Modern Modeling and Pricing of Interest Rates Derivatives Day 2 - Sessions 1 & 2: Beyond SABR & pricing convex products

Vladimir Sankovich [1]
Qinghua Zhu [2]

London, June 6-7, 2018

[1] vsankovich@drwholdings.com
[2] qzhu@drwholdings.com

Summary

- SABR expansion flaws and how to fix them
  - patching the tails
  - ZABR
- Convexity adjustments revised
  - Libor in arrears
  - CMS
- European correlation “exotics”
  - CMS spreads
  - Quantos
  - Options on averages and swaptions on variable notional swaps
  - CMS more in details

The SABR model

In many ways, the SABR model [HKLW02] has been a standard tool for the last few years

$$\begin{cases} dS_t = z_t \, \alpha \, S_t^{\beta} \, dW_t \\ dz_t = z_t \, \nu \, dZ_t \\ z_0 = 1 \end{cases} \tag{1}$$

where W_t and Z_t are Brownian processes in the chosen measure, $\langle dW_t, dZ_t \rangle = \rho dt$, and α, β, ρ, ν are constant parameters.

In [HKLW02] the authors used the singular perturbations technique ($\alpha \to \epsilon\alpha$, $\nu \to \epsilon\nu$, $\epsilon \ll 1$) and a lengthy derivation to come up with a simple approximation for the lognormal and normal volatilities generated by the model.

These expressions were valid for

- short expiries,
- strikes close to ATM,

but, because of their simplicity, they became industry standard, and they started being applied at every regime.

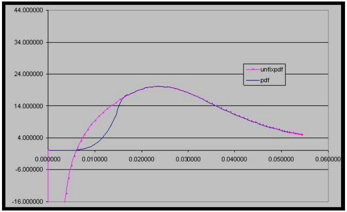
The problem is that for long maturities these expansions break up and don't generate sensible prices – in fact, they can be arbitraged.

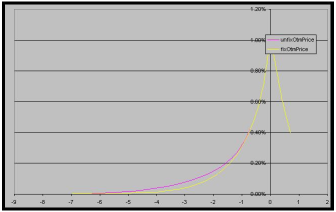
SABR improving on the original expansion, but issues remain

The original expansion got improved in several works:

- improving the leading order for the implied vol expression [BBF04],
- obtaining an expression that is correct first-order in time (homogeneously correct in strike... this is the role of the parallel transport....) [HL05],
- obtaining a second-order in time accurate expression, by using the kernel expansion up to 2nd order [Pau09].

However, the results being still expansions they kept on being prone to breaking down producing arbitrages like these





Another problem not necessarily related to arbitrages is that the right tail of the smile decays really slowly, to the point where the second moment of the distribution is barely limited, which causes all sorts of issues when used as is in CMS pricing by replication.

No-arbitrage conditions

The forward price of a put option expiring at t_s struck at x is

$$\pi(x) = \mathbb{E}[(x - S_{t_s})^+] = \int (x - z)^+ f_{S_{t_s}}(z)\mathrm{d}z\,, \tag{2}$$

where $f_{S_{t_s}}$ is the density for S_{t_s} in the pricing measure. By taking the first and the second derivative of the price wrt the strike:

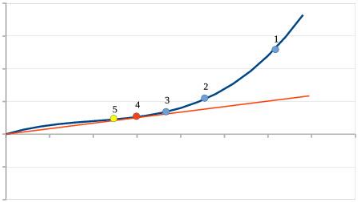
$$\begin{aligned} \pi'(x) &= \int \theta(x - z) f_{S_{t_s}}(z)\mathrm{d}z = F_{S_{t_s}}(x) \\ \pi''(x) &= \int \delta(x - z) f_{S_{t_s}}(z)\mathrm{d}z = f_{S_{t_s}}(x)\,. \end{aligned} \tag{3}$$

Continuum of option prices $\Longleftrightarrow$ terminal distribution and density.

So, π must be positive, monotone increasing, and convex.

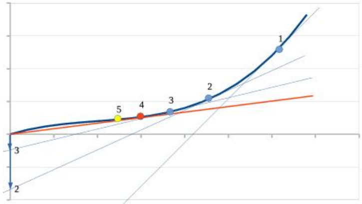
No-arbitrage conditions – where do we start fixing?

Let's say the thick blue line in the sketch below would want to represent the price of a put as a function of the strike.



- It's positive, it's monotonic, yet it isn't convex: the yellow point 5 is an inflection point, and the function is concave between 0 and point 5. Thus, the density for the underlying in that region would be negative, which can't be. This corresponds to an arbitrage – take 3 strikes a, b, $\frac{a+b}{2}$ between 0 and point 5, and you'll be making money. . .
- If we want the price of a put to be 0 for $K = 0$ (more on this later!), then problems actually start earlier than point 5: if we want to touch 0 in a convex way, we have to leave the blue curve at point 4, and start moving along the red tangent line.
- If we're trying to devise an algorithm to fix the arbitrages we have to start operating at least at point 4, well before we enter the region of non-convexity.

No-arbitrage conditions – the Legendre transform



Drawing tangent lines to the price curve, as above, we see that the distance btw the origin and the intercept of the tangent line with the y-axis decreases as we move down the curve, until it becomes 0 for point 4, where we then have no choice other than going straight to 0 following the tangent. Notice how this price behavior requires a finite probability mass to be concentrated at 0.

The distance between the origin and the intercept with the y-axis of a tangent to the curve in the point $(x, \pi(x))$ is related to the Legendre transform of the price, and it corresponds to

$$\pi^L(x) \equiv x\pi'(x) - \pi(x) \quad \Rightarrow \quad \pi^{L\prime}(x) = x\,\pi''(x)\,, \tag{4}$$

thus, π^L will have to be increasing for $x > 0$. If one assumes the density not to have support in the negative region, then $\pi^L(0) = 0$, and being π^L increasing, π^L must also be positive.

A strategy for patching SABR

A possible approach could be to only use SABR for a certain region around the forward, borrowing its dynamical behavior, to then extrapolate it somehow in an arbitrage-free way also adding control to the tail behavior.

For small strikes, given what we've seen so far about π, π', π'', and π^L, a strategy to find the point we should start fixing SABR from could be to look for the zero of π^L by moving in fractions of a Newton step, whilst monitoring the other conditions.

For high strikes, typically there are no arbitrages, and the problem is more that of controlling the speed of decay of the density.

