ullint2b.tex pm Wed 7.2.2018

III. SHORT-RATE MODELS

1. Possible model choices

This approach relies on the fact that the zero coupon curve at each time, or equivalently the zero bond curve

$$T \mapsto P(t,T) := E_t^{\mathbb{Q}}[\exp\{-\int_t^T r_s ds\}],$$

is completely characterised by the probabilistic/dynamic properties of r.

So we write a model for r, the initial point of the curve $T \mapsto L(t,T)$ for $T = t$ at every instant t. Typically, an SDE for r is chosen:

$$dr_t = b(t, r_t)dt + \sigma(t, r_t)dS_t,$$

where $b(t, r_t)$ is the local mean, $\sigma(t, r_t)$ is the local standard deviation (SD), and the stochastic process (S_t) is the driving noise (stochastic change). We shall confine ourselves in this course to the most important case, when S is chosen to be Brownian motion (BM):

$$dr_t = b(t, r_t)dt + \sigma(t, r_t)dW_t \qquad (r)$$

('W for Wiener').

Each choice of model for r needs to be evaluated in the light of the requirements for interest-rate modelling set out at the end of Chapter II above. We confine ourselves here to four of the most important ones. These are – using α to denote a (multi-dimensional) parameter:

1. *Vasicek model (1977)*:

$$dx_t = k(\theta - x_t)dt + \sigma dW_t, \qquad \alpha = (k, \theta, \sigma). \qquad (Vas)$$

2. *Cox-Ingersoll-Ross (CIR) model, 1985.*

$$dx_t = k(\theta - x_t)dt + \sigma\sqrt{x_t}dW_t, \qquad \alpha = (k, \theta, \sigma), \quad 2k\theta > \sigma^2. \qquad (CIR)$$

3. *Affine term-structure models (ATM)*:

$$R(t,T) = \alpha(t,T) + \beta(t,T)r_t.$$

4. *Exponential Vasicek*:

$$x_t = \exp\{z_t\}, \qquad dz_t = k(\theta - z_t)dt + \sigma dW_t, \qquad \alpha = (k, \theta, \sigma). \qquad (ExpV)$$

Each model has important consequences, which must be kept in mind when choosing a particular short-rate model.

We mention also the *Dothan/Rendleman and Bartter* model:

$$dx_t = ax_t dt + \sigma x_t W_t, \qquad \alpha = (a, \sigma).$$

This has the form of the SDE for geometric Brownian motion (GBM: MATL480, VI), with solution

$$x_t = x_0 \exp\{(a - \frac{1}{2}\sigma^2)t + \sigma W_t\}.$$

So (as with the Black-Scholes model of MATL480), the solution is *log-normal*. For bond prices, this leads to integrals that have to be done numerically. See BM, §3.2.2. This is not an ATM.

2. Vasicek model, 1977

$$dx_t = k(\theta - x_t)dt + \sigma dW_t, \qquad \alpha = (k, \theta, \sigma). \qquad (Vas)$$

We have met this SDE before (MATL480, V.6, Week 5a), in the form of the *Ornstein-Uhlenbeck (OU) process* (replace x_t by $x_t - \theta$ to reduce to (OU) there). This model has several attractive properties. It can be solved explicitly (as we did before). Its solution is Gaussian – indeed, it is stationary Gaussian Markov. The model is *mean-reverting*: the mean $E[r_t] \to \theta$ as $t \to \infty$, with a velocity depending on k (relaxation time $1/k$), while the variance does not explode.
Drawbacks: (a) As rates are Gaussian, they can take negative values (I.5).
(b) Gaussianity does not fit with observed market data.

Mean reversion and interest rates.

This behaviour is consistent with that of interest rates traditionally – that is, before the Crash of 2007/08. Then, interest rates varied with the state of the business cycle, and showed a tendency to revert to their long-term mean ('central push'); see I.3.

Pricing ZCBs under Vasicek.

