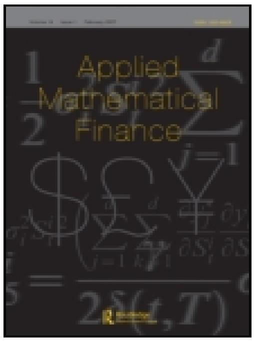

Applied Mathematical Finance

ISSN: 1350-486X (Print) 1466-4313 (Online) Journal homepage: www.tandfonline.com/journals/ramf20

Calibration of the SABR Model in Illiquid Markets

Graeme West

To cite this article: Graeme West (2005) Calibration of the SABR Model in Illiquid Markets, Applied Mathematical Finance, 12:4, 371-385, DOI: 10.1080/13504860500148672

To link to this article: https://doi.org/10.1080/13504860500148672

Published online: 17 Feb 2007.

Applied Mathematical Finance,
Vol. 12, No. 4, 371–385, December 2005

Calibration of the SABR Model in Illiquid Markets

GRAEME WEST

Financial Modelling Agency, South Africa, and Programme in Advanced Mathematics of Finance, School of Computational & Applied Mathematics, University of the Witwatersrand, South Africa

(Received 14 July 2004; in revised form 9 November 2004)

ABSTRACT *Recently the SABR model has been developed to manage the option smile which is observed in derivatives markets. Typically, calibration of such models is straightforward as there is adequate data available for robust extraction of the parameters required asinputs to the model. The paper considers calibration of the model in situations where input data is very sparse. Although this will require some creative decision making, the algorithms developed here are remarkably robust and can be used confidently for mark to market and hedging of option portfolios.*

KEY WORDS: SABR model, equity derivatives, volatility skew calibration, illiquid markets

Why another Skew Model?

Vanilla OTC European options or European futures options are priced and often hedged using respectively the Black–Scholes or Black model. In these models there is a one-to-one relation between the price of the option and the volatility parameter σ, and option prices are often quoted by stating the implied volatility σ_{imp}, the unique value of the volatility which yields the option price when used in the formula. In the classical Black–Scholes–Merton world, volatility is a constant. But in reality, options with different strikes require different volatilities to match their market prices. This is the market skew or smile. Typically, although not always, the word skew is reserved for the slope of the volatility/strike function, and smile for its curvature.

Handling these market skews and smiles correctly is critical for hedging. One would like to have a coherent estimate of volatility risk, across all the different strikes and maturities of the positions in the book.

The development of local volatility models in Dupire (1994, 1997), Derman & Kani (1994), Derman *et al.* (1996) and Derman & Kani (1998) was a major advance in handling smiles and skews. Another crucial thread of development is the

Correspondence Address: Graeme West, Financial Modelling Agency, 19 First Ave East, Parktown North, 2193, South Africa. Email: graeme@finmod.co.za

1350-486X Print/1466-4313 Online/05/040371–15
DOI: 10.1080/13504860500148672

stochastic volatility approach, for which the reader is referred to Hull & White (1987), Heston (1993), Lewis (2000), Fouque *et al.* (2000), Lipton (2003), and finally Hagan *et al.* (2002), which is the model we will consider here.

Local volatility models are self-consistent, arbitrage-free, and can be calibrated to precisely match observed market smiles and skews. Currently these models are the most popular way of managing smile and skew risk. Possibly they are often preferred to stochastic volatility models for computational reasons: the local volatility models are tree models; to price with stochastic volatility models typically means Monte Carlo. However, it has recently been observed (Hagan *et al.*, 2002) that the dynamic behaviour of smiles and skews predicted by local volatility models is exactly opposite the behaviour observed in the marketplace: local volatility models predict that the skew moves in the opposite direction to the market level, in reality, it moves in the same direction. This leads to extremely poor hedging results within these models, and the hedges are often worse than the naive Black model hedges, because these naive hedges are in fact consistent with the smile moving in the same direction as the market.

To resolve this problem, Hagan *et al.* (2002) derived the SABR model. The model allows the market price and the market risks, including vanna and volga risks, to be obtained immediately from Black's formula. It also provides good, and sometimes spectacular, fits to the implied volatility curves observed in the marketplace. More importantly, the SABR model captures the correct dynamics of the smile, and thus yields stable hedges.

