On Parameter Estimation of the Schwartz-Smith Short-Term/Long-Term Model

Xin Tai and Minyue Fu

Abstract—The short-term/long-term model proposed by Schwartz and Smith in 2000 is widely used in modeling commodity prices. A key and nontrivial problem in this modeling technique is how to estimate the model parameters. This paper considers the parameter estimation problem based on the maximum likelihood criterion and proposes a method to simplify the task. Two components are contained in the proposed method: one to do with re-parametrization and one to do with separating the parameter set so that one part can be solved directly using least-squares and another part using nonlinear optimization. The effectiveness of the proposed method is demonstrated via numerical tests.

I. INTRODUCTION

Dynamic behaviors of commodities prices are important determinants of economic fundamentals. Hence, modeling and analysis of commodity price dynamics receive a great amount of attention by both academics and practitioners, and have become an important area of financial economics. Since the Black-Scholes model proposed in 1973 [1], many alternative models have been developed for modeling commodity prices.

Brennan and Schwartz [2] proposed in 1985 a one-factor model to describe the behavior of copper prices, and the optimal decisions for managing the exploration of a mine was analyzed based on this model. Gibson and Schwartz [3] proposed the first two-factor model in commodities in 1990 to analyze oil prices, in which stochastic factors were used to describe the convenience yield and the spot price. In 1997, Schwartz [4] used a set of models including one-, two- and three-factor model to analyze the behavior of commodity prices. Schwartz and Smith [5] proposed an alternative two-factor model, which becomes very popular. Cortazar and Naranjo [6] proposed an N-factor Gaussian model to analyze the oil futures prices in 2006. However, how to estimate the model parameters becomes an important problem in this modeling technique. This has been studied by many researchers, e.g. [7] and [8].

In this paper we propose a simplified method to estimate the parameters in the short-term/long-term model based on the maximum likelihood criterion. Our method has the advantage of giving the true maximum likelihood but having low computational complexity. We first reparameterize the model parameters and then separate the new parameters into two subsets, one solvable directly using least-squares and one solvable using nonlinear optimization. As a result, very accurate parameter estimates can be given without heavy computation. A numerical test is given to illustrate its reliability.

II. SHORT-TERM/LONG-TERM MODEL

The two-factor model proposed by Schwartz and Smith [5] assumes that the logarithm of the price \( S_t \) (including spot and futures prices) is separated into two components: a short-term deviation (known as mean-reverting) part \( \chi_t \) and a long-term equilibrium part \( \xi_t \), i.e.,

\[
\ln S_t = \chi_t + \xi_t,
\]

and that both parts are driven by a first-order linear stochastic process. More specifically,

\[
\begin{align*}
\frac{d\chi_t}{d\tau} &= -(\kappa\chi_t + \lambda\chi_t)dt + \sigma_d d\xi_t, \\
\frac{d\xi_t}{d\tau} &= (\mu\xi_t - \lambda\xi_t)dt + \sigma_d d\xi_t,
\end{align*}
\]

where \( \kappa > 0 \) is the mean-reverting parameter representing the time constant of the transient response to price disturbances, \( \mu\xi > 0 \) represents the expected (linear) growth rate of the log price \( \ln(S_t) \), \( \lambda\xi \) and \( \lambda\xi \) present the investment risks due to interest costs for borrowing, both of which equal to zero in the spot price case, \( d\xi_t \) and \( d\xi_t \) are increments of standard Brownian motions representing fluctuations in the spot price, and the two are allowed to correlate with

\[
d\xi_t d\xi_t = \rho_{\xi\xi} dt, \quad \rho_{\xi\xi} \in (-1, 1)
\]

We denote the state and output of the model by \( x_t = [\chi_t \xi_t]^T \). Without loss of generality, it is assumed that the price sampling interval \( \Delta t = 1 \). Then the continuous-time model (2) and (3) can be converted into a discrete-time model as follows:

\[
\begin{align*}
x_{t+1} &= c + Gx_t + w_t, \\
y_t &= d + F'x_t + v_t,
\end{align*}
\]

where

\[
x_t = \begin{bmatrix} \chi_t \\ \xi_t \end{bmatrix}, \quad c = \begin{bmatrix} 0 \\ \mu\xi \end{bmatrix}, \quad G = \begin{bmatrix} e^{-\kappa} & 0 \\ 0 & 1 \end{bmatrix}
\]

This work was supported by Centre of Complex Dynamic Systems and Control at University of Newcastle, Australia.