Recall (MATL480, Prob/Soln 5b Q2) our solution to the Ornstein-Uhlenbeck process (OU), equivalently, to Vasicek, and (MATL480, Prob/Soln 5b Q1) that for $\mathbf{a}$ a constant vector, linear combinations $\mathbf{a}^T\mathbf{X}$ of a Gaussian random vector $\mathbf{X}$ are Gaussian:

$$\mathbf{X} \sim N(\mu, \Sigma) \quad \Rightarrow \quad \mathbf{a}^T\mathbf{X} \sim N(\mathbf{a}^T\mu, \mathbf{a}^T\Sigma\mathbf{a}).$$

We used this there to show that for f deterministic, $\int_0^t f(s)dW_s$ is Gaussian. The same argument shows that, when the spot rate $r = (r_t)$ is Gaussian, as here in the Vasicek model,

$$X := -\int_0^t r_s ds \text{ is Gaussian :} \qquad X \sim N(M, V^2),$$

say. Then $D(t,T) = e^X$ is log-normal, $LN(M, V^2)$, and (MATL480, Prob/Soln 4b Q1, Q2) has mean

$$P(t,T) = E[D(t,T)] = E[e^X] = \exp\{M + \frac{1}{2}V^2\}.$$

One can find M and V, and so show that ([BM, 3.2.1], [Z, p.126]) the bond price P is given by

$$\begin{aligned} P(t,T) &= \exp\{-R(t,T)(T-t)\} \\ &= A(t,T)\exp\{-B(t,T)r_t\}, \end{aligned}$$

where

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

$$A(t,T) = \exp\{(\theta - \frac{\sigma^2}{2\kappa^2})[B(t,T) - T + t] - \frac{\sigma^2}{4\kappa}B(t,T)^2\} :$$

$$P(t,T) = \exp\{(\theta - \frac{\sigma^2}{2\kappa^2})[B(t,T) - T + t] - \frac{\sigma^2}{4\kappa}B(t,T)^2 - B(t,T)r_t\}.$$

For details, we refer to the original Vasicek paper,
O. VASICEK, An equilibrium characterization of the term structure. *J. Financial Economics* **5** (1977), 177-188.

Observe that $R(t,T)$ here is an *affine* function of r, that is, of the form

$$r \mapsto a + br.$$

We shall return to this aspect later (III.4).

One can calculate the price of an *option* on a ZCB in the Vasicek model. Since a caplet can be seen as a put option on a zero bond, one can thus calculate the price of a *caplet* (this involves a combination of the results above and calculations resembling those giving the Black-Scholes formula). We omit details; see e.g. [BM, 3.2.1]. They involve the *bivariate normal* distribution (MATL480, Prob/Soln 2b).

The Vasicek model is the simplest of the widely-used models in interest-rate theory, and pricing caplets is a fairly basic task here, to which we return later. It is abundantly clear from the above that, even for this fairly simple task in a fairly simple model, one needs quite a lot of mathematics! Indeed, to do Mathematical Finance demands a good knowledge of, in particular, Probability and Statistics, and computing – R for Statistics, MATLAB, C/C++ (or C#) for general computing and data-handling.

One also sees from the above that there are links (indeed, profound links) between *normality* (or *Gaussianity*) and *linearity*. Where one has both, one may have the *analytic tractability* to enable one to do calculations, as here.

3. Cox-Ingersoll-Ross (CIR) model, 1985

This is given by the SDE (for the short rate $r_t = y_t$)

$$dy_t = \kappa[\mu - y_t]dt + \nu\sqrt{y_t}dW_t, \qquad \alpha = (\kappa, \mu, \nu).$$

Subject to reasonable restrictions on the parameters, this model gives *positive* interest rates. The instantaneous rates are not now Gaussian, but have a *non-central chi-squared* (χ^2) distribution. As with Vasicek, CIR is *mean-reverting*; the mean tends to the central value μ at a speed depending on κ, and the variance does not explode. Although it can be handled analytically to some extent, it is *less tractable* than Vasicek, especially for extensions to the multifactor case with correlation. But, CIR usually fits market data better than Vasicek. The parameters have interpretations as in Vasicek:
μ: long-term mean reversion level;
κ: speed of mean reversion;
ν: volatility.
We quote:

$$E[y_t] = y_0 e^{-\kappa t} + \mu(1 - e^{-\kappa t}),$$

$$var(y_t) = y_0 \frac{\nu^2}{\kappa}(e^{-\kappa t} - e^{-2\kappa t}) + \mu\frac{\nu^2}{2\kappa}(1 - e^{-\kappa t})^2.$$

So the process has a limit (or equilibrium, or *ergodic*) distribution as t increases, around the limiting mean μ and with the limiting variance $\mu\nu^2/2\kappa$. So:
the larger μ, the larger the long-term average interest rate;
the larger κ, the faster the convergence;
the larger ν, the larger the volatility – but, κ and ν fight each other in their influence on the volatility.

