ullint3a.tex am Wed 13.2.2019

III.7 (continued)

Optimization methods

These include:
local methods – e.g., gradient method;
global methods – e.g., simulated annealing, genetic algorithms, ...
For such endogenous methods, the resulting fit is usually poor.

Exogenous methods

The usual remedy here is to pass from an endogenous model, with parameters constants, to an exogenous model with parameters *functions* – varying with time. For example, for Vasicek,

$$dr_t = \kappa(\theta - r_t)dt + \sigma dW_t \mapsto \kappa(\theta(t) - r_t)dt + \sigma dW_t.$$

The function $\theta(t)$ of time can be defined from the initial market curve $T \mapsto L(0,T)$ so that the model is *exact* for the initial time 0. The remining parameters may be used to calibrate caps/swaptions data (we don't need to *price* caps, as we *know* their prices –they are very liquid – but we want to have them 'inside the model', to help us price more difficult things).

Exogenous methods: Examples

Dynamics of $r_t = x_t$ under the risk-neutral measure:
Ho-Lee:

$$dx_t = \theta(t)dt + \sigma dW_t. \tag{Ho - Lee}$$

Hull-White (extended Vasicek)

$$dx_t = \kappa(\theta(t) - x_t)dt + \sigma dW_t. \tag{Hull - White/Vas}$$

Hull-White (Extended CIR)

$$dx_t = \kappa(\theta(t) - x_t)dt + \sigma\sqrt{x_t}dW_t. \tag{Hull - White/CIR}$$

Black-Derman-Toy (Extended Dothan)

$$x_t = x_0 \exp\{u(t) + \sigma(t)W_t\}. \tag{BDT}$$

Black-Karasinski (Extended exponential Vasicek)

$$x_t = \exp\{z_t\}, \qquad dz_t = \kappa(\theta(t) - z_t)dt + \sigma dW_t. \qquad (BK)$$

CIR++ (Shifted CIR model, Brigo & Mercurio, 2000)

$$r_t = x_t + \phi(t;\alpha), \qquad dx_t = \kappa(\theta - x_t)dt + \sigma\sqrt{x_t}dW_t. \qquad (CIR++)$$

Now parameters can be used to fit volatility structures.
Note. As always, one has conflicting pressures: too few parameters, and we cannot fit without distortion; too many parameters, and we risk *over-fitting* – treating randomness in our data with 'too much respect', so that it becomes fixed rather than transient, and adding new parameters that may not mean much (or even anything).

Summary of model performance
Dn = distribution (I use "d/n" in my own notes);
ABP = analytical bond prices (a 'good thing');
AOP = analytical bond-option prices (ditto);
MR = mean reverting;
Mult = tractable multi-factor extension;
$r > 0$: just that.

Model
Vasicek: D/n N (normal); ABP; AOP; Multi; MR.
CIR: D/n non-central chi-squared; ABP; AOP; Multi; MR; $r > 0$.
Dothan: D/n e^N (log-normal); ABP; MR; $r > 0$.
Exponential Vasicek: D/n e^N; MR; $r > 0$.
Ho-Lee: D/n N; ABP; AOP; Multi: $r > 0$.
Hull-White (Vasicek): D/n N; ABP; AOP; Multi; MR.
Hull-White (CIR): D/n n-c χ^2; MR; $r > 0$.
BDT: D/n e^N; MR; $r > 0$.
Black-Karasinski: D/n e^N; MR; $r > 0$.
CIR++ (Brigo-Mercurio): D/n n-c χ^2; ABP; AOP; Multi; MR; $r > 0$.

Comments
Ho-Lee: very tractable; stylized, simplistic; negative rates.
Hull-White (Vasicek): very tractable, easy to implement and calibrate; trees easy; Monte-Carlo possible; can have negative rates; can give pathological

calibrations under some market conditions.
Hull-White (CIR): Not tractable; numerical problems, etc.
BDT: Intractable; some mean reversion but linked to volatility; excellent distribution; good calibration to distributions implied by market rates; explosion problem in continuous version –

$$E[B_t] = E[\exp\{\int_0^t r_s ds\}] = \infty.$$

Needs *trinomial* trees (below). No Monte Carlo possible.
Black-Karasinski: Intractable; mean reversion; excellent distribution and good calibration as above; explosion problem (as in all log-normal short-rate models); needs trinomial trees; no M-C.
CIR++: Tractable; many formulas; easy to implement and calibrate; trees not easy but feasible; M-C possible; positive rates; can give pathological calibrations under some market conditions (as with most one-dimensional short-rate models).

