Abstract: The theory of the term structure of interest rates is of fundamental importance in financial engineering. In this paper, Square-Root Unscented Kalman Filter (SRUKF) algorithm is used to estimate a popular term structure model of interest rates. Simulation tests are conducted based on SRUKF and Extended Kalman Filter (EKF) respectively. The numerical results show that both approaches are capable of tracking changes in term structure and demonstrate the superior performance of SRUKF-based estimator.

Key-Words: Square-Root Unscented Kalman Filter, Extended Kalman Filter, Term Structure of Interest Rates, Vasicek Model

1 Introduction

The term structure of interest rates describes the relationship between bond rates of different terms and it is of great interest to people due to its position of a leading indicator for economic activities.

The modeling and estimation of the dynamics of term structure is a main issue in term structure analysis. According to the number of factors, term structure models are classified into one-factor and multi-factor models. One-factor models have often been analyzed in the literature for their simplicity. What’s more, there is evidence showing that almost 90 percent of the variation in the changes of the yield curve of bond rates is attributable to the variation in the first factor based on principle component analysis. For example, for most one-factor models, the factor is generally taken to be the instantaneous short rate [1].

Most modeling approaches are to divide the stochastic movement of the instantaneous short rate into two parts using a stochastic differential equation. The first part is the drift of the process, which is deterministic. The second part is the volatility component of the process, which is the random part. Examples are one-factor Vasicek model [2], multi-factor extensions of Vasicek model [3], two-factor Brennan-Schwartz model [4], one-factor Cox-Ingersoll-Ross model [5] and two-factor extensions of Cox-Ingersoll-Ross model [6], among many others.

Although the modeling progress keeps going, the estimation techniques are relatively immature because of the complexity in the models. For many models, the unknown probability distribution of bond rates or yields and the unobserved state variables pose the challenge. Filtering is a natural approach when the underlying state is unobserved. Kalman Filter (KF) is an optimal filter for recursive estimation for unobserved state variables [7]. KF applies where the process and measurement noises follow Gaussian distributions and the system is linear. Moreover, a wide variety of financial models are nonlinear, so it is necessary to use nonlinear filtering algorithms in which Extended Kalman Filter (EKF) is one of the most popular [8]. But when a model presents highly nonlinear nature, EKF-based estimator may diverge mainly because EKF handles nonlinear systems based on an approximate first-order Taylor series expansion around the mean values. To compensate this, many options are proposed. Examples are Particle Filter (PF) [9], Unscented Kalman Filter (UKF) [10] and Square-Root Unscented Kalman Filter (SRUKF) [11]. Due to huge computational expense of PF, this paper considers Unscented Filter. Furthermore, to prevent the covariance matrix from becoming non-positive semi-definite, the Square-Root form is adopted.

2 SRUKF Implementation

In UKF, the most computationally expensive part lies in calculating new sigma points during time update every time and this requires the computation of a matrix square-root of the state covariance matrix [12]. However, in SRUKF, such matrix square-root will be propagated directly, avoiding the re-factorizing at
each re-sampling point. Consequently, numerical stability is well kept. The system model is assumed to be of the form:

\[ x_{t+1} = f(x_t, k) + w_t \]  
\[ y_t = h(x_t, k) + v_t \]

where equation (1) is state equation, (2) is observation equation, the two equations form a space-state model, \( f(\cdot) \) and \( h(\cdot) \) are both nonlinear functions, \( x_t \) is state vector, \( y_t \) is observation vector, \( w_t \) and \( v_t \) are zero-mean, white and Gaussian noises and their joint covariance matrix is:

\[
E\left[\begin{pmatrix} w_t^T \\ v_t^T \end{pmatrix}\right] = \begin{pmatrix} Q & 0 \\ 0 & R \end{pmatrix}
\]

(3)

1) The initial condition:

\[
\hat{x}_0 = E(x_0)
\]

(4)

\[
S_0 = \text{chol}\left\{E\left[(x_0 - \hat{x}_0)(x_0 - \hat{x}_0)^T\right]\right\}
\]

(5)

where \( \text{chol} \) means Cholesky factorizing [11]. All symmetric nonnegative definite matrices have Cholesky factors. If \( A = XX^T \), where \( A \) is a symmetric nonnegative definite matrix, \( X \) is a triangular matrix, then \( X \) is a triangular Cholesky factor, indicated as \( X = \text{chol}(A) \).