The bond price P is given here by

$$\begin{aligned} P(t,T) = \exp\{-R(t,T)(T-t)\} &= A(t,T)\exp\{-B(t,T)r_t\} \\ &= \exp\{-\alpha(t,T) - B(t,T)r_t\}, \end{aligned}$$

say, with A (or α), B that can be calculated (see e.g. [BM, (3.24), (3.25)]). Again, the exponent – or $R(t,T)$ – here is affine in r.

4. Affine Term-structure Models (ATM).

Affine term-structure models (ATM for short) are those for which the continuously compounded spot rate $R(t,T)$ (II.1) is an *affine function of the spot rate* r_t:

$$R(t,T) = \alpha(t,T) + \beta(t,T)r_t. \qquad (ATM)$$

Equivalently, by II.1 (R),

$$P(t,T) = A(t,T)\exp\{-B(t,T)r_t\}.$$

These have special and useful features. As we have seen, both the Vasicek and CIR models are of this type: *Vasicek and CIR are ATM.*

Recall (II.1) that the bond prices $P(t,T)$ can be expressed as

$$P(t,T) = \exp\{-\int_t^T f(t,u)du\},$$

in terms of the *instantaneous forward rate*

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

So for affine models,

$$f(t,T) = -\frac{\partial}{\partial T}\log A(t,T) + \frac{\partial B(t,T)}{\partial T}r_t.$$

So the stochastic differential is of the form

$$df(t,T) = \{\cdots\}dt + \frac{\partial B(t,T)}{\partial T}\sigma(t,r_t)dW_t,$$

where $\sigma(t,r_t)$ is the diffusion coefficient in the short-rate dynamics for r. So the volatility for f in an affine model is

$$\sigma_f(t,T) = \frac{\partial B(t,T)}{\partial T}\sigma(t,r_t).$$

Write the risk-neutral dynamics for the short rate r_t as

$$dr_t = b(t,r_t)dt + \sigma(t,r_t)dW_t.$$

It turns out that a general way to form affine models is to take both the functions b and σ^2 to be *affine themselves*:

$$b(t,x) = \lambda(t)x + \eta(t), \qquad \sigma^2(t,x) = \gamma(t)x + \delta(t).$$

For then, the functions A and B can be obtained from the functions λ, η, γ, δ above by solving the following differential equations (DEs):

$$\frac{\partial}{\partial t}B(t,T) + \lambda(t)B(t,T) - \frac{1}{2}\gamma(t)B(t,T)^2 + 1 = 0, \qquad B(T,T) = 1,$$

$$\frac{\partial}{\partial t}\log A(t,T) - \eta(t)B(t,T) + \frac{1}{2}\delta(t)B(t,T)^2 = 0, \qquad A(T,T) = 1.$$

The first equation is a *Riccati* DE. It needs to be solved numerically in general, but in the particular cases of the Vasicek model,

$$\lambda(t) = -\kappa, \qquad \eta(t) = \kappa\theta, \qquad \gamma(t) = 0, \qquad \delta(t) = \sigma^2,$$

or the CIR model,

$$\lambda(t) = -\kappa, \qquad \eta(t) = \kappa\theta, \qquad \gamma(t) = \sigma^2, \qquad \delta(t) = 0,$$

the DEs are explicitly solvable, and give (of course!) the solutions for A and B that we found before.

So affinity in the coefficients gives affinity in the term structure. The converse does not hold in general, but it does hold in the time-homogeneous case:

$$b(t,x) = b(x), \qquad \sigma(t,x) = \sigma(x),$$

giving

$$b(x) = \lambda x + \eta, \qquad \sigma^2(x) = \gamma x + \delta$$

for suitable constants $\lambda, \kappa, \gamma, \delta$, giving affine coefficients.

5. Exponential Vasicek model

Here our process is $x = (x_t)$, where $x_t = \exp\{z_t\}$ and $z = (z_t)$ is as in the Vasicek model. As we have studied this, we can transfer our conclusions from z to x by taking exponentials. But note that there is an important change: x_t is now *positive*, being an exponential. So the exponential Vasicek model has positive interest rates (I.5).