The easiest way to patch SABR is to choose a sensible functional form for the price for small/high strikes, to be glued to the original price curve by imposing continuity conditions up to 2nd order. The requirements for the functional form are

- fast to compute,
- fast to solve for the parameters from the continuity conditions,
- fast to check that the extrapolant obtained is monotonic, convex, and vanishing at zero/infinity,
- for high strikes, have an extra parameter providing control on the speed of decay,
- guarantee that for every sensible boundary condition (e.g. π, π', π'') there is a sensible extrapolant,
- if there is any need to transition to a different functional form, or to activate new parts of it, make sure the fitting procedure guarantees a smooth transition.

Functional forms – small strikes

A decent functional form to extrapolate prices for $0 < x < t$, t being the threshold, could be (see for example [SDK08])

$$g(x) = e^{a+b\,(x-t)+\mu\ \ln(x)}\,. \tag{5}$$

The extrapolant satisfies the required conditions if and only if $\mu > 1$.

In the case of SABR, though, the above would be very deficient, as the model generates prices that require a finite probability mass at zero or thereabouts to be explained. The easy fix is to add a term $z \cdot x$ in $g(x)$. One can see that z must be chosen within $[0, g(t)/t]$, and that the condition for a good extrapolant is still $\mu > 1$.

Having extra freedom with z has a big benefit: if with $z = 0$ we come up with $\mu_{z=0} \leq 1$, we can always find a z such that $\mu_z > 1$.

As always, with great power comes great responsibility, and one has to devise a fitting strategy that removes the freedom in choosing z whilst guaranteeing the transitions smooth between corner cases, if there are any.

Functional forms – large strikes

An example of extrapolant for the call prices for large strikes $x > t$, t the threshold, could be

$$C_\mu(x) = \exp(\mu x + a_\mu + b_\mu/x + c_\mu/x^2)\,, \tag{6}$$

Clearly, one must now require $\mu < 0$. On top of that, one can prove that the extrapolant is a good one if and only if

$$g(z) = 4\,z^6\,c_\mu^2 + 4\,z^5\,c_\mu\,b_\mu + (b_\mu^2 + 6\,c_\mu)\,z^4 + (-4\,\mu\,c_\mu + 2\,b_\mu)\,z^3 - 2\,\mu\,z^2\,b_\mu + \mu^2 \tag{7}$$

has no roots in $[0,\,1/t]$, which one can check relatively quickly with something like the Jenkins-Traub algorithm.

The parameter μ can be used as a handle to control the speed of decay. It can be useful to determine it by requiring the extrapolant to match the original price at some other threshold t', to get a measure of the speed of decay of the original smile, which can then be leveraged to get a faster decay, for example.

Negative rates

The original model assumes positive rates, and so does the expansion as LN vol. However, market prices had started violating these assumptions long ago already:

- we started seeing tradable prices that could only be extrapolated in a convex way by letting the price be non-zero for negative rates,
- the forwards started going lower and lower, with normal volatilities staying the same, to the point that only a fraction of a standard deviation was enough to take you to the negative region,

Even with positive forwards, the expansion started breaking down more and more, to the point where

- prices needed fixing for a wide range of strikes,
- fixed prices were basically dictated by how much probability we were accumulating at zero, and little else, with the consequence of risk drying up.

One could not fix SABR as it was anymore.

Shifting SABR

The easiest way to fix this situation was to "shift" the model, i.e. to assume $S_t + \delta$ to follow the SABR dynamics, rather than S_t, where δ is some given constant shift.

This way the model is capable of generating non-zero prices for options whenever $K > -\delta$.

The shift δ must

- help explaining prices seen for very low / negative strikes,
- cater for possible future moves in the forwards (pretty much same as above),
- be big enough to move the hard boundary far away for its presence not to "stress" the expansion too much (one can use the standard deviation corresponding to the atm normal vol as a guidance).

At the first, whether the observed prices with their large implied shifts were due to people really thinking a 30y swap rate could go that low, or more to people busy bringing their SABR analytics back to a working state was open to debate.

So after timid approaches with small shifts, people started going all the way down to more than 1%, and then further.

Issues shifting SABR

Some limitations with shifting the model are:

- We'd like to express a view of the behavior for low fwds (≈SLN) and high fwds (≈Normal) within the same model. Right now the fwds are low, they have shift and maybe some biggish beta because they want a fatter right tail (they think the rates should eventually move up). If in the near future the fwds move back up again, they'll likely have to reduce the shift and/or beta.
- When changing beta/shift, parameters and risk profile change. Moreover, your option pricing doesn't include these readjustments you already know you'd be doing in case the fwds moved up/down.

Ultimately, shifting the model changes the local vol term of SABR from S_t^β to $(S_t + \delta)^\beta$, just to let S_t go negative – we'd like to set a hard boundary further down, so not to have to move it around in case the fwds went lower, and still retain control on the local vol.

SABR more recent developments

Results about exact pricing under SABR for the $\rho = 0$ case started appearing

- price as a 3D integral of special functions (very slow) [Isl09],
- improving the above, reducing it to a 2D integral of elementary functions [AS12], which can be approximated as a 1D integral [AKS13].

These results can be applied to the more general $\rho \neq 0$ case through the so called proxying techniques, whereby the a zero-correlation model is built that reproduces the original model prices well. The downside of this approach is that it's not blazing fast, and the mapping is only accurate if done in a strike dependent way, at which point there's no guarantee it's arbitrage free.

In [Dou12], a good approximation for the SABR implied probability density is derived, which is guaranteed to be arbitrage-free. The probability mass at zero is accounted for explicitly and separately, which is key to getting to a good result (this is probably the main shortcoming of the original SABR derivation). Unfortunately, there's no way to determine how much probability one should put in there, and the author suggests to use MC, which would not be practical.

In [BT13], an approximation for a SABR model with non-zero correlation, $\beta = 0$, and absorption boundary condition at 0 explicitly handled, is presented and proposed as the basis model to use with the proxying technique. Again no arbitrage-free guarantee.

The ZABR approach

In [AH11], Andreasen and Huge re-propose an argument by Balland [Bal06] and Lewis [Lew07] to find the approximate short-time equivalent local vol for the SABR model: starting with

$$\begin{cases} dS(t) = z(t)\,\sigma(S(t))\,dW(t) \\ dz(t) = z(t)\,\nu\,dZ(t) \\ z_0 = 1 \end{cases}, \tag{8}$$

where $\sigma(\cdot)$ is a **general** local vol function, one looks for the local vol $\vartheta(k)$ such that the pure local vol model

$$dS(t) = \vartheta(S(t))\,dW(t) \tag{9}$$

gives the same prices for European options, in the limit of small expiry.
For SABR one can find

$$\vartheta(k) = J(y)\,\sigma(k)\,, \tag{10}$$

with

$$y = \int_k^{S_0} \frac{1}{\sigma(u)} du\,,$$
$$J(y) = \sqrt{1 + \nu^2 y^2 - 2\,\rho\,\nu\,y}\,. \tag{11}$$

SABR eqv local vol

The short-time eqv local vol is a very strong signature of a model – two different models leading to the same the short-time eqv local vol will give very similar European prices. This can be used to proxy a model with another one, and it's very much what the mapping techniques mentioned before do.

