ESTIMATING EXPONENTIAL-AFFINE MODELS OF THE TERM STRUCTURE

MARK FISHER AND CHRISTIAN GILLES

Preliminary and Incomplete
Do not quote without permission

Comments welcome

Abstract. We show how to estimate any model of the term structure of interest rates in the affine-exponential class, which includes the Vasicek (1977), Cox, Ingersoll, Jr., and Ross (1985), and Longstaff and Schwartz (1992) models, among many others. For most models in this class, analytical expressions for both bond prices and the conditional distribution for the state variables are not available. However, there are (a) fast and accurate numerical solutions to the bond-price partial differential equation and (b) closed-form expressions for the first two conditional moments for the state variables. We show how to construct a quasi-maximum likelihood estimator using (a) and (b) based on the maximum likelihood estimator of Chen and Scott (1993). We discuss extensions to other estimation techniques.

1. Introduction

The term structure of interest rates contains information about the expected path of the short term rate. When the unbiased expectations hypothesis holds, for example, the curve of instantaneous forward rates is indistinguishable from the expected path of future short rates. The evidence clearly shows, though, that the forward premium (the difference between the forward rate and the expected future spot rate) does not equal zero, and does not even remain constant, so that the implications of the unbiased expectations hypothesis are soundly rejected. The challenge of the term structure modeling, then, is to account for the behavior of forward premiums, so that the term structure reveals its information about the expected path of the short rate.

Duffie and Kan (1995) characterize a class of arbitrage-free term structure models in which zero-coupon yields are affine functions of Markovian state variables. This exponential-affine class of models, which is both analytically and numerically tractable, can also account for the empirical behavior of forward premiums. We show how to estimate any model in this class, which includes Vasicek (1977), Cox,

Date: September 19, 1996.
JEL Classification. G12.
Key words and phrases. Exponential affine models, term structure, expectations hypothesis.
The views expressed herein are the authors' and do not necessarily reflect those of the Board of Governors or the Federal Reserve System. We thank Greg Duffee for helpful discussions.

Ingersoll, Jr., and Ross (1985), and Longstaff and Schwartz (1992) models, among many others. For most models in this class, analytical expressions for both bond prices and the conditional distribution for the state variables are not available. However, there are (a) fast and accurate numerical solutions to the bond-price partial differential equation (PDE) and (b) closed-form expressions for the first two conditional moments for the state variables. We show how to construct a quasi-maximum likelihood estimator using (a) and (b) based on the maximum likelihood estimator of Chen and Scott (1993). We discuss extensions to other estimation techniques.

2. Arbitrage-free models of the term structure

The state-price deflator. Let $V(t)$ be the value at time t of an asset. The absence of arbitrage implies the existence of a strictly positive *state-price deflator*, $m(t)$, such that for any $s \geq t$

$$m(t)\, V(t) = E_t[m(s)\, V(s)], \tag{2.1}$$

as long as the asset pays no dividends between t and s. We can think of the state-price deflator as the marginal utility of wealth. With this interpretation, the left-hand side of (2.1) is the marginal cost in utility terms of buying the asset at time t and the right-hand side is the expected marginal benefit in utility terms of selling the asset at time s. Thus, (2.1) is a first-order condition for equilibrium. The marginal-utility-weighted asset price, $z(t) := m(t)\, V(t)$, is also called the *deflated* asset price. Restating equation (2.1) in terms of the dynamics of deflated asset prices, the absence of arbitrage implies that $z(t)$ is a martingale; that is, the current value equals the expected future value, $z(t) = E_t[z(s)]$, and its expected change is zero.

Assume that the state price deflator can be represented as an Ito process:

$$\frac{dm(t)}{m(t)} = -r(t)\, dt - \lambda(t)^\top dW(t), \tag{2.2}$$

where $r(t)$ is a scalar, $\lambda(t)$ is a length-d column vector, and $W(t)$ is a length-d column vector of independent Brownian motions. Moreover, let us write the dynamics for (a strictly positive) $V(t)$ as follows:

$$\frac{dV(t)}{V(t)} = \mu_V(t)\, dt + \sigma_V(t)^\top dW(t). \tag{2.3}$$

Ito's lemma delivers the process for $z(t)$ in terms of (2.2) and (2.3) using Ito's lemma:

$$\frac{dz(t)}{z(t)} = \mu_z(t)\, dt + \sigma_z(t)^\top dW(t), \tag{2.4}$$

where

$$\mu_z(t) = \mu_V(t) - r(t) - \lambda(t)^\top \sigma(t) \tag{2.5}$$

and

$$\sigma_z(t) = \sigma_V(t) - \lambda(t).$$

Then (2.5) and the martingale property $\mu_z(t) = 0$ imply

$$\mu_V(t) = r(t) + \lambda(t)^\top \sigma(t). \tag{2.6}$$

We see that $r(t)$ is the short-term risk-free interest rate[1] and $\lambda(t)$ is the market price of risk.