Shifted Vasicek model

This is defined by

$$r_t = x_t + \phi(t, \alpha), \qquad dx_t = \mu(x_t; \alpha)dt + \sigma(x_t; \alpha)dW_t,$$

with x_0 a *further parameter*. For details, see [BM. 3.8.4, p.100-102].

Path-dependent products

These are derivatives whose payout at maturity T depends on interest rates at a number of earlier times $t_i < T$. The payout cannot then be decomposed into a sum of payouts each depending on a single interest rate at that time. In such cases, it is usually necessary to price by Monte-Carlo simulation.

Bermudan swaptions

These are swaptions that can be exercised every year, rather than at a single maturity T. Monte-Carlo simulation is not suitable here. Indeed, simulating forward in time does not help us to find the optimal exercise strategy. Instead, we proceed as with American options and the Snell envelope: start from the final maturity time, and use binomial trees and backward induction.
Note. To remember the term: Bermuda is in the North Atlantic, between

America and Europe, so Bermudan options/swaptions, which can be exercised early at *some* times, are between American (early exercise possible at any time) and European (no early exercise).

Trinomial trees

For more complicated models, such binomial trees do not work, and have to be replaced by *trinomial trees*. For details, see e.g. [BM]:
3.3.3 for Hull-White;
3.5.2 for BK;
3.9.1 for CIR;
3.10 for expVas;
4.1.2 for G2++.

8. Multidimensional models: How many factors?

The subject of interest-rate theory is *infinite-dimensional*: the object of interest is the yield curve, an infinite-dimensional object. But, this is driven by the source of randomness – driving noise. In the models above, this is Brownian motion (BM) *on the line*, BM = $W(\mathbb{R})$. One can capture more of what is actually observed in real markets by moving to a higher-dimensional driving noise.

Recall:
(i) the formula

$$(dW_t)^2 = dt,$$

based on Lévy's quadratic-variation theorem and basic to the Itô calculus;
(ii) the *bivariate normal distribution*, with five parameters: two means, two variances, and one correlation

$$\rho \in (-1, 1)$$

($\rho = \pm 1$ is possible, but degenerate, so we exclude them here).

These can be used to extend the Vasicek model, and make it more flexible – so better able to capture cap and swaption structures seen in the market.

Two-factor Vasicek

This is defined by

$$dx_t = \kappa_x(\theta_x - x_t)dt + \sigma_x dW_1(t),$$

$$dy_t = \kappa_y(\theta_y - y_t)dt + \sigma_y dW_2(t),$$

$$dW_1(t)dW_2(t) = \rho dt,$$

$$r_t = x_t + y_t + \phi(t, \alpha), \qquad \alpha = (k_x, \theta_x, \sigma_x, x_0; \kappa_y, \theta_y, \sigma_y, y_0).$$

This extra flexibility is valuable: with a one-factor (Vasicek) model, we have for the continuously-compounded spot rate $R(t, T)$

$$corr_0(R(1y, 2y), R(1y, 30y)) = 1,$$

because there is only *one* source of randomness, W. But this is an extreme and degenerate case (compare degrees Fahrenheit and Centigrade!). With a two-factor model, we can have much more reasonable correlations between two such different things.

The question arises as to how many factors – sources of randomness – we should include. Too few, and the model is not capable of capturing some of the essential features of what we see in the market. Too many, and the model risks becoming intractable, and showing *over-interpretation* – failing to average out randomness in the data (so fixing it in the model), and cluttering the model up with spurious parameters etc. The areas of Statistics relevant here include Multivariate Analysis (SMF, Ch. III) in general, and Principal Component Analysis (PCA) in particular (SMF, III.5), and regression (SMF, Ch. IV); see also
N. H. BINGHAM and John M. FRY, *Regression: Linear models in statistics*, Springer, 2010.
In what follows, we focus mainly on two-factor and three-factor models.

Also relevant here is the *tenor structure*. Government bonds, for instance, could be issued at any time t and for any maturity T. In practice, they are issued only at *some* times t_i with *some* maturities T_j. So the bond market is actually *finite-dimensional*. Furthermore, the number of products is in the hundreds, but the amounts traded are in trillions; so, interest-rate derivatives are *highly liquid*, so we know their price accurately. This is part of the motivation for *market models* (Ch. V).

