Continuous Time Finance Notes, Spring 2004 – Section 4, *Feb. 18, 2004*

Notes by Robert V. Kohn, Courant Institute of Mathematical Sciences. For use in connection with the NYU course Continuous Time Finance.

This section begins our discussion of interest rate models. You should already have some familiarity with basic terminology (e.g. bond prices, instantaneous forward rates), instruments (e.g. caps and captions, swaps and swaptions), and their valuation using Black's formula. If you need to review these topics, Hull is excellent; see also my Derivative Securities lecture notes, sections 10 and 11 (warning: the notation there is slightly different from the present notes). Today we'll get started by discussing relatively simple one-factor models of the short rate (Vasicek, Cox-Ingersoll-Ross). Next time we'll discuss the Hull-White (also known as modified Vasicek) model, a richer one-factor short rate model that can be calibrated to an arbitrary initial yield curve. After that we'll turn to the Heath-Jarrow-Morton theory.

I've adopted Baxter & Rennie's notation. However my pedagogical strategy is different from theirs. I'm starting with short-rate models, because they're easier. Baxter & Rennie start with HJM, specializing afterward to short-rate models, because it's more efficient. The Vasicek model is by now quite standard; treatments can be found (with slightly different organization and viewpoints) in Brigo & Mercurio, Lamberton & Lapeyre, and Avellaneda & Laurence. For a good survey of the "big picture," I recommend reading chapter 1 of R. Rebonato, *Modern pricing of interest-rate derivatives: the Libor market model and beyond* (2nd edition, 2002, on reserve in the CIMS library).

****

Basic terminology. The time-value of money is expressed by the *discount factor*

$$P(t, T) = \text{value at time } t \text{ of a dollar received at time } T.$$

This is, by its very definition, the price at time t of a zero-coupon (risk-free) bond which pays one dollar at time T. If interest rates are stochastic then $P(t, T)$ will not be known until time t; its evolution in t is random, and can be described by an SDE. Note however that $P(t, T)$ is a function of *two* variables, the initiation time t and the maturity time T. The dependence on T reflects the *term structure* of interest rates; we expect $P(t, T)$ to be relatively smooth as a function of T, since interest is being averaged over the time interval $[t, T]$. We usually take the convention that the present time is $t = 0$ – thus what is observable now is $P(0, T)$ for all $T > 0$.

There are several equivalent ways to represent the time-value of money. The *yield* $R(t, T)$ is defined by

$$P(t, T) = e^{-R(t,T)(T-t)};$$

it is the unique constant interest rate that would have the same effect as $P(t, T)$ under continuous compounding. Evidently

$$R(t, T) = -\frac{\log P(t, T)}{T - t}.$$

The *instantaneous forward rate* $f(t,T)$ is defined by

$$P(t,T) = e^{-\int_t^T f(t,\tau)\,d\tau};$$

fixing t, it is the deterministic time-varying interest rate that describes all the loans starting at time t with various maturities. Clearly

$$f(t,T) = -\frac{\partial \log P(t,T)}{\partial T}.$$

The *instantaneous interest rate* $r(t)$, also called the short rate, is

$$r(t) = f(t,t);$$

it is the rate earned on the shortest-term loans starting at time t. Notice that the yields $R(t,T)$ and the instantaneous forward rate $f(t,T)$ carry the same information as the entire family of discount factors $P(t,T)$; however the short rate $r(t)$ carries much less information, since it is a function of just one variable.

Why is $f(t,T)$ called the instantaneous forward rate? To explain, we start by observing that the ratio

$$P(0,T)/P(0,t)$$

has an important interpretation: it is the discount factor (for time-t borrowing, with maturity T) that can be locked in at time 0, at no cost, by a combination of market positions. In fact, consider the following portfolio:

(a) long a zero-coupon bond worth one dollar at time T (present value $P(0,T)$), and

(b) short a zero-coupon bond worth $P(0,T)/P(0,t)$ at time t (present value $-P(0,T)$).

