Implementing a Stochastic Model for Oil Futures Prices

Gonzalo Cortazar[a], Eduardo S. Schwartz[b]

[a] *Departamento de Ingeniería Industrial y de Sistemas, Escuela de Ingeniería,*
Pontificia Universidad Católica de Chile, Vicuña Mackenna 4860, Santiago, Chile
[b] *Anderson School at UCLA, 110 Westwood Plaza, Los Angeles, CA 90095-1481, USA*

.

June 2002

Abstract

This paper develops a parsimonious three-factor model of the term structure of oil futures prices that can be easily estimated from available futures price data. In addition, it proposes a new simple spreadsheet implementation procedure. The procedure is flexible, may be used with market prices of any oil contingent claim with closed form pricing solution, and easily deals with missing data problems. The approach is implemented using daily prices of all futures contracts traded at the New York Mercantile Exchange between 1991 and 2001. In-sample and out-of-sample tests indicate that the model fits the data extremely well. Though the paper concentrates on oil, the approach can be used for any other commodity with well-developed futures markets.

1. Introduction

This paper develops a parsimonious three-factor model of the term structure of oil futures prices that can be easily estimated from available futures price data. In addition, it proposes a new simple spreadsheet implementation procedure. The procedure is flexible, may be used with market prices of any oil contingent claim with closed form pricing solution, and easily deals with missing data problems. The approach is implemented using daily prices of all futures contracts traded at the New York Mercantile Exchange between 1991 and 2001. In-sample and out-of-sample tests indicate that the model fits the data extremely well. Though the paper concentrates on oil, the approach can be used for any other commodity with well-developed futures markets.

In the last ten years there has been an increasing interest both by academics and practitioners in understanding the stochastic behavior of oil prices. This interest comes from increasing price volatility, which makes predictions more difficult. Also, modern asset pricing models of oil-related real assets such as oil-field development or extraction projects, consider these assets as real options contingent on the oil price level and its stochastic process (Paddock et al., 1988). In addition, pricing of new oil-linked financial instruments requires knowledge of the stochastic behavior of the underlying asset. Finally, commodity price risk, which can have a huge impact on a firm's profits (Culp and Miller (1994)), may be successfully hedged to the extent that the stochastic process for the underlying commodity is known.

It is well known that the futures price of a financial asset that pays no dividends is equal to the spot price of the asset plus the interest rate (carrying costs) over the life of the futures contract. Any dividend payment on the financial asset should be subtracted from the carrying costs. In the case of commodities, futures prices are normally lower than the spot price plus the interest rate over the life of the futures contract. This "shortfall", which is like an implicit dividend that accrues to the holder of the spot commodity but not to the holder of the futures contract, is what is known as the convenience yield of the commodity (Brennan, 1991).

Stochastic models of the behavior of commodity prices differ on the role played by the convenience yield and on the number of factors used to describe uncertainty. Early models assumed a constant convenience yield and a one-factor Brownian

motion (Brennan and Schwartz, 1985). This random walk specification for commodity prices was used until a decade ago, when mean reversion began to be included as a response to the evidence that volatility of futures returns declines with maturity. One-factor mean reverting models can be found for example in Ross (1995), Schwartz (1997), Cortazar and Schwartz (1997), Laughton and Jacoby (1993 and 1995).

An undesirable implication of one-factor models, however, is that all futures returns are perfectly correlated, a fact that defies empirical evidence. To account for a more realistic stochastic behavior, two-factor models with mean reversion were introduced. Examples are Gibson and Schwartz (1990), Schwartz and Smith (2000) and Schwartz (1997).

Enlightening as they may be, these stochastic models have been adopted rather slowly by practitioners. One possible reason for this could be that even though two-factor models behave reasonably well most of the time, for some market conditions they behave poorly, making daily estimations somewhat unreliable. Also, most parameter estimation procedures proposed in the literature are rather involved and require extensive data aggregation, which translates into substantial information loss.

In this paper we propose a new three-factor model that performs much better than existing two-factor models. We also suggest an implementation procedure that uses a basic spreadsheet to calibrate the model, which considerably simplifies the application of the approach. Even though three-factor models seem to be necessary to explain day-to-day variations in commodity futures term structures, there are very few examples of this type of models in the literature. Schwartz (1997) presents a three-factor model, but its third factor is calibrated using bond prices instead of commodity futures prices. Cortazar and Schwartz (1994) also develop a three-factor model but use a no-arbitrage approach more in the spirit of Heath et al. (1992). The three-factor model proposed in this paper is related to Schwartz (1997) but all three factors are calibrated using only commodity prices.

We also show a model implementation that can be seen as a simpler alternative to the Kalman filtering estimation procedure typically proposed in the literature. Even though this procedure exhibits strong and desirable econometric properties, it places rather high implementation requirements. In typical implementations of the Kalman

filter procedure (Schwartz, 1997) missing data problems are so severe that many market transactions are either discarded or aggregated with others of close but different maturities, with great information loss. Also, it becomes increasingly more complex to include contracts with nonlinear pricing expressions such as options in the estimation.

Our proposed alternative procedure handles very easily the above restrictions: it makes full use of all commodity-linked market asset prices and is very simple to implement with a spreadsheet. This could greatly expand the use of this type of models by practitioners.

This paper is organized as follows. The model is developed in Section 2. Section 3 explains the proposed estimation procedure. Section 4 implements the approach using oil futures prices from 1991 to 2001. Finally, Section 5 concludes.

2. The Model

Given that our proposed three-factor model can be seen as an extension of Schwartz's (1997) two-factor model originally proposed in Gibson and Schwartz (1990), we start by discussing the latter.

2.1 The Two-Factor Schwartz (1997) Model

Defining S as the spot price of oil and δ as the instantaneous convenience yield, Schwartz (1997) proposes the following oil price dynamics:

$$dS = (\mu - \delta)Sdt + \sigma_1 Sdz_1 \tag{1.}$$

$$d\delta = \kappa(\alpha - \delta)dt + \sigma_2 dz_2 \tag{2.}$$

with

$$dz_1 \text{ and } dz_2 \sim N(0, dt^{1/2}) \tag{3.}$$

$$dz_1 dz_2 = \rho dt \tag{4.}$$

In this model, μ represents the long-term total return on oil (price appreciation plus convenience yield), κ the mean reverting coefficient and α the long-term convenience yield. σ_1 and σ_2 are the volatilities of the spot price and the convenience yield dynamics, respectively. ρ is the correlation between the stochastic processes.

For purposes of valuation we use the risk-neutral or risk-adjusted stochastic processes. To obtain the risk-adjusted processes for the spot and the convenience yield, Schwartz (1997) substitutes μ by the risk free interest rate r in the former, and deducts the market price of convenience yield risk λ from the expected drift in the latter:

$$dS = (r - \delta)Sdt + \sigma_1 Sdz_1^* \qquad (5.)$$

$$d\delta = [\kappa(\alpha - \delta) - \lambda]dt + \sigma_2 dz_2^* \qquad (6.)$$

$$dz_1^* dz_2^* = \rho dt \qquad (7.)$$

Using standard arguments it is easy to show that futures prices must satisfy the following partial differential equation:

$$\frac{1}{2}\sigma_1^2 S^2 F_{SS} + \sigma_1\sigma_2\rho S F_{S\delta} + \frac{1}{2}\sigma_2^2 F_{\delta\delta} + (r-\delta)S F_S + [\kappa(\alpha-\delta) - \lambda]F_\delta - F_T = 0 \quad (8.)$$

subject to the boundary condition:

$$F(S, \delta, T=0) = S \qquad (9.)$$

As shown in Jamshidian and Fein (1990) and Bjerksund (1991), the solution to the above equation is:

$$F(S,\delta,T) = S \cdot \exp\left[-\delta\frac{1-e^{-\kappa T}}{\kappa} + \left(r - \hat{\alpha} + \frac{1}{2}\frac{\sigma_2^2}{\kappa^2} - \frac{\sigma_1\sigma_2\rho}{\kappa}\right)T + \frac{1}{4}\sigma_2^2\frac{1-e^{-2\kappa T}}{\kappa^3} + \left(\hat{\alpha}\kappa + \sigma_1\sigma_2\rho - \frac{\sigma_2^2}{\kappa}\right)\frac{1-e^{-\kappa T}}{\kappa^2}\right] (10.)$$

$$\hat{\alpha} = \alpha - \frac{\lambda}{\kappa}$$

2.2 A Parsimonious Two-Factor Model

In this section we modify the model described in the previous one to obtain a more -
parsimonious representation of the two-factor model. This modification is essential
for understanding the extension to a three-factor model we develop in the next
section.