Zero-coupon bonds. Let $p(t,T)$ be the value at time t of a default-free zero-coupon bond that pays one unit of account at time T; *i.e.*, $p(T,T) = 1$. Using (2.1), we can write the following expression for the term structure of interest rates:

$$p(t,T) = E_t\left[\frac{m(T)}{m(t)}\right]. \tag{2.7}$$

Now, write the dynamics for zero-coupon bond prices as follows:

$$\frac{dp(t,T)}{p(t,T)} = \mu_p(t,T)\,dt + \sigma_p(t,T)^\top dW(t). \tag{2.8}$$

Then, the absence-of-arbitrage condition (2.6) implies

$$\mu_p(t,T) = r(t) + \lambda(t)^\top \sigma_p(t,T). \tag{2.9}$$

Thus, modeling the term structure of interest rates in an arbitrage-free way can be reduced to modeling the dynamics of the state-price deflator, which in turn can be reduced to specifying the dynamics of $r(t)$ and $\lambda(t)$. In the next section, we write $r(t)$ and $\lambda(t)$ in terms of a vector of state variables and we specify the dynamics of those state variables.

Markovian state variable representations. Let $X(t)$ be a length-d vector of variables, and let the dynamics of $X(t)$ be given by

$$dX(t) = \mu_X(t)\,dt + \sigma_X(t)^\top dW(t). \tag{2.10}$$

A Markovian state variable representation of the term structure follows from specifying the short rate $r(t)$, the market price of risk $\lambda(t)$, the drift $\mu_X(t)$, and the diffusion $\sigma_X(t)$ as functions of the state vector$X(t)$, so we can write:

$$r(t) = \mathcal{R}(X(t)), \tag{2.11a}$$

$$\lambda(t) = \Lambda(X(t)), \tag{2.11b}$$

$$\mu_X(t) = \mu_X(X(t)) \tag{2.11c}$$

and

$$\sigma_X(t) = \sigma_X(X(t)) \tag{2.11d}$$

Under this specification, bond prices have the following form:

$$p(t,T) = P(X(t), T-t). \tag{2.12}$$

[1]Assume there is a money market account, $\beta(t)$, with drift $\mu_\beta(t)$ and no instantaneous risk; that is, its process is $d\beta(t) = \beta(t)\,\mu_\beta(t)\,dt$ and μ_β can be interpreted as the riskless short-term rate. Letting $V(t) = \beta(t)$ in (2.6) implies that $r(t) = \mu_\beta(t)$, so that the relative drift of $m(t)$ is indeed the negative of the riskless short rate.

Thus our strategy is this: Given $\mu_X(x)$, $\sigma_X(x)$, $\mathcal{R}(x)$, and $\Lambda(x)$, solve for $P(x,\tau)$ using the no-arbitrage condition (2.9).

Since bond prices are functions of the state vector, Ito's lemma yields analytical expressions for their drifts and diffusions. Let

$$P_\tau(x,\tau) := \frac{\partial\, P(x,\tau)}{\partial\, \tau},$$

$$P_x(x,\tau) := \begin{pmatrix} \frac{\partial\, P(x,\tau)}{\partial\, x_1} \\ \vdots \\ \frac{\partial\, P(x,\tau)}{\partial\, x_d} \end{pmatrix},$$

and

$$P_{xx}(x,\tau) := \begin{pmatrix} \frac{\partial^2\, P(x,\tau)}{\partial\, x_1^2} & \cdots & \frac{\partial^2\, P(x,\tau)}{\partial\, x_1\, \partial\, x_d} \\ \vdots & \ddots & \vdots \\ \frac{\partial^2\, P(x,\tau)}{\partial\, x_d\, \partial\, x_1} & \cdots & \frac{\partial^2\, P(x,\tau)}{\partial\, x_d^2} \end{pmatrix}.$$

Then the process (2.8) becomes

$$\frac{dp(t,T)}{p(t,T)} = \mu_P(X(t), T-t)\, dt + \sigma_P(X(t), T-t)^\top dW(t),$$

where

$$\mu_P(x,\tau) = \frac{\mu_X(x)^\top P_x(x,\tau) + \frac{1}{2}\, \mathrm{tr}\left[P_{xx}(x,\tau)\, \sigma_X(x)^\top \sigma_X(x)\right] - P_\tau(x,\tau)}{P(x,\tau)}, \tag{2.13a}$$

$$\sigma_P(x,\tau) = \frac{\sigma_X(x)\, P_x(x,\tau)}{P(x,\tau)}, \tag{2.13b}$$

and $\mathrm{tr}[a]$ is the trace of matrix a. In view of (2.13), the absence-of-arbitrage condition (2.9) turns into a partial differential equation:

