No. 1021

Analytical Valuation of Swap Yield Curves

by

Hideyuki Takamizawa and Isao Shoji

January 2003

Analytical Valuation of Swap Yield Curves

Hideyuki Takamizawa and Isao Shoji *

November 13, 2002

Abstract

In modeling the term structure in a multivariate framework, the following properties should be considered; time-varying volatilities of changes in state variables, the flexible correlation structure of state variables, and risk premiums that change sings over time. We propose a model that accommodates all these properties without conflict. In order to retain tractability for such a model, we appeal to an approximation technique, which obtains analytical approximation of conditional moments of arbitrary diffusions. Since the term structure is expressed as conditional expectation to the discounted payoff, a tractable model with adequate specification is derived by the method. We compare performance of the proposed model with that of affine term structure models, based on descriptive power to both in- and out-of-sample data on U.S. swap yields in the cross-section.

1 Introduction

Studies on interest rates by using affine term structure models reveal that the key to explaining observed data is the following properties; time-varying volatilities of changes in state variables, the realistic correlation structure, and risk premiums that change signs over time; see Duffie and Singleton (1997), Dai and Singleton (2000, 2002), and Duffee (2002). However, it is well known that affine models have the structural limitation that increasing the number of state variables which contribute to time-varying volatilities, say m, makes the correlation structure and risk premiums inconsistent with actual data. Specifically, Dai and Singleton (2000) shows that among affine models admissible in a three-factor framework the model with $m = 1$ has the best descriptive power to U.S. swap yields. With $m = 2$ and $m = 3$, the performance worsens because the instantaneous correlation matrix of state variables contains less non-zero elements so that interactions of yields

*Institute of Policy and Planning Sciences, University of Tsukuba, Tsukuba Ibaraki 305-8573, Japan. e-mail: htakamiz@sk.tsukuba.ac.jp, shoji@sk.tsukuba.ac.jp

cannot be captured consistently with observed data. Moreover, Duffee (2002) proposes an "essentially" affine model in which risk premiums are designed to change sings over time so as to explain the observed behavior of excess returns on bonds. In the study, the model with $m = 0$ has the best descriptive power to U.S. Treasury-bond yields in a three-factor framework. Increasing m reduces free parameters associated with changing signs of the risk premiums. Thus, the performance deteriorates.

In this paper, we propose a multi-factor model that can accommodate all the key properties implied by actual data with the small number of state variables. The proposed model distinguishes from existing models in that

[P1] all state variables assumed in the model contribute to time-varying volatilities,

[P2] the instantaneous correlation matrix of state variables contains enough non-zero elements to be consistent with data,

[P3] risk premiums change signs over time.

In such a setting that is more general than assumed in affine models, we do not have closed-form solutions to the valuation equation for the term structure. Therefore, in order to retain tractability for such a model, we appeal to an approximation technique proposed by Shoji (2002a), which obtains analytical approximation of conditional moments of arbitrary diffusions as solutions to a system of ordinary differential equations. Since the term structure is expressed as conditional expectation to the discounted payoff, a tractable model with adequate specification is derived by the method. We apply the proposed model to valuation for the U.S. fixed-for-variable interest rate swap yields. We compare performance of the proposed model with that of affine term structure models, based on descriptive power to both in- and out-of-sample data on swap yields in the cross-section.

The rest of the paper is organized as follows. Section 2 introduces an approximation formula for conditional moments of diffusion processes that is used as a pricing model. Section 3 briefly describes the payoff of interest-rate swap contracts to which the approximation formula is applied. Section 4 discusses adequate specification and illustrates the derivation of the model. Section 5 shows empirical results of performance comparison. Section 6 concludes.

2 Approximation Formula of Conditional Moments

We explain an approximation of conditional moments which is a key to modeling swap yield curves. For simplicity we consider a one-dimensional stochastic process $\{X_t\}$ satisfying the following stochastic differential equation (SDE):

$$dX_t = f(X_t)dt + \sigma(X_t)dB_t \, , \tag{1}$$

where $\{B_t\}$ is a standard Brownian motion. Suppose that up to nth conditional moments of X_t at time s are required. Define

$$\Psi_s(t) = (\psi_{1,s}(t), \cdots, \psi_{n,s}(t)) , \tag{2}$$

where

$$\psi_{m,s}(t) = E[(X_t - X_s)^m | \mathcal{F}_s] .$$

Let the functions f and $g = \sigma^2$ satisfy the conditions given in Appendix 1. Then, for some $\mathcal{F}_s$ measurable matrix A and vector b,