A closer look at the Schwartz (1997) model shows that it can be rewritten in a simpler
way with fewer parameters. We can define a new state variable y as the demeaned
convenience yield δ by subtracting from it the long-term convenience yield α.

$$y = \delta - \alpha \tag{11.}$$

We substitute Equation (11) in Equation (1) and define ν as the long-term price return
(price appreciation) on oil obtained by deducting the long-term convenience yield α
from the long term total return μ. That is:

$$\nu = \mu - \alpha \tag{12.}$$

Equations (1) and (2) then become:

$$dS = (\nu - y)Sdt + \sigma_1 Sdz_1 \tag{13.}$$

$$dy = -\kappa y dt + \sigma_2 dz_2 \tag{14.}$$

In this formulation of the model we treat both factors S and y as non-traded state
variables and, therefore, to transform the original processes ((13) and (14)) into the
risk-adjusted processes we assign one risk premium to each process. Defining λ_i as
the risk premium associated with the factors, the risk-adjusted processes are:

$$dS = (\nu - y - \lambda_1)Sdt + \sigma_1 Sdz_1^* \tag{15.}$$

$$dy = (-\kappa y - \lambda_2)dt + \sigma_2 dz_2^* \tag{16.}$$

$$dz_1^* dz_2^* = \rho dt \tag{17.}$$

It must be stressed that this new model formulation, which has one parameter less than Schwartz (1997) has the same explanatory power but it is more parsimonious and is the basis of the three-factor model we develop later. Also with this formulation it becomes clear why Schwartz (1997) did not estimate the risk free interest rate from futures prices, but rather obtained it from bond data.

Another benefit of this new formulation is that it is more intuitive for practitioner's use. Returns are now defined in terms of the long-term price appreciation, which is more in line with industry practice than using the long-term convenience yield concept.

We now turn to the explanatory power issue of the two-factor model. Even though Schwartz (1997) presents reasonable mean square errors when applying the two-factor model to a set of market prices, these averages hide the fact that for some days the model is unable to capture the behavior of market prices. This may be unacceptable for some model applications, such as its use to support trading. An example of a poor fit between market prices and the two-factor model is illustrated in Figure 1 by comparing model and observed market prices for oil futures traded at NYMEX on January 8th, 1999.

The benefits of using the three-factor model described in the next section are illustrated in Figure 2, where we use it to explain the same market prices shown before, finding a remarkably better fit. This performance improvement, which may become important in many cases, is obtained without worsening the fitness for other dates for which the two-factor model behaved reasonably well.

2.3 A Three-Factor Model for Oil Prices.

We now present the proposed three-factor model. We start with our reformulated version of the two-factor Schwartz (1997) model defining as state variables the commodity spot price, S, and the demeaned convenience yield, y. As in the two-factor model, the instantaneous expected spot price return is ν minus y and the expected long-term spot price return is ν. Thus y can also be defined as the deviation of spot price returns from its long-term value. This return deviation y is modeled as reverting with a coefficient k to its long-term value, which is by definition zero.

In our model we consider as a third risk factor the long-term spot price return, ν, allowing it to be stochastic and to mean revert to a long-term average $\bar{\nu}$, with mean reversion coefficient a and volatility σ_3. Recall that Schwartz (97) also defined a three-factor model, but the third factor was the risk free interest rate estimated from bond data. So, in that model futures market prices were used only to estimate the parameters of a two-factor model. In our model, on the other hand, the third factor is more general and may include, but is not limited to, perturbations in the risk free interest rate. In addition, all model parameters are estimated using only futures prices.

The dynamics of the state variables for the proposed three-factor model for oil prices is then:

$$dS = (\nu - y)Sdt + \sigma_1 Sdz_1 \tag{18.}$$

$$dy = -\kappa y dt + \sigma_2 dz_2 \tag{19.}$$

$$d\nu = a(\bar{\nu} - \nu)dt + \sigma_3 dz_3 \tag{20.}$$

with

$$dz_1 dz_2 = \rho_{12} dt \tag{21.}$$

$$dz_1 dz_3 = \rho_{13} dt \tag{22.}$$

$$dz_2 dz_3 = \rho_{23} dt \tag{23.}$$

Defining λ_i as the risk premium for each of the three risk factors, we can write the risk-adjusted processes, used for valuation purposes, as:

$$dS = (\nu - y - \lambda_1)Sdt + \sigma_1 Sdz_1^* \tag{24.}$$

$$dy = (-\kappa y - \lambda_2)dt + \sigma_2 dz_2^* \tag{25.}$$

$$dv = a\left((\bar{v}-v)-\lambda_3\right)dt + \sigma_3 dz_3^*$$
(26.)

$$dz_1^* dz_2^* = \rho_{12}dt$$
(27.)

$$dz_1^* dz_3^* = \rho_{13}dt$$
(28.)

$$dz_2^* dz_3^* = \rho_{23}dt$$
(29.)

Using standard arguments it can be shown that futures prices must satisfy the following partial differential equation:

$$\frac{1}{2}\sigma_1^2 S^2 F_{SS} + \frac{1}{2}\sigma_2^2 F_{yy} + \frac{1}{2}\sigma_3^2 F_{vv} + \sigma_1\sigma_2\rho_{12}SF_{Sy} + \sigma_1\sigma_3\rho_{13}F_{Sv} + \sigma_2\sigma_3\rho_{23}F_{yv} +$$

$$(v-y-\lambda_1)SF_S + (-\kappa y - \lambda_2)F_y + a\left((\bar{v}-v)-\lambda_3\right)F_v - F_T = 0$$
(30.)

subject to the terminal boundary condition:

F(S, y, v, T = 0) = S
(31.)

which gives the following solution for the futures price:

$$F(S,y,v,T) = S\exp\left[\begin{array}{l} -y\dfrac{1-e^{-\kappa T}}{\kappa} + v\dfrac{1-e^{aT}}{a} + -(\lambda_1)T + \dfrac{\lambda_2 - \sigma_1\sigma_2\rho_{12}}{\kappa^2}\left(\kappa T + e^{-\kappa T} - 1\right) \\[2ex] +\dfrac{\sigma_2^2}{4\kappa^3}\left(-e^{-2\kappa T} + 4e^{-\kappa T} + 2\kappa T - 3\right) + \dfrac{a\bar{v} - \lambda_3 + \sigma_1\sigma_3\rho_{13}}{a^2}\left(aT + e^{-aT} - 1\right) \\[2ex] -\dfrac{\sigma_3^2}{4a^3}\left(e^{-2aT} - 4e^{-aT} - 2aT + 3\right) - \dfrac{\sigma_2\sigma_3\rho_{23}}{\kappa^2 a^2(\kappa + a)}\left(\begin{array}{l}\kappa^2 e^{-aT} + \kappa a e^{-aT} + \kappa a^2 T \\ + \kappa a e^{-\kappa T} + a^2 e^{-\kappa T} \\ - \kappa a e^{-(\kappa+a)T} - \kappa^2 \\ - \kappa a - a^2 + \kappa^2 aT\end{array}\right) \end{array}\right]$$
(32.)

To determine the volatility of futures returns implied by this model we apply Ito's lemma and obtain:

$$dF(S, y, v, T) = F_S dS + F_y dy + F_v dv + F_T dT + \frac{1}{2} F_{SS} dS^2 \tag{33.}$$

from where we obtain:

$$\sigma_F^2(T) = E\left[\left(\frac{dF}{F}\right)\right]^2 = \sigma_1^2 + \sigma_2^2 \frac{(1-e^{-\kappa T})^2}{k^2} + \sigma_3^2 \frac{(1-e^{-aT})^2}{a^2} - 2\sigma_1\sigma_2\rho_{12}\frac{(1-e^{-\kappa T})}{k}$$
$$+ 2\sigma_1\sigma_3\rho_{13}\frac{(1-e^{-aT})}{a} - 2\sigma_2\sigma_3\rho_{23}\frac{(1-e^{-aT})(1-e^{-\kappa T})}{ak} \tag{34.}$$

which converges, as T goes to infinity, to:

$$\sigma_F^2(T \to \infty) = \sigma_1^2 + \frac{\sigma_2^2}{k^2} + \frac{\sigma_3^2}{a^2} - \frac{2\sigma_1\sigma_2\rho_{12}}{k} + \frac{2\sigma_1\sigma_3\rho_{13}}{a} - \frac{2\sigma_2\sigma_3\rho_{23}}{ak} \tag{35.}$$

Equations (34) and (35) imply a volatility term structure decreasing with maturity which converges to a positive constant. This is consistent with a mean reverting non-stationary process.

3. Model Implementation

In this section we suggest a new way of estimating the parameters of the model that has many advantages relative to the more complicated methods employed in the literature. We are looking for an estimation procedure that is both accurate and simple enough to facilitate the task of updating parameter estimates on a regular basis as new market information arrives.

Standard commodity model implementations (Schwartz, 1997, Schwartz and Smith, 2000) make use of rather complex Kalman filtering methods that jointly optimize time series and cross sectional data. The main advantage of this procedure is that, in

addition to the point estimates of the parameters, it provides statistical confidence intervals.

The Kalman filter approach, however, has relatively high implementation requirements. One of the main issues is the handling of missing observations in the data, which is particularly important when using financial instruments, like futures, that only trade for specific maturities. In typical implementations of this procedure contracts with close, but different maturities are grouped together to complete the data sets used in the estimation. The consequence of these groupings is that a substantial amount of data is discarded with its associated information loss. This problem is particularly acute when new contracts are introduced in the market and there are no complete time series data for a given maturity. For example, oil markets currently trade futures contracts for over thirty-five different maturities while only a few years ago they only traded futures contracts for less than twenty different maturities.

Current implementations of the Kalman filter methodology for this type of models require linearity in the pricing expressions, which restricts its applications to linear payoffs such as futures. Nonlinear Kalman filter estimations do exist in the literature, but are much harder to apply. Even though in the current implementation of our model we only use futures prices in the estimation, it would be easy to include contracts with nonlinear payoffs such as options.

Our proposed estimation procedure makes use of all available data while keeping a fairly intuitive and simple approach. We implement our estimation procedure in a simple general-purpose spreadsheet using the standard Solver of Excel.

