University of Piraeus
Department of Banking and Finance

MSc in Banking and Finance

PRICING OF COMMODITIES FUTURES: AN EMPIRICAL STUDY

Student Name

DIMITRIS TASSIS

Supervisor:

GEORGE SKIADOPOULOS

Committee:

A. Antzoulatos G. Skiadopoulos N. Apergis

JANUARY 2013

Abstract

In the current thesis, methods and models for pricing commodities contracts are presented and studied. More specifically, a model which was created by Gibson and Schwartz (1990) is studied in depth, analysed and implemented using Matlab. Moreover, results of implementation are presented and analysed. Implementation is based on a large data set of more than 20 years futures prices of crude oil commodities.

Keywords

Commodities, Convenience Yield, Term Structure, Market Price of Risk, Futures.

Acknowledgements

I would like to express my greatest gratitude to the people who have helped and supported me throughout my thesis. I am grateful to my professor George Skiadopoulo for his continuous support for the project.

I want to thank my parents for their undivided support and my friends Fotis, Spiros, George and Pantelis. Last but not least I want to thank Vanessa who appreciated me for my work and motivated me to make all the things possible.

List of Figures

List of Tables

Table of Contents

1. Introduction

In the current thesis, methods and models for pricing commodities contracts are presented and studied. More specifically, in the initial part, a literature review is conducted with emphasis on commodities pricing models and then a model which was created by Gibson and Schwartz (1990) is studied in depth, analysed and implemented. Moreover, results of implementation are presented and analysed. Implementation is based on a large data set of more than 20 years futures prices of crude oil commodities. In the next sections of the Chapter the subject and objectives as well as the structure of the Thesis are presented in more detail.

1.1. Thesis Subject and Objectives

Main subject of this Thesis is the study and implementation of methods for pricing commodities futures. Strong emphasis is given on two dimensional models (based on crude oil spot price and convenience yield) and a representative implementation of one of them has been carried out, tested and evaluated. The main objectives of the Thesis can be summarized as follows:

1. Present a comprehensive literature review of existing models for pricing commodities futures.
2. Present the theoretical foundation of a well known two-factor model (Gibson and Schwartz, 1990) based on spot prices and convenience yield. Also consider and study practical ways to implement the model.
3. Implement the presented model using Matlab software platform and conduct a detailed qualitative and quantitative analysis on model parameters.
4. Evaluate in depth and discuss thoroughly on performance of the implemented model.

1.2. Thesis Structure

In the current Chapter we state the main subject and objectives of the Thesis as well as the structure of its content. In Chapter 2 an introduction to the market traded commodities and futures as well as a comprehensive literature review on the subject of the Thesis are presented. In Chapter 3 we present and explain in detail the definition and rational of a two-factor model created by Gibson and Schwartz (1990). In Chapter 4 our implementation approach of the model is explained and analyzed. Moreover, the model has been evaluated and tested on a large data set and a detailed analysis on its performance is discussed. Finally, Chapter 5 contains a synopsis of the study together with the main conclusions derived.

2. Formulation and Rational of Commodity Futures Pricing Models

2.1. Introduction to Market Traded Commodities and Futures

In recent years the commodities markets have changed a lot in the view of the trading volume, the range of the commodities and the variety of the derivative instruments. Commodity contingent claims have become increasingly popular and the over-the-counter market for them has grown remarkably. As a result, the commodity price of risk has received more attention. Therefore, the modeling of price processes and the development of new contingent claims pricing approaches are important area of research today.

Commodities have characteristics that make them differ from financial assets. The main difference is that any transaction in commodities may be physical (delivery of the commodity) or financial (no exchange of the underlying good but only a cash flow from one party to the other). Also, the returns of the commodities assets have usually negative correlation with the returns of financial assets resulting in diversification benefits. Commodity spot and futures prices are mean reverting. In equilibrium, supply increases if the price of the commodity is high and as a consequence higher cost producers enter the market. Correspondingly, supply decreases if the price is low as some of the higher cost producers exit the market. Moreover, seasonal effects introduced by the nature of the production process are typical for many commodities resulting in volatility in price levels.

For the purpose of understanding futures prices, it is convenient to divide commodity futures contracts into the following two categories according to the underlying asset: investment and consumption commodities. Investment commodities, e.g. gold and silver, are held for investment purposes by a significant number of investors. Consumption commodities, e.g. oil, are held primarily for consumption purposes. In the case of consumption commodities, it is not possible to obtain the futures price as a function of the spot price and other observable variables. Hence, a parameter known as convenience yield becomes important and we will it thoroughly discuss it in the next sections of the current thesis.

The major world commodities markets are the Chicago Board of Trade (CBOT), the Chicago Merchantile Exchange (CME) and the New York Merchantile Exchange (NYMEX) in the U.S, the London Metal Exchange (LME) in Europe and finally the India Commodity and Derivatives Exchange, the Tokyo Commodity Exchange (TOCOM), the Singapore Merchantile Exchange (SMX) and the Shangai Futures Exchange in Asia.

2.2. Commodity Futures Pricing: Literature Review

This thesis reviews the literature on the term structure models of commodity prices. The term structure is defined as the relationship between the spot price and the futures price for any delivery date. Term structure models of commodity prices aim to reproduce the futures prices observed in the market as accurately as possible. They provide useful information for hedging or investment decision, because they synthesize the information available in the market and the operators' expectations concerning the future. This information is very useful for management purposes: it can be used to hedge exposures in the physical market, to adjust the stock level or the production rate. It can also be used to undertake arbitrage transactions, to evaluate derivatives instruments based on futures contracts etc.

The methodology that is applied in the term structure models of commodity prices is the above. Firstly, the state variables are selected and their dynamics are specified. The most common state variables that we are taking into account are the spot price, the convenience yield, the long-term price and the interest rate. Afterwards, applying Ito's Lemma in the price of a futures contract, which is a function of the state variables and the time, we are obtaining the dynamic behavior of the futures price. Finally, an arbitrage reasoning and a construction of a hedging portfolio lead to the valuation equation that characterizes the model and the solution of the model is obtained, whenever it is possible.

Term structure models are based in two fundamental theories. The first one is the theory of storage developed by Kaldor (1939), Working (1949) and Brennan (1958). According to that theory the difference between contemporaneous spot and futures prices (basis) is explained on the basis of interest foregone in storing commodity product, warehousing costs and a convenience yield on inventory. Also, there is a negative relation between the convenience yield and the level of inventories. The second one is the theory of normal backwardation (the futures price is below the spot price) which was first developed by Keynes (1930). This theory views commodity futures price as a combination of an expected risk premium and a forecast of the future spot price. It has been argued that if hedgers tend to hold short positions and speculators tend to hold long positions, the futures price will be below the expected spot price. This is because speculators require compensation for the risks they are bearing.

Fama and French (1987) examine the theory of storage for valuing commodity futures. They calculate the standard deviation for various commodities. It is observed that for commodity products which present strong seasonal variation in supply and demand (agricultural products) and commodities which are conveyed to high storage costs (animal products) present high standard deviation of the basis, a fact that is aligned to the storage theory.

Schwartz (1997) tests the performance of three mean reverting models for commodity pricing. The Kalman filter methodology is used to estimate the parameters of the three models for two commercial commodities, copper and oil, and one precious metal, gold.

The first model, which is a one factor model, assumes that the logarithm of the spot price of the commodity follows a mean reverting process of the Ornstein-Uhlenbeck type. The use of a mean reversion process for the spot price allows taking into account the behavior of operators. When the spot price is lower than its long run mean, the industrials, expecting a rise in the spot price, reconstitute their inventories, whereas the producers reduce their production rate. The increasing demand and the simultaneous reduction of supply have a rising influence on the spot price. Conversely, when the spot price is higher than its long run mean, industrials try to reduce their surplus inventories and producers increase their production rate, pushing the spot price to lower levels.

Firstly, assuming that the commodity spot price follows the stochastic process:

$$dS/S = k(\mu - lnS)dt + \sigma dz \qquad (2.1)$$

Defining X = lnS and applying Ito's Lemma:

$$dX = k(a - X)dt + \sigma dz \qquad (2.2)$$

$$\alpha = \mu - \frac{\sigma^2}{2\kappa} \qquad (2.3)$$

Under the equivalent martingale measure the above equation can be written as:

$$dX = k(a^* - X)dt + \sigma dz^* \qquad (2.4)$$

$$a^* = a - \lambda \qquad (2.5)$$

An arbitrage reasoning and the construction of a hedging portfolio lead to the valuation equation of the futures prices, which is:

$$1/2\, \sigma^2 S^2 F_{ss} + k(\mu - \lambda - lnS)SF_s - F_T = 0 \qquad (2.6)$$

Where λ is the market price of risk and the terminal boundary condition associated with the equation is:

$$F(S, 0) = S \qquad (2.7)$$

The solution to the above differential equation is:

$$F(S,T) = exp\left[e^{-kT} lnS + (1 - e^{-kT})a^* + \frac{\sigma^2}{4\kappa}(1 - e^{-2kT})\right] \qquad (2.8)$$

The second model, which is a two factor model, is almost the same with the one that we will describe in the next sections of this thesis (Gibson and Schwartz (1990). It assumes that the spot price S and the instantaneous convenience yield follow a joint stochastic process with constant correlation. The main difference is that the convenience yield is brought into the spot price as a dividend yield. The dynamics of the above state variables are:

$$dS/S = (\mu - \delta)dt + \sigma_1 dz_1 \qquad (2.9)$$

$$d\delta = k(a - \delta)dt + \sigma_2 dz_2 \qquad (2.10)$$

Also, dz_1 and dz_2 are correlated increments to standard Brownian processes where $dz_1 dz_2 = \rho dt$