As we have seen for SABR, the Taylor expansion of the short-time eqv LV in the strike identifies the role of the model parameters in determining level, slope, curvature of the smile

$$\sigma(S_0) + \frac{1}{2}\left[\rho\,\nu + \sigma'(s)\right](K - S_0) + \\ + \frac{1}{12\,\sigma(S_0)}\left[(2-3\rho^2)\nu^2 + (2\sigma(S_0)\sigma''(S_0) - \sigma'(S_0)^2)\right](K - S_0)^2 \qquad (12)$$

The diffusion time $x = \int_k^{S_0} \frac{1}{\vartheta(u)} \mathrm{d}u$ is a very interesting quantity, as it maps things back to a diffusion with volatility 1. For SABR one obtains

$$x = \frac{1}{\nu} \ln \frac{J(y) - \rho + \nu\, y}{1 - \rho}\,. \qquad (13)$$

The implied normal vol values obtained as $v = \frac{S_0 - k}{x}$ are very similar to those obtained with the customary formulae in [HKLW02].

Extrapolating the short-time local vol to long expiries

Following [AH11], the idea is to assume the eqv short-time local vol $\vartheta(k)$ to be valid for any time t, and to then plug it into the fwd pricing PDE

$$
\begin{aligned}
c_t(t,k) &= \frac{1}{2}\vartheta(k)^2\, c_{kk}(t,k)\,,\\
c(0,k) &= (s-k)^+\,,
\end{aligned}
\tag{14}
$$

so to obtain prices at whatever maturity.

In [AH11] the approximation is taken even further by replacing (14) with its 1-step finite difference approximation

$$
[1-\frac{1}{2}\,t\,\theta(k)^2\,\partial_k^2]c(t,k) = (s-k)^+\,, \tag{15}
$$

where $\theta(k)$ is a version of $\vartheta(k)$ adjusted in a way to retrieve the right result for normal options.

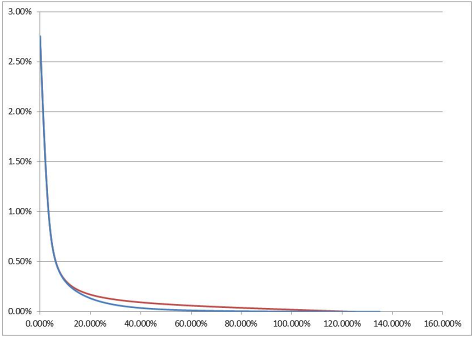
The above "procedure" is guaranteed to return arbitrage free prices, and so is its finite difference approximation. The prices will reflect the dynamics of the model through the short-time eqv LV. In practice, prices from the standard SABR and this method are very much compatible up until SABR starts breaking down badly.

Practical details

- **Sampling:** When discretizing Eq (15), it's worth noting that x is pretty much normally distributed (it is for short maturities), hence something like uniform sampling btw -5 and 5 in x works well. In practice, we'll end up stopping earlier on the low strikes because of absorption, and we'll deliberately stop earlier on high strikes too because otherwise we'd be going too far (for long maturities, $k = 1000$ and beyond... no wonder CMS explodes).
- **Boundary conditions:** We impose the price to match the intrinsic at $k = 0$ (or at the shift), and we also impose the price to be 0 at k_{max}, using absorption – this way we have an idea of how much density we've got up there and we put a stop to crazy values.
- **Interpolating the discrete solution:** Rather than generating prices and then splining them somehow, it's worth plugging the cubic spline expression directly in the ODE (15), using "not-a-knot" conditions. The linear system is only slightly more complicated, there are very minor chances of easily detectable arbitrages when not using enough points, but it's worth doing.
- Organise the calculations to avoid calling expensive functions like sqrt and exp.
- What local vol?

Tail control in ZABR

For the local vol, one can for example use 3-regimes of *beta* – low, med, high fwds. In particular, modulating the local vol for high strikes controls the speed of decay of the pdf (i.e. CMS prices).
For example, this is what you get ...



Summary

- SABR expansion flaws and how to fix them
  - patching the tails
  - ZABR
- Convexity adjustments revised
  - Libor in arrears
  - CMS
- European correlation “exotics”
  - CMS spreads
  - Quantos
  - Options on averages and swaptions on variable notional swaps
  - CMS more in details

Libor in arrears

In the previous sessions, we've seen how nowadays one strips (at least) a curve of discount factors $D(t, T_2)$, and a curve of the forwards for a given tenor basis τ

$$F(t, T_1, T_2) = \mathbb{E}_t^{T_{2c}}\left[L(t_s, T_1, T_2)\right] . \tag{16}$$

We've also seen that the price at t of the cap(floor)lets on the Libor setting at t_s for the period $[T_1, T_2]$, and paying at T_2, corresponds to

$$Cplt(t, K; T_1, T_2) = D(t, T_2)\mathbb{E}_t^{T_{2c}}\left[(L(t_s, T_1, T_2) - K)^+\right] , \tag{17}$$

so these instruments can be used to determine the density of that Libor rate in the T_2-forward fully collateralised measure.

All the above refers to instruments paying at T_2 – with the density at hand, one can price whatever payout on $L(t_s, T_1, T_2)$ paying at T_2 (fully collateralised).

However, what if we simply wanted to pay $L(t_s, T_1, T_2)$ at the beginning of the period T_1, rather than at the end T_2? Risk neutral pricing would simply give for the flow's value at t

$$D(t, T_1)\mathbb{E}_t^{T_{1c}}\left[L(t_s, T_1, T_2)\right] , \tag{18}$$

but now we have to find the Libor's expectation under the new measure T_{1c}-fwd.