We now briefly describe the proposed implementation procedure. Assume we want to calibrate the model using futures prices for N dates t_i, with $i = 1,.....,N,$ and that for each date t_i there are M_i contracts with different maturities. Conceptually the procedure works as follows: for a given initial set of parameters we use the cross section of futures prices to estimate the state variables for that particular day. Then we use the estimated state variables for the whole sample period and the full cross section and time series of observed futures prices to estimate a new set of parameter values. With this new set of parameters we repeat the procedure until convergence.

More concretely, for a given set of model parameters, Ω, we can estimate the state variables for each day i, as the solution to the following minimization problem:

$$\{S(t_i), y(t_i), V(t_i)\} \in \underset{S(t_i), y(t_i), V(t_i)}{\operatorname{argmin}} \sum_{j=1}^{M_i} \left(Ln\hat{F}_{ij}\left(S(t_i), y(t_i), V(t_i), T_j - t_i; \Omega\right) - LnF_{ij} \right)^2 \quad (36.)$$

$$\text{for } i = 1, 2, ... N$$

with

$$\Omega : \{\kappa, a, \bar{V}, \lambda_1, \lambda_2, \lambda_3, \sigma_1, \sigma_2, \sigma_3, \rho_{12}, \rho_{13}, \rho_{23}\}$$

where $\hat{F}_{ij}$ is the model price and F_{ij} the observed market price at date i for maturity j.

Since the logarithm of the futures model price is linear in the state variables, Equation (36) implies that the problem of estimating the unobserved state variables is equivalent to performing a simple linear regression. We can therefore obtain estimates for the unobserved state variables, given the cross section of observed futures prices and a set of model parameters, using least squares regression. This estimation methodology of the unobserved state variables uses all price information available each day.

To find the optimal set of parameter values, we solve the following minimization problem:

$$\underset{\{\Omega\}}{\min} \sum_{i=1}^{N} \sum_{j=1}^{M_i} \left(Ln\hat{F}_{ij}\left(S(t_i), y(t_i), V(t_i), T_j - t_i; \Omega\right) - LnF_{ij} \right)^2$$

$$s.t. \qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad (37.)$$

$$\{S(t_i), y(t_i), V(t_i)\} \in \underset{S(t_i), y(t_i), V(t_i)}{\operatorname{argmin}} \sum_{j=1}^{M_i} \left(Ln\hat{F}_{ij}\left(S(t_i), y(t_i), V(t_i), T_j - t_i; \Omega\right) - LnF_{ij} \right)^2$$

$$\text{for } i = 1, 2, ... N$$

and to volatility, correlation and risk premiums consistent with the time series properties of the state variables.

In practice, given that some of the parameters can be estimated directly from the time series of the state variables, we can further simplify the estimation by writing the optimization problem using an even smaller number of parameters. The first thing to notice is that the Equation (32) can be rewritten maintaining three (modified) state variables but reducing the number of parameters from twelve to seven.

The state variables are S, and the modified state variables $\hat{y}$ and $\hat{v}$ defined as:

$$\hat{y} = y + \frac{\lambda_2}{\kappa} \tag{38.}$$

$$\hat{v} = v - \lambda_1 + \frac{\lambda_2}{\kappa} \tag{39.}$$

The seven parameters are now:

$$\Omega' : \{\kappa a \beta \sigma_2^2 \sigma_3^2 \sigma_{12} \sigma_{23}\}$$

with (40.)

$$\beta = \frac{a\overline{v}^* - \lambda_3 + \sigma_{13}}{a^2} \tag{41.}$$

$$\overline{v}^* = \overline{v} - \lambda_1 + \frac{\lambda_2}{\kappa} \tag{42.}$$

$$\sigma_{13} = \sigma_1 \sigma_3 \rho_{13} \tag{43.}$$

With these modifications the futures price Equation (32) becomes:

$$F(S, \hat{y}, \hat{v}, T) = S \exp\left[-\hat{y}\frac{1 - e^{-\kappa T}}{\kappa} + \hat{v}\frac{1 - e^{-aT}}{a} + B(T)\right] \tag{44.}$$

$$B(T) = -\frac{\sigma_{12}}{\kappa^2}\left(\kappa T + e^{-\kappa T} - 1\right)$$

$$+ \frac{\sigma_2^2}{4\kappa^3}\left(-e^{-2\kappa T} + 4e^{-\kappa T} + 2\kappa T - 3\right) + \beta\left(aT + e^{-aT} - 1\right) \tag{45.}$$

$$- \frac{\sigma_3^2}{4a^3}\left(e^{-2aT} - 4e^{-aT} - 2aT + 3\right) - \frac{\sigma_{23}}{\kappa^2 a^2 (\kappa+a)}\begin{pmatrix} \kappa^2 e^{-aT} + \kappa a e^{-aT} + \kappa a^2 T \\ + \kappa^2 aT + \kappa a e^{-\kappa T} + a^2 e^{-\kappa T} \\ - \kappa a e^{-(\kappa+a)T} - \kappa^2 \\ - \kappa a - a^2 \end{pmatrix}$$

The optimization problem then reduces to:

$$\min_{\{\Omega'\}} \sum_{i=1}^{N} \sum_{j=1}^{M_i} \left(Ln\hat{F}_{ij}\left(S(t_i), \hat{y}(t_i), \hat{v}(t_i), T_j - t_i; \Omega' \right) - LnF_{ij} \right)^2$$

s.t. $\tag{46.}$

$$\left\{ S(t_i), \hat{y}(t_i), \hat{v}(t_i) \right\} \in \operatorname*{argmin}_{S(t_i), \hat{y}(t_i), \hat{v}(t_i)} \sum_{j=1}^{M_i} \left(Ln\hat{F}_{ij}\left(S(t_i), \hat{y}(t_i), \hat{v}(t_i), T_j - t_i; \Omega' \right) - LnF_{ij} \right)^2$$

$$\text{for } i = 1, 2, ... N$$

and also subject to restrictions such that volatility and correlations are consistent with the time series properties of the state variables.

This modification considerably simplifies the estimation procedure since we only have to estimate seven parameters. In what follows we will be using this modified model in all the calculations.

The proposed procedure provides results reasonably close to those of the more formal Kalman filter estimation. For example, applying both estimation procedures to the weekly oil futures prices traded at NYMEX between 1990 and 1995 (similar to Panel B in Table 1 of Schwartz (1997)) we find very similar mean square errors (MSE). Figure 3 shows that the estimated values for the state variables are almost indistinguishable between the two procedures. Figure 4 shows that the percentage differences between the estimated values, using both procedures, for the six and twelve-month futures prices are very small and generally smaller than 0.5%.

Even though this method has the disadvantage of not providing distributions for parameter estimates, its simplicity, accuracy, flexibility and easier use of the complete data set, makes it a valid alternative for many applications.

4. Results

Having defined the model and the calibration procedure, we now present the results of applying it to recent oil futures price data. In this section we first define the data used, then we analyze in-sample model behavior, to finally show the results of applying the model to out-of-sample data.

The data used are 76,121 daily prices corresponding to all futures contracts traded at the NYMEX from January 2nd 1991 to December 30th 2001. We do not aggregate data but use each available transaction separately. The number of different contracts traded each day at NYMEX has been increasing over the years, starting with 22 different contracts with a maximum maturity of 3 years, to end the sample period with 35 different contracts with maturity from 1 month to 7 years. Table 1 describes the futures data used.

To test the model we split the data into two sets. Data-set 1 includes only transactions from January 2nd 1991 to December 30th 1998. We will use this data set to make our initial model calibration and to analyze the in-sample properties of our model. Data-set 2 includes the whole set from January 2nd 1991 to December 30th 2001, where the last three years will be used later to make the out-of-sample analysis of the model.

Our in-sample results using Data-set 1 are reported in Figures 5 to 7. Figure 5 shows the dollar root mean square error (RMSE) between model and market prices across maturities. It can be seen that with the exceptions of the shortest and the longest maturity, the model estimates a futures price with a RMSE in the range of $0.05 - $0.15.

Figure 6 presents the RMSE analysis in percentage terms across maturities. The average percentage RMSE is 0.42% and again it can be seen that the worse fit is for the shortest and the longest maturity contracts. All other contacts have a model RMSE between 0.25 and 0.75%.

15

Figure 7 presents another measure of model goodness-of-fit by comparing the implied model volatility for different maturities with historical futures return volatility. Even though the model seems to exhibit a slight upward biased volatility for long-term contracts, it is remarkable how closely it tracks historical volatility for contracts with maturities up to 3 years.

To make our out-of-sample analysis we use Data-set 2 and analyze the performance of the model over the last three years (1999 to 2001). Our initial model calibration, for transactions in the first quarter of 1999, is done with Data-set 1. The model is then recalibrated on a quarterly basis (on the last day of March, June, September and December of each year) and parameter values are not updated for the following three months. This out-of-sample analysis is done using only past information, available at the time of the calibration and represents an upper bound for the errors. More frequent calibrations could reduce model errors, since more recent prices would be included.

Results for the three-year out-of-sample data are now presented. We start by noting that the model is very flexible in that it can deal with very different shapes of the term structure of futures prices. To illustrate this point we select arbitrary dates in the sample period in which futures oil prices exhibit either backwardation or contango. Figure 8 is representative of a day with a high degree of backwardation, while Figure 2 is an example of a day with prices that exhibit a strong contango. These figures are illustrations of how well the model may be used to explain very different term structures for oil futures prices.