The Model

Stochastic volatility models are in general characterized by the use of two driving correlated Brownian motions, one which determines the increments to the underlying process and the other determines the increments to the volatility process. For example, the model of Hull & White (1987) can be summarized as follows:

$$\mathrm{d}F = \phi F\,\mathrm{d}t + \sigma F\,\mathrm{d}W_1 \tag{1}$$

$$\mathrm{d}\sigma^2 = \mu\sigma^2\,\mathrm{d}t + \xi\sigma^2\,\mathrm{d}W_2 \tag{2}$$

$$\mathrm{d}W_1\,\mathrm{d}W_2 = \rho\,\mathrm{d}t \tag{3}$$

where ϕ, μ and ξ are time and state dependent functions, and $\mathrm{d}W_1$ and $\mathrm{d}W_2$ are correlated Brownian motions.

Similarly, the model of Heston (1993) proceeds with the pair of driving equations

$$\mathrm{d}F = \mu F\,\mathrm{d}t + \sigma F\,\mathrm{d}W_1 \tag{4}$$

$$\mathrm{d}\sigma = -\beta\sigma\,\mathrm{d}t + \delta\,\mathrm{d}W_2 \tag{5}$$

$$\mathrm{d}W_1\,\mathrm{d}W_2 = \rho\,\mathrm{d}t \tag{6}$$

where this time, μ, β and δ are constants.

As another example, the models of Fouque *et al.* (2000) are variations on the following initial set-up:

$$\mathrm{d}F = \mu F\,\mathrm{d}t + \sigma F\,\mathrm{d}W_1 \tag{7}$$

$$\mathrm{d}y = \alpha(m - y)\mathrm{d}t + \beta\,\mathrm{d}W_2 \tag{8}$$

$$\mathrm{d}W_1\,\mathrm{d}W_2 = \rho\,\mathrm{d}t \tag{9}$$

where this time α, m and β are constants, and for example $y=\ln\sigma$. Here, the process for y is a mean reverting Ornstein–Uhlenbeck process.

The model we consider here is known as the stochastic $\alpha\beta\rho$ model, or SABR model. Here

$$\mathrm{d}F = \alpha F^{\beta}\,\mathrm{d}W_1 \tag{10}$$

$$\mathrm{d}\alpha = v\alpha\,\mathrm{d}W_2 \tag{11}$$

$$\mathrm{d}W_1\,\mathrm{d}W_2 = \rho\,\mathrm{d}t \tag{12}$$

where the factors F and α are stochastic, and the parameters β, ρ and v are not. α is a 'volatility-like' parameter: not equal to the volatility, but there will be a functional relationship between this parameter and the at the money volatility, as we shall see in due course. The constant v is to be thought of as the volatility of volatility, a model feature which acknowledges that volatility obeys well-known clustering in time. The parameter $\beta \in [0, 1]$ determines the relationship between futures spot and at the money volatility: $\beta \approx 1$ indicates that the user believes that if the market were to move up or down in an orderly fashion, the at the money volatility level would not be significantly affected. $\beta \ll 1$ indicates that if the market were to move then at the money volatility would move in the opposite direction. The closer to 0 the more pronounced would be this phenomenon. Furthermore, the closer β is to 1 (respectively, 0) the more lognormal- (respectively, normal-) like is the stochastic model.

The Option Pricing Formula

A desirable feature of any local or stochastic volatility model is that the model will reproduce the prices of the vanilla instruments that were used as inputs to the calibration of the model. Material failure to do so will make the model not arbitrage free and render it almost useless.

