi 2}$	-0.3792 (0.0506)	-0.0342 (0.0362)	-0.0026 (0.0453)
$\rho_{\chi 1\chi 2}$	0.8614 (0.0166)	0.8537 (0.0110)	0.6264 (0.0298)
μ_ξ’	-0.0857 (0.0089)	-0.0522 (0.0018)	-0.0857 (0.0078)
$\lambda_{\chi 1}$	0.2387 (0.0693)	0.1373 (0.0552)	0.0193 (0.0570)
$\lambda_{\chi 2}$	0.0055 (0.1080)	-0.0697 (0.0670)	-0.0861 (0.0593)
σ_η	0.0282 (0.0002)	0.0220 (0.0001)	0.0368 (0.0003)
Log-likelihood	26516.0	39276.7	24425.9
AIC	26490.0	39250.7	24399.9
SIC	26436.2	39195.4	24346.2

TABLE 10

THE JOINT MODEL WITHOUT COMMON LONG-TERM TREND FOR PAIRS OF COMMODITIES

The table presents the results for the Schwartz and Smith (2000) two-factor model for pairs of commodities without common long-term trend. Standard errors in parentheses.

	WTI Crude Oil and Gasoline	WTI Crude Oil and Heating Oil	Heating Oil and Gasoline
Contracts	F1, F3, F5, F7, F9	F1, F4, F7, F11, F15, F18	F1, F3, F5, F7, F9
Period	06/30/1997 to 04/24/2006	09/09/1996 to 09/18/2006	06/30/1997 to 04/24/2006
Number obs.	461	522	461
$\mu_{\xi 1}$	0.1519 (0.0354)	0.1474 (0.0272)	0.1602 (0.0489)
$\mu_{\xi 2}$	0.1305 (0.0449)	0.1476 (0.0292)	0.1306 (0.0428)
k_1	1.4839 (0.0980)	1.1543 (0.0292)	1.0777 (0.1033)
k_2	1.4855 (0.0846)	1.3473 (0.0292)	1.5721 (0.1025)
$\sigma_{\xi 1}$	0.1975 (0.0065)	0.1459 (0.0037)	0.3055 (0.0240)
$\sigma_{\xi 2}$	0.3193 (0.0136)	0.2992 (0.0083)	0.5043 (0.0319)
$\sigma_{\chi 1}$	0.2505 (0.0121)	0.1520 (0.0037)	0.2498 (0.0129)
$\sigma_{\chi 2}$	0.5021 (0.0186)	0.3180 (0.0081)	0.4947 (0.0198)
$\rho_{\xi 1 \xi 2}$	-0.2521 (0.0666)	-0.1000 (0.0422)	-0.7273 (0.0496)
$\rho_{\xi 1 \chi 1}$	0.3307 (0.0431)	0.8215 (0.0178)	-0.0828 (0.0597)
$\rho_{\xi 1 \chi 2}$	0.2101 (0.0436)	0.0050 (0.0372)	0.3624 (0.0502)
$\rho_{\xi 2 \chi 1}$	0.1010 (0.0477)	0.1281 (0.0368)	0.3513 (0.0523)
$\rho_{\xi 2 \chi 2}$	0.3279 (0.0365)	0.7310 (0.0207)	-0.0285 (0.0511)
$\rho_{\chi 1 \chi 2}$	-0.6186 (0.0440)	-0.0924 (0.0370)	-0.5888 (0.0512)
$\mu_{\xi 1}$’	-0.1235 (0.0105)	-0.0550 (0.0026)	-0.1330 (0.0177)
$\mu_{\xi 2}$’	-0.0589 (0.0107)	-0.0508 (0.0022)	-0.0625 (0.0118)
$\lambda_{\chi 1}$	-0.0586 (0.0593)	0.0126 (0.0575)	-0.1105 (0.0921)
$\lambda_{\chi 2}$	0.0979 (0.1007)	-0.0231 (0.0615)	0.0952 (0.0958)
σ_{η}	0.0235 (0.0002)	0.0212 (0.0001)	0.0277 (0.0002)
Log-likelihood	27789.2	39575.2	26429.1
AIC	27751.2	39537.2	26391.1
SIC	27672.7	39456.3	26313

TABLE 11

THE JOINT MODEL WITH COMMON LONG-TERM TREND FOR THE THREE COMMODITIES

The table presents the results for the Schwartz and Smith (2000) two-factor model assuming a common long-term trend for all three commodities. Standard errors in parentheses.