Libor in arrears II

The product could be expressed in the T_2-fwd measure by observing that at t_s the flow has value $L(t_s, T_1, T_2) \cdot D(t_s, T_1)$ so that by risk neutral pricing, the flow's value at t can be written as

$$D(t, T_2)\mathbb{E}_t^{T_{2c}}\left[\frac{L(t_s, T_1, T_2) \cdot D(t_s, T_1)}{D(t_s, T_2)}\right], \tag{19}$$

The two equations above imply

$$L^{arrears}(t, T_1, T_2) \equiv \mathbb{E}_t^{T_1}[L(t_s, T_1, T_2)] = \mathbb{E}_t^{T_{2c}}\left[L(t_s, T_1, T_2) \cdot \frac{D(t, T_2)\, D(t_s, T_1)}{D(t, T_1)\, D(t_s, T_2)}\right]. \tag{20}$$

The term $\frac{D(t,T_2)\, D(t_s,T_1)}{D(t,T_1)\, D(t_s,T_2)}$ is the so called Radon-Nykodym derivative for changing the measure from T_2 to T_1, and the above is a particular case of a general rule for moving from the measure associated to a numeraire N_t to that associated to a numeraire M_t:

$$\forall V,\ \mathbb{E}_t^N\left[\frac{V_s}{N_s}\right] = \frac{V_t}{N_t} = \frac{M_t}{N_t}\frac{V_t}{M_t} = \frac{M_t}{N_t}\mathbb{E}_t^M\left[\frac{V_s}{M_s}\right] \Longrightarrow \mathbb{E}_t^N[U_s] = \mathbb{E}_t^M\left[U_s\frac{M_t}{N_t}\frac{N_s}{M_s}\right]. \tag{21}$$

From the above one can see how a Radon-Nykodym derivative is positive, and its expectation is 1, by all means it changes the probability density to the new measure.

Libor in arrears III

Assuming the Libor-OIS basis to be non-stochastic

$$
\begin{aligned}
b_{t_s} &= L(t_s, T_1, T_2) - \frac{1}{\tau}\left[\frac{D(t_s, T_1)}{D(t_s, T_2)} - 1\right] = b_t = \mathbb{E}_t^{T_{2c}}\left[b_{t_s}\right] = \\
&= \mathbb{E}_t^{T_{2c}}\left[L(t_s, T_1, T_2)\right] - \frac{1}{\tau}\left[\frac{D(t, T_1)}{D(t, T_2)} - 1\right], \qquad (22)
\end{aligned}
$$

and the Libor in arrears *convexity adjustment* becomes

$$
\begin{aligned}
& L^{arrears}(t, T_1, T_2) - \mathbb{E}_t^{T_{2c}}\left[L(t_s, T_1, T_2)\right] = \\
&= \mathbb{E}_t^{T_{2c}}\left[L(t_s, T_1, T_2) \cdot \left(\frac{D(t, T_2)\, D(t_s, T_1)}{D(t, T_1)\, D(t_s, T_2)} - 1\right)\right] = \\
&= \tau\, \frac{D(t, T_2)}{D(t, T_1)}\, \mathbb{E}_t^{T_{2c}}\left[L(t_s, T_1, T_2) \cdot \left(L(t_s, T_1, T_2) - \mathbb{E}_t^{T_{2c}}[L(t_s, T_1, T_2)]\right)\right] = \\
&= \tau\, \frac{D(t, T_2)}{D(t, T_1)}\, \mathrm{Var}_t^{T_{2c}}[L(t_s, T_1, T_2)]\,. \qquad (23)
\end{aligned}
$$

Libor in arrears – conclusions

Expressing the RN derivative as a function of Libor itself we concluded that

$$L^{arrears}(t, T_1, T_2) = \mathbb{E}_t^{T_{2c}}\left[L(t_s, T_1, T_2)\right] + \tau \frac{\mathrm{Var}_t^{T_{2c}}[L(t_s, T_1, T_2)]}{1 + \tau \cdot (\mathbb{E}_t^{T_{2c}}[L(t_s, T_1, T_2)] - b_t)} . \quad (24)$$

- The above can be explicitly calculated from the terminal density for Libor obtained from the associated cap/floorlets.
- Given that the convexity correction depends on the 2nd moment of the distribution, it's all about the behavior of the model for high strikes, which needs to be kept in check, to make sure the size of the adjustment is sensible.
- Assuming a flat normal volatility across expiries, the size of the convexity adjustment will grow linearly with the fixing date of libor. Therefore, the adjustment will be more important the further out we are from value date. For these expiries the assumption of non-stochastic Libor-OIS basis is not that bad.
- As mentioned, the density of Libor in the T_1-fwd measure is the product of the RN derivative - which is a function of Libor - and the density of Libor in the T_2-fwd measure. With this density at hand, we can price any payoff on Libor being paid at T_1.

Swap rate density – from the annuity to the T-fwd measure

Considerations analogous to the ones we've seen for Libor rates also apply to swap rates. We've seen how

$$Payer(t, K) = A(t)\,\mathbb{E}_t^A\left[\frac{(S(t_s) - K)^+ A(t_s)}{A(t_s)}\right] = A(t)\,\mathbb{E}_t^A[(S(t_s) - K)^+]\,, \qquad (25)$$

so, from a continuum of swaptions' prices, typically coming from a model fit to market prices, we can infer the density of the underlying swap rate in the annuity measure.

In some cases, one may want to see the the value of a swap rate materialise right away, at a given time T, rather than having it paid out as an annuity over time – this is what a CMS flow boils down to: the value of a swap rate paid out at a given time.

For such a flow, risk-neutral valuation gives the value at t

$$V(t) = D(t, T)\,\mathbb{E}_t^{T_c}\left[S(t_s)\right]\,, \qquad (26)$$

and once again we have an example of a product whose pricing depends on a change of measure, in this case from that associated to the annuity $A(t)$ to that associated to T zc-bond $D(t, T)$.

CMS

Valuing in the annuity measure gives the value at t

$$V(t) = A(t)\,\mathbb{E}_t^A\left[\frac{S(t_s)D(t_s,T)}{A(t_s)}\right]. \tag{27}$$

The forward value of the flow is the so called CMS rate, and it reads

$$\text{CMS}(t) = \mathbb{E}_t^{T_c}[S(t_s)] = \frac{V(t)}{D(t,T)} = \mathbb{E}_t^A\left[S(t_s)\,\frac{A(t)}{D(t,T)}\,\frac{D(t_s,T)}{A(t_s)}\right]. \tag{28}$$

The term $\frac{A(t)}{D(t,T)}\frac{D(t_s,T)}{A(t_s)}$ is yet another RN derivative, in this case for changing the measure from the annuity measure to the T-forward. The difference between this expectation in the T-forward measure and that in the A-measure (i.e. $S(t)$, given that S is martingale in that measure), is called CMS convexity adjustment.

Many different products can be defined whose pricing ultimately depends on knowing the density of the swap rate in this T-fwd measure, like CMS cap/floors/spreads.

