What do market-calibrated stochastic processes indicate about the long-term price of crude oil?

Warren J. Hahn a,⁎, James A. DiLellio b, James S. Dyer a

a The University of Texas at Austin, 1 University Station, B6000, Austin, TX 78712, USA
b Pepperdine University, 18111 Von Karman Ave, Irvine, CA 92612, USA

A R T I C L E   I N F O

Article history:
Received 28 May 2013
Received in revised form 3 April 2014
Accepted 13 April 2014
Available online 25 April 2014

JEL classification:
C52
C53
Q47

Keywords:
Oil prices
Futures markets
Stochastic processes
Kalman filter
Forecasting

A B S T R A C T

Stochastic process models of commodity prices are important inputs in energy investment evaluation and planning problems. In this paper, we focus on modeling and forecasting the long-term price level, since it is the dominant factor in many such applications. To provide a foundation for our modeling approach we first evaluate the empirical characteristics of crude oil price data from 1990 to 2013 using unit root and variance ratio tests. Statistical evidence from these tests shows only weak support for the applicability of stationary mean-reverting type processes up through 2004, with non-stationary Brownian motion type processes becoming more plausible when the data from 2005 to 2013 is added. We then apply a Kalman filtering method with maximum likelihood approach to estimate the model parameters for both a single-factor Geometric Brownian motion (GBM) process as well as the two-factor Schwartz and Smith (2000) process. The latter process decomposes the spot price into unobservable factors for the long-term equilibrium level and short-term deviation, and it accommodates aspects of both a GBM process and a mean-reverting process. Both empirical and simulated data are analyzed with these models, and we quantify the increases in both the drift rate and volatility of these processes that result from developments in the crude oil markets since the middle of the last decade. We conclude by comparing and contrasting both historical accuracy and forecasts from the parameterized models, and show that when the emphasis is on the long-term expectations, a single factor GBM forecast may be sufficient.

© 2014 Elsevier B.V. All rights reserved.

1. Introduction

Given that crude oil is generally expected to play a significant role in meeting the world’s energy needs for the foreseeable future, forecasting the price of this commodity will be important for planning continuing oil exploration and production investments. Additionally, oil price forecasting can provide insights for determining the potential impact of energy costs on the broader economy and developing appropriate policies for the eventual transition to alternative energy sources. These issues are more likely to depend on the long-term expectations for crude oil price rather than short-term fluctuations; therefore, a primary objective of this paper is to evaluate different approaches for developing long-term forecasts based on the most recent developments in crude oil prices.

There are several different approaches to developing longer-term forecasts for commodity prices, including many types of econometric models, equilibrium models, and expert survey forecasts. In this paper, we use an approach that is based upon calibrating some of the commonly-used stochastic process models with data from the commodities markets. Schwartz (1997), Schwartz and Smith (2000), Manoliu and Tompaidis (2002), and others describe how the parameters for these types of process models can be obtained with the Kalman filter and maximum likelihood estimation, and evaluate the performance of these models for capturing the dynamics of futures prices. We extend this previous work by first expanding the scope of the parameterization study, including an update through the last decade of data that includes several significant developments in the crude oil market. We also conduct back-testing of forecasts at different points in time, and validate the parameterization process through a simulation study. We then use the capability of the Kalman filter to isolate the unobservable long-term component of the crude oil price process, and use this information to develop long-term forecasts.

It is important to distinguish this work from other related empirical research, and to note what we believe are important characteristics of the approach we utilize. First, there are several studies which evaluate the historical performance of futures prices as direct predictors of future spot prices (e.g., Alquist and Kilian, 2010). We instead use futures prices to characterize different stochastic price models, and then use those models to generate forecasts of spot prices, so that the relationship between futures and spot prices is established within the context of a risk-neutral valuation framework. In this approach, futures prices are equal to the expected future spot price under a risk-neutral stochastic process (Duffie, 1992).
There are also other studies which use econometric methods to model oil market structure or construct equilibrium models to capture the relationships between oil price and fundamental factors such as demand, supply, storage and other variables (e.g., Alquist and Killian, 2010; Chevillon and Riffart, 2009; Dees, et. al., 2007; Kaufmann et. al., 2008; Kaufmann and Ullman, 2009; Pindyck, 2001). In addition to testing these fundamental relationships, still other research has the added objective of identifying breaks or changes in the market structure over time (e.g., Fan and Xu, 2011; Kapetanios and Tzavalis, 2010; Krichene, 2002; Miller and Ratti, 2009).

Fundamental supply and demand relationships, as well as new features of the increasingly complex global oil market are present in the approach that we use as well, but they are specified by the consensus of market participants as they set futures prices through their transactions. The aggregation of these factors into futures prices makes their individual effects unobservable, and would seem to limit the usefulness of the price information set. However, the application of Kalman filtering allows us to recover those effects through optimal estimation of the stochastic process parameters, so that we can decompose prices into, for example, short and long term components, where both components are affected by supply, demand and other market factors. The Kalman filter also allows us to deal with changes in the market structure over time as well, since it is a recursive procedure for estimating the state variables at a given time, based on the information available at that time, thereby enabling continuous estimation as new information becomes available (Schwartz, 1997).

Using this approach, we seek to address two basic research objectives. First, based on the current data, we wish to investigate which of the most commonly-used stochastic process forms is most appropriate, and what are the most likely model parameters. In particular, we want to determine whether the typical assumption of stationary oil prices is still valid, in light of the market developments over the past decade. To our knowledge, this is the first work to re-parameterize the forms of stochastic processes that we consider with the Kalman filter using data from this period. Given the results for this first objective, the secondary purpose of this paper is to outline the implications for forecasting oil prices over the longer term. We believe that this work is the only application of these parameterized models for this purpose.

In the next section, we begin with a description and comparison of some of the popular forms of commodity price process models. Section 3 presents an empirical analysis of the historical crude oil price data to evaluate the potential fit with the modeling frameworks discussed in section 2. In section 4, we address our first research objective as we discuss in detail the stochastic process parameter specification and obtain parameter estimates for what appear to be the most appropriate processes for modeling the long-term price of crude oil using both actual futures data and synthetic data obtained through simulation. In section 5 we address our second research objective by presenting our forecasts using the parameter estimates obtained in section 4 and compare and contrast the results. We conclude in section 6 with a summary of findings and implications.