X. Tai is with State Key Lab. of Industrial Control Technology, Zhejiang University, 310027, P. R. China. taixin@zju.edu.cn

M. Fu is with the School of Electrical Engineering and Computer Science, University of Newcastle, Callaghan, NSW 2308, Australia. Minyue.Fu@newcastle.edu.au
The likelihood function used by Schwartz and Smith is not explicitly given in their paper. This, however, can be worked out from the estimation algorithm used in the paper. Indeed, from (12), we get

\[ R_{t+1} = G(R_t - R_tF(F'R_tF + V)^{-1}F'R_t)G' + W \]

Using the fact that

\[ R_t - R_tF(F'R_tF + V)^{-1}F'R_t = (R_t^{-1} + FV^{-1}F')^{-1} \]

(the well-known matrix inversion lemma), the above becomes

\[ R_{t+1} = G(R_t^{-1} + FV^{-1}F')^{-1} + W \]

It is also straightforward to show that \( A_t \) can be rewritten as

\[ A_t = (R_t^{-1} + FV^{-1}F')^{-1}F' \]

which also implies that

\[ I - A_t F' = (R_t^{-1} + FV^{-1}F')^{-1}R_t^{-1} \]

Using (13)-(15), the recursion in (12) can be simplified to

\[
\hat{x}_t = (R_t^{-1} + FV^{-1}F')^{-1} \cdot \{R_t^{-1} c + G\hat{x}_{t-1} + FV^{-1}(y_t - d)\} \\
\hat{y}_t = d + F'\{c + G\hat{x}_{t-1}\} \\
R_{t+1} = G(R_t^{-1} + FV^{-1}F')^{-1} + W
\]

where \( \hat{x}_{t-1} \) is the one-step-ahead prediction of \( x_t \) with \( R_t \) being the covariance of the corresponding prediction error, \( \hat{y}_t \) is the one-head prediction of \( y_t \) with \( Q_t \) being the covariance of the corresponding prediction error, \( A_t \) is the Kalman gain, \( \hat{x}_t \) is the updated estimate of \( x_t \) with \( C_t \) being the covariance of the corresponding estimation error.

One important feature of recursive Kalman filtering is that the recursion of \( C_t \) and \( R_t \) is independent of the observation data. Indeed, from (12), we get

\[ R_{t+1} = G(R_t - R_tF(F'R_tF + V)^{-1}F'R_t)G' + W \]

The solution to the optimal one-step-ahead prediction is given by a recursive Kalman filter as follows:

\[
\hat{x}_{t|t-1} = c + G\hat{x}_{t-1} \\
\hat{y}_t = d + F'\hat{x}_{t|t-1} \\
R_t = GC_{t-1}G' + W \\
Q_t = F'R_tF + V \\
A_t = R_tFQ_t^{-1} \\
\hat{x}_t = \hat{x}_{t|t-1} + A_t(y_t - \hat{y}_t) \\
C_t = R_t - A_tFQ_tA_t' 
\]

where \( \hat{x}_{t|t-1} \) is the one-step-ahead prediction of \( x_t \) with \( R_t \) being the covariance of the corresponding prediction error, \( \hat{y}_t \) is the one-head prediction of \( y_t \) with \( Q_t \) being the covariance of the corresponding prediction error, \( A_t \) is the Kalman gain, \( \hat{x}_t \) is the updated estimate of \( x_t \) with \( C_t \) being the covariance of the corresponding estimation error.

A. Kalman Filtering

The solution to the optimal one-step-ahead prediction is given by a recursive Kalman filter as follows:

\[
\hat{x}_{t|t-1} = c + G\hat{x}_{t-1} \\
\hat{y}_t = d + F'\hat{x}_{t|t-1} \\
R_t = GC_{t-1}G' + W \\
Q_t = F'R_tF + V \\
A_t = R_tFQ_t^{-1} \\
\hat{x}_t = \hat{x}_{t|t-1} + A_t(y_t - \hat{y}_t) \\
C_t = R_t - A_tFQ_tA_t' 
\]
B. Parameter Estimation

We now consider the parameter estimation problem. We will take the assumption that the data set is sufficiently long so that the recursive Kalman filter can be approximated using the steady-state Kalman filter.

The proposed method contains two components: one to do with re-parametrization and one to do with separating the parameter set so that one part can be solved directly using least-squares and another part using nonlinear optimization. As a result, the parameter estimation problem is much more manageable.