$$\Psi_s(t) = \int_s^t \exp(A(t-u))\, b\, du + R_X , \tag{3}$$

where R_X has an order, $O((\Delta t)^{(n+3)/2})$, with $\Delta t = t - s$. In addition, if A is invertible,

$$\Psi_s(t) = A^{-1}\{\exp(A\Delta t) - I\}\, b + R_X . \tag{4}$$

These equations show that $\Psi_s(t)$ can be approximated by the above analytical formula with approximation error of order $O((\Delta t)^{(n+3)/2})$. This formulation exactly holds when f is at most linear and g is at most quadratic; that is,

$$\Psi_s(t) = \int_s^t \exp(A(t-u))\, b\, du . \tag{5}$$

In addition, if A is invertible,

$$\Psi_s(t) = A^{-1}\{\exp(A\Delta t) - I\}\, b . \tag{6}$$

In the same way, the approximation formula can be extended to multi-variate case, which is actually used for modeling the swap yield curve. In the next section, we identify the payoff of interest-rate swap contracts to which the formula is applied.

3 The Payoff of Swap Yields

Swap contracts are not default-free although their contractual features reduce credit risks substantially; see, for example, Litzenberger (1992), and Duffie and Huang (1996). In modeling default events, we employ the reduced-form approach for valuing credit-risky claims developed by Duffie and Singleton (1999). This approach is straightforward in the following sense; with certain technical conditions and assumptions, risk components are collected into the risk-adjusted instantaneous rate and payoffs of risky claims are treated as if they are riskless. Hence we can price defaultable instruments in the same way as default-free ones.

The risk-adjusted interest rate R_t is the sum of the risk-free rate r_t and the mean-loss rate, which is a product of the instantaneous default probability h_t and the loss-rate L_t

in the event of default. Therefore, in the reduced-form model, it is important how R_t is specified. There are basically two approaches. The first approach is to specify R_t directly by a function of state variables; see Duffie and Singleton (1997), and Dai and Singleton (2000). The other approach is to specify r_t and h_tL_t separately to extract some of the information on default events; see Duffee (1999), Collin-Dufresne and Solnik (2001), and Liu et al. (2000). Because the purpose of this paper is to propose an analytical model in which [P1] – [P3] are compatible, not to analyze credit-risk components, we adopt the first approach.

Let $\{X_t\}$ be d-dimensional stochastic processes and R be a function defined on $\mathbf{R}^d$ into $\mathbf{R}$. Then, price of a defaultable zero-coupon bond, $D(t,T)$, can be obtained by

$$D(t,T) = E^Q\left[\exp\left(-\int_t^T R(X_u)du\right)\Bigg|\mathcal{F}_t\right], \tag{7}$$

where the conditional expectation $E^Q[\cdot|\mathcal{F}_t]$ is taken with respect to the equivalent martingale measure.

In order to simplify the expression for the payoff of swap contracts, we assume the following; (i) the stochastic discount factor to the payouts at time T is of the form $\exp(-\int_t^T R_u du)$, (ii) the set $\{D(t,T); t \leq T\}$ is used for composing floating rates as well as for discounting a series of fixed payouts, and (iii) exchanges of payouts between the fixed and floating legs take place semiannually and at the same time the floating rate paid in the next six months is determined. With these assumptions, the floating rate paid at time t_i can be expressed by $\{1 - D(t_{i-1}, t_i)\}/D(t_{i-1}, t_i)$, and setting the initial value of the swap contract to zero leads to the following simple expression for the fixed rate (swap yield) C:

$$C = \frac{2\{1 - D(t, t_{2n})\}}{\sum_{i=1}^{2n} D(t, t_i)}. \tag{8}$$

Therefore, valuing swap yields is equivalent to valuing a set of defaultable zero-coupon bonds at appropriate maturities. Define $Z_{s,t} \equiv \exp(-\int_s^t R(X_u)du)$, and $D(t,T)$ is the first conditional moment of $Z_{t,T}$ in the equivalent martingale measure. Applying the Ito formula to $Z_{s,t}$ leads to

$$dZ_{s,t} = -R(X_t)Z_{s,t}\,dt. \tag{9}$$

Therefore, jointly with the d-dimensional processes $\{X_t\}$ whose dynamics are also given by the SDEs, we calculate $E^Q[Z_{t,T}|\mathcal{F}_t]$ by using (3) or (4). In the next section, we illustrate the application of the approximation formula to the zero-coupon bond valuation after determining adequate processes for state variables.