$$\mathcal{R}(x) = \hat{\mu}_X(x)^\top \left(\frac{P_x(x,\tau)}{P(x,\tau)}\right) + \frac{1}{2}\, \mathrm{tr}\left[\left(\frac{P_{xx}(x,\tau)}{P(x,\tau)}\right) \sigma_X(x)^\top \sigma_X(x)\right] - \left(\frac{P_\tau(x,\tau)}{P(x,\tau)}\right), \tag{2.14}$$

where

$$\hat{\mu}_X(x) := \mu_X(x) - \sigma_X(x)^\top \Lambda(x). \tag{2.15}$$

Note that the drift μ_X and the price of risk Λ affect bond prices only through their effect on $\hat{\mu}_X$, which is often called the *risk-adjusted drift*. To justify this term, observe that $\hat{\mu}$ is the drift of the state variables that bond prices would imply under the assumption that the price of risk is zero. Duffie and Kan (1995) do not even introduce μ or Λ; instead, they postulate an affine form for the risk-adjusted drift $\hat{\mu}(x)$ directly.

The approach we have taken allows us to use the information in the whole term structure to identify the process for the short rate. This contrasts with the time series approach which uses only observations of the short rate process itself. The conditional expectation of $r(s)$ given the information at time $t \leq s$ is given by

$$E[r(s) \mid X(t)] = E[\mathcal{R}(X(s)) \mid X(t)].$$

We see that forecasting the short rate is tantamount to forecasting the state variables, which requires knowing the drift of $X(t)$, $\mu_X(x)$. Note that the risk-adjusted drift, which is sufficient to explain the shape of the yield curve at any point in time, is useless for forecasting yields; for this purpose, we need the actual drift. Our strategy requires a time series of term structures as panel data. The cross-sections (each yield curve) contains information about $\hat{\mu}_X$, σ_X, and R. The time series of any given yield (in particular, but not limited to, that of the short rate) contains information about μ_X and σ_X. Together they can identify all of the parameters in the model, including the parameters of Λ.

3. Exponential-affine term-structure models

First we review the setting so far. $X(t)$ is a length-d column vector of *state variables* or *factors*,[2] that obeys the following process:

$$dX(t) = \mu_X(X(t))\,dt + \sigma_X(X(t))^\top dW(t), \tag{3.1}$$

where $\mu_X(x)$ is a length-d column vector and $\sigma_X(x)$ is a $d \times d$ matrix. Let the short-term risk-free interest rate and the market price-of-risk vector depend on the state variables:

$$r(t) = \mathcal{R}(X(t))$$

and

$$\lambda(t) = \Lambda(X(t)).$$

Then the process for the state-price deflator is given by

$$\frac{dm(t)}{m(t)} = -\mathcal{R}(X(t))\,dt - \Lambda(X(t))^\top dW(t). \tag{3.2}$$

Now, define

$$\mathcal{M}(x) := \begin{pmatrix} \mathcal{R}(x) \\ \mu_X(x) \end{pmatrix} \quad \text{and} \quad \mathcal{S}(x) := \begin{pmatrix} \Lambda(x) & \sigma_X(x) \end{pmatrix},$$

where $\mathcal{M}(x)$ is $(d+1)\times 1$ and $\mathcal{S}(x)$ is $d \times (d+1)$. Then the resulting term structure model falls in the exponential-affine class if and only if

$$\mathcal{M}(x) \text{ and } \mathcal{S}(x)^\top \mathcal{S}(x) \text{ are affine in } x. \tag{3.3}$$

[2]We will use the two terms interchangeably.

In particular, we can write

$$\mathcal{R}(x) = r_0 + \sum_{i=1}^{d} r_i\, x_i \tag{3.4a}$$

$$\mu_X(x) = b_0 + \sum_{i=1}^{d} b_i\, x_i, \tag{3.4b}$$

$$\sigma_X(x)^\top \sigma_X(x) = G_0 + \sum_{i=1}^{d} G_i\, x_i, \tag{3.4c}$$

and

$$\sigma_X(x)^\top \Lambda(x) = h_0 + \sum_{i=1}^{d} h_i\, x_i, \tag{3.4d}$$

where the r_i are scalars, the b_i and h_i are $d \times 1$ vectors and the G_i are $d \times d$ matrices. Condition (3.3) is stronger than that given by Duffie and Kan, whose drift condition is only that $\hat{\mu}_X = \mu_X - \lambda^T \sigma_X$ be affine, whereas we require that both the actual drift μ_X and the risk adjustment $\lambda^T \sigma_X$ be affine. Under our conditions, the risk-adjusted drift is also affine, since