Credit risk.

As we have mentioned, there is in fact no such thing as a *risk-free* interest rate. In practice, there is always some risk of default – *credit risk*. We discuss this in detail in Ch. VII. This combines well with spot rates: in the presence of credit risk with default rate λ_t, the spot rate r_t is increased to $r_t + \lambda_t$ (*Lando's formula*). This gives the *credit spread* as λ_t.

IV. FORWARD-RATE MODELS

1. The Heath-Jarrow-Morton (HJM) model

Recall (II.2, W2a) the *forward LIBOR rate* at time t between T and S $(S > T > t)$,

$$F(t,T,S) = \Big(\frac{P(t,T)}{P(t,S)} - 1\Big)/(S-T) = -\frac{1}{P(t,S)}.\frac{P(t,T)-P(t,S)}{T-S},$$

which makes the FRA contract to lock in at time t the interest rates between T and S fair. When S collapses to T, we get the *instantaneous forward rates*:

$$f(t,T) = \mathrm{lim}_{S\downarrow T} F(t,T,S) = -\frac{1}{P(t,T)}.\frac{\partial P(t,T)}{\partial T} = -\frac{\partial}{\partial T}\log P(t,T).$$

So

$$P(t,T) = \exp\{-\int_t^T f(t,s)ds\}. \qquad (*)$$

When further T collapses to t we get the spot rate (short rate):

$$\lim_{T\downarrow t} f(t,T) = r_t.$$

For, when $\epsilon > 0$ is small, by $(*)$

$$P(t,t+\epsilon) = \exp\{-\int_t^{t+\epsilon} f(t,s)ds\} \sim \exp\{-\epsilon f(t,t)\}.$$

Also,

$$P(t,t+\epsilon) = E_t[\exp\{-\int_t^{t+\epsilon} r_s ds\}] \sim E_t[\exp\{-\epsilon r_t\}] \sim \exp\{-\epsilon r_r\};$$

this follows from continuity of r_t in t, which holds in the models of III W2b above (or more generally, if r_t is continuous in mean). Comparing, the result follows.

We re-write $(*)$ for reference as

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

As we have seen in Ch. III on modelling the spot rate r_t, there are limitations here:
(i) The spot rate is not observable (so trying to model it may not be the best approach);
(ii) We do not get a particularly good fit with one-factor models (most of the ones we considered). Two-factor models give a better fit, at the cost of greater complexity, etc.

Here we change approach, and model the *forward rate* $f(t,T)$. This is not observable either! – so it is not clear yet that this will advance us. Indeed, it seems that this might be even worse, as it is r that is more fundamental:

$$f(t,T) = -\frac{\partial}{\partial T} \log E_t[\exp\{-\int_t^T r_s ds\}],$$

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

Heath, Jarrow and Morton (1992) – HJM – assumed that, for a given maturity T, the instantaneous forward rate $f(t,T)$ evolves, under a given measure, according to the following diffusion process:

$$df(t,T) = \alpha(t,T)dt + \sigma(t,T)dW_t,$$

with initial condition

$$f(0,T) = f^M(0,T),$$

where

$$T \mapsto f^M(0,T)$$

is the market instantaneous forward curve at time $t = 0$, and $W = (W_1, \cdots, W_N)$ is an N-dimensional BM. Here $\sigma(t,T) = (\sigma_1(t,T), \cdots, \sigma_N(t,T))$ and $\alpha(t,T)$ are adapted processes, and

$$\sigma(t,T)dW_t = \sum_1^N \sigma_i(t,T)dW_i(t)$$

is the dot (scalar) product of the two vectors on the LHS.

The fundamental result of HJM is that, *if the model has no arbitrage* (is NA), then under the risk-neutral measure the dynamics of f must be of the form

$$df(t,T) = \sigma(t,T)\Big(\int_t^T \sigma(t,s)ds\Big)dt + \sigma(t,T)dW_t. \qquad (HJM)$$

As the coefficient of dt is the (local) mean or drift, and this shows that *the drift is determined by the (local) volatility or diffusion coefficient.*

The SDE (HJM) is called the *Heath-Jarrow-Morton drift condition.* We defer its proof to the next chapter, after we have more detailed tools on change of numeraire.