2. Commodity price process models

The most basic models used for commodity prices are simple one-factor stochastic processes. Perhaps the most common of these models is a Geometric Brownian motion (GBM) process, in which commodity prices \( P \) evolve according to the stochastic differential equation

\[
D X = \mu dt + \sigma dZ,
\]

where \( X = \ln(P) \), \( \mu \) is the expected rate of change or drift rate of the process over an increment of time \( dt \), \( \sigma \) is the process volatility, and \( dZ = \sigma dW \) is a random increment of a standard Brownian motion process with \( W \sim \text{Normal}(0,1) \). The GBM model implies that \( \ln(P) \) is normally distributed with mean \( \ln(P_0) + \mu t \) and variance \( \sigma^2 t \), where \( P_0 \) is the price at \( t = 0 \).

The economic assumptions behind the GBM model are that commodity prices are expected to increase over time at a continuous rate, due to inflation and other growth factors, with the variance of prices also increasing in relation to time. A GBM model of prices also implies the assumption that markets are efficient, so that all relevant past price information is impounded in current prices and that future price movements are conditionally independent of past price movements. Under these assumptions, rational investors drive a non-stationary process with the expectation of normally-distributed returns and a lognormal distribution of prices. The GBM process is the most commonly assumed model of prices in the markets for equities and other financial assets that are traded by investors that are generally assumed to have such expectations. A GBM model of prices is simple to implement, flexible to use, and depends on a limited number of parameters. However, some research (e.g., Pindyck and Rubinfeld, 1991) has shown cases when prices were not modeled well with a GBM, and were instead shown to exhibit mean reverting behavior over time.

Mean reverting processes are an alternative type of Markov process where the sign and degree of the drift are dependent on the current level of the variable being modeled, which reverts to a long-term equilibrium level that we typically assume is the long-term mean. The economic assumptions behind this model are that, unlike the GBM model of constant expected growth of commodity prices, prices will tend to increase or decrease depending on the relationship between the price at a given time and the long-term equilibrium price level. In terms of a market hypothesis, the assumption implicit in mean reverting processes is that the price discovery process is one of so-called rational expectations. Under the rational expectations hypothesis, an inverse relationship between spot prices and the slope of the futures price curve indicates that investors expect mean reversion in spot prices, since it implies a lower expected future spot when prices increase and vice-versa. To illustrate the practical reasons behind these expectations, if we suppose that the current price rises above the long-term equilibrium value, investors might expect additional production capacity to be brought on-line and/or use of substitutes to be increased. These activities would result in downward pressure on commodity prices, forcing them back toward the long-term equilibrium level. Conversely, when prices are below the long-term equilibrium price, investors might expect capacity to be reduced and/or commodity use to be shifted away from substitutes causing the price of the commodity to rise.

The simplest form of mean reverting process is the one factor Ornstein–Uhlenbeck process, also called an arithmetic mean reverting process, which has the form

\[
dX_t = \kappa (\bar{X} - X_t) dt + \sigma dz_t.
\]

For commodity price modeling, \( X_t \) is the log of price, \( \kappa \) is the mean reversion coefficient, \( \bar{X} \) is the log of the long term mean price, \( \sigma \) is the process volatility and \( dz_t \) is an increment of a standard Brownian motion process. The log of price is commonly used since it is generally assumed that commodity prices are lognormally distributed. This is convenient, because the price cannot be negative and it also allows future price movements to be modeled based on the stochastic behavior of returns.

The expected value and variance of the Ornstein–Uhlenbeck process are given by Eqs. (1) and (2):

\[
E[X_t] = \bar{X} + (X_0 - \bar{X})e^{-\kappa T}
\]

\[
\text{Var}[X_t] = \frac{\sigma^2}{2\kappa} \left( 1 - e^{-2\kappa T} \right).
\]

Eqs. (1) and (2) show that when \( T \to \infty \), then \( E[X_t] \to \bar{X} \) and \( \text{VAR}[X_t] \to \frac{\sigma^2}{2\kappa} \), as opposed to a GBM where the variance approaches \( \infty \).
there are several variations of this type of model (e.g., Bhattacharya, 1978; Dixit and Pindyck, 1994), and Dias (2005) provides a survey of some of these different processes and their application to oil price modeling.

Several researchers including Schwartz (1997) and Hilliard and Reis (1998) have shown through empirical analyses that simple one-factor models may be inadequate for modeling the complex behavior of commodity prices and therefore may oversimplify their linkages to the different market hypotheses that may coexist among market participants. As a result, models with multiple factors have been developed to simultaneously apply aspects of different single-factor models. Examples of models that utilize multiple factors include Gibson and Schwartz (1990), Schwartz (1997), Schwartz and Smith (2000), and others.

There are two basic types of two-factor models. The first approach, used by Gibson and Schwartz (1990), Schwartz (1997), and Ribeiro and Hodges (2004) is to model price with a GBM as the first factor, and nest within the drift function of the price process a mean-reverting process for convenience yield. Hull and White (1994) also use a variation of this approach in their two-factor model; however they use their fitted mean-reverting formulation as the process for the first factor instead of a GBM. The second approach is to decompose price into factors for the long-term mean, which is specified with a GBM process, and the short-term deviation from the long-term mean, which is modeled as a one-factor mean-reverting process. Schwartz and Smith’s short-term/long-term model is the primary example of this approach. This approach is more computationally convenient for some applications, because the two factors are connected only by the correlation of their increments. It also provides straightforward estimates of long-term equilibrium and short-term deviations, which for our purposes allows a more direct comparison to single-factor models.

Lastly, the short-term and long-term processes allow models to reflect the price discovery process on multiple time scales, reflecting different behaviors of rational participants in efficient markets. For example, the short term process may reflect the process as perceived by active traders in the futures markets, while the long term process can accommodate longer term trends in demand and investments in production capacity.

The basic rationale for the Schwartz and Smith model is to draw on the valid arguments of both primary single-factor models. Prices at a basic level should be expected to grow at a constant rate over time with variance increasing in proportion to time, which is behavior that can be modeled with a GBM. In the short term, however prices will also be affected by supply and demand conditions. Since these effects are expected to be short-lived in efficient markets, they would go away over time, which can be modeled with a process reverting to a mean of zero. The Schwartz and Smith model accommodates both types of behavior by introducing a bifurcation of the time horizon into the short- and long-term processes.