$$\hat{\mu}_X(x) = (b_0 - h_0) + \sum_{i=1}^{2} (b_i - h_i)\, x_i.$$

The central implication of teh affine structure summarized in (3.3) is that, setting aside the question of existence, bond prices have the form

$$P(x, \tau) = \exp\left(-A(\tau) - B(\tau)^\top x\right). \tag{3.5}$$

Using (3.5), the PDE (2.14), which is the condition for absence of arbitrage applied to bond prices, can be written in terms of the parameters of $\mathcal{M}$ and $\mathcal{S}^\top \mathcal{S}$. The building blocks for this specialization of the general PDE are

$$\frac{P_\tau(x,\tau)}{P(x,\tau)} = -A'(\tau) - B'(\tau)^\top x,$$

$$\frac{P_x(x,\tau)}{P(x,\tau)} = -B(\tau),$$

and

$$\frac{P_{xx}(x,\tau)}{P(x,\tau)} = B(\tau)\, B(\tau)^\top.$$

Plugging these expression into (2.14), there results

$$A'(\tau) + B'(\tau)^\top x - B(\tau)^\top \hat{\mu}_X(x) + \frac{1}{2}\, B(\tau)^\top \sigma_X(x)^\top \sigma_X(x)\, B(\tau) - \mathcal{R}(x) = 0. \tag{3.6}$$

The reason for imposing an affine structure now becomes clear. PDE's are notoriously difficult to solve, even numerically. But the PDE (3.6) decomposes into $d+1$ ordinary differential equations (ODEs). To see this, write (3.6) using the explicitly affine representations (3.4) and collect terms in x_i: The coefficient of each x_i, as well as the constant term, must be identically zero. Thus, we have

$$A'(\tau) = r_0 + B(\tau)^\top (b_0 - h_0) - \frac{1}{2}\, B(\tau)^\top G_0\, B(\tau) \tag{3.7a}$$

and

$$B_i'(\tau) = r_i + B(\tau)^\top (b_i - h_i) - \frac{1}{2}\, B(\tau)^\top G_i\, B(\tau) \tag{3.7b}$$

subject to

$$A(0) = B_i(0) = 0 \tag{3.7c}$$

for $i = 1$ to d, where the initial conditions follow from the requirement that $p(T,T) = 1$. It is the simple structure of (3.7) that gives exponential-affine models their tractability—both analytically and numerically. Notwithstanding this relative tractability, only a small fraction of potential models in this class have been solved analytically. Even for some models that have been solved analytically, the solution can be so complicated as to make implementation impracticable.[3] Fortunately, (3.7) is a set of first-order (quadratic) ODEs that can easily be solved by standard numerical techniques.

4. Maximum likelihood estimation using the conditional distribution of yields

In the previous section, we studied the link between the state variables and zero-coupon bond prices (or equivalently zero-coupon yields) at a single point in time. In this section, we use the dynamics of the state variables to establish the link between one yield curve and another over time; in fact, we obtain an analytical expression for the conditional factor transitions. We also briefly review a method in Pearson and Sun (1994), who show how to use these transitions to obtain maximum-likelihood estimates of multi-factor CIR models.

Start with the relationship between zero-coupon yields and the state variables. Define the yield to maturity as

$$y(t,T) := -\frac{\log(p(t,T))}{T-t}. \tag{4.1}$$

In view of (3.5), in the exponential-affine case yields to maturity satisfy:

$$y(t,T) = \frac{1}{T-t}\, \Big(A(T-t) + B(T-t)^\top X(t) \Big) .$$

Select a set of d distinct maturities $\{\tau_1, \cdots, \tau_d\}$, and consider the corresponding vector of yields at time t, $Y(t)$:

$$Y(t) := \begin{pmatrix} y(t, t+\tau_1) \\ \vdots \\ y(t, t+\tau_d) \end{pmatrix} .$$

[3]See Chen (1995) for an example of how complicated a closed-form solution can get.

The relationship between the vector of yields and state variables is given by

$$Y(t) = \mathcal{A} + \mathcal{B}\, X(t),$$

where

$$\mathcal{A} = \begin{pmatrix} A(\tau_1)/\tau_1 \\ \vdots \\ A(\tau_d)/\tau_d \end{pmatrix}$$

and

$$\mathcal{B} = \begin{pmatrix} B(\tau_1)/\tau_1 \\ \vdots \\ B(\tau_d)/\tau_d \end{pmatrix} = \begin{pmatrix} B_1(\tau_1)/\tau_1 & \cdots & B_d(\tau_1)/\tau_1 \\ \vdots & \ddots & \vdots \\ B_1(\tau_d)/\tau_d & \cdots & B_d(\tau_d)/\tau_d \end{pmatrix}.$$