Table 2 provides the estimated quarterly parameter values from 12/98 to 12/01 for the three-factor model of oil prices. These results are obtained by solving the optimization problem for the parameters described earlier and by computing the volatility and correlation coefficients from the time series of state variables. Note that most of the parameters are quite stable over time. The only ones that are not stable are the risk premiums associated with the factors. These are the same parameters that turn to be statistically insignificant in the Kalman filter estimation (Schwartz, 1997).

Using the parameter values in Table 2 we compute the out-of-sample model errors over the three-year period reported in what follows. Table 3 compares the RMSE (in

dollars and percentage) for both the in-sample and the out-of-sample data sets. As expected, the out-of-sample data set induces higher errors than the in-sample one, in this case by a factor of approximately two.

Figures 9 and 10 plot the average errors and the one standard deviation interval of the futures out-of sample errors between 1999 and 2001 for different maturities. It can be seen that for most maturities the mean error is very close to zero and the standard deviation of the errors is less than 1% (or less than US$0.25). Note that these results were obtained recalibrating the model every quarter; obviously, more frequent parameter recalibration would increase even more the goodness-of-fit of the model.

4. Conclusions

In this article we develop a parsimonious three-factor model of the term structure of futures oil prices, which fits the data extremely well. In addition, and very importantly, we propose an implementation procedure that significantly simplifies the estimation methods proposed in the literature. These factors make the proposed approach ideally suited for practical applications in the valuation and hedging of real and financial oil-contingent claims. The method can be used for other commodities as well; for example, we have also implemented the model using copper futures price data with similar results.

The approach presented in the paper has been used in practice for almost three years to provide an estimate of the term structure of oil futures prices to an oil company. It is also currently used by the website www.riskamerica.com to provide daily estimates of the oil and copper futures curves.

Acknowledgments

References

BJERKSUND, P. (1991) "Contingent claims evaluation when the convenience yield is stochastic: Analytical results", Working Paper, Norwegian School of Economics and Business Administration.

BRENNAN, M.J. (1991) "The price of convenience and the valuation of commodity contingent claims", In D. Lund y B. Øksendal (eds.) Stochastic models and option values, Elsevier, North Holland.

BRENNAN, M.J. y SCHWARTZ, E.S. (1985) "Evaluating natural resources investments", Journal of Business, Vol. 58, N° 2, 135-157.

CORTAZAR, G. y SCHWARTZ, E.S. (1994) "The valuation of commodity-contingent claims", The Journal of Derivatives, Vol. 1, 27-35.

CORTAZAR, G. y SCHWARTZ, E.S. (1997) "Implementing a real option model for valuing an undeveloped oil field", International Transactions in Operational Research, Vol. 4, N° 2, 125-137.

CORTAZAR, G., SCHWARTZ, E., RIERA F. (2000) "Market-based Forecasts of Commodity Prices using Futures", 4th Annual Conference Real Options: Theory Meets Practice International, Cambridge University/ROG Cambridge, July 7-8.

CULP, C.L. y MILLER, M.H. (1994) "Hedging a flow of commodity deliveries with futures: Lessons from Metallgesellschaft", Derivatives Quarterly, Vol. 1, N° 1, 7-15.

GIBSON, R. y SCHWARTZ, E.S. (1990) "Stochastic convenience yield and the pricing of oil contingent claims", The Journal of Finance, Vol. 45, N° 3, 959-976.

HEATH, D., JARROW, R. y MORTON, A. (1992) "Bond pricing and the term structure of interest rates: A new methodology for contingent claims valuation", Econometrica, Vol. 60, N° 4, 77-105.

JAMSHIDIAN, F. y FEIN, M. (1990) "Closed form solutions for oil futures and European options in the Gibson Schwartz model: A comment", Working Paper, Merrill Lynch Capital Markets.

LAUGHTON, D.G. y JACOBY, H.D. (1993) "Reversion, timing options, and long-term decision-making", Financial Management, Vol. 22, N° 3, 225-240.

LAUGHTON, D.G. y JACOBY, H.D. (1995) "The effects of reversion on commodity projects of different length", In L. Trigeorgis (ed.) Real options in capital investments: Models, strategies, and applications, Praeger Publisher, Westport, Conn.

PADDOCK, J., SIEGEL, D. y SMITH, J. (1988) "Option valuation of claims on real assets: The case of offshore petroleum leases", Quarterly Journal of Economics, Vol. 53, N° 3, 479-508.

SCHWARTZ, E.S. (1997) "The stochastic behavior of commodity prices: Implications for valuation and hedging", The Journal of Finance, Vol. 52, N° 3, 923-973.

SCHWARTZ, E.S. y SMITH, J.E. (2000) "Short-term variations and long-term dynamics in commodity prices", Management Science, Vol. 46, 893-911.

Table 1
Oil Data
Daily Observations from 01/02/1991 to 12/30/2001

Futures Contract	Mean Price (Standard Error)	Mean Maturity (Standard Error)	Number of Observations
F1	US$ 20.78 (4.73)	0.11 (0.02) years	2736
F2	20.63 (4.48)	0.19 (0.02)	2738
F3	20.49 (4.27)	0.27 (0.02)	2737
F4	20.36 (4.07)	0.36 (0.02)	2738
F5	20.25 (3.89)	0.44 (0.02)	2738
F6	20.15 (3.72)	0.52 (0.02)	2737
F7	20.05 (3.56)	0.61 (0.02)	2738
F8	19.96 (3.42)	0.69 (0.02)	2735
F9	19.88 (3.28)	0.77 (0.02)	2734
F10	19.81 (3.16)	0.86 (0.02)	2722
F11	19.74 (3.04)	0.94 (0.02)	2726
F12	19.68 (2.94)	1.03 (0.02)	2725
F13	19.63 (2.83)	1.11 (0.02)	2716
F14	19.58 (2.73)	1.19 (0.02)	2713
F15	19.55 (2.64)	1.28 (0.02)	2702
F16	19.50 (2.55)	1.36 (0.02)	2703
F17	19.48 (2.47)	1.44 (0.02)	2697
F18	19.47 (2.44)	1.52 (0.02)	2559
F19	19.53 (2.63)	1.61 (0.02)	1923
F20	19.52 (2.57)	1.69 (0.02)	1890
F21	19.48 (2.51)	1.77 (0.02)	1891
F22	19.44 (2.44)	1.86 (0.02)	1915
F23	19.47 (2.37)	1.94 (0.02)	1872
F24	19.46 (2.32)	2.02 (0.02)	1820
F25	19.51 (2.31)	2.11 (0.02)	1656
F26	19.50 (2.22)	2.19 (0.02)	1644
F27	19.50 (2.17)	2.28 (0.02)	1617
F28	19.49 (2.08)	2.36 (0.02)	1595
F29	19.56 (2.04)	2.44 (0.02)	1091
F30	19.54 (1.93)	2.52 (0.02)	631
F36	19.38 (1.71)	2.81 (0.14)	2592
F48	19.14 (1.76)	3.54 (0.28)	1575
F60	19.33 (1.62)	4.54 (0.28)	1316
F72	19.41 (1.50)	5.56 (0.28)	1150
F84	19.63 (1.33)	6.54 (0.28)	1049
		TOTAL	76121

Table 2.:
Quarterly Parameter Values for the three-factor model of oil prices
from 12/98 to 12/01 using data since 01/91

	01/91-12/98	1/91-3/99	1/91-6/99	1/91-9/99	1/91-12/99	1/91-3/00	1/91-6/00	1/91-9/00	1/91-12/00	1/91-03/01	1/91-6/01	1/91-9/01	1/91-12/01
K	1.959	1.913	1.863	1.831	1.841	1.849	1.775	1.682	1.675	1.697	1.733	1.731	1.684
a	0.788	0.685	0.686	0.724	0.707	0.667	0.578	0.487	0.455	0.472	0.491	0.492	0.495
σ_1	0.368	0.370	0.369	0.366	0.368	0.368	0.367	0.367	0.371	0.373	0.373	0.378	0.381
σ_2	0.717	0.677	0.680	0.692	0.696	0.687	0.653	0.627	0.623	0.626	0.629	0.630	0.627
σ_3	0.240	0.204	0.213	0.235	0.230	0.215	0.189	0.170	0.162	0.162	0.162	0.161	0.169
ρ_{12}	0.705	0.727	0.717	0.699	0.698	0.697	0.719	0.737	0.746	0.745	0.746	0.751	0.739
ρ_{23}	0.594	0.554	0.590	0.631	0.621	0.602	0.595	0.603	0.584	0.571	0.553	0.549	0.572
ρ_{13}	-0.050	-0.058	-0.032	-0.012	-0.020	-0.036	-0.001	0.043	0.043	0.033	0.014	0.017	0.022
λ_{12}	0.014	0.059	0.078	0.115	0.131	0.142	0.162	0.160	0.152	0.156	0.152	0.143	0.122
λ_{23}	0.227	0.217	0.193	0.206	0.225	0.177	0.168	0.152	0.190	0.174	0.159	0.153	0.141
λ_{13}	0.062	0.017	0.004	-0.019	-0.022	-0.038	-0.045	-0.033	-0.017	-0.023	-0.027	-0.025	-0.017
	0.042	0.036	0.045	0.042	0.042	0.054	0.051	0.061	0.060	0.064	0.065	0.064	0.065

