1 Introduction

The SABR model (Hagan et al. 2002) has emerged in the last years as a reference stochastic volatility framework for modelling the swaption implied volatility at different strikes. However, many implications of the model behaviour in a market context have so far been overlooked.

First of all we test empirically the capability of the model to be consistent with current swaption market quotations. In particular we perform testing on different possible implementations of the model, in order to assess how model behaviour changes when practical considerations on implementation feasibility and parameter stability lead to specific constraints, for example on the number of independent stochastic volatility factors considered. We also compare SABR behaviour with the one of traditional stochastic volatility models, and assess some consequences on the pricing of non-vanilla structures.

Secondly, SABR has been proposed as a model for a generic asset forward price under its natural probability measure. This leaves aside issues about modelling jointly different assets, which is instead relevant for application of the model to term-structure derivatives. In particular it does not consider how the model dynamics changes when the characteristics of a financial product force us to model the evolution of a plurality of interest rates under a common measure, which cannot be the natural measure for all the rates considered. Neglecting these considerations would lead to a model allowing arbitrage. In this paper we compute and analyze these dynamics when SABR distributional assumptions are applied to a Libor market model (BGM) framework, and we assess their practical relevance and impact. This allows us also to address the issue of the relationships between forward Libor and swap rates when they are
jointly modelled according to SABR dynamics, detecting approximations for relating the two markets analogous to those typically used for more standard dynamics.

Thirdly, the SABR model has been put forward based on considerations of the desired behaviour of a smile model in hedging, however the behaviour of the SABR model in hedging for different configurations of its parameters has not been addressed. In this work we perform a simple analysis of the behaviour of this model compared to local volatility models, and in particular we extend the analysis taking explicitly into account the possible presence of correlation between stochastic volatility and underlying asset. Results are not in line with the general market wisdom.

## 2 Empirical tests of SABR

The standard implementation of SABR as an approximated Swap Model assumes that each swap rate is associated to a different stochastic volatility factor. In this case calibration of a Swap Market Model does not require a global optimization: each rate is calibrated separately to the corresponding quoted smile.

However, if one needs to price products depending on the entire curve, this parameterization will be computationally complex both in theoretical terms such as computing dynamics (each volatility would be correlated to its own rate, and also the volatilities would be correlated with each other) and in practical terms such as Monte Carlo simulation. In addition this approach gives little synthetic information on the main factors driving the market.

So we also consider the case in which all rates are driven by a single stochastic volatility factor, which can be differently correlated to the different rates (the same volatility parameters are common to all rates).

We consider 3 maturities and 3 tenors with 9 different strikes each.

<table>
<thead>
<tr>
<th>Maturities</th>
<th>1y</th>
<th>5y</th>
<th>10y</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tenor</td>
<td>2y</td>
<td>5y</td>
<td>10y</td>
</tr>
<tr>
<td>Strikes&gt;ATM</td>
<td>+25</td>
<td>+50</td>
<td>+100</td>
</tr>
<tr>
<td>Strikes&lt;ATM</td>
<td>-25</td>
<td>-50</td>
<td>-100</td>
</tr>
</tbody>
</table>

First of all, we assess the difference between using a fully parameterized model as in the first case above, with a huge number of parameter, not tractable and little informative on the relevant market drivers, and a more synthetic and tractable model with one single stochastic volatility factor, that may be less satisfactory in terms of calibration capabilities. We use square volatility differences as objective function to minimize in calibration.

<table>
<thead>
<tr>
<th></th>
<th>Mean Error</th>
<th>Mean Error, $\beta = 1$</th>
</tr>
</thead>
<tbody>
<tr>
<td>SABR with as many stoch. vol. as rates</td>
<td>0.32 bps (0.00%)</td>
<td>4.53 bps (0.05%)</td>
</tr>
<tr>
<td>SABR with one single stoch vol</td>
<td>15.59 bps (0.16%)</td>
<td>17.06 bps (0.17%)</td>
</tr>
</tbody>
</table>

The fitting, here expressed as an average over all quotations considered, is good for all model specifications. It is virtually perfect for the fully parameterized
model. However there can be some reasons for considering a synthetic model
1) Tractability, which has already been considered above.
2) Scarce (or counterintuitive) effect on some exotic prices. Let us see the CMS
swaps with \( n = 5, 10 \) (payments up to five or ten years), \( c = 5, 10 \) (the CMS
leg pays 5/10 year length swap rates) and \( \Delta = 0.25 \) (yearly swap rates are paid
every three months).