Its present value is 0, and its holder pays $P(0,T)/P(0,t)$ at time t and receives one dollar at time 0. Thus the holder of this portfolio has "locked in" $P(0,T)/P(0,t)$ as his discount rate for borrowing from time t to time T. This ratio is called the *forward term rate* at time 0, for borrowing at time t with maturity T. Similarly, $P(t,T_2)/P(t,T_1)$ is the forward term rate at time t, for borrowing at time T_1 with maturity T_2. The associated yield is

$$-\frac{\log P(t,T_2) - \log P(t,T_1)}{T_2 - T_1}.$$

In the limit $T_2 - T_1 \to 0$ we get $-\partial \log P(t,T)/\partial T = f(t,T)$.

What's our goal? As usual, we want a no-arbitrage-based framework for pricing and hedging options. A simple example would be a call option on a zero-coupon bond. If the option's holder has the right to purchase a zero-coupon bond at time t, paying one dollar at maturity T, for price K, then his payoff at time t is $(P(t,T) - K)_+$. The task of pricing a caplet is almost the same: consider an option that caps the term interest rate $\mathcal{R}$ for lending between times T_1 and T_2 at rate R_0. For a loan with principal one dollar, the option's payoff at time T_2 is

$$\Delta t \max\{\mathcal{R} - R_0, 0\}$$

where $\Delta t = T_2 - T_1$ and $\mathcal{R}$ is the actual term interest rate in the market at time T_1 (defined by $P(T_1, T_2) = 1/(1 + \mathcal{R}\Delta t)$). The discounted value of this payoff at time T_1 is

$$(1 + R_0\Delta t) \max \left\{ \frac{1}{1 + R_0\Delta t} - \frac{1}{1 + \mathcal{R}\Delta t}, 0 \right\}$$

since

$$1 - \frac{1 + R_0\Delta t}{1 + \mathcal{R}\Delta t} = \frac{(\mathcal{R} - R_0)\Delta t}{1 + \mathcal{R}\Delta t}.$$

So the caplet is equivalent to $(1 + R_0\Delta t)$ put options on a zero-coupon bond at time T_1, paying one dollar at maturity T_2, with strike price $1/(1 + R_0\Delta t)$. A cap consists of a collection of caplets; it is equivalent by this argument to a portfolio of puts on zero-coupon bonds.

Let's step back a minute. If this is the first time you think about interest-based instruments, you may wonder how the interest rate can be simultaneously risk-free and random. For short-rate models this is easy to understand. We already saw the continuous-time version on HW1, problem 1: if, under the risk-neutral probability, the short rate r solves a stochastic differential equation of the form $dr = \alpha(r, t)\, dt + \beta(r, t)\, dw$ then the value at time t of a dollar received at time T is

$$P(t, T) = E\left[e^{-\int_t^T r(s)\, ds} \mid \mathcal{F}_t \right].$$

Moreover $P(t, T)$ can be determined by solving the PDE

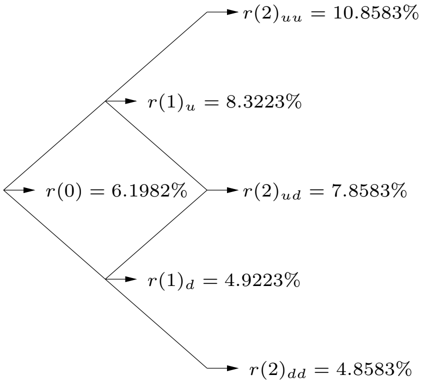
$$V_t + \alpha V_r + \tfrac{1}{2}\beta^2 V_{rr} - rV = 0$$

with final-time condition $V(T, r) = 1$ for all r at $t = T$; the value of $P(t, T)$ is then $V(t, r(t))$.

To make this very concrete, let's briefly discuss how something similar can be done with a binomial tree. (My treatment follows one in the textbook by Jarrow & Turnbull.) The basic idea is shown in the figure: each node of the tree is assigned a risk-free rate, different from node to node; it is the one-period risk-free rate for the binomial subtree just to the right of that node.

What probabilities should we assign to the branches? It might seem natural to start by figuring out what the subjective probabilities are. But why bother – what we need for option pricing are the risk-neutral probabilities. Moreover recall that there is always some freedom in designing a binomial tree to fit a specified stochastic process; the continuum limit is governed by the central limit theorem, so what matters is that we get the means and variances right. We can therefore take the convention that the risk-neutral probability of going up at any branch is $q = 1/2$, then choose the "drifts" and "volatilities" implicit in the nodes to mimic the desired stochastic process. This procedure leads to the Black-Derman-Toy framework for valuing interest-based instruments; briefly, the approach involves

- restricting attention to the risk-neutral interest rate process;
- assuming the risk-neutral probability is $q = 1/2$ at each branch; and



- choosing the interest rates at the various nodes so that the values of $P(0,T)$ obtained using the tree match those observed in the marketplace for all T.