Valuing CMS

When valuing a payoff g on the swap rate S_{t_s}, paying at T, we end up calculating

$$\mathbb{E}_t^A \left[g(S(t_s)) \frac{A(t)}{D(t,T)} \frac{D(t_s,T)}{A(t_s)} \right] = \mathbb{E}_t^A \left[g(S(t_s)) \, \mathbb{E}^A \left[\frac{A(t)}{D(t,T)} \frac{D(t_s,T)}{A(t_s)} | S_{t_s} \right] \right] . \quad (29)$$

The S_{t_s}-conditional expectation of the RN derivative in the above expression can be approximated assuming a non-stochastic Libor-OIS basis, assuming all the accrual fractions in the annuity to be the same, and adopting a constant yield to maturity approximation for discount factors:

$$\begin{aligned} \mu(S_{t_s}) = \frac{A(t)}{D(t,T)} \mathbb{E}^A \left[\frac{D(t_s,T)}{A(t_s)} | S_{t_s} \right] &\approx \eta \cdot \frac{(1+\tilde{S}/m)^{m\cdot(T-T_0)}}{\sum_{i=1}^{n} \frac{(1/m)}{(1+\tilde{S}/m)^i}} \\ &= \eta \cdot (1+\tilde{S}/m)^{m\cdot(T-T_0)} \frac{\tilde{S}}{1-\left(\frac{1}{1+\frac{\tilde{S}}{m}}\right)^n} , \end{aligned} \quad (30)$$

where $\tilde{S} = S_{t_s} - b_t$, m denoted the fixed leg frequency, n its number of coupons, and η is chosen to make sure $\mu(S)$ is correctly normalised, i.e. it integrates to 1 against the implied swap rate density in the annuity measure.

Valuing CMS

The product of $\mu(S_{t_s})$ with the density of the swap rate in the annuity measure gives the density of the swap rate in the T-fwd measure, which we can use to value any payoff on the swap rate being paid at T.

$$D(t,T)\mathbb{E}_t^T\left[g(S_{t_s})\right] = D(t,T)\mathbb{E}_t^A\left[g(S_{t_s})\mu(S_{t_s})\right] . \tag{31}$$

As for Libor in arrears, the change of measure brings in non-linearities, which bring in the second moment of the density into play, and hence a high sensitivity on how one extrapolates its smiles.

Summary

- SABR expansion flaws and how to fix them
  - patching the tails
  - ZABR
- Convexity adjustments revised
  - Libor in arrears
  - CMS
- European correlation “exotics”
  - CMS spreads
  - Quantos
  - Options on averages and swaptions on variable notional swaps
  - CMS more in details

CMS spreads

So far, we've only considered options written on a single rate. However, people may want to take exposure on the relative movements of the curve, like steepening or flattening. This translates on spreads between swap rate 20Y and 10Y. This is what CMS spread options are traded for – they pay at some time T the payoff $(S_2 - S_1 - K)^+$, where S_1 and S_2 are two swap rates of different tenors, setting at the same time t_s.

As we've seen in the previous slides, the terminal densities of two swap rates can be obtained under the same T-forward measure, by measure changing their respective terminal densities extracted from their swaption's prices. These densities can be the coupled with a copula function into a joint density, which can then be used to price whatever payoff on the two rates by integrating.

A copula is the mathematical object that describes how marginal distributions are glued together into a joint distribution – it describes the correlation structure of the joint distribution.

Sklar theorem and copulas

Sklar theorem (1959): for any given joint distribution function $F_{X_1 \dots X_n}$, with marginals F_{X_i} there exists a copula function $C(u_1, \dots, u_n) : [0,1]^n \rightarrow [0,1]$ such that

$$F_{X_1 \dots X_n}(x_1, \dots, x_n) = C(F_{X_1}(x_1), \dots, F_{X_n}(x_n)) \;, \tag{32}$$

where C is unique in the case of continuous marginals. Differentiating wrt to all the variables we find the joint density

$$f_{X_1 \dots X_n}(x_1, \dots, x_n) = c(F_{X_1}(x_1), \dots, F_{X_n}(x_n)) \cdot f_{X_1}(x_1) \cdots f_{X_n}(x_n) \;, \tag{33}$$

where the *copula density* $c(u_1, \dots, u_n)$ has been defined as

$$c(u_1, \dots, u_n) = \frac{\partial^n}{\partial u_1 \dots \partial u_n} C(u_1, \dots, u_n) \tag{34}$$

Given some marginals, we can glue them together into a joint density using a copula function.

Gaussian Copula

The Gaussian copula is the copula function of joint Gaussian variables, it describes their correlation structure, and as such it is parameterised by their "linear" (or Pearson's) correlation $\rho_{ij} = \langle \frac{(x_i - \bar{x}_i)}{\sigma_i} \frac{(x_j - \bar{x}_j)}{\sigma_j} \rangle$.

The very same copula can be used to stitch together non-Gaussian marginals, by borrowing the correlation structure typical of joint Gaussian variates.

For two random variables X and Y, the Gaussian copula density takes the form

$$c_\rho(x, y) = \frac{\phi_\rho(u, v)}{\phi(u)\phi(v)}\,, \quad \text{with } u = \Phi^{-1}(F_X(x))\,,\ v = \Phi^{-1}(F_Y(y))\,, \tag{35}$$

where ϕ and ϕ_ρ denote the univariate and the bivariate Gaussian probability densities

$$\phi(u) = \frac{1}{\sqrt{2\pi}} e^{-u^2/2}\,, \qquad \phi_\rho(u, v) = \frac{1}{2\pi\sqrt{1-\rho^2}} e^{-\frac{u^2+v^2-2\rho uv}{2(1-\rho^2)}}\,, \tag{36}$$

Φ is the Gaussian cumulative distribution, and F_X, F_Y are the cumulative distributions of X and Y, respectively.

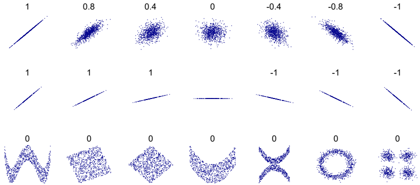
A joint density for X and Y, having marginals F_X and F_X, can then be defined by

$$f^\rho_{XY}(x, y) = f_X(x) f_Y(y) c_\rho(x, y)\,. \tag{37}$$

Linear correlation and dependence

Linear correlation is a measure of linear dependence. It's often used as a measure of dependence, but this is only sensible for elliptical distributions (like multivariate Gaussian), but often random variables are not distributed like that.

Even though linear correlation is invariant for linear and strictly increasing transformations, a simple non-linear strictly monotonic transformation of the variates can make the linear correlation lose any meaning as a measure of dependence, so it's not robust enough in that sense.



Copulas and dependence

By their nature, copulas are invariant for monotonic transformations of the variates, and thus are the natural ingredient for defining robust measures of dependence (better say concordance).