	WTI Crude Oil, Gasoline and Heating Oil
Contracts	F1, F3, F5, F7 and F9
Period	06/30/1997 to 04/24/2006
Number obs.	461
$\mu_{\xi 1}$	0.2062 (0.0432)
k_1	1.7671 (0.0470)
k_2	2.2432 (0.0629)
k_2	2.0721 (0.0554)
$\sigma_{\xi 1}$	0.1730 (0.0045)
$\sigma_{\xi 1}$	0.2642 (0.0086)
$\sigma_{\chi 1}$	0.3138 (0.0094)
$\sigma_{\chi 2}$	0.3398 (0.0099)
$\rho_{\xi 1 \xi 2}$	0.1504 (0.0433)
$\rho_{\xi 1 \chi 1}$	0.1186 (0.0415)
$\rho_{\xi 1 \chi 2}$	0.0558 (0.0415)
$\rho_{\xi 2 \chi 1}$	0.7164 (0.0227)
$\rho_{\xi 2 \chi 2}$	0.6639 (0.0248)
$\rho_{\chi 1 \chi 2}$	0.6564 (0.0255)
$\mu_{\xi 1}$’	-0.0935 (0.0058)
$\mu_{\xi 2}$’	0.1705 (0.0591)
$\lambda_{\chi 1}$	-0.1312 (0.0741)
$\lambda_{\chi 2}$	-0.2183 (0.0682)
σ_η	0.0308 (0.0002
Log-likelihood	39150.2
AIC	39112.2
SIC	39033.7

TABLE 12

IN-SAMPLE PREDICTIVE ABILITY

TWO-FACTOR MODEL FOR EACH COMMODITY SEPARATELY

The table presents the mean error (real minus predicted value) and the root mean squared error (RMSE) in order to analyze the in-sample predictive power ability of the Schwartz and Smith (2000) two-factor model for the three commodities separately. The time period is 06/30/1997 to 04/24/2006 (461 weekly observations for each commodity).

WTI			Heating Oil			Gasoline		
Contract	Mean	RMSE	Contract	Mean	RMSE	Contract	Mean	RMSE
F1	0.0005	0.0428	F1	0.0004	0.0469	F1	-0.0001	0.0528
F3	-0.0012	0.0369	F3	-0.0010	0.0413	F3	0.0004	0.0502
F5	0.0003	0.0319	F5	0.0001	0.0397	F5	-0.0003	0.0449
F7	0.0009	0.0283	F7	0.0010	0.0328	F7	0.0000	0.0324
F9	-0.0006	0.0267	F9	-0.0006	0.0392	F9	0.0001	0.0503

TABLE 13

IN-SAMPLE PREDICTIVE ABILITY

TWO-FACTOR MODEL FOR PAIRS OF COMMODITIES WITH COMMON LONG-TERM TREND

The table presents the mean error (real minus predicted value) and the root mean squared error (RMSE) in order to analyze the in-sample predictive power ability of the Schwartz and Smith (2000) two-factor model for pairs of commodities assuming a common long-term trend. The time period is 06/30/1997 to 04/24/2006 (461 weekly observations for each commodity) when using the contracts F1, F3, F5, F7 and F9 and 09/09/1996 to 09/18/2006 (522 weekly observations for each commodity) when using the contracts F1, F4, F7, F11, F15 and F18.

