IDENTIFICATION OF AFFINE TERM STRUCTURES FROM YIELD CURVE DATA

SHIN ICHI AIHARA
Tokyo University of Science, Suwa
Toyohira 5000-1, Chino, Nagano, Japan
aihara@rs.suwa.tus.ac.jp

ARUNABHA BAGCHI
FELab and Department of Applied Mathematics
University of Twente, P. O. Box 217
7500AE Enschede, The Netherlands
a.bagchi@ewi.utwente.nl

Received 6 October 2008
Accepted 15 April 2009

We consider a slight perturbation of the Hull-White short rate model and the resulting modified forward rate equation. We identify the model coefficients by using the martingale property of the normalized bond price. The forward rate and the system parameters are then estimated by using the maximum likelihood method.

Keywords: Interest rate models; affine term structure; forward curves; Kalman filter; MLE.

1. Introduction

There is a vast literature [4, 8, 9] on estimating parameters of short rate models in finance. One popular approach is to take a short rate model that leads to an exponential-affine expression for the corresponding bond price. The yield is then easy to calculate. Artificial noises are then added to yields of different maturities and the maximum likelihood method is used to estimate the model parameters. The likelihood function involves the filtered states which are given by the Kalman filter equations. Adding artificial noises is essential for the Kalman filter algorithm to work. As a byproduct one also gets the minimum variance estimate of the short rate, which is otherwise not observed in the market. An early survey on the use of filtering theory in finance is the paper of Brigo and Hanzon [6].

There are two fundamental drawbacks to this approach. The first is the artificial noises added to yields of different maturities, which is hard to justify. This is done
only for the purpose of the filtering algorithm to work. The other difficulty is the robustness of the model. If a model is very close to one of the usual short rate models, but not exactly matches that model, there is no guarantee that the bond price will still be of the exponential-affine form.

We, therefore, approach the problem from a different perspective. We start with the usual Hull-White model of the short rate, and assume that a slightly different model will lead to a slightly perturbed bond price of the usual one derived from the Hull-White model. We also consider this perturbation to be generated by an infinite dimensional noise, as it should depend on all times to maturity. In this approach, we do not need to add artificial noises to bond yield of different maturities in order to use the Kalman filtering algorithm. We estimate forward rates as solutions of the resulting filtering problem, and estimate model parameters by maximizing the derived likelihood functional.

2. A New Model for the Short Rate

Suppose that the short rate evolves according to the Hull-White model:

\[ dr(t) = \{\Theta(t) - ar(t)\}dt + \sigma_r dW_r(t) \]  

(2.1)

where \( \{W_r(t), t \geq 0\} \) is a scalar standard Brownian motion, \( \Theta(t) \) is a deterministic function of time, while \( a \) and \( \sigma_r \) are constants. It then follows that the bond price \( P(t, T), 0 < T < \hat{T} \), is given by

\[ P(t, T) = \exp \left\{ - \int_0^{T-t} [A(t, x) + B(t, x)r(t)]dx \right\}, \]  

(2.2)

where \( x \) denotes the time-to-maturity and the deterministic functions \( A(t, x) \) and \( B(t, x) \) are given in [5].

Let us now assume that the short rate does not exactly follow (2.1), but is very close to this model. This will cause the bond price to be somewhat perturbed from the formula given in (2.2). Suppose that this perturbed bond price has the following expression:

\[ P(t, T) = \exp \left\{ - \int_0^{T-t} [A(t, x) + B(t, x)r(t)]dx + \int_0^t \sigma dw(s, x + t - s)dx \right\} \]  

(2.3)

where \( r(t) \) follows equation (2.1) with \( \Theta(t) \) a random function of time, the constant \( \sigma \) denotes the magnitude of the perturbation and \( w(t, x) \) is a two parameter Brownian motion process represented by

\[ w(t, x) = \sum_{k=1}^{\infty} \frac{1}{\lambda_k} e_k(x) \beta_k(t) \]

where \( e_k \) is a sequence of differentiable functions forming an orthonormal basis in \( L(0, \hat{T}) \) and \( \{\beta_k(t)\} \) are mutually independent Brownian motion processes. In the sequel, we set \( H = L^2(0, \hat{T}) \) with the inner product \( (., .) \). Hence the Brownian motion
process \( w(t, \cdot) \) is regarded as the H-valued BMP with its incremental covariance operator \( Q \):
\[
E\{(\phi_1, w(t))(w(t), \phi_2)\} = (\phi_1, Q\phi_2)t, \quad \text{for } \phi_1, \phi_2 \in H
\]
with
\[
\text{Tr}\{Q\} = \sum_{k=1}^{\infty} \frac{1}{\lambda_k^2} < \infty.
\]
We also represent this kernel by
\[
Q = \int_T^0 q(x, y)(\cdot)dy.
\]
Hence we have
\[
\int_0^t \sigma dw(s, x + t - s) = \sum_{k=1}^{\infty} \int_0^t \sigma \frac{1}{\lambda_k} v_k(x + t - s) d\beta_k(s).
\]
The above stochastic integral term denotes the modeling error between the affine term structure constructed by \( A \) and \( Br \) and the true term structure.

Hence the usual forward process \( f(t, x) \) which is defined by \( P(t, T) = \exp\{-\int_0^{T-t} f(t, x)dx\} \) is given by
\[
f(t, x) = A(t, x) + B(t, x)r(t) + \int_0^t \sigma dw(s, x + t - s).
\]

See [1] for the general form of hyperbolic type formulation for \( f(t, x) \) and the structure of \( w(s, x + t - s) \).

