



Nova Economia

ISSN: 0103-6351

ISSN: 1980-5381

Nova Economia

Aiube, Fernando Antonio Lucena; Levy, Ariel  
Recent movement of oil prices and future scenarios  
Nova Economia, vol. 29, no. 1, 2019, January-April, pp. 223-248  
Nova Economia

DOI: 10.1590/0103-6351/4159

Available in: <http://www.redalyc.org/articulo.oa?id=400465639012>

- How to cite
- Complete issue
- More information about this article
- Journal's webpage in redalyc.org

UABM [redalyc.org](http://www.redalyc.org)

Scientific Information System Redalyc  
Network of Scientific Journals from Latin America and the Caribbean, Spain and  
Portugal

Project academic non-profit, developed under the open access initiative

# Recent movement of oil prices and future scenarios

*Movimentos recentes dos preços do petróleo e os cenários futuros*

*Fernando Antonio Lucena Aiube <sup>(1)</sup>*

*Ariel Levy <sup>(2)</sup>*

<sup>(1)</sup> Universidade do Estado do Rio de Janeiro

<sup>(2)</sup> Universidade Federal Fluminense

## Abstract

The recent movement of oil prices has brought many forecasts about what is coming in the near future. This is natural since the plunge in prices has been dramatic after 2014 and oil is an essential source of energy worldwide. This paper examines the probabilities of spot price scenarios. We model prices through stochastic processes focusing on the Schwartz-Smith model. The calibration is based on the term structure of future prices. Since the conditional distribution is log-normal we define the probability of a certain value of the spot price in a given time horizon. We found that the recovery of crude oil prices will be slow in the next four years. Moreover, the scenario of prices under US\$ 20/barrel has the same probability as being greater than US\$ 50/barrel. The methodology has many applications, mainly for government planning and for oil companies in their capital budget decisions.

## Keywords

oil prices; commodity models; Kalman filter.

**JEL Codes** C1; C6; G13.

## Resumo

*Os movimentos recentes dos preços do petróleo trouxeram várias especulações sobre as previsões futuras. Este é um fato normal pois o petróleo é a fonte de energia mais relevante no mundo e teve uma enorme queda em seus preços após 2014. Este artigo examina as probabilidades de cenários para o preço à vista. Os preços foram modelados por processos estocásticos usando o modelo de Schwartz e Smith. A calibração do modelo é baseada na estrutura a termo dos preços futuros. A partir do fato de que a distribuição do preço à vista é log-normal, definimos as probabilidades de valores do preço à vista para diferentes horizontes de tempo. Concluímos que a recuperação dos preços será lenta nos próximos quatro anos. Além disso, a probabilidade de preços inferiores a US\$ 20/barril é a mesma que a probabilidade de preços superiores a US\$ 50/barril. A metodologia tem várias aplicações práticas para autoridades no planejamento do setor de energia e para empresas de petróleo no processo de aprovação de projetos de investimento.*

## Palavras-chave

*preços do petróleo; modelos para commodities; filtro de Kalman.*

**Códigos JEL** C1; C6; G13.

# 1 Introduction

Crude oil is by far the most important source of energy in the world. It is the most traded commodity in the physical market and also in future exchanges. It has an enormous impact on virtually all economies worldwide. The oil price is affected by different aspects: *a)* fundamentals of supply and demand, *b)* geopolitical instability in the Middle East, *c)* competition of existing energy sources, *d)* new technological developments less dependent on oil, *e)* restrictions on fossil fuel consumption due to environmental concerns, *f)* the role of OPEC (Organization of the Petroleum Exporting Countries) which makes political decisions in the interests of its members trying to control the direction of prices, among others (see Baumeister & Kilian, 2016). The consequence in such circumstances is highly volatile oil prices. It is well documented that the volatility of commodities is much higher than that of equities (Demilaray & Ulusoy, 2014; Mensi, Beljid, Boubaker, & Managi, 2013; Miffre & Brooks, 2013) (see Table 1 comparing standard deviations of commodities and equities). The ups and downs of oil prices affect countries in different ways. Producers suffer from low prices in their balance of payments while importers have to spend more of their budgets and suffer an inflationary process as a consequence of high oil prices, for example. The consequences of these movements have imposed great difficulties for economies that depend on oil prices. Since the 70s, the dynamics of several economies have changed as a result while many communities have modified their livelihoods, attracted by the expectation of development coming from oil discoveries and then facing dependency relationships with this commodity. In other words, these economies are exposed to the volatility of oil prices. There are many examples worldwide in which economies are seriously affected by the decline in oil prices. For instance, in Brazil, despite other issues, the regional economy of the state of Rio de Janeiro has suffered from strong deficits since oil prices plunged recently.

Focusing on recent history, the oil price has fluctuated greatly. Just before the financial crisis in 2008, the oil price skyrocketed reaching more than US\$ 140/barrel in July. This effect was claimed by researchers as the result of equity funds investing in commodities markets (also known as the financialization of commodities markets). The literature on financialization of commodities is huge, see for example Cheng and Xiong (2014), Creti,

Joëts, and Mignon (2013), Sadorsky (2014), Silvennoinen and Thorp (2013), Irwin and Sanders (2011), among others. Three months later, following the Lehman Brothers bankruptcy, oil dropped to US\$ 40/barrel. The recovery took more than a year and oil prices ranged between US\$ 80/barrel and US\$ 100/barrel for a long period. However, in the second half of 2014 an unpredicted drop pushed prices to US\$ 40/barrel again. In 2015 it plunged even more, to US\$ 35/barrel. And in the beginning of 2016, WTI (West Texas Intermediate) oil was traded at under US\$ 30/barrel on NYMEX (New York Mercantile Exchange). These recent declines in oil prices make governments and the industry continuously ask whether the floor has already been reached or not. Many oil industry experts have made forecasts based on a variety of methods.

There are several limitations on forecasting prices and their volatility, as reported in the economics and finance literature (see for example, Baumeister & Kilian, 2016 and Kilian, 2014). Despite the difficulties involved in modelling, there are many efforts devoted to obtaining some evidence about this problem and mitigating the consequences of such variation. Furthermore, better knowledge of the behaviour of oil prices, such as the short-term trend, would help governments and companies to make economic decisions. Several attempts with different methodological approaches are used with this aim (see Alquist, Kilian, & Vigfusson 2013 and Hamilton and Wu, 2014). Bini, Canever and Denardim (2015), analyzed the integration of energy and exchange rate with agricultural commodities in the Brazilian economy. They found that oil prices and exchange rate are the variables that are most integrated into agricultural prices. Many of them use an econometric approach relating the oil price to exogenous variables. In this case a natural difficulty is modelling and forecasting these exogenous variables. Another direction is the use of stochastic processes, following finance modelling that deals with derivatives and future prices. In this direction, the main difficulty is calibrating complex time series.