The last bullet – the *calibration* of the tree to market information – is of course crucial. It will be discussed at length in the class Interest Rate and Credit Models. Here we won't try to do any calibration. But let's be sure we understand how the tree determines $P(0,T)$ for all T. As an example let's determine $P(0,3)$, the value at time 0 of a dollar received at time 3, for the tree shown in the figure. We take the convention that the interest rates shown in the tree are rates per time period (equivalently: we suppose the time period is $\Delta t = 1$).

Consider first time period 2. The value at time 2 of a dollar received at time 3 is $P(2,3)$; it has a different value at each time-2 node. These values are computed from the fact that

$$P(2,3) = e^{-r\Delta t}[\tfrac{1}{2}P(3,3)_{\text{up}} + \tfrac{1}{2}P(3,3)_{\text{down}}] = e^{-r\Delta t}$$

so

$$P(2,3) = \begin{cases} e^{-r(2)_{uu}} = .897104 & \text{at node } uu \\ e^{-r(2)_{ud}} = .924425 & \text{at node } ud \\ e^{-r(2)_{dd}} = .952578 & \text{at node } dd. \end{cases}$$

Now we have the information needed to compute $P(1,3)$, the value at time 1 of a dollar received at time 3. Applying the rule

$$P(1,3) = e^{-r\Delta t}[\tfrac{1}{2}P(2,3)_{\text{up}} + \tfrac{1}{2}P(2,3)_{\text{down}}]$$

at each node gives

$$P(1,3) = \begin{cases} e^{-r(1)_u}(\frac{1}{2}\cdot .897104 + \frac{1}{2}\cdot .924425) = .838036 & \text{at node } u \\ e^{-r(1)_d}(\frac{1}{2}\cdot .924425 + \frac{1}{2}\cdot .952578) = .893424 & \text{at node } d. \end{cases}$$

Finally we compute $P(0,3)$ by applying the same rule:

$$\begin{aligned} P(0,3) &= e^{-r\Delta t}[\tfrac{1}{2}P(1,3)_{\text{up}} + \tfrac{1}{2}P(1,3)_{\text{down}}] \\ &= e^{-r(0)}[\tfrac{1}{2}\cdot .838036 + \tfrac{1}{2}\cdot .893424] = .8137. \end{aligned}$$

****

General orientation. Let's briefly review our understanding of options on equities in a constant-interest-rate environment. The underlying is assumed to solve a stochastic differential equation; in the simplest one-factor setting the model is determined by specifying the drift and volatility. They determine a unique risk-neutral measure, and (equivalently) a unique market price of risk. Every option – indeed, every tradeable – has a unique price, forced upon us by the existence of replicating portfolios and the principle of no-arbitrage. Do the prices we get this way match the marketplace? Well, that depends what we use for the volatility. We often use a constant volatility, e.g. the implied volatility for at-the-money options; the prices obtained this way do *not* match market prices exactly (the implied volatility of an option depends, in practice, on its strike and maturity). Though imperfect, the theory is nevertheless useful, because it gives approximate hedge portfolios. To do better, one can make the theory more complicated – e.g. by looking for a time-and-price-dependent volatility that prices all European options (with all strikes and maturities) correctly at time 0; the resulting "local vol" model would then be used to price more exotic options (still at time 0).

In modeling interest rates we have a similar tradeoff between simplicity and accuracy. There are three basic viewpoints:

(a) *Simple short rate models.* Historically these came first. A standard example, which we'll discuss in detail, is the Vasicek model, which assumes that under the risk-neutral probability the short rate solves

$$dr = (\theta - ar)\,dt + \sigma\,dw \tag{1}$$

with θ, a, and σ constant and $a > 0$. This mean-reverting diffusion is known to physicists as an Ornstein-Uhlenbeck process. (We don't have to start with the risk-neutral process. It is equivalent to assume the subjective short-rate process has the form (1) and the market price of risk is constant. Then the risk-neutral process has the form (1) as well, with a new value of θ.)