**Theorem 2.1.** The factor process \( f(t, x) \) given in (2.6) can be decomposed as
\[
f(t, x) = f_r(t, x) + e^{-ax}R(t) + f_w(t, x),
\]
where
\[
\frac{\partial f_r(t, x)}{\partial t} = \frac{\partial f_r(t, x)}{\partial x} + q_a(x), \quad f_r(0, x) = f_o(x),
\]
\[
dR(t) = -aR(t)dt + \sigma_r dW_r(t), \quad R(0) = 0
\]
and
\[
\frac{\partial f_w(t, x)}{\partial t} = \frac{\partial f_w(t, x)}{\partial x} dt + \sigma dw(t, x), \quad f_w(0, x) = 0
\]
and where
\[
q_a(x) = \sigma^2 e^{-ax} \int_0^x e^{-ay}dy + \sigma^2 \int_0^x q(x, y)dy.
\]

**Proof.** Noting that the exact spot rate \( r^e(t) \) is given by \( f(t, 0) \), we have
\[
r^e(t) = A(t, 0) + B(t, 0)r(t) + \int_0^t \sigma dw(s, t - s),
\]
with \( r^r(0) = r(0) \). It should be noted that the \( r^r(t) \)-process belongs to \( L^2(\Omega; C([0, t_f]; R^1)) \) from the following estimate.

\[
E \left\{ \sup_{0 \leq s \leq t_f} \left| \int_0^t \sum_{k=1}^\infty \frac{1}{\lambda_k} c_k(t - s) d\beta_k(s) \right|^2 \right\} = \int_0^{t_f} \sum_{k=1}^\infty \left( \frac{1}{\lambda_k} c_k(t_f - s) \right)^2 ds \leq \sigma^2 T r(Q) t_f.
\]

Now we identify \( A \) and \( B \) by using the fact that the discount process \( \tilde{P}(t, T) = P(t, T) / \exp\{\int_0^t r^r(s) ds\} \) must be a martingale.

Noting that

\[
d \left[ \int_0^{T-t} \left\{ \int_0^t \sigma dw(s, x + t - s) \right\} dx \right] = - \left\{ \int_0^t \sigma dw(s, T - s) \right\} dt + \int_0^{T-t} d \left\{ \sum_{k=1}^\infty \int_0^t \frac{1}{\lambda_k} c_k(x + t - s) d\beta_k(s) \right\} dx \\
= - \left\{ \int_0^t \sigma dw(s, T - s) \right\} dt \\
+ \left\{ \int_0^{T-t} \left\{ \sum_{k=1}^\infty \frac{1}{\lambda_k} \frac{\partial c_k(x + t - s)}{\partial t} d\beta_k(s) \right\} dx \right\} dt \\
+ \sum_{k=1}^\infty \left\{ \int_0^{T-t} \sigma \frac{1}{\lambda_k} c_k(x) dx \right\} d\beta_k(t)
\]

(2.12)

and

\[
\int_0^{T-t} \left\{ \sum_{k=1}^\infty \int_0^t \frac{1}{\lambda_k} \frac{\partial c_k(x + t - s)}{\partial t} d\beta_k(s) \right\} dx \\
= \int_0^{T-t} \left\{ \sum_{k=1}^\infty \int_0^t \frac{1}{\lambda_k} \frac{\partial c_k(x + t - s)}{\partial x} d\beta_k(s) \right\} dx \\
= \sum_{k=1}^\infty \int_0^t \frac{1}{\lambda_k} (e_k(T - s) - e_k(t - s)) d\beta_k(s) \\
= \int_0^t \sigma dw(s, T - s) - \int_0^t \sigma dw(s, t - s),
\]

(2.13)

we have

\[
(2.12) = - \left\{ \int_0^t \sigma dw(s, t - s) \right\} dt + \sum_{k=1}^\infty \left\{ \int_0^{T-t} \sigma \frac{1}{\lambda_k} c_k(x) dx \right\} d\beta_k(t).
\]

(2.14)
By using Ito’s formula, the differential form of \( P(t, T) \) becomes

\[
\frac{dP(t, T)}{P(t, T)} = \left[ A(t, T - t) - \int_0^{T-t} \frac{\partial A(t, x)}{\partial t} dx \right.
\]

\[
- \int_0^{T-t} B(t, x) dx \Theta(t) + \frac{\sigma^2}{2} \left( \int_0^{T-t} B(t, x) dx \right)^2
\]

\[
+ \frac{1}{2} \sum_{k=1}^{\infty} \left( \int_0^{T-t} \sigma \frac{1}{\lambda_k} \epsilon_k(x) dx \right)^2 dt
\]

\[
+ \left[ B(t, T - t) - \int_0^{T-t} \frac{\partial B(t, x)}{\partial t} dx + \int_0^{T-t} B(t, x) dx a \right] r(t) dt
\]

\[
- \int_0^{T-t} B(t, x) dx \sigma_r dW(t) - \sum_{k=1}^{\infty} \left( \int_0^{T-t} \sigma \frac{1}{\lambda_k} \epsilon_k(x) dx \right) d\beta_k(t).
\]

The differential form of the discounted bond price \( \bar{P}(t, T) = P(t, T) / \exp \{ \int_0^t r(s) ds \} \) becomes

\[
\frac{d\bar{P}(t, T)}{\bar{P}(t, T)} = \left[ A(t, T - t) - A(t, 0) - \int_0^{T-t} \frac{\partial A(t, x)}{\partial t} dx \right.
\]

\[
- \int_0^{T-t} B(t, x) dx \Theta(t) + \frac{\sigma^2}{2} \left( \int_0^{T-t} B(t, x) dx \right)^2
\]

\[
+ \frac{1}{2} \sum_{k=1}^{\infty} \left( \int_0^{T-t} \sigma \frac{1}{\lambda_k} \epsilon_k(x) dx \right)^2 dt
\]

\[
+ \left[ B(t, T - t) - B(t, 0) - \int_0^{T-t} \frac{\partial B(t, x)}{\partial t} dx + \int_0^{T-t} B(t, x) dx a \right] r(t) dt
\]

\[
- \int_0^{T-t} B(t, x) dx \sigma_r dW(t) - \sum_{k=1}^{\infty} \left( \int_0^{T-t} \sigma \frac{1}{\lambda_k} \epsilon_k(x) dx \right) d\beta_k(t).
\]