Given observed yields, $Y(t)$, we can then easily solve for state variables:

$$X(t) = \mathcal{B}^{-1}\,(Y(t) - \mathcal{A}).$$

Suppose that the conditional distribution of the state variables is known and given by

$$f_X(X(s) \mid X(t)) \quad \text{for} \quad t \le s.$$

Note the distribution f_X involves the drift $\mu_X(x)$ itself rather than its risk-adjusted counterpart $\hat{\mu}_X(x)$. The conditional distribution for the yields follows directly from f_X and involves the Jacobian of the transformation from $Y(t)$ to $X(t)$. Since the transformation is

$$\mathcal{X}(y) := \mathcal{B}^{-1}\,(y - \mathcal{A}),$$

its Jacobian is

$$J = \det\left(\frac{\partial\,\mathcal{X}(y)}{\partial\, y}\right) = \det\left(\mathcal{B}^{-1}\right) = \frac{1}{\det(\mathcal{B})},$$

which implies that the conditional distribution of the yields is

$$f_Y(Y(s) \mid Y(t)) = \frac{1}{\det(\mathcal{B})}\; f_X(\mathcal{X}(Y(s)) \mid \mathcal{X}(Y(t))).$$

Given a set of observations at times $\{t_1, \cdots, t_n\}$, the log-likelihood function is given by

$$\mathcal{L} = \sum_{i=1}^{n} \log\left(f_Y(Y(t_i) \mid Y(t_{i-1}))\right), \tag{4.2}$$

where $t_0 = -\infty$ so that $f_Y(Y(t_1) \mid Y(t_0))$ is the unconditional distribution of $Y(t_1)$.

If the model were well-specified and yields observed without errors, the choice of maturities for the yields would not matter because any choice would imply the same time series for the factors. But in fact, the model is not necessarily well specified, and the assumption that yields are observed without errors is not strictly tenable. In large sample, the difference would be hardly noticeable, but in small samples different results would obtain when different maturities are used. Since there is information contained in each of the yields, efficiency requires taking it into account by using more yields than the number required to infer the level of the state variables. One approach that was implemented by Chen and Scott (1993) is to assume that the additional yields are measured with error. While the asymmetry entailed in this assumption is not well justified, computationally this provides a convenient framework.

Suppose, then, that there are m additional yields that are measured with error, besides the d yields that are observed without errors

$$\tilde{Y}(t) = \tilde{\mathcal{A}} + \tilde{\mathcal{B}}\,\mathcal{X}(Y(t)) + \varepsilon(t),$$

where

$$\tilde{Y}(t) = \begin{pmatrix} y(t, t+\tilde{\tau}_1) \\ \vdots \\ y(t, t+\tilde{\tau}_m) \end{pmatrix}.$$

Let the conditional distribution of measurement errors be given by

$$h(\varepsilon(t_i) \mid \varepsilon(t_{i-1})).$$

Then the log-likelihood function with measurement errors is given by

$$\mathcal{L}^* = \mathcal{L} + \sum_{i=1}^{n} \log\left(h(\varepsilon(t_i) \mid \varepsilon(t_{i-1}))\right). \tag{4.3}$$

5. Quasi-maximum likelihood

The technique outlined in the previous section requires knowing the conditional distribution for the state variables, f_X. In many cases, the conditional distribution is not known. Even in cases where there is an analytic solution to the bond price formula, there may not be an expression for the factor transitions.[4] Quasi-maximum likelihood, by contrast, requires knowing only the first two conditional moments. It can be justified on a GMM basis.

Define the multivariate normal distribution

$$\hat{f}_X(X(s) \mid X(t)) := (2\,\pi)^{-\frac{d}{2}}\,|V_X|^{-\frac{1}{2}} \times \exp\left\{-\frac{1}{2}\Big(X(s) - E_X\Big)^{\top} V_X^{-1}\Big(X(s) - E_X\Big)\right\},$$

where

$$E_X := E[X(s) \mid X(t)] \quad \text{and} \quad V_X := V[X(s) \mid X(t)].$$

Then let

$$\hat{f}_Y(Y(s) \mid Y(t)) := \frac{1}{\det(\mathcal{B})}\,\hat{f}_X(\mathcal{X}(Y(s)) \mid \mathcal{X}(Y(t))).$$

Replacing f_Y with $\hat{f}_Y$ in (4.2) and (4.3) and maximizing produces a quasi-maximum likelihood estimator. In order to make this estimator operational, we need expressions for the first two conditional moments of $X(t)$.

[4]See Chen (1996).

Conditional moments. Here we present closed-form solutions for the first and second conditional moments for affine state-variables. The solutions are general and do not depend on diagonalizability as in Duan and Simomato (1995).