4 Adequate Specification and Application of the Approximation Formula

A. The number of state variables

The dataset contains U.S. at-market fixed-for-variable swap yields of two, three, five, seven and ten years to maturity. Weekly sampled data (Wednesdays) covers the period from January 1993 to May 2001 (439 observations). Data source is Bloomberg.

We apply the principal component analysis (PCA) to the first difference of above data. The result shows that the first factor explains about 95.65% of the total variation in yield change and that the marginal contribution of the second factor is nearly 3.26%. Therefore, 98.91% of the total variation is explained by the first two principal components. Therefore, we employ the two-factor economy in explaining swap yield curves.

The correlation of the first difference of the ten-year yield with that of the first factor is 0.965, whereas the correlation of the first difference of the spread (measured by the ten-year minus two-year yields) with that of the second factor is 0.978. These interpretations are consistent with the former studies. In empirical analysis presented later in Section 5, we estimate model parameters by using these variables as proxies for unobservable states. That is, the level factor X is given by the ten-year yield and the slope factor Y is by the spread between the ten-year and two-year yields.

B. Specification

Basically, differences of model-implying prices arise from the following three components; state-variable processes (in the actual measure), factor risk premiums [1] and the risk-adjusted instantaneous rate. We specify the three components so as to be consistent with [P1] – [P3].

(i) *processes of state variables*

We assume the following processes, whose diffusion coefficients are similar to the model proposed by Brennan and Schwartz (1979):

$$\begin{aligned} dX &= (\kappa_1 X^3 + \theta_1)dt + \sigma_1 X dB_1 \,, \\ dY &= (\kappa_2 Y + \theta_2)dt + \sigma_2 Y dB_1 + \sigma_3 X dB_2 \,. \end{aligned} \tag{10}$$

The first variable X is modeled consistently with behavior of the level factor, so that the possibility of negative values is excluded with $\kappa_1 < 0$, $\theta_1 > 0$ and $X_0 > 0$. The reason we include the cubic term in the drift is that later we specify the risk premium by a nonlinear function whose order is higher than 1 (the linear term). Thus, in order to preclude the

[1] We mean factor risk premiums by the product of a vector of market prices of risk and a matrix of diffusion coefficients of state variables, as is referred to by CIR (1985, eq. (21))

probability that the process X explodes in finite time, we put higher order term. The second variable Y corresponds to the slope factor. It can take positive and negative values due to the term $\sigma_3 X dB_2$. The instantaneous covariance matrix is

$$\begin{pmatrix} \sigma_1^2 X^2 & \sigma_1\sigma_2 XY \\ \sigma_1\sigma_2 XY & \sigma_2^2 Y^2 + \sigma_2^3 X^2 \end{pmatrix}.$$

[P1] and [P2] are satisfied; both X and Y contribute to time-varying volatilities and the instantaneous correlation between X and Y is specified. It is noted that if we avoid negative values of Y, we replace XdB_2 with YdB_2 in the diffusion term of Y, which becomes dependent only on the level of Y. This specification is useful if we relate both state variables directly to the level of interest rates. Thus, the covariance structure of the model can generate various features implied by actual data.

(ii) *factor risk premiums*

In the first place, factor risk premiums must be consistent with arbitrage-free conditions. We follow the condition imposed by Stanton (1997) and Jiang (1998); risk premiums are zero whenever diffusion terms of state variables are zero. This condition is natural since risk premiums are compensation which investors require for taking risks associated with uncertain variation of future states. We choose a vector of factor risk premiums $\Lambda(X,Y)$ in the following:

$$\Lambda(X,Y) = (\lambda_1 XY,\ \lambda_2 Y)\,. \tag{11}$$

Since Y takes positive and negative values, risk premiums can change signs over time, satisfying [P3]. Another advantage of the specification is that the risk premium attributable to X is a nonlinear function, so that we incorporate the finding by Dai and Singleton (2000) that the nonlinear structure of risk premiums better explains actual data. As noted before, even for the nonlinear specification, X in the equivalent martingale measure does not explode if we obtain $\kappa_1 < 0$. Therefore, the estimate of λ_1 is free from sign constraint.

(iii) *The risk-adjusted instantaneous rate R*

We parameterize R by a linear function of state variables as

$$R(X,Y) = w_1 X + w_2 Y\,. \tag{12}$$

While we consider an alternative specification such as a quadratic function of state variables, the descriptive power of the model deteriorates. We refer to the model whose components are given by (10) – (12) as (Basic).