Table 3.:
RMSE for In-Sample (1991 to 1998) and
Out-of-Sample (1999 to 2001) Data

Contract	In-Sample		Out-of-Sample	
Type	Dollars	Percentage	Dollars	Percentage
F1	0.22	1.03%	0.43	1.61%
F2	0.06	0.30%	0.13	0.51%
F3	0.10	0.52%	0.17	0.67%
F4	0.12	0.61%	0.20	0.80%
F5	0.11	0.55%	0.19	0.75%
F6	0.08	0.43%	0.17	0.68%
F7	0.06	0.33%	0.15	0.60%
F8	0.05	0.26%	0.14	0.58%
F9	0.04	0.23%	0.12	0.49%
F10	0.04	0.25%	0.11	0.47%
F11	0.05	0.27%	0.12	0.52%
F12	0.05	0.28%	0.11	0.50%
F13	0.05	0.28%	0.10	0.46%
F14	0.05	0.28%	0.09	0.43%
F15	0.05	0.28%	0.09	0.44%
F16	0.05	0.27%	0.09	0.44%
F17	0.05	0.26%	0.09	0.43%
F18	0.05	0.26%	0.09	0.43%
F19	0.05	0.28%	0.09	0.44%
F20	0.05	0.28%	0.10	0.45%
F21	0.05	0.27%	0.10	0.47%
F22	0.05	0.26%	0.11	0.51%
F23	0.05	0.26%	0.13	0.57%
F24	0.05	0.27%	0.14	0.63%
F25	0.05	0.28%	0.15	0.67%
F26	0.06	0.30%	0.16	0.70%
F27	0.06	0.33%	0.16	0.74%
F28	0.07	0.38%	0.17	0.76%
F29	0.07	0.36%	0.17	0.77%
F30	0.06	0.33%	0.16	0.74%
F36	0.10	0.53%	0.14	0.68%
F48	0.13	0.68%	0.16	0.77%
F60	0.14	0.72%	0.18	0.87%
F72	0.13	0.69%	0.25	1.30%
F84	0.23	1.18%	0.44	2.21%
All	0.083	0.425%	0.172	0.763%
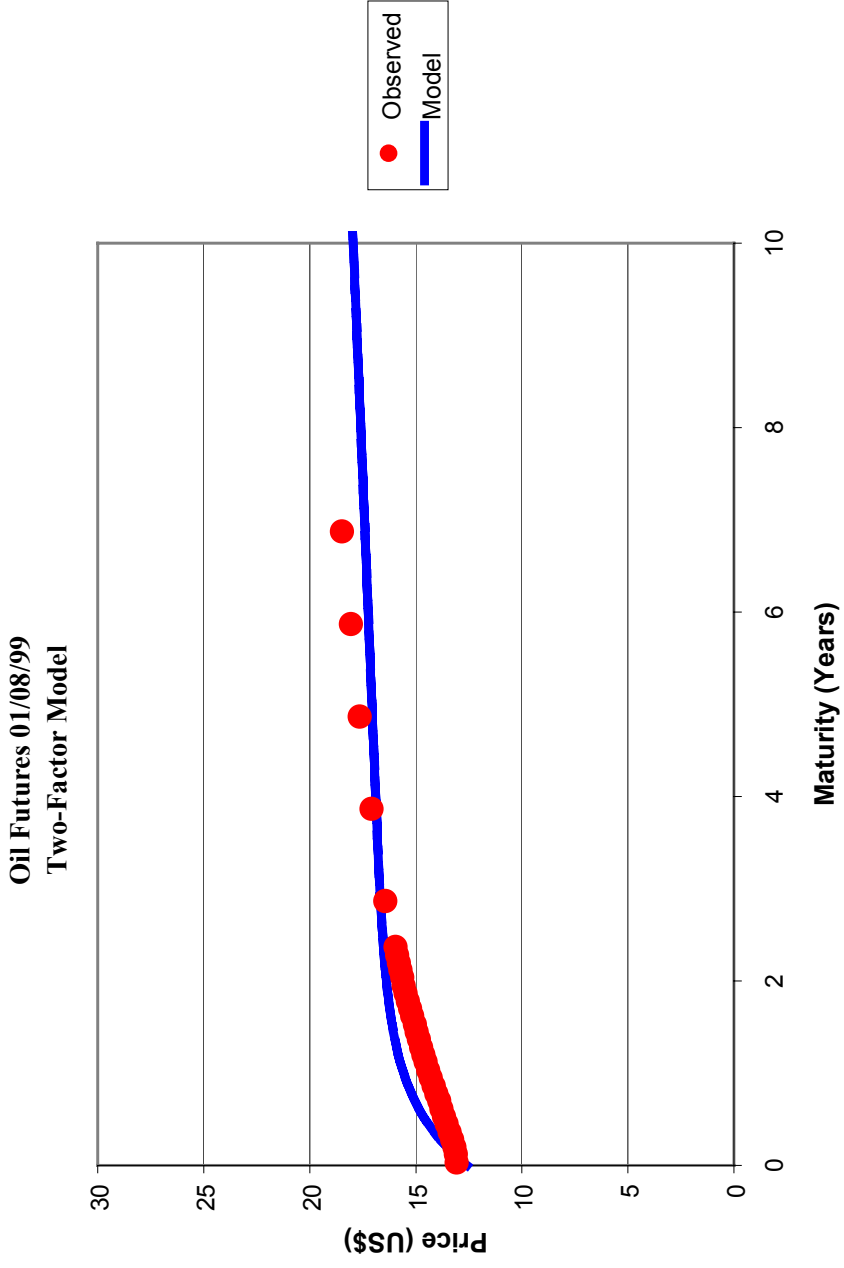
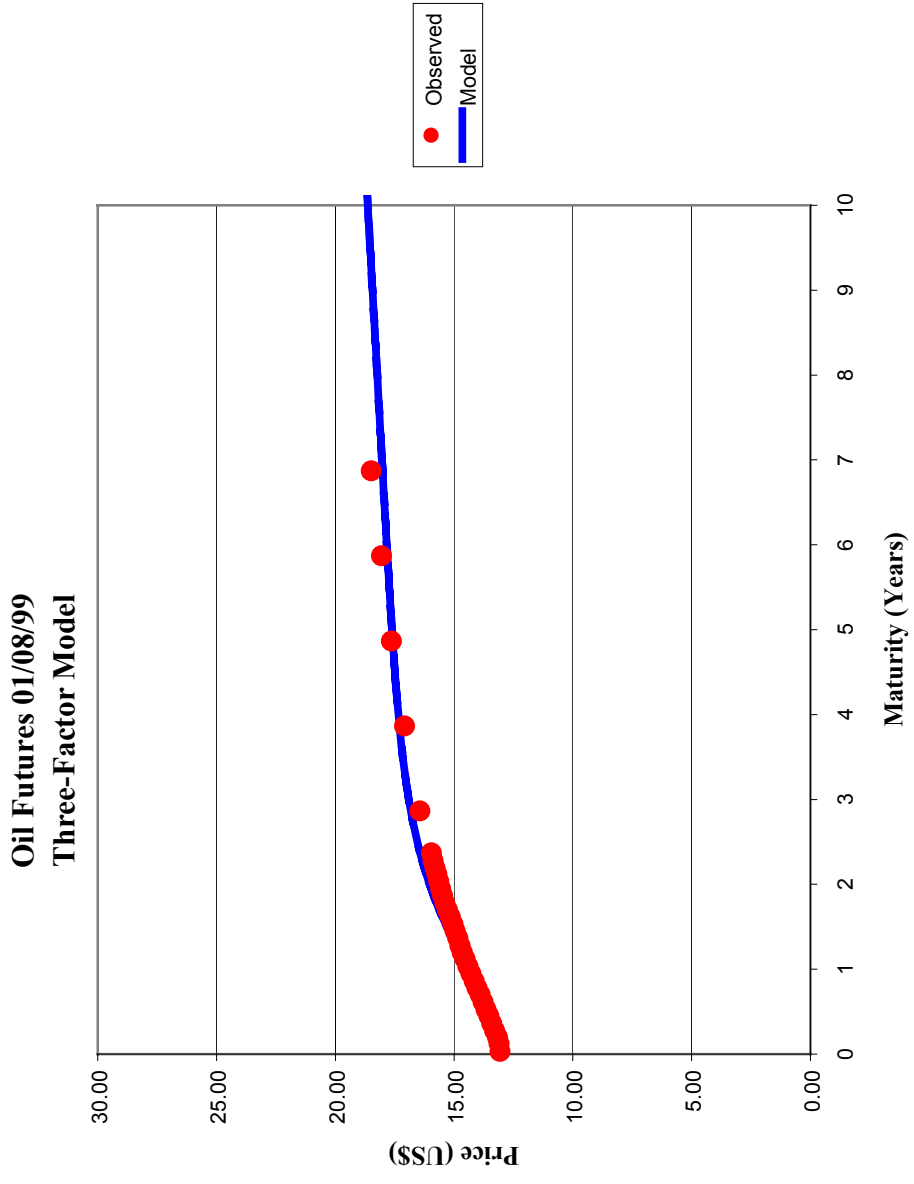Figure 1: Oil Futures 01/08/99 using a Two-Factor Model