The stochastic process for the factors under the equivalent martingale measure can be expressed as:

$$dS = (r - \delta)Sdt + \sigma_1 S dz_1^* \tag{2.11}$$

$$d\delta = [k(a - \delta) - \lambda]dt + \sigma_2 dz_2^* \tag{2.12}$$

$$dz_1^* dz_2^* = \rho dt \tag{2.13}$$

The third model, which is a three factor model, extends the two factor model by adding a third state variable: the instantaneous interest rate. By assuming mean reverting process for the interest rate instead of being constant, it is possible to obtain a closed form solution for futures prices. Nevertheless the inclusion of stochastic interest rates in the commodity price models does not have a significant impact in the pricing of commodity futures. The addition of the stochastic interest rate as a third stochastic factor was also followed by Miltersen and Schwartz (1998) and Hilliard and Reis (1998). The dynamics of the above state variables are:

$$dS = (r - \delta)Sdt + \sigma_1 dz_1^* \tag{2.14}$$

$$d\delta = \kappa(\hat{\alpha} - \delta)dt + \sigma_2 dz_2^* \tag{2.15}$$

$$dr = a(m^* - r)dt + \sigma_3 dz_3^* \tag{2.16}$$

$$\mathrm{dz}_1^* \mathrm{dz}_2^* = \rho_1 \mathrm{dt} \tag{2.17}$$

$$\mathrm{dz}_2^* \mathrm{dz}_3^* = \rho_2 \mathrm{dt} \tag{2.18}$$

$$\mathrm{dz}_1^* \mathrm{dz}_3^* = \rho_3 \mathrm{dt} \tag{2.19}$$

Where α and m^* are the speed of adjustment coefficient and the risk adjusted mean short rate of the interest rate process.
Futures prices must then satisfy the partial differential equation:

$$\tfrac{1}{2}\sigma_1^2 S^2 F_{ss} + \tfrac{1}{2}\sigma_2^2 F_{\delta\delta} + \tfrac{1}{2}\sigma_3^2 F_{rr} + \sigma_1\sigma_2\rho S F_{S\delta} + \sigma_2\sigma_3\rho_2 F_{\delta r} + \sigma_1\sigma_3\rho_3 S F_{Sr} + (r - \delta)SF_S + \kappa(\hat{\alpha} - \delta)F_\delta + \alpha(m^* - r)F_r - F_T = 0 \tag{2.20}$$

Subject to the terminal boundary condition:

$$F(S, \delta, r, T) = S \tag{2.21}$$

The solution to the above differential equation is:

$$F(S,\delta,r,T) = Sexp\left[-\delta\frac{1-e^{-\kappa T}}{\kappa} + \frac{r\left(1-e^{-aT}\right)}{a} + C(T)\right] \qquad (2.22)$$

Where

$$C(T) = \frac{(\kappa\hat{\alpha} + \sigma_1\sigma_2\rho_1)\left((1-e^{-\kappa T}) - \kappa T\right)}{\kappa^2} - \frac{\sigma_2^2(4(1-e^{-\kappa T}) - (1-e^{-2\kappa T}) - 2\kappa T)}{4\kappa^3}$$
$$- \frac{(\alpha m^* + \sigma_1\sigma_3\rho_3)\left((1-e^{-\alpha T}) - \alpha T\right)}{\alpha^2}$$
$$- \frac{\sigma_3^2(4(1-e^{-\alpha T}) - (1-e^{-2\alpha T}) - 2\alpha T)}{4\alpha^3}$$
$$+ \sigma_2\sigma_3\rho_2\left(\frac{(1-e^{-\kappa T}) + (1-e^{-\alpha T}) - \left(1-e^{-(\kappa+\alpha)T}\right)}{\kappa\alpha(\kappa+\alpha)}\right.$$
$$\left.+ \frac{\kappa^2(1-e^{-\alpha T}) + \alpha^2(1-e^{-\kappa T}) - \kappa\alpha^2 T - \alpha\kappa^2 T}{\kappa^2\alpha^2(\kappa+\alpha)}\right)$$

(2.23)

The second and the third model perform better than the first one both for the short and the long term futures contracts pricing. The second and the third model are of equivalent performance with a slice difference for the third model that fits the data slightly better for the long term futures contracts.

Ribeiro and Hodges (2004) introduce a new reduced form two-factor model for commodity futures prices and futures valuation that extends Schwartz's (1997) two factor model by adding two new features. First they replace the Ornstein-Uhlenbeck process by a Cox-Ingersoll-Ross (CIR) process that ensures the non-negativity of the convenience yield a fact that rules out arbitrage possibilities. Second, instead of a constant volatility they consider that the spot price volatility is proportional to the square root of the convenience yield level. This assumption implies that the spot price volatility depends on inventory levels of the commodity and it is aligned to the storage theory. This model performs slightly better from the Schwartz's two factor model. The dynamics of the above state variables are:

$$dS/S = (\mu - \delta)dt + \sigma_1\sqrt{\delta}dz_1 \qquad (2.24)$$

$$d\delta = \alpha(\mu - \delta)dt + \sigma_2\sqrt{\delta}dz_2 \qquad (2.25)$$

A second application of term structure models of commodity prices is the investment decision. The use of term structure models in the case of investment decision is rather intuitive. With such a model, it is possible to estimate a futures price for any expiration date. Thus, such a model enables the valuation of net cash flows associated with an investment project.

Brennan and Schwartz (1985) develop a model for evaluating natural resources investment. The aim of the paper is to value the uncertain cash flow stream generated by an investment project using a self-financing portfolio, whose cash flows replicate those which are to be valued. The construction of the portfolio rests on the assumption that the convenience yield on the output commodity can be written as a function of the output price alone and that the interest rate is non-stochastic. More precisely, the paper considers a mine producing a commodity whose output price can be modeled by a geometric Brownian motion and it determines the optimal behavior to run the mine (at what rate to produce the output commodity) or to close the mine ready to re-open it later, or even to abandon it.

Cortazar and Schwartz (1997) use a one-factor model based on mean reverting spot price, in which the convenience yield is variable and depends on the deviation of the spot price to a long-term average price. Using this model, they estimate the value of the field at different stages: before the development, during the development and during the production.

Schwartz and Smith (2000) apply their short-term / long-term model to some hypothetical real options problem. They consider two real options: the option to defer investment for long-term investment and the development option for a short-term project. They show that in the short-term project, the values and policies are sensitive to both state variables and the value increases with both the short-deviations and the equilibrium price. However, the value and policies of the long-term project are insensitive to the short-term deviations.

Finally, Cortazar,Schwartz and Cassassus (2001) collapse price and geological-technical uncertainty into a new factor model. Using this model, they determine the value of several real options such as investment schedules for all exploration stages and timing options for the development decision.

3. Definition and Implementation Approach of a Two-Factor Pricing Model

In this Chapter, we present and analyze the two-factor model that was established by Gibson and Schwartz (1990). This model is the basis for what we have implemented, which is presented in the next Chapter.

3.1 Introduction

A fundamental assumption of every two-factor model is that, there is not one single source of uncertainty which has impact on the commodity (crude oil in our case) value, but also a second one. This assumption leads to a two-factor (two-dimensional) futures pricing model, which by definition should be more accurate, but at the same time more complex, compared to one factor models.

In this section we introduce and review a two factor model which was established by Gibson and Schwartz (1990) to price futures on crude oil. The two main factors of uncertainty in the proposed model are (1) the spot price of the crude oil and (2) the stochastic convenience yield. Convenience yield is viewed as a net dividend yield that accrues to the physical owner of the commodity but not to the owner of a contract for future delivery. Gibson and Schwartz built their model on the assumption that for crude oil, convenience yield requires a stochastic representation.

Results that came out from the implementation of this model have proven that it performs well, especially in valuing short term crude oil future contracts. Moreover, the proposed model can justify and explain the "intrinsic" difference of spot and futures contracts in volatility of the price and also what they call decreasing maturity pattern (Samuelson effect). It happens because a shock affecting the nearby contract price has an impact on succeeding prices that decreases as maturity increases. Indeed, as futures contracts reach their expiration date, they react much stronger to information shocks, due to the convergence of futures prices to spot prices upon maturity.

The authors claim that the proposed model is extensible and can be used for more complex crude oil financial securities, where the payoff structure is a linear function of the spot price of crude oil. However, it has drawbacks. Firstly, it does not prevent the convenience yield from taking negative values, a fact that opposes the non-arbitrage theory. Negative convenience yield for oil means negative “insurance costs” for refiners, and that in turn means they can buy crude oil, pay for all of its storage costs and hedge it by selling the next month's future. Secondly, the model assumes that the volatilities of the spot price and the convenience yield are constant and so the correlation among them. The above is in contrast to the theory of storage which claims that the volatility of the spot price depends on the level of the spot price and the convenience yield. When the inventories are rare, S is high. In this situation, any change in the demand has an important impact on the spot price, because inventories are not sufficiently abundant to absorb the price fluctuations.

3.2 Definition of the Model and its Parameters

In this section we present the foundation of the model for pricing crude oil futures, established by Gibson and Schwartz (1990). As it was already stated in the previous section, the main assumption of the model is that the crude oil future prices depend only on the crude oil spot prices, the instantaneous net convenience yield of the crude oil and the time to maturity of the future contracts. Additionally, the authors base their model on the fundamental assumption that the spot prices and convenience yield of the crude oil follow a joint stochastic process, which can be expressed as follows:

$$\frac{dS}{S} = \mu dt + \sigma_1 dz_1 , \tag{3.1}$$

$$d\delta = k(\alpha - \delta)dt + \sigma_2 dz_2 , \tag{3.2}$$

where dz_1 and dz_2 are correlated increments of standard Brownian motion processes with $dz_1 \cdot dz_2 = \rho dt$. With ρ is denoted the correlation coefficient between the two Brownian motion processes.

The specific formulation of the Brownian motion processes defined above, is based on the facts that: (1) the spot price of the crude oil follows a lognormal distribution, and (2) the future convenience yields of the crude oil follow a mean reverting stochastic process of the Ornstein-Uhlenbeck type. When the Ornstein-Uhlenbeck process is applied to convenience yield is relied in the assumption that there is a level of inventories, which satisfies the needs of industry under normal conditions. When the convenience yield is low, the inventories are abundant and the operators sustain a high storage cost compared with the benefits related to holding the raw materials. Therefore, if they are rational, they try to reduce these surplus inventories. Conversely, when the inventories are rare the operators tend to reconstitute them. Such a formulation is a good illustration of the fact that the convenience yield is implicit revenue associated with physical inventories.