Note also that, as z_t is normal (Gaussian), x_t is *log-normal*.

Note. The log-normal distribution is somewhat peculiar, even pathological. Recall the moment-generating function (MGF) of a random variable X: if $X \sim N(\mu, \sigma^2)$,

$$M(t) = M_X(t) := E[e^{tX}] = \exp\{\mu t + \frac{1}{2}\sigma^2 t^2\},$$

for t real. In particular, for the standard normal, $X \sim N(0, 1)$,

$$M(t) = \exp\{\frac{1}{2}t^2\}.$$

If we write

$$\mu_n := E[X^n]$$

for the nth moment of X, expand the exponential, $e^{tX} = \sum_0^\infty t^n X^n/n!$ (we assume all moments exist, as they do in the normal case we are interested in here), interchange of $E[.]$ and Σ give (by Fubini's theorem from Measure Theory – quote)

$$M(t) = \sum_0^\infty \mu_n t^n/n!$$

For those who know Complex Analysis: the mathematics of power series, as here, is essentially *complex* rather than *real* analysis (for background, see e.g. my homepage, link to M2P3 Complex Analysis, Ch. II). When the power series here has *positive* radius of convergence R, the above is justified (Cauchy-Taylor Theorem). But when $R = 0$, things are more complicated. For $R > 0$, the distribution is uniquely determined by its moments (example:

normal distribution, $R = \infty$). But when $R = 0$, this may not hold – and it *doesn't* hold in the log-normal case: the log-normal is the simplest example of a distribution *not uniquely determined by its moments*. This is thought by some to underlie some of the difficulties one encounters in Mathematical Finance, in particular, in the Black-Scholes theory.

6. Vasicek model (continued): Objective measure; econometrics, statistics, historical estimation

We can consider the objective measure $\mathbb{Q}^0$-dynamics of the Vasicek model as a process of the form

$$dr_t = [\kappa\theta - (\kappa + \lambda\sigma)r_t]dt + \sigma dW_t^0, \qquad r(0) = r_0.$$

Here λ is a new parameter, corresponding to the *market price of risk*. Compare these $\mathbb{Q}^0$-dynamics to the *risk-neutral* $\mathbb{Q}$-dynamics

$$dr_t = \kappa(\theta - r_t)dt + \sigma dW_t, \qquad r(0) = r_0.$$

Both are expressed by linear Gaussian SDEs, which correspond for $\lambda = 0$, the new parameter in the first (constant here, though not in general).

Pass between these two dynamics by a *Girsanov change of measure* (MATL480, VI.3):

$$\frac{d\mathbb{Q}}{d\mathbb{Q}^0}|\mathcal{F}_t = \exp\{-\frac{1}{2}\int_0^t \lambda^2 r_s^2 ds + \int_0^t \lambda r_s dW_s^0\}.$$

We get a spot-rate process which is tractable under both measures.

Note.

In traditional finance (MATL480), we begin with the objective measure $\mathbb{Q}^0$, and then pass to the risk-neutral measure and dynamics by adding parameters. Here we go in the reverse direction, because when *pricing* one starts with the risk-neutral dynamics:

Statistics, historical estimation, econometrics: objective measure:

$$dr_t = [\kappa\theta - (\kappa + \lambda\sigma)r_t]dt + \sigma dW_t^0, \qquad r(0) = r_0; \qquad (Obj)$$

Pricing: risk-neutral measure:

$$dr_t = \kappa(\theta - r_t)dt + \sigma dW_t, \qquad r(0) = r_0. \qquad (RiskN)$$

It is clear why we need a tractable dynamics in $(RiskN)$: claims are *priced* that way, and we have to be able to price things if we are going to trade in them in (at least, in any quantity: of course, some things are bought and sold as 'one-offs', under exceptional circumstances, such as 'fire sales' of assets by firms with cash-flow problems, or bankrupt firms). The reason why we also need tractable dynamics in (Obj) is to be able to do statistics on past (historical) data.