The specification for competing closed-form models are provided in Table 1. We label them as (Vasicek) and (CIR), respectively, for simplicity. Of course, there are many alternative specifications for each category, however, they are given based on the same assumptions as in (i) – (iii); X is the level factor which remains nonnegative (if possible),

Y is the slope factor which can change sings, and risk premiums are consistent with the arbitrage-free condition discussed in Stanton (1997) and Jiang (1998).

C. *Application of the approximation formula*

Define a conditional moment of $(X_T, Y_T, Z_{t,T})$ at time t under the equivalent martingale measure by

$$\psi_{l,m,n,t}(T) = E^Q[(X_T - X_t)^l (Y_T - Y_t)^m (Z_{t,T} - Z_{t,t})^n | \mathcal{F}_t] , \tag{13}$$

where l, m, n are nonnegative integers. Then, $\psi_{0,0,1,t}(T)$ is of our interest. To obtain this, we calculate up to the second conditional moments, *i.e.*, $1 \le l + m + n \le 2$. Consider all the combinations of (l, m, n) that satisfy the above constraint. Then,

$$\begin{aligned} \Psi_t(T) \quad = \quad & (\psi_{1,0,0,t}(T),\ \psi_{0,1,0,t}(T),\ \psi_{0,0,1,t}(T), \\ & \psi_{2,0,0,t}(T),\ \psi_{0,2,0,t}(T),\ \psi_{0,0,2,t}(T), \\ & \psi_{1,1,0,t}(T),\ \psi_{1,0,1,t}(T),\ \psi_{0,1,1,t}(T)) , \end{aligned}$$

and we use (4) to obtain the analytical approximation of $\Psi_t(T)$. The corresponding matrix A and the vector b are given in Appendix 2. The reason why we also consider the second conditional moments is that we can utilize information on the covariance structure of state variables for valuing defaultable zero-coupon bonds.

5 Empirical Analysis

As noted in Section 4.A, we estimate model parameters by using proxy variables for (X, Y). That is, the level factor X is proxied by the ten-year yield while the slope factor Y is by the spread between the ten-year and two-year yields. Because the main purpose of the empirical analysis is to compare performance of competing models, the relatively rough approach for estimation works well as long as the same approach is applied to all models. On the other hand, the merit of this is significant for reducing computational burden.

We estimate parameters by the quasi-maximum likelihood method. We exploit information on both time-series and cross-section of swap yields simultaneously. As indicated by Chen and Scott (1993), this scheme is particularly useful in estimating risk premium parameters of affine models; otherwise, we could not distinguish them from those of drift terms. Let Θ be a parameter vector and S_t be a vector of state variables, *i.e.*, $S_t = (X_t, Y_t)$. Furthermore, we introduce a vector of measurement errors, ϵ_t, defined as the difference between the actual and model-implying yields; we assume

$$\epsilon_t | S_t \sim i.i.d(0, V) ,$$

where V is a diagonal matrix with the i-th element v_i^2. Then, the log-likelihood function is given by

$$\sum_t \{f_1(\Theta; S_{t+\Delta t}, S_t) + f_2(\Theta; \epsilon_{t+\Delta t}, S_{t+\Delta t})\} \ , \tag{14}$$

where

$$\begin{aligned} f_1(\Theta; S_{t+\Delta t}, S_t) &= -\frac{1}{2}\ln|\Sigma_t(\Theta)| \quad -\frac{1}{2}\{S_{t+\Delta t} - M_t(\Theta)\}^T \Sigma_t(\Theta)^{-1} \{S_{t+\Delta t} - M_t(\Theta)\} \ , \\ f_2(\Theta; \epsilon_{t+\Delta t}, S_{t+\Delta t}) &= -\frac{1}{2}\ln|V(\Theta)| \quad -\frac{1}{2}\epsilon_{t+\Delta t}(\Theta)^T V(\Theta)^{-1}\epsilon_{t+\Delta t}(\Theta) \ , \end{aligned}$$

and where M_t and Σ_t are the conditional mean vector and the conditional covariance matrix for $S_{t+\Delta t}$, respectively, derived in the actual measure. These are obtained basically by utilizing the approximation formula given by (4). Note that (Vasicek) and (CIR) have the exact forms, hence given by (6).