WTI AND HEATING OIL			WTI AND GASOLINE			HEAT. OIL AND GASOLINE		
WTIL CRUDE OIL			WTI CRUDE OIL			HEATING OIL		
Contract	Mean	RMSE	Contract	Mean	RMSE	Contract	Mean	RMSE
F1	0.0066	0.0452	F1	0.0104	0.0502	F1	0.0034	0.0513
F4	-0.0002	0.0344	F3	0.0048	0.0381	F3	0.0004	0.0435
F7	0.0022	0.0276	F5	0.0047	0.0340	F5	0.0007	0.0459
F11	0.0033	0.0255	F7	0.0054	0.0349	F7	0.0017	0.0500
F15	0.0026	0.0239	F9	0.0051	0.0377	F9	0.0011	0.0509
F18	0.0003	0.0238	-	-	-	-	-	-
HEATING OIL			GASOLINE			GASOLINE		
Contract	Mean	RMSE	Contract	Mean	RMSE	Contract	Mean	RMSE
F1	-0.0003	0.0460	F1	-0.0017	0.0542	F1	-0.0042	0.0564
F4	-0.0043	0.0441	F3	0.0008	0.0495	F3	-0.0015	0.0516
F7	0.0003	0.0431	F5	0.0008	0.0464	F5	-0.0010	0.0517
F11	0.0016	0.0378	F7	0.0007	0.0374	F7	-0.0005	0.0453
F15	0.0015	0.0310	F9	-0.0008	0.0460	F9	-0.0013	0.0435
F18	-0.0020	0.0298	-	-	-	-	-	-

TABLE 14

IN-SAMPLE PREDICTIVE ABILITY

TWO-FACTOR MODEL FOR PAIRS OF COMMODITIES WITHOUT COMMON LONG-TERM TREND

The table presents the mean error (real minus predicted value) and the root mean squared error (RMSE) in order to analyze the in-sample predictive power ability of the Schwartz and Smith (2000) two-factor model for pairs of commodities without common long-term trend. The time period is 06/30/1997 to 04/24/2006 (461 weekly observations for each commodity) when using the contracts F1, F3, F5, F7 and F9 and 09/09/1996 to 09/18/2006 (522 weekly observations for each commodity) when using the contracts F1, F4, F7, F11, F15 and F18.

WTI AND HEATING OIL			WTI AND GASOLINE			HEAT. OIL AND GASOLINE		
WTIL CRUDE OIL			WTI CRUDE OIL			HEATING OIL		
Contract	Mean	RMSE	Contract	Mean	RMSE	Contract	Mean	RMSE
F1	0.0018	0.0440	F1	0.0005	0.0443	F1	0.0004	0.0474
F4	-0.0034	0.0345	F3	-0.0012	0.0375	F3	-0.0010	0.0417
F7	0.0002	0.0272	F5	0.0003	0.0324	F5	0.0001	0.0396
F11	0.0017	0.0245	F7	0.0009	0.0291	F7	0.0010	0.0327
F15	0.0011	0.0224	F9	-0.0006	0.0277	F9	-0.0005	0.0397
F18	-0.0014	0.0225	-	-	-	-	-	-
HEATING OIL			GASOLINE			GASOLINE		
Contract	Mean	RMSE	Contract	Mean	RMSE	Contract	Mean	RMSE
F1	0.0018	0.0455	F1	-0.0001	0.0525	F1	-0.0001	0.0528
F4	-0.0035	0.0440	F3	0.0003	0.0485	F3	0.0003	0.0492
F7	0.0004	0.0434	F5	-0.0003	0.0443	F5	-0.0003	0.0442
F11	0.0015	0.0393	F7	0.0000	0.0314	F7	0.0000	0.0313
F15	0.0016	0.0302	F9	-0.0000	0.0489	F9	-0.0000	0.0499
F18	-0.0017	0.0280	-	-	-	-	-	-

TABLE 15

DESCRIPTIVE STATISTICS FOR CRACK SPREAD OPTIONS

The Table shows the main descriptive stats of the crack-spread options. The time period is January 2004 to January 2007. For gasoline vs. WTI crack-spread options there are four maturating dates: March, April, August 2006 and January 2007, with twelve exercise prices, from 5 to 16 dollars. For heating oil vs. WTI crack spread options there are data corresponding to contracts maturating from January 2004 to January 2007, with only two exercises prices available: 5 and 8 dollars.