Note the contrast with the results of Ch. III on modelling r. There, our SDEs were of the form

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

so the whole risk-neutral dynamics was free: b and σ there had no link due to NA.

Condition (HJM) can be useful in studying NA-properties of models. But when we need to write a concrete model to price and hedge financial products, most of the useful models coming out of HJM are the already-known short-rate models seen earlier (Ch. III), and their multi-factor extensions, which we shall see next (these are particular HJM models, especially Gaussian models) – or the *market models* we shall see later (Ch. V). Even though market models do not necessarily need the HJM framework for their derivation, HJM can serve as a unifying framework in which all categories of NA interest-rate models can be expressed.

HJM and credit risk

The HJM framework may be applied to credit risk (VI below). See e.g. [MG] R. Maksymiuk and D. Gatarek, Applying HJM to credit risk. *Risk* **12**:5 (1999), 67 - 68.

2. Multi-dimensional models and correlations

Recall the Vasicek model: the evolution of the spot-rate process r is given by the linear Gaussian SDE

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

Recall also the Vasicek (more generally, affine) bond-price formula

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

from which all rates can be computed in terms of t. In particular, the continuously-compounded spot rates are given by the following affine transformation of r:

$$R(t,T) = -\frac{\log P(t,T)}{T-t} = -\frac{\log A(t,T)}{T-t} + \frac{B(t,T)}{T-t} r_t =: a(t,T) + b(t,T) r_t.$$

Consider now a payoff depending on the joint distribution of two such rates at time t: for example, $T_1 = t+1$ years and $T_2 = t+10$ years. This would then depend on the *joint* law of the one-year and ten-year continuously-compounded spot rates at time t. So the *correlation* between these two rates plays a crucial role. Now for the Vasicek model, this is 1:

$$Corr(R(t,T_1), R(t,T_2) = Corr(a(t,T_1) + b(t,T_1)r_t, a(t,T_2) + b(t,T_2)r_t) = 1,$$

as there is only *one* source of randomness here. So at each time t, all the maturities in the curve are perfectly correlated: for example, the 30-year interest rate and the 3-month interest rate at the same instant. This means that a shock to the interest-rate curve at time t is transmitted equally through all maturities, and the curve, when its initial point – the spot rate r_t – is shocked, moves almost rigidly in the same direction. This sounds unrealistic in theory, and is observed to be unrealistic in practice also. So a more satisfactory model of curve evolution is needed.

One-factor models such as HW, BK, CIR++ etc. may still prove useful when the product to be priced does not depend on the correlations of different rates, but depends at every instant on a single rate of the whole interest-rate curve – say the six-month rate, for example. Such models may still give an acceptable approximation, e.g. for risk-management purposes, when the rates that jointly influence the payoff are close (say, the six-month and one-year rates). Indeed, the real correlation between such near rates is usually high, so the perfect correlation of a one-factor model may not be unacceptable in principle. But in general, we need to move to a model allowing for more realistic correlation patterns. This can be achieved with *multifactor* models, and in particular with *two-factor models.* Suppose for instance that we replace the Gaussian Vasicek model with its two-factor (additive) version (G2++):

$$r_t = x_t + y_t, \tag{G2++}$$

where

$$dx_t = \kappa_x(\theta_x - x_t)dt + \sigma_x dW_1(t),$$
$$dy_t = \kappa_y(\theta_y - y_t)dt + \sigma_y dW_2(t)$$

and

$$dW_1(t)dW_2(t) = \rho dt.$$

As we shall see, this leads to models with the bond prices affine functions of the *two* factors x and y,

$$P(t,T) = A(t,T)\exp\{-B_x(t,T)x_t - B_y(t,T)y_t\}.$$

This now leads to correlations of the form

$$Corr(R(t,T_1), R(t,T_2) = Corr(b_x(t,T_1)x_t + b_y(t,T_1)y_t, b_x(t,T_2)x_t + b_y(t,T_2)y_t),$$

and this is no longer identically 1, but depends crucially on the correlation between the factors x and y, which in turn depends (among other things) on the instantaneous correlation ρ in their joint dynamics. How much flexibility is gained here in the correlation structure, whether this is worthwhile and whether it suffices for practical purposes remains to be seen (and will depend on the context in which the model is to be used). But this is clearly a step forward.