23

Figure 2: Oil Futures 01/08/99 using a Three-Factor Model

State Variables Values using Kalman Filter and SpreadSheet Optimization Oil Data 1990-1995

Legend:
- Ln S (SSheet)
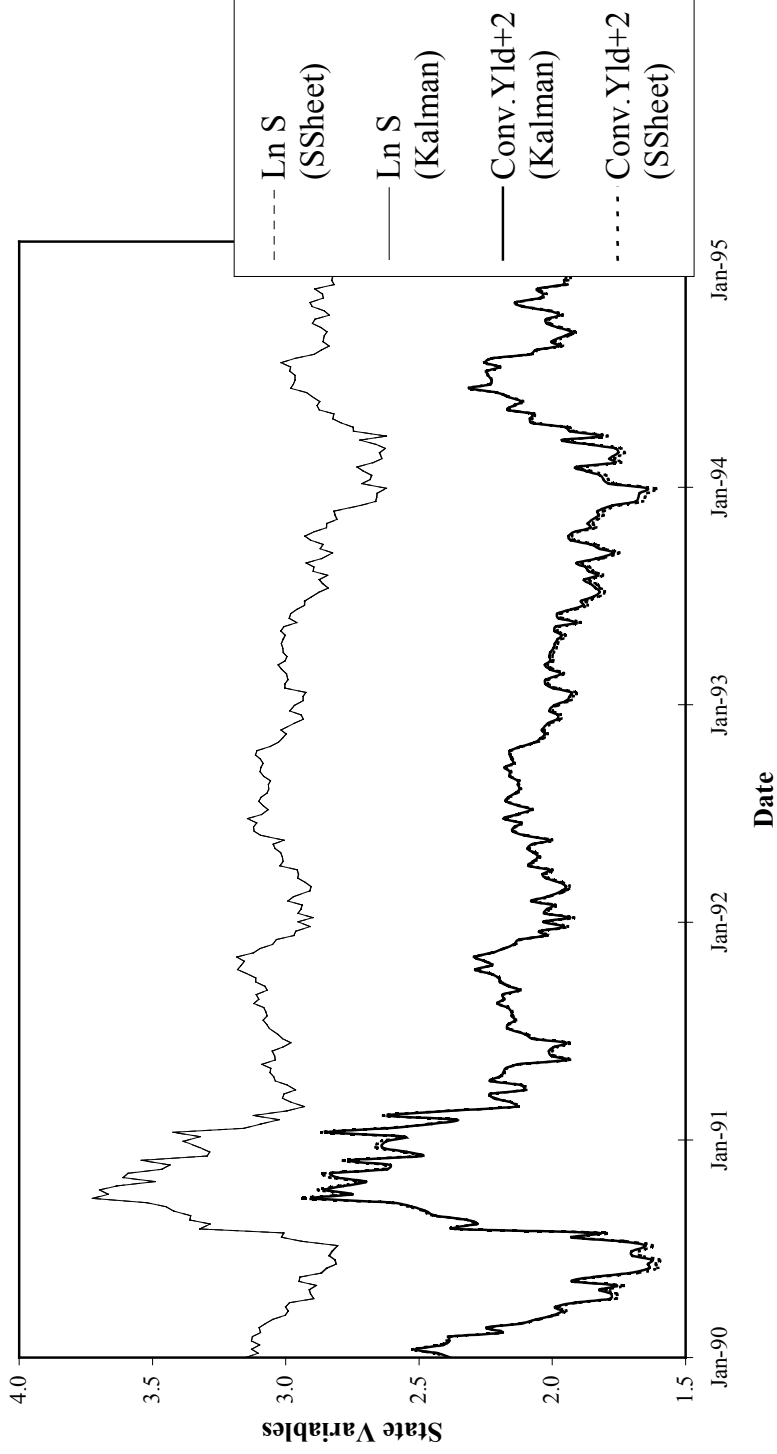- Ln S (Kalman)
- Conv.Yld+2 (Kalman)
- Conv.Yld+2 (SSheet)

Figure 3 . State variables values of the two-factor model in Schwartz (1997) (Log Spot and Convenience Yields) using Kalman Filter and our spreadsheet optimization procedure

25

**Futures Price Difference using Kalman Filter and SpreadSheet Optimization
Oil Data 1990-1995**
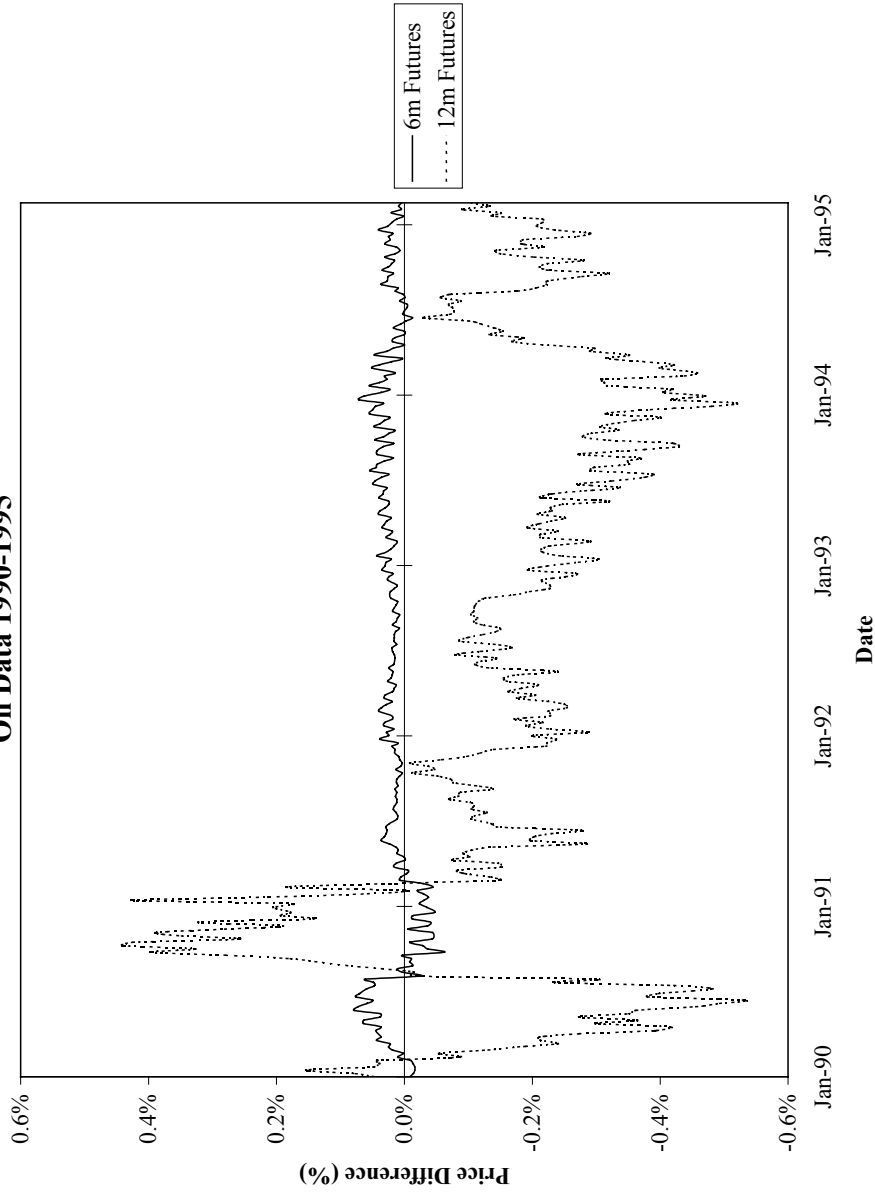
Figure 4 . Percentage differences in 6 month and 12 month futures price estimation using Kalman Filter (Schwartz, 1997) and the Spreadsheet optimization procedure proposed in this paper