Following the nomenclature of Schwartz and Smith (2000), the long-term equilibrium price and deviation from the equilibrium price at any point are denoted as \( \chi_t \) and \( \chi_t \) respectively. Then, the log of price \( X_t \) is the sum of the two factors:

\[
X_t = e^{\chi_t + \kappa_t},
\]

where the two processes are:

\[
d\chi_t = \mu_t dt + \sigma_t d\zeta_t \quad \text{GBM for long-term mean price}
\]

\[
d\zeta_t = \kappa(0 - \zeta_t) dt + \sigma_t d\zeta_t \quad \text{Mean-reverting process for the deviation}
\]

and the increments of the two processes are correlated:

\[
d\zeta_t d\zeta_t = \rho_{\zeta_t} dt.
\]

This formulation indicates that there are five parameters required to specify this model: \( \mu_t, \sigma_t, \sigma_t, \kappa, \) and \( \rho_{\zeta_t} \).

There are many other types of commodity price models that have been discussed in the literature as well, including models with more than two factors (e.g., Cassassus and Collin-Dufresne, 2005; Schwartz, 1997) and models that include a jump factor. However, the basic types that we have discussed in this section are the most commonly cited models in the literature and are also the most used in practical applications. Consequently, they are the focus of our discussion in the remainder of this paper.

3. Empirical evidence

The decision about which of the basic types of stochastic processes discussed in Section 2 to apply to modeling commodity prices for a particular application is a complicated issue. As Dixit and Pindyck (1994) suggest, in order to select an appropriate stochastic process for modeling commodity price or any other variable, the best approach is to rely on theoretical considerations, such as equilibrium mechanisms, as well as statistical tests. In this section, we provide up-to-date results for crude oil using two of these tests.

The most common statistical test for determining whether a GBM or mean reverting process is most appropriate is the unit root test (Dickey and Fuller, 1981). For example, Pindyck (1999) applies a version of the Dickey–Fuller unit root test to evaluate several oil, coal and natural gas price time series. The null hypothesis for this test posits that a unit root exists, in which case the time series is not stationary; therefore if it can be rejected, then there is support for the alternative claim of mean-reversion in the time series.

We applied the standard Dickey–Fuller test to a time series consisting of the weekly average estimated spot prices for West Texas Intermediate Crude Oil (WTI) during the period from January, 1990 to January, 2013 (Fig. 1). We also investigated a truncated version of the data (January, 1990 to December, 2004) which predated the recent run-up in oil prices, since mean reversion appears to be less plausible after this period (Meade, 2010).

We also applied the augmented Dickey–Fuller test (ADF), which can improve the power of the standard test if a trend exists in the data, and the Phillips–Perron (PP) unit root test, first without a trend component, and then with a trend component. An advantage of the PP tests over the Dickey–Fuller tests is that the PP tests are robust to general forms of heteroskedasticity in the error term. The test statistics for all four of these tests are shown in Table 1, along with their corresponding critical values for the 5% and the 10% significance levels. These tests results are in general agreement, and indicate little support for mean reversion in the data. The only exceptions are the PP tests in the 1990–2004 data set, which indicate only weak support (10% significance level) for mean reversion over that time horizon.

Finally, we ran the KPSS inverse test for stationarity, with the results also shown in Table 1. For this test, we obtained significant test statistics for both data sets, leading to rejection of the null hypothesis that the time series was stationary in both cases. All hypothesis testing was conducted using the Econometrics Toolbox™ within MATLAB.

![Average weekly estimated spot price (WTI)](image-url)
Although these results do not support mean reversion of prices when the recent data is included, a well-known issue with unit root tests is their low power. Their rationale for proving mean reversion, which is developed by ruling out the competing hypothesis (i.e., that a GBM model is appropriate) with a high (e.g., 90% or higher) degree of certainty, is difficult to do in many instances. For example, Pindyck (1999) tested many time series, and only an extremely long time series for oil (96 years, annual average price) indicated conclusively that there was no unit root, thus rejecting the appropriateness of a GBM process. For shorter duration series in that same study, the presence of a unit root could not be rejected, even though the series graphically appear to exhibit mean reversion. For these series, Pindyck (1999) points out that the failure to reject the presence of a unit root does not necessarily prove that a series follows a random walk. Rather, it leaves open the question as to which process is most appropriate. Thus, failure to reject the random walk hypothesis does not strictly preclude the existence of auto-regression (mean reversion) in the variable of interest.

As an alternative, Pindyck (1999) suggests that investigating the extent to which price shocks are permanent might be more informative than looking for a unit root in checking for a random walk or for mean reversion. Under mean-reversion, price shocks tend to dissipate due to the constant force of reversion. This is contrary to the case of a GBM, where price shocks are permanent. Pindyck (1999) utilizes variance ratio tests to test this condition, which measure the extent to which the variance of a series grows with the lag of the test. The variance ratio can be expressed as:

$$ R_k = \frac{1}{k} \times \frac{\text{Var}(P_{t+k} - P_t)}{\text{Var}(P_{t+1} - P_t)} \quad (4) $$

The $\text{Var}(\ldots)$ terms in Eq. (4) represent the variance of the series of lagged (by $k$ periods) differences in the price series $P$. In the case of a GBM, as the variance grows linearly with $k$, the ratio $R_k$ should approach 1 as $k$ increases. On the other hand, under mean reversion, the variance is bounded to a limit as $k$ increases. Therefore the variance ratio above should decay for large values of the lag ($k$), indicating that price shocks are not permanent and that prices are mean reverting.

We applied the above variance ratio test to our series of crude oil prices, with the results shown in Fig. 2 below. As with the series that Pindyck (1999) evaluated, the ratios for the full 1990–2013 series initially grow with lag $k$, which is consistent with both the GBM and mean reversion assumptions. But then they begin to decrease, with the higher lag ratios returning to a value near 1. The truncated 1990–2004 series, however, remains in a generally decreasing pattern below a ratio of 1. This is more supportive of the assumption of a mean reverting process for modeling crude prices during this time horizon.

The implication of these statistical tests is that oil prices may have demonstrated some degree of mean-reversion during the period up until about a decade ago. However, including data since that time produces test results that make non-stationary models of prices more plausible. Our findings confirm and extend what Askari and Krichene (2008) found in their study of oil prices during the 2002–2006 time horizon. We will therefore consider both model types in the following section on model parameterization.