PANEL A: HEATING OIL VS WTI (35 SERIES, 1735 OBSERVATIONS)

	% sample	Mean number of observations/series	Mean K ($)	Mean price ($)
Put options	62.86	48.5	7.45	0.6
Call options	37.14	51.38	7.31	1.67

PANEL B: GASOLINE VS WTI (18 SERIES, 820 OBSERVATIONS)

	% sample	Mean number of observations/series	Mean K ($)	Mean price ($)
Put options	22.22	57.75	10.00	2.31
Call options	77.78	42.07	10.86	2.73

TABLE 16

CRACK-SPREAD OPTION VALUATION RESULTS

ERROR DESCRIPTIVE STATISTICS

The table presents several metrics in order to analyze the in-sample predictive power ability of the models under study: the joint model with common trend, the joint model without common trend and the model for commodities separately. The time period is 2004-2007 for gasoline vs WTI crack spread options and 2006-2007 for heating oil vs WTI crack spread options (daily observations).

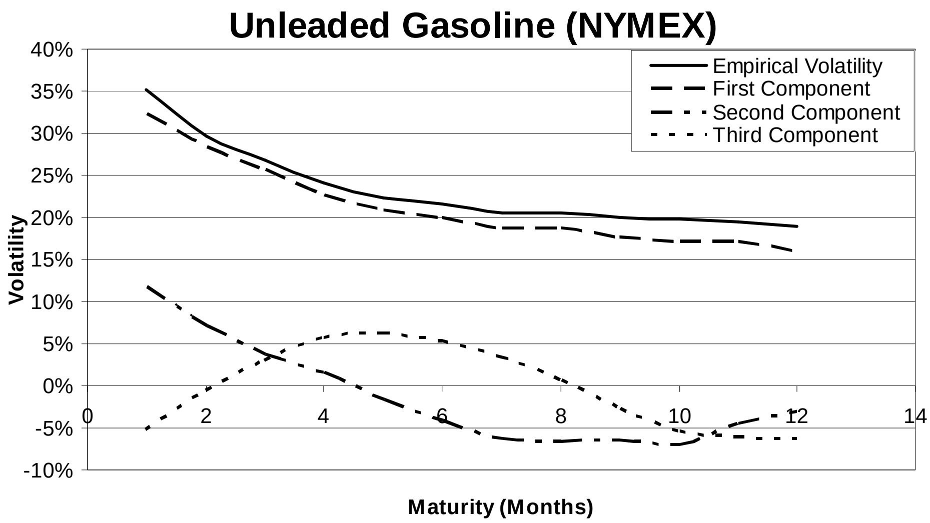
PANEL A: HEATING OIL VS WTI CRACK SPREAD OPTIONS

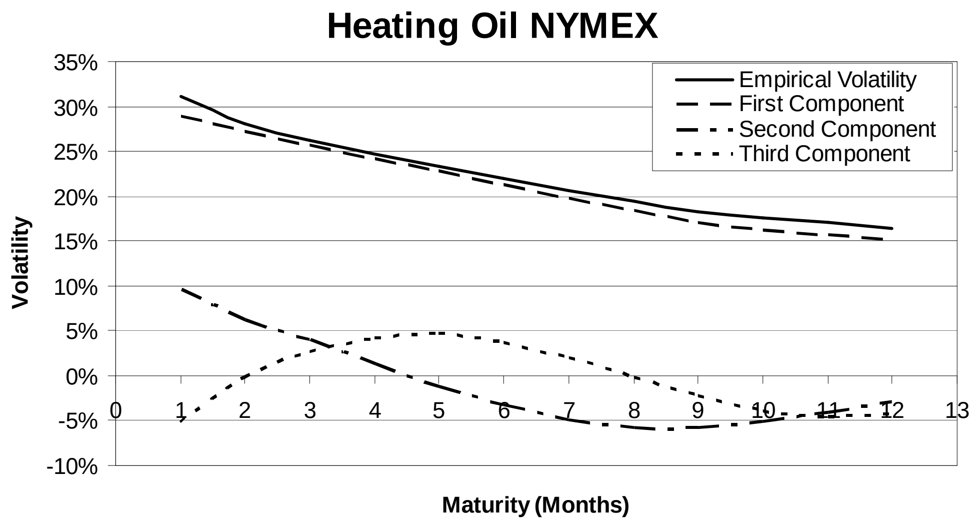
STATISTIC	3 Factor Model (Joint Model with Common Trend)	4 Factor Model (Joint Model without Common Trend)	2 Factor Model (Uncorrelated Model)
Mean Bias (real – predicted)	0.7925	-0.1670	1.7083
RMSE	0.4347	1.9728	2.0784
Bias Standard Deviation	1.8066	0.4013	1.1837
Median Bias	0.0036	0.1640	1.4912
Root Median Squared Error	0.3937	0.2619	1.4912