The recent literature on oil prices is vast. Just to mention some of these articles with various approaches and focuses, we cite: *a)* Mu and Ye (2012), who try to forecast oil prices in a long-term horizon, *b)* Cologni and Manera (2008), who analyze the effect of oil price shocks on macro-variables for the G-7 nations using a SVAR model, *c)* Kilian and Hicks (2013), who investigate whether the surge in real oil prices between 2003 and 2008 was caused by positive shocks in all industrial commodities reflecting high

growth in emerging Asia, *d*) Baumeister and Kilian (2014), who show that the use of combinations of forecasts improves oil prices forecasting, *e*) Tokic (2015), who suggests that the collapse of oil prices in 2014 was triggered by the falling euro–US dollar exchange rate.

However, there is no absolute answer regarding the best approach to understand the behaviour and the impacts of oil prices on overall economies. As mentioned, researchers have devoted efforts using different approaches such as Vector Autoregressive (VAR), Structural Vector Autoregressive (SVAR), time series models (ARMA-GARCH) and continuous time models (stochastic processes).

The main contribution of this study is the forecasting scenarios based on the probability distribution of spot prices and on the term structure of future prices. This is a fast and reliable way to tackle this problem. Oil companies can take advantage of this methodology for their capital budget decisions. Government can also use it for short-term planning. In many cases government revenue is based on royalties from oil production. In general, the use of quantitative methods can help government and regulators to have a reliable response in their strategic planning. The methodology we use here is an example that can guide central authorities. The quantitative methodologies in finance and economics have grown tremendously in the last thirty years. This is a consequence of the expansion of the financial industry (more financial integration among markets), and international trading activity. Furthermore, the demands for growth and welfare have challenged academics to point out new directions in economic science based on quantitative methods.

The approach we are using is within the context of mathematical finance. And the estimation procedure is inserted in the discipline of financial econometrics. In the following, we contextualize our approach in a historical perspective.

The literature on the modelling of prices through stochastic processes has grown since the seminal papers of Black and Scholes (1973) and Merton (1973) who define the price of European options. They used the geometric Brownian motion to model the price of stocks traded on exchanges. Using the concepts of no arbitrage they found the price of a call/put option(s) subscribed on a stock. Since then, a new discipline called mathematical finance (quantitative finance) has gained important attention from researchers. This discipline involves concepts of Stochastic Calculus,

Probability and Measure Theory, Stochastic Processes, Statistical Inference and Bayesian Methods. The main goal is pricing derivatives (or contingent claims). The first step in this direction is to model the underlying security through a stochastic process. At the time of the Black and Scholes and Merton's models, the geometric Brownian motion was an accepted process to describe the price of stock traded on exchanges. At the same time an important literature on the econometrics of financial time series (financial econometrics) has emerged. New modelling methodologies became available. For example, the behaviour of the conditional volatility of financial time series gained a new interpretation with the GARCH (Generalized Autoregressive Conditional Heteroscedasticity) models of Engle (1982) and Bollerslev (1986). These advances enabled an understanding of the stochastic behaviour of the volatility that was considered constant by Black and Sholes (1973), and Merton (1973). Heston (1993) extended the former work considering two stochastic processes to model stock prices. The basic equation of the spot prices included a time-varying volatility. The volatility was modelled using a mean-reverting process. In this manner a more appropriate behaviour of the volatility could be captured and the result was a more reliable price of the derivative. In general, more complex models mean difficulties on the estimation of the parameters involved in the model. The option price value found in the Heston model is a semi-analytical formula. Another example in quantitative finance is the model of interest rates using stochastic processes. A classic example is the Vasicek (1977) model. Following this direction, derivatives on fixed income markets can also be priced. Many others derivatives can be priced following the advances of mathematical finance and financial econometrics. In the last two decades, the literature on mathematical finance has matured. On this topic we cite Musiela and Rutkowski (2005), Shreve (2004a) and Shreve (2004b).

As mentioned before, more complex models require the use of more complex methodologies for the estimation procedure. Financial econometrics advanced in the estimation of complex models using new developments and also using tools from other disciplines. For example, non-observable variables in economic models began to be estimated using the Kalman filter. The Kalman filter is an estimation procedure developed by Kalman (1960) to deal with the navigation and control of vehicles in engineering science.

The econometrics of the Kalman filter can be found in the textbooks of Harvey (1989), Hamilton (1994), Durbin and Koopman (2001), Shumway and Stoffer (2006), among others. The applications of the Kalman filter in economics can be found in Pasricha (2006). The applications in finance are cited in the next Section where we carry out a literature review. The basics of this methodology is writing the model in state-space form. All ARMA models can be described in this form. The algorithm of the filter estimates, at each time step, the hidden variable and simultaneously the hyperparameters of the model. This methodology can be used in linear and Gaussian models. This is the case of the model we use in this paper. To deal with non-linear and non-Gaussian models, the literature has advanced in the last decade in the direction of Bayesian methodology. The cost of Bayesian methods is the computational time on the inference process. However, faster computers have enabled the use of Bayesian methods. In this context, the Markov chain Monte Carlo (MCMC) has been widely used in finance literature. A classical textbook on Bayesian data analysis is Gelman *et al.* (2013). On inference using MCMC we cite the textbooks of Migon and Gamerman (2014) and Gamerman and Lopes (2006). The stochastic volatility models were further developed with inference based on MCMC methodology. See for example these improvements in Qu and Perron (2013), Laurini and Mauad (2015), to mention a few. The other direction on the estimation of non-linear and non-Gaussian models is based on sequential Monte-Carlo methods also called bootstrap filters or particle filters. A textbook on this subject is Doucet, De Freitas and Gordon (2001) and examples of the applications in finance can be found in Javaheri, Lautier and Galli (2003) and Aiube, Baidya and Tito (2008).

Many of these academic developments were driven by the demands of the financial industry. As academia responded positively on the modelling of complex problems, the financial industry evolved creating new instruments to support the growing demands of risk management and hedging problems. Nowadays the volume of derivatives negotiated daily on exchanges worldwide is enormous. Agents can hedge their exposures through different type of derivatives such as options, futures, swaps, options on futures, etc.

However, all these developments in the financial industry did not come without criticism. Many financial disasters were attributed to the use of the quantitative models on competitive daily trading. A discussion of the

use of mathematical models and the 2008 financial crisis is found in an interesting magazine article written by Shreve (2009). In his words "... A good model provides insight and a guide to action. A good user of a model understands its limitations ...". The market liquidity is one important assumption considered in many models. Without liquidity there is no counterpart to trade and the agent position cannot be undone. One important point is to understand the model limitations and its careful use.

All these advances were also used to model commodity prices. The use of stochastic processes to model futures and spot prices began in the 80s. The modelling strategies evolved following the new estimation procedures that became available. In the next section we describe the evolution of commodity prices models.