Table 2 shows parameter estimates, which are reasonable for all models. In particular, two points that are consistent with the former studies are worth noting. First, the different speed of mean reversion is realized; the first factor X proxied by the ten-year yield has much slower mean reversion than the second factor Y proxied by the spread. This is true even for (Basic) since κ_1 is the coefficient of the cubic term. However, as noted before at extremely high levels the drift term is strictly negative, so that the speed of mean reversion is very fast. In this sense, this specification and resulting estimates are consistent with the findings by Aït-Sahalia (1996) and Stanton (1997); nonlinearity in the drift is the key to explaining observed properties of interest rates that they behave like a random walk over the wide range of their level but preserve stationarity over the global range. Second, $\sigma_2 > 0$ is obtained, implying that the instantaneous correlation between the two variables is positive. If instead we take Y as the negative of the spread between the ten-year and two-year yields, which is documented by Duffie and Singleton (1997), and Duffee (1999), $\sigma_2 < 0$ holds thus negative correlation is realized.

Table 3 exhibits performance comparison by using in-sample data. Panel A shows mean log-likelihood values evaluated at parameter estimates presented in Table 2. We find that while (Vasicek) and (CIR) are equally fitted to both time-series and cross-sectional properties, (Basic) has the much higher explanatory power in terms of the likelihood value. On the other hand, Panel B shows the size of measurement errors, which are multiplied by 10000 to be expressed in basis point (bp, 1bp = 1/10000). We again find that the fit to the cross-sectional data, which is in practice more important for descriptive power of pricing models, is better for (Basic) than for closed-form models. More precisely, while for the three-year yield the difference of the size of errors between closed-from models (2.8 bp) and (Basic, 2.6 bp) is not obvious, the differences widen for longer maturity yields; for closed-form models the size of errors increases from 5.9 bp (five-year yield) to 11.0 bp (seven-year yield), whereas for (Basic) it is around 3.1 bp for both maturity yields. On average, (Basic) has the error of the size 3.0 bp, which is less than half of (Vasicek, 6.6 bp) and (CIR, 6.6 bp).

Table 4 exhibits the performance comparison by using out-of-sample data. Root mean squared errors (RMSE) of each model are calculated, which is also expressed in basis point, given the parameter estimates presented in Table 2 and realizations of (X, Y) over the out-of-sample period. Although the relative goodness of fit is slightly worsens for (Basic), the overall pattern of results is the same as in Table 3; the average RMSE of (Basic, 3.7 bp) is about 40% smaller than those of (Vasicek, 6.2bp) and (CIR, 6.2 bp). Therefore, we find the evidence that considering [P1] – [P3] together improves the goodness of fit of the model.

6 Concluding Remarks

We propose a multi-factor model that can incorporate all of the important properties implied by actual interest-rate data without causing structural limitations. In order to derive the analytical model of swap yields in such a more general setting, we employ a method of approximating conditional moments proposed by Shoji (2002a,b). By using data on U.S. swap yields, we show that the adequately specified model in a two-factor framework outperforms the existing closed-form models remarkably. This goodness of fit can be utilized, for example, for inferring default processes by specifying the risk-adjusted instantaneous rate properly, which is left for future research.

Appendix 1

This appendix gives a sketch of proof of the approximation formula in Section I. The details are given in Shoji (2002b). In the first place we assume the followings:

1. For every n $\sup_{0\leq t<\infty} E[|X_t^n|] < \infty$.
2. For every n, there exist a constant L_n and an integer N_n, which may depend on n, such that,

$$\begin{aligned} |f^{(n)}(X)| &< L_n(1+|X|^{N_n}), \\ |g^{(n)}(X)| &< L_n(1+|X|^{N_n}), \end{aligned}$$

   where $f^{(n)}$ and $g^{(n)}$ are n-th derivatives of f and g with $f^{(0)} = f$ and $g^{(0)} = g$.

For m $(1 \leq m \leq n)$ application of the Itô's formula gives,

$$\begin{aligned} (X_t - X_s)^m &= \int_s^t m(X_u - X_s)^{m-1} f(X_u)du + \int_s^t \frac{m(m-1)}{2}(X_u - X_s)^{m-2} g(X_u)du \\ &\quad + \int_s^t m(X_u - X_s)^{m-1}\sigma(X_u)dB_u \,. \end{aligned} \tag{15}$$