\[
\begin{array}{cccc}
X_{5.5} & X_{10.5} & X_{5.10} & X_{10.10} \\
\text{Market CMS Spread} & 47.7 & 44.7 & 70.9 & 64.75 \\
\text{SABR with one single vol} & 42.05 & 40.5 & 66.4 & 63.3 \\
\text{SABR with as many vols as rates} & 42.1 & 39.4 & 66.5 & 61.3 \\
\end{array}
\]

3) The error of a synthetic model can be strongly reduced if we exclude illiquid
and non-significant quotations.

\[
\begin{array}{ccccc}
\text{Tenor=5} & \text{Strikes} & \text{Mat=1y} & 5y & 10y \\
\text{SABR bps Errors} & & & & \\
-200 & 78.7692 & 8.4558 & 59.7989 & \\
-100 & 13.7067 & 7.0278 & 9.4287 & \\
-50 & 30.7829 & 4.8987 & 12.9009 & \\
-25 & 31.4949 & 2.5373 & 20.4647 & \\
0 & 27.5593 & 0.4115 & 24.5359 & \\
+25 & 19.5776 & 1.4732 & 24.6896 & \\
+50 & 9.7622 & 3.5537 & 21.1655 & \\
+100 & 13.1901 & 4.2430 & 3.5364 & \\
+200 & 51.3264 & 2.6875 & 49.6430 & \\
\end{array}
\]
Tenor 5, maturity 1. Market: Dotted line. SABR: Continuous line.

If we exclude -200bps strike, we find that the error is

<table>
<thead>
<tr>
<th>Mean Error</th>
<th>SABR one vol (without lowest strike)</th>
<th>8.38 bps (0.08%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean Error</td>
<td>SABR one vol (with lowest strike)</td>
<td>17.06 bps (0.17%)</td>
</tr>
</tbody>
</table>

More than half of the error was due to an illiquid strike, not very significant when using implied volatility for measuring errors.

Now we compare SABR with term structure models, characterized by a more elaborated mean reverting process for the variance. We consider a Heston-Wu and Zhang Libor Model with exogenous correlation. For Heston-WZ the calibration is less immediate, since it does not allow for a closed-form implied volatility formula. For stability problems, for Heston-WZ we minimize squares of relative price difference, while the error is still expressed in terms of implied volatility.

### SABR Swap Parametric Form vs Heston Libor Term Structure Model

<table>
<thead>
<tr>
<th>Mean Error</th>
<th>SABR Swap Parametric Form</th>
<th>8.38 bps (0.08%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean Error</td>
<td>HESTON-WZ Libor Model</td>
<td>5.99 bps (0.06%)</td>
</tr>
</tbody>
</table>

It appears that a mean reverting structure of volatility allows a better fitting to market data even using a single stochastic volatility factor.

Using the output of above calibrations as initial guess, volatility minimization is stable also for Heston-WZ. Using volatility minimization for Heston-WZ too (so now WZ and SABR are fully comparable), the fitting is improved.
<table>
<thead>
<tr>
<th></th>
<th>Mean Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>HESTON-WZ LM</td>
<td>4.01 bps (0.04%)</td>
</tr>
<tr>
<td>SABR</td>
<td>8.38 bps (0.08%)</td>
</tr>
</tbody>
</table>

Heston-WZ outperforms SABR, in particular around the money, but at the cost of an increased computational burden (about 2.5 minutes in Matlab).
Calibration to swaption quotations with Tenor 5, maturity 1.
SABR results are on the left, Heston-WZ on the right.
Market: Dotted line. Models: Continuous line.
3 SABR in Libor Model framework

Following for example Wu and Zhang (2005), but considering SABR assumptions on the dynamics, we can start from the following dynamics under the reference risk neutral measure.

\[ dF_k(t) = \mu^Q_k \ dt + \alpha_k \sqrt{V oI(t)} \ dt + F_k(t)^\beta \ dZ^Q_k(t) \]

\[ E[dZ_i(t)dZ_j(t)] = \rho_{ij} dt \]

\[ dV oI(t) = vV oI(t) dW^Q(t), \quad V oI(0) = 1, \]

\[ E[dW(t)dZ_k(t)] = \rho^V_k dt \quad \forall k. \]

A peculiarity of the SABR model is that the volatility is directly modelled, rather than variance, as in more standard approaches. Therefore we can first express the dynamics in terms of variance. Application of Ito's Lemma leads to

\[ dV oI^2(t) = v^2 dV oI^2(t) dt + 2vV oI(t) dW^Q(t). \]

Notice that assuming a lognormal martingale for the volatility process \( V oI(t) \), like in SABR under the reference measure, leads to process for the variance \( V oI^2(t) \) that, although not a martingale, is still a lognormal process.