Using the assumption that the crude oil contingent claim price *B* is a continuous and twice differentiable function of *S* and δ, the Lemma of Ito can be applied in order to define its instantaneous price change:

$$\begin{aligned} dB &= B_S dS + B_\delta d\delta - B_r dt + \frac{1}{2} B_{SS} (dS)^2 + \frac{1}{2} B_{\delta\delta} (d\delta)^2 + B_{S\delta} dS d\delta \\ &= \left\{ - B_r - \frac{1}{2} B_{SS} \sigma_1^2 S^2 + B_{S\delta} S \rho \sigma_1 \sigma_2 + \frac{1}{2} B_{\delta\delta} \sigma_2^2 + B_S \mu S + B_\delta [k(\alpha - \delta)] \right\} dt \\ &\quad + \sigma_1 S B_S dz_1 + \sigma_2 S B_\delta dz_2 \end{aligned} \quad (3.3)$$

Based on "perfect market" assumptions which imply absence of arbitrage and considering that interest rates are not stochastic, the authors have shown that the future price of the contingent crude oil claim should satisfy the following partial differential equation (pde):

$$\frac{1}{2} B_{SS} \sigma_1^2 S^2 + B_{S\delta} S \rho \sigma_1 \sigma_2 + \frac{1}{2} B_{\delta\delta} \sigma_2^2 + B_S S (r - \delta) + B_\delta [k(\alpha - \delta) - \lambda \sigma_2] - B_r - rB = 0 \quad (3.4)$$

where with λ is denoted the market price per unit of convenience yield risk which is considered to be a function of *S*, δ, and *t*.

Finally, it follows that the price F of a future on one barrel of crude oil with maturity time T should satisfy the following pde:

$$\frac{1}{2}F_{SS}\sigma_1^2 S^2 + F_{S\delta}S\rho\sigma_1\sigma_2 + \frac{1}{2}F_{\delta\delta}\sigma_2^2 + F_S S(r-\delta) + F_\delta\left[k(\alpha-\delta) - \lambda\sigma_2\right] - F_r = 0 \qquad (3.5)$$

with the following initial condition:

$$F(S,\delta,0) = S\,. \qquad (3.6)$$

The above pde has been solved analytically and a closed form solution has been obtained by Lautier and Galli (2005). The solution can be stated as follows:

$$F(S,\delta,\tau) = S(t)\times e^{-H\delta + B(\tau)}, \qquad (3.7)$$

where

$$H = \frac{1-e^{-k\tau}}{k}, \qquad (3.8)$$

$$B(\tau) = \left(r - \hat{a} + \frac{1}{2}\frac{\sigma_2^2}{k^2} - \frac{\sigma_1\sigma_2\rho}{k}\right)\times\tau + \frac{1}{4}\sigma_2^2\,\frac{1-e^{-2k\tau}}{k^3} + \left(\hat{a}k + \sigma_1\sigma_2\rho - \frac{\sigma_2^2}{k}\right)\times\left(\frac{1-e^{-k\tau}}{k^2}\right) \qquad (3.9)$$

with

$$\hat{a} = a - \frac{\lambda}{k}, \qquad (3.10)$$

and

$$\tau = T - t. \qquad (3.11)$$

Obviously, in order the above model to be of practical use, we need to estimate the variables and parameters that are part of it. These are the Convenience Yield Risk Market Price λ and the parameters k, α, σ_1, ρ, and σ_2 which are part of the joint stochastic processes followed by the spot price and the convenience yield. Moreover, the convenience yield δ must be calculated beforehand and become input parameter in the above pde. All these parameter calculations and estimations are presented in the next sections of the Chapter.

3.3 Estimation of Convenience Yield

As already discussed in Section 3.2, in order to be able to calculate the crude oil futures prices, we need to obtain the approximate values of the spot price S and the convenience yield δ, as none of these variables can be actually observed.

Regarding the crude oil spot prices, based on the assumption made by Gibson and Schwartz, the settlement price of the closest maturity crude oil future trading on the New York Mercantile Exchange, can be used as a very good approximation.

Regarding the convenience yield, the method proposed by the authors to approximate it, has been based on the strong mean reverting tendency of the short term (two - six months duration) annualized forward convenience yields. Taking into account the theory of storage, where there is an inverse relationship between the level of inventories and the relative net convenience yield, and also the fact that crude oil markets are highly volatile it has been proved that the convenience yield will remain finite. Moreover, due to the nature of the crude oil markets, the changes of convenience yield will be more evident during oil market turmoil.

The process that is used by the authors to calculate the instantaneous convenience yield of crude oil is strongly based on the well established relationship between the futures and the spot price of a commodity when both interest rates and convenience yield are considered to be constant. This relationship can be expressed as:

$$F(S,T) = Se^{(r-\delta)(T-t)}. \qquad (3.12)$$

Based on the above relation, we can determine the annualized monthly future convenience yields, using futures with successive monthly maturities, as follows:

$$\delta_{T-1,T} = r_{T-1,T} - 12\ln\left(\frac{F(S,T)}{F(S,T-1)}\right), \tag{3.13}$$

where $\delta_{T-1,T}$ represents the T-1 periods ahead annualized one month future convenience yield and $r_{T-1,T}$ stands for the T-1 periods ahead annualized one month risk-free future interest rate. From the results presented by the authors, it is evident that the volatility of the convenience yield of crude oil is very high, compared to the spot price of crude oil which is quite stable and seems to follow a random walk.

However, in our implementation of the proposed model, we have used an alternative approach to calculate the implied convenience yield from observed futures prices. The proposed calculation is again based on equation:

$$F(S,T) = Se^{(r-\delta)(T-t)} \tag{3.14}$$

where, as it has already been mentioned, the spot price S is approximated by the closest to maturity future price. Using the equation above to evaluate future prices F_1 (second closest to maturity) and F_2 (third closest to maturity) with corresponding maturities T_1 and T_2 and interest rates r_1 and r_2, we get:

$$F_1 = F_0 e^{(r_1-\delta)(T_1)}, \tag{3.15}$$

$$F_2 = F_0 e^{(r_2-\delta)(T_2)}, \tag{3.16}$$

where F_0 is the closest to maturity future price.

To calculate the implied convenience yield between now and the next month we divide the above two relations and get:

$$\frac{F_2}{F_1} = \frac{e^{(r_2-\delta)(T_2)}}{e^{(r_1-\delta)(T_1)}}. \tag{3.17}$$

Solving the above equation to obtain the value of implied convenience yield δ, we take the following solution:

$$\delta = \frac{r_1 T_1 - r_2 T_2}{T_1 - T_2} + \frac{\ln(F_2/F_1)}{T_1 - T_2}. \tag{3.18}$$

Finally, to calculate the risk-free interest rates, the author uses the more suitable libor rates, more precisely the libors with maturity respective to the remaining days to maturity of each future (interpolation). To obtain such a libor rate, if it does not exist, the author interpolated between the two libor rates which enclose the maturity of the future under consideration.

3.4 Estimation of Model Parameters

Before calculating the parameters of the model, we should analyze the behavior of the crude oil spot price time series, to validate that the lognormal distribution assumption is supported by the data. The authors of the model were based on a sample of 5 years weekly price data (from January 1984 to November 1988) and applied a linear regression. The results provided evidence that data support the conjectured assumption. Moreover, the historical standard deviation σ_1 of the logarithmic returns proved to be stable within sub-periods with relatively low volatility over the entire period captured.

Since the stochastic processes of the crude oil spot prices and the future convenience yields have residuals which are correlated, the authors have applied a Seemingly Unrelated Regression (SUR) model to estimate the parameters of interest k, α, σ_2, and ρ. In more detail the SUR model is based on the following equations:

$$\delta_t - \delta_{t-1} = \alpha k + k\delta_{t-1} + e_t, \tag{3.19}$$

$$\ln\left(\frac{S_t}{S_{t-1}}\right) = a + b\ln\left(\frac{S_{t-1}}{S_{t-2}}\right) + e_t\,. \tag{3.20}$$

According to what the authors noted, the results made evident a very strong mean reverting pattern (a high value of *k*) of the convenience yield. The value of the long term mean of the convenience yield proved to be fairly stable across time, while σ_2, proved to be a function of the number of sharp oil price declines or increases observed. One last thing that was noted was that the correlation of the residuals of the two processes proved the positive relationship between the unexpected changes of spot prices and convenience yields of crude oil.

3.5 Estimation of Convenience Yield Risk Market Price

An important parameter of the proposed model is the market price of convenience yield risk λ, which is considered to be constant for a period of time. The market price of risk is the difference between the expected rate of return of the underlying asset and the riskless interest rate, reported to the quantity of risk measured by the volatility.

It is essential to understand the meaning of the market price of risk because commodity markets are not complete. If a market is complete a derivative asset can be duplicated by a combination of other existing assets. If the latter are sufficiently traded to be arbitrage free evaluated, they can constitute a hedging portfolio whose behavior replicates the derivatives behavior. Their proportions are fixed such as there are no arbitrage opportunities and the strategy is risk-free. Then, in equilibrium, the return of the portfolio must be the risk free rate. The valuation is made in a risk neutral world so it does not depend on the attitude toward risk of the operators. Commodity markets are far from being free of arbitrage opportunities. Thus, valuation will probably not be realized in a risk-neutral world for commodities markets and several risk neutral probabilities may coexist.

To estimate λ, empirical data of crude oil futures contracts have been used by the authors and compared to their theoretical prices, for a period of almost 5 years (January 1984 to November 1988). To calculate the theoretical futures prices they used approximation methods, as the closed form solution of the pde presented in (eq. 3.7) was formulated later by Lautier and Galli (2005).

In more detail, the calculation process starts with three arbitrary values of λ, computes for each one of them the sum of squared errors and then estimates a new λ^* by assuming that the sum of squared errors is a second order polynomial in λ and setting λ^* equal to the value that minimizes the polynomial. The process is repeated until two successive values of λ^*, lead to respective mean root squared pricing errors which differ by less than one cent. Using the period from January 1984 to November 1988 (a total of 2,180 weekly futures prices) the estimated value of λ was founded by the authors to be equal to -1.796. The negative value of λ implies that it pays to bear convenience yield risk.

4 Model Implementation and Testing

In order to check validity and performance of the proposed by Gibson and Schwartz model, we have implemented and test it using real crude oil futures prices for contracts that are actually traded in NYMEX. In this Chapter we present and analyze our implementation. It should be mentioned that we have used a very large sample of futures data, which refer to time periods very different from those that Gibson and Schwartz did. In the next sections, we explain our results and compare them towards those derived by the Gibson and Schwartz. As software implementation platform, we have used Matlab, due to the advantages that this package presents regarding manipulation of large data sets and specific functions that are required for this type of implementations.

4.1 Introduction

For the current implementation, the data that is used consists of daily traded crude oil futures prices from January 2nd 1990 to September 27th 2012, crude oil spot prices from June 27th 1990 to July 26th 2012, and libor rates from January 2nd 1990 to September 27th 2012 which where provided by Bloomberg.