Consider Taylor expansions of f and g around X_s of order $n-m+2$ and $n-m+3$, respectively. Except for the coefficient, m, the first term of the right hand side of (15) can be expressed as,

$$\begin{aligned}\int_s^t (X_u - X_s)^{m-1} f(X_u) du &= \int_s^t f^{(0)}(X_s)(X_u - X_s)^{m-1} du + \int_s^t f^{(1)}(X_s)(X_u - X_s)^m du \\ &\quad + \cdots + \int_s^t \frac{f^{(n-m+1)}(X_s)}{(n-m+1)!}(X_u - X_s)^n du \\ &\quad + \int_s^t \frac{f^{(n-m+2)}(\xi)}{(n-m+2)!}(X_u - X_s)^{n+1} du\end{aligned}$$

where $\xi = (1-\theta)X_s + \theta X_u$ for some $0 \le \theta \le 1$. In the same way the second term of (15) is expressed as a Taylor series up to $n-m+3$. Replace the first and second terms of (15) with these expressions and then apply conditional expectation $E[\cdot|\mathcal{F}_s]$ to the both sides of (15), $\psi_{m,s}(t)$ can be expressed as summation of two parts; the first is a linear combination of $\psi_{k,s}(t)$ $(0 \le k \le n)$ and the second corresponds to conditional expectation of residual terms of the Taylor expansion. Thanks to the assumption and Hölder's inequality the latter is proven to have order $O((\Delta t)^{(n+3)/2})$. This also holds for all $1 \le m \le n$. Arranging the linear combinations for all m so as to be consistent with $\Psi_s(t)$, the resulting matrix and vector come to be A and b, whose components are all $\mathcal{F}_s$ measurable. That is,

$$\Psi_s(t) = A \int_s^t \Psi_s(u) du + b(t-s) + R. \tag{16}$$

Here if A is invertible,

$$\int_s^t \exp(A(t-u))\, b\, du = A^{-1}\{\exp(A(t-s)) - I\}\, b.$$

Occasionally if f and g are at most linear and quadratic, respectively, there is no need to consider residual term because $n-m+2 > 1$ and $n-m+3 > 2$; the former implies residual term is zero because of linearity of f and so does the latter.

Appendix 2

Extension to the multi-variate case is straightforward. Applying the Ito formula to $(X_t - X_s)^l (Y_t - Y_s)^m (Z_{s,t} - Z_{s,s})^n$ leads to nine terms (except for martingale parts), and the rest of the procedures follows as in Appendix 1. Up to the αth conditional moments where $1 \le l+m+n \le \alpha$, considering all the combinations of (l,m,n) which satisfy the above constraint constitutes elements of the vector $\Psi_s(t)$. The matrix A and the vector b also satisfy the equation (16). In (Basic), we specify

$$\begin{aligned} dX_t &= \mu_{1,t}\, dt + \sigma_1 X_t\, dB_1^Q, \\ dY_t &= \mu_{2,t}\, dt + \sigma_2 Y_t\, dB_1^Q + \sigma_3 X_t\, dB_2^Q, \\ dZ_{s,t} &= -R_t\, Z_{s,t}\, dt. \end{aligned}$$

where $\mu_{1,t} = \kappa_1 X_t^3 + \theta_1 - \lambda_1 X_t Y_t$, $\mu_{2,t} = \kappa_2 Y_t + \theta_2 - \lambda_2 Y_t$, and $R_t = w_1 X_t + w_2 Y_t$. For $\alpha = 2$, the resulting matrix $A = [A_{i,j}]_{1\le i,j\le 9}$ and the vector $b = [b_i]_{1\le i\le 9}$ are, respectively,

$$
\begin{array}{llll}
A_{11} = 3\kappa_1 X_s^2 - \lambda_1 Y_s & A_{12} = -\lambda_1 X_s & A_{14} = 3\kappa_1 X_s & A_{17} = -\lambda_1 \\
A_{22} = \kappa_2 - \lambda_2 & & & \\
A_{31} = -w_1 & A_{32} = -w_2 & A_{33} = -R_s & \\
A_{38} = -w_1 & A_{39} = -w_2 & & \\
A_{41} = 2\mu_{1,s} + 2\sigma_1^2 X_s & A_{44} = 2A_{11} + \sigma_1^2 & A_{47} = 2A_{12} & \\
A_{51} = 2\sigma_3^2 X_s & A_{52} = 2\mu_{2,s} + 2\sigma_2^2 Y_s & A_{54} = \sigma_3^2 & A_{55} = 2A_{22} + \sigma_2^2 \\
A_{63} = -2R_s & A_{66} = -2R_s & A_{68} = -2w_1 & A_{69} = -2w_2 \\
A_{71} = \mu_{2,s} + \sigma_1\sigma_2 Y_s & A_{72} = \mu_{1,s} + \sigma_1\sigma_2 X_s & A_{75} = A_{12} & A_{77} = A_{11} + A_{22} + \sigma_1\sigma_2 \\
A_{81} = -R_s & A_{83} = \mu_{1,s} & A_{84} = -w_1 & \\
A_{87} = -w_2 & A_{88} = -R_s + A_{11} & A_{89} = A_{12} & \\
A_{92} = -R_s & A_{93} = \mu_{2,s} & A_{95} = -w_2 & \\
A_{97} = -w_1 & A_{99} = -R_s + A_{22} & &
\end{array}
$$