As for the forward rates, computing no-arbitrage dynamics follows closely the derivation typical of the reference lognormal model. Consider for example the spot-Libor measure, allowing us to express in a synthetic way the joint dynamics of forward rates. We obtain that under spot-Libor measure \( Q^D \)

\[ dF_k(t) = \alpha_k V oI(t) F_k^\beta(t) \left( V oI(t) \sum_{j=\gamma(t)+1}^{k} \frac{\tau_j \rho_{k,j} \alpha_j F_j^\beta(t)}{1 + \tau_j F_j(t)} dt + dZ_k^D(t) \right) \]

However what is of more interest to us, since it enters calibration, is the dynamics of forward rates and their volatilities under the associated pricing measures, namely the forward measures.

Via the standard change of numeraire toolkit we find the following forward measure \( Q^k \) Dynamics:

\[ dW^k(t) = dW^Q(t) + \xi_k(t) V oI(t) dt, \quad (1) \]

\[ \xi_k(t) = \sum_{j=1}^{k} \frac{\tau_j F_j(t) \alpha_j \rho_{k,j}^V}{1 + \tau_j F_j(t)} \]

7
so that
\[
\begin{align*}
\frac{d\text{Vol}^2(t)}{dt} &= v^2\text{Vol}^2(t)\frac{dW^k(t)}{dt} + 2v\text{Vol}^2(t) \left[ \frac{dW^k(t)}{dt} - \xi_k(t) \text{Vol}(t) \right] dt \\
&= v^2\text{Vol}^2(t)dt + 2v\text{Vol}^2(t)dW^k(t) - 2v\text{Vol}^2(t)\xi_k(t) dt
\end{align*}
\]

It is clear that, for obtaining a process of the same kind as the starting one and keep the tractability of a lognormal process, we need to freeze to zero all stochastic quantities in the correction to be made to the stochastic shock for measure change:

\[
dW^k(t) \approx dW^Q(t) + \hat{\xi}_k(t)\text{Vol}(0) dt,
\]

(2)

\[
\hat{\xi}_k(t) = \sum_{j=1}^k \frac{\tau_j F_j(0) \alpha_j \rho_j \text{Vol}(t)}{1 + \tau_j F_j(0)}
\]

While this approximation may be acceptable since these drifts tend to be not very high, still notice that actually (2) involves one more approximation than Wu and Zhang (2005) procedure, where the choice of a square root process for volatility allows to avoid the freezing of \(\text{Vol}(t)\), keeping the same distribution that the process has under the initial reference risk-neutral measure.

With the freezing in (2), the process for the stochastic variance becomes (recall \(\text{Vol}(0) = 1\))

\[
\frac{d\text{Vol}^2(t)}{dt} = \left[ v^2 - 2v\hat{\xi}_k(t) \right] \text{Vol}^2(t)dt + 2v\text{Vol}^2(t)dW^k(t)
\]

Using Ito’s lemma now to reverse to \(\text{Vol}(t)\), we obtain

\[
\begin{align*}
\frac{d\text{Vol}(t)}{dt} &= \left[ \frac{1}{2}v^2 - v\hat{\xi}_k(t) \right] \text{Vol}(t)dt + v\text{Vol}(t)dW^k(t) - \frac{1}{8v\text{Vol}^3}4v^2\text{Vol}^4(t)dt \\
&= -v\hat{\xi}_k(t)\text{Vol}(t)dt + v\text{Vol}(t)dW^k(t)
\end{align*}
\]

again a lognormal process, although not driftless.