### 4. Parameterizing stochastic process models

The spot price of crude oil for one factor-models and the state variables for multi-factor models (e.g., $\xi_t$ and $\chi_t$ for the Schwartz and Smith model) are unobservable. Therefore, it is necessary to link them in some way to observable information in order to determine the parameters above that define the stochastic process, and we can use futures price data for this purpose. Although futures markets are primarily in existence to provide a way to transfer risks, another benefit of these markets is that they also impart information about commodity prices that can be used to specify a diffusion model.

We will focus our discussion of this approach on the Schwartz and Smith model, since it also includes a single-factor GBM model as a special case. We extend their work by first significantly expanding (from 1990–1996 to 1990–2013) the time horizon for the empirical evaluation of the two-factor model performance, and then also numerically validating the parameterization of the model with the Kalman filter maximum likelihood approach. We will then use these parameterization results to compare and contrast the forecasting properties of the two-factor model versus a single-factor model in the following section.

There are various proposed methods for parameterizing this model, including an implied approach proposed by Schwartz and Smith (2000) and implemented by Jafarizadeh and Bratvold (2012). However, in this work we utilize a Kalman filter-based approach to obtain the parameter estimates because this method is the most commonly used for this type of application, and because it provides us with error statistics that are useful for inference about model fit and forecasting.

#### 4.1. State space formulation for the Kalman filter

Under risk-neutral valuation, the Schwartz and Smith model should be transformed with the addition of two risk-premium parameters, $\lambda_y$ and $\lambda_x$, to adjust the drift of each process and produce a risk-neutral price model. The function of these two parameters, called the short-term deviation risk premium and equilibrium risk premium respectively, is to transform the two processes so that cash flows generated from...
the model can be discounted at the risk-free rate. The resultant formulation for the two-factor process then becomes:

\[ d\kappa_t = (\mu_\kappa - \lambda_\kappa) dt + \sigma_\kappa d\zeta_t \]

\[ dx_t = (-\kappa x_t - \lambda_x) dt + \sigma_x d\zeta_t. \]

Denoting the risk-neutral drift as \( \mu_x \equiv \mu_\kappa - \lambda_x \), the first equation can be written as:

\[ d\kappa_t = \mu_x dt + \sigma_x d\zeta_t. \]

The result is that the log of future spot price is normally distributed with the following revised mean and variance:

\[ E[\ln(x_t)] = e^{-\kappa t} x_0 + \xi_0 - (1 - e^{-\kappa t}) \lambda_x / \kappa + \mu_x t \]

\[ \text{Var}[\ln(x_t)] = (1 - e^{-2\kappa t}) \sigma_x^2 / 2\kappa + \sigma_x^2 T + 2 \left( 1 - \exp^{-\kappa T} \right) \gamma_{\kappa} \sigma_x \sigma_y / \kappa. \]

A complete derivation of these formulas can be found in Schwartz and Smith (2000).

Under risk-neutral valuation, the futures prices will equal the expected spot prices (Black, 1976). Therefore the expectation and variance can be used to derive the following expression for futures prices:

\[ \ln \left( F_{T,0} \right) = e^{-\kappa T} x_0 + \xi_0 + A(T). \]  

Here, \( F_{T,0} \) is the current market price for a futures contract at time to maturity \( T \), and,

\[ A(T) = \mu_x T - (1 - e^{-\kappa T}) \frac{\lambda_x}{\kappa} + 2 \left( 1 - e^{-\kappa T} \right) \gamma_{\kappa} \sigma_x \sigma_y / \kappa. \]  

In this case, there are seven parameters required to specify the model: \( \kappa, \sigma_x, \mu_x, \xi_0, \gamma_{\kappa}, \sigma_y \) and \( \lambda_x \).

The primary method for estimating the parameters in the two-factor model is Kalman filtering with maximum likelihood. The Kalman filter is a recursive procedure for optimally estimating unobserved state variables based on observations that depend on these state variables (Kalmim, 1960). In this case, the Kalman filter can be applied to estimate the unobservable state variables \( \chi_t \) and \( \xi_t \) in the Schwartz and Smith model using Eqs. (5) and (6). It is then possible to calculate the likelihood of a set of observations given a particular set of parameters. By varying the parameters and re-running the Kalman filter, the parameters that maximize the likelihood function can be identified. A detailed description of this technique can be found in Harvey (1989).

For the Kalman filter, the stochastic process must be represented in a state space formulation. This consists of a transition equation to describe the evolution of the state variable or variables over time and a measurement equation to relate the state variable(s) to the observable data. Schwartz and Smith (2000) specify the transition equation for the two-factor model as:

\[ x_t = c + \mathbf{G} x_{t-1} + \omega, \quad t = 1, \ldots, n_T \]

where

\[ x_t = [\chi_t, \xi_t] \] is a 2 \times 1 vector of state variables;
\[ c = [0, \mu_\Delta \kappa \omega] \] is a 2 \times 1 vector;
\[ G = \begin{bmatrix} e^{-\Delta \kappa} & 0 \\ 0 & 1 \end{bmatrix} \] is a 2 \times 2 matrix of state variables;
\[ \omega \] is a 2 \times 1 vector of normally distributed disturbances with \( \text{E}[\omega] = 0 \) and \( \text{Var}(\omega) = \sigma_\omega^2 \); and
\[ W = \text{Cov}(\omega) = \begin{bmatrix} (1 - e^{-2\kappa \Delta}) \gamma_{\kappa}^2 & (1 - e^{-2\kappa \Delta}) \gamma_{\kappa} \sigma_x \\ (1 - e^{-2\kappa \Delta}) \gamma_{\kappa} \sigma_x & \sigma_y^2 \end{bmatrix}. \]

\[ \Delta t \] is the length of time steps, and \( n_T \) is the number of time periods.

The corresponding measurement equation, which is based on Eqs. (5) and (6), is:

\[ y_t = d_t + F_t \kappa_t + v_t, \quad t = 1, \ldots, n_T \]

where,

\[ y_t = [\ln(F_{T_1}), \ldots, \ln(F_{T_n})] \] is a \( n \times 1 \) vector of observed futures prices with maturities \( T_1, T_2, \ldots, T_n \);
\[ d_t = [A(T_1), \ldots, A(T_n)] \] is a \( n \times 1 \) vector;
\[ F_t = [e^{-\kappa T_1}, \ldots, e^{-\kappa T_n}] \] is a \( n \times 2 \) matrix;
\[ v_t \] is a \( n \times 1 \) vector of normally distributed disturbances (measurement errors) with \( \text{E}[v_t] = 0 \) and \( \text{Cov}(v_t) = \Sigma_v \).