Say we are given a series $r_0, r_1, \cdots, r_n$ of daily observations of a proxy of r_t – say, a monthly rate, $r_t \sim L(t, t+1m)$: to use this information in our model, we estimate the model parameters. (For background on estimation of parameters in Statistics, see MATL374 Statistical Methods in Actuarial Science, or e.g. NHB, homepage, link to SMF (Statistical Methods for Finance), Ch. I, in particular, *maximum-likelihood estimation* (ME) – MATL374; NHB, SMF I.) Now data are collected in the *real world*, under the real-world – objective – measure $\mathbb{Q}^0$. So what we can estimate from such historical observations is the $\mathbb{Q}^0$-dynamics, via estimates of the objective parameters κ, λ, θ, σ. By contrast, to price *derivatives*, we use the risk-neutral measure $\mathbb{Q}$. So calibration of the model to derivative prices, reflecting the current market prices of such derivatives, involves the $\mathbb{Q}$-dynamics, and the parameters κ, θ, σ – not including λ as above.

One can combine the two. For instance, as σ is the same in both, one could estimate σ from historical data by (MLE), and κ, θ by calibration to market prices.

MLE for Vasicek

$$dr_t = [b - ar_t]dt + \sigma dW_t^0,$$

with b, a constants. The solution is, as above, for $s < t$

$$r_t = r_s + \frac{b}{a}(1 - e^{-a(t-s)}) + \sigma \int_s^t e^{-a(t-u)} dW_u^0.$$

So (as the Ornstein-Uhlenbeck or Vasicek process is Markov)

$$r_t | r_s = r_t | \mathcal{F}_s \sim N(r_s + \frac{b}{a}(1 - e^{-a(t-s)}), \frac{\sigma^2}{2a}(1 - e^{-2a(t-s)})). \qquad (*)$$

Write δ for the time-step in the data (1 month, in the example above);

$$\beta := b/a, \qquad \alpha := e^{-a\delta}, \qquad V^2 := \frac{\sigma^2}{2a}(1 - e^{-2a(t-s)}).$$

The MLEs of these can be shown to be:

$$\hat{\alpha} = \frac{n\sum_1^n r_i r_{i-1} - \sum_1^n r_i \sum_1^n r_{i-1}}{n\sum_1^n r_i^2 - (\sum_1^n r_{i-1})^2},$$

$$\hat{\beta} = \frac{\sum_1^n (r_i - \hat{\alpha} r_{i-1})}{n(1-\hat{\alpha})},$$

$$\hat{V}^2 = \frac{1}{n}\sum_1^n [r_i - \hat{\alpha} r_{i-1} - \hat{\beta}(1-\hat{a})^2]^2.$$

This gives the MLE estimates of the δ-transition probabilities of the spot-rate process r under the objective measure $\mathbb{Q}^0$; this allows, for instance, simulation of r over δ-spaced future time-instants.

Monte-Carlo simulation (MATL484, Computational Methods in Financial Mathematics)

From $(*)$ above, when we know the spot rate r at a time t_i, we can use M-C to simulate forward in time, to the next such time-point t_{i+1}. This is very convenient. By contrast, we will meet later models that do not lend themselves to M-C.

7. Spot rate: Choice of model

There are a number of questions we should ask when choosing a model for the spot-rate r:

(a) Does the dynamics imply positive rates r_t for each t? This is desirable, but (I.5) not as essential as in the past.

(b) What distribution does the dynamics imply for r? – fat-tailed, etc.?

(c) Are bond prices

$$P(t,T) = E_t[\exp\{-\int_t^T r_s ds\}]$$

(and so spot rates, forward rates and swap rates) explicitly computable from the dynamics?

(d) Are bond-option (and cap, floor, swaption) prices explicitly computable from the dynamics?

(e) Is the model mean-reverting (i.e. mean tends to long-time limit, and variance does not explode)?

(f) What do the volatility structures implied by the model look like?
(g) Does the model allow for explicit short-rate dynamics under the forward measures?
(h) How suitable is the model for Monte-Carlo simulation?
(i) How suited is it for building trees (recombining lattices)?
(j) Do the dynamics allow historical estimation techniques to be used for parameter estimation?

If we have the initial curve

$$T \mapsto P(0, T),$$

for our model to incorporate this curve (our 'initial condition'), we need to choose the parameters so that the curve fits future market curves as closely as possible. This is *calibration of the model to market data*. This is an *optimization* problem – e.g., for the Vasicek case above. With too few parameters, some shapes of curve can never be obtained. In such cases, and to also calibrate *caplet* data, *exogeneous* term-structure models (below) are usually used.