One example is Kendall's tau

$$\tau = \mathbb{P}\{(X - X')(Y - Y') > 0\} - \mathbb{P}\{(X - X')(Y - Y') < 0\}\,, \qquad (38)$$

for (X, Y) and (X', Y') i.i.d. according to the bivariate distribution of interest. If C is the copula of such distribution, one could see that

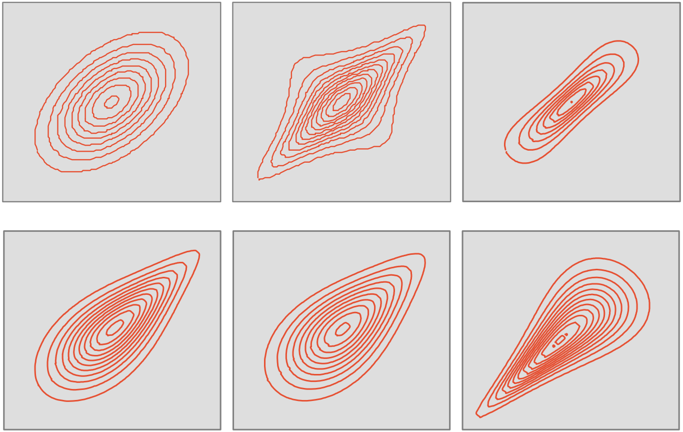
$$\tau = 4\mathbb{E}[C(U, V)] - 1\,, \quad U, V \sim \mathcal{U}_{[0,1]}\,. \qquad (39)$$

Another example is Spearmans's rho

$$\rho = 3\mathbb{P}\{(X - X')(Y - Y') > 0\} - \mathbb{P}\{(X - X')(Y - Y') < 0\} =$$
$$= 12 \int\int_{[0,1]^2} u\, v\, \mathrm{d}C(u, v) - 3\,. \qquad (40)$$

Both the above are concordance measure.

Copulas examples



Valuing CMS spreads – let's integrate

Once the terminal densities have been obtained, and a copula function has been chosen, pricing, for example, a CMS spread option resorts to integrating

$$\int \int f_{S_1}(s_1) \cdot f_{S_2}(s_2) \cdot c(s_1, s_2) \cdot (s_2 - s_1 - K)^+ \cdot \mathrm{d}s_1 \cdot \mathrm{d}s_2 \tag{41}$$

- The nice thing about copulas is that they guarantee consistency with the marginals and coherence across the board. However, the latter is not the case anymore if we start marking a correlation smile (for example, with Gaussian copula).
- One trick to extend copulas is to do power-copulas

$$C(u_1, u_2) \to u_1^{1-\theta_1} u_2^{1-\theta_2} C(u_1^{\theta_1}, u_2^{\theta_2}) \tag{42}$$

- Power-t copulas seem to be able to be used to mark the correlation smile.
- Copulas have an upper and lower bound, $C_-(u_1, u_2) = \max(0, u_1 + u_2 - 1)$ and $C_+(u_1, u_2) = \min(u_1, u_2)$, which can be used to determine upper and lower bounds on prices.
- The integral above can be re-jiggled in different ways that may be more amenable for numerical integration.
- With SABR marginals, some care is needed dealing with the finite probability accumulated at zero (or wherever that is after the shift) – in fact, digi-spread options are very revealing about how one handles that sort of thing.

Gaussian Copula – Quanto

As mentioned in a previous session, paying a flow in a currency different from the natural one (what is natural?) brings in a quanto convexity adjustment, which as always comes from a change of measure. If we have a payoff $g(S)$ paid out at T in some foreign currency, the forward value of the payment will be

$$\mathbb{E}_t^{T,f}[g(S_{t_s})] = \mathbb{E}_t^{T,d}[g(S_{t_s}) \cdot X_T/X_t] = \frac{1}{\mathbb{E}[X]} \int\int g(s)\, x\, f_{SX}(s,x)\, \mathrm{d}s\, \mathrm{d}x \ , \qquad (43)$$

where X_t is the FX fwd rate, which is a T,d martingale, and $f_{SX}(s,x)$ the joint T-fwd density for S_{t_s} and X_T.

Even in this case one could tackle the problem with copulas, by decomposing $f_{SX}(s,x) = f_S(s)\, f_X(x)\, c(s,x)$. The easiest thing one could possibly do is to assume X to be lognormal, and the copula to be Gaussian

$$F_X(x) = \Phi\left(\frac{\ln x - \mu}{\sigma}\right) \ , \quad \text{and } f_X(x) = \frac{1}{\sigma x}\,\phi\left(\frac{\ln x - \mu}{\sigma}\right) \ . \qquad (44)$$

With a bit of algebra one can see that the quantoed probability density for S turns out to be a neat translation of the Gaussian-mapped quantiles of the original density

$$F_{S^q}(s) = \Phi\left(\Phi^{-1}(F_S(s)) - \sigma\rho\right) \ . \qquad (45)$$

Swaptions with variable notional

One can decompose a swap with varying notional as

$$\tilde{A}_t\,(\tilde{S}_t - K) = \sum_i A_t^{(i)}\,(S_t^{(i)} - K)\,, \tag{46}$$

so that, by assuming the variability of the ratio between any two annuities to be small enough for the ratio to be frozen at its initial value, one has

$$\tilde{S}_t - K \approx \sum_i \frac{A_0^{(i)}}{\tilde{A}_0}\,(S_t^{(i)} - K) = \sum_i w_i\,S_t^{(i)} - K\,, \quad \text{with } w_i = \frac{A_0^{(i)}}{\tilde{A}_0}\,, \tag{47}$$

where some of the weights may be negative, but their sum will be 1.

Under this approximation, all the A_i-measures are equivalent, and the present value of a swaption on this swap becomes

$$V = \tilde{A}_0 \cdot \mathbb{E}^{\tilde{A}}\left[(\tilde{S}_t - K)^+\right] \approx \tilde{A}_0 \cdot \mathbb{E}^{\tilde{A}}\left[\left(\sum_i w_i\,S_t^{(i)} - K\right)^+\right]. \tag{48}$$

Swaptions with variable notional - basket of normals

One way to come up with a solution is to try to capture the volatility of the constituent swaps at the so-called "exercise boundary"

$$K^{(i)} = \mathbb{E}[S^{(i)} | \tilde{S} = K] \, . \tag{49}$$

Starting from an the assumption of a basket of normals with vols σ'_i

$$K^{(i)} = \mathbb{E}[S^{(i)} | \tilde{S} = K] = S_0^{(i)} + \frac{\langle S^{(i)} \, , \tilde{S} \rangle}{\langle \tilde{S}, \, \tilde{S} \rangle} \, (K - \tilde{S}_0) = S_0^{(i)} + \frac{\sigma'_i \, \sum_j \rho_{ij} \, w_j \, \sigma'}{\sum_{ij} w_i \, w_j \, \sigma'_i \, \sigma'_j \, \rho_{ij}} (K - \tilde{S}_0) \, , \tag{50}$$

where the σ'_i are some sensible levels for the volatility. The describes an iteration $\sigma'_i \longrightarrow \sigma_i = \sigma_i(K^{(i)})$ which may converge to a point which would be some sort of exercise boundary.