The advantage of such a model is that it leads to explicit formulas. Moreover for some short-rate models – including Vasicek – we can see relatively easily that Black's formula is valid, since $P(t,T)$ has lognormal statistics under the forward risk-neutral measure.

The disadvantage of such a model is that it has just a few parameters. So there's no hope of calibrating it to match the entire yield curve $P(0,T)$ observed in the marketplace at time 0. For this reason Vasicek and its siblings are rarely used in practice. The main purpose of discussing it is to warm up.

(b) *Richer short-rate models.* The conceptual simplicity of a short-rate model is attractive. But we want to calibrate it to the full yield curve $P(0,T)$ observed at time 0. So the model must depend on a function of one variable. A standard example, which we'll discuss in detail soon, is the extended Vasicek model, also called the Hull-White model:

$$dr = (\theta(t) - ar)\,dt + \sigma\,dw \tag{2}$$

where a and σ are still constant but θ is a function of t. We'll show in due course that when θ satisfies

$$\theta(t) = \frac{\partial f}{\partial T}(0,t) + af(0,T) + \frac{\sigma^2}{2a}(1 - e^{-2at}) \tag{3}$$

the Hull-White model correctly reproduces the entire yield curve at time 0.

Hull-White retains many of the advantages of the basic Vasicek model. It still leads to explicit formulas; moreover it is still consistent with Black's formula (this is important, as a practical matter, since Black's formula is widely used in practice). Also it can be approximated by a recombining trinomial tree (very convenient for numerical use).

However Hull-White still has a major disadvantage: it gives us little freedom in modeling the *evolution* of the yield curve. Tomorrow's yield curve will be different from today's; but if we insist on getting the yield curve right at time 0, Hull-White we can only play with a and σ to try and fit this evolution. That, of course, is very constraining.

(c) *One-factor Heath-Jarrow-Morton.* We'd like a theory that can be calibrated precisely to the time-0 yield curve, but also permits a rich family of possible assumptions about the evolution of the yield curve. The one-factor HJM framework achieves precisely this goal. Rather than work in terms of a short-rate, it specifies the evolution of the instantaneous forward rate $f(t,T)$ by solving an SDE in t:

$$d_t f(t,T) = \alpha(t,T)\,dt + \sigma(t,T)\,dW_t. \tag{4}$$

The initial data $f(0,T)$ should naturally be taken from market data. The volatility $\sigma(t,T)$ in (4) must be specified – it is what determines the model. We'll see presently that Vasicek corresponds to the choice $\sigma(t,T) = \sigma e^{-a(T-t)}$. It turns out that the drift $\alpha(t,T)$ cannot be specified independently – the fact that all bonds (with all maturities) have the same market price of risk will give us an expression for α in terms of σ. (This is the analogue, for HJM, of the fact that for equities the risk-neutral process always has drift r.)

Every theory has advantages and disadvantages, and HJM is no exception. For one thing, it gives us too much freedom: how, in practice, should we choose $\sigma(t,T)$? Another disadvantage is the difficulty of using HJM numerically: only a few special cases (mainly corresponding to familiar short-rate models such as Hull-White and Black-Derman-Toy) give Markovian short-rates, which can be modelled using recombining trees.

The preceding viewpoints are in order of increasing complexity, and also in historical order. It is remarkable how recent these theories are. The fundamental work on equity-based options (e.g. the papers by Black, Scholes, and Merton) was done in the mid 70's, and the use of simple short-rate models like Vasicek dates from about the same time. But the development of Hull-White and its siblings (e.g. Black-Derman-Toy) dates from about 1990. And the fundamental papers by Heath, Jarrow, and Morton were published in the same year. In brief: the modern viewpoint on interest-rate modeling is not quite 15 years old!

The preceding discussion has focused entirely on one-factor models. It is of course possible (and valuable) to consider models with multiple sources of randomness (multifactor short-rate and multifactor HJM models). But let's not get carried away. We have already bitten off a lot; it's time to chew and swallow it.

****

The Vasicek model. Vasicek's model (1) has the advantage that we can work out almost everything explicitly. Moreover many of the methods used to analyze it extend straightforwardly to the Hull-White model (2).