**Model in-sample RMSE in Dollars
Oil Futures 1991 to 1998**
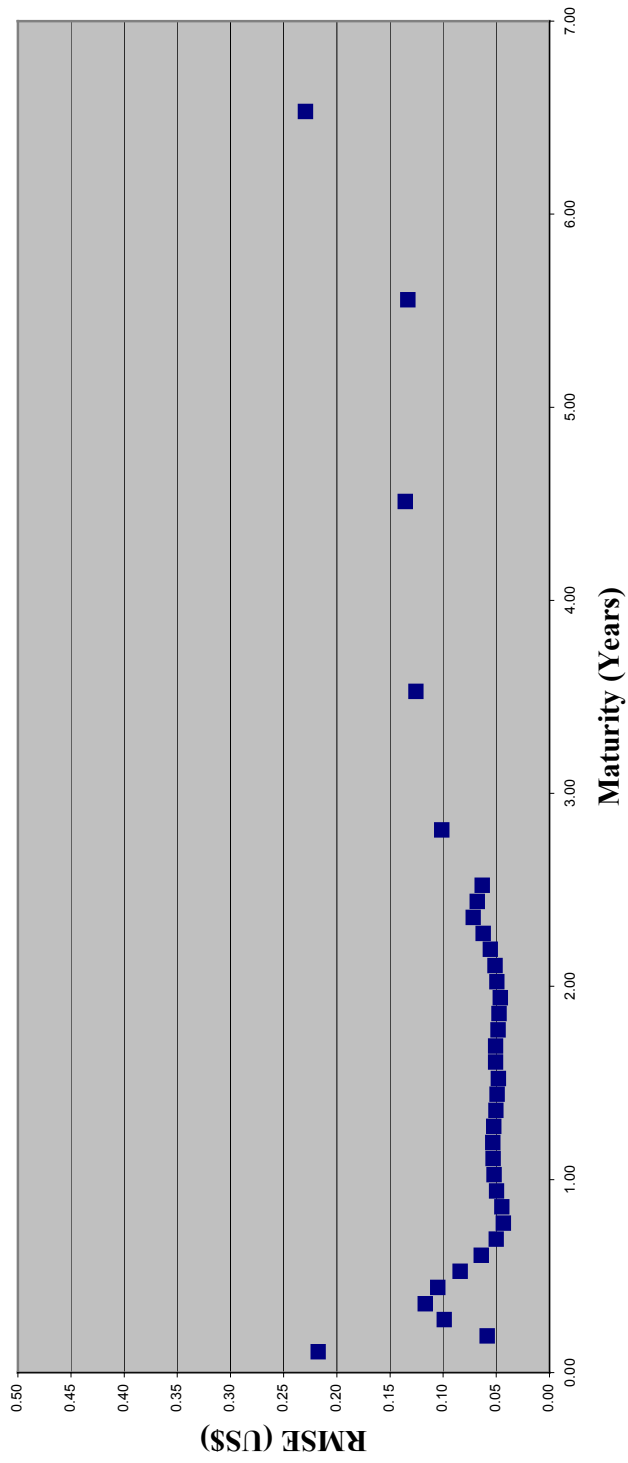
Figure 5: In-Sample Dollar Root Mean Square Errors in Oil Futures 01/91 to 12/98

27

Model in-sample RMSE in Percentages
Oil Futures 1991 to 1998

Maturity (Years)

RMSE (%)
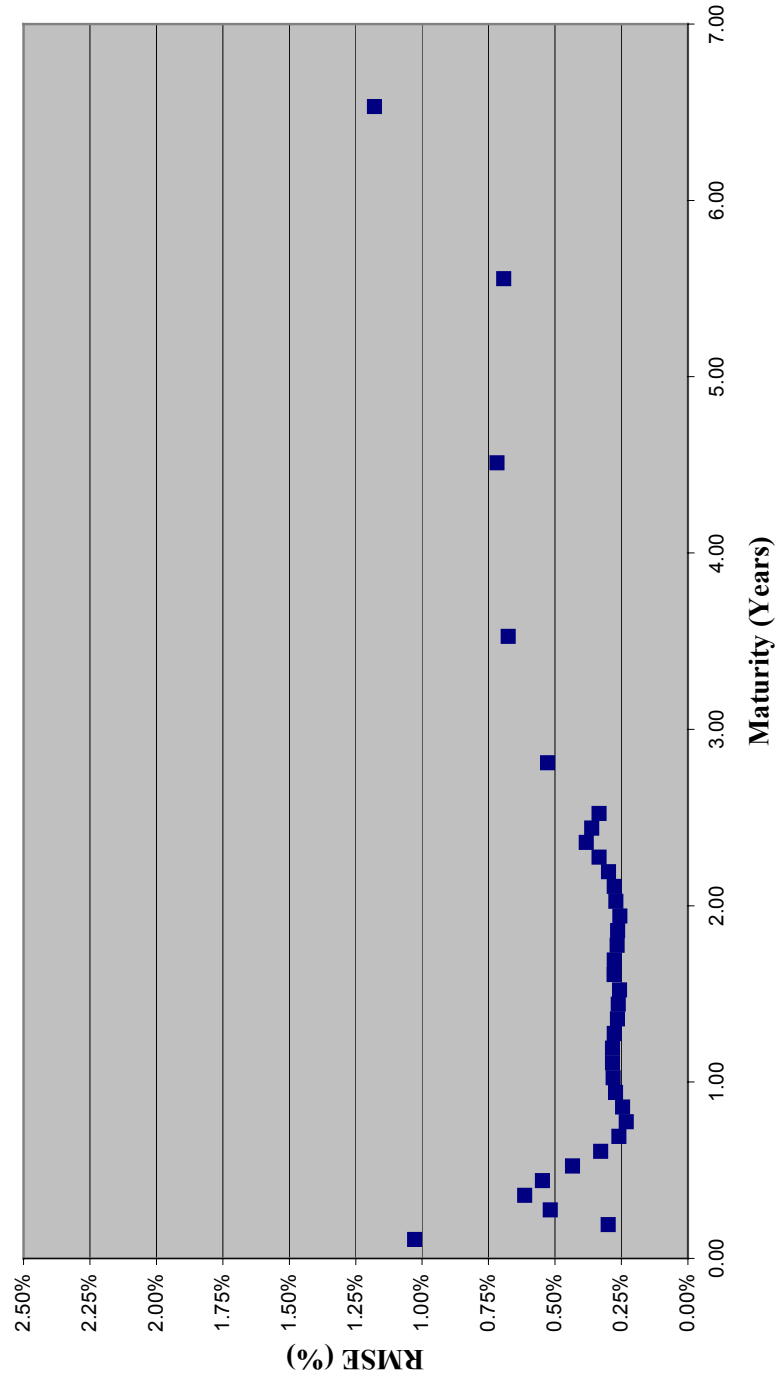
Figure 6: In-Sample Percentage Root Mean Square Errors in Oil Futures 01/91 to 12/98