### 3.0.1 Swap Rate dynamics

When we move to the swap measure, again the procedure for swap rates follows closely the standard LMM derivation. We assume the local exponent \(\beta\) to be the same across different rates, as done often in SABR implementation. Since

\[
S_{a,b}(t) = \sum_{j=a+1}^{b} \frac{P(t, T_i) \tau_i}{\sum_{i=a+1}^{b} P(t, T_i) \tau_i} F_j(t) = \sum_{j=a+1}^{b} w_i(t) F_j(t)
\]

(3)

by Ito’s Lemma we have that under \(Q^{\alpha,\beta}\)

\[
\frac{dS_{a,b}(t)}{dt} = \sum_{j=a+1}^{b} \frac{\partial S_{a,b}(t)}{\partial F_j(t)} F_j^\beta(t) \alpha_j \text{Vol}(t)dZ^a,b_j(t).
\]

(4)
We define

\[
\gamma_j(t) = \frac{\partial S_{a,b}(t)}{\partial F^j(t)} \frac{F^j(t)}{S_{a,b}(t)} =: \tilde{\omega}_j(t),
\]

\[
dS_{a,b}(t) = S_{a,b}^j(t) V o l(t) \sum_{j=a+1}^b \gamma_j(t) \alpha_j dZ_{a,b}^j(t).
\]

The process is still not tractable since the \(\gamma_j(t)\) are state dependent. So we approximate them with \(\gamma_j(0)\). Now we need to express the forward rate volatility functions as one single swap rate volatility function, and the forward rate stochastic drivers as one single swap rate stochastic driver. Thus we set

\[
\alpha_{a,b} = \frac{1}{S_{a,b}^j(0)} \sqrt{\sum_{k=a+1}^b \sum_{h=a+1}^b \gamma_k(0) \alpha_k \gamma_h(0) \alpha_h \rho_{k,h}}, \quad dZ_{a,b}(t) = \frac{\sum_{j=a+1}^b \gamma_j(0) \alpha_j dZ_{a,b}^j(t)}{\alpha_{a,b}(t)}
\]

or

\[
\alpha_{a,b} = \frac{1}{S_{a,b}^j(0)} \sqrt{\sum_{k=a+1}^b \sum_{h=a+1}^b \tilde{\omega}_k(0) F_{\beta}^j(0) \alpha_k \tilde{\omega}_h(0) F_{\beta}^j(0) \alpha_h \rho_{k,h}}.
\]

Notice that

\[
\mathbb{E} \left[ \sum_{j=a+1}^b \gamma_j(0) \alpha_j dZ_{a,b}^j(t) \cdot \sum_{j=a+1}^b \gamma_j(0) \alpha_j dZ_{a,b}^j(t) \right]
\]

\[
= \sum_{k=a+1}^b \sum_{h=a+1}^b \gamma_k(0) \gamma_h(0) \alpha_k \alpha_h \mathbb{E} \left[ dZ_{a,b}^j(t) dZ_{a,b}^j(t) \right]
\]

\[
= \sum_{k=a+1}^b \sum_{h=a+1}^b \gamma_k(0) \gamma_h(0) \alpha_k \alpha_h \rho_{k,h} dt = \alpha_{a,b}^2 dt
\]

\[
\mathbb{E} \left[ dZ_{a,b}^j(t) dZ_{a,b}^j(t) \right] = \frac{\alpha_{a,b}^2 dt}{\alpha_{a,b}^2} = dt
\]

In case one first freezes the \(w_i(t)\) to zero in (3), then

\[
S_{a,b}(t) = \sum_{j=a+1}^b w_i(0) F_i(t)
\]
with dynamics

\[ dS_{a,b}(t) = \sum_{j=a+1}^{b} \frac{\partial S_{a,b}(t)}{\partial F_j(t)} F_j^\beta(t) \alpha_j \text{Vol}(t) dZ_{a,b}^j(t) \]

\[ = \sum_{j=a+1}^{b} w_j(0) F_j^\beta(t) \alpha_j \text{Vol}(t) dZ_{a,b}^j(t) \]

\[ = S_{a,b}^\beta(t) \sum_{j=a+1}^{b} \frac{w_j(0)}{S_{a,b}^\beta(t)} \alpha_j \text{Vol}(t) dZ_{a,b}^j(t) \]

At this point one freezes the \( F_j(t) \) to \( F_j(0) \), \( S_{a,b}(t) \) to \( S_{a,b}(0) \) and finds

\[ \alpha_{a,b} = \frac{1}{S_{a,b}^\beta(0)} \sqrt{\sum_{k=a+1}^{b} \sum_{h=a+1}^{b} w_k(0) F_k^\beta(0) \alpha_k w_h(0) F_h^\beta(0) \rho_{k,h}}. \]

The resulting dynamics is, in any case

\[ dS_{a,b}(t) = S_{a,b}^\beta(t) \alpha_{a,b} \sqrt{\text{Vol}(t)} dZ_{a,b}(t), \]

analogous to the forward rate dynamics we started from.