For the case of a single-factor GBM process, the transition equation for the spot price \( \xi_t \) is:

\[ x_t = c + x_{t-1} + \omega, \quad t = 1, \ldots, n_T \]

where,

\[ x_t = [\xi_t] \] is a scalar for the state variable (the spot price);
\[ c = [\mu \Delta \kappa \omega] \] is a scalar;
\[ \omega \] is a normally-distributed disturbance with \( \text{E}[\omega] = 0 \) and \( \text{Var}(\omega) = \sigma_\omega^2 \).

The corresponding measurement equation in this case is:

\[ y_t = d_t + f_t x_t + v_t, \quad t = 1, \ldots, n_T \]

where,

\[ y_t = [\ln(F_{T_1}), \ldots, \ln(F_{T_n})] \] is a \( n \times 1 \) vector of observed futures prices with maturities \( T_1, T_2, \ldots, T_n \);
\[ d_t = [A(T_1), \ldots, A(T_n)] \] is a \( n \times 1 \) vector;
\[ f_t = [1, \ldots, 1] \] is a \( n \times 1 \) vector;
\[ v_t \] is a \( n \times 1 \) vector of serially uncorrelated normally-distributed disturbances with \( \text{E}[v_t] = 0 \) and \( \text{Cov}(v_t) = \Sigma_v \).

With a state space formulation and a set of historically observed futures prices for different maturities, the Kalman filter can be run recursively beginning with a prior distribution of the initial values of the state variables \( [\xi_0, \xi_0] \) or \( [\kappa_0, \omega_0] \). For the GBM model with only \( \xi_t \), there are only three parameter models to estimate \( (\xi_0, \mu_\kappa, \mu_\omega) \), as compared to seven for the full two-factor model. In either case, the terms in the covariance matrix \( \Sigma_v \) for the measurement errors for each of the futures contract maturities in the data must also be estimated. The measurement errors can be simplified with the common assumption that they are not correlated with each other, so that \( V \) is a diagonal matrix with elements \( \sigma_v^2 \) as in Schwartz (1997) and Schwartz and Smith (2000). The general approach is to maximize the log-likelihood function for a joint normal distribution:

\[ \ln(L) = -\frac{1}{2} \sum_{t=1}^{n_T} \ln[F_t] - \frac{1}{2} \sum_{t=1}^{n_T} \left( y_t - \hat{y}_{t-1} \right)^T F_t^{-1} \left( y_t - \hat{y}_{t-1} \right) + \text{const}. \]
The results shown above for the first case (1990–1995) indicate good overall agreement with the estimates obtained by Schwartz and Smith (2000). While our results duplicated most of those results, we might also expect some differences due to slight differences in data. We did not have access to their source, Knight-Ridder financial services, instead using data that we obtained from Bloomberg. Given that global optimization of a function of seven variables is a challenging computational problem, slight differences between the estimates could also result from the use of different optimization routines. A Gauss optimization routine was used in the Schwartz and Smith (2000) study, whereas we used a MATLAB routine in this study.

If we consider the confidence intervals around parameter estimates from Schwartz and Smith (2000), only our estimates for short-term volatility and equilibrium volatility are outside the 95% intervals; however these two parameters also have very small standard errors. We also obtained a slightly higher log-likelihood score, but this again may be due to some differences in the optimization routine and data source.

The results from fitting the data over the expanded time horizon from January 1990 to December 2004 indicate two major changes to the parameter estimates, as shown in Table 2. The equilibrium drift rate increases significantly, from near zero to approximately 4%, while the correlation between increments of the long-term equilibrium and short-term deviation decreases significantly and in fact becomes slightly negative. When factoring in the risk-neutral equilibrium drift rates, which do not change substantially, we can infer from the increase in the equilibrium drift rate that the long-term risk premium has also increased significantly. This was likely due to some signs of a price run-up toward the end of the data set (i.e., even prior to 2005), and the uncertainty about the long-term equilibrium level. All of the other parameter estimates are similar to the estimates from the 1990–1995 data set. As expected with the additional data, the standard errors for the estimates are all lower, with the exception of that for the drift rate for the risk-neutral equilibrium process, which remained small. The log-likelihood score is about three times the score for the 1990–1995 data set, which is expected, since it is dependent on the amount of data fitted and this aligns with the improvement in the parameter estimates.

When we add the remaining available data, up through the last week of January 2013, to the data set, we notice some marked changes in the two-factor model parameter estimates. Most notably, we see that the equilibrium drift rate is again increased, to double the value from the 1990–2004 data set. Given that the risk-neutral equilibrium drift rate remains near zero, we can again infer that the long-term risk premium has also increased significantly, indicating even more uncertainty about long-term future prices. We also see an increase in uncertainty regarding short-term deviations, as both the short-term risk premium and the volatility of the short term deviation process increase. Standard errors stay about the same or are further decreased for the parameter estimates in this case, and the log-likelihood score is again increased with the additional data.

To review the evolution of the underlying state variables, , and , are plotted for each of the successive two-factor cases in Fig. 4. The equilibrium levels for each case are shown at the top of the plot, and the

\[
F_t = E \left[ (y_{t|t-1}) (y_{t|t-1}) \right].
\]

4.2. Parameter estimates

We used MATLAB routines to numerically determine the estimates of the parameters for the Schwartz and Smith two-factor process, as well as the simplified GBM model with only the single factor , based on a crude oil futures data price set. This routine includes modules to read in and manipulate the futures data, to return the likelihood function based on the Kalman filter, and to maximize the likelihood function using the fminunc optimization routine in the MATLAB Optimization Toolbox™. The futures data set covers the period from January 1990 to January 2013 and includes mid-week prices for contracts with 1-, 3-, 5-, 9-, 13-, and 17-month maturities (Fig. 3). We used three time horizons of particular interest in our parameterization study: 1) the period from 1990 to 1995, which approximately corresponds to the futures data used in the Schwartz and Smith (2000) paper; 2) the period from 1990 to 2004 which, as discussed above, approximately predates the price run-up over the past few years; and 3) the period from 1990 up through the last week of January, 2013.