Volatility of Oil Futures Returns 1991 to 1998
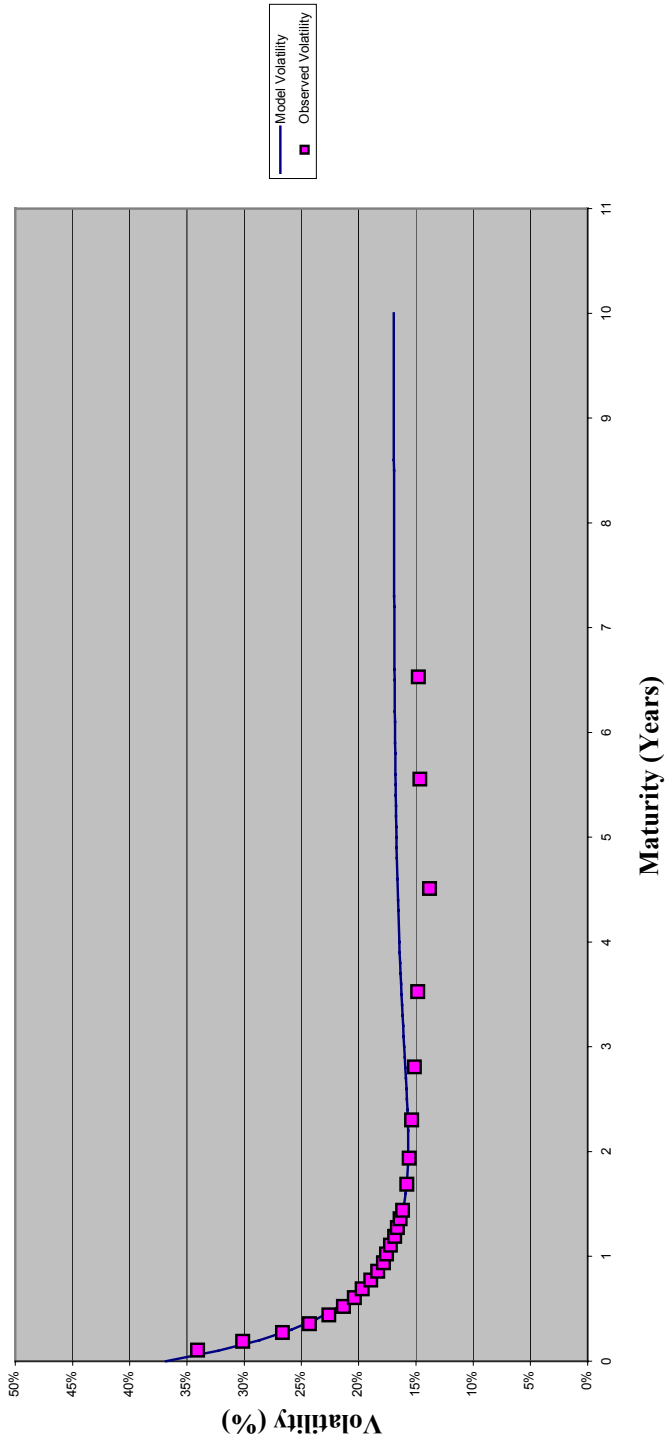
Figure 7: Model and Historical Volatility of Oil Futures Returns 01/91 to 12/98
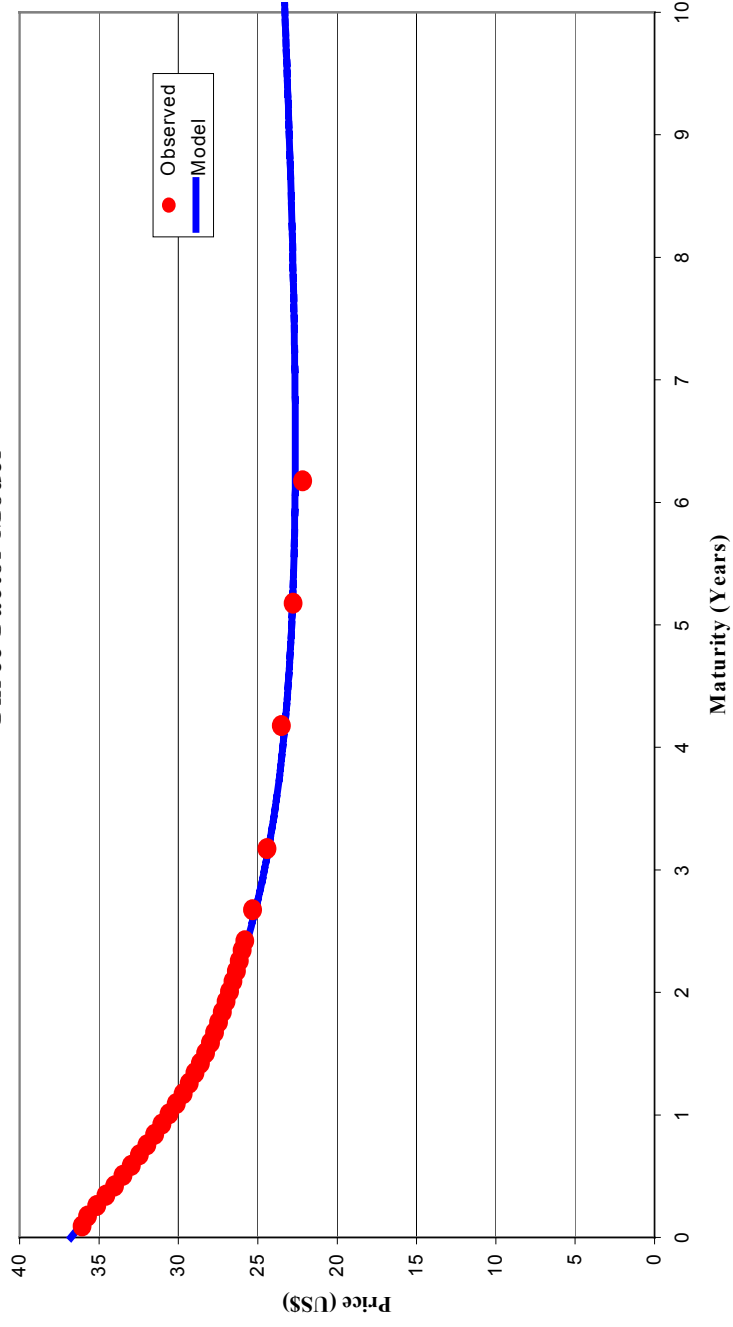
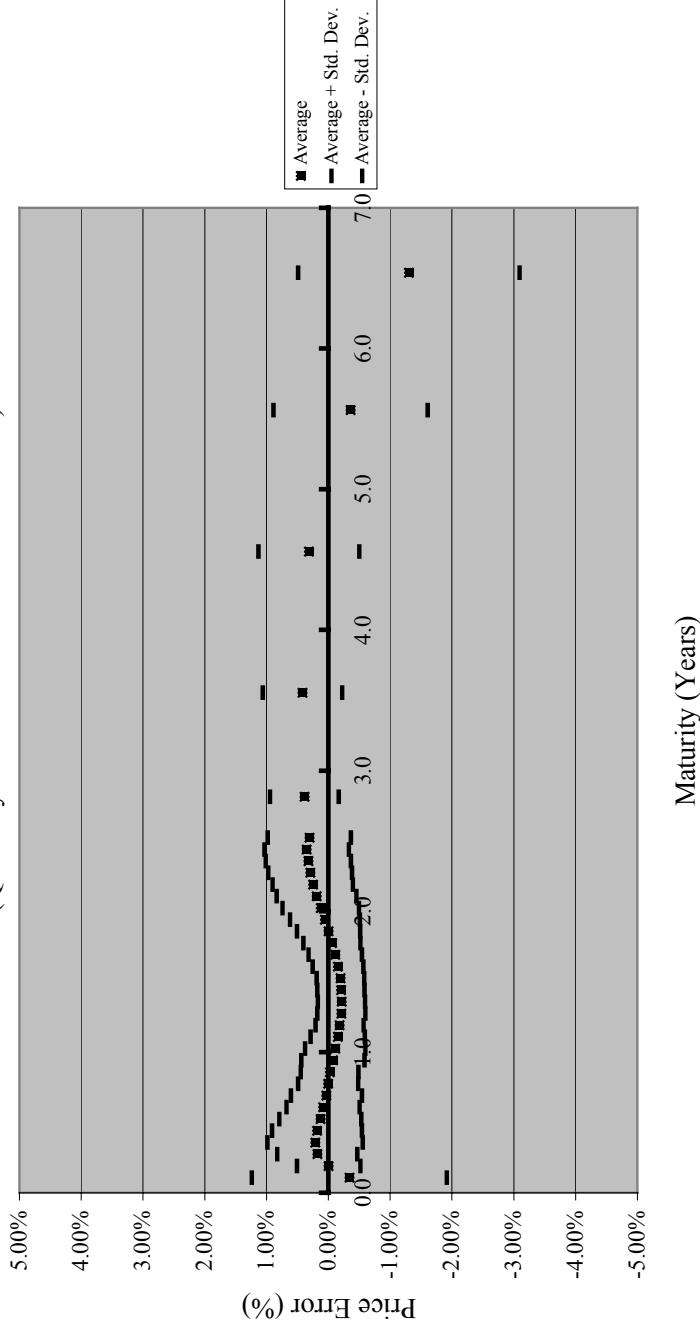Figure 8: Oil Futures Market and Model Prices for 10/12/00

Figure 9 Percentage Out-of-sample Errors in Daily Oil Futures from 1999 to 2001.

The graph presents the mean errors and the interval plus and minus one standard deviation.
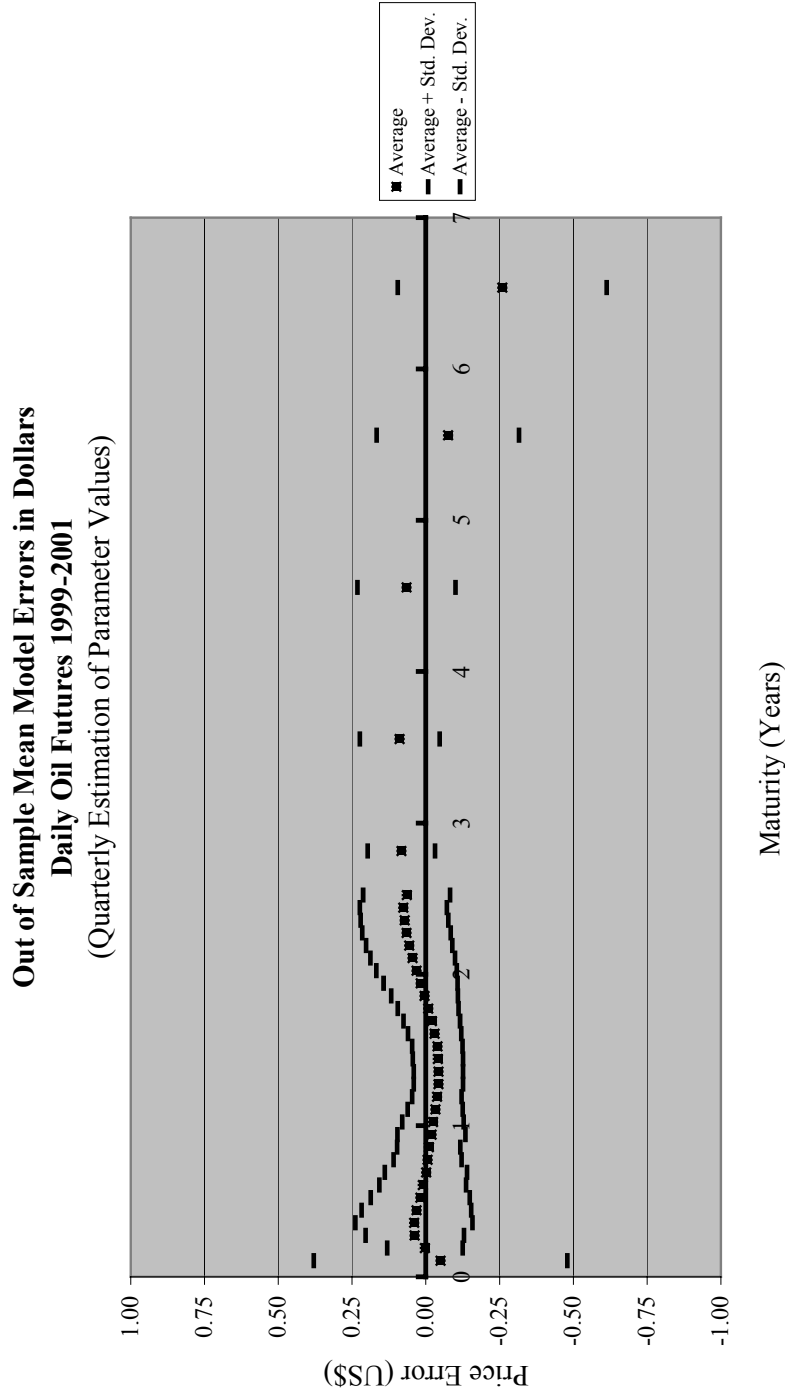
Figure 10 Dollar Out-of-sample Errors in Daily Oil Futures from 1999 to 2001.

The graph presents the mean errors and the interval plus and minus one standard deviation