Before we proceed further, we need to work out the likelihood function after Kalman filtering. It is a well-known property of the recursive Kalman filter that the prediction error

\[ e_t = y_t - \hat{y}_t \] (19)

is a zero-mean Gaussian white noise sequence with covariance \( Q_t = F'R_t F + V \) which becomes \( Q = F'R'F + V \) at steady state. Therefore, the probability of \( Y = \hat{Y} \) is proportional to

\[ \prod_{t=1}^{N} \frac{1}{\sqrt{\det(Q_t)}} \exp \left( -\frac{1}{2} (y_t - \hat{y}_t)'Q_t^{-1}(y_t - \hat{y}_t) \right) \] (20)

Taking the log of the above, maximizing the likelihood function is the same as minimizing

\[ \sum_{t=1}^{N} \left\{ \ln(\det(Q_t)) + (y_t - \hat{y}_t)'Q_t^{-1}(y_t - \hat{y}_t) \right\} \] (21)

To account for the potential errors in the transient period when a steady-state Kalman filter is used, we can truncate the sequence of \( \{e_t\} \) to start from some \( N_0 > 0 \). We also replace \( Q_t \) by \( Q \), so the resulting maximum likelihood estimation problem becomes

\[ \min_{\theta} J(\theta) = (N - N_0 + 1) \ln(\det(Q)) + \sum_{t=N_0}^{N} (y_t - \hat{y}_t)'Q^{-1}(y_t - \hat{y}_t) \] (22)

1) Re-parametrization of \( W \): Examining (18) shows that \( R \) and \( W \) form a one-one mapping in the set of \( 2 \times 2 \) positive definite matrices. Indeed, for any \( W > 0 \), there is a unique solution \( R > 0 \) to (18), as explained before. On the other hand, for any \( R > 0 \), \( (R^{-1} + F'F)^{-1} < R \), which means that the solution to \( W \) in (18), which is given by

\[ W = R - G(R^{-1} + F'F)^{-1}G' \] (23)

is unique and positive definite.

The observation above means that we can use \( R \) instead of \( W \) as parameters. This has the advantage that instead of solving the nonlinear equation (18), we only need to solve (23), which is much easier.

2) Re-parametrization of \( x_t \): Recall from (9) that \( y_t \) has the mean of \( e^{-\kappa t}X_0 + \xi_0 + \mu_\xi t \). To bring this term out explicitly in \( \hat{y}_t \), we define

\[ \hat{X}_t = X_0 - e^{-\kappa t}X_0 \]
\[ \hat{\xi}_t = \xi_0 - \xi_0 - \mu_\xi t \] (24)

and the new state \( \hat{x}_t = [\hat{X}_t \hat{\xi}_t]' \). Then, the discrete-time two-factor model (5) becomes

\[ \hat{x}_t = G\hat{x}_{t-1} + w_t, \hat{x}_0 = 0 \] (25)
\[ y_t = \hat{d}_t + F'\hat{x}_t + \nu_t \] (26)

where

\[ \hat{d}_t = d + (e^{-\kappa t}X_0 + \xi_0 + \mu_\xi t)h \] (27)

with \( h = [1 \ 1 \ \cdots \ 1]' \). The corresponding steady-state Kalman filter is further simplified to

\[ \hat{x}_t = K\hat{x}_{t-1} + A(y_t - \hat{d}_t), \hat{x}_0 = 0, \]
\[ \hat{y}_t = \hat{d}_t + D\hat{x}_{t-1} \] (28)

where

\[ K = (R^{-1} + F'F)^{-1}R^{-1}G \]
\[ A = (R^{-1} + F'F)^{-1}F' \]
\[ D = F'G \] (29)

3) Partial Parameter Estimation via Least-Squares: We now separate the set of parameters into two subsets: \( \theta_1 = [\mu_\xi \ \mu_\xi \ \lambda_\xi \ \xi_0 \ \xi_0]' \) and \( \theta_2 = [\kappa, R, V]' \). It turns out that \( \theta_1 \) can be efficiently optimized via least-squares. To see this, we note that \( \theta_1 \) enters only into \( \hat{d}_t \) and do so linearly (see (7), (10) and (27)), the filter (28) operates linearly on \( \hat{d}_t \). Therefore, \( \hat{d}_t \) is linear in \( \theta_1 \). In computing \( \hat{y}_t \), we can easily express it as

\[ \hat{y}_t = \phi_0 \theta_1 + \psi_t \] (30)

where \( \phi_0 \) is an \( n \times 5 \) matrix and \( \psi_t \) is an \( n \times 1 \) vector. The exact expressions for \( \phi_0 \) and \( \psi_t \) can be worked out from (28), but omitted here.