Again, the question arises: how many factors should we use in practice? The choice of number of factors involves a compromise between numerically efficient computation (keep the number low), and the capacity of the model to fit realistic covariance/correlation patterns and structures satisfactorily in most concrete situations.

Empirical evidence

Usually, historical analysis of the whole yield curve, based on PCA or factor analysis, suggests that under the objective measure *two* components can explain $85\% - 90\%$ of the variations in the yield curve. See for example, Table 1 (p.45) in
Farid JAMSHIDIAN and Yu ZHU, Scenario simulation: Theory and methodology, *Finance and Stochastics* **1** (1997), 43 - 67.
They consider JPY, USD and DEM data (this is pre-Euro! – Germany had the Deutschmark till 2000). They showed that one PC explains 68 - 76 % of the total variation, whereas three PCs can explain 93 - 94 %. A related analysis is carried out in Ch. 3 of Rebonato's book [R1] in interest-rate models (his Table 3.2) for the UK market.
Here things seem more optimistic: one principal component explains 92 % of the total variance, whereas two PCs already explain 99.1 %. In some

studies, an interpretation is given to the PCs, in terms of *average level*, *slope* and *curvature* of the zero-coupon curve; again, see e.g. Jamshidian and Zhu (1997).

To summarise: in the objective – real – world, a two- or three-dimensional process (at least) is needed to model the evolution of the whole zero-coupon curve realistically. When we move from $\mathbb{P}$-measure to $\mathbb{Q}$-measure – from the objective world to the risk-neutral world – the covariance structure does not change: only the *drift* changes when we use Girsanov's theorem. We conclude that two- or three-dimensional models will be needed to get satisfactory results. We focus on these here, for their good tractability and implementability.

Our first model of this kind will be an *additive* model (which we will again call G2++), of the form

$$r_t = x_t + y_t + \phi(t), \qquad (G2++)$$

where $\phi(t)$ is a deterministic shift, added in order to fit exactly the initial zero-coupon curve,. The main advantage of this over the CIR++ model of III.7 is that we need to take the correlation ρ between the two Brownian motions W_1, W_2 there to be 0 to obtain an analytically tractable model, whereas here we do not need to do so. The reason for this is that in the CIR++ case, with ρ non-zero we obtain square-root non-central chi-square processes. These are much harder to handle than linear Gaussian processes; it would not be possible to compute bond prices analytically, and the distribution of r would become intractable. The reason why G2++ is so much preferable to CIR++ here is that the extra correlation parameter ρ gives us much more modelling flexibility. Moreover, $\rho < 0$ allows for a *humped* volatility curve of the instantaneous forward rates, as seen in practice. Indeed, if we consider at time t the graph of the T-function

$$T \mapsto \sqrt{var(df(t,T)/dt)},$$

where the instantaneous forward rate $f(t,T)$ comes from the G2++ model, it can be seen that for $\rho = 0$ this function is decreasing and concave upwards. The function can only assume a humped shape when $\rho < 0$. Now since humped-shaped curves are seen in market behaviour, this is an important advantage for G2++.

In the reverse direction, CIR++ does have advantages over G2++. The distribution of the short rate there is that of the sum of two independent

non-central chi-square random variables, and as such has *fatter tails* than that of the Gaussian laws in G2++. Since nearly all financial data also have fatter tails than Gaussian, this is desirable – and may more than offset the loss of tractability. Also, CIR++ spot rates are affine transformations of such non-central chi-squares, and are closer to the *log-normal* than the Gaussians for the same rates implied by G2++. Of course, log-normally distributed random variables are positive (as in the Black-Scholes model: prices are log-normal and positive, log-prices and returns are normal and change sign). So this is a second advantage for CIR++ over G2++ – but the ability of G2++ to model humped-shaped curves is very important. One cannot have both; one has to choose; the choice will depend on the context – what one is trying to model, and why. In weighing up pros and cons for, say, two-factor models, one might ask questions such as:
Is the model flexible enough to be calibrated to a large set of swaptions, or even to caps and swaptions at the same time?
How many swaptions can be calibrated satisfactorily?
What evolution of the term structure of volatilities is implied by the calibrated model?
Is this realistic?
How can one implement trees for the model?
Is Monte-Carlo simulation feasible?
Can the model be used for products depending on *more* than an interest-rate curve, taking into account correlations between different interest-rate curves, exchange rates, etc?
Here, we will focus mainly on G2++, and address some of these questions.