For the last of these three time horizons, we applied the Kalman filter parameterization approach for both the Schwartz and Smith two-factor model and a single factor GBM. The stochastic process parameter estimates, measurement errors (, ), and log-likelihood scores for each of these cases are shown in Table 2.

The results shown above for the first case (1990–1995) indicate good overall agreement with the estimates obtained by Schwartz and Smith (2000). While our results duplicated most of those results, we might also expect some differences due to slight differences in data. We did not have access to their source, Knight-Ridder financial services, instead using data that we obtained from Bloomberg. Given that global optimization of a function of seven variables is a challenging computational problem, slight differences between the estimates could also result from the use of different optimization routines. A Gauss optimization routine was used in the Schwartz and Smith (2000) study, whereas we used a MATLAB routine in this study.

If we consider the confidence intervals around parameter estimates from Schwartz and Smith (2000), only our estimates for short-term volatility and equilibrium volatility are outside the 95% intervals; however these two parameters also have very small standard errors. We also obtained a slightly higher log-likelihood score, but this again may be due to some differences in the optimization routine and data source.

The results from fitting the data over the expanded time horizon from January 1990 to December 2004 indicate two major changes to the parameter estimates, as shown in Table 2. The equilibrium drift rate increases significantly, from near zero to approximately 4%, while the correlation between increments of the long-term equilibrium and short-term deviation decreases significantly and in fact becomes slightly negative. When factoring in the risk-neutral equilibrium drift rates, which do not change substantially, we can infer from the increase in the equilibrium drift rate that the long-term risk premium has also increased significantly. This was likely due to some signs of a price run-up toward the end of the data set (i.e., even prior to 2005), and the uncertainty about the long-term equilibrium level. All of the other parameter estimates are similar to the estimates from the 1990–1995 data set. As expected with the additional data, the standard errors for the estimates are all lower, with the exception of that for the drift rate for the risk-neutral equilibrium process, which remained small. The log-likelihood score is about three times the score for the 1990–1995 data set, which is expected, since it is dependent on the amount of data fitted and this aligns with the improvement in the parameter estimates.

When we add the remaining available data, up through the last week of January 2013, to the data set, we notice some marked changes in the two-factor model parameter estimates. Most notably, we see that the equilibrium drift rate is again increased, to double the value from the 1990–2004 data set. Given that the risk-neutral equilibrium drift rate remains near zero, we can again infer that the long-term risk premium has also increased significantly, indicating even more uncertainty about long-term future prices. We also see an increase in uncertainty regarding short-term deviations, as both the short-term risk premium and the volatility of the short term deviation process increase. Standard errors stay about the same or are further decreased for the parameter estimates in this case, and the log-likelihood score is again increased with the additional data.

To review the evolution of the underlying state variables, , and , are plotted for each of the successive two-factor cases in Fig. 4. The equilibrium levels for each case are shown at the top of the plot, and the

\[
\frac{\lambda \mathbf{X}}{\mathbf{S}} = \begin{bmatrix} 0.0420 & 0.0007 & 0.03742 & 0.0015 & 0.0394 & 0.0010 & 0.0362 & 0.0007 & 0.0958 & 0.0020 \\
0.2860 & 0.0114 & 0.33510 & 0.0137 & 0.3525 & 0.0091 & 0.3116 & 0.0066 \\
0.0040 & 0.0006 & 0.00430 & 0.0002 \\
0.0060 & 0.0002 & 0.0068 & 0.0001 \\
0.0000 & 0.0001 & 0.0002 \\
0.0040 & 0.0006 & 0.00430 & 0.0002 \\
0.0070 & 0.0001 \\
5140 & 6230 & 18,110 & 28,056 & 18,805 \\
\end{bmatrix}
\]

\[
\text{log L} = 5140
\]
values can be read on the left axis. The short-term deviations are shown at the bottom of the plot, and their values can be read on the right secondary axis. The targets to which the short-term deviations revert (given by \( -\lambda \chi /\kappa \)) are also shown on the plot as flat lines, and can be read on the right vertical axis as well.

Each of the series for the short-term deviation appear to be mean reverting, and the figure shows how the reversion target has migrated downward due primarily to the increasing short term risk premium \( \lambda \). As the data series progresses through time, the equilibrium level is more recognizable as a GBM with increasing drift rate.

The length of time at which we would expect to have a given percentage of deviation remaining is \(-\ln(\% \text{ remaining})/\kappa\); given the range of values for \( \kappa \) in Table 2 (\(-1 \text{ to } 1.5\)). Using this relationship, only 10% of a given short-term fluctuation will be remaining after a period of two years, which thus serves as the point that divides the time horizon into short- and long-term for forecasting purposes. Furthermore, the empirical evidence in Section 3 showed that a GBM process may be a plausible approximation for the 1990–2013 data, and Schwartz and Smith (2000) noted the predominance of the long-term equilibrium process for longer-term investment and policy decision-making. Therefore, we ran an additional case with the 1990–2013 data with uncertainty in the equilibrium price only and without explicitly modeling short-term effects. These results, shown in the far right column of Table 2, indicate a lower equilibrium drift rate (5% compared to just over 8% for the equilibrium in the two-factor case). However, the equilibrium drift rate for the risk-neutral process is much lower, approaching \(-6\%\), implying a higher long-term risk premium. This might be expected in the absence of a short-term deviation factor. We will further investigate the implications of this modeling alternative in the next section on forecasting.

4.3. Validation of the estimation process using simulated data

In addition to applying the Kalman filter estimation process to empirical data, we also investigated the robustness and sensitivity of this process over a range of different conditions using simulated futures data. Our simulation of the short-term/long-term model was developed using a representative set of the seven process parameters estimated from fits to several different data sets used during our study. By assuming these parameters perfectly defined the underlying process, we simulated futures prices for contracts with 1-, 3-, 5-, 9-, 13-, and 17-month maturities using the exact discretizations defined by Davis (2012) for the long-term equilibrium

\[
\xi_{t+\Delta t} = \left( \xi_t - \frac{\lambda}{\kappa} \right) + \left( \mu - \frac{\lambda}{\kappa} \right) \Delta t + \sigma \sqrt{2 \Delta t} \varepsilon_t,
\]

and short-term deviation

\[
\chi_{t+\Delta t} = \left( \chi_t + \frac{\lambda}{\kappa} \right) e^{-\kappa \Delta t} + \sigma \sqrt{1 - \exp(-2\kappa \Delta t)} \varepsilon_t,
\]

where \( \varepsilon_t \approx N(0, 1) \) and \( \varepsilon_t \approx N(\mu_t, 1 - \rho^2) \) are the correlated increments of the processes.