An explicit formula for $r(t)$. It is easy to solve (1) explicitly: we have

$$d(e^{at}r) = e^{at}\,dr + ae^{at}r\,dt = \theta e^{at}\,dt + e^{at}\sigma\,dw,$$

so

$$e^{at}r(t) = r(0) + \theta\int_0^t e^{as}\,ds + \sigma\int_0^t e^{as}dw(s).$$

which simplifies to

$$r(t) = r(0)e^{-at} + \frac{\theta}{a}(1-e^{-at}) + \sigma\int_0^t e^{-a(t-s)}\,dw(s). \tag{5}$$

That calculation could have started at any time; thus

$$r(t) = r(s)e^{-a(t-s)} + \frac{\theta}{a}(1-e^{-a(t-s)}) + \sigma\int_s^t e^{-a(t-\tau)}\,dw(\tau). \tag{6}$$

We observe from (5) that $r(t)$ is Gaussian (the stochastic integral is Gaussian, because each term of the approximating Riemann sum is Gaussian, and sums of Gaussians are Gaussian.) Its mean is clearly

$$E[r(t)] = r(0)e^{-at} + \frac{\theta}{a}(1-e^{-at}),$$

and the variance is

$$Var\,[r(t)] = \sigma^2 E\left[\left(\int_0^t e^{-a(t-s)}\,dw(s)\right)^2\right] = \sigma^2\int_0^t e^{-2a(t-s)}\,ds = \frac{\sigma^2}{2a}(1-e^{-2at}).$$

We can use the preceding calculation to see that $P(t,T)$ has lognormal statistics. Indeed, by definition

$$P(t,T) = E\left[e^{-\int_t^T r(s)\,ds}\,|\,\mathcal{F}_t\right]. \tag{7}$$

Using (6) (with t and s interchanged) we have

$$P(t,T) = A(t,T)e^{-B(t,T)r(t)} \tag{8}$$

where

$$B(t,T) = \int_t^T e^{-a(s-t)}\,ds, \quad \text{and} \quad A(t,T) = E\left[e^{-\int_t^T \left\{\frac{\theta}{a}(1-e^{-a(s-t)})+\sigma\int_t^s e^{-a(s-\tau)}\,dw(\tau)\right\}ds}\right].$$

Since $A(t,T)$ and $B(t,T)$ are deterministic and $r(t)$ is Gaussian, (8) shows that $P(t,T)$ is lognormal.

An explicit formula for $P(t,T)$. One approach is to evaluate $A(t,T)$ and $B(t,T)$ from the definitions given above. The calculation of B is easy; that of A is less trivial – an amusing exercise in stochastic integration. You'll find the details of an equivalent calculation on pages 128-129 of Lamberton & Lapeyre.

Here let's implement another approach, which has the nice feature of being somewhat more general. (For example, a similar calculation works for the Cox-Ingersoll-Ross model $dr = (\theta - ar)\,dt + \sigma\sqrt{r}\,dw$, though the distribution of r is *not* Gaussian in this case.) Recall that $P(t,T) = V(t, r(t))$ where V solves the PDE

$$V_t + (\theta - ar)V_r + \tfrac{1}{2}\sigma^2 V_{rr} - rV = 0$$

with final-time condition $V(T,r) = 1$ for all r at $t = T$. Motivated by the preceding discussion, let's look for a solution of the form

$$V = A(t,T)e^{-B(t,T)r}.$$

To satisfy the PDE, A and B should satisfy

$$A_t - \theta AB + \tfrac{1}{2}\sigma^2 AB^2 = 0 \quad \text{and} \quad B_t - aB + 1 = 0$$

with final-time conditions

$$A(T,T) = 1 \quad \text{and} \quad B(T,T) = 0.$$

Solving for B first, then A, we get

$$B(t,T) = \frac{1}{a}(1 - e^{-a(T-t)})$$

and

$$A(t,T) = \exp\left[\left(\frac{\theta}{a} - \frac{\sigma^2}{2a^2}\right)(B(t,T) - T + t) - \frac{\sigma^2}{4a}B^2(t,T)\right].$$

The desired formula for $P(t,T)$ is of course

$$P(t,T) = A(t,T)e^{-B(t,T)r(t)}.$$

Evaluating the associated term structure, and its volatility. Vasicek has only three parameters, so of course its term structure is rather special. For later comparison with HJM, let's identify it by calculating $f(0,T)$. The calculation is slightly tedious but entirely straightforward, using the definition $f(0,T) = -\partial \log P(0,T)/\partial T$ and our formula for $P(0,T)$. The final result is:

$$f(0,T) = \frac{\theta}{a} + e^{-aT}(r_0 - \frac{\theta}{a}) - \frac{\sigma^2}{2a^2}(1 - e^{-aT})^2.$$

Note that this is consistent with (3).