2) Computation of the set of sigma points:

\[
X_{t+1} = \left\{ (x_{t+1} - \hat{x}_{t+1})_i (x_{t+1} - \hat{x}_{t+1})_j \right\}
\]

\[
= \left\{ \hat{x}_{t+1} + \sqrt{L + \lambda S} \hat{x}_{t+1} - \sqrt{L + \lambda S} \right\}_i
\]

(6)

where \( i = 1, 2, ..., N \), \( j = 1, 2, ..., L \), \( \lambda = \alpha^2 (L + \kappa) - L \) is a scaling parameter which can be adjusted to enhance the approximation of the distribution of the state vector, \( \alpha \) is a positive scaling parameter which can be made arbitrarily small to minimize the higher order effects, \( \kappa \) is a secondary scaling parameter. There are \( 2L + 1 \) sigma points to be required and \( L \) is the dimension of the state vector.

3) Time updates:

The transformed set propagates through the nonlinear state equation:

\[
X_{t+1} = f(X_{t+1})
\]

The predicted mean is calculated as:

\[
\hat{x}_t = \sum_{i=0}^{2L} W_i^{(m)} X_{i,t+1}
\]

(8)

where \( W_i^{(m)} = \frac{\lambda}{L + \lambda} \), \( W_i^{(m)} = \frac{1}{2L + 2\lambda} \), \( i = 1, 2, ..., 2L \).

The Cholesky form of the covariance is predicted using:

\[
S_t = q r \left[ W_i^{(c)} (X_{1:2L,t+1} - \hat{x}_t) \right] \sqrt{Q}
\]

(9)

\[
S_t = \text{cholupdate}\left\{ S_{0}, X_{0,t} - \hat{x}_{0}, W_i^{(c)} \right\}
\]

(10)

where \( q r \) represents the QR decomposition of the matrix and \( \text{cholupdate} \) represents the Cholesky factor update. The QR decomposition of a matrix \( A \in \mathbb{R}^{N \times N} \) is given by \( A' = QR \), where \( Q \in \mathbb{R}^{N \times N} \) is orthogonal, \( R \in \mathbb{R}^{N \times M} \) is upper triangular and \( N \geq M \) [13].

3) Measurement updates:

The sigma points propagate through the nonlinear measurement equation:

\[
Y_{t+1} = h(X_{t+1})
\]

The predicted mean observation is calculated as:

\[
\hat{y}_t = \sum_{i=0}^{2L} W_i^{(m)} y_{i,t+1}
\]

(12)

The innovation Cholesky covariance is given by:

\[
S_h = q r \left[ W_i^{(c)} (Y_{1:2L,t} - \hat{y}_t) \right] \sqrt{R}
\]

(13)

\[
S_h = \text{cholupdate}\left\{ S_{y}, Y_{0,t} - \hat{y}_t, W_i^{(c)} \right\}
\]

(14)

The cross-covariance matrix of \( x \) and \( y \) is determined by:

\[
P_{x,y} = \sum_{i=0}^{2L} W_i^{(c)} (X_{i,t+1} - \hat{x}_t)(Y_{i,t+1} - \hat{y}_t)^T
\]

(15)

where \( W_i^{(c)} = \frac{\lambda}{L + \lambda} + (1 - \alpha^2 + \beta) \), \( W_i^{(c)} = \frac{1}{2L + 2\lambda} \), \( i = 1, 2, ..., 2L \), \( \beta \) is an extra degree of freedom scalar parameter used to incorporate extra prior knowledge of the distribution of the state variable.

The Kalman gain is:

\[
K_t = \left( \frac{P_{x,y}}{S_h^T} \right) S_h
\]

(16)

The update mean is:

\[
\hat{x}_t = \hat{x}_t + K_t (y_t - \hat{y}_t)
\]

(17)

The update Cholesky factor is:

\[
S_t = \text{cholupdate}\left\{ S_t, K_t S_h, -1 \right\}
\]

(18)

3 Numerical Example of Vasicek Model

In this section, we introduce the state-space formulation of an interest rate stochastic volatility model - Vasicek model and use this example to illustrate the effectiveness of SRUKF.

As a popular one-factor stochastic volatility model, Vasicek model is widely used in term structure literature. The model is given by the following stochastic differential equation:

\[
dr_t = k(\theta - r_t)dt + \beta dr_t
\]

where \( k \), \( \theta \), \( \beta \) are all positive constants, \( k \) is the mean-reverting intensity, \( \theta \) is the long-run average of the instantaneous interest rate \( r_t \), which is the state variable, \( \beta \) is the volatility parameter of the
process, \( dz \) denotes an independent Wiener process and \( dz = \sqrt{dt} \beta_0, \beta \sim N(0,1) \).

The process specified by the aforementioned stochastic differential equation is defined in continuous time, while the observed data are sampled at discrete time intervals. Before applying directly the algorithm of SRUKF, we try to put equation (19) into discrete form. If we use daily data to do the estimation, we can set \( dt = 1/365 \). Then we get:

\[
\begin{align*}
  r_{t+1} &= r_t + \left[ k(\theta - r) \right]/365 + \left( \beta \epsilon \right)/\sqrt{365} \\
  &= (1-k/365) r_t + (k \theta)/365 + \left[ \beta \epsilon \right]/\sqrt{365} + w_t \\
  &= (1-k/365) r_t + (k \theta)/365 + w_t,
\end{align*}
\]

where \( w_t = \left[ \beta \epsilon \right]/\sqrt{365} \). \( \epsilon \) is zero-mean, white Gaussian noise with unit variance.