For convenience, we used the same initial factors \((\xi_0, \varepsilon_0)\) observed from the historical data in early 1990. Several simulated one month futures contract prices appear in Fig. 5 over a period of 1160 weeks, or approximately 22 years. The figure shows that identical stochastic process parameters can lead to very different time series for price. The time series all exhibit a behavior somewhat similar to our historical data set, where prices initially grow slowly, with some noticeable short term deviations. But, significant departures can occur in later portions of the time series. In particular, while three out of the four selected cases never pass above $200/bbl, one of the cases easily exceeds this value over the simulated time frame. However, all four of the simulated series finished between $50/bbl and $150/bbl.

From these four sets of simulated futures prices, we re-ran the Kalman filter with maximum likelihood to demonstrate if the parameters that we used in the simulation could be recovered, and thereby confirm the robustness of the parameter estimates. Indeed, the parameter estimates relevant to long-term investment planning were recoverable, but required inserting a small disturbance of 0.001 squared to each of the diagonal elements of the measurement error covariance matrix. This wasn’t an issue when maximizing the likelihood function of the two factor model for real data, because small modeling errors existed due to imperfections in the price following a short-term/long-term process. For the simulated cases however, modeling error is nonexistent, and maximizing likelihood attempts to minimize the measurement errors and thereby provide a best fit for futures prices. This becomes clear when examining the table of parameter estimates obtained from the simulated data. For reference, the representative set of parameter values that we used to simulate futures data is listed in the column labeled “Real data” in Table 3.

The recovered measurement error estimates \(s_1, \ldots, s_6\) (the diagonal elements of the measurement error matrix \(V\)) from each of the simulated cases in Table 3 are all about the same size as the values in the first column, and they are extremely small. Previously, the measurement

\footnote{It is important to emphasize that in some cases the short-term fluctuation can be very impactful, e.g. the short-term investment problem discussed in Schwartz and Smith (2000). In those cases we would advocate using both factors in forecasting. The distinction in this work is that we are primarily concerned with policy and other decisions that are driven by long-term prices, such as the long-term investment problem discussed in Schwartz and Smith (2000).}

\footnote{See the Appendix for a discussion of the recursive estimation procedure within the Kalman filter, and the modifications necessary for working with synthetic data.}
error for the one month maturities was noticeably large (0.0366) versus the other maturities (ranging between 0.0097 and 0.0000). More importantly, in each of the four simulated data sets, all the parameters associated with the long-term equilibrium \( \xi \) were recovered successfully. For example, \( \mu_4 \), \( \sigma_4 \), and \( \mu_5 \) were all very close to the values found from the real futures data. Unfortunately, this was not the case with the parameters associated with the short term deviation \( \chi \), such as \( \kappa \), \( \lambda_4 \), and \( \omega_4 \). Additionally, the correlation parameter was not in alignment with the value obtained from the real futures data. These results suggest that the robustness that we expected only appears in the long-term parameter estimates.

The final simulated results investigate the sensitivity of the parameter estimates to longer history time series or to additional maturities. Table 4 summarizes the results from the parameter estimation from the two factor model where 6 futures maturities are simulated over half (11 years) or up to 10 times (223 years) of our original data set duration. As Schwartz and Smith (2000) suggested, having the longer duration time series is helpful in improving the accuracy of the parameter estimates, especially for the equilibrium drift rate, \( \mu_t \).

Table 5 shows the parameter estimates and the associated accuracy when additional maturities are added for 2, 3, 5 and 7 years. Because the current expected spot price has less dependency on longer duration contract maturities, the standard errors of the parameter estimates show little to no improvement. This finding also confirms what Schwartz and Smith (2000) noted empirically with their two different data sets of differing maturities.

5. The equilibrium level and forecasting prices

In this section we discuss the secondary purpose of this paper, which is to forecast the long-term oil prices. The distinction between short- and long-term can be defined in many different ways; however, as discussed in the previous section, we define short-term as 2 years or less, based on our range of estimates for the mean-reverting coefficient \( \kappa \). As Schwartz and Smith (2000) demonstrated with their example of a long-term investment, the value is primarily affected by the long-term equilibrium price in their two-factor model since the short-term deviation is expected to revert to zero in the non-risk neutral case. We specifically wanted to investigate whether a long-term equilibrium price forecast obtained when the short-term deviation factor was included in the model was significantly different from a single-factor GBM price forecast. If significant differences were found to exist, we wished to determine which approach might be preferred in light of the empirical evidence shown in Section 3.

We obtained long-term (10 year) forecasts from the two-factor model using parameters from the Kalman filter fits to each of the three time horizons and the following equation for spot price:

\[
\ln(E[S_t]) = e^{-\frac{\kappa}{2} t} + \xi_0 + \mu_t t + \frac{1}{2} \left( 1 - e^{-\frac{\kappa}{2} t} \right) \frac{\sigma^2}{\kappa} + \frac{1}{2} \left( 1 - e^{-\frac{\kappa}{2} t} \right) \sigma^2_k \left( 1 - e^{-\frac{\kappa}{2} t} \right) \frac{P_A \sigma^2}{N} k. \tag{7}
\]

These forecasts, along with their 10th and 90th percentile envelopes are shown in Fig. 6. We used 10 year forecasts here based on an average oilfield decline rate (8%) and assumed discount rate of 5%, which dictates that 75–80% of the present value will be realized within this timeframe.

The estimated spot price is also shown in this figure, and with a few exceptions, it remained within the confidence envelope over the forecast time horizon. The figure also illustrates how each successive forecast has a higher drift rate for the expected value of the process. One might question, however, whether the most recent forecast shown in Fig. 6 would hold up against economic realities over the long
term. Oil prices in excess of $200/barrel would likely have significant negative impact on economies around the world, which would likely put downward pressure on oil prices from the demand side. Supply would likely also significantly increase, as many marginal sources of production would become economically viable, which would also put downward pressure on prices. As a result, most official forecasters such as the IEA (2012) project import prices to be significantly below $200/barrel in 2020. A reason for this apparent disconnect is that the two-factor model includes the short-term deviation, and thus, the forecasts shown in Fig. 6 are by definition a combination of the two variables as shown in Eq. (3).