PANEL B: GASOLINE VS WTI CRACK SPREAD OPTIONS

STATISTIC	3 Factor Model (Joint Model with Common Trend)	4 Factor Model (Joint Model without Common Trend)	2 Factor Model (Uncorrelated Model)
Mean Bias (real – predicted)	-0.5950	-0.6609	0.9696
RMSE	1.0751	0.9381	1.3856
Bias Standard Deviation	0.8954	0.6657	0.9899
Median Bias	-0.4704	-0.5250	0.8522
Root Median Squared Error	0.5580	0.5565	0.8607

FIGURE 1. PRINCIPAL COMPONENT ANALYSIS FOR GASOLINE, HEATING OIL AND CRUDE OIL SEPARATELY





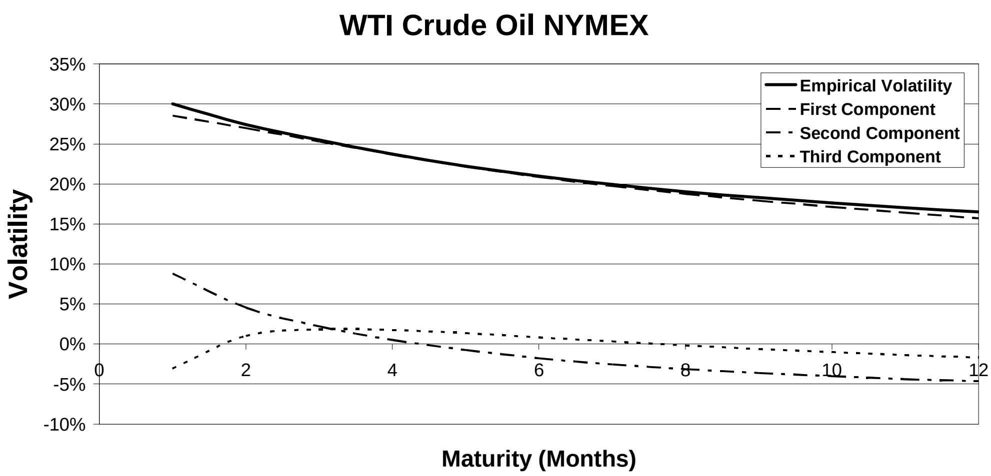
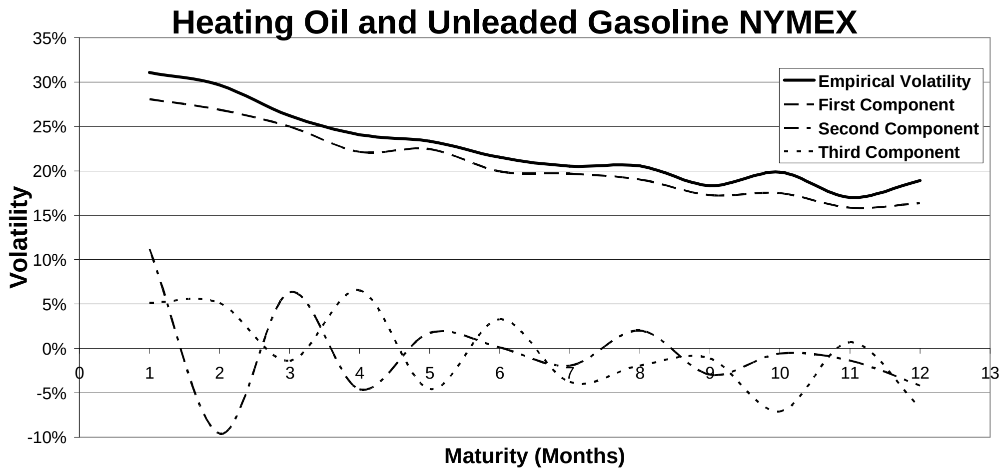
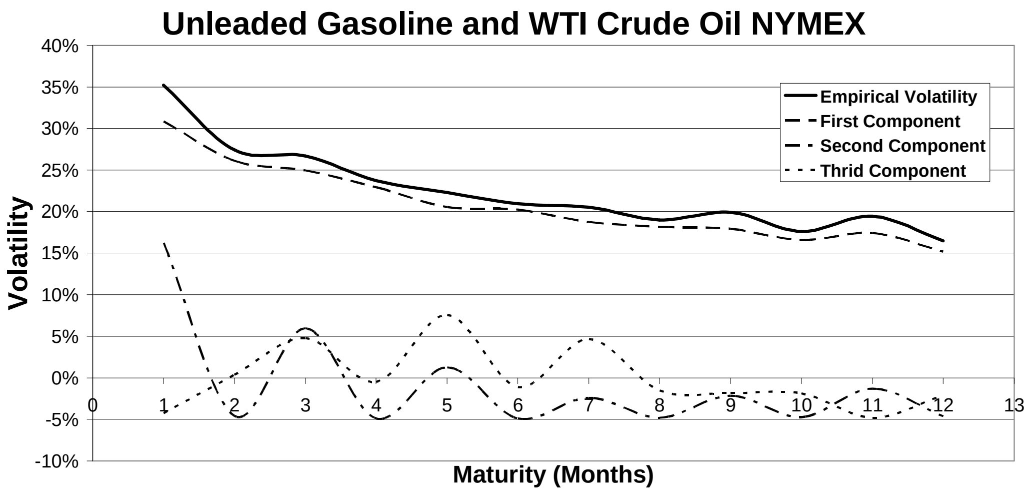