The least-squares solution to the minimization of \( J(\theta) \) with respect to \( \theta_1 \) is then given by

\[ \theta_1^* = \left( \sum_{t=N_0}^{N} \phi_0'Q^{-1}\phi_0 \right)^{-1} \sum_{t=N_0}^{N} \phi_0'Q^{-1}(y_t - \psi_t) \] (31)

and the resulting \( J(\theta) \) is given by

\[ J(\theta_2) = (N - N_0 + 1) \ln(\det(Q)) + \Upsilon(Q) \] (32)

where

\[ \Upsilon(Q) = \sum_{t=N_0}^{N} (y_t - \psi_t)'Q^{-1}(y_t - \psi_t) \]
\[ - \left( \sum_{t=N_0}^{N} (y_t - \psi_t)'Q^{-1}\phi_0 \right) \left( \sum_{t=N_0}^{N} \phi_0'Q^{-1}\phi_0 \right)^{-1} \]
\[ \cdot \left( \sum_{t=N_0}^{N} \phi_0'Q^{-1}(y_t - \psi_t) \right) \] (33)

We caution that the parameters \( \xi_0 \) and \( \xi_0 \) can not be estimated accurately using the method above. The reason is that the effect of the initial condition \( x_0 \) on the estimate \( \hat{y}_t \) is transient and will decay exponentially. When the number of samples \( N \) becomes large, the cost function \( J(\theta) \) becomes insensitive to \( x_0 \), which makes the accurate estimate of \( x_0 \) difficult. More precisely, we first see from (27) that the effect...
of $\chi_0$ on $\hat{d}_i$ diminishes exponentially. Then, we see from (28) that the effect of $d_\tau$ on $\hat{y}_{i+\tau}$ also diminishes exponentially (because $K$ has eigenvalues strictly within the unit circle). The two facts above together mean that the effect of $\xi_0$ on $\hat{y}_i$ also diminishes exponentially, we note that the component of $\xi_0$ in $\hat{d}_i$ is a constant vector $h\xi_0$. At steady state, the response of $\hat{y}_i$ to this component is

$$\hat{y}_0 = (I - D(I - K)^{-1}A)h\xi_0$$

which can be obtained from (28) by setting $\hat{y}_i = \hat{y}_{i-1}$. Using (29) and the fact that $h\xi_0 = F'G(0 \xi_0)'$, it is straightforward to verify that $\hat{y}_0 = 0$ for any $\xi_0$. This confirms the claim that the effect of $\xi_0$ on $\hat{y}_i$ diminishes exponentially.

Inaccurate estimation of $\chi_0$ is typically not problematic because its effect on $\hat{y}_i$ (and hence $\hat{y}_i$) is transient only. This difficulty is also inherent because its effect is transient only. The difficulty in estimating $\xi_0$ is not inherent, it is due to the use of the Kalman filter based method. In other words, accurate estimation of $\xi_0$ is possible because its information is permanently present in $y_i$. Despite the fact that $\xi_0$ are $\xi_0$ can be inaccurately estimated, it does not affect computing the minimum of $J(\theta)$ over $\theta_1$.

4) Recursive Partial Parameter Estimation: The solution to $\theta_1$ in (31) can be computed recursively using a standard recursive least-squares method, as explained below.

Defining the recursion

$$P_t = (P_{t-1} + \phi_t'Q^{-1}\phi_t)^{-1}$$

with $P_{N_0} = \gamma I$ for some large $\gamma$, it is straightforward to verify that

$$P_t = P_{t-1} - P_{t-1}\phi_t'(Q + \phi_t P_{t-1}\phi_t')^{-1}\phi_t P_{t-1}$$

(35)

(which follows the matrix inversion lemma) and that

$$P_t \to \left(\sum_{t=N_0}^t \phi_t'Q^{-1}\phi_t\right)^{-1}, \text{ as } \gamma \to \infty$$

(36)

when $t$ is large. Defining

$$\theta_{t,i} = P_t \sum_{t=N_0}^t \phi_t'Q^{-1}(y_t - \psi_t)$$

(37)

and using (35), we have

$$\theta_{t,i} = (I - \alpha_t \phi_t)\theta_{t-1,i} + \alpha_t(y_i - \psi_i)$$

(38)

with $\theta_{1,N_0} = 0$, where

$$\alpha_t = P_{t-1}\phi_t'(Q + \phi_t P_{t-1}\phi_t')^{-1}$$

(39)

It is clear that as $\gamma \to \infty$, $\theta_{1,i} \to \theta_1^*$ in (31).

C. Parameter Estimation Algorithm

The proposed parameter estimation method is summarised in the algorithm below:

Step 1: Initialize $\theta_0$, i.e., initialize $\kappa, R$ and $V$.