Recalling the results of our study with simulated data which showed that the recovered short-term process parameters were not as well estimated as were the long-term parameters, we followed the example from other work where short- and long-term components were isolated in order to assess commodity price effects (e.g., Lescaroux, 2009; Suenaga and Smith, 2011). Thus, we next obtained a forecast using the parameters from a single-factor GBM model without sacrificing accuracy for long-term planning and valuation. Both Pindyck (2001) and Postali and Picchetti (2006) provide affirmative arguments for this approach based upon evidence of low speeds of mean reversion in prices. Therefore, we also obtained a forecast using the parameters from a single-factor GBM model in the last columns from Table 2, with the results shown in Fig. 8. In this case, both short- and long-term price effects are included in determining the process parameters. This also means that the short-term parameters cannot be excluded in developing the forecast, which is an important distinction relative to the forecasts shown in Fig. 7.

By comparing the forecasts in Figs. 7 and 8, we can see that the forecast envelope is somewhat larger for the GBM model forecast. This is due to the fact that the non-decomposable single state variable includes both short- and long-term risk factors. Nonetheless, the respective expectations of the two forecasts are fairly similar. Thus, if the purpose for obtaining an oil price forecast is to focus on the long-term expectations with less emphasis on the range of uncertainty, then this analysis suggests that a single factor GBM forecast might be sufficient.

### Table 5

<table>
<thead>
<tr>
<th>Parameter estimates and measurement errors with additional futures maturities.</th>
<th>-22 years, 5 futures maturities</th>
<th>-22 years, 10 futures maturities</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Est.</td>
<td>Std. err</td>
</tr>
<tr>
<td>$\mu_\xi$</td>
<td>0.09022</td>
<td>0.0053</td>
</tr>
<tr>
<td>$\kappa$</td>
<td>1.05221</td>
<td>0.0001</td>
</tr>
<tr>
<td>$\lambda_\xi$</td>
<td>1.19068</td>
<td>0.0430</td>
</tr>
<tr>
<td>$\sigma_\xi$</td>
<td>0.58781</td>
<td>0.0004</td>
</tr>
<tr>
<td>$\alpha_\xi$</td>
<td>0.20641</td>
<td>0.0032</td>
</tr>
<tr>
<td>$\rho_\xi\chi$</td>
<td>0.20629</td>
<td>0.0462</td>
</tr>
<tr>
<td>$\rho_\xi\tau$</td>
<td>-0.02337</td>
<td>0.0007</td>
</tr>
<tr>
<td>$s_1$</td>
<td>3.29E-09</td>
<td>4.30E-06</td>
</tr>
<tr>
<td>$s_2$</td>
<td>-2.19E-09</td>
<td>3.49E-06</td>
</tr>
<tr>
<td>$s_3$</td>
<td>7.34E-10</td>
<td>3.26E-06</td>
</tr>
<tr>
<td>$s_4$</td>
<td>3.10E-09</td>
<td>3.27E-06</td>
</tr>
<tr>
<td>$s_5$</td>
<td>-3.36E-09</td>
<td>3.56E-06</td>
</tr>
<tr>
<td>$s_6$</td>
<td>-4.60E-09</td>
<td>4.02E-06</td>
</tr>
<tr>
<td>$s_7$</td>
<td>n/a</td>
<td>n/a</td>
</tr>
<tr>
<td>$s_8$</td>
<td>n/a</td>
<td>n/a</td>
</tr>
<tr>
<td>$s_9$</td>
<td>n/a</td>
<td>n/a</td>
</tr>
<tr>
<td>$s_{10}$</td>
<td>n/a</td>
<td>n/a</td>
</tr>
</tbody>
</table>

The resulting effect can be seen in the reduction in forecasted prices in all three forecasts relative to the corresponding forecasts in Fig. 6. The forecasts in Fig. 7 are still derived by separating price into its short- and long-term components, but this approach provides a more economically plausible forecast for decision-makers acting on behalf of stakeholders who will primarily be affected by oil price behavior over the long run.
6. Conclusions

In this paper, we have developed long term forecasts for oil prices based upon market calibrated stochastic processes, given the different modeling alternatives and an up-to-date data set. We have found that it is indeed difficult to determine statistically whether prices have historically been stationary (mean reverting) or non-stationary, and thus the rationale for two-factor models holds true. However, our analysis also shows that the data since the middle of the last decade has moved somewhat more directionally toward fitting within the framework of non-stationary processes. This presents the possibility of simpler price modeling with a GBM process when the application of the models is longer term planning and forecasting.

We have also conducted an extensive study of parameter estimation for different model types using futures data, progressing through several time horizons up through the last week of January 2013. These results show how the parameters reflect the changing nature of the futures data, and the implications for the state variable(s) modeled. We also showed the robustness of the Kalman filter approach to modeling unobservable variables, not only with our empirical data, but also with a study of synthetic futures data generated with Monte Carlo simulation. These results confirm that additional data indeed reduces the parameter estimation errors and increases the log-likelihood scores. We also confirmed that including additional maturities does not significantly improve model fitting performance.

Finally we have determined the implications of each of these intermediate results for the ultimate purpose of forecasting. Most significantly, short-term effects can significantly affect the long-term forecast in two-factor models. Two alternatives for forecasts that could be used for long-term planning purposes are a two-factor model forecast with the short-term process inactivated and a single factor forecast with the GBM form typically used to model long-term equilibrium prices. Similar expectations can be derived using both of these approaches. However the latter will include short-term effects in the overall uncertainty associated with the forecast. Although this approach has historically been a less accurate model of long-term prices, our forecasts show that it may suffice in some long-term planning applications, and has the benefit of being a simpler model to fit and implement relative to a two-factor model. The focus area of additional studies will be on improving our understanding of why short term parameters were not successfully recovered from our synthetic data sets. This may provide further evidence of their relevance in long-term planning applications.

Acknowledgments

We wish to thank John Butler for his very helpful comments and suggestions, and the Energy Management and Innovation Center at the University of Texas at Austin for providing the data for this study. We also wish to thank James E. Smith for assisting us with a MATLAB algorithm for implementing a Kalman filter.

Appendix A. Supplementary data

Supplementary data to this article can be found online at http://dx.doi.org/10.1016/j.eneco.2014.04.007.

References