

INTEREST RATES AND FX MODELS

4. SABR and Smiles

Andrew Lesniewski

Baruch College
New York

September 24, 2013

Contents

1	Introduction	2
2	Stochastic volatility and the SABR model	2
2.1	Implied volatility	3
2.2	Calibration of SABR	5
2.3	Arbitrage in the SABR model	6
3	Volatility cube	7
3.1	ATM swaption volatilities	8
3.2	Stripping cap volatility	9
3.3	Adding the third dimension	10

1 Introduction

The volatility skew models that we have discussed so far improve on Black's models but still fail to reflect the market dynamics. One issue is, for example, the "wing effect" exhibited by the implied volatilities of some maturities (especially shorter dated) and tenors which is not captured by these models: the implied volatilities tend to rise for high strikes forming the familiar "smile" shape. Among the attempts to move beyond the locality framework are:

- (a) *Stochastic volatility models*. In this approach, we add a new stochastic factor to the dynamics by assuming that a suitable volatility parameter itself follows a stochastic process.
- (b) *Jump diffusion models*. These models use a broader class of stochastic processes (for example, *Levy processes*) to drive the dynamics of the underlying asset. These more general processes allow for discontinuities ("jumps") in the asset dynamics.

Because of time constraints we shall limit our discussion to an example of approach (a), namely the SABR stochastic volatility model.

2 Stochastic volatility and the SABR model

The SABR model is an extension of the CEV model in which the volatility parameter is assumed to follow a stochastic process. Its dynamics is explicitly given

by:

$$\begin{aligned} dF(t) &= \sigma(t)C(F(t))dW(t), \\ d\sigma(t) &= \alpha\sigma(t)dZ(t). \end{aligned} \tag{1}$$

Here $F(t)$ is the forward rate process which, depending on context, may denote a LIBOR forward or a forward swap rate¹, and $\sigma(t)$ is the stochastic volatility parameter. The process is driven by two Brownian motions, $W(t)$ and $Z(t)$, with

$$\mathbb{E}[dW(t)dZ(t)] = \rho dt,$$

where the correlation ρ is assumed constant. The diffusion coefficient $C(F)$ is assumed to be of the CEV type:

$$C(F) = F^\beta. \tag{2}$$

Note that we assume that a suitable numeraire has been chosen so that $F(t)$ is a martingale. The process $\sigma(t)$ is the stochastic component of the volatility of F_t , and α is the volatility of $\sigma(t)$ (the *volvol*) which is also assumed to be constant. As usual, we supplement the dynamics with the initial condition

$$\begin{aligned} F(0) &= F_0, \\ \sigma(0) &= \sigma_0, \end{aligned} \tag{3}$$

where F_0 is the current value of the forward, and σ_0 is the current value of the volatility parameter.

Note that the dynamics (1) requires a boundary condition at $F = 0$. One usually imposes the absorbing (Dirichlet) boundary condition.

2.1 Implied volatility

As in the case of the CEV model, the analysis of the SABR model requires solving the terminal value problem for the backward Kolmogorov equation associated with the process (1). Namely, the valuation function $B = B(t, f, \sigma)$ is the solution to

$$\begin{aligned} \frac{\partial}{\partial t} B + \frac{1}{2} \sigma^2 \left(f^{2\beta} \frac{\partial^2}{\partial f^2} + 2\alpha\rho f^\beta \frac{\partial^2}{\partial f \partial \sigma} + \alpha^2 \frac{\partial^2}{\partial \sigma^2} \right) B &= 0, \\ B(T, f, \sigma) &= \begin{cases} (f - K)^+, & \text{for a call,} \\ (K - f)^+, & \text{for a put.} \end{cases} \end{aligned} \tag{4}$$

¹The SABR model specification is also used in markets other than interest rate market, and thus $F(t)$ may denote e.g. a crude oil forward.

This is a more difficult problem than the models discussed above. Except for the special case of $\beta = 0$, no explicit solution to this model is known. The general case can be solved approximately by means of a perturbation expansion in the parameter $\varepsilon = T\alpha^2$, where T is the maturity of the option. As it happens, this parameter is typically small and the approximate solution is actually quite accurate. Also significantly, this solution is very easy to implement in computer code, and lends itself well to risk management of large portfolios of options in real time.