A better way to approach the problem would be trying to come up with an effective SDE for $\tilde{S}_t$ given those for S^i_t.

SABR averaging

$$\begin{cases} dS_t^i = z_t^i\, g^i(S_t^i)\, dW_t^i \\ dz_t^i = z_t^i\, \nu^i\, dZ_t^i \\ z_0^i = 1 \end{cases} \tag{51}$$

where $\langle dW_t^i, dW_t^j \rangle = \gamma_{ij} dt$, $\langle dZ_t^i, dZ_t^j \rangle = \xi_{ij} dt$, and $\langle dZ_t^i, dW_t^j \rangle = \rho_{ij} dt$, we want to find a SABR-like equation for their weighted average

$$S_t = \sum_i w^i\, S_t^i \ . \tag{52}$$

Let's define σ_t and W_t such that

$$dS_t = \sum_i w^i\, dS_t^i = \sum_i w^i\, z_t^i\, g^i(S_t^i)\, dW_t^i \doteq \sigma_t\, dW_t \ . \tag{53}$$

i. e. , by defining $f_t^i = w^i\, g^i(S_t^i)$,

$$\sigma_t^2 = \sum_{ij} f_t^i\, f_t^j\, \gamma_{ij}\, z_t^i\, z_t^j \ , \qquad\qquad dW_t = \frac{1}{\sigma_t} \sum_i f_t^i\, z_t^i\, dW_t^i \ . \tag{54}$$

We want to project the above expression for dS_t onto a SABR form

$$dS_t = z_t\, g(S_t)\, dW_t \ , \tag{55}$$

where z_t doesn't depend on the S_t^i's, $z_0 = 1$, and z_t will be projected onto a lognormal diffusion.

SABR averaging II

The problem can be approached by using Markovian projection (for this specific example, see [KW10]):

- Set z_t to the value of σ_t with all the S_t^i's frozen at $t = 0$, and normalised so that $z_0 = 1$
$$z_t^2 = \frac{1}{\sigma_0^2} \sum_{ij} f_0^i \, f_0^j \, \gamma_{ij} \, z_t^i \, z_t^j \, , \tag{56}$$
- Find the expression of $g(S_t)$ by imposing the marginals corresponding to the two equations for $\mathrm{d}S_t$ above to be the same. By Győngy's lemma, this will happen if
$$\mathbb{E}[\sigma_t^2 \,|\, S_t] = \mathbb{E}[z_t^2 \, g^2(S_t) \,|\, S_t] = \mathbb{E}[z_t^2 \,|\, S_t] \, g^2(S_t) \, . \tag{57}$$
- The two conditional expectations above can be approximated with their short-time expressions: all the stochastic variables are approximately normal, and functions on them can be linearised around the initial conditions (and will then be normal themselves)
- Remember that for two jointly normal variates a and b with correlation ρ the following holds (we've used it already for the basket of normal case)
$$\mathbb{E}[a - \bar{a} \,|\, b] = \frac{\langle a, b \rangle}{\langle b, b \rangle} (b - \bar{b}) \, . \tag{58}$$
- Project z_t onto a lognormal.

SABR averaging III

The result is

$$g(S_t) \simeq \sigma_0 + \frac{1}{\sigma_0} \sum_{ij} f_0^j \, \gamma_{ij} \, f_0^{i\prime} \, \mathbb{E}[S_t^i - S_0^i | \, S_t] \tag{59}$$

$$\simeq \sigma_0 + (S_t - S_0) \frac{1}{\sigma_0^3} \sum_{ij} f_0^i \, f_0^j \, \gamma_{ij} \, f_0^{i\prime} \, \frac{1}{w^i} \sum_h f_0^h \, \gamma_{ih} \; , \tag{60}$$

$$\nu^2 = \frac{1}{\sigma_0^4} \sum_{ijhk} f_0^i \, f_0^j \, \gamma_{ij} \, \nu^i \, \xi_{ih} \, f_0^h \, f_0^k \, \gamma_{hk} \, \nu^h \; , \tag{61}$$

$$\rho = \frac{\langle \mathrm{d}Z_t, \mathrm{d}W_t \rangle}{\mathrm{d}t} \simeq \frac{1}{\nu \sigma_0^2} \sum_{ij} f_0^i \, f_0^j \, \gamma_{ij} \, \nu^i \, \frac{1}{\sigma_0} \sum_h f_0^h \, \rho_{ih} \; . \tag{62}$$

So for the usual SABR we have $g^i(x) = \alpha_i \, x^{\beta_i}$, and similarly we'll assume the local vol functional for the average to be $g(x) = \alpha \, x^\beta$. Plugging these expression in Eq. (60), and matching up the zero and first order, one gets

$$\beta = \frac{S_0}{\sigma_0^4} \sum_{ij} f_0^i \, f_0^j \, \gamma_{ij} \, f_0^{i\prime} \, \frac{1}{w^i} \sum_h f_0^h \, \gamma_{ih} \; , \qquad \alpha = \sigma_0 \, S_0^{-\beta} \; . \tag{63}$$

SABR averaging – what correlation completion?

Following [JK09], from a set of standard normal variables x_i, y_i, driving the underlyings and the stoch vols, respectively, with known correlations defined as

$$\langle x_i, x_j \rangle = \rho_{ij} \; , \langle x_i, y_i \rangle = \gamma_i \; , \tag{64}$$

we have to come up with a choice for $\langle x_i, y_j \rangle$ and $\langle y_i, y_j \rangle$.
Take the component of the stoch vol driver orthogonal to the underlying

$$z_i = \frac{y_i - \gamma_i \cdot x_i}{\sqrt{1 - \gamma_i^2}} \; , \qquad \qquad \langle x_i, z_i \rangle = 0 \, . \tag{65}$$

We will assume the stoch vol drivers z_i's to have no dependency to any of the x_i's,

$$\langle x_i, z_j \rangle = 0 \quad \Longrightarrow \quad \langle x_i, y_j \rangle = \gamma_j \cdot \rho_{ij} \, . \tag{66}$$