Conditional expectation. Since $X(s) = X(t) + \int_{v=t}^{s} dX(v)$,

$$\begin{aligned} E_t[X(s)] &= X(t) + \int_{v=t}^{s} E_t[dX(v)] \\ &= X(t) + \int_{v=t}^{s} E_t[\mu_X(X(v))]\, dv. \end{aligned}$$

Referring to (3.4b), we can write $\mu_X(x) = b_0 + B\, x$, where $B := (b_1, \ldots, b_d)$, and thus $E_t[\mu_X(X(v))] = b_0 + B\, E[X(v) \mid X(t)]$. Then we can write

$$\hat{X}(t, s) = X(t) + \int_{v=t}^{s} [b_0 + B\, \hat{X}(t, v)]\, dv, \tag{5.1}$$

where

$$\hat{X}(t, s) := E[X(s) \mid X(t)].$$

Differentiating both sides of (5.1) with respect to s produces

$$\hat{X}'(t, s) = b_0 + B\, \hat{X}(t, s), \tag{5.2a}$$

subject to

$$\hat{X}(t, t) = X(t), \tag{5.2b}$$

where

$$\hat{X}'(t, s) := \frac{\partial}{\partial s} X(t, s).$$

Equation (5.2a) is a set of first-order, linear differential equations that must satisfy boundary condition (5.2b). The solution to the system (5.2) is

$$\hat{X}(t, T) = \mathit{\Phi}(T - t)\, X(t) + \mathcal{D}(T - t)\, b_0. \tag{5.3}$$

where

$$\mathit{\Phi}(\tau) := e^{B\, \tau} \tag{5.4}$$

is the fundamental matrix and

$$\mathcal{D}(\tau) := \int_0^{\tau} \mathit{\Phi}(s)\, ds. \tag{5.5}$$

Conditional variance. Let $V_t[Y]$ denote the conditional variance of Y, and let $v(t,T) := V_t[X(T)]$. We can derive an expression for $v(t,T)$ as follows. Applying Ito's lemma to (5.3) for fixed T, the dynamics of the conditional expectation are given by

$$d\hat{X}(t,T) = \hat{\sigma}_X(t,T)^\top dW(t), \tag{5.6}$$

where

$$\hat{\sigma}_X(t,T) := \sigma_X(X(t))\,\Phi(T-t)^\top.$$

Note that

$$X(T) = \hat{X}(t,T) + \int_{s=t}^{T} d\hat{X}(s,T) = \hat{X}(t,T) + \int_{s=t}^{T} \hat{\sigma}_X(s,T)^\top dW(s).$$

Thus

$$v(t,T) = V_t\left[\int_{s=t}^{T} \hat{\sigma}_X(s,T)^\top dW(s)\right] = E_t\left[\int_{s=t}^{T} \hat{\sigma}_X(s,T)^\top \hat{\sigma}_X(s,T)\,ds\right], \tag{5.7}$$

where the second equality is shown in Duffie (1996).[5] Using (3.4c), we can write

$$E_t\left[\hat{\sigma}_X(s,T)^\top \hat{\sigma}_X(s,T)\right] = \Phi(T-s)\,F(t,s)\,\Phi(T-s)^\top,$$

where,

$$F(t,s) := G_0 + \sum_{\ell=1}^{d} G_\ell\,\hat{X}_\ell(t,s). \tag{5.8}$$

Therefore, we can write

$$v(t,T) = \int_t^T \Phi(T-s)\,F(t,s)\,\Phi(T-s)^\top\,ds. \tag{5.9}$$

See Fisher and Gilles (1996) for further results.

6. Yield factor models

Let us consider models where the factors are yields themselves. This is discussed in Duffie and Kan (1995), which in fact is titled "A yield factor model of interest rates."

Given $Y(t) = \mathcal{A} + \mathcal{B}\,X(t)$, we have

$$dY(t) = \mathcal{B}\,dX(t) = \mu_Y(Y(t))\,dt + \sigma_Y(Y(t))^\top dW(t),$$

where $\mu_Y(y)$ and $\sigma_Y(y)^\top \sigma_Y(y)$ are affine in y.[6]

[5] Equation (5.7) is a generalization of Duffie's equation (1) on page 84. He cites Protter (1990) for a proof.

[6] See the Appendix.

