ullint7a.tex am Wed 14.3.2018

VI.3 (continued)

The spread – or intensity – λ behaves just like an interest rate in discounting. So we can now use techniques we used before (III) in studying short-rate models r for credit models for λ. This is very convenient – and is a good reason for studying short-rate models (III) before we progress to the more widely used but more complicated market models (V).

Again just as with short-rates, intensities can be constant, deterministic or stochastic.

Constant. $\lambda_t \equiv \gamma$

(so as to reserve the symbol λ for $t \mapsto \lambda_t$.

Deterministic. $\lambda_t = \gamma(t)$, for some deterministic function $\gamma(.)$. This is a model with a term structure of credit spreads, but without credit-spread volatility.

Stochastic. Here $\lambda = (\lambda_t)$ is a full stochastic process – e.g., a Cox process (above). This allows us to model both the term structure and the volatility of credit spreads.

CDS pricing with constant $\lambda_t = \gamma$.

Assume as an approximation that the CDS premium leg pays continuously. Instead of paying $(T_i - T_{i-1})R$ at T_i as in the standard CDS, given that there has been no default before T_i, we approximate this premium leg by assuming that it pays Rdt in $[t, t+dt)$ given no default before $t+dt$. This amounts to replacing the CDS pricing formula of VI.2 (receiver case, so cash-flows for the payer case are reversed; spot CDS with $T_a = 0$, today),

$$\begin{aligned} CDS_{0,b}(0, R, LGD; \{(\tau > .)\}) \quad &= \quad R[-\int_0^{T_b} P(0,T)(t - T_{\beta(t)-1}) d_t \mathbb{Q}(\tau \geq t) \\ &\quad + \sum_{i=1}^{b} P(0,T_i)\alpha_i \mathbb{Q}(\tau \geq T_i)] \\ &\quad + LGD[\int_0^{T_b} P(0,t) d_t \mathbb{Q}(\tau \geq t)], \end{aligned}$$

with (accrual term on RHS vanishes as payments are now continuous)

$$\begin{aligned} CDS_{0,b}(0,R,LGD;\{(\tau>.)) &= R\int_0^{T_b} P(0,t)\mathbb{Q}(\tau\geq t)dt \\ &\quad +LGD[\int_0^{T_b} P(0,t)d_t\mathbb{Q}(\tau\geq t)]. \end{aligned}$$

But for constant intensity γ, as above

$$\mathbb{Q}(\tau>t)=e^{-\gamma t},\qquad d_t\mathbb{Q}(\tau>t)=-\gamma e^{-\gamma t}dt=-\gamma\mathbb{Q}(\tau>t)dt.$$

Substituting, this now becomes

$$CDS_{0,b}(0,R,LGD;\{(\tau>.))=(R-\gamma LGD)\int_o^{T_b} P(0,t)\mathbb{Q}(\tau\geq t)dt.$$

If we inset the market CDS rate

$$R=R_{0,b}^{mktMID}(0),$$

then the CDS present value should be 0. So

$$\gamma=R_{0,b}^{mktMID}(0)/LGD.$$

This shows that the CDS premium rate R is indeed a sort of credit spread, or intensity, as γ is.

Example: FIAT.

CDS of FIAT trades at 300bps for 5y, with recovery 0.3. What is a quick rough estimate for the risk-neutral probability that FIAT survives 10 years?

$$\gamma=R_{0,b}^{mktMID}(0)/LGD=\frac{300/10,000}{1-0.3}=4.29\%.$$

For survival for 10y:

$$\mathbb{Q}(\tau>10y)=\exp\{-10\gamma\}=\exp\{-10\times 0.0429\}=65.1\%.$$

Default between 3 and 5 years:

$$\mathbb{Q}(\tau>3y)-\mathbb{Q}(\tau>5y)=\exp\{-3\times 0.0429\}-\exp\{-5\times 0.0429\}=7.2\%.$$

CDS pricing with time-dependent intensity $\lambda_t = \gamma(t)$.

We now consider a deterministic time-varying intensity $\gamma(t)$, which we take positive and piecewise-continuous. Define

$$\Gamma(t) := \int_0^t \gamma(u)du,$$

the *hazard function.* Then in the notation above,

$$\begin{aligned}
\mathbb{Q}(s < \tau \leq t) &= \mathbb{Q}(s < \Lambda^{-1}(\xi) \leq t) \qquad (\tau = \Gamma^{-1}(\xi), \quad \xi \sim E(1)) \\
&= \mathbb{Q}(\Gamma(s) < \xi \leq \Gamma(t)) \\
&= \mathbb{Q}(\xi > \Gamma(s)) - \mathbb{Q}(\xi > \Gamma(t)) \\
&= \exp\{-\Gamma(s)\} - \exp\{-\Gamma(t)\}.
\end{aligned}$$

Reduced-form models are the models most commonly used in the market to infer implied default probabilities from market quotes. The market instruments mostly used here are CDS and bonds.

To implement, we use the stripping algorithm sketched earlier (CDS stripping), but not taking into account that the probabilities are expressed as exponentials of the deterministic intensity $\gamma(.)$, assumed piecewise constant. By adding iteratively CDSs with longer and longer maturities, at each step we will strip a new part of the intensity $\gamma(t)$ associated with the last added CDS, while keeping the previous values of $\gamma(.)$, for earlier times, that were used to fit CDSs with shorter maturities.

CIR++ stochastic intensity model

We model the stochastic intensity as

$$\lambda_t = y_t + \psi(t; \beta), \qquad t \geq 0,$$

where the intensity has a random component y and a deterministic component ψ chosen to fit the CDS term structure. For y we take a *jump-CIR model (JCIR)*:

$$dy_t = \kappa(\mu - y_t)dt + \nu\sqrt{y_t}dZ_t + dJ_t, \qquad \beta = (\kappa, \mu, \nu, y_0), \quad 2\kappa\mu > \nu^2.$$

The jumps are taken independent of everything else, with exponential arrival times with intensity η and exponential jump-size with some parameter. We confine ourselves here to no jumps, $J = 0$; for the jump case, see e.g. [BM, 22.8]. Then with no jumps:

$$dy_t = \kappa(\mu - y_t)dt + \nu\sqrt{y_t}dZ_t,$$

where, as in III.3:
κ is the speed of mean reversion,
μ is the long-term mean reversion level,
ν is the volatility.
We refer there for the mean and variance.

We need to keep ψ, and hence λ, *positive*, as is required for intensity models. For this, see again [BM, 22.8].

If we can read off from market data the implied hazard functions Γ^{Mkt}, as in the Lehman example (below), then the model agrees with this if

$$\exp\{-\Gamma^{Mkt}(t)\} = \exp\{-\psi(t;\beta)\}E[\exp\{-\int_0^t y_s ds\}].$$

Note:
1. This is only possible if λ is strictly positive.
2. To perform this calibration, we need to be able to evaluate the expectation above analytically.
The only known diffusion model used in interest-rates theory satisfying both these constraints is CIR. Then,

$$\exp\{-\Gamma^{Mkt}(t)\} = \mathbb{Q}(\tau > t) = \exp\{-\psi(t;\beta)\}E[\exp\{-\int_0^t y_s ds\}].$$

But here, the expectation on the RHS is just the CIR bond price with short-rate y_t, and we know this analytically from III.3. Write it as $P^y(0, t, y_0; \beta)$.

As in the short-rate case, λ is calibrated to the market-implied hazard function Ψ^{Mkt} if we set

$$\Psi(t;\beta) := \Gamma^{Mkt}(t) + \log P^y(0, t, y_0; \beta)).$$

Here we choose the parameters β to have a positive function ψ, as before.

4. Firm value (structural) models

In VI.3 we considered intensity models. These are called *reduced* models, in contrast to the models we turn to now, the firm-value or structural models. These have more economics content, and so are more linked to the big picture – the wider world, and the uncertainty driving it, whereas the intensity models here in VI.4 are defined by

$$\tau = \Lambda^{-1}(\xi),$$

but say nothing about how these relate to the underlying economic causes of default.

Our stylised model is as follows: $V(t)$ is a stochastic process representing the *value* of the firm;
$t \mapsto H(t)$ is a barrier representing debt and safety covenants. This is often estimated using balance-sheet data: short-term debt, long-term debt etc.

Note.

Balance sheets depend on *double-entry book-keeping*. This goes back to Renaissance Italy (Luca Pacioli (1445-1514), *Summa*, 1494). The essence of this is that every transaction has two aspects: one as a debit, the other as a credit (compare Newton's Laws of Motion, one of which is: to every action, there is an equal and opposite reaction). **Every student of Mathematical Finance should learn double-entry book-keeping** – basic, and very useful. (You can teach yourself in an afternoon – just get a book on it.)

Now the default time τ is the first time that the value V touches the safety barrier H. Compare the ruin problem in insurance: MATL480, VII.

Comparison.
Intensity models: suited to :
modelling credit spreads; easier to calibrate to corporate bond or CDS data; suited to refined relative-value pricing – CDS options, bonds with optionalities, etc.
Structural models.
These are more suited to fundamental pricing and risk analysis than to relative value pricing. They are difficult to calibrate accurately to CDS or bonds. They are used for, e.g., equity return swaps with conterparty risk, total rate

of return swaps, counterparty risk in any equity product, and Equity Default Swaps.
They extend more naturally to multi-name situations (with a portfolio of assets) than intensity models do (via copulas – below).

Merton's model.
The first structural model is that of Merton (1974). Here the value V of the firm is assumed to follow a geometric Brownian motion:

$$dV(t) = mV(t)dt + \sigma V(t)dW(t), \qquad (GBM)$$

$$m = \mu \text{ under } \mathbb{P}, \qquad m = r - k \text{ under } \mathbb{Q};$$

here μ is the return under the historical measure, r the risk-free interest rate, k the payout ratio and σ the volatility. As in MATL480 VI, the dynamics are log-normal. This seems to conform fairly well with actual data.
The value V is decomposed into the equity part S and the debt part D:

$$V(t) = D(t) + S(t), \qquad \text{Firm value} = \text{Debt value} + \text{Equity}.$$

The debt structure is that $D(t)$ is a zero-coupon debt with maturity $\overline{T}$ and face value L. If at maturity the firm value $V > L$, then the debt is repaid in full and the firm survives. If however $V < L$ at maturity, default occurs. So in this model, default can only happen at the debt maturity $\overline{T}$.
Because we know how to solve the SDE (GBM) (MATL480, VI.1), we can read off the firm value at T (and so the probability of default when $T = \overline{T}$):

$$P(V(T) < L) = \Phi\Big(\frac{\log(L/V(0)) - (M - \frac{1}{2}\sigma^2)T}{\sigma\sqrt{T}}\Big) = \Phi(-d_2),$$

as in the calculation for the Black-Scholes formula). The hazard rate is

$$\lim_{T \downarrow 0} P(\tau \le T)/T = 0.$$

Compare this with the standard constant-hazard rate intensity model, where

$$\mathbb{Q}(\tau \le t) = 1 - e^{-\gamma T}, \;\Rightarrow\; \lim_{T\downarrow 0} \mathbb{Q}(\tau \le T)/T = \gamma > 0.$$

This is an important difference: basic structural models such as Merton's have zero short-term credit spreads, while intensity models have positive short-term credit spreads. This is a modelling advantage of intensity models. Merton's model does not work well for very short maturities.

For calibration: we quote that this works well if the barrier has the form

$$H(t) = \frac{H}{V_0} E[V_t] \exp\{-B \int_0^t \sigma_s^2 ds\}.$$

Because of its similarity to the Black-Scholes option-pricing formula, the structural approach above is also referred to as the *option-theoretic approach*.

Optimal time to default

When a firm is under pressure, and it realises there is a real danger of bankruptcy, the board may have some leeway as to *when to go bankrupt*. Too early, and the chance to survive to 'fight another day', when circumstances may have changed in one's favour, may be missed. Too late, and the board may decide to 'admit defeat', rather than 'throw good money after bad'. Furthermore, trading while insolvent is illegal. How best to proceed in this 'grey zone' varies according to bankruptcy law, which varies from country to country (e.g., USA, filing under Chapter 11 for protection from creditors). Early work here was done by Black (of Black-Scholes) and Cox (of Cox-Ross-Rubinstein):
[BC] F. BLACK and J. C. COX, Valuing corporate securities: Some effects of bond indenture provisions. *J. Finance* **31** (1976), 351 - 367.

5. Dependence and copulas: Tails and diversification

Higher dimensions; joint and marginal distributions

We often have to look at a number of different assets together. With stocks, we will hold a *portfolio* of different assets, by *Markowitzian diversification*, chosen to have lots of *negative correlation*. With interest rates, we may be dealing with a number of different currencies, and/or with a number of different products denominated in the same currency. We need to be able to handle the resulting probability theory in higher dimensions.

If $X = (X_1, \ldots, X_n)$ is a random variable taking values in n-dimensional space – a random n-vector – then its distribution function F is defined as above, but coordinatewise. If $x = (x_1, \ldots, x_n)$, $y = (y_1, \ldots, y_n)$, we write

$$x \leq y \text{ iff } x_1 \leq y_1, \ldots, x_n \leq y_n.$$

Then

$$F(x) := P(X \leq x) = P(X_1 \leq x_1, \ldots, X_n \leq x_n).$$

This is also called the *joint* distribution of $(X_1, \ldots, X_n)$, while

$$F_i(x_i) := P(X_i \leq x_i), \qquad i = 1, \ldots, n$$

is called the *marginal* distribution of X_i. Note that letting the jth argument $x_j \to \infty$ eliminates the condition $X_j \leq x_j$, and so leaves the joint distribution of the Xs with X_j omitted. So the joint distribution of a random vector determines the joint distribution of any subvector, and the marginals of its coordinates, just by letting unwanted arguments go to $+\infty$. In sum: *the joint determines the marginals.*

To summarise: the *joint* distribution tells us how the variables behave *together* (interactions, dependence, correlation, ...);
the *marginal distributions* tell us how the variables behave *separately*;
the *joint* determines the *marginals*. But the converse does *not* hold, except when the variables are *independent* – do not interact with each other. From the present point of view this is a trivial case: we should then look at them *separately*, and we are then back in *one* dimension, for each asset separately.

The question arises as to how we can split the information in the *joint* into two parts:
(i) the information in the *marginals*;
(ii) the remaining information, which governs dependence and interaction.
It turns out that we can do this, by using *copulas* and *Sklar's theorem*, below.

Probability Integral Transformation (PIT).

As F is non-decreasing, it has an inverse function. We use

$$F^{-1}(x) := \inf\{x : F(x) \geq t\} = \min\{x : F(x) \geq t\}$$

(also non-decreasing, but left-continuous – so the infinum is attained, i.e. is a minimum). Write $X \sim F$ to mean that the random variable X has distribution F. Then if $U[0,1]$ is the uniform distribution above (probability = length) and $U \sim U[0,1]$, then U is uniformly distributed on $[0,1]$; we shall use this standard notation below. The *Probability Integral Transformation (PIT)* uses U and F to generate X:

$$X := F^{-1}(U) \sim F. \qquad (PIT).$$

Proof. $P(X = F^{-1}(U) \leq x) = P(U \leq F(x)) = F(x)$. //

The PIT is very useful in the context of Simulation (using computers to generate random numbers); see e.g. NHB > SMF > IS, I and p.2. It means that we only need random number tables for the uniform distribution $U[0,1]$, and can then use (PIT) to transform this data to have distribution F.

Copulas

The question arises of how to go in the reverse direction. It is helpful to think of the information in the joint distribution as composed of two parts: one on the marginals, the other on the *dependence* between the coordinates – often of great statistical importance! One needs a function that *couples* the marginals together to form the joint. This is called the *copula.*

A *copula* C in n dimensions is a probability distribution function on (= supported by – all its probability mass is on) the unit n-cube $[0,1]^n$.

Sklar's Theorem (A. SKLAR, 1958). If $F(x)$ is a joint distribution in n dimensions, with marginals $F_i(x_i)$, there exists an n-dimensional copula C with

$$F(x) = F(x_1, \ldots, x_n) = C(F_1(x_1), \ldots, F_n(x_n)).$$

Conversely, given any copula C and marginals F_i, this formula gives a joint distribution F with marginals F_i. The correspondence between F and C is unique if the marginals F_i are continuous.

Gaussianity, tails and extremes

The basic importance of *Gaussian* distributions is obvious throughout this course and its predecessor MATL480. In particular, we use *Brownian motion* as our basic model for driving noise, and this produces Gaussians (normals) – or lognormals, simply related to them.

There are two basic problems with Gaussianity in financial modelling:
(i) Gaussians are symmetric. Real financial data are not. They show *skew* – basically, a reflection of the profound asymmetry between *profit* and *loss.* Big losses are lethal; big profits are just nice to have.
(ii) Gaussians have *extremely thin tails* – 'minus the log-density' grows fast (quadratically). Real financial data have *much thicker tails.* How much

thicker depends on the context; e.g., typically,
(i) monthly returns are (roughly) Gaussian (aggregational Gaussianity – Central Limit Theorem, CLT);
(ii) daily returns have 'minus log density' growing much more slowly (linearly – hyperbolic distributions), so the tails are *much fatter* than Gaussian tails;
(iii) high-frequency returns (tick data) have 'minus log density' growing only logarithmically – so their tails are *much fatter still*.

Because large losses are lethal, one needs to be very careful about how one models the probability of a large loss. This involves the (extreme) *tails* of the relevant distributions – *extreme values*, or just *extremes*, of the corresponding random variables (big losses, in this case).

We need to *diversify*, and hold a number of assets.

So: we need to know how the move from one to higher dimensions interacts with the distributions we use, and in particular, how the probabilities of big losses behave here.

There are various ways to measure the interaction, or dependence, of such probabilities for different variables occurring together. They reveal that, as a direct result of the extreme thinness of Gaussian tails:
if the joint distribution is Gaussian, the tails are effectively independent.
For background and references, see e.g.
[BS] N. H. BINGHAM and Rafael SCHMIDT, Interplay between distributional and temporal dependence: an empirical study with high-frequency asset returns. *The Shiryaev Festschrift: From stochastic calculus to mathematical finance*, Springer, 2006, 69-90.

This means that, for Gaussians, **Markowitzian diversification is completely useless for protecting against the possibility of large losses**.

This basic and important fact was not appreciated by higher management of financial institutions pre-Crash (see below), and was one of the predisposing factors that led to it. Not only did management not know this – which lulled them into a false sense of security – they felt reassured by mathematics (a pity, as mathematics – above – could have saved them). Misapplication of a formula, due to David X. Li, involving the *Gaussian copula*, was a major factor leading up to the Crash. This became notorious as *the formula that killed Wall Street*:

$$C_\rho(u, v) = \Phi_\rho(\Phi^{-1}(u), \Phi^{-1}(v)).$$

For further background on the Gaussian copula, see e.g.
Felix SALMON, Recipe for disaster: The formula that killed Wall Street.

Wired Magazine 17:3, March 2009;
David X. LI, On default correlation: a copula function approach. *J. Fixed Income* **9**:4 (2000), 43-54;
D. MacKENZIE and T. SPEARS, 'The formula that killed Wall Street'? The Gaussian copula and the material culture of modelling. Internet.

For more on copulas in finance, see e.g.
[CLV] Umberto CHERUBINO, Elisa LUCIANO and Walter VECCHIATO, *Copula models in finance*. Wiley, 2004.

Quantitative risk management (QRM)

The subject of QRM is extremely important, and has become more so since the Crash. See for example
[MFE] Alexander J. McNEIL, Rüdiger FREY and Paul EMBRECHTS, *Quantitative risk management: Concepts, techniques, tools*, revised ed. Princeton University Press, 2015 (1st ed. 2005).
This book covers copula methods, the Gaussian copula and its asymptotic independence. The revised edition incorporates the changes needed since the Crash.

Build-up to the Crash

Various factors were involved here (see 7b for more on this).
(a) The sub-prime mortgage expansion in the US housing market led to many mortgages being extended to borrowers who would not have been judged credit-worthy before, and some of whom were not.
(b) Such 'sub-prime assets' were packaged up into job-lots, and sold. This process is a form of *securitization* (7b).
(c) Banks typically thought that holding numbers of such assets would increase their security, compared with the security of an individual asset. They imagined that they were *diversifying* (in the style of Markowitz: MATL480), and so making things safer.
(d) When the bubble burst, many of these assets went into free fall, all responding to the same thing – the bursting of the bubble (remember: bubbles do burst, eventually). The market then went into crisis mode. So then, it was the *tails* of the distributions that mattered.
(e) In many cases, banks were relying on Gaussian copulas, which as we have seen are useless for diversification in a crisis, *because their tails are so thin*

– much thinner than real financial data. Senior bankers were shocked by this; some complained that they had been assured that such a combination of asset failures would occur only extremely rarely (hundreds or thousands of years), whereas such things were happening in quick succession.
(f) In effect, so far from diversifying, packaging up such low-grade assets simply meant that the worst 'infected', or 'polluted', the others, hence the term *toxic debt* (7b).
Comments
1. All this is rather reminiscent of the *Lloyds of London scandal*, that hit the UK insurance market in the 1990s. This was not only a very serious crisis in the industry, it had a devastating effect on *confidence* in the UK insurance industry. Recall that *the UK pioneered insurance* – Lloyd was British!
2. In both cases – insurance in the 1990s, the Crash in 2007/08 – leaders of the financial institutions involved were dangerously under-informed or mis-informed. These matters involve serious mathematics, and this lay well beyond the technical competence of many at board level, even CEOs.
3. The main problem was not so much that the mathematics used by the technical staff of the institutions was wrong – Li's Gaussian copula formula is not wrong, though it is very dangerous when wrongly used, out of its proper context. The problem was lack of proper liaison between company employees with mathematical and statistical competence, and those in higher management, whose backgrounds were typically banking.