Noting that

\[
A(t, T - t) - A(t, 0) = \int_0^{T-t} \frac{\partial A(t, x)}{\partial x} dx
\]

and

\[
B(t, T - t) - B(t, 0) = \int_0^{T-t} \frac{\partial B(t, x)}{\partial x} dx,
\]
we derive the following two equations for supporting the arbitrage free condition:

\[
\frac{\partial A(t, x)}{\partial t} = \frac{\partial A(t, x)}{\partial x} - B(t, x)\Theta(t) + \sigma^2 B(t, x) \int_0^x B(t, y) dy + \sigma^2 \int_0^x q(x, y) dy
\]  

(2.16)

and

\[
\frac{\partial B(t, x)}{\partial t} = \frac{\partial B(t, x)}{\partial x} + aB(t, x)
\]

with the boundary conditions:

\[
A(t, 0) = 0 \quad \text{and} \quad B(t, 0) = 1.
\]

It is easy to show that

\[
B(t, x) = e^{-ax},
\]

and

\[
\sum_{k=1}^{\infty} \int_0^T \sigma \frac{1}{\lambda_k} e_k(y) dy \sigma \frac{1}{\lambda_k} e_k(x) = \int_0^T q(y, x) dy.
\]

Hence the forward process \( f(t, x) \) satisfies the following stochastic partial differential equation:

\[
df(t, x) = \frac{\partial f(t, x)}{\partial x} dt + \left\{ \sigma^2 e^{-ax} \int_0^x e^{-ay} dy + \sigma^2 \int_0^x q(x, y) dy \right\} dt + e^{-ax} \sigma dW_r(t) + \sigma dw(t, x),
\]

(2.17)

\[
f(0, x) = f_o(x).
\]

(2.18)

By using the Ito formula, we find that the process \( f_w(t, x) = \int_0^t \sigma dw(s, x + t - s) \) is a solution of

\[
df_w(t, x) = \frac{\partial f_w(t, x)}{\partial x} dt + \sigma dw(t, x), f_w(0, x) = 0.
\]

(2.19)

Noting that \( f_r(t, x) \) and \( R(t) \) are solutions of (2.8) and (2.9), (2.7) can be derived.

**Proposition 2.1.** The function \( \Theta(t) \) can be represented by

\[
\Theta(t) = a f(0, t) + \frac{\partial f(0, t)}{\partial x} + e^{-at} \int_0^t \left[ \sigma^2 e^{-as} e^{-a(t-s)} + \sigma^2 ae^{at} \int_0^{t-s} q(s, t, x) dy \right. \\
\left. + \sigma^2 e^{-at} q(t-s, t-s) + \sigma^2 e^{-at} \int_0^{t-s} \frac{\partial q(t-s, y)}{\partial x} dy \right] ds.
\]

**Proof.** From (2.6), we have

\[
A(t_1, x) = f(t_1, x) - e^{-a(t_1-t)} \int_0^{t_1} \sigma dw(s, x + t_1 - s).
\]
Hence it follows from (2.16) that

$$A(t, x) = \left\{ f(t_1, x + t - t_1) - e^{-a(x + t - t_1)}r(t_1) - \int_0^{t_1} \sigma dw(s, x + t - s) \right\}$$

$$+ \int_{t_1}^{t} \left[ -e^{-a(x + t - s)} \Theta(s) + \sigma^2 e^{-a(x + t - s)} \int_0^{x + t - s} e^{-a y} dy \right.$$

$$\left. + \sigma^2 \int_0^{x + t - s} q(x + t - s, y) dy \right] ds.$$

(2.20)

It follows from $A(t, 0) = 0$ that

$$e^{a(t_1 - t)} - e^{at} f(t_1, t - t_1) + e^{at} \int_0^{t_1} \sigma dw(s, t - s)$$

$$= \int_{t_1}^{t} \left[ -e^{ax} \Theta(s) + \sigma^2 e^{as} \int_0^{x - s} e^{-a y} dy + \sigma^2 e^{at} \int_0^{t - s} q(t - s, y) dy \right] ds.$$  (2.21)

Differentiating the above equation with respect to $t$, we have

$$-ae^{at} f(t_1, t - t_1) - e^{at} \frac{\partial f(t_1, t - t_1)}{\partial x} + ae^{at} \int_0^{t_1} \sigma dw(s, t - s)$$

$$+ e^{at} \int_0^{t_1} \sigma d \left( \frac{\partial w(s, t - s)}{\partial x} \right)$$

$$= -e^{at} \Theta(t) + \int_{t_1}^{t} \left[ \sigma^2 e^{as} e^{-a(t - s)} + \sigma^2 e^{at} \int_0^{t - s} q(t - s, y) dy \right.$$

$$\left. + \sigma^2 e^{at} \int_0^{t - s} \frac{\partial q(t - s, y)}{\partial x} dy \right] ds$$

Taking a limit as $t_1 \to t$, we get

$$\Theta(t) = af(t, 0) + \frac{\partial f(t, 0)}{\partial x} - a \int_0^{t} \sigma dw(s, t - s) - \int_0^{t} \sigma d \left( \frac{\partial w(s, t - s)}{\partial x} \right).$$  (2.22)

We also have for $t_1 \to 0$,

$$\Theta(t) = af(0, t) + \frac{\partial f(0, t)}{\partial x} + e^{-at} \int_0^{t} \left[ \sigma^2 e^{as} e^{-a(t - s)} + \sigma^2 e^{at} \int_0^{t - s} q(t - s, y) dy \right.$$

$$\left. + \sigma^2 e^{at} q(t - s, y) + \sigma^2 e^{at} \int_0^{t - s} \frac{\partial q(t - s, y)}{\partial x} dy \right] ds.$$  

$\square$

3. The Observation Mechanism

In practice, we only observe the yield curve data from the market, and do not precisely know the forward rate process. The observed yield is given by

$$Y(t, T - t) = -\frac{1}{T - t} \log P(t, T).$$
Setting the time-to-maturity $\tau = T - t$ as constant, we have

$$Y(t, \tau) = -\frac{1}{\tau} \log P(t, t + \tau)$$

$$= \frac{1}{\tau} \int_0^\tau f(t, x) \, dx$$

$$= \frac{1}{\tau} \int_0^\tau \{f_r(t, x) + e^{-ax}R(t) + f_w(t, x)\} \, dx.$$  \hspace{1cm} (3.1)

Hence for $i = 1, 2, \ldots, m$

$$dY(t, \tau_i) = \frac{1}{\tau_i} \int_0^{\tau_i} (df_r(t, x) + e^{-ax}dR(t) + df_w(t, x)) \, dx,$$

$$= \frac{1}{\tau_i} [f_r(t, \tau_i) - f_r(t, 0)] \, dt + \frac{1}{\tau_i} \int_0^{\tau_i} q_a(x) \, dx \, dt$$

$$+ \frac{1 - e^{-\alpha \tau_i}}{\alpha \tau_i} [-aR(t) \, dt + \sigma_r dW_r(t)] + \frac{1}{\tau_i} [f_w(t, \tau_i) - f_w(t, 0)] \, dt$$

$$+ \frac{1}{\tau_i} \sum_{j=1}^\infty \frac{1}{\lambda_j} \int_0^{\tau_i} e_j(x) \, dx \, d\beta_j(t)$$

$$= \left[H_i \{f_r(t, \cdot) + f_w(t, \cdot\} - aG_i(a)R(t)\} \, dt + F_i(a) \, dt$$

$$+ G_i(a)\sigma_r dW_r(t) + \sigma K_i dW(t)\right]$$  \hspace{1cm} (3.2)

where $\forall \phi(x) \in C([0, \hat{T}])$ (continuous functions),

$$H_i \phi = (\phi(\tau_i) - \phi(0)) \frac{1}{\tau_i}$$

$$G_i(a) = \frac{1 - e^{-\alpha \tau_i}}{\alpha \tau_i}$$

$$F_i(a) = \frac{1}{\tau_i} \int_0^{\tau_i} q_a(x) \, dx$$

$$= \frac{1}{\tau_i} \int_0^{\tau_i} \left[\sigma_r^2 e^{-ax} \int_0^x e^{-ay} \, dy + \sigma^2 \int_0^x q(x, y) \, dy\right] \, dx$$

and

$$K_i dW(t) = \frac{1}{\tau_i} \sum_{j=1}^\infty \frac{1}{\lambda_j} \int_0^{\tau_i} e_j(x) \, dx \, d\beta_j(t).$$  \hspace{1cm} (3.3)

Now we set the $m$-dimensional observation;

$$\tilde{Y}_{(m)}(t) = [Y(t, \tau_1), Y(t, \tau_2), \ldots, Y(t, \tau_m)]'.$$  \hspace{1cm} (3.4)
The differential form of $\tilde{Y}_{(m)}(t)$ becomes
\[ d\tilde{Y}_{(m)}(t) = [H(f_r(t, \cdot) + f_w(t, \cdot)) - aG(a)R(t)]dt + F(a)dt \]
\[ + \sigma_r G(a)dW_r(t) + \sigma Kdw(t), \]
where
\[ Kdw(t) = \frac{1}{\tau_i} \left[ \sum_{k=1}^{\infty} \frac{1}{\lambda_k} \int_{0}^{\tau_i} e_k(x)dx \beta_k(t) \right]_{m \times 1} \]
with\(^1\)
\[ KQK^* = \left[ \sum_{k=1}^{\infty} \frac{1}{\lambda_i^2} \int_{0}^{\tau_i} e_k(x)dx \int_{0}^{\tau_j} e_k(x)dx \right]_{m \times m} \]
\[ = \left[ \frac{1}{\tau_i \tau_j} \int_{0}^{\tau_i} \int_{0}^{\tau_j} q(x,y)dx dy \right]_{m \times m}. \]

### 3.1. Observation in the real world

By the filtering problem we mean the minimum mean square estimate of $f(t,x)$ based on the knowledge of $Y_{(m)}(s); 0 \leq s \leq t$, [11, 12]. Our data used in the filtering for the factor process is a yield curve. From [8, 9] and our experimental studies for the US-bond data in [2], we have to include the market price of risk term in the factor model. Here these terms are assumed to be proportional to the square root of the noise covariances of the factor model. We specify the risk premium terms as
\[ \mu_r \sigma_r g_a(x) = \mu_r \sigma_r e^{-ax} \]
and
\[ \mu_q \sigma \tilde{q}_{1/2}(x) = \mu_q \sigma \sum_{i=1}^{M} \frac{1}{\lambda_i} e_i(x), \quad \text{for } M > 1. \]

Mathematically speaking, we construct new Brownian motion processes $\tilde{W}_r(t)$ and $\tilde{w}(t, x)$ by using the Girsanov’s theorem. See Appendix A for details. In the filtering problem, we need to change the terms $q_a(x)$ and $F(a)$ as stated below. Now we reset $q_a(x)$ as $q_{a\mu}(x)$:
\[ q_{a\mu}(x) = \sigma^2 e^{-ax} \int_{0}^{x} e^{-ay}dy + \sigma^2 \int_{0}^{x} q(x,y)dy - \mu_r \sigma_r e^{-ax} - \mu_q \sigma \sum_{i=1}^{M} \frac{1}{\lambda_i} e_i(x) \]
\[ = q_a(x) - \mu_r \sigma_r g_a(x) - \mu_q \sigma \tilde{q}_{1/2}(x), \]
where
\[ q_a(x) = e^{-ax} \quad \text{and} \quad \tilde{q}_{1/2}(x) = \sum_{i=1}^{M} \frac{1}{\lambda_i} e_i(x). \]

\(^1\) $A^*$ denotes the adjoint of the operator $A$. If $A$ is a matrix, $*$ denotes the transpose.
\( F(a) \) is also reset as \( F_{\alpha\mu} \)

\[
F_{\alpha\mu} = F(a) - \mu_r \sigma_r K g_a - \mu_q \sigma K \tilde{q}_{1/2}(x). \tag{3.9}
\]

Our objective now is to estimate the unknown system parameters. Our first difficulty is the covariance kernel \( q(x, y) \). If we can parameterize it with one or more parameter(s) say \( c \), then the parameters we need to estimate are \( a, \sigma_r, \sigma, \mu_r, \mu_q \) and \( c \). The obvious approach is to use the method of maximum likelihood, for which we need to calculate the likelihood functional from the observation data \( \{ \vec{Y}(m)(t); 0 \leq t \leq t_f \} \). Since the observation noise covariance \( \sigma_r^2 G(a) G^*(a) + \sigma^2 KQK^* \) is unknown, we do not have an obvious likelihood functional in our situation [3]. Since our model is linear and Gaussian, we may circumvent this problem by working with a quasi likelihood functional as proposed in [3]. We can, in fact, replace the observation noise covariance appearing as weight in the usual likelihood functional expression by any positive definite matrix which we take here to be the identity matrix \( I \). The noise covariance matrix does appear in the quasi likelihood functional through the filtered states and may be estimated from there.

The quasi likelihood functional for our problem is then

\[
QL(t, \vec{Y}(m), I) = \int_0^t \left( \begin{array}{c}
\hat{f}_r(s) \\
\hat{R}(s) \\
\hat{f}_w(s)
\end{array} \right) ^* d\vec{Y}(m)(s) - \frac{1}{2} \int_0^t \left\| \begin{array}{c}
\hat{f}_r(s) \\
\hat{R}(s) \\
\hat{f}_w(s)
\end{array} \right\| _{R_m}^2 \, ds, \tag{3.10}
\]

where \( \hat{f}_r(s), \hat{R}(s) \) and \( \hat{f}_w(s) \) are the minimum mean square estimates of the states \( f_r(s), R(s) \) and \( f_w(s) \) given by the observation data \( \sigma(\vec{Y}(m)(\tau); 0 \leq \tau \leq s) \) and \( \mathcal{H}(a) = [H, -aG(a), H] \).

The MLE of the unknown parameters is then given by

\[
\begin{bmatrix}
\hat{a} \\
\hat{\sigma}_r^2 \\
\hat{\sigma}_q^2 \\
\hat{\mu}_r \\
\hat{\mu}_q
\end{bmatrix} = \arg \max QL(t, \vec{Y}(m), I) \tag{3.11}
\]

where we set the function form of \( q(x, y; c) \) as in (2.4) and \( c \) is an unknown parameter.

4. The Filtering Problem

Before studying our filtering problem, we consider the usual affine term structure case (2.2). In this case the observation data (yield curve) becomes
for $i = 1, 2, \ldots, m$

$$
dY(t, \tau_i) = \left[ \frac{1}{\tau_i} \{ f_r(t, \tau_i) - f_r(t, 0) \} + \frac{1}{\tau_i} \int_0^{\tau_i} q_{\mu\mu}(x) dx - \frac{1 - e^{-a\tau_i}}{a\tau_i} aR(t) \right] dt \\
+ \sigma_r \frac{1 - e^{-a\tau_i}}{a\tau_i} dW_r(t).
$$

(4.1)

Hence the covariance matrix of the observation noise related to $W_r(t)$ becomes

$$
\sigma_r G(a) G^*(a) = \sigma_r^2 \left[ \frac{1 - e^{-a\tau_i}}{a\tau_i} \frac{1 - e^{-a\tau_j}}{a\tau_j} \right]_{m \times m}
$$

(4.2)

and we get

$$
\sigma_r^2 \det [G(a) G^*(a)] = 0.
$$

(4.3)

Due to this singularity of the covariance matrix (4.2) we can not apply the Kalman filtering theory \cite{11, 12} to the usual affine term structure case (2.2). Hence to avoid this difficulty, one usually adds an artificial small observation noise to the yield curve data as in \cite{8}. As mentioned before, this is not easy to justify.

From our formulation of the yield curve (3.5), the yield curve data itself contains the noise term $KQK^*$. Hence assuming the condition (4.4) below for the covariance kernel $q(x, y)$ of $Q$ (this condition is numerically checked by using the real data in Example 5.1 later), we derive the Kalman filter for estimating the forward rate process without adding any artificial observation noise.

Now we summarize the system and observation mechanism in the usual vector notation;

$$
d \begin{bmatrix} f_r(t, x) \\ R(t) \\ f_w(t, x) \end{bmatrix} = \begin{bmatrix} \frac{\partial f_r(t, x)}{\partial x} \\ -aR(t) \\ \frac{\partial f_w(t, x)}{\partial x} \end{bmatrix} dt + \begin{bmatrix} q_{\mu\mu}(x) \\ 0 \\ 0 \end{bmatrix} dt + \begin{bmatrix} 0 \\ \sigma_r dW_r(t) \\ \sigma_dW(t, x) \end{bmatrix},
$$

with

$$
dY_{(m)}(t) = H(a) \begin{bmatrix} f_r(t, \cdot) \\ R(t) \\ f_w(t, \cdot) \end{bmatrix} dt + F_{\mu\mu} dt + \sigma_r G(a) dW_r(t) + \sigma K dW(t, \cdot),
$$

where

$$
H(a) = [H, -aG(a), H].
$$

Under the following assumption (see [3]):

$$
\sigma_r^2 G(a) G^*(a) + \sigma^2 KQK^* > 0,
$$

(4.4)
we can derive the optimal filtering equations. Before writing down the optimal filtering equations, we first show the feasibility of the above assumption by using the US-treasury bond data.

Example 4.1. We will check the assumption (4.4) numerically by using the US-treasury bond data. As shown in Fig. 1, we used the 7-dimensional yield curve data, i.e., \( \vec{Y}_7(t) \). It is well known that \[ \lim_{n \to \infty} \frac{1}{n \Delta t} \sum_{i=1}^{n} (\vec{Y}_7(t_{i+1}) - \vec{Y}_7(t_i))(\vec{Y}_7(t_{i+1}) - \vec{Y}_7(t_i))^* = \sigma^2_r G G^* + \sigma^2 K Q K^* \text{ a.s.} \] (4.5)

In this numerical example, we set \( \Delta t = t_{i+1} - t_i = 0.0027 \text{ year} = 1 \text{ day} \). From (4.5), we get \( \sigma^2_r G G^* + \sigma^2 K Q K^* \) numerically as shown in Fig. 2. As the determinant of the obtained \( \sigma^2 G G^* + \sigma^2 K Q K^* \) matrix becomes positive, it is also possible to calculate the inverse of \( \sigma^2 G G^* + \sigma^2 K Q K^* \) matrix as is also shown in Fig. 3.

Usually we would construct the Kalman filter for \( f_r(t, x), R(t) \) and \( f_w(t, x) \). The process \( f_r(t, x) \) does not contain the random additive noise but the initial \( f_o(x) \) is random and unknown. However the filter gain equation then consists of 6 equations, which are not easy to apply in practical situations.

In this section, we present an empirical method for estimating the \( f_r(t, x) \)-process and separate out this process for the filtering algorithm.
Identification of Affine Term Structures 271

Fig. 2. Numerically obtained $\sigma^2_{\text{GG}} + \sigma^2_{\text{KQ}}$.

Fig. 3. Numerically obtained $\left(\sigma^2_{\text{GG}} + \sigma^2_{\text{KQ}}\right)^{-1}$.
4.1. Empirical estimation for $f_r(t, x)$

In practice, from the initial yield curve $\vec{Y}_m(0)$, one can construct the initial factor curve $f_o(x)$ [2]:

- Construct the whole yield curve $\hat{Y}(0, x)$ from $\vec{Y}_m(0)$ by using the spline interpolation method.
- The initial estimate for $f_o(x)$ is given by
  \[
  \hat{f}_o(x) = x \frac{d\hat{Y}(0, x)}{dx} + \hat{Y}(0, x).
  \]

Hence the estimate $\hat{f}_r(t, x)$ is a solution of
\[
\frac{\partial \hat{f}_r(t, x)}{\partial t} = \frac{\partial \hat{f}_r(t, x)}{\partial x} + q_{a\mu}(x), \quad \hat{f}_r(0, x) = \hat{f}_o(x).
\] (4.6)

4.2. Optimal estimates for $R(t)$ and $f_w(t, x)$

This is basically the Kalman filtering problem. In our filtering problem, the system and observation noises have a correlation:

\[
E \begin{bmatrix} \sigma_r & 0 \\ 0 & \sigma \end{bmatrix} d\begin{bmatrix} W_r(t, w(t, x)) \\ \sigma_r G(a) \sigma K \end{bmatrix} dt = \begin{bmatrix} \sigma^2_r G^*(a) \\ \sigma^2 Q K^* \end{bmatrix} dt.
\]

Under condition (4.4) it is possible to derive the filtering equation from [11] in p. 269. The optimal estimates for $R(t)$ and $f_w(t, x)$ are given by

\[
d \begin{bmatrix} \hat{R}(t) \\ \hat{f}_w(t, x) \end{bmatrix} = \begin{bmatrix} -a\hat{R}(t) \\ \frac{\partial \hat{f}_w(t, x)}{\partial x} \end{bmatrix} dt
+ \left( \mathcal{P}(t) \begin{bmatrix} -aG^*(a) \\ H^* \end{bmatrix} + \begin{bmatrix} \sigma_r^2 G(a) \\ \sigma^2 K \end{bmatrix} \right) (\sigma^2_r G^* G + \sigma^2 K Q K^*)^{-1}
\times \begin{bmatrix} d\hat{Y}_m(t) - \mathcal{H}(a) \hat{R}(t) \\ \hat{f}_w(t, \cdot) \end{bmatrix} dt - F_{a\mu} dt,
\] (4.7)

where $\hat{f}_r(t, x)$ is a solution of (4.6)

\[
Q K^* = \left[ \frac{1}{\tau_1} \int_{0}^{\tau_1} q(x, y) dy, \ldots, \frac{1}{\tau_m} \int_{0}^{\tau_m} q(x, y) dy \right],
\] (4.8)

\[
\mathcal{P}(t) = \begin{pmatrix} P_R(t) & P_{RW}(t) \\ P_{WR}(t) & P_w(t) \end{pmatrix},
\] (4.9)
The kernel equations are given by

\[
\frac{dp_R(t)}{dt} = -2ap_R(t) + \sigma_r^2 \left[ -p_R(t) \frac{1 - e^{-a\tau_i}}{\tau_i} + (p_R(t, \tau_i) - p_{Rw}(t, 0)) \frac{1}{\tau_i} \right. \\
+ \left. \sigma_r^2 \frac{1 - e^{-a\tau_i}}{\tau_i} \left( \sigma_r^2 GG^* + \sigma_r^2 KK^* \right)^{-1} \right] \times \left[ -p_R(t) \frac{1 - e^{-a\tau_j}}{\tau_j} + (p_{Rw}(t, \tau_j) - p_{Rw}(t, 0)) \frac{1}{\tau_j} \right. \\
+ \left. \sigma_r^2 \frac{1 - e^{-a\tau_j}}{\tau_j} \right]_{m \times 1},
\]

(4.10)

\[
\frac{\partial p_w(t, x, y)}{\partial t} = \frac{\partial p_w(t, x, y)}{\partial x} + \frac{\partial p_w(t, x, y)}{\partial y} + \sigma_r^2 q(x, y) \\
- \left[ -p_{Rw}(t, x) \frac{1 - e^{-a\tau_i}}{\tau_i} + (p_w(t, x, \tau_i) - p_w(t, x, 0)) \frac{1}{\tau_i} \right. \\
+ \left. \sigma_r^2 \int_0^{\tau_i} q(x, y)dy \right]_{1 \times m} \left( \sigma_r^2 GG^* + \sigma_r^2 KK^* \right)^{-1} \\
\times \left[ -p_{Rw}(t, y) \frac{1 - e^{-a\tau_j}}{\tau_j} + (p_w(t, \tau_j, y) - p_w(t, 0, y)) \frac{1}{\tau_j} \right. \\
+ \left. \sigma_r^2 \int_0^{\tau_j} q(x, y)dx \right]_{m \times 1},
\]

(4.11)

\[
\frac{\partial p_{Rw}(t, x)}{\partial t} = -ap_{Rw}(t, x) + \frac{\partial p_{Rw}(t, x)}{\partial x} - \left[ -p_R(t) \frac{1 - e^{-a\tau_i}}{\tau_i} + (p_R(t, \tau_i) \\
- p_{Rw}(t, 0)) \frac{1}{\tau_i} \right. \\
+ \left. \sigma_r^2 \frac{1 - e^{-a\tau_i}}{\tau_i} \left( \sigma_r^2 GG^* + \sigma_r^2 KK^* \right)^{-1} \right] \times \left[ -p_{Rw}(t, x) \frac{1 - e^{-a\tau_j}}{\tau_j} + (p_w(t, \tau_j, x) - p_w(t, 0, x)) \frac{1}{\tau_j} \right. \\
+ \left. \sigma_r^2 \int_0^{\tau_j} q(x, y)dy \right]_{m \times 1},
\]

(4.12)

with \( p_R(0) = p_w(0, x, y) = p_{Rw}(0, x) = 0 \).

**Remark 4.1.** From above results, we obtain the estimate of \( \Theta(t) \) from (2.22)

\[
\hat{\Theta}(t) = a \hat{f}_r(t, 0) + \frac{\partial \hat{f}_r(t, 0)}{\partial x},
\]

(4.13)
5. Simulation Studies

5.1. Simulated data

Before simulating the factor process \( f(t, x) \), we need to set the values of parameters \( \{a, \sigma, Q\} \). The covariance operator \( Q \) is set as

\[
\bar{\sigma}^2 Q(c)\phi = \sigma^2 \sum_{j=1}^{20} \frac{1}{j^3} \exp(-cj) \sin\left(\frac{i\pi x}{30}\right) \int_0^{30} \exp(-cy) \sin\left(\frac{i\pi y}{30}\right) \phi(y) dy
\]

\[
= \int_0^{30} \sigma^2 q(x, y; c)\phi(y) dy, \quad \forall \phi \in L^2(0, T).
\] (5.1)

In Example 5.1, we already got the value of \( \sigma^2 GG^* + \sigma^2 KQK^* \) from US bond data. In order to fit this value, we set true parameters as given in Table 1.

After generating the factor process \( f(t, x) \), (2.19), (2.8) and (2.9) are simulated where the market price of risk coefficients are set as

\[ \mu_r = 2.6, \quad \mu_q = 0.4 \]

in order to obtain the similar shape of the yield curve as shown in Fig. 1. The simulated yield curve data is shown in Fig. 4, with the forward process \( f(t, x) \) in Fig. 5.

<table>
<thead>
<tr>
<th>Table 1. True parameters.</th>
</tr>
</thead>
<tbody>
<tr>
<td>( a )</td>
</tr>
<tr>
<td>---------------</td>
</tr>
<tr>
<td>3.3114</td>
</tr>
</tbody>
</table>

Fig. 4. Simulated yield curves.
Identification of Affine Term Structures

5.2. MLE results

To realize the MLE for $a, \sigma_r, c$ and $\sigma$, we used the data obtained in Example 5.1 for $\sigma_r^2 GG^* + \sigma^2 KQK^*$. Setting

$$Cov(\bar{Y}) = \{\text{Obtained value of } \sigma_r^2 GG^* + \sigma^2 KQK^*\},$$

the quasi-likelihood $QL$ becomes $QL(t, \bar{Y}(m), Cov(\bar{Y}))$. To obtain the MLE numerically, we used the genetic algorithm in MATLAB GA toolbox. Setting

$t_f = \text{final time} = 0.027 \times 250,$

$2.0 \leq a \leq 4, \quad 0.1 \leq \sigma_r \leq 0.3,$

$0.1 \leq c \leq 0.2, \quad 0.5 \leq \sigma \leq 0.7,$

$2.0 \leq \mu_r \leq 3.0, \quad 0.2 \leq \mu_q \leq 0.5,$

we got the estimated parameters shown in Table 2. During the running time of GA-algorithm, we show the change of $QL$ in Fig. 6.

Remark 5.1. To obtain the MLE results above, the range intervals for the unknown variables are quite important and depend on the required CPU time. To perform this identification, first we used the standard statistical procedure for guessing the initial guesses. However the GA-algorithm experimentally works well for the wide intervals for unknown parameters but it takes a very long CPU time for computing these estimates. In this simulation study, we used the Windows 64-bit.
Fig. 6. Evolutions of quasi-likelihood.

Fig. 7. Estimated $f(t,x)$ process.
Table 2. Estimated parameters shown in Fig. 7.

<table>
<thead>
<tr>
<th>Variable #1</th>
<th>Variable #2</th>
<th>Variable #3</th>
<th>Variable #4</th>
<th>Variable #5</th>
<th>Variable #6</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\hat{\alpha}$</td>
<td>$\hat{\sigma}_r$</td>
<td>$\hat{\beta}$</td>
<td>$\hat{\sigma}$</td>
<td>$\hat{\rho}_r$</td>
<td>$\hat{\mu}$</td>
</tr>
<tr>
<td>3.4256</td>
<td>0.2674</td>
<td>0.1288</td>
<td>0.5592</td>
<td>2.7932</td>
<td>0.2095</td>
</tr>
</tbody>
</table>

PC with Intel Core 2 Extreme (CPU X9770, 320GHz) and 8.00 GB Ram. To get the above results, it took about 120 minutes to obtain these graphical outputs.

By using the estimated parameters, the estimated $f(t,x)$ process is demonstrated in Fig. 7. At the points $x = 0.1, 0.15, \text{ and } 1.6$, we also present the true and estimated $f(t,\cdot)$ processes in Figs. 8–10, respectively.

5.3. MLE for US-bond data

Encouraged by the good results of the preceding simulation studies, we apply our MLE algorithm to the real US-bond data. We used the same initial data used in Sec. 6.2. The estimates of the parameters are shown in Table 3. During the running time of GA-algorithm by using MATLAB GA-toolbox, we got the graphical output shown in Fig. 11 at each generation. Hence, we can stop this procedure, if we agree that the change of the fitness value does not move drastically.

The estimated $f(t,x)$, $f(t,0.1)$ and $f(t,19)$ are also shown in Figs. 12–14 respectively.

Fig. 8. True and estimated $f(t,0.1)$ processes.
Fig. 9. True and estimated $f(t,0.5)$ processes.

Fig. 10. True and estimated $f(t,1.6)$ processes.
Table 3. Estimated parameters for US-bond data in Fig. 11.

<table>
<thead>
<tr>
<th>Variable #1</th>
<th>Variable #2</th>
<th>Variable #3</th>
<th>Variable #4</th>
<th>Variable #5</th>
<th>Variable #6</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.9610</td>
<td>0.2311</td>
<td>0.1163</td>
<td>0.6988</td>
<td>2.0358</td>
<td>0.2036</td>
</tr>
</tbody>
</table>

6. Conclusion

In this paper a completely new approach is proposed for modeling interest rates. Traditionally, the effort is to come up with increasingly complicated models for short rates while retaining the exponential affine structure of the bond price. We argue instead that the real behavior of short rate may deviate slightly from some standard model used in the literature. This leads to a slightly perturbed bond price. We model this perturbation to take into account correlation of bond prices of different maturities. Imposing arbitrage free condition specifies our model.

The problem extensively studied in short rate models is the estimation of model parameters. One popular approach is to use the method of maximum likelihood which involves, for linear models, estimating the short rate by means of the Kalman filter. To apply the Kalman filter, one is forced to add artificial noises to bond prices. This is absolutely unnecessary in our approach. We study in details this parameter estimation problem using both simulated and real US treasury data.
We consider perturbation of the simple Hull-White model for short rates. We believe that the real issue is an appropriate perturbation of some short rate model, rather than the complication of the original short rate model.

It is well known that the Cox-Ingersoll-Ross model and the quadratic term structure model are suitable to capture the complex market behavior, see Geyer.
1. There is no serious technical complication in extending our approach to more complicated short rate models proposed by [7] and [10]. The parameter estimation problem, however, would then involve the nonlinear filter. In practice, this would involve approximating the filter by some classical techniques like extended Kalman filter, or else, use some particle filter algorithm.

**Acknowledgments**

The authors are grateful to Professor Farshid Jamshidian for his insightful comments, and for his pointing out an error in an earlier version. The authors also would like to thank an anonymous referee for some very useful suggestions and critical comments.

**Appendix**

In this appendix, we set the risk premium terms as the abstract form $\mu_r(t)$ and $\mu_r(t,x)$ which satisfy

$$E \left\{ \frac{1}{2} \int_0^t |\mu_r(s)|^2 \, ds \right\} < \infty$$

(A.1)

and

$$E \left\{ \frac{1}{2} \int_0^t \|\mu_g(s,\cdot)\|^2 \, ds \right\} < \infty$$

(A.2)
for every \( t \) and \( H = L^2(0, \hat{T}) \). Then we can define a Martingale measure \( \mathcal{P}_R \)(the real world measure);

\[
\frac{d\mathcal{P}_R}{d\mathcal{P}} = \exp \left\{ - \int_0^t \left( Q^{1/2} \mu_q(s, \cdot), dw(s) \right)_H - \int_0^t \mu_r(s) dW_r(s) \right. \\
- \left. \frac{1}{2} \int_0^t \left\{ ||Q^{1/2} \mu_q(s, \cdot)||_H^2 + |\mu_r(s)|^2 \right\} ds \right\}
\] (A.3)

where

\[
Q^{1/2} = \sum_{i=1}^{\infty} \frac{1}{\lambda_i} \phi_i(e_i, \cdot)
\]

and

\[
(\phi_1, \phi_2) = \int_0^{\hat{T}} \phi_1(x) \phi_2(x) dx.
\]

Hence under the measure \( \mathcal{P}_R \)

\[
\tilde{w}(t, \cdot) = w(t, \cdot) + \int_0^t Q^{1/2} \mu_q(s, \cdot) ds
\]

and

\[
\tilde{W}_r(t) = W_r(t) + \int_0^t \mu_r(s) ds
\]

are Brownian motion processes and we have

\[
df_{w\mu}(t, x) = \left[ \frac{\partial f_{w\mu}(t, x)}{\partial x} - \sigma \mu_q(t, x) \right] dt + \sigma d\tilde{w}(t, x), \quad f_{w\mu}(0, x) = 0 \quad (A.4)
\]

and

\[
dR_{\mu}(t) = -a R_{\mu}(t) dt - \sigma_r d\tilde{W}_r(t). \quad (A.5)
\]

Now we set the simple forms of these terms as

\[
\mu_r(t) = \mu_r \quad \text{(Constant)}
\]

and

\[
\mu_q(t, x) = \mu_q \sum_{i=1}^{M} e_i(x)
\]

where \( \mu_q \) and \( M \) are constants. Hence it is easy to show that

\[
||\mu_q(t, x)||_H^2 = \mu_q^2 M
\]

i.e., the condition (A.2) is satisfied.
References