In Vasicek model, the instantaneous interest rate \( r_t \) is the unobserved state variable, while the corresponding bond price data can be observed. The corresponding bond pricing formula is:

\[
P_t(r_t, \mathbb{R}, \tau) = A_t(r_t, \tau)e^{-\gamma t/2} + \beta_2 \epsilon_t
\]

(21)

where \( \mathbb{R} \) is a set containing model parameters, \( \tau(t, T) = T-t \) denotes term with \( t \) representing the time the bond starts and \( T \) representing the time the bond matures, \( \epsilon_t \) is zero-mean, white Gaussian noise with unit variance, \( P_t(r_t, \mathbb{R}, \tau) \) is the bond price. Here the measurement noise \( \epsilon_t \) is added to equation (21) because the equation involves the problem of estimation. The noise \( \epsilon_t = \beta_2 \epsilon_t \) is zero-mean, white Gaussian noise with variance \( \beta_2^2 \). In equation (21),

\[
A_t(r_t, \tau) = \exp \left\{ \gamma B_t(r_t, \tau) - \frac{\beta_2^2 B_t^2(r_t, \tau)}{4k} \right\}
\]

(22)

\[
B_t(r_t, \tau) = \frac{1}{k} (1 - e^{-\gamma \tau})
\]

(23)

\[
\gamma = \theta + \beta_2 \phi - \frac{\beta_2^2}{2k}
\]

(24)

where \( \phi \) denotes the market price of risk.

Therefore, equation (20) and equation (21) are the discrete time state-space specifications of Vasicek model.

For the sake of comparison, SRUKF and EKF are respectively applied to the system in 100 simulations. During the simulation, we set \( k = 0.1, \theta = 0.2, \sigma_\beta = 1, \sigma_\gamma = 1, \phi = 0.16 \) based on historical experience. In addition, the process noise covariance matrix is \( Q = 0.0001 \) and the measurement noise covariance matrix is \( R = 0.21 \). The three parameters of SRUKF estimator are assumed to be \( \alpha = e^{-1} \) (usually \( e - 2 \leq \alpha \leq 1 \)), \( \beta = 2 \) (for Gaussian distribution, \( \beta = 2 \) is optimal), \( \kappa = 0 \) (usually set to either 0 or 3-L).

The abilities of SRUKF-based estimator and EKF-based estimator to track the term structure are presented in Fig. 1 and Fig. 2 separately.

![Fig.1 Estimation of term structure using SRUKF](image1)

![Fig.2 Estimation of term structure using EKF](image2)

The two figures show an example of the state estimates produced by SRUKF and EKF respectively compared to the truth. The displayed results indicate that EKF estimator demonstrates larger errors than SRUKF one does.

In order to further quantify the performance of the two estimators, the mean and variance of RMSE have been generated. The average results of RMSE are shown in Table 1:

<table>
<thead>
<tr>
<th>algorithms</th>
<th>SRUKF</th>
<th>EKF</th>
</tr>
</thead>
<tbody>
<tr>
<td>RMSE(mean)</td>
<td>0.63532</td>
<td>0.97559</td>
</tr>
<tr>
<td>RMSE(var)</td>
<td>0.38323</td>
<td>1.7298</td>
</tr>
</tbody>
</table>

Table 1 RMSE of 100 times simulation
The simulation result shows that SRUKF estimator seems to have stronger noise power immunity and higher precision than EKF estimator has. Such better numerical performance is mainly because that SRUKF estimator can achieve third-order accuracy and EKF estimator approximates the linearity by first-order approximation through Taylor expansion series.

4 Conclusion
SRUKF and EKF are two filters aiming at solving state and/or parameter estimation problems of nonlinear system. Compared with EKF estimator, SRUKF estimator is derivative-free and need not calculate Jacobian matrix associated with EKF algorithm and sometimes difficult to calculate in case of large systems. For SRUKF, it only needs relatively simple algebra calculation. This fact makes it much easier for SRUKF-based estimation. Simulation results tell us that SRUKF-based estimator provides higher precision because of the higher-order approximation than the EKF one. What’s more, the square-root form embedded in SRUKF helps effectively reduce the covariance of important weights and guarantees positive definiteness of the underlying state covariance. SRUKF estimator exhibits improved accuracy and numerical performance relative to the EKF one and consequently can be a better alternative in case of nonlinear system estimation than the EKF estimator.

References: