ullint5a.tex am Wed 28.2.2018

8. LMM: Calibration to market data

Market models were introduced in 1997, and have become by far the most important models in use in interest-rate theory. So they deserve a study in some depth and detail. Note that they have been around for twenty years now, ten pre-Crash and ten post-Crash. So they have stood the test of time, and shown the flexibility needed to adjust to the very different conditions of the world post-Crash.

One of the main reasons why market models have taken over centre-stage – and why they got their name – is that they allow practitioners to *calibrate their models to market data*. Such data is very extensive. The market in interest-rate products amounts to trillions of dollars; most of this is in the few hundred most heavily traded products. So these are highly liquid. The ability to perform well in trading – principally, pricing and hedging – depends on, among other things, the ability to *model* (above – Probability and Stochastics), and to handle and interpret *data* (Statistics, and Numerical Analysis). Small advantages matter (recall that practitioners think in terms of *basis points (bp)* – hundredths of a percent. But such small fractions of trillions is still a very large amount ... .

For more on market models generally, see e.g. [BM, Part III Ch. 6-8]; for more on calibration, see e.g. [BM, 6.4, 6.17, Ch. 7].

Inputs:
standard liquid products:
FRAs, swaps, caps, ...
Model:
LMM: correlations ρ, volatilities σ, ...
Outputs:
Exotic products: ratchet caps, constant maturity swaps (CMS), ...
Prices, hedges and risk.

LMM is natural for caps, and SMM is natural for swaptions. We choose LMM, and adapt it to price swaptions later.

Recall: under numeraire $P(.,T_i) \neq P(.,T_k)$,

$$dF_k(t) = \mu_k^i(t)F_k(t)dt + \sigma_k(t)F_k(t)dZ_k, \qquad dZdZ^T = \rho dt.$$

Model specification: Choice of $\sigma_k(t)$ and ρ.
The LMM is completely specified once we specify the $\sigma_k(t), \rho_{ij}$ and $F_k(0)$ for all i, j, k in the tenor structure – that is, for all the relevant times, at which we have data.

Parametrisation of instantaneous covariances

Divide the relevant time-range into intervals $(0, T_0], (T_0, T_1], \cdots, (T_{M-2}, T_{M-1}]$. With these intervals labelling the *rows* in the covariance matrix, and the forward rates $F_1(t), \cdots, F_M(t)$ labelling the *columns*, the entries *above* the diagonal will correspond to *expired* products, so are omitted. The matrix is thus 'diagonal + sub-diagonal'. Drawing a picture will produce a *ziggurat* shape.
General piecewise constant (GPC) vols are of the form

$$\sigma_k(t) = \sigma_{k,N(t)}, \qquad T_{N(t)-2} < t \leq T_{N(t)-1}.$$

Separable piecewise constant (SPC):

$$\sigma_k(t) = \Phi_k.\psi_{k-(N(t)-1)}.$$

Parametric linear-exponential (LE) vols:

$$\sigma_i(t) = \Phi_i.\psi(T_{i-1} - t; a, b, c, d) = \Phi_i.([a(T_{i-1} - t) + d]e^{-b(T_{i-1}-t)} + c)).$$

Caplet volatilities

Recall that under numeraire $P(., T_i)$,

$$dF_i(t) = \sigma_i(t)F_i(t)dZ_i, \qquad dZdZ^T = \rho dt.$$

Caplet: strike rate K, reset T_{i-1}, payment T_i: payoff $\tau_i(F_i(T_{i-1}) - K)_+$ at T_i.
This corresponds to a *call option* on F_i, which is lognormal under $\mathbb{Q}_i$. This gives *Black's formula* with Black volatility parameter

$$v_{i-cap}^2 := \frac{1}{T_{i-1}} \int_0^{T_{i-1}} \sigma_i(t)^2 dt;$$

v_{i-cap} is the T_{i-1}-*caplet volatility.*
Only the volatilities σs have any impact on caplet (and so on cap) prices; the

correlations ρ have no effect.
GPC vols:

$$v_{i-cap}^2 = \frac{1}{T_{i-1}} \sum_{j=1}^{i} (T_{j-1} - T_{j-2}) \sigma_{ij}^2.$$

LE vols:

$$T_{i-1} v_{i-cap}^2 = \Phi_i^2 \int_0^{T_{i-1}} (([a(T_{i-1} - t) + d] e^{-b(T_{i-1} - t} + c)^2 dt.$$

For GPC, caplet volatilities can be computed very simply, as follows. Take the volatilities matrix (ziggurat-shaped):
(a) Square each entry.
(b) For each row, sum the squared terms, each multiplied by the year-fraction expiry-to-maturity τ for that volatility.
(c) Take the total in (b) and divide it by the caplet reset time (= sum of all τs used in that row).
(d) Take the square root.

Term structure of caplet volatilities (TSOV)

The term structure of volatility (TSOV) at time T_j is a graph of expiry times T_{k-1} against average volatilities $V(T_j, T_{k-1})$ of the related forward rates $F_k(t)$ up to that expiry time itself, i.e. for $t \in (T_j, T_{k-1})$. So at time $t = T_j$, this is a graph of points

$$\{(T_{j+1}, V(T_j, T_{j+1})), (T_{j+2}, V(T_j, T_{j+2})), \cdots, (T_{M-1}, V(T_j, T_{M-1})), \},$$

$$V^2(T_j, T_{k-1}) = \frac{1}{T_{k-1} - T_j} \int_{T_j}^{T_{k-1}} \sigma_k^2(t) dt, \qquad k > j + 1.$$

In the ziggurat matrix, we can easily compute the future TSOV at time T_j much as before:
(a) Square each entry. Starting from the column corresponding to the desired future time, in each row add up all the squares up to the diagonal, each multiplied by the relevant year-fraction τ.
(b) Take the total in (a); divide by the sum of the τ used; take the square root.
(c) To compute the TSOV at all future times, a calculation with cumulative sums of squares going backwards is ideal.

Example.

Implement with:
Semi-annual tenors, $T_i - T_{i-1} = 6m$;
Instantaneous correlation matrix estimated historically, first fitted on the full-rank parametric form

$$\rho_\infty + (1 - \rho_\infty) \exp\{-\alpha|i - j|\},$$

and then possibly fitted to a reduced-rank correlation (no impact on caps, but needed for ratchets etc.; more on this later).

Three specimen choices:
First plot: $\Phi = 1$ (homogeneous in time-to-expiry).
Second plot: $\psi = 1$ (homogeneous in time).
Third plot: intermediate (neither Φ nor ψ set to 1).

If traders have no view on future TSOV, they prefer a stationary model, so use the first plot, $\Phi = 1$.

In the second plot, with $\psi = 1$, the TSOV 'collapses onto its own tail' – 'flattens out to zero; – as time evolves. Traders do not like this: it assumes future volatilities will be much lower than current ones. Unless there is good economic reason to believe this (there isn't!), this choice should be avoided, although it is the one that makes calculations and calibration easiest, and makes terminal and instantaneous correlations equal (see below).

So realistically, we will get something like the third plot – roughly stationary, but showing some evolution, with the Φs close to 1 but not exactly 1.

Terminal and instantaneous correlation

Swaptions depend on *terminal* correlations among forward rates. For example, the swaption whose underlying is $S_{1,3}$ depends on

$$corr(F_2(T_1), F_3(T_1)).$$

This terminal correlation depends both on the *instantaneous* correlation $\rho_{2,3}$ *and* on the way the $T_1 - T_2$ and $T_2 - T_3$ caplet volatilities are decomposed into instantaneous vols $\sigma_2(t)$ and $\sigma_3(t)$ for $t \in [0, T_1]$. Under GPC vols,

$$corr(F_2(T_1), F_3(T_1)) \sim \rho_{2,3} \frac{\sigma_{2,1}\sigma_{3,1} + \sigma_{2,2}\sigma_{3,2}}{v_1\sqrt{T_1}\sqrt{\sigma_{2,1}^2 + \sigma_{2,1}^2}},$$

in our previous notation. No such formula holds for general short-rate models! This shows again why market models matter – and why they are worth taking the trouble to study in detail.

For more on calibration in practice, see
[Sid] J. SIDENIUS, LIBOR market models in practice. *J. Computational Finance* **3**:3 (2000), 5-26.

9. Instantaneous correlation: parametric forms

Swaptions depend on *terminal* correlation among forward rates (ρs and σs). How do we model ρ? But first, what general patterns would we like the correlation matrix to show?

Corrrelation matrices have 1 on the diagonal, and are symmetric, so we confine attention to the *subdiagonal* part. As we move *away* from the diagonal, we expect the entries to *decrease*: the movement of the 6m-1y rate will be more correlated with that of the 1y-1y6m rate than with that of the 9y-9y6m rate. The entries are also expected to *increase* along the subdiagonals. For, the curve is expected to move more rigidly (more correlated) for long maturities than for short ones. We are very sensitive to expectations of changes between now and in 6 months. We do not have the predictive capability to be sensitive to the difference between 30y from now and 30y6m.

To reflect all this, various parametric forms have been proposed. We turn now to some of the principal ones. Note that we are dealing here with *matrices*. These are studied in Linear Algebra in Mathematics – the study of vector spaces, linear transformations between them, matrices, determinants etc. Depending on your background here, you may wish to revise what you have learned, or learn for yourself (that is what libraries are for!). You may find the following link on my homepage useful:
NHB, SMF (Statistical Methods for Finance), Ch. III (Multivariate Analysis).
Multivariate Analysis forms an important part of Statistics. Numerical Linear Algebra forms an important part of Numerical Analysis. Both are stiff with matrix theory.

You will find the idea of the *rank* of a matrix important. This is the maximum number of linearly independent rows (or columns). If this is as big as it could be given the size of the matrix (which need not be square), the matrix has *full rank*. This is the general, or typical, or non-degenerate case. Otherwise the matrix has *defective rank*. Exceptional cases are usually of this

kind. For example: the multivariate normal distribution $N(\mu, \Sigma)$ splits into the full-rank case – the covariance matrix Σ is non-singular, and there is a density (given by Edgeworth's theorem, below), and the defective-rank case. This is degenerate in the original dimensionality, but not in the appropriate lower dimension – so one should start again, and work there. Example: temperature in Centigrade and Fahrenheit. Each determines the other (just match up freezing and boiling points of water) – so the situation is really *one*-dimensional, rather than two.

The Multinormal Density; Edgeworth's theorem.

If $\mathbf{X}$ is n-variate normal, $N(\mu, \mathbf{\Sigma})$, its density (in n dimensions) need not exist (e.g. the singular case $\rho = \pm 1$ with $n = 2$). But if $\mathbf{\Sigma} > \mathbf{0}$ (so $\mathbf{\Sigma}^{-1}$ exists), $\mathbf{X}$ has a density. The link between the multinormal density below and the multinormal CF (or MGF) is due to the English statistician F. Y. Edgeworth (1845-1926) in 1893; see e.g. SMF, IV.3.

Theorem (Edgeworth, 1893)). If μ is an n-vector, $\mathbf{\Sigma} > \mathbf{0}$ a symmetric positive definite $n \times n$ matrix, then
(i)

$$f(\mathbf{x}) := \frac{1}{(2\pi)^{\frac{1}{2}n} |\mathbf{\Sigma}|^{\frac{1}{2}}} \exp\{-\frac{1}{2}(\mathbf{x}-\mu)^T \mathbf{\Sigma}^{-1}(\mathbf{x}-\mu)\}$$

is an n-dimensional prob. density function (of a random n-vector $\mathbf{X}$, say);
(ii) $\mathbf{X}$ has CF $\phi(\mathbf{t}) = \exp\{i\mathbf{t}^T\mu - \frac{1}{2}\mathbf{t}^T\mathbf{\Sigma}\mathbf{t}\}$, MGF $\phi(\mathbf{t}) = \exp\{\mathbf{t}^T\mu + \frac{1}{2}\mathbf{t}^T\mathbf{\Sigma}\mathbf{t}\}$;
(iii) $\mathbf{X}$ is multinormal $N(\mu, \mathbf{\Sigma})$.

The following paper has been important for what follows:
[SC] J. SCHOENMAKERS and B. COFFEY: Systematic generation of parametric correlation structures for the LIBOR market model. *Internat. J. Theor. Appl. Finance* **6** (2003), 507-519.
You can see how recent some of the major developments are! You should bear this in mind when choosing a book. We have drawn here on [BM, 6.9.1]. This material is important, but not mentioned in most earlier books.

Full-rank parametric forms for instantaneous correlations ρ

Schoenmakers and Coffey propose a finite sequence

$$1 = c_1 < c_2 < \cdots < c_M, \qquad c_1/c_2 < c_2/c_3 < \cdots < c_{M-1}/c_M,$$

and they set (F here stands for Full (Rank))

$$\rho^F(c)_{ij} := c_i/c_j, \qquad i \le j, \qquad i,j = 1,\cdots,M. \qquad (SC-c)$$

So the correlation between changes in adjacent rates is

$$\rho^F_{i+1,i} = c_i/c_{i+1};$$

these are all < 1, and are *increasing* in i. Both these are desirable features, in view of the above.

So: under (SC), *the subdiagonal of the correlation matrix $\rho^F(c)$ is increasing when moving from NW to SE.* Interpretation: as we move along the yield curve, the larger the tenor, the more correlated changes in adjacent forward rates become. This corresponds (not only to the expectation above, but also) to the experienced fact that the forward curve tends to flatten, and to move in a more correlated way, for large maturities than for small ones.

The number of parameters needed for a Schoenmakers-Coffey matrix is M, rather than the $\frac{1}{2}M(M-1)$ parameters needed for a general correlation matrix of the same size. Schoenmakers and Coffey showed (we quote this) that any matrix as in $(SC-c)$ is permissible here, and gave an alternative parametrisation, $(SC-\Delta)$ below:

Theorem (Schoemmakers-Coffey, 2003).
(i) Any matrix $C = (c_{ij})$ satisfying $(SC-c)$ is a genuine correlation matrix – symmetric, positive semi-definite and with 1s on the diagonal.
(ii) This parametrisation can be characterised alternatively in terms of

$$\Delta_2,\cdots,\Delta_M \ge 0 :$$

$$c_i = \exp\{\sum_{j=2}^{i} j\Delta_j + \sum_{j=i+1}^{M}(i-1)\Delta_j\}. \qquad (SC-\Delta)$$

Some useful particular cases of SC parametrisations are as follows. First, take all the Δs zero except for the last two. Then changing notation one has:

$$\rho_{i,j} = \exp\{-\frac{|i-j|}{M-1}\Big(-\log\rho_\infty + \eta\frac{M-1-i-j}{M-2}\Big)\}.$$

This is a promising choice for parametrisation of correlation: it is:
(a) two-parameter;

(b) stable (numerically – small changes in the c-parameters cause small changes in the two parameters, ρ_∞ and η;
(c) full rank;
(d) increasing along subdiagonals.
Note that $\rho_\infty = \rho_{1,M}$ is the correlation between the rates furthest apart, whereas η is related to the first non-zero Δ, which comes at the end:

$$\eta = \frac{1}{2}\Delta_{M-1}(M-1)(M-2).$$

A three-parameter form is obtained with the Δs following a straight line (two parameters) until the last one, whose value is the third parameter. This leads to a stable, full-rank, 3-parameter parametrisation increasing along subdiagonals,

$$\begin{aligned}\rho_{i,j} &= \exp\{-|i-j|\Big(\beta - \frac{\alpha_2}{6M-18}(i^2+j^2+ij-6i-6j-3M^2+15M-7)\\ &\quad + \frac{\alpha_1}{6M-18}(i^2+j^2+ij-3Mi-3Mj+3i+3j+3M^2-65M+2)\Big)\},\end{aligned}$$

which we call $(SC3)$ (3-parameter SC). However, experience has shown that the final Delta, Δ_{M-1}, is nearly always close to zero. So we lose little, but gain in simplicity, by moving to the two-parameter version of $(SC3)$:

$$\rho_{i,j} = \exp\{-|i-j|\Big(-\log\rho_\infty + \eta\frac{(i^2+j^2+ij-3Mi-3Mj+3i+3j+2M^2-M-4)}{(M-2)(M-3)}\Big)\}. \qquad (SC2)$$

As before, $\rho_\infty = \rho_{1,M}$, whereas η is related to the steepness of the straight line in the Δs.
Full-rank, two-parameter, exponentially decreasing parametrisation

Schoenmakers and Coffey also introduced the model

$$\rho_{i,j} = \rho_\infty + (1-\rho_\infty)\exp\{-\beta|i-j|\}, \qquad \beta \geq 0.$$

Here ρ_∞ still represents the correlation between the ends, but only asymptotically (let $j \to \infty$).
Rebonato's model.

Rebonato (1999) introduced the model

$$\rho_{i,j} = \rho_\infty + (1-\rho_\infty)\exp\{-|i-j|(\beta - \alpha(\max(i,j)-1))\}.$$

This still has the desirable property of being increasing along subdiagonals. However, the matrix is not positive definite for all values of the three parameters $\alpha, \beta, \rho_\infty$.

Note that the last Schoenmakers-Coffey model above (2003) is a *simplification* of the Rebenato model above (1999). It is also *superior*, in that the matrix is always positive definite.

Reducing the rank

In M dimensions, we have an instantaneous correlation matrix ρ. We may be able to factorise ρ, at least approximately, as

$$\rho = BB^T,$$

with B an $M \times n$ matrix with n much less than M:

$$n << M.$$

This reduces the complexity enormously! So if we can do this, we should. In terms of the relevant SDEs and driving noise, the replacement is

$$dZdZ^T - \rho dt \mapsto BdW(BdW)^T = BB^T dt.$$

Note: Regression.

In regression, we encounter similar 'long thin matrices'. The *design matrix* $A = (a_{ij})$ is $n \times p$, where n is the sample size (as large as possible) and p is the number of parameters (as small as possible) – ao $p << n$. There are *two* associated square matrices:
(i) the *information matrix* $C := A^T A$ ('$(p \times n) \times (n \times p)$', so $p \times p$);
(ii) the *projection matrix* $P := AC^{-1}A^T$ ('$(n \times p) \times (p \times p) \times (p \times n)$', so $n \times n$), also called the *hat matrix* H, as it takes the data y into the fitted values $\hat{y} = Py$.
So one needs to be careful about *matrix size* and *transposes*!. For background, see e.g.
[BF] N. H. BINGHAM and J. M. FRY, *Regression: Linear models in statistics*, Springer, 2010, §3.4, §3.6.

Eigenvalues

As ρ is a real positive-definite (PD) symmetric matrix, it can be written

$$\rho = PHP^T,$$

where P is orthogonal,

$$P^TP = PP^T = I_M,$$

and H is a diagonal matrix of the eigenvalues of ρ, which are positive (see e.g. SMF, III.1). The columns of P are the eigenvectors of ρ, and the same order as their corresponding eigenvalues. Letting Λ be the diagonal matrix whose entries are the square roots of the corresponding entries of H, so $\Lambda\Lambda = H$, and with

$$A := P\Lambda,$$

we can write $\Lambda = H^{\frac{1}{2}}$ (or $\sqrt{H}$ – for background on such matrix square roots and their inverses, see e.g. SMF III.1), and we have

$$AA^T = \rho, \qquad A^TA = H.$$

We can now try to mimic the rank-M decomposition $\rho = AA^T$ by a suitable rank-n decomposition, with B an $M \times n$ matrix as above and BB^T a rank-n correlation matrix, with $n << M$. This will result in a great decrease in computation, and be worth some approximation to achieve.

Eigenvalue-zeroing

To do this, we rank the eigenvalues (which are positive) in *decreasing order*. For a rank-n approximation, we neglect the $M - n$ smallest (in effect, replacing them by zero), retaining only the n largest. This relates to the statistical technique of *principal components analysis (PCA)* (SMF, III), and is called *eigenvalue-zeroing* in mathematical finance. If Λ_n is the diagonal matrix so obtained, write

$$B_n := P\Lambda_n.$$

The resulting candidate correlation matrix is then

$$\overline{\rho}^{(n)} := B_nB_n^T.$$

However, $\overline{\rho}^n$ may not have 1s on its diagonal. So, we interpret it as a *covariance* matrix, and pass to the corresponding *correlation* matrix $\rho^{(n)}$ by

$$\rho_{ij}^{(n)} := \overline{\rho}_{ij}^{(n)} / \sqrt{\overline{\rho}_{ii}^{(n)} \overline{\rho}_{jj}^{(n)}}.$$

According to the *Eckart-Young theorem* (SMF, III), this $\rho^{(n)}$, which is a rank-n approximation to the original rank-M matrix ρ, is the *best* rank-n approximation (in the sense of the Frobenius norm for matrices). See e.g. SMF, III.

One method of implementing this rank reduction is in terms of *Rebonato's angles* (1999). For a detailed study, with numerical examples, see [BM, 6.9.2,3].