Step 2: Compute $F, G, W, K, A$ and $D$ using (6), (7), (23) and (29). Also compute $Q = F'RF + V$.

Step 3: Express $\tilde{d}_i = \Delta \theta_1 + \delta$ using (7) and (27), then compute $\Delta$ and $\delta$.

Step 4: Run the steady-state Kalman filter (28) to compute $\phi_t$ and $\psi_t$ in (30) for $t = N_0, 2, \ldots, N$ (where $N_0$ can be chosen to be, say 10% of $N$).

Step 5: Compute the optimal $\theta_1$ and $J(\theta_2)$ using (31) and (32).

Step 6: Tune $\kappa, R, V$ using any nonlinear optimization method (e.g., Newton Gradient search) until further reduction of $J(\theta_2)$ is negligible.

IV. NUMERICAL TESTS

In this section some simulation results will be given to illuminate how this method works. We first test the method on a set of artificially generated data using a known two-factor model. The purpose of this exercise is to see how well the method estimated the parameters. We then apply the proposed method on the NYMEX crude oil futures contract data which were used in [5], and we show that our method gives somewhat better estimates than those given in [5].

A. Generation of Artificial Data

A set of artificial data are generated using the two-factor model (equation (5)), with the values of parameters as in Table I. Fig. 1 shows the state variables, and Fig. 2 shows the outputs of the futures contracts with maturities (0,1,5,9,13,17) months (Obviously, the first one is the spot price). We generate 500 data samples (i.e., $N = 500$).

**Table I**

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Values</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\kappa$</td>
<td>1.49</td>
</tr>
<tr>
<td>$\sigma_x$</td>
<td>0.286</td>
</tr>
<tr>
<td>$\lambda_x$</td>
<td>0.157</td>
</tr>
<tr>
<td>$\mu_x$</td>
<td>-0.0125</td>
</tr>
<tr>
<td>$\sigma_z$</td>
<td>0.145</td>
</tr>
<tr>
<td>$\mu_z$</td>
<td>0.0115</td>
</tr>
<tr>
<td>$\rho_{xz}$</td>
<td>0.300</td>
</tr>
<tr>
<td>$\rho_{x0}$</td>
<td>0.5</td>
</tr>
<tr>
<td>$\xi_0$</td>
<td>2.5, $i = 1,2,\ldots,n$</td>
</tr>
<tr>
<td>$\xi_0^*$</td>
<td>0.005</td>
</tr>
</tbody>
</table>

**Fig. 1.** Generated States of the Two-Factor Model.
B. Testing Estimation of $\theta_1$

We first want to see how the estimation of $\theta_1$ works. For this purpose, we choose $\theta_2$ to be the correct values and apply the least squares method to estimate $\theta_1$. The results are shown in Fig. 3 to Fig. 7, from which we can see that the estimated values converge well as time evolves.

C. Testing Estimation of $\theta_2$

With some initial values, the parameters in $\theta_2$ were estimated via Newton Gradient method. From the results in Table II, we can see that, when the size of the data increases, the estimated values of $\theta_2$ also approach their correct values. However, it should be noted that in this table we give the results of $\sigma_x, \sigma_z, \rho_{xz}$ directly instead of $R$ or $W$.

D. Application to NYMEX Crude Oil Prices

Now we apply the proposed method to the NYMEX crude oil futures contracts with maturities in 1, 5, 9, 13 and 17 months, from 1/2/90 to 2/17/95. This is the same set of data used in [5]. The purpose of our exercise is to see how well we can estimate the parameters in a two-factor model.

We note that [5] gives a set of estimated parameters using a numerical method. However, their estimates do not contain the values for $\chi_0$ and $\xi_0$. In order to see how our method works, we first apply our method to estimate all the parameters. We then take the parameter values given in [5] and use the least-squares method to obtain the optimal values for $\chi_0$ and $\xi_0$. Finally, the cost function $J(\theta)$ is compared using the two sets of parameters. The results are shown in Table III, from which we see that the second set of parameters has a less value of the likelihood function, which means it can let the model to be more likely to the actual data. (i.e., a smaller value for the cost function).
V. CONCLUSIONS

This paper proposes a method to simplify the parameter estimation problem in the short-term/long-term model of Schwartz and Smith. The resulting optimization problem is still nonlinear, but with much less number of parameters to search. Some numerical tests are given to show this method works. The proposed method is tested on a set of NYMEX oil data used in [5] and favorable comparison is shown.

REFERENCES