$$
\begin{array}{lll}
b_1 = \mu_{1,s} & b_2 = \mu_{2,s} & b_3 = -R_s \\
b_4 = \sigma_1^2 X_s^2 & b_5 = \sigma_2^2 Y_s^2 + \sigma_3^2 X_s^2 & b_6 = \sigma_1\sigma_2 X_s Y_s
\end{array}
$$

and the rest of the elements are zero. Certainly, A and b are $\mathcal{F}_s$ measurable.

Reference

1. Brennan, Michael J., and Eduardo. S. Schwartz, 1979, A continuous time approach to the pricing of bonds, *Journal of Banking and Finance* 3, 133-155.

2. Chen, Ren-Raw, and Louis Scott, 1993, Maximum likelihood estimation for a multifactor equilibrium model of the term structure of interest rates, *Journal of Fixed Income* 3, 14-31.

3. Collin-Dufresne, Pierre, and Bruno Solnik, 2001, On the term structure of default premia in the swap and LIBOR markets, *Journal of Finance* 56, 1095-1115.

4. Dai, Qiang, and Kenneth J. Singleton, 2000, Specification analysis of affine term structure models, *Journal of Finance* 55, 1943-1978.

5. Dai, Qiang, and Kenneth J. Singleton, 2002, Expectation puzzles, time-varying risk premia, and affine models of the term structure, *Journal of Financial Economics* 63, 415-441.

6. Duffee, Gregory R., 1999, Estimating the price of default risk, *Review of Financial Studies* 12, 197-226.

7. Duffee, Gregory R., 2002, Term premia and interest rate forecasts in affine models, *Journal of Finance* 57, 405-443.

8. Duffie, Darrell, and Ming Huang, 1996, Swap rates and credit quality, *Journal of Finance* 51, 921-949.

9. Duffie, Darrell, and Kenneth J. Singleton, 1997, An econometric model of the term structure of interest-rate swap yields, *Journal of Finance* 52, 1287-1321.

10. Duffie, Darrell, and Kenneth J. Singleton, 1999, Modeling term structures of defaultable bonds, *Review of Financial Studies* 12, 687-720.

11. Jiang, George J., 1998, Nonparametric modeling of U.S. interest rate term structure dynamics and implications on the prices of derivative securities, *Journal of Financial and Quantitative Analysis* 33, 465-497.

12. Litzenberger, Robert H., 1992, Swaps: plain and fanciful, *Journal of Finance* 47, 831-850.

13. Liu, Jun, Francis A. Longstaff, and Ravit E. Mandell, 2000, The market price of credit risk: An empirical analysis of interest rate swap spreads, Working Paper, UCLA.

14. Shoji, Isao, 2002a, Approximation of conditional moments of diffusion processes, *International Journal of Computational and Numerical Analysis and Applications* 1, 163-190.

15. Shoji, Isao, 2002b, Asymptotic expansion of transition probability density functions of diffusion processes, working paper, University of Tsukuba.

16. Stanton, Richard, 1997, A nonparametric model of term structure dynamics and the market price of interest rate risk, *Journal of Finance* 52, 1973-2002.

Table 1. Specification of Competing Models

State-variable processes are given in the equivalent martingale measure (drift terms are adjusted with risk premiums). Risk premiums are determined by following the arbitrage-free condition; risk premiums are zero whenever the diffusion terms of state processes are zero.