Crude oil futures (West Texas Intermediate) are traded in the New York Merchantile Exchange (NYMEX) under the ticker symbol CL in US dollars per barel. They are delivered in January, February, March, April, May, June, July, August, September, October, November and December under the respective ticket symbols CLF, CLG, CLH, CLJ, CLK, CLM, CLN, CLQ, CLU, CLV, CLX, CLZ. Each contract has a size of 1000 barrels.

The first step of our implementation is the preparation of the data. The data have been downloaded from Bloomberg in excel spreadsheets and have been suitably prepared in order to facilitate usage of the required Matlab structures. Although we have not spent effort to reorganize the derived data, we have made some minor data manipulation, such as eliminating blank columns in order to be able to automate parsing of the spreadsheets and populating the suitable Matlab structures. Moreover, we have automatically transformed the dates to numbers (changing the format of the cells that include the dates), to facilitate automatic date matching while parsing the various columns of the spreadsheets. The files that have been used to provide data to Matlab, along with a short description of each one of them are presented in the Table 4.1 below.

Table 4.1: Input Files for the Implementation of the Model using Matlab.

File No	File Name	Description of Data
1.	m_daily crude oil.xls	Daily spot data, required for the calculation of market price convenience yield risk.
2.	m_weekly crude oil.xls	Weekly spot data that was derived from the daily data, keeping the prices of every end of week (Friday) and eliminating the others.
3.	m_libor.xls	Full history of libor rates. Missing data has been completed as follows: (1) 2-weeks libor rates have been updated with 1-month libor rates where missing, (2) 1-week libor rates have been updated with 2-weeks libor rates where missing, and (3) overnight libor rates have been updated with 1-week libor rates where missing (parsing has been done in this order).
4.	m_futures 1994-1990.xls	Futures prices as listed by NYMEX, starting with January 1994 expiry in the first 2 columns (date in numeric format in the first and price in the second), then February 1994, etc., and continuing until December 1990 expiry, with the same logic.
5.	m_futures 2000-1995.xls	Futures prices as listed by NYMEX, starting with January 2000 expiry in the first 2 columns (date in numeric format in the first and price in the second), then February 2000, etc., and continuing until December 1995 expiry, with the same logic.
6.	m_futures 2006-2001.xls	Futures prices as listed by NYMEX, starting with January 2006 expiry in the first 2 columns (date in numeric format in the first and price in the second), then February 2006, etc., and continuing until December 2001 expiry, with the same logic.

7.	m_futures 2012-2007.xls	Futures prices as listed by NYMEX, starting with January 2012 expiry in the first 2 columns (date in numeric format in the first and price in the second), then February 2012, etc., and continuing until December 2007 expiry, with the same logic.
8.	m_daily_convenience_yield_data.xls	Daily data required to calculate implied convenience yield. More specifically, 1st column contains all working dates from 1990 (converted to numbers), 2nd and 3rd columns contain respectively maturity year and month of the second closest future contract, 4th column contains number of days from current date to maturity T1, 5th and 6th columns contain respectively maturity year and month of the third closest future contract, 7th column contains number of days from current date to maturity T2.
9.	m_daily_eval_data.xls	Daily data required to evaluate the accuracy of the implemented model. More specifically, 1st column contains all working dates from 1990 (converted to numbers), 2nd and 3rd columns contain respectively maturity year and month of the second closest future contract, 4th column contains number of days from current date to maturity T1, 5th and 6th columns contain respectively maturity year and month of the third closest future contract, 7th column contains number of days from current date to maturity T2, 8th and 9th columns contain respectively maturity year and month of the fourth closest future contract, 10th column contains number of days from current date to maturity T3.

Before starting calculations required by the proposed method, we have conducted a linear regression on lognormals of spot prices to test the validity of the model. This regression has been executed twice, one for the full set of crude oil spot prices data (Period A) and one for the first half of the period, 1990-2000 (Period B). The statistics of both regressions are presented in Tables 4.2 below. From the data of these Tables it becomes evident that in both cases the model is valid. More specifically, statistics t-stat and DW-stat show that in both cases the model performs well, as *b* is statistically important, while R^2 increases with the number of observations (N).

Table 4.2: Regression statistics for spot prices based on the model $\ln\frac{S_{t+1}}{S_t} = a + b\ln\frac{S_t}{S_{t-1}} + e_t$.

Period	*a*	*b*	t-stat	DW-stat	R^2	*N*
A: 1990-2012	-0.0892	0.0016	1.0317	2.0017	0.0080	1152
B: 1990-2000	-0.0691	0.0009	0.4145	1.9964	0.0048	549

Moreover, based again on the same samples, we have calculated the mean value, the standard deviation and the annualized standard deviation of the crude oil spot prices. Annualized standard deviation becomes input variable for the calculations required by the model as will be presented in detail in the next sections. The calculated values for both periods are summarized in the Table 4.3 below. We can notice from the numbers reported on the table that the mean value is small in both cases and significantly smaller for Period B compared to Period A. Moreover, standard deviation for both cases is close to that reported by Gibson and Schwartz (1990), while it is mentionable that Period B seems to be slightly less volatile compared to Period A, as annualized standard deviation for Period B is about 1,9% smaller than those of Period A.

Table 4.3: Mean price and standard deviation for both periods.

Period	Mean (μ)	Annualized Standard Deviation (σ_1)
A: 1990-2012	0.0015	37.76%
B: 1990-2000	8.22×10^{-4}	35.85%

The figure 4.1 (a) below, displays the weekly crude oil spot prices as a function of time for Period A. As our observations cover a period of almost 22 years, we can notice on the graph that there is a very large difference from the minimum (about 15$) to the maximum (about 150$) of the period. We can also notice that in specific subperiods there are very sharp movements of the spot price. Similar observations can be made also for Period B (see Figure 4.1 (b)), although the min and max values of the spot prices within this period present smaller difference.

Figure 4.1 (a): Weekly crude oil spot prices as a function of time for Period A.

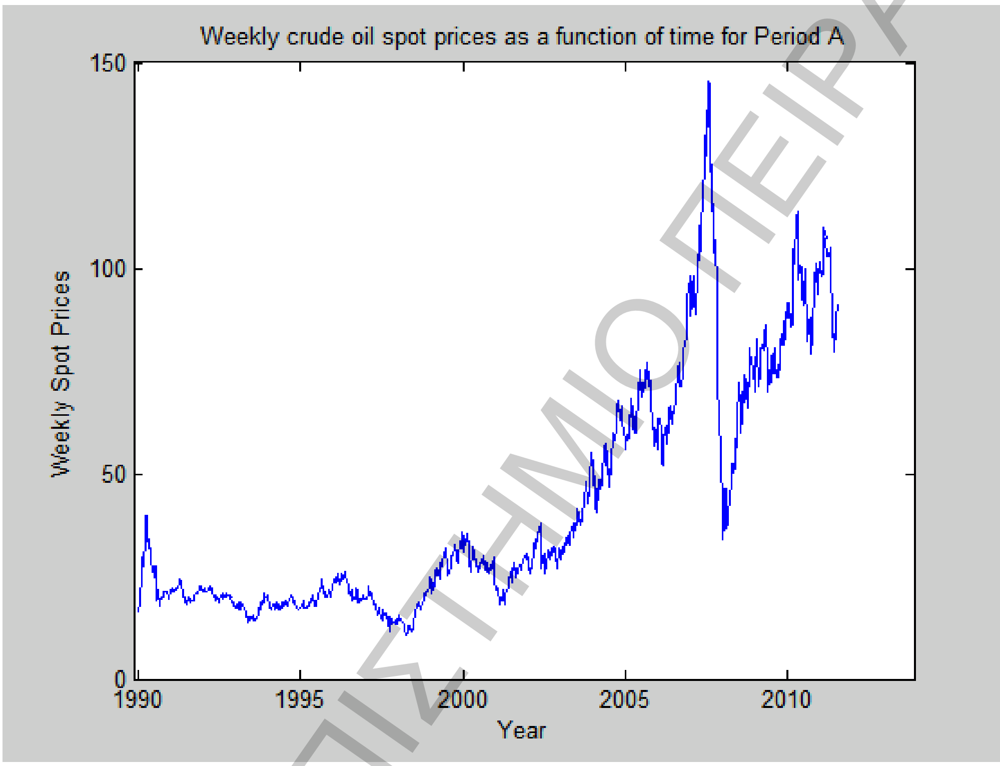
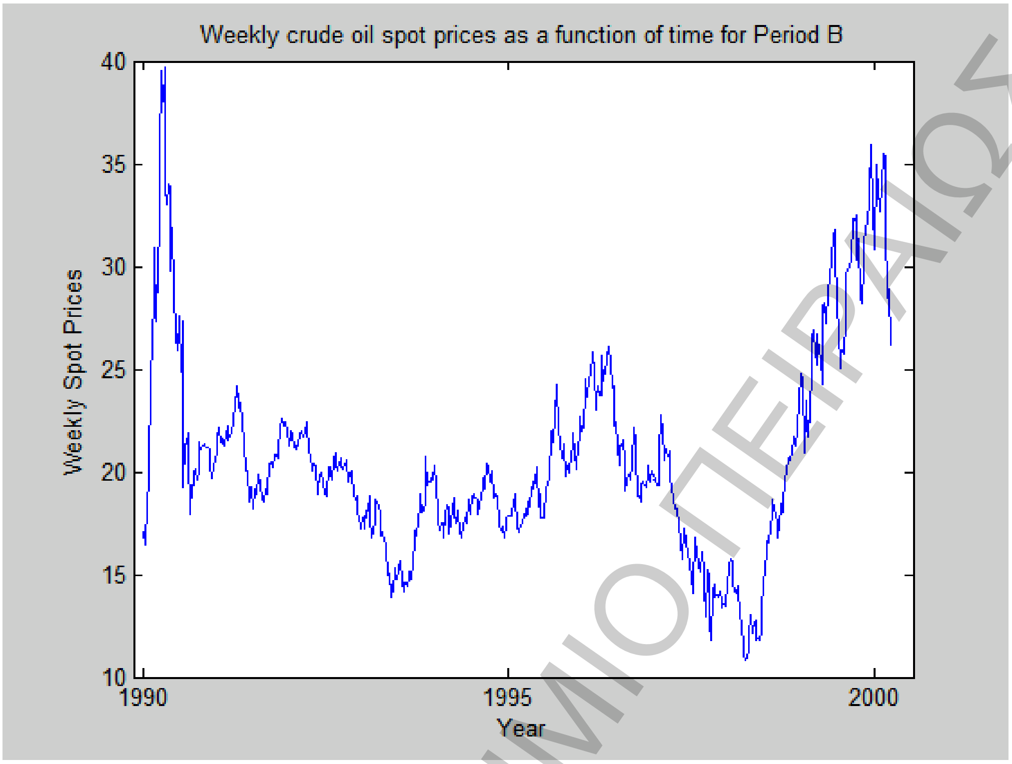


Figure 4.1 (b): Weekly crude oil spot prices as a function of time for Period B.



This pattern of spot prices implies that, it is not probably optimal to base implementation of our model on large periods of time, but instead split in subperiods and examine each subperiod independently. This argument is also supported by the dramatic market changes, especially during the last 3 years due to the global financial crisis. The approach of splitting into subperiod has been followed also by Gibson and Schwartz (1990) leading to better results and is a subject for further research in the future.

Finally, it should be noted that all the above preparation steps are necessary to continue with the calculations required for the convenience yield, which are described in detail in the next section.

4.2 Calculation of Convenience Yield

To calculate the convenience yield for crude oil, we have applied the technique that was presented in Section 3.3 for the full set of data of crude oil futures prices (Period A), but also for the first half of the period, 1990-2000 (Period B).

For performing the required calculations, we have used, among others, as input the futures prices of the second closest to maturity future F1. To identify the appropriate futures data, in the input files, we parse all four files with crude oil future prices and try to match the date in each file with dates in a suitably prepared working matrix (Matlab variable conv_yield). More specifically, to automatically find the required future price, we have used a formula that selects the appropriate column of the spreadsheet and then searches in this column until it finds a date that matches with the date for which we need to populate the conv_yield working matrix. For example for future contracts which mature in September 1991, the column with the required data is given by 120 - (5 - (1994 - **1991**)) * 24 + **9** * 2 = 90 (where with bold characters are mentioned the year and the month of the given maturity date which are parameters of the formula). The formula is slightly adjusted to be effective for each one of the four files. The four formulas are presented in Table 4.4 below, using as parameters for maturity year *Mat_Year* and for maturity month *Mat_Month*.

Table 4.4: Formula for the automatic selection of the appropriate columns of input files.

	File Name	Formula
1.	m_futures 1994-1990.xls	120 - (5 - (1994 - Mat_Year)) * 24 + Mat_Month * 2
2.	m_futures 2000-1995.xls	144 - (6 - (2000 - Mat_Year)) * 24 + Mat_Month * 2
3.	m_futures 2006-2001.xls	144 - (6 - (2006 - Mat_Year)) * 24 + Mat_Month * 2
4.	m_futures 2012-2007.xls	144 - (6 - (2012 - Mat_Year)) * 24 + Mat_Month * 2

The process described above is applied for both future prices, F1 (2^{nd} closest to maturity) and F2 (3^{rd} closest to maturity). Finally, as a precaution measure, in order to avoid malfunction of the program in case of missing values (blank cells) in any cell of the input files, we update any missing future value with the previous non-missing future value starting from the older ones.

To provide more credible results data **filtrations** have also been conducted which exclude the last three days before maturity as behavior of the future price changes significantly in this period. Moreover, dates with very small volumes of traded future contracts, have been excluded from our sample in order to provide more accurate results and obtain a more effective implementation of the proposed model.

Apart from the future prices F1 and F2, to calculate the convenience yield we need the appropriate interest rates for remaining durations of both F1 and F2. These are based on the libor rates which have been derived from input file “m_libor.xls”. The appropriate libor rates have been selected using as a basis the number of days to maturity of each future contract. With an approach similar to what has been described before, for selection of future prices, the two closest libors are selected for both F1 and F2 and then an interpolation process is applied to determine the applicable interest rate for each maturity. Before applying the interpolation, as was explained before for the selected future prices, in order to avoid malfunction of the implemented program in case of missing values (blank cells) in any place of the input files, we update any missing libor value with the previous non-missing libor value starting from the older ones. Interpolation, to find the applicable interest rates for futures F1 and F2, is presented in Section 3.3,

$$\delta = \frac{r_1 T_1 - r_2 T_2}{T_1 - T_2} + \frac{\ln\left(F_2 / F_1\right)}{T_1 - T_2}. \tag{4.1}$$

It should be mentioned that the final result for the calculated convenience yields, after applying the above process, are divided by 100, to be transformed from a percentage to a number (between zero and one) in order to become aligned with the rest of the parameters in the required calculations. Calculations take place on a daily basis and the calculated convenience yields are displayed in the Figure 4.2 below.

As with the above method we have calculated the convenience yield on a daily basis, we then apply a selection of the end of week convenience yields to create a weekly sample (stored in Matlab conv_yield matrix), which we then use for regression purposes. Figures 4.3 (a) and 4.3 (b) display the evolution of the weekly convenience yield over time for Periods A and B correspondingly, while in Figures 4.4 (a) and 4.4. (b) evolution of the weekly convenience yield over time is displayed in the same graph with the evolution of weekly crude oil spot prices over time for Periods A and B correspondingly.

Figure 4.2: Daily convenience yield as a function of time.

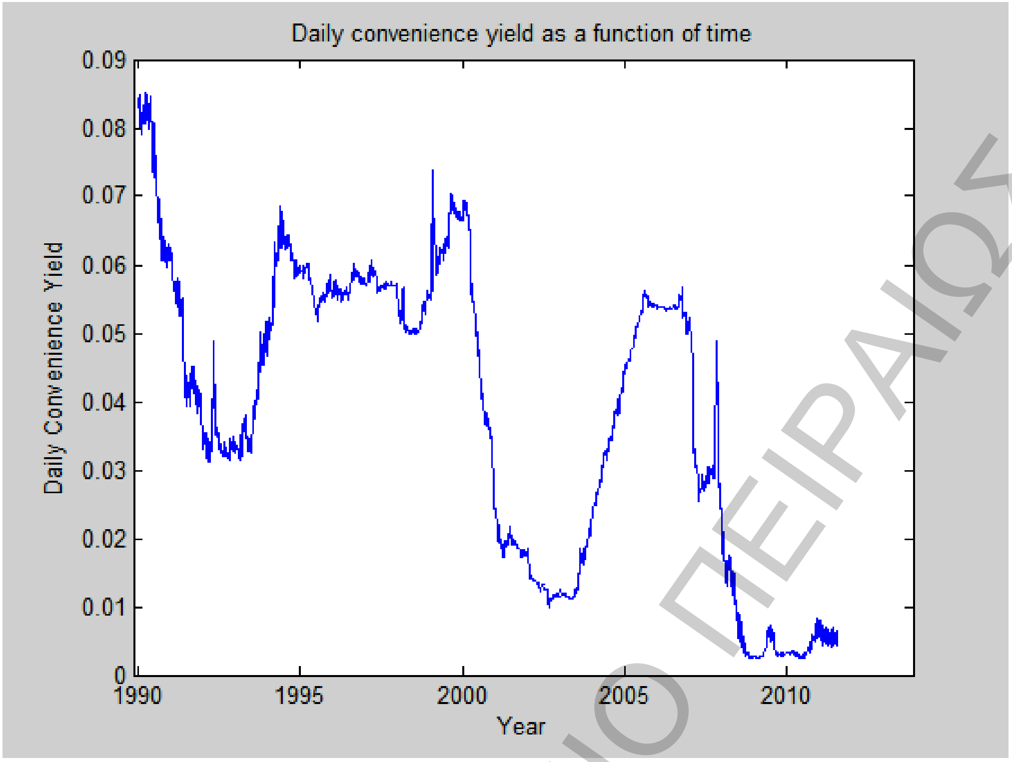


Figure 4.3 (a): Weekly convenience yield as a function of time for Period A.

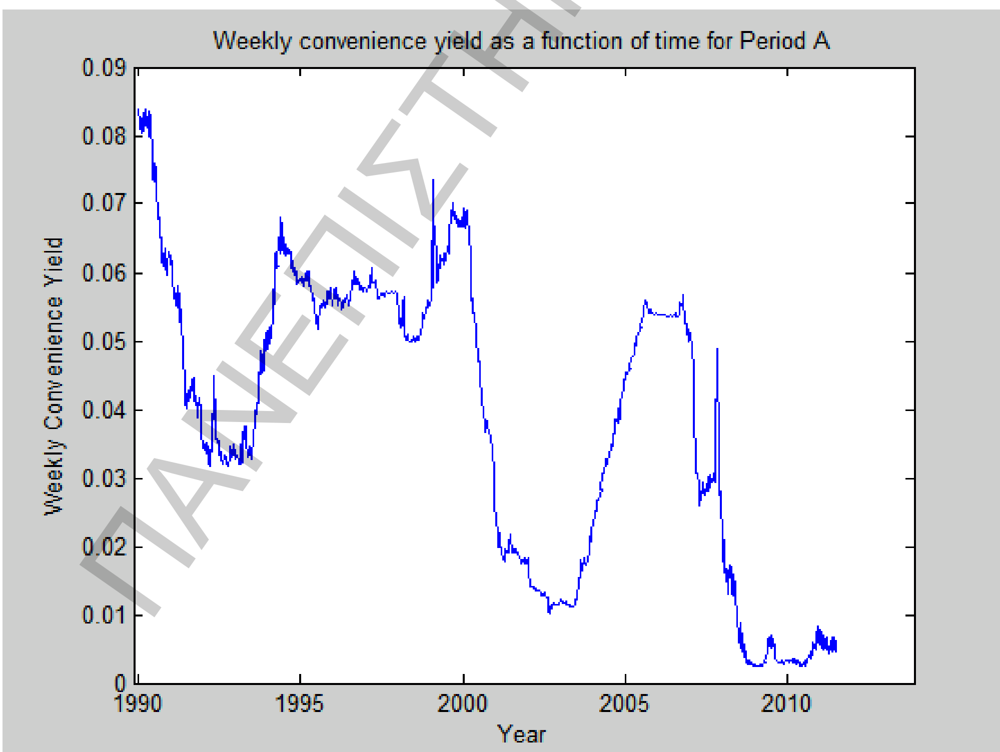


Figure 4.3 (b): Weekly convenience yield as a function of time for Period B.

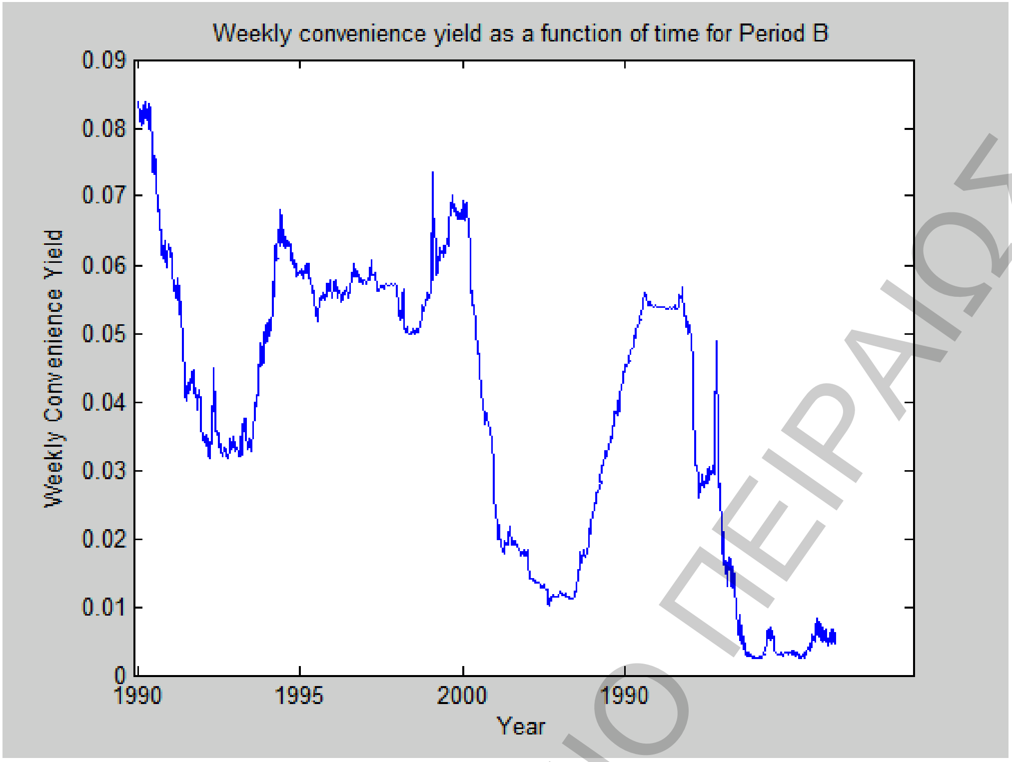


Figure 4.4 (a): Crude oil spot prices and convenience yield as a function of time for Period A.

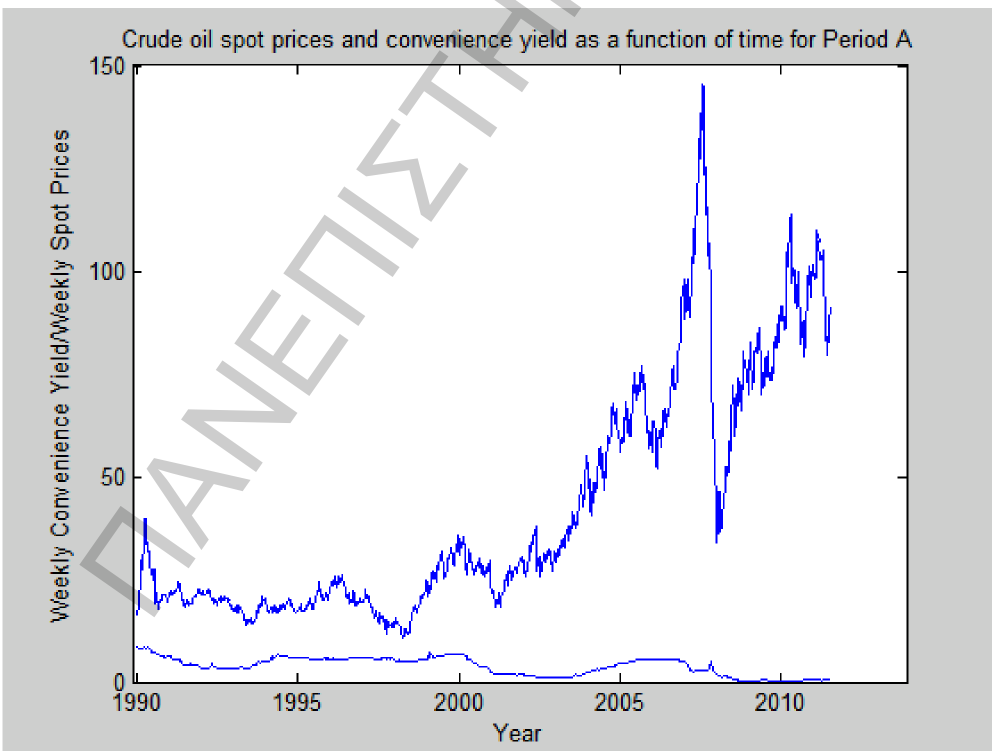
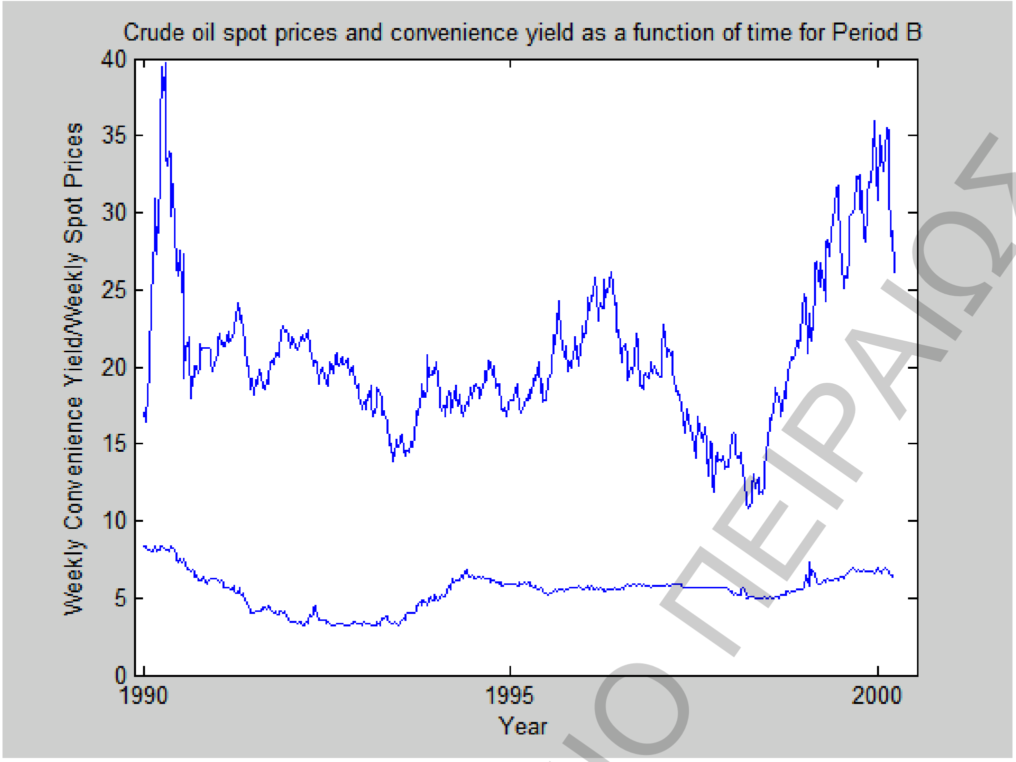


Figure 4.4 (b): Crude oil spot prices and convenience yield as a function of time for Period B.



What can be derived by the previous graphs is that (1) convenience yield fluctuations follow the pattern of the corresponding spot prices, (2) the range of convenience yield is from about 0,4% to 8,5% for Period B, while it is smaller for Period A excluding Period B (0,2% to 6%), (3) spot prices have increase significantly over the last decade while at the same period convenience yield has been significantly decreased, (4) convenience yields tend to decrease over time, and are very low (less than 0,5%) in the last three years.

4.3 Calculation of Model Parameters

As we have already mentioned, a number of parameters should be estimated, to make the model practically implementable. These parameters are k, α, σ_2 (standard deviation) and ρ (correlation). It has been proposed by Gibson and Schwartz (1990) that these estimations are conducted by performing a multivariate Seemingly Unrelated Regression (SUR). Using SUR, calculation of the parameters is feasible and seems to be realistic since the stochastic processes of the spot price of crude oil and of the future price convenience yields have correlated residuals. The regression model consists of the following two components:

$$\delta_t - \delta_{t-1} = ak - k\delta_{t-1} + \varepsilon_t \,, \tag{4.2}$$

$$\ln\left(\frac{S_t}{S_{t-1}}\right) = \alpha + b\ln\left(\frac{S_{t-1}}{S_{t-2}}\right) + \varepsilon_t \,. \tag{4.3}$$

To perform a multivariate seemingly unrelated regression with Matlab, we create a vector of the "previous" values (right hand side) and a vector of the "current" values (left hand side). We use a special function to "convert" the regression to seemingly unrelated (function "convert2sur") with a specific structure of the Design vector and exploit Matlab's multivariate regression function ("mvnrmle"). The estimations of the above mentioned parameters come out as the result of the conducted regression. As is proposed by Gibson and Schwartz (1990), the value of k is annualized. After completing the regression and in order to obtain the correlation ρ, we transform the covariance (outcome of the conducted regression) to a correlation matrix using the specific Matlab function "corrcov". Last but not least, we calculate the mean value, as well as the standard deviation and the annualized standard deviation of the weekly convenience yield.

The derived parameters from the process just described, as well as the statistics of the conducted regression for (eq. 4.2) of the model have been estimated for both the full set of weekly data (Period A) and the first half of the period, i.e. 1990-2000 (Period B) and the results are presented in the Table 4.5 below.

Table 4.5: Calculated parameters of the model.

Period	k	a	σ_2	ρ	DW	R^2	N
A: 1990-2012	0.0181	0.0343	0.1569	-0.0444	2.1089	0.1332	1152
B: 1990-2000	0.2007	0.0494	0.0880	-0.0538	2.2931	0.1869	549

From a first view and comparing with the results of presented in Gibson and Schwartz (1990) it seems that most of the parameter values vary significantly. For example k and σ_2 are much smaller. An explanation for this is the entirely different and larger data set that we have used in our implementation. Not only our data set was much larger, but also market behavior has changed significantly over the last two decades. The spot prices increased a lot and interest rates declined rapidly. A small evidence for this change could come when comparing the value of k presented in Table 4.5. We can easily notice that the value of k is more than 10 times larger for Period B compared with Period A. This fact shows the trend of this parameter in the last two decades. Of course, further analysis and research should be conducted, to justify the pattern, for example by splitting the data into smaller time periods, estimating the parameters period by period and comparing the results. This analysis is not included in the scope of the current Thesis and is left for future research.

4.4 Calculation of Convenience Yield Risk Market Price

The market price of the convenience yield risk (λ), is another very important parameter of the proposed model and must be calculated to make the model functional as it can not be observed in the markets. In this section we describe the method that has been used to calculate λ. The proposed algorithm has been based on a similar method originally proposed in (Gibson and Schwartz, 1990), but includes also some new logic. The proposed algorithm assumes that λ is constant within one time period and uses the market future prices of all closest to maturity futures within the selected period to achieve the calculation. The underlying idea is to compare these future market prices with the corresponding theoretical values calculated by the model for 3 different values of λ.

These values are initially assumed to be arbitrary and they are updated during execution of the algorithm. In more detail this update substitutes the worst of the initial selections with a more appropriate one, which provides smaller squared error. More specifically, the steps of the Algorithm are the following:

1. Start with 3 arbitrary values of λ, i.e. λ_1, λ_2 and λ_3.
2. Compute for each of the selected λ_i the sum of squared errors of all closest to maturity future prices. In more detail this steps implies the following:
    a. Calculate the sums of squared errors for all futures, i.e. Error (λ_1), Error (λ_2), Error (λ_3).
    b. Assume that calculated errors can be derived from a second order polynomial of λ, i.e. $Error(\lambda) = a_1\lambda^2 + a_2\lambda + a_3$ and solve the 3 × 3 linear system with the values of Error calculated in (3a) to find a new λ. Consider as solution the root that minimizes the polynomial, i.e. $-a_2/2a_1$.
    c. Substitute with the new calculated λ, that current value of λ which is less close to the calculated one.
3. Repeat the process to compute the next appropriate λ.
4. Terminate after a sufficient number of iterations (e.g. 10 iterations), as after that number the coefficient matrix of the linear system to be solved becomes ill-conditioned and the solution of it is subject to large errors.

What differentiates our proposed algorithm from the one that is proposed in (Gibson and Schwartz, 1990) is the 4th step, where we propose a fixed number of steps instead of iterating until minimizing the squared error. The proposed algorithm has been tested on the given data sets and seems to perform well achieving a four decimal digits accuracy only with 10 iterations.

Finally, the above Algorithm has been applied to both the full set of weekly data (Period A) and the first half of the period, 1990-2000 (Period B). The derived results that are presented in Table 4.6 below:

Table 4.6: Market price of convenience yield risk.

Period	Convenience Yield Risk Market Price
A: 1990-2012	-0.4746×10^{-3}
B: 1990-2000	-0.0611

Comparing with the corresponding results of Gibson and Schwartz (1990), the values of λ obtained are much smaller, but agree on sign (negative numbers). For Period A the absolute value of λ is more than ten times smaller compared to that of Period B, which leads us to the conclusion that the value of λ depends heavily on the data sets and has significantly changed during the last two decades. Again this is subject to further research and can be part of future work to be conducted on the subject.

4.5 Model Performance

The performance of a model measures the ability to reproduce the term structure of commodity prices. To assess the performance, criteria values are needed. We first present these criteria.

Two criteria are usually retained to measure the performance of the model of a term structure model: the **mean pricing error** (MPE) and the **root mean squared error** (RMSE).

The MPE is defined as follows:

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

Where N is the number of observations, $\tilde{F}(n,\tau)$ is the estimated futures price for maturity τ at the date n, and $F(n,\tau)$ is the observed futures price. The MPE measures the estimation bias for a given maturity. If the estimation is good, the MPE should be very close to zero.

Using the same notation, the RMSE is, for a given maturity τ:

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

The RMSE can be considered as an empirical variance, which measures the estimation stability. This second criterion is considered as more representative because price errors can offset themselves and the MPE can be low even if there are strong correlations.

In this section we evaluate the implemented model, in terms of accuracy, and compare the results with those obtained in (Gibson and Schwartz, 1990). Moreover, we discuss on the performance of the model and pinpoint its advantages and disadvantages. To be able to measure errors and accuracy of the model we have created in Matlab the matrix "eval_futures" which holds the necessary input for our comparisons and also stores the derived outputs that are calculated in order to evaluate the model.

Our evaluation is twofold. First we take into account the full data set (Period A) and calculate the mean and squared errors between the model calculated daily future prices and the market prices provided in the input data files. This error calculation is done for the second, third and fourth closest to maturity future contracts. Second, we use the first half of the data (Period B) to calculate the parameters of the model (k, α, σ_1, σ_2, ρ and λ) and then we feed these estimations to the model to calculate the future prices for the second half of the time period (2001-2012) and the mean and squared errors for the first, second, third and fourth closest to maturity future contracts, on a daily basis. In both of these evaluations, we compare our results with those of Gibson and Schwartz (1990) and comment on the comparison. Moreover the two evaluations are compared to each other and the accuracy obtained is explained and justified.

In the next two sections, we present the details of each evaluation approach, as well as the derived results after execution of the program for each evaluation case.

4.5.1 Evaluation for Time Period 1990-2012

Although many of the comments and discussion of the previous sections are already part of an evaluation, in this section we go deeper and explain in more detail the results and errors that are produced when applying the proposed model to the full data set (Period A).

As we already explained, in what has been implemented, a special Matlab matrix named "eval_futures" has been used to facilitate evaluation. This matrix is first populated with future prices of the second, third and fourth closest to maturity future prices f2, f3 and f4. Again, as was done in previous cases, we complete potential missing values with the previous non-missing values for all f2, f3 and f4.

Moreover, to apply our model, it is necessary to populate "eval_futures" with the required libor rates. The approach for this population process is similar to what was applied for the calculation of convenience yield (see Section 4.2) and leads to the appropriate interest rate per future contract after execution of the required interpolation[1]. One more time, we complete potential missing values of interest rates with the previous non-missing values before interpolating.

Having the input data required in our "eval_futures" structure, we can then calculate future prices f2, f3, f4 using the analytical solution of the model (eqs. 3.7-3.11) and results are stored back to "eval_futures".

In Figures 4.5(a), (b), (c), estimated and observed futures prices for futures f2, f3 and f4 respectively, are presented at the same plot.

[1] Calculation assumes 30-day months.

Figure 4.5 (a): Estimated and observed futures prices for the closest to maturity future f1 for the period 1990-2012

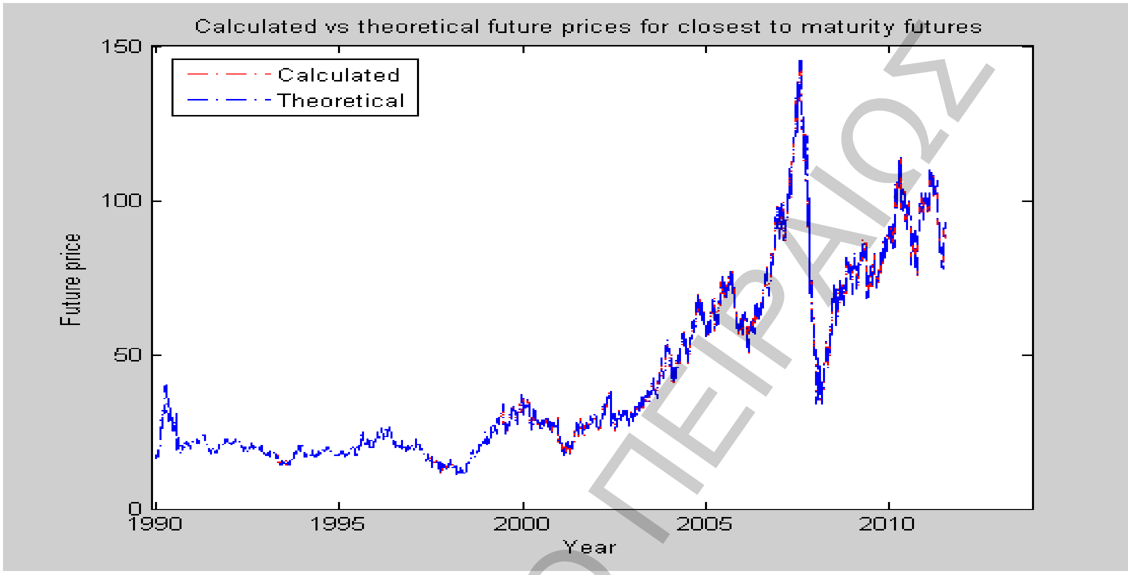


Figure 4.5 (b): Estimated and observed futures prices for the second closest to maturity future f2 for the period 1990-2012