An HJM model is characterized by the volatility of the instantaneous forward rate. We'll show in due course that Vasicek is a special case of HJM. Preparing for this, let's identify now the volatility of f, i.e. let's find $\sigma(t,T)$ such that

$$d_t f(t,T) = (\text{stuff})\, dt + \sigma(t,T)\, dw.$$

This is easy: we have $\log P(t,T) = \log A(t,T) - B(t,T)r(t)$, so

$$f(t,T) = -\partial_T \log A(t,T) + \partial_T B(t,T) r(t),$$

so by Ito's formula

$$\sigma(t,T) = \sigma \partial_T B(t,T) = \sigma e^{-a(T-t)}.$$

Validity of Black's formula. To show Black's formula is valid for options on zero-coupon-bonds, we must show that $P(t,T)$ is lognormal under the forward-risk-neutral measure. (Remember: this is the measure under which tradeables normalized by $P(t,T)$ are martingales.) We already know it is lognormal under the risk-neutral measure, but here we're interested in a different measure, associated with a different numeraire.

Let's review how change-of-numeraire works in the one-factor setting. The risk-neutral measure is associated with the risk-free money-market account B as numeraire (by definition $dB = rB\, dt$ with $B(0) = 1$). Let N be another numeraire, and suppose $\overline{Q}$ is the associated equivalent martingale measure. We can only use tradeables as numeraires, so the risk-neutral process for N is

$$dN = rN\, dt + \sigma_N N\, dw$$

where w is a Brownian motion under the risk-neutral measure. By Ito we have

$$d(B/N) = B\, d(N^{-1}) + N^{-1}\, dB$$

(there is no $dBd(N^{-1})$ term since $dB = rB\, dt$). A bit of algebra gives

$$d(B/N) = (B/N)\sigma_N^2\, dt - (B/N)\sigma_N\, dw.$$

When we use N as numeraire, the associated martingale measure $\overline{Q}$ is characterized by the fact that B/N is a $\overline{Q}$-martingale, i.e.

$$d(B/N) = -(B/N)\sigma_N\, d\overline{w}$$

where $\overline{w}$ is a $\overline{Q}$-Brownian motion. It follows that

$$d\overline{w} = -\sigma_N\, dt + dw.$$

What's the point of all this? We need to write the SDE for the Vasicek process under the forward-risk-neutral measure, i.e. when the numeraire is $P(t,T)$. Recall that $P(t,T) = A(t,T)e^{-B(t,T)r(t)}$, so from Ito the volatility of $P(t,T)$ is $-B(t,T)\sigma$. Therefore the preceding calculation gives

$$d\overline{w} = \sigma B(t,T)\, dt + dw.$$

We conclude that the SDE for the interest rate is

$$dr = (\theta - ar)\, dt + \sigma\, dw = [\theta - ar - \sigma^2 B(t,T)]\, dt + \sigma\, d\overline{w}$$

where $\overline{w}$ is a Brownian motion under the forward-risk-neutral measure. This SDE can be solved much as we did before (the fact that the drift depends on time doesn't disturb the calculation at all). Explicit formulas are a bit tedious to obtain – but they're not needed. From the structure of the calculation, we see quite easily (just as in the risk-neutral case) that the short rates and bond prices are lognormal.

****

Aside from being too restrictive (not enough parameters), Vasicek has one other awkward feature: since $r(t)$ is Gaussian, there is a positive probability that it is negative. This is of course unrealistic: interest rates must be positive. The Cox-Ingersoll-Ross model mentioned above was introduced to fix this problem: one can show that if $\theta \geq \sigma^2/2$ then r stays positive (with probability one). The form of $P(t,T)$ can be found by a separation-of-variables calculation much like the one done above. The statistics of $r(t)$ can also be characterized. See e.g. Avellaneda and Laurence for some analysis of this model.