(Vasicek)		
dX	$=$	$(\kappa_1 X + \theta_1 - \lambda_1)dt + \sigma_1 dB_1^Q$
dY	$=$	$(\kappa_2 Y + \theta_2 - \lambda_2)dt + \sigma_2 dB_1^Q + \sigma_3 dB_2^Q$
$R(X,Y)$	$=$	$w_1 X + w_2 Y$
(CIR)		
dX	$=$	$(\kappa_1 X + \theta_1 - \lambda_1 X)dt + \sigma_1 \sqrt{X} dB_1^Q$
dY	$=$	$(\kappa_2 Y + \theta_2 - \lambda_2 X)dt + \sqrt{X}(\sigma_2 dB_1^Q + \sigma_3 dB_2^Q)$
$R(X,Y)$	$=$	$w_1 X + w_2 Y$
(Basic)		
dX	$=$	$(\kappa_1 X^3 + \theta_1 - \lambda_1 XY)dt + \sigma_1 X dB_1^Q$
dY	$=$	$(\kappa_2 Y + \theta_2 - \lambda_2 Y)dt + \sigma_2 Y dB_1^Q + \sigma_3 X dB_2^Q$
$R(X,Y)$	$=$	$w_1 X + w_2 Y$

Table 2. Parameter Estimates of Competing Models

Parameters are estimated by the quasi-maximum likelihood method. Log-likelihood functions from time-series and cross-sections of swap yields are maximized simultaneously. The ten-year swap yield and the spread between the ten-year and two-year swap yields are used as proxy variables for state variables. Models are fitted to the three-year, five-year and seven-year swap yields. In-sample data covers the period from January 1993 to December 1998 (313 obs.). Standard errors are in parenthesis.

	(Vasicek)	(CIR)	(Basic)
κ_1	-0.0005	-0.1224	-0.8593
	(0.0013)	(0.0591)	(0.1519)
κ_2	-1.7737	-1.6175	-0.4043
	(0.3644)	(0.2231)	(0.1662)
θ_1	-0.0030	0.0019	0.0005
	(0.0038)	(0.0004)	(0.0001)
θ_2	0.0123	0.0032	0.0003
	(0.0039)	(0.0004)	(0.0002)
σ_1	0.0091	0.0357	0.1440
	(0.0003)	(0.0013)	(0.0053)
σ_2	0.0010	0.0041	0.1465
	(0.0003)	(0.0010)	(0.0228)
σ_3	0.0045	0.0177	0.0661
	(0.0002)	(0.0007)	(0.0022)
w_1	1.0156	1.0680	0.9863
	(0.0031)	(0.1291)	(0.0038)
w_2	-3.6593	-3.3504	-1.4308
	(0.7183)	(0.4337)	(0.0243)
λ_1	-0.0043	-0.1203	-4.9298
	(0.0038)	(0.0620)	(0.0974)
λ_2	0.0098	-0.0287	-0.2505
	(0.0040)	(0.0918)	(0.1686)

Table 3. In-sample performance comparison

Panel A shows mean log-likelihood values evaluated at parameter estimates presented in Table 2. Panel B displays parameter estimates of the covariance matrix of measurement errors, given by $\epsilon_t \sim iid(0, V)$ where V is a 3×3 diagonal matrix with the ith v_i^2. Each model is fitted to the three-year (3Y), five-year (5Y) and seven-year (7Y) swap yields, so that the measurement errors correspond to the same order. 'Ave.' stands for the average of v_i $(i = 1, 2, 3)$. In-sample data covers the period from January 1993 to December 1998 (313 obs.).

	(Vasicek)	(CIR)	(Basic)
	Panel A: Mean log-likelihood		
	33.84	33.83	35.82
	Panel B: Measurement errors ($\times$ 10000, in bp)		
v_1 (3Y)	2.84	2.83	2.64
v_2 (5Y)	5.91	5.93	3.10
v_3 (7Y)	11.07	11.00	3.13
Ave.	6.61	6.58	2.96

Table 4. Out-of-sample performance comparison

Root mean squared errors (RMSE) are expressed in basis point (bp). Each model is fitted to the three-year (3Y), five-year (5Y) and seven-year (7Y) swap yields, given parameter estimates presented in Table 2 and realizations of state variables, (X, Y), in the put-of-sample period. 'Ave.' stands for the average of the three RMSEs. Out-of-sample data covers the period from January 1999 to May 2001 (126 obs.).

model	Vasicek			CIR			Basic		
	RMSE	S.D.	Bias	RMSE	S.D.	Bias	RMSE	S.D.	Bias
3Y	3.63	3.31	-1.48	3.67	3.39	-1.41	3.80	3.35	-1.79
5Y	5.85	4.97	3.09	5.40	4.79	2.49	4.01	3.54	-1.89
7Y	9.09	7.07	5.72	9.47	6.57	6.82	3.21	2.63	-1.84
Ave.	6.19			6.18			3.67		