A significant feature of the SABR model is that the prices of vanilla instruments can be recovered from the model in closed form (up to the accuracy of a series expansion). This is dealt with in detail in Hagan *et al.* (2002, Appendix B). Essentially it is shown there that the price of a vanilla option under the SABR model is given by the appropriate Black formula, provided the correct implied volatility is used. For given α, β, ρ, v and τ, this volatility is given by:

$$\sigma(X,F)=\frac{\alpha\left(1+\left(\frac{(1-\beta)^2}{24}\frac{\alpha^2}{(FX)^{1-\beta}}+\frac{1}{4}\frac{\rho\beta v\alpha}{(FX)^{(1-\beta)/2}}+\frac{2-3\rho^2}{24}v^2\right)\tau\right)}{(FX)^{(1-\beta)/2}\left[1+\frac{(1-\beta)^2}{24}\ln^2\frac{F}{X}+\frac{(1-\beta)^4}{1920}\ln^4\frac{F}{X}\right]}\frac{z}{\chi(z)} \tag{13}$$

$$z=\frac{v}{\alpha}(FX)^{(1-\beta)/2}\ln\frac{F}{X} \tag{14}$$

$$\chi(z)=\ln\left(\frac{\sqrt{1-2\rho z+z^2}+z-\rho}{1-\rho}\right) \tag{15}$$

Although the formula appears fearsome, it is closed form, so practically instantaneous. This formula of course can be viewed as a functional form for the volatility skew, and so, when this volatility skew is observable, we have some sort of error minimisation problem, which, subject to the caveats raised in Hagan *et al.* (2002), is quite elementary. The thesis of this article is the same calibration problem in the absence of an observable skew, in which case, we need a model to infer the parametric form of the skew given a history of traded data.

Note as in Hagan *et al.* (2002) that if $F=X$ then the z and $\chi(z)$ terms are removed from the equation, as then $\frac{z}{\chi(z)}=1$ in the sense of a limit[1], and so

$$\sigma(F,F)=\frac{\alpha\left(1+\left(\frac{(1-\beta)^2}{24}\frac{\alpha^2}{F^{2-2\beta}}+\frac{1}{4}\frac{\rho\beta v\alpha}{F^{1-\beta}}+\frac{2-3\rho^2}{24}v^2\right)\tau\right)}{F^{1-\beta}} \tag{16}$$

The Market we Consider for this Analysis

We consider the equity futures market traded at the South African Futures Exchange. For details of the operation of this market the reader is referred to SAFEX (2004), West (2005, Chapter 10).

This market is characterized by an illiquidity that is gross compared to other markets. We will focus on the TOP40 (the index of the biggest shares, as determined by free float market capitalisation and liquidity) futures options contracts. Contracts exist for expiry in March, June, September and December of each year. Among these, the following March contract is the most liquid, along with the nearest contract. Nevertheless, the March contract only becomes liquid in anything like a meaningful manner about two years before expiry. Packages of options on the March 2004 contract traded a total of perhaps 800 times, for which there were about double that number of different strikes. (By a package, we mean not only a single trade, but a collar, butterfly, condor, etc.) The full set of strike and volatility history is not published. Nevertheless, we sourced, via one of the largest brokers, a significant portion of the history (possibly 70% or more), which we have taken as a representative sample for the purposes of building our model.

Despite this, there is a significant derivatives market, chiefly comprising over the counter structures sold by the merchant banks to asset and other wealth managers. The banks need to hedge their exposures, and they do a significant amount of this in the exchange market. Furthermore, as is usual, the relevant models of the skew, which will be applied equally to over-the-counter products as well as exchange traded products, will be parameterized via exchange traded information. Thus it is necessary to have a robust model of the derivative skew for mark to market and hedging of positions.

Until April 2001 SAFEX calculated margin requirements on a flat volatility. At that time a skew was introduced into the mark to market and margining of exchange positions, although most players were aware of the skew and had (and have) models of the skew since significantly before that time.[2] The construction of the skew was initially supposed to be via an auction system, but has become merely a monthly poll, of moneyedness and corresponding addition or subtraction of volatility basis points, from the quoted at the money volatility. So here we note explicitly that bids and offers for the usual set of at and away from the money strikes simply do not exist in this market. Being merely a poll, the derivatives desks do not have to 'put their money where their mouth is' concerning their contribution to the poll, and so, although not grossly inaccurate, it is common knowledge that the exchange quoted skew cannot be used for trading, and by preference should not be used for mark to market (although many risk/back office/audit functions, in order to satisfy the requirement of sufficient independence from the front office, might do so). It is the scenario described here that led to the requirement from some major players in the South African market for an 'accurate' and 'objective' skew construction methodology, which the model described here is aimed at providing.