The parameter α shows a persistent stable term structure as a function of the swaption maturity and the tenor of the underlying swap. On a given market snapshot, the highest α 's is located in the upper left corner of the volatility matrix (short expirations and short tenors), and the lowest one is located in the lower right corner (long expirations and long tenors). α is a monotone decreasing function of both the option expiry and the underlying swap tenor. A typical range of values of α is $0.2 \lesssim \alpha \lesssim 2$.²

There is no known closed form option valuation formula in the SABR model. Instead, one takes the following approach. We force the valuation formula to be of the form

$$\begin{aligned} P^{\text{pay}}(T, K, F_0, \sigma_n) &= A_0(T_0, T) \sigma_n \sqrt{T} \left(d_+ N(d_+) + N'(d_+) \right), \\ P^{\text{rec}}(T, K, F_0, \sigma_n) &= A_0(T_0, T) \sigma_n \sqrt{T} \left(d_- N(d_-) + N'(d_-) \right), \\ d_{\pm} &= \pm \frac{F_0 - K}{\sigma_n \sqrt{T}}, \end{aligned} \quad (5)$$

given by the normal model, with the (normal) implied volatility σ_n depending on the SABR model parameters. A detailed analysis of the model dynamics shows that the implied normal volatility is then approximately given by:

$$\begin{aligned} \sigma_n(T, K, F_0, \sigma_0, \alpha, \beta, \rho) &= \alpha \frac{F_0 - K}{\delta(\zeta)} \left\{ 1 + \left[\frac{2\gamma_2 - \gamma_1^2}{24} \right. \right. \\ &\quad \times \left. \left(\frac{\sigma_0 C(F_{\text{mid}})}{\alpha} \right)^2 + \frac{\rho\gamma_1}{4} \frac{\sigma_0 C(F_{\text{mid}})}{\alpha} + \frac{2 - 3\rho^2}{24} \right] \varepsilon + \dots \left. \right\}, \end{aligned} \quad (6)$$

where F_{mid} denotes a conveniently chosen midpoint between F_0 and K (such as

²On a few days at the height of the recent financial crisis the value of α corresponding to 1 month into 1 year swaptions was as high as 4.7.

$(F_0 + K)/2$), and

$$\gamma_1 = \frac{C'(F_{\text{mid}})}{C(F_{\text{mid}})},$$

$$\gamma_2 = \frac{C''(F_{\text{mid}})}{C(F_{\text{mid}})}.$$

The “distance function” entering the formula above is given by:

$$\delta(\zeta) = \log \left(\frac{\sqrt{1 - 2\rho\zeta + \zeta^2} + \zeta - \rho}{1 - \rho} \right),$$

where

$$\begin{aligned} \zeta &= \frac{\alpha}{\sigma_0} \int_K^{F_0} \frac{dx}{C(x)} \\ &= \frac{\alpha}{\sigma_0(1-\beta)} \left(F_0^{1-\beta} - K^{1-\beta} \right). \end{aligned} \quad (7)$$

A similar asymptotic formula exists for the implied lognormal volatility σ_{ln} . Namely,

$$\begin{aligned} \sigma_{\text{ln}}(T, K, F_0, \sigma_0, \alpha, \beta, \rho) &= \alpha \frac{\log(F_0/K)}{D(\zeta)} \left\{ 1 + \left[\frac{2\gamma_2 - \gamma_1^2 + 1/F_{\text{mid}}^2}{24} \right. \right. \\ &\quad \times \left. \left(\frac{\sigma_0 C(F_{\text{mid}})}{\alpha} \right)^2 + \frac{\rho\gamma_1}{4} \frac{\sigma_0 C(F_{\text{mid}})}{\alpha} + \frac{2 - 3\rho^2}{24} \right] \varepsilon + \dots \Big\}. \end{aligned} \quad (8)$$

2.2 Calibration of SABR

For each option maturity and underlying we have to specify 4 model parameters: $\sigma_0, \alpha, \beta, \rho$. In order to do it we need, of course, market implied volatilities for several different strikes. Given this, the calibration poses no problem: one can use, for example, Excel’s Solver utility.

It turns out that there is a bit of redundancy between the parameters β and ρ . As a result, one usually calibrates the model by fixing one of these parameters:

- (a) Fix $\beta \leq 1$, say $\beta = 0.5$, and calibrate σ_0, α, ρ . This choice of calibration methodology works quite well under “normal” conditions. In times of distress, such as 2008 - 2009, the height of the recent financial crisis, the choice of $\beta = 0.5$ occasionally led to extreme calibrations of the correlation parameters ($\rho = \pm 1$). As a result, some practitioners choose high β ’s, $\beta \approx 1$ for short expiry options and let it decay as option expiries move out.

- (b) Fix $\rho = 0$, and calibrate σ_0, α, β .

Calibration results show interesting term structure of the model parameters as functions of the maturity and underlying. As we already mentioned, typical is the shape of the parameter α which start out high for short dated options and then declines monotonically as the option expiration increases. This indicates presumably that modeling short dated options should include a jump diffusion component.