### 3.0.2 Volatility and Correlations under the swap measure

Now we need to compute the change of drift for the volatility under a generic swap measure. Again following WZ, we obtain

\[ dW^a,b(t) = dW^Q(t) + \xi_{a,b}(t) \text{Vol}(t) dt, \]

\[ \xi_{a,b}(t) = \sum_{j=a+1}^{b} w_j(t) \xi_j(t) \]

As we did before, we move to the following approximation

\[ dW^a,b(t) \approx dW^Q(t) + \hat{\xi}_{a,b}(t) \text{Vol}(0) dt, \]

\[ \hat{\xi}_{a,b}(t) = \sum_{j=a+1}^{b} w_j(0) \hat{\xi}_j(t) \]

and we find

\[ d\text{Vol}(t) = -v \xi_{a,b}(t) \text{Vol}(t) dt + v \text{Vol}(t) dW^a,b(t). \]

Now we also need to compute the correlation \( \rho_{a,b}^V \) between the swap rate and the stochastic volatility factor (that, notice, is always the same process). It is

\[ \rho_{a,b}^V dt = \mathbb{E} \left[ dW^a,b(t) dZ_{a,b}^a(t) \right] = \mathbb{E} \left[ dW^a,b(t) dZ_{a,b}^b(t) \right] \]

\[ = \frac{\sum_{j=a+1}^{b} \gamma_j(t) \alpha_j(t) \mathbb{E} \left[ dW^a,b(t) dZ_j^a(t) \right]}{\alpha_{a,b}} = \frac{\sum_{j=a+1}^{b} \gamma_j(t) \alpha_j(t) \rho_{a,b}^V dt}{\alpha_{a,b}} \]
This is an important parameter in implementation (recall that instantaneous correlations do not change across equivalent measures).

Both forward and swap rates turn out to be (CEV) martingales under their natural measure, and under any swap or forward measure the volatility is a process

\[ dVol(t) = \nu \mu(t) Vol(t) dt + \nu Vol(t) dW(t), \]

where the drift is measure dependent but deterministic. This is a geometric brownian motion, and this allows us to use via approximation the SABR implied volatility formula.

4 Hedging tests

SABR model was introduced based on the consideration that local volatility models, a popular choice for interest rate modelling with smile, have an undesirable dynamic behaviour. In fact, when we are using a model for forward rates or forward swap rates and we change the level of the underlying, for example from \( F \) to \( F + \varepsilon \) we would like the smile to move in the same direction, since this appears to be the usual market behaviour. Indicating with \( \sigma_F(K) \) the implied volatility curve predicted by the model when the underlying forward is \( F \), would like to observe that

\[ F \rightarrow F + \varepsilon \Rightarrow \sigma_{F+\varepsilon}(K) = \sigma_F(K - \varepsilon). \]  

(5)

This means that if we move the underlying to the right, namely increasing it, also the implied volatility curve will move to the right, so that, for example, a property such as “the minimum of the smile curve is at the strike corresponding to ATM” is preserved by changes in the underlying. In fact (5) implies

\[ F \rightarrow F + \varepsilon \Rightarrow \sigma_{F+\varepsilon}(F + \varepsilon) = \sigma_F(F). \]

On the contrary, local volatility models ( to be precise, those models where the implied volatility is a function of the underlying only, with no direct dependence on time) have a different behaviour, summarized by:

\[ F \rightarrow F + \varepsilon \Rightarrow \sigma_{F+\varepsilon}(K) = \sigma_F(K + \varepsilon), \]

the very opposite of the desired behaviour.

Not in contrast with the above analysis, one can show with simple examples that the local volatility part of SABR behaves as in local volatility models, so that SABR behave as desired in terms of hedging behaviour only when stochastic volatility effects overwhelm local volatility effects, and in particular when the skew is determined by the correlation between rates and stochastic volatility, so \( \rho \neq 0 \), possibly with the local exponent \( \beta \) set to 1 like in a standard lognormal model. This has practical implications, since one popular choice among market practitioner for implementing the SABR model is to set correlation to zero.
Following this result, we consider how standard hedging tests, such as the one of Hagan et al. (2002), can be modified for taking explicitly into account that

\[ \mathbb{E}[dW_V \, dW_P] = \rho \, dt, \]

rather than being zero. Simple numerical tests show that in this case the model behaviour resembles more closely the behaviour of local volatility models. We conclude making a few considerations on the consequences of this fact on hedging assumptions, including examples of behaviour of local volatility models in out-of-the-model hedging.