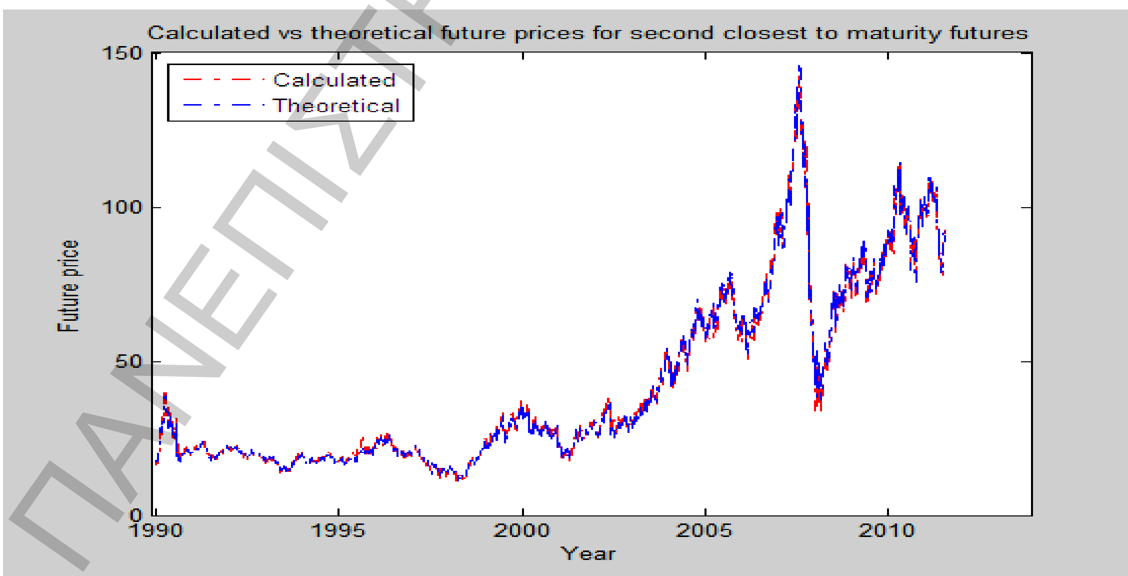
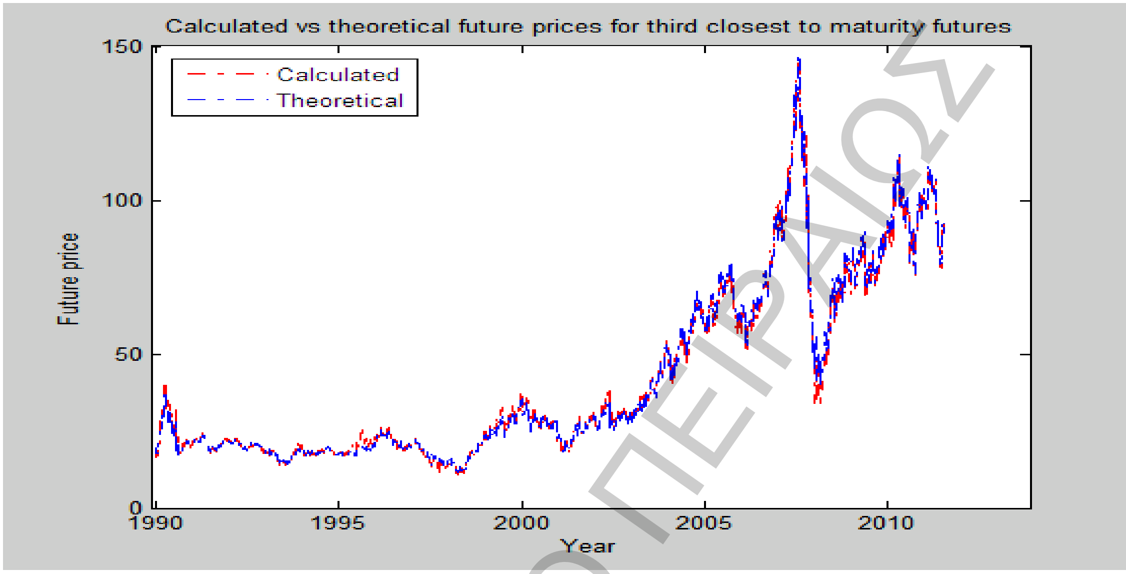


Figure 4.5 (c): Estimated and observed futures prices for the third closest to maturity future f3 for the period 1990-2012



Finally, mean and squared errors are calculated. A summary of the model errors for this evaluation case, consisting of the derived mean and squared errors are presented in Table 4.7 below for all three future prices.

Table 4.7: Period A mean and squared errors for futures of the first, second, third and fourth closest maturities.

Error Type	f1	f2	f3	f4
Mean error	0.0099	0.5723	1.0132	1.4251
Squared error	0.0579	0.8889	1.4880	1.9869

It becomes evident from Table 4.7 that:

1. Errors increase with the increase of time to maturity. For example, the first closest to maturity future price is approximated with a mean error that is about 57 times smaller compared to that of the second closest to maturity future price and about 101 times smaller compared to that of the third closest to maturity

future price. The case is similar when taking into account the squared errors. This finding is in line with what is presented in (Gibson and Schwartz, 1990).

2. Squared errors are larger than the mean errors. This is also aligned with the results obtained in (Gibson and Schwartz, 1990).
3. The results obtained in our implementation compared to those in (Gibson and Schwartz, 1990) are much better when considering second closest to maturity futures and similar for the third and fourth closest to maturity futures contracts.

4.5.2 Evaluation for Time Period 2001-2012

In this evaluation case, we split the time period into two sub-periods (1990-2000 and 2001-2012) from which the first one plays the role of the input data provider to calculate the model parameters, while the second one serves as another sample to be used for evaluating calculation accuracy by using the model parameters which were calculated from the data of the first sub-period. The calculation approach end estimation of errors is identical to that of the first evaluation case and the only thing that changes it the time periods where the various calculations are applied. A summary of the model input parameters for this evaluation case, as well as the derived mean and squared errors are presented in Table 4.8 below.

Table 4.8: Mean and squared errors for futures of the first, second, third and fourth closest maturities based on parameters calculated from Period B data set.

Error Type	f1	f2	f3	f4
Mean error	0.0104	0.7541	1.3254	1.7773
Squared error	0.0159	1.1026	1.8323	2.3698

It becomes evident from Table 4.8 that:

1. Errors increase with the increase of time to maturity. For example, the first closest to maturity future price is approximated with a mean error that is about 74 times smaller compared to that of the second closest to maturity future price

about 130 times smaller compared to that of the third closest to maturity future price and about 170 times smaller compared to that of the fourth closest to maturity future price. The case is similar when taking into account the squared errors. This finding is in line with what is presented in (Gibson and Schwartz, 1990).

2. Squared errors are larger than the mean errors. This is also aligned with the results obtained in (Gibson and Schwartz, 1990).
3. The results obtained in our implementation compared to those in (Gibson and Schwartz, 1990) are much better when considering second closest to maturity futures and similar for the third and fourth closest to maturity futures contracts.

Comparing results between the two evaluation cases, we can obviously see that, with the exception of the squared error for future f1, the model performs better in the first case. This is in line with our previous comments regarding subperiod splitting and it seems that as we move from an older to a more recent time period the parameters of the model are changing significantly, influencing negatively the obtained accuracy of the model.

5.Conclusion

In this thesis, a literature review of existing models for pricing commodities futures has been presented along with the theoretical foundation of a well known two-factor model (Gibson and Schwartz, 1990) based on spot prices and convenience yield. The model was implemented using Matlab and a detailed qualitative and quantitative analysis on calculated parameters was conducted. Moreover, performance of the model was evaluated for an extensive data set of market data. Data that has been used consists of daily futures prices from January 2nd 1990 to September 27th 2012, spot prices from June 27th 1990 to July 26th 2012 and libor rates from January 2nd 1990 to September 27th 2012.

We have noticed that it is not probably optimal to base implementation of our model on large time periods. Instead, splitting into subperiods and examining each subperiod independently would be more appropriate. This argument is also supported by the dramatic market changes, especially during the last 3 years due to the global financial crisis. Additionally, the approach of splitting into subperiods has been followed also by Gibson and Schwartz leading to better results and is a subject for further research in the future.

Convenience yield has been calculated for the given data set and it becomes evident that, its fluctuations follow the pattern of the corresponding spot prices. Convenience yield varies from 0,4% to 8,5% for Period B (2001-2012), while it is smaller for Period A (full data set) -excluding Period B (0,2% to 6%). Spot prices have increased significantly over the last decade while at the same period convenience yield has been significantly decreased. Moreover, convenience yield tend to decrease over time, and are very low (less than 0,5%) in the last three years.

Finally, performance of the implemented model has been assessed and findings can be summarized as follows:

- With the exception of the squared error for future f1, the model performs better for the first Period. It seems that as we move to a more recent time period, the parameters of the model are changing significantly, influencing negatively the obtained accuracy of the model.
- Errors increase with the increase of time to maturity.

- Squared errors are larger than the mean errors.
- Our results, compared to those in (Gibson and Schwartz, 1990), are much better for the 2nd closest to maturity futures and similar for the 3rd and 4th closest to maturity futures contracts.
- The model performs better for the Period A as we move to more recent data, the parameters of the model are changing significantly, influencing negatively the obtained model accuracy.

6. Bibliography

J. Keynes (1930), A treatise on Money, The applied Theory of Money, Londres

Working, H. (1948), The theory of the price of storage, American Economic Review 39, 1254-1262

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

Brennan and E. Schwartz (1982), An equilibrium Model of Bond Pricing and a Test of Market Efficiency, Journal of Financial and Quantitative Analysis, 17(3), 301-329

Brennan and E. Schwartz (1985), Evaluating Natural Resource Investments, Journal of Business, 58, 135-137

Cox, J. E. Ingersoll and S. A. Ross (1985) , A Theory of the Term Structure of Interest Rates, Econometrica, 53, 385-408.

E.F Fama and K.R French (1987). Commodity futures prices: some evidence of forecast power, premiums and the theory of storage, Journal of Business, 60(1), 55-74

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

Gibson and E. Schwartz (1990), Stochastic Convenience Yield and the Pricing of Oil Contingent Claims, The journal of Finance, 45 (3), 959-976.

H. Reisman (1991), Movements of term structure of commodity futures and pricing of commodity claims

A. Deaton and and G. Laroque (1992). On the behavior of commodity prices, Review of Economic Studies, 59(198), 1-24

G. Cortazar and E. Schwartz (1994), The valuation of commodity contingent claims, Journal of Derivatives, 1(4), 27-39

G. Cortazar and E. Schwartz (1997), Implementing a real option model for valuing an undeveloped oil field, International Transactions in Operational Research, 4(2), 125-137

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

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

K. Miltersen, E. Schwartz (1998), Pricing of Options on commodity futures with stochastic term structures of convenience yields and interest rates, Journal of Financial and Quantitative Analysis, 33(1), 33-59

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

C. Carter (1999), Commodity futures market: A survey. Australian Journal of Agricultural and Resource Economics, 43(2),209-248

E. Schwartz and J. Smith (2000), Short-Term Variations and Long-Term Dynamics in Commodity Prices, Management Science, 46, 893-911

S.Neftci (2000), An Introduction to the Mathematics of Financial Derivatives, Academic Press, Second Edition

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

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

M. Nielsen and E. Schwartz (2004), Theory of Storage and the Pricing of Oil Contingent Claims, Review of Derivatives Research

Ribeiro D.R and Hodges S.D, (2004), A two factor model for commodity prices and futures valuation

D. Lautier (2005), Term Structure Models Of Commodity Prices

D. Lautier and A. Galli (2005), Dynamic hedging strategies and commodity risk management Final report, Cerna.

Gorton G. Hayasi F. and Rouwenhorst G. (2008), The fundamentals of commodity futures returns

J. Hull (2008), Options, Futures and Other Derivatives, Prentice Hall, Seventh Edition

P. Brandimarte (2008), Introduction to Numerical Methods in Finance and Economics, John Wiley & Sons

A. Trolle, E. Schwartz (2009), Unspanned stochastic volatility and the pricing of commodity derivatives, The review of Financial Studies, 22(1)