In this study, we use the Schwartz-Smith (2000) model. It is a two-factor model that considers the observation of the term structure of futures prices. In other words, we observe the prices of future contracts and filter this information obtaining the spot price series. The general assumption on commodity models is that the spot price is a non-observable variable (this is better explained in Section 3). At the same time as the spot price estimation, we calibrate the model hyperparameters using the maximum likelihood. With all variables estimated, we can forecast scenarios of the spot price  $S_t$ . Using a closed-form solution for the conditional distribution of  $S_t$  one can compute the probability  $Pr(S_{t+\Delta t})$ . This means that one can evaluate the probability of the spot price being less than or equal to some threshold value at time  $t + \Delta t$  given the information available until time  $t$ . These price scenarios can help firms, government, regulatory agencies and traders to make better decisions. We justify the choice of the model and its details in Section 3.

The text is organized as follows. Section 2 presents a brief review of the literature on modelling commodity prices through stochastic processes. Section 3 details the model used. Section 4 presents the data. Section 5 reports the results, and Section 6 concludes.

## 2 Review of the literature

The study of commodity prices is well documented in the economic and finance literature. There are several approaches to analyze the behaviour



of commodity prices. As mentioned, a relevant methodology that has been used in finance in the last thirty years is modelling prices through stochastic processes. The use of stochastic processes was extended to model commodity prices. In the following studies, we also mention the application of the stochastic process on modelling interest rates and also the use of the Kalman filter on its estimation.

One of the initial works in this direction was the study by Brennan and Schwartz (1985) on modelling the options involved in the management of a natural resource. The authors used a single stochastic factor to describe the future prices. In a subsequent work, Gibson and Schwartz (1990) modelled oil prices following two stochastic processes, one for the convenience yield and the other for the spot prices.

Schwartz (1997) analyzed the behaviour of commodity prices through three different models. The first model was based on one stochastic equation to describe the spot price. The second model uses two stochastic variables: the convenience yield and the spot price. The third model includes one more stochastic variable, the interest rate. The models were estimated through the Kalman filter. The author concluded that the two-variable model better described the empirical data and that the inclusion of the interest rate added little additional information.

Miltersen and Schwartz (1998) proposed a model in which the interest rate and the convenience yield are stochastic variables. Using Gaussian stochastic equations, they could work analytically and developed equations for the future prices and computed closed formulas for option prices on futures. Although the main goal of the paper is on development of pricing formulas, one can use these equations to obtain the previous model of Schwartz.

Babbs and Nowman (1999) used the Kalman filter to calibrate the Vasicek model on interest rates. In this model the interest rate is modelled using a mean-reverting process. They used one, two and three factor models to calibrate the term structure of interest rates in the US market. They conclude that the two- and three-factor models have similar performance.

Casassus and Collin-Dufresne (2005) modelled the convenience yield as a function of the spot price and the interest rate. The interest rate is also modelled as a mean reverting process. The model for the future price is an affine function. The coefficients of this function are found through the solution of the Ricatti equations. The authors observed a strong reversion of

spot prices and concluded that this is due to the effects of the convenience yield and to the time-varying risk premium.

One relevant property used in this type of model is the mean reversion process. This is a stylized fact in commodity models. There are empirical works dedicated to verifying this property of commodity prices. For example, the mean reversion of commodity prices is verified in Bessembinder, Coughenour, Seguin and Smoller (1995) and Pindyck (1999). The use of this type of process has an economic appeal since commodity prices fluctuate following cycles in the economy. Growth and recession periods are reflected in commodity prices in the long run. Also, the short-term prices are affected by temporary disruptions or transient shocks. For example, oil prices are affected by hurricanes in the Gulf region, and also by the level of stocks. The mean-reverting processes are also called Ornstein-Uhlenbeck process.

Cortazar and Naranjo (2006) proposed an N-factor Gaussian model to investigate the stochastic behaviour of oil future prices. The estimation was based on the Kalman filter. For oil prices the authors conclude that three factors were able to explain the term structure of future prices and four factors to explain the term structure of the volatility. They also used daily copper prices and reported a good performance of the model.

Schwartz and Smith (2000) modelled oil prices through two stochastic processes. This is the model we use in this work and the details are left to the next Section. However, in the following we give examples of many studies using this model.

Manoliu and Tompaidis (2002) used this model to analyze the gas market in the US. The estimation was based on the Kalman filter. Cartea and Williams (2008) used it to analyze natural gas prices in the UK. They also estimated the model using the Kalman filter. These results were used to analyze interruptible supply contracts of natural gas. Kolos and Ronn (2008) studied the risk premium in energy markets. One of their assumptions was modelling the term structure of volatility using the Schwartz-Smith model. Sørensen (2002) investigated seasonality in agricultural commodities. The author applied the model to analyze the prices of corn, wheat and soybean. Lucia and Schwartz (2002) analyzed electricity markets.

There are also extensions to the Schwartz-Smith model. Escribano, Peña and Villaplana (2011) studied the electricity market considering mean reversion, seasonality, GARCH behaviour and jump process. They used the model to analyze the markets in different countries. Villaplana (2004)

introduced a jump component to the short-term stochastic variable. Hence the model lost its Gaussian property. In line with this work, Aiube *et al.* (2008) carried out the estimation of the non-Gaussian model using the particle filter methodology. As mentioned earlier, this type of model requires more computation time to be estimated.

The use of stochastic processes on modelling financial time series has many applications such as *a)* pricing derivatives, *b)* forecasting prices, *c)* pricing real options on investment decisions, *d)* management of financial portfolios, *e)* valuation, management and pricing contracts of real assets under uncertainty. Here our focus is on extracting information from spot price distribution to infer probabilities on forecasting prices.

### 3 The model

As we mention in the Introduction there are many approaches to forecasting prices. One of them is the use of continuous time finance that treats the dynamic of prices as stochastic processes. Commodity prices can also be modelled in this way. In Section 2, we described different types of commodity models available in the literature. The complexity of the models has grown since the first models. The increase in complexity is introduced to better capture the empirical data. However, as mentioned, more complex models mean greater difficulties in the estimation. The main goal of this article is to show the use of a reliable model capable of helping agents, government and companies on strategic planning and decision processes. Our model choice is based on the following: *a)* Schwartz and Smith's model is a well-known model in commodity literature. We mentioned many published articles using this model in the previous Section. *b)* The model is linear and Gaussian and the natural methodology available on estimation is the Kalman filter algorithm which is well established in economics. These features make it much less complex than non-Gaussian and/or non-linear models. *c)* In the model the spot price is decomposed into two latent variables describing the near-term (such as convenience yield or transitory shocks) and long-term effects in commodity prices making it more orthogonal. This point is explained below. *d)* The short-term component is modelled following a mean-reverting process. This feature enables it to capture the shocks as transitory effects, such as momentary scarcity

or natural disasters (hurricanes), for example. *e)* Bernard, Khalaf, Kichian and McMahon. (2008) used seven different models of the GARCH family and Schwartz and Smith (2000) to analyze the properties of oil prices and found that the Schwartz-Smith model provided a better fit with minor forecasting errors. To sum up we follow the same strategy as in Cartea and Williams (2008). They mention that more complex models could be used such as models that include jumps or stochastic volatility on the dynamics of commodities. However, this will make the model lose the Gaussian property and the estimation would be more complicated. In other words, using a Gaussian model we cannot capture properly spikes and jumps that sometimes are observed in empirical data. This is a limitation of the model we are using, nonetheless the estimation is much easier and direct to implement with a reliable response.