2.3 Arbitrage in the SABR model

The explicit implied volatility given by formulas (6) or (8) make the SABR model easy to implement, calibrate, and use. These implied volatility formulas are usually treated as if they were exact, even though they are derived from an asymptotic expansion which requires that $\alpha^2 T \ll 1$. The unstated argument is that instead of treating these formulas as an accurate approximation to the SABR model, they could be regarded as the exact solution to some other model which is well approximated by the SABR model. This is a valid viewpoint as long as the option prices obtained using the explicit formulas for σ_n (or σ_{ln}) are arbitrage free. There are two key requirements for arbitrage freeness of a volatility smile model:

- (i) Put-call parity, which holds automatically since we are using the same implied volatility σ_n for both calls and puts.
- (ii) The terminal probability density function implied by the call and put prices needs to be positive.

To explore the second condition, note that call and put prices can be written quite generally as

$$\begin{aligned} P^{\text{call}}(T, K) &= \int_{-\infty}^{\infty} (F - K)^+ G_T(F, K) dF, \\ P^{\text{put}}(T, K) &= \int_{-\infty}^{\infty} (K - F)^+ G_T(F, K) dF, \end{aligned} \tag{9}$$

where $G_T(F, K)$ is the risk-neutral probability density at the exercise date (including the delta function from the Dirichlet boundary condition). Since

$$\frac{d^2}{dx^2} x^+ = \delta(x), \tag{10}$$

it follows that

$$\begin{aligned}\frac{\partial^2}{\partial K^2} P^{\text{call}}(T, K) &= \frac{\partial^2}{\partial K^2} P^{\text{put}}(T, K) \\ &= G_T(F, K) \\ &\geq 0,\end{aligned}\tag{11}$$

for all K . So the explicit implied volatility formula (6) can represent an arbitrage free model only if

$$\begin{aligned}\frac{\partial^2}{\partial K^2} P^{\text{call}}(T, K, F_0, \sigma_n(K, \dots)) &= \frac{\partial^2}{\partial K^2} P^{\text{put}}(T, K, F_0, \sigma_n(K, \dots)) \\ &\geq 0.\end{aligned}\tag{12}$$

In other words, there cannot be a “butterfly arbitrage”. As it turns out, it is not terribly uncommon for this requirement to be violated for very low strike and long expiry options. The problem does not appear to be the quality of the call and put prices obtained from the explicit implied volatility formulas, because these usually remain quite accurate. Rather, the problem seems to be that implied volatility curves are not a robust representation of option prices for low strikes. It is very easy to find a reasonable looking volatility curve $\sigma_n(K, \dots)$ which violates the arbitrage free constraint in (11) for a range of values of K .

3 Volatility cube

We can organize the implied swaption volatilities by:

- (a) Option expiration.
- (b) Tenor of the underlying swap.
- (c) Strike on the option.

This three dimensional object is referred to as the *volatility cube*. Swaption market is relatively liquid but, on a given day, only a limited number of structures trade. Out of over a hundred elements of the volatility matrix, the market provides information for certain benchmark maturities (1 month, 3 months, 6 months, 1 year, ...), underlyings (1 year, 2 years, 5 years, ...), and strikes (ATM, ± 50 bp, ...) only. The process of volatility cube construction requires performing intelligent interpolations for the remaining expiries and tenors in a way that does not allow for arbitrage.

3.1 ATM swaption volatilities

The market quotes swaption volatilities for certain standard maturities and underlyings. Table 1 contains the December 13, 2011 snapshot of the matrix of lognormal (Black) at the money swaption volatilities for the standard expirations and underlyings.

	1Y	2Y	3Y	4Y	5Y	7Y	10Y	15Y	20Y	30Y
1M	71.7%	68.2%	61.8%	56.0%	50.6%	49.7%	47.3%	44.4%	43.4%	43.0%
3M	73.6%	68.8%	61.1%	55.9%	51.5%	49.6%	47.2%	44.0%	42.8%	42.8%
6M	72.2%	70.0%	60.0%	55.6%	51.9%	49.0%	46.1%	42.9%	42.1%	41.9%
1Y	73.8%	65.5%	57.1%	52.9%	49.7%	46.6%	43.8%	40.8%	40.3%	40.0%
2Y	73.7%	59.0%	51.4%	47.2%	45.4%	42.0%	39.4%	37.0%	36.6%	36.4%
3Y	57.8%	49.0%	44.3%	41.8%	40.4%	37.9%	36.0%	34.3%	34.1%	33.4%
4Y	46.0%	41.6%	39.2%	37.8%	36.8%	35.1%	33.9%	32.7%	32.4%	32.1%
5Y	39.7%	37.8%	36.6%	35.6%	34.8%	33.5%	32.6%	31.7%	31.6%	31.3%
7Y	34.4%	33.4%	32.4%	31.7%	31.2%	30.7%	30.7%	30.0%	29.7%	29.2%
10Y	29.8%	29.4%	29.1%	28.8%	28.7%	28.8%	29.2%	28.2%	27.5%	26.8%

Table 1: Swaption ATM lognormal volatilities

Alternatively, Table 2 contains the matrix of normal at the money swaption volatilities expressed, for ease of readability, in basis points.