FIGURE 2. PRINCIPAL COMPONENT ANALYSIS FOR GASOLINE, HEATING OIL AND CRUDE OIL IN PAIRS





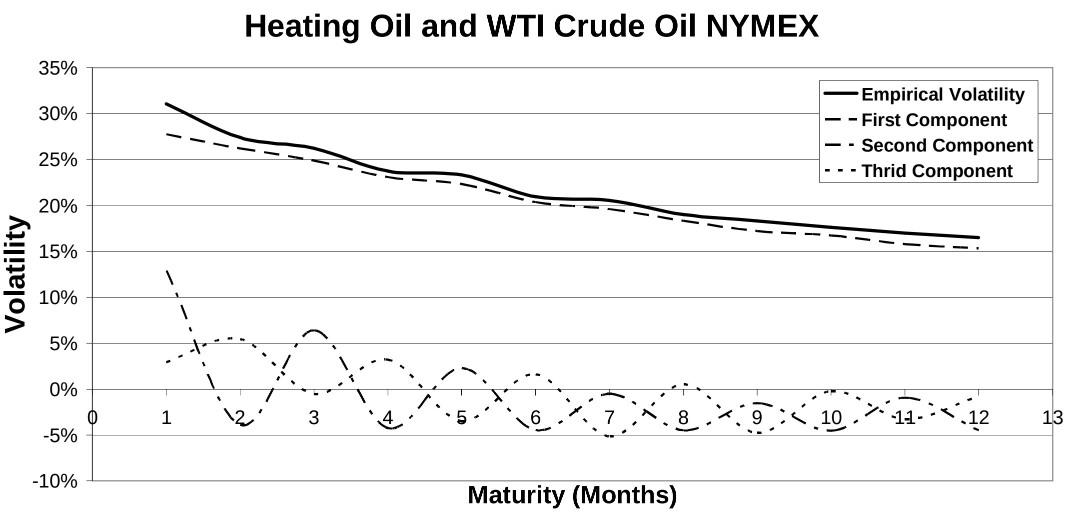


FIGURE 3. PRINCIPAL COMPONENT ANALYSIS FOR GASOLINE, HEATING OIL AND CRUDE OIL JOINTLY