Another important issue is why commodity models rely on future prices observation. In general commodity models consider the spot price as a non-observable variable. The reason is the following. The spot price is the price quoted for immediate settlement on a transaction. In commodity markets this condition is rarely observed. Once the transaction occurs the commodity will be shipped and the delivery will be much later, days or even months. On the delivery the transaction is settled. The price of this transaction is collected by specialized firms that make pools among traders. Rigorously speaking this is not a spot transaction. The price involved in the transaction is not formed on a competitive basis. Moreover, in many cases the traders inform the price but there is no commitment to buy or sell the product. The price formed on a competitive basis is the price coming from exchanges where there are thousands of transactions daily. Hence the literature on finance adopts the future prices as the observation variable. The consequence is that the spot price needs to be estimated from futures. That is why the filtering methodology is used to estimate the spot price. The nature of the spot prices in commodity markets can be observed in the report of the US-EIA 2002 (Energy Information Administration). Mu (2007) who studied the prices in the natural gas market devoted attention to this point and refers to the same report.

In the Schwartz-Smith (2000) model, the spot price  $S_t$  is modelled as the exponential sum of two latent (or non-observable) variables:  $\chi_t$  and  $\xi_t$  (equation (1a)). These variables have their dynamics expressed in equations (1b) and (1c), respectively. The model is written as

$$S_t = e^{\chi_t + \xi_t} \quad (1a)$$

$$d\chi_t = -\kappa\chi_t dt + \sigma_\chi dB_{\chi_t} \quad (1b)$$

$$d\xi_t = \mu_\xi dt + \sigma_\xi dB_{\xi_t} \quad (1c)$$

The  $\chi_t$  is called the short-term variations of the spot price. It is a mean-reverting process (Ornstein-Uhlenbeck) in which  $\kappa$  is the speed of reversion and  $\sigma_\chi$  is the volatility of the process. The second latent variable,  $\xi_t$ , is modelled as a geometric Brownian motion with drift  $\mu_\xi$  and volatility  $\sigma_\xi$ .  $dB_{\chi_t}$  and  $dB_{\xi_t}$  are increments of the standard Brownian process and they are correlated as  $dB_{\chi_t} B_{\xi_t} = \rho dt$ .

The first latent variable captures the short-term effects, or transient effects and is related to the convenience yield. For example, for variations in stocks, hurricanes, etc., all these effects are captured by  $\chi$ . The long-term equilibrium price is represented by  $\xi$  and means changes in fundamentals that persist. This decomposition is more orthogonal than that in Gibson and Schwartz's (1990) paper. This is a desirable condition since effects in one variable have little influence in the other.

The next step is to change the probability measure. Equations (1b)-(1c) are written in the real measure, or physical measure, or else historical measure (in technical terms this is called the  $\mathbb{P}$  measure). To price a derivative, we need to change it to an equivalent measure. This equivalent measure is called the equivalent martingale measure (EMM) or risk neutral measure (in technical terms this is called the  $\mathbb{Q}$  measure). This transformation is possible once the market is free of arbitrage, which is the assumption when pricing a derivative. And this measure is unique if the market is complete. To do this transformation one needs to include the market price of risk in the drift of each equation. These two new parameters are supposed to be constant.

Writing the model in the new probability measure (the measure used to price derivatives) or in the EMM we have

$$d\chi_t = (-\kappa\chi_t - \lambda_\chi) dt + \sigma_\chi d\tilde{B}_{\chi_t} \quad (2a)$$

$$d\xi_t = (\mu_\xi - \lambda_\xi) dt + \sigma_\xi d\tilde{B}_{\xi_t} \quad (2b)$$

In these equations  $\lambda_{\chi}$ ,  $\lambda_{\xi}$  are the market prices of risk of each state (or latent) variable and  $d\tilde{B}$  means the increment of the standard Brownian process in the new probability (EMM). To simplify the notation, we use  $\mu_{\xi}^* = \mu_{\xi} - \lambda_{\xi}$ , meaning the drift in the EMM.

One more assumption in this model is the existence of a term structure of future prices. This is true in most of the commodity markets. In other words, the future price  $F_{t,T_j}$  is the observed variable, where  $T_j$  is the maturity of the contract and  $j=1, \dots, m$ .

One can use the well-known equation in which the future price at time  $t$  is the conditional expected value of spot price at time  $T$  under the pricing probability, that is  $F_{t,T_j} = E^Q(S_{T_j} | \mathcal{F}_t)$ , where  $\mathcal{F}_t$  means the available information at time  $t$ . The equation above means that the expectations of  $S_T$  are being computed under EMM with all information available at time  $t$ .

Using this equation, it can be proved that the future price is given by

$$\ln(F_{t,T_j}) = \exp(-\kappa(T_j - t))\chi_t + \xi_t + A(T_j - t), j = 1, \dots, m \quad (3)$$

where

$$A(T_j - t) = \mu_{\xi}^*(T_j - t) - (1 - \exp(-\kappa(T_j - t)))\lambda_{\xi}/\kappa + \frac{1}{2} \left[ 1 - \exp(-2\kappa(T_j - t)) \right] \frac{\sigma_{\chi}^2}{2\kappa} + \sigma_{\xi}^2(T_j - t) + 2 \left( 1 - \exp(-\kappa(T_j - t)) \right) \frac{\rho\sigma_{\chi}\sigma_{\xi}}{\kappa}$$

The details of this demonstration can be found in the original work of Schwartz and Smith (2000), in Manoliu and Tompaidis (2002) or in Aiube and Samanez (2014).

We need to estimate the latent variables  $\chi_t$  and  $\xi_t$  and also the hyperparameters of the model, which we collected in the vector  $\Theta$  as  $\Theta = (\kappa, \sigma_{\chi}, \mu_{\xi}, \sigma_{\xi}, \rho, \lambda_{\chi}, \mu_{\xi}^*, s_j)$ . Here we define  $s_j$  as the standard deviation of the difference between observed future price and the future price modelled in equation (3) for each contract  $j$ . This error term is assumed to be Gaussian as usual in this type of modelling. Since future prices are observable, equation (3) is called the measurement equation (future prices are the input data). Equations (1b) and (1c) are called transition equations and they include the state (or latent) variables that will be estimated from the filtering methodology. Under all these considerations one can observe that

we are dealing with an affine Gaussian model for future prices. As a usual procedure in this type of modelling, we applied the Kalman filter to estimate the latent variables and simultaneously maximized the likelihood to estimate the hyperparameters. Since we estimate the spot price we know its empirical distribution. Moreover, the spot price has a lognormal distribution as defined in equation (1a). Hence, one can compute the mean and variance of the distribution. To see the details, first note that the solutions of equations (1b) and (1c) for the state variables  $\chi_t$  and  $\xi_t$  are given by

$$\chi_T = e^{-\kappa(T-t)} \chi_t + \sigma_\chi \int_t^T e^{-\kappa(T-u)} dB_{\chi_u} \quad (4a)$$

$$\xi_T = \xi_t + \mu_\xi (T-t) + \sigma_\xi \int_t^T dB_{\xi_u} \quad (4b)$$

Second, the conditional moments of the distribution under the original probability are defined as  $\Lambda_{T|F_t} = E(S_T | F_t)$  and  $\Sigma_{T|F_t} = Var(S_T | F_t)$ . Computing these values, one can write

$$\Lambda_{T|F_t} = \exp \left\{ e^{-\kappa(T-t)} \chi_t + \mu_\xi (T-t) + \frac{\delta_{T|F_t}}{2} \right\} \quad (5a)$$

$$\Sigma_{T|F_t} = \exp \{ 2e^{-\kappa(T-t)} \chi_t + 2\xi_t + 2\mu_\xi (T-t) \} \times \left[ \exp(2\delta_{T|F_t}) - \exp(\delta_{T|F_t}) \right], \quad (5b)$$

where  $\delta_{T|F_t}$  is the conditional variance of  $\chi_T + \xi_T$  at time  $T$  given the information until  $t$  and is written as

$$\delta_{(T|F_t)} = \frac{\sigma_\chi^2}{2\kappa} (1 - e^{-2\kappa(T-t)}) + \sigma_\xi^2 (T-t) + 2(1 - e^{-2\kappa(T-t)}) \frac{\rho\sigma_\chi\sigma_\xi}{\kappa}.$$

The details of these computations can be found in the Appendix. Computing the mean and variance from equations (5a) and (5b) we can evaluate the probability of the spot price being less or equal to any threshold value.

## 4 Data

The analysis consists of estimating the hyperparameters of the model and computing the mean and variance of spot prices  $S_T$ . Hence one can compute the quantiles of the distribution of  $S_T$  for different time horizons

evaluating the implied probabilities. The process was done using the parameters calibrated for the entire sample, called the full sample (1/2/1985-12/1/2015). Then we re-run the calibration considering the period after the 2008 financial crisis (9/15/2008 through 12/1/2015), called the sub-sample. Thus, we considered two different samples. In each one we obtained the hyperparameters to compute the implied probabilities following the log-normal distribution of spot prices.

To perform the calibration, we sampled a complete panel of the WTI daily crude oil future prices traded on NYMEX from 1/2/1985 to 12/1/2015. The sample consists of five contracts with different maturities:  $F_1$ ,  $F_3$ ,  $F_5$ ,  $F_7$ ,  $F_9$ . To avoid the natural turbulence during the days close to maturity, the sample was rolled over three days before the expiration of the contract. Table 1 and Table 2 exhibit the descriptive statistics of the future prices for each contract for both samples. Note that the average price of oil considering a daily sample of more than 30 years is around US\$ 42/barrel. This is not too far from the recent oil prices.

Table 1 Descriptive statistics for oil future prices (full sample)

	<b>F1</b>	<b>F3</b>	<b>F5</b>	<b>F7</b>	<b>F9</b>
<b>Mean</b>	42.50	42.58	42.53	42.42	42.31
<b>Minimum</b>	10.42	10.58	10.84	10.88	10.92
<b>Maximum</b>	145.29	146.13	146.68	146.93	146.86
<b>Std. dev.</b>	30.38	30.71	30.89	30.98	30.99
<b>Skewness</b>	1.00	0.98	0.96	0.95	0.94
<b>Kurtosis</b>	2.69	2.59	2.53	2.49	2.45

Notes: Figures are in US\$/barrel.

Table 2 Descriptive statistics for oil future prices (sub-sample)

	<b>F1</b>	<b>F3</b>	<b>F5</b>	<b>F7</b>	<b>F9</b>
<b>Mean</b>	81.44	82.44	83.00	83.29	83.46
<b>Minimum</b>	33.98	39.65	41.03	42.26	43.36
<b>Maximum</b>	113.93	114.71	114.82	114.73	114.38
<b>Std. dev.</b>	19.78	18.73	17.90	17.18	16.56
<b>Skewness</b>	-0.65	-0.66	-0.68	-0.70	-0.71
<b>Kurtosis</b>	2.27	2.25	2.29	2.35	2.42

Notes: Figures are in US\$/barrel.



## 5 Estimation results

Table 3 presents the estimation results for the two periods considered. To save space we do not show the standard error deviation for each measurement equation of future prices. They are available on request (that is the  $s_j$  parameter in the  $\Theta$  vector above). One can observe that the drift of the long-term ( $\mu_\xi$ ) is not significant even in the entire sample. This is not a surprise, since the drift in the geometric Brownian motion is the most difficult parameter to estimate. It depends on even longer time series to find a good estimation. This same finding is present in Schwartz and Smith (2000), Manoliu and Tompaidis (2000) and Sørensen (2002).

Table 3 Estimation results for the periods analyzed

Parameter	Full sample (7762 obs)	Sub-sample (1820 obs)
<b>Panel A</b>		
$\kappa$	1.5751 (0.0000)	1.0350 (0.0000)
$\sigma_\chi$	0.2696 (0.0000)	0.2313 (0.0000)
$\mu_\xi$	0.0219 (0.4917)	-0.0917 (0.3380)
$\sigma_\xi$	0.1780 (0.0000)	0.2640 (0.0000)
$\rho$	0.1210 (0.0000)	0.2110 (0.0000)
$\lambda_\chi$	0.0360 (0.0709)	-0.1170 (0.0000)
$\mu_\xi^*$	-0.0348 (0.0000)	-0.0785 (0.0000)
<b>Panel B</b>		
final state variable $\chi$	-0.23369	-0.12170
final state variable $\xi$	3.94938	3.86714
MLE function	108937.16	34032.30

Notes: Panel A presents the results of the calibration, with *p*-values in parentheses, and Panel B presents the final value of the state variables and the likelihood function.

The short-term market price of risk ( $\lambda_\chi$ ) in period 1 is significant only at the 10% level (note that this parameter is not used to compute the probabilities in the  $\mathbb{P}$ -measure). All other parameters are significant. From the results above, one can say that the analysis in the real measure is better computed considering a detrended model.

Also, comparing the  $\sigma_\xi$  in both periods, it can be concluded that the long-term volatility, in the period after the economic crisis, is much higher

than that for the entire period. This fact is consistent with the empirical high volatility in oil prices observed in the recent period.

The half-life of the shocks is given by  $-\ln(0.5) / \kappa$ . This means the time expected for short-term deviations ( $\chi$ ) to halve. One can observe that  $\kappa$  is smaller in the second period, meaning that shocks are more persistent. In other words, after the financial crisis the transient shocks take longer to dissipate.

From equations (5a) and (5b), we compute the conditional mean and variance of the distribution of spot price ( $S_T$ ) considering different time horizons  $T - t$ . Table 4 reports the results for the full sample considering three different scenarios of the spot prices, i.e. we compute the probabilities of  $S_T$  being less than or equal to US\$ 20/barrel, US\$ 30/barrel and US\$ 50/barrel.

For example, the  $Pr(S_T \leq 30)$  four years from now is 48.89%. From these results, one can say that there is a small change in the probabilities regarding the time horizon for US\$20/barrel or US\$ 30/barrel. Also, even for the US\$ 50/barrel, the change in probabilities varies less than 3% between one year and four years from now. This result can be interpreted as follows. Considering the entire history (information) of crude oil prices until December 2015 and the hypothesis embedded in the model, i.e. the log-normality of the spot prices, we conclude that the recovery of oil prices will be slow in the near future with a predominant scenario of being less than US\$ 50/barrel. Furthermore, the probability of being around US\$ 30/barrel is about 50%, regardless of the time horizon considered. Note that the probability of spot prices being greater than US\$ 50/barrel is given by  $Pr(S_T > 50) = 1 - Pr(S_T \leq 50)$ . From Table 4, for example, taking the four years scenario, we have  $Pr(S_T > 50) = 1 - 0.5673 = 0.4327$ . This probability is almost the same as the probability of it being less than US\$ 20/barrel (from Table 4,  $Pr(S_T \leq 20) = 0.4268$ ).

Table 4 Probabilities of  $S_T$  obtained from the full sample

Time horizon	$Pr(S_T \leq 20)$	$Pr(S_T \leq 30)$	$Pr(S_T \leq 50)$
one year	41.63	49.54	59.49
two years	42.25	49.29	58.17
four years	42.68	48.89	56.73

Notes: The entries are expressed in percentages.

Table 5 exhibits the results considering the sub-sample (after the 2008 economic crisis), i.e. the probabilities were computed using the parameters in the second column of Table 3. The conclusions are the same as for period 1. Further, we have some evidence that changing the sample does not alter the odds significantly. It seems to be a consequence of the fact that the long-term drift is not significant in both samples. To verify this fact, we computed the probabilities inserting the usual values for the long-term drift. There is a small change, however the overall conclusions do not change in the forecast scenarios of short-term prices. In longer horizons the impact is higher.

Table 5 Probabilities of  $S_T$  obtained from the sub-sample

Time horizon	$Pr(S_T \leq 20)$	$Pr(S_T \leq 30)$	$Pr(S_T \leq 50)$
one year	42.40	49.19	57.76
two years	42.75	48.71	56.25
four years	42.76	47.96	54.54

Notes: The entries are expressed in percentages.

This means that no matter whether one considers the most recent history of oil prices or not, the scenario that oil prices will be under US\$ 50/barrel is predominant in the next four years.

## 6 Conclusions

In this paper, we analyzed the implied probabilities of the spot prices of crude oil being under some specific levels for different time horizons. For this purpose, we used the Schwartz-Smith model. Although the methodology seems to be complex, it is simpler than those that consider non-Gaussian models. So, it has limitations but on the other hand it has advantages. The algorithm of the Kalman filter used to estimate the parameters is well-known in the economics and finance literature. The main motivation is to define probabilities to oil price scenarios, bringing agility to the decision processes of government, regulators and companies.

We calibrated the model based on two different samples. The first encompasses the entire period for the term structure of future prices available in the NYMEX for the WTI crude oil. The second considers the period

after the 2008 economic crisis. We found, from the first sample, that the probability of oil prices being under US\$ 50/barrel is about 57% over a horizon of four years. Using the second sample we found almost the same result. The model response in the short-term seems to be robust to the size of different samples collected. The long-term drift was found to be not significant in both periods. This fact is also the same as observed in many different studies, by different authors, using this model. This is also in accordance with the theoretical results of the asymptotic distribution of drift estimator (see Campbell, Lo and MacKinley, 1997). This result seems to affect the short-term forecast. We performed computations using typical values for the long-term drift and the response did not change significantly in the short-term.

The recent plunge in oil prices in the second half of 2014 seems to be linked to the property of mean reversion. From Table 1, one can observe that the long run average of oil prices in our sample is around US\$ 42.50/barrel for the whole period. And based on the results we obtained, this scenario has a probability of more than 50%.

From the results above, we conclude that the prevalent scenario for oil prices is that of slow recovery. One interesting point is that the probability of the spot price being less than US\$ 20/barrel is almost the same as it being greater than US\$ 50/barrel. Hence, the alarming scenario in the media of oil prices plunging even further below the current level, has the same probability as the opposite scenario of recovery. This is a remarkable point for agents who are anchored in oil prices. To sum up, government (especially economies that have a strong dependence on oil prices) and oil companies, interested in forecasting scenarios can take advantage of this analysis to come up with fast and reliable information. Furthermore, all the analysis above can also be applied to other types of commodities that have a term structure of future prices. For example, exporting countries of agricultural commodities can undertake this study by including a seasonal component in the spot price equation.

An interesting point is the inclusion of macro-variables in the model such as global supply, aggregate demand, etc. (we are indebted to an anonymous referee for this suggestion). Some technical difficulties will arise when mixing data with different sample frequency (daily prices and quarterly variables, for example). This is a topic for future research.

## References

- Aiube, F., Baidya, T., & Tito, E. (2008). Analysis of commodity prices with the particle filter. *Energy Economics*, 30(2), 597–605.
- Aiube, F. A. L. & Samanez, C. P. (2014). On the comparison of Schwartz and Smith's two-and three-factor models on commodity prices. *Applied Economics*, 46(30).
- Alquist, R., Kilian, L., & Vigfusson, R. (2013). Forecasting the price of oil. *Handbook of economic forecasting*, vol. 2, pp. 427–507, Elsevier, Oxford, UK.
- Babbs, S. H. & Nowman, K. B. (1999). Kalman filtering of generalized Vasicek term structure models. *Journal of Financial and Quantitative Analysis*, 34, 115–130.
- Baumeister, C. & Kilian, L. (2014). Equation forecasting the real price of oil in a changing world: A forecast combination approach. *Journal of Business & Economic Statistics*.
- Baumeister, C. & Kilian, L. (2016). Forty years of oil price fluctuations: Why the price of oil may still surprise us. *The Journal of Economic Perspectives*, 30, 139–160.
- Bernard, L., Khalaf, L., Kichian, M., & MaMahon, S. (2008). Oil prices: Heavy tails, mean reversion, and the convenience yield. Cahiers de recherche 801: 08-01.
- Bessembinder, H., Coughenour, J., Seguin, P., & Smoller, M. (1995). Mean reversion in equilibrium asset prices: Evidence from the futures term structure. *The Journal of Finance*, 50(1), 361–375.
- Bini, D. A., Canever, M. D., & Denardim, A. A. (2015). Correlação e causalidade entre os preços de commodities e energia. *Nova Economia*, 25 (1), 143–160.
- Black, F. & Scholes, M. (1973). The pricing of options and corporate liabilities. *Journal of Political Economy* 81, 3, 637–654.
- Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, 31, 307–327.
- Brennan, M. & Schwartz, E. (1985). Evaluating natural resource investments. *The Journal of Business*, 58(2), 135–157.
- Campbell, J. Y., Lo, A., & MacKinlay, A. C. (1997). *The econometrics of financial markets*. Princeton University Press.
- Cartea, A. & Williams, T. (2008). U.K. Gas Markets: The market price of risk and applications to multiple interruptible supply contracts. *Energy Economics*, 30(3), 829–846.
- Casassus, J. & Collin-Dufresne, P. (2005). Stochastic convenience yield implied from commodity futures and interest rates. *The Journal of Finance*, 60(5), 2283–2331.
- Cheng, I.-H. & Xiong, W. (2014). Financialization of commodity markets. *Annual Review of Financial Economics*, 6(1), 419–441.
- Cogni, A. & Manera, M. (2008). Oil prices, inflation and interest rates in a structural cointegrated var model for the G-7 countries. *Energy Economics*, 30(3), 856–888.
- Cortazar, G. & Naranjo, L. (2006). *Journal of Future Markets*, 26, 243–268.
- Creti, A., Joëts, M., & Mignon, V. (2013). On the links between stock and commodity markets' volatility. *Energy Economics*, 37, 16–28.
- Demilaray, S. & Ulusoy, V. (2014). Links between commodity futures and stock market: Di-

- versification benefits, financialization and financial crisis MRPA Paper, online at <http://mp.ra.ub.uni-muenchen.de/59727/>
- Doucet, A., De Freitas, N., & Gordon, N. (2001). *Sequential Monte Carlo methods in practice*. Springer Verlag.
- Durbin, J. & Koopman, S. (2001). *Time series analysis by state space methods*. Oxford University Press.
- Engle, R. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. *Econometrica: Journal of the Econometric Society*, 987–1007.
- Escribano, A., Peña, I., & Villaplana P. (2011). *Oxford Bulletin of Economics and Statistics*, 73, 622–650.
- Gamerman, D. & Lopes, H.F. (2006). *Markov chain Monte Carlo: Stochastic simulation for Bayesian inference*. CRC Press.
- Gelman, A., Carlin, J. B, Stern, H. S., David, B., Vehtari, A., & Rubin, D. B. (2013). *Bayesian data analysis*. CRC Press.
- Gibson, R. & Schwartz, E. (1990). Stochastic convenience yield and the pricing of oil contingent claims. *Journal of Finance*, 959–976.
- Hamilton, J. (1994). *Time series analysis*. Princeton University Press.
- Hamilton, J. & Wu, C. (2014). Risk premia in crude oil futures prices. *Journal of International Money and Finance*, 42, 9–37.
- Harvey, A. (1989). *Forecasting, structural time series models and the Kalman filter*. Cambridge University Press
- Heston, S. (1993). A closed-form solution for options with stochastic volatility with applications to bond and currency options. *Review of Financial Studies*, 327–343.
- Irwin, S. H. & Sanders, D. R. (2011). Index funds, financialization, and commodity futures markets. *Applied Economic Perspectives and Policy*, 33(1), 1–31.
- Javaheri A., Lautier, D., & Galli, A. (2003). Filtering in finance. *Wilmott*, 3, 67–83.
- Kalman, R. (1960). A new approach to linear filtering and prediction problems. *Journal of Basic Engineering*, 82, 1, 35–45.
- Kilian, L. (2014). Oil price shocks: Causes and consequences. *Annual Review of Resource Economics*, online at [resource.annualreviews.org](http://resource.annualreviews.org)
- Kilian, L. & Hicks, B. (2013). Did unexpectedly strong economic growth cause the oil price shock of 2003–2008? *Journal of Forecasting*, 32(5), 385–394.
- Kolos, S. & Ronn, E. (2008). Estimating the commodity market price of risk for energy prices. *Energy Economics*, 30(2), 621–641.
- Laurini, M. P. & Mauad, R. B. (2015). A common jump factor stochastic volatility model. *Finance Research Letters*. Elsevier.
- Lucia, J. & Schwartz, E. (2002). Electricity prices and power derivatives: Evidence from the Nordic Power Exchange. *Review of Derivatives Research*, 5, 5–50.
- Manoliu, M. & Tompaidis, S. (2002). Energy futures prices: Term structure models with Kalman filter estimation. *Applied Mathematical Finance*, 9, 21–43.

- Mensi, W., Beljid, M., Boubaker, A., & Managi, S. (2013). *Correlations and volatility spillovers across commodity and stock markets: Linking energies, food and gold*. MPRA paper No. 44395, <http://mpra.ub.uni-muenchen.de/44395/>
- Merton, R. (1973). Theory of rational option pricing. *The Bell Journal of Economics and Management Science*, 141-183.
- Miffre, J. & Brooks, C. (2013). Do long-short speculators destabilize commodity futures markets? *International Review of Financial Analysis*, 30, 230–240.
- Migon, H.S. & Gamerman, D. (2014). *Statistical inference: An integrated approach*. CRC Press.
- Miltersen, K. & Schwartz, E. S. (1998). Pricing of options on commodity futures with stochastic term structure of convenience yields and interest rates. *Journal of Finance and Quantitative Analysis*, 33, 33–49.
- Mu, X. (2007). Weather, Storage, and Natural Gas Price Dynamics: Fundamentals and Volatility. *Energy Economics*, 29(1), 46–63.
- Mu, X. & Ye, H. (2012). Small trends and big cycles in crude oil prices. *USAAE Working Paper*, 36.
- Musiela, M. & Rutkowski, M. (2005). *Martingale methods in financial modelling*, 2<sup>nd</sup> edition. Springer Verlag.
- Pasricha, G. K. (2006). 52, MPRA, paper 22734, Online at <http://mpra.ub.uni-muenchen.de/22734/>.
- Pindyck, R. S. (1999). The long-run evolution of energy prices. *The Energy Journal*, 20(2), 1–27.
- Qu, Z. & Perron, P. (2013). A stochastic volatility model with random level shifts and its applications to S&P500 and NASDAQ return indices, 16, (3). *The Econometrics Journal*, 309–339.
- Sadorsky, P. (2014). Modeling volatility and correlations between emerging market stock prices and the prices of copper, oil and wheat. *Energy Economics*, 43, 72–81.
- Schwartz, E. (1997). The stochastic behavior of commodity prices: Implications for valuation and hedging. *Journal of Finance*, 52, (3) 923–973.
- Schwartz, E. & Smith, J. (2000). Short-term variations and long-term dynamics in commodity prices. *Management Science*, 46, (7) p. 893–911.
- Shreve, S. E. (2004a). *Stochastic Calculus for Finance: Continuous-Time Models*. Springer.
- Shreve, S. E. (2004b). *Stochastic calculus for finance: The binomial asset pricing model*. Springer.
- Shreve, S. E. (2009). Did faulty mathematical models cause the financial fiasco? *Analytics Magazine*. <http://www.math.cmu.edu/users/shreve/ModelRisk.pdf>
- Shumway, R. & Stoffer, D. (2006). *Time series analysis and its applications*, 2<sup>nd</sup> edition. Springer Verlag.
- Silvennoinen, A. & Thorp, S. (2013). Financialization, crisis and commodity correlation dynamics. *Journal of International Financial Markets, Institutions and Money*, 24, 42–65.
- Sørensen, C. (2002). Modeling seasonality in agricultural commodity futures. *Journal of Futures Markets*, 22(5).
- Tokic, D. (2015). The 2014 oil bust: Causes and consequences. *Energy Policy*, 85, 162–169.
- Vasicek, O. (1977). An equilibrium characterization of the term structure. *Journal of Financial Economics*, 5(2), 177–188.