We can eliminate $X(t)$ from the formula for bond prices:

$$\begin{aligned} P_Y(y,\tau) &:= P(\mathcal{X}(y),\tau) \\ &= \exp\Big(-A(\tau) - B(\tau)^\top \mathcal{X}(y)\Big) \\ &= \exp\Big(-A_Y(\tau) - B_Y(\tau)^\top y\Big), \end{aligned}$$

where

$$A_Y(\tau) = A(\tau) - B(\tau)^\top \mathcal{B}^{-1}\,\mathcal{A}$$

and

$$B_Y(\tau)^\top = B(\tau)^\top \mathcal{B}^{-1}.$$

We still have the restrictions $A_Y(0) = 0$ and $B_Y(0) = 0$ (from $p(T,T) = 1$). But there are $2\,d$ *additional* restrictions that arise from the following relationship:

$$y(t, t+\tau) = \frac{1}{\tau}\Big(A_Y(\tau) + B_Y(\tau)^\top Y(t)\Big). \tag{6.1}$$

For each τ_i in the set of maturities that makes up $Y(t)$, (6.1) is an identity that implies

$$A_Y(\tau_i) = 0 \quad \text{and} \quad \frac{B_Y(\tau_i)}{\tau_i} = e_i,$$

where e_i is a vector of zeros with a one in the i-th position. These additional restrictions make it difficult to solve the PDE.

7. Previous and other estimation techniques

Exponential-affine models:

- Brown and Dybvig (1986). Nonlinear least squares estimates of one-factor CIR short rate using cross-section information only.
- Brown and Schaefer (1994) and Brown and Schaefer (1996). Nonlinear least squares estimates of one-factor CIR *real* short rate using cross-section information only.
- Gibbons and Ramaswamy (1993). GMM estimates of one-factor CIR short rate model using unconditional moments.
- Pearson and Sun (1994). Maximum likelihood estimates of two-factor CIR using bond prices to infer state variables.
- Chen and Scott (1993). Maximum likelihood estimates of one-, two-, and three-factor CIR, extending Pearson and Sun to include yields measured with error.
- Ball and Torous (1996). Uses Kalman filter estimates of one-factor CIR short rate to overcome unit root problems.
- Fisher and Gilles (1996). GMM estimates based on Campbell–Shiller-type regression moments.

Other (univariate and time series approaches):

- Chan, Karolyi, Longstaff, and Sanders (1992).

- Pagan, Hall, and Martin (1995). Review time series literature on the term structure and compare with finance approach.

Kalman Filter. Assume *all* yields are measured with error. We can write the *measurement equation*:

$$Y(t) = \mathcal{A} + \mathcal{B}\, X(t) + \varepsilon(t),$$

where is a $(d+m)$-dimensional vector. Given

$$\hat{X}(t_i, t_{i+1}) = E_{t_i}[X(t_{i+1})]$$

and

$$v(t_i, t_{i+1}) = V_{t_i}[X(t_{i+1})],$$

we can write the *transition equation*:

$$X(t_{i+1}) = \hat{X}(t_i, t_{i+1}) + v(t_i, t_{i+1})^{1/2}\, \eta(t_{i+1}),$$

where $\eta(t_{i+1})$ is a vector of zero-mean and unit variance error terms and $v(t_i, t_{i+1})^{1/2}$ is the Cholesky decomposition of $v(t_i, t_{i+1})$. For a Gaussian exponential-affine model, the Kalman filter provides the optimal solution to prediction, updating, and evaluating the likelihood. For a non-Gaussian model the Kalman filter delivers inconsistent parameter estimates and how to proceed is an open question

See Duan and Simomato (1995), Chen and Scott (1993), and Buraschi (1996) for implementations of the Kalman filter.

Appendix A. The affine structure in more detail

Duffie and Kan (1995) showed that if $\sigma_X(x)^\top \sigma_X(x)$ is affine in x, then

$$\sigma_X(x)^\top \sigma_X(x) = \Omega^\top D(x)\, \Omega, \tag{A.1}$$

where Ω is an invertible constant $d \times d$ matrix and $D(x)$ is a diagonal matrix,

$$D(x) = \begin{pmatrix} u_1(x) & 0 & \cdots & 0 \\ 0 & u_2(x) & \cdots & 0 \\ \vdots & & \ddots & \vdots \\ 0 & \cdots & 0 & u_d(x) \end{pmatrix},$$

where

$$u_i(x) = \alpha_i + \beta_i^\top x.$$

Moreover, we can write

$$D(x) = D_0 + \sum_{i=1}^{d} D_i\, x_i,$$

where each D_i is a positive semi-definite $d \times d$ matrix. Thus we can write $G_i = \Omega^\top D_i\, \Omega$. Also, given (A.1), we can write $H_i = \Omega^\top D_i\, \delta$ for some constant vector δ. Note that an implication of this affine structure is that

$$\Big(\sigma_X(x)^\top \sigma_X(x)\Big)^{-1} \sigma_X(x)^\top \Lambda(x) = \Omega^{-1}\, \delta \tag{A.2}$$

is independent of x.