We still have freedom to choose $\langle z_i, z_j \rangle$ as a correlation matrix. Given there's no real reason here why the stoch vols should be correlated in a very different way from the underlyings, we expect their correlation levels to be rather similar, i.e. $\langle z_i, z_j \rangle \approx \rho_{ij}$, which leads to the completion

$$\langle y_i, y_j \rangle = \left(\sqrt{(1 - \gamma_i^2)(1 - \gamma_j^2)} + \gamma_i \cdot \gamma_j \right) \cdot \rho_{ij} \; . \tag{67}$$

Valuing CMS more in detail

For analysing the calculation of CMS one can also follow [CP12]

$$\mu(S_{t_s}) = \frac{A(t)}{D(t,T)}\,\mathbb{E}^A\left[\frac{D(t_s,T)}{A(t_s)}|S_{t_s}\right] = \frac{A(t)}{D(t,T)}\,\frac{\mathbb{E}^{t_s}\left[D(t_s,T)|S_{t_s}\right]}{\mathbb{E}^{t_s}\left[A(t_s)|S_{t_s}\right]}\,. \tag{68}$$

where for the last step we have used the change of measure under conditional expectation

$$\mathbb{E}^A[X|\mathcal{G}] = \frac{\mathbb{E}^B[X\,Z|\mathcal{G}]}{\mathbb{E}^B[Z|\mathcal{G}]}\,,\quad Z = \mathrm{d}A/\mathrm{d}B \tag{69}$$

with $A \to A(t)$, $B \to D(t,t_s)$, $X = \frac{D(t_s,T)}{A(t_s)}$, $Z = A(t_s)/D(t_s,t_s)$, $D(t_s,t_s) = 1$.

So in the end one has to find some sensible approximation for terms like

$$\tilde{D}(S_{t_s};t_s,T) = \mathbb{E}^{t_s}\left[D(t_s,T)|S_{t_s}\right]\,, \tag{70}$$

and then again one can resort to the simple like a constant yield to maturity approximation and to assuming a non-stochastic Libor-OIS basis, i.e.

$$\tilde{D}(S_{t_s};t_s,T) \approx \frac{1}{(1+\frac{S_{t_s}-b_t}{m})^{m\cdot(T-t_s)}}\,, \tag{71}$$

which, by assuming all the accrual fractions in the annuity to be the same, will give Eq. (30) again.

Valuing CMS more in detail – where's my risk?

Although very convenient, the approximation for $\mu(S_{t_s})$ in Eq. (30) has a shortcoming: it is not representative of the fact that a CMS should show some dependency on the covariance of the swaps with shorter tenors, which would come in through the RN derivative, and which would have to show up somehow in the expression of $\mu(S_{t_s})$.

The issue has been analysed in [CP12] starting from

$$A_i(t_s) = A_{i-1}(t_s) + \tau_i D(t_s, T_i) = A_{i-1}(t_s) + \tau_i(1 - S_i(t_s)A_i(t_s)) =$$
$$= \frac{\tau_i + A_{i-1}(t_s)}{1 + \tau_i\, S_i(t_s)} = \sum_{l=1}^{i} \tau_l \prod_{m=l}^{i} \frac{1}{1 + \tau_m\, S_m(t_s)}\,, \qquad (72)$$

and hence obtaining an expression for discount factors as functions of the swap rates

$$D(t_s, T_i) = 1 - S_i(t_s)A_i(t_s) = 1 - S_i(t_s) \sum_{l=1}^{i} \tau_l \prod_{m=l}^{i} \frac{1}{1 + \tau_m\, S_m(t_s)}\,, \qquad (73)$$

which is then used in conjunction with a stylised model for the joint density of the S_i's, representing their correlations ρ_{ij} and volatilities σ_i, to come up with an expression for $\tilde{D}(S_{t_s}; t_s, T)$, and ultimately with a $\mu(S_{t_s}; \sigma_i, \rho_{ij})$ (basis omitted for clarity).

CMS, decorrelation, and cash settled swaptions

Having a bit more coherence btw the view on corrs and vols and the measure changed density for the swap rate might be a good thing when considering products that do depend on such correlations, like CMS spreads.

Cash-settled swaptions are a CMS-like product whose payout containing $A(S) = \frac{1}{S}\left(1 - \left(\frac{m}{m+S}\right)^n\right)$ is designed to counteract convexity by simplifying away with the change of measure $\mu(S_{t_s})$. The more the two things are decorrelated, the less the convexity will be reduced – swap rates decorrelation and stochastic basis play a major role in determining this (participants in the EUR market started quoting zero-wide collars and cash vs physical swaptions because of this).

- Jesper Andreasen and Brian Huge, *ZABR – expansions for the masses*, http://ssrn.com/abstract=1980726 (2011).
- A. Antonov, M. Konikov, and M. Spector, *SABR spreads its wings*, Risk (2013), 58–63.
- A. Antonov and M. Spector, *Advanced analytics for the SABR model*, http://ssrn.com/abstract=2026350 (2012).
- Phillipe Balland, *Forward smile*, ICBI Global Derivatives Presentation (2006).
- Henri Berestycki, Jérôme Busca, and Igor Florent, *Computing the implied volatility in stochastic volatility models*, Communications on Pure and Applied Mathematics **10** (2004), no. 57, 1352–1373.
- Philippe Balland and Quan Tran, *SABR goes normal*, Risk (2013), 76–81.
- Simon Cedervall and Vladimir Piterbarg, *Cms: covering all bases*, Risk (2012).
- Paul Doust, *No-arbitrage SABR*, The Journal of Computational Finance **15** (2012), no. 3, 3–31.
- Patrick S. Hagan, Deep Kumar, Andrew S. Lesniewski, and Diana E. Woodward, *Managing smile risk*, Wilmott Magazine (2002), 84–108.
- Pierre Henry-Labordére, *A general asymptotic implied volatility for stochastic volatility models*, http://arxiv.org/abs/cond-mat/0504317 (2005).

- Othmane Islah, *Solving SABR in exact form and unifying it with libor market model*, http://ssrn.com/abstract=1489428 (2009).
- Peter Jäckel and Christian Kahl, *Positive semi-definite correlation matrix completion*.
- Jörg Kienitz and Manuel Wittke, *Option valuation in multivariate sabr models*, Tech. Report 272, Quantitative Finance Research Centre, University of Technology, Sidney, February 2010.
- Alan L. Lewis, *Geometries and smile asymptotics for a class of stochastic volatility models*, UC Santa Barbara Presentation (2007).
- Louis Paulot, *Asymptotic implied volatility at the second order with application to the SABR model*, http://ssrn.com/abstract=1413649 (2009).
- Benaim Shalom, Matthew Dodgson, and Dherminder Kainth, *An arbitrage-free method for smile extrapolation*, Tech. report, Royal Bank of Scotland, 2008.