Villaplana, P. (2004). *Pricing power derivatives: a two-factor jump-diffusion approach*. In 8<sup>th</sup> Real Options Conference, Montreal, Canada.

### About the authors

*Fernando Antonio Lucena Aiube – faiube@gmail.com*

Faculdade de Ciências Econômicas, Universidade do Estado do Rio de Janeiro, Rio de Janeiro, Brasil.  
ORCID: <https://orcid.org/0000-0003-3627-3484>.

*Ariel Levy – alevy@id.uff.br*

Faculdade de Administração e Ciências Contábeis, Universidade Federal Fluminense, Niterói, Rio de Janeiro, Brasil.  
ORCID: <https://orcid.org/0000-0003-3557-1201>.

### About the article

Submission received on July 16, 2017. Approved for publication on May 03, 2018.



## APPENDIX

This Appendix provides the computations of the conditional mean and variance of  $S_T$  given the information until time  $t$ . Here expected value and variance will be taken under the  $\mathbb{P}$ -measure. We begin by computing the first and second moments of the state (latent) variables,  $\chi_T$  and  $\xi_T$  in the model. First, note from equation (1a) that  $S_T$  is the exponential of a sum of Gaussian distributions, hence  $S_T$  is log-normal. The mean and variance of the distribution are computed below.

The solutions for equations (1b) and (1c) when solved between time  $t$  and time  $T$  are given by equations (4a) and (4b). We start using these results

$$\chi_T = \chi_t e^{-\kappa(T-t)} + \sigma_\chi \int_t^T e^{-\kappa(T-u)} dB_{\chi_u} \quad (6a)$$

$$\xi_T = \xi_t + \mu_\xi (T-t) + \sigma_\xi \int_t^T dB_{\xi_u} \quad (6b)$$

The mean and variance of  $\chi_T$  and  $\xi_T$  are straightforward. Using Ito's isometry, we can write

$$E(\chi_T | \mathcal{F}_t) = \chi_t e^{-\kappa(T-t)} \quad (7a)$$

$$Var(\chi_T | \mathcal{F}_t) = \frac{\sigma_\chi^2}{2\kappa} (1 - e^{-2\kappa(T-t)}) \quad (7b)$$

$$E(\xi_T | \mathcal{F}_t) = \xi_t + \mu_\xi (T-t) \quad (7c)$$

$$Var(\xi_T | \mathcal{F}_t) = \sigma_\xi^2 (T-t) \quad (7d)$$

Now we compute the covariance between both variables. Since we know the expected values from equations (7a) and (7c), we have

$$\chi_T \xi_T = \left[ E(\chi_T | \mathcal{F}_t) + \sigma_\chi \int_t^T e^{-\kappa(T-u)} dB_{\chi_u} \right] \left[ E(\xi_T | \mathcal{F}_t) + \sigma_\xi \int_t^T dB_{\xi_u} \right].$$

Multiplying all terms into brackets, taking the expected value and noting that we have a quadratic covariation, we get

$$Cov(\chi_T, \xi_T | \mathcal{F}_t) = \frac{\rho \sigma_\xi \sigma_\chi}{\kappa} (1 - e^{-\kappa(T-t)}). \quad (8)$$

To compute the conditional expected value and variance of  $S_T$ , one needs to sum both variables  $\chi_T$  and  $\xi_T$ . First, using the results from (7a) and (7c), we have that

$$E(\chi_T + \xi_T | \mathcal{F}_t) = E(\chi_T | \mathcal{F}_t) + E(\xi_T | \mathcal{F}_t) = e^{-k(T-t)} \chi_t + \xi_t + \mu_\xi (T-t) \quad (9)$$

And from equations (7b), (7d) and (8), we obtain the conditional variance

$$\begin{aligned} Var(\chi_T + \xi_T | \mathcal{F}_t) &= Var(\chi_T | \mathcal{F}_t) + Var(\xi_T | \mathcal{F}_t) + 2Cov(\chi_T, \xi_T | \mathcal{F}_t) = \\ &= \frac{\sigma_\chi^2}{2\kappa} (1 - e^{-2k(T-t)}) + \sigma_\xi^2 (T-t) + 2(1 - e^{-2k(T-t)}) \frac{\rho \sigma_\chi \sigma_\xi}{\kappa} \end{aligned} \quad (10)$$

To simplify the notation, we will call equation (10) as  $\delta_{T|F_t}$ . Note that conditional expected value of  $S_T$  is given by

$$E(S_T | \mathcal{F}_t) = E(e^{\chi_T + \xi_T} | \mathcal{F}_t) = e^{E(\chi_T + \xi_T | \mathcal{F}_t) + \frac{1}{2} Var(\chi_T + \xi_T | \mathcal{F}_t)}, \quad (11)$$

by substituting equations (9) and (10) into equation (11), we have

$$\Lambda_{T|F_t} = E(S_T | \mathcal{F}_t) = \exp \left\{ e^{-k(T-t)} \chi_t + \xi_t + \mu_\xi (T-t) + \frac{1}{2} \delta_{T|F_t} \right\}. \quad (12)$$

Now we will compute the conditional variance of  $S_T$ , which is given by

$$Var(S_T | \mathcal{F}_t) = E(S_T^2 | \mathcal{F}_t) - E^2(S_T | \mathcal{F}_t) \quad (13)$$

Note that  $E(S_T^2 | F_t)$  is simply

$$E(S_T^2 | \mathcal{F}_t) = \exp \left\{ 2e^{-k(T-t)} \chi_t + 2\xi_t + 2\mu_\xi (T-t) + 2\delta_{T|F_t} \right\} \quad (14)$$

and that  $E^2(S_T | \mathcal{F}_t)$  is

$$E^2(S_T | \mathcal{F}_t) = \exp \left\{ 2e^{-k(T-t)} \chi_t + 2\xi_t + 2\mu_\xi (T-t) + \delta_{T|F_t} \right\} \quad (15)$$

Taking equations (14) and (15) and plugging them into equation (13) we finally have

$$\begin{aligned}\Sigma_{T|\mathcal{F}_t} = \text{Var}(S_T|\mathcal{F}_t) &= \exp\left\{2e^{-k(T-t)}\chi_t + 2\xi_t + 2\mu_\xi(T-t)\right\} \times \\ &\times \left[\exp(2\delta_{T|\mathcal{F}_t}) - \exp(\delta_{T|\mathcal{F}_t})\right]\end{aligned}\tag{16}$$