	1Y	2Y	3Y	4Y	5Y	7Y	10Y	15Y	20Y	30Y
1M	48	48	51	58	65	85	101	111	114	116
3M	51	49	53	61	69	88	102	111	113	116
6M	51	52	56	66	75	91	103	110	112	115
1Y	52	56	63	74	82	94	104	108	110	111
2Y	72	76	83	88	94	99	103	104	104	104
3Y	91	93	95	98	100	101	103	102	101	99
4Y	101	101	101	102	102	102	103	100	98	97
5Y	105	104	104	104	104	103	102	99	97	95
7Y	103	102	101	100	100	99	98	94	92	89
10Y	98	97	96	95	94	93	92	87	84	81

Table 2: Swaption ATM normal volatilities (in basis points)

The matrix of at the money volatilities should be accompanied by the matrix of forward swap rates, calculated from the rate market snapshot. These are the at the money strikes for the corresponding swaptions. This is illustrated by Table 3.

	1Y	2Y	3Y	4Y	5Y	7Y	10Y	15Y	20Y	30Y
1M	0.67%	0.70%	0.82%	1.04%	1.29%	1.72%	2.13%	2.49%	2.62%	2.71%
3M	0.69%	0.72%	0.87%	1.10%	1.35%	1.77%	2.18%	2.52%	2.64%	2.72%
6M	0.71%	0.76%	0.94%	1.19%	1.45%	1.86%	2.25%	2.57%	2.67%	2.75%
1Y	0.72%	0.87%	1.13%	1.41%	1.66%	2.03%	2.38%	2.66%	2.74%	2.80%
2Y	1.02%	1.33%	1.64%	1.90%	2.10%	2.39%	2.66%	2.84%	2.88%	2.90%
3Y	1.64%	1.96%	2.21%	2.39%	2.52%	2.72%	2.91%	3.00%	3.00%	3.00%
4Y	2.27%	2.50%	2.64%	2.75%	2.84%	2.98%	3.09%	3.11%	3.09%	3.07%
5Y	2.72%	2.84%	2.92%	2.99%	3.06%	3.15%	3.21%	3.18%	3.14%	3.11%
7Y	3.08%	3.16%	3.21%	3.26%	3.29%	3.31%	3.29%	3.22%	3.17%	3.13%
10Y	3.40%	3.41%	3.41%	3.40%	3.38%	3.33%	3.26%	3.18%	3.14%	3.11%

Table 3: Forward swap rates (ATM strikes)

3.2 Stripping cap volatility

A cap is a basket of options of different maturities and different moneynesses. For simplicity, the market quotes cap / floor prices in terms of a single number, the *flat volatility*. This is the single volatility which, when substituted into the valuation formula (for all caplets / floorlets!), reproduces the correct price of the instrument. Clearly, flat volatility is a dubious concept: since a single caplet may be part of different caps it gets assigned different flat volatilities. The process of constructing actual implied caplet volatility from market quotes is called *stripping* cap volatility. The result of stripping is a sequence of ATM caplet volatilities for maturities ranging from one day to, say, 30 years. Convenient benchmarks are 3 months, 6 months, 9 months, The market data usually include Eurodollar options and OTC caps and floors.

There are various methods of stripping at the money cap volatility. Among them we mention:

Bootstrap. One starts at the short end and moves further trying to match the prices of Eurodollar options and spot starting caps / floors. This method tends to produce a jagged shape of the volatility curve.

Optimization. This method produces a smooth shape of the cap volatility curve but is somewhat more involved. We use a two step approach: in the first step fit the caplet volatilities to the *hump function*:

$$H(t) = (\alpha + \beta t) e^{-\lambda t} + \mu. \quad (13)$$

Generally, the hump function gives a qualitatively correct shape of the cap volatility. Quantitatively, the fit is insufficient for accurate pricing and should be refined. An good approach is to use smoothing B-splines. Once α, β, λ , and μ have been

calibrated, we use cubic B-splines in a way similar to the method explained in Lecture 1 in order to nail down the details of the caplet volatility curve.

3.3 Adding the third dimension

The third dimension of the volatility cube is the strike dependence of volatility. It is best characterized in terms of the parameters of the smile model such as the parameters α , β , and ρ of the SABR model. We can conveniently organize these parameters in matrices of the same dimensions as the at the money volatility matrices.

References

- [1] Gatheral, J.: *The Volatility Surface: A Practitioner's Guide*, Wiley (2006).
- [2] Hagan, P., Kumar, D., Lesniewski, A., and Woodward, D.: Managing smile risk, *Wilmott Magazine*, **September**, 84 - 108 (2002).
- [3] Hagan, P., Kumar, D., Lesniewski, A., and Woodward, D.: Arbitrage free SABR, *Wilmott Magazine* (2013).