Given $Y(t) = \mathcal{A} + \mathcal{B}\, X(t)$, we have

$$dY(t) = \mathcal{B}\, dX(t) = \mu_Y(Y(t))\, dt + \sigma_Y(Y(t))^\top dW(t),$$

where

$$\begin{aligned} \mu_Y(y) &= \mathcal{B}\, \mu_X(\mathcal{X}(y)) \\ &= \mathcal{B}\left(a - b\, \mathcal{B}^{-1}\, \mathcal{A}\right) + \mathcal{B}\, b\, \mathcal{B}^{-1}\, y \\ &= a_Y + b_Y\, y \end{aligned}$$

and

$$\begin{aligned} \sigma_Y(y)^\top \sigma_Y(y) &= \mathcal{B}\, \sigma_X(\mathcal{X}(y))^\top \sigma_X(\mathcal{X}(y))\, \mathcal{B}^\top \\ &= \Big(\mathcal{B}\, \Omega^\top\Big)\, D(\mathcal{X}(y))\, \Big(\Omega\, \mathcal{B}^\top\Big) \\ &= \Omega_Y^\top D_Y^2(y)\, \Omega_Y, \end{aligned}$$

where

$$u_{Yi}(y) = \left(\alpha_i - \beta_i\, \mathcal{B}^{-1} \mathcal{A}\right) + \beta_i\, \mathcal{B}^{-1} y = \alpha_{Yi} + \beta_{Yi}\, y.$$

We see that dynamics for $Y(t)$ is affine in $Y(t)$ just as the dynamics for $X(t)$ is affine in $X(t)$.

References

Ball, C. A. and W. N. Torous (1996). Unit roots and the estimation of interest rate dynamics. *Journal of Empirical Finance 3*, 215–238.

Brown, R. and S. Schaefer (1994). The term structure of real interest rates and the cox, ingersoll, and ross model. *Journal of Financial Economics 35*, 3–42.

Brown, R. and S. Schaefer (1996). Ten years of the real term structure: 1984–1994. *Journal of Fixed Income* March.

Brown, S. and P. Dybvig (1986). The empirical implications of the Cox, Ingersoll, Ross theory of the term structure of interest rates. *Journal of Finance 51*, 617–630.

Buraschi, A. (1996). How large is the inflation risk premium in the u.s. nominal term structure? IFA Working Paper 232–1996, Institute of Finance and Accounting, London Business School.

Chan, K., A. Karolyi, F. A. Longstaff, and A. Sanders (1992). An empirical comparison of alternative models of the short-term interest rate. *Journal of Finance 47*(3), 1209–1227.

Chen, L. (1996). *Interest rate dynamics, derivatives pricing, and risk management*, Volume 435 of *Lecture Notes in Economics and Mathematical Systems*. New York: Springer-Verlag.

Chen, R.-R. (1995). A two-factor, preference-free model for interest rate sensitive claims. *The Journal of Futures Markets 15*, 345–372.

Chen, R.-R. and L. Scott (1993). Maximum likelihood estimation for a multifactor equilibrium model of the term structure of interest rates. *Journal of Fixed Income* December, 14–31.

Cox, J. C., J. E. Ingersoll, Jr., and S. A. Ross (1985). A theory of the term structure of interest rates. *Econometrica 53*(2), 385–407.

Duan, J.-C. and J.-G. Simomato (1995). Estimating and testing exponential-affine term structure models by Kalman filter. Working paper, McGill University.

Duffie, D. (1996). *Dynamic Asset Pricing Theory* (second ed.). Princeton, New Jersey: Princeton University Press.

Duffie, D. and R. Kan (1995). A yield-factor model of interest rates. Working paper, Graduate School of Business, Stanford University. Forthcomming: *Mathematical Finance*.

Fisher, M. and C. Gilles (1996). Term permia in exponential-affine models of the term structure. Working paper, Board of Governors of the Federal Reserve System, Washington, DC.

Gibbons, M. and K. Ramaswamy (1993). A test of the Cox, Ingersoll, and Ross model of the term structure. *Review of Financial Studies 6*(3), 619–658.

Longstaff, F. A. and E. Schwartz (1992). Interest rate volatility and the term structure: A two-factor general equilibrium model. *Journal of Finance 47*, 1259–1282.

Pagan, A. R., A. D. Hall, and V. Martin (1995). Modeling the term structure. Working paper, Australian National University.

Pearson, N. and T.-S. Sun (1994). Exploiting the conditional density in estimating the term structure: An application to the Cox, Ingersoll, and Ross model. *Journal of Finance 49*(4), 1279–1304.

Protter, P. (1990). *Stochastic Integration and Differential Equations*. New York: Springer-Verlag.

Vasicek, O. (1977). An equilibrium characterization of the term structure. *Journal of Financial Economics 5*, 177–188.

(Mark Fisher) Research and Statistics, Board of Governors of the Federal Reserve System, Washington, DC 20551

E-mail address: mfisher@frb.gov

(Christian Gilles) Monetary Affairs, Board of Governors of the Federal Reserve System, Washington, DC 20551

E-mail address: cgilles@frb.gov