One point that needs to be noted here is that futures options are American and fully margined, that is, the buyer of options does not pay an outright premium for the option, but is subject to margin flow, being the difference in the mark to market values on a daily basis. It can be shown (see West, 2005, Chapter 10) that the appropriate option pricing formula in this setting is

$$V_C = FN(\mathrm{d}_1) - XN(\mathrm{d}_2) \tag{17}$$

$$V_P = XN(-\mathrm{d}_2) - FN(-\mathrm{d}_1) \tag{18}$$

$$\mathrm{d}_{1,2} = \frac{\ln\frac{\mathrm{F}}{\mathrm{X}} \pm \frac{1}{2}\sigma^2\tau}{\sigma\sqrt{\tau}} \tag{19}$$

$$\tau = T - t \tag{20}$$

It can be shown that it is sub-optimal to exercise either calls or puts early, and so one should not be surprised that the option pricing formula are 'Black like' even though the option is American. Furthermore, the fact that the options are fully margined has the attractive consequence that the risk-free rate does not appear in the pricing formulae. This is indeed fortunate as the South African yield curve itself is subject to a paucity of data compared to many markets, and hence may require some art in construction, which will typically be proprietary.

The Interpretation of Parameters

The β Parameter

As in Hagan *et al.* (2002), (16) shows that

$$\ln \sigma(F, F) = \ln \alpha - (1-\beta)\ln F + \cdots \tag{21}$$

and so the value of β is estimated from a log-log plot of $\sigma(F, F)$ and F. Some empirical analysis suggests that the value of β depends on time regimes: whether the contract is far, middle, or near to expiry. (We use these terms informally. By far, we mean about two years to expiry, near is perhaps six months or less, middle is in between.) The contract does not trade meaningfully until at least two years to expiry; although the underlying futures may trade, there will be little or no option activity. See Figure 1 where we see that the quality of a fitted regression line as in (21) would very much depend on a data selection criterion. This naturally suggests a time weighted regression, very much as for exponentially weighted moving average volatility calculations.[3] In Figure 2 we show the resultant evolution of the β value and the evolution of the correlation coefficient, which is also calculated using time weighting.

One feature we can note is that the value of β deteriorates towards 0 as the contract draws to expiry. This was a feature found to be common to all expiries.

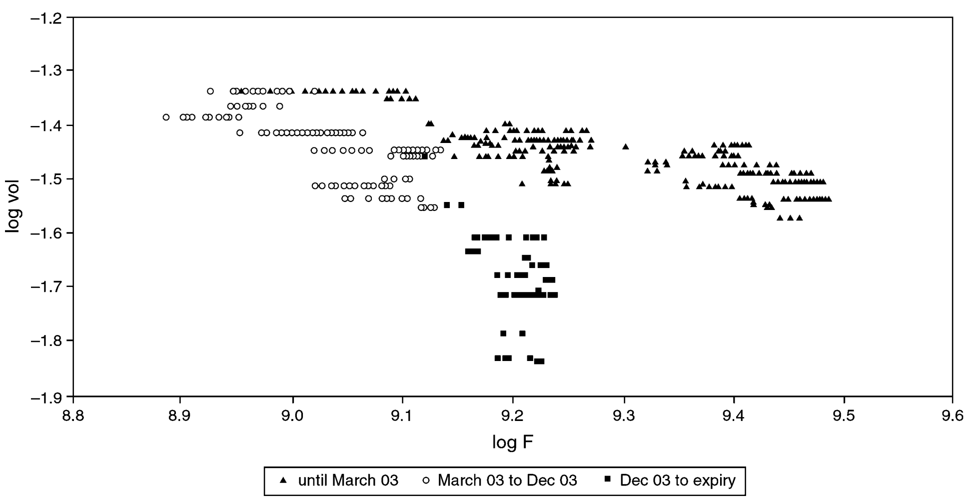

Figure 1. A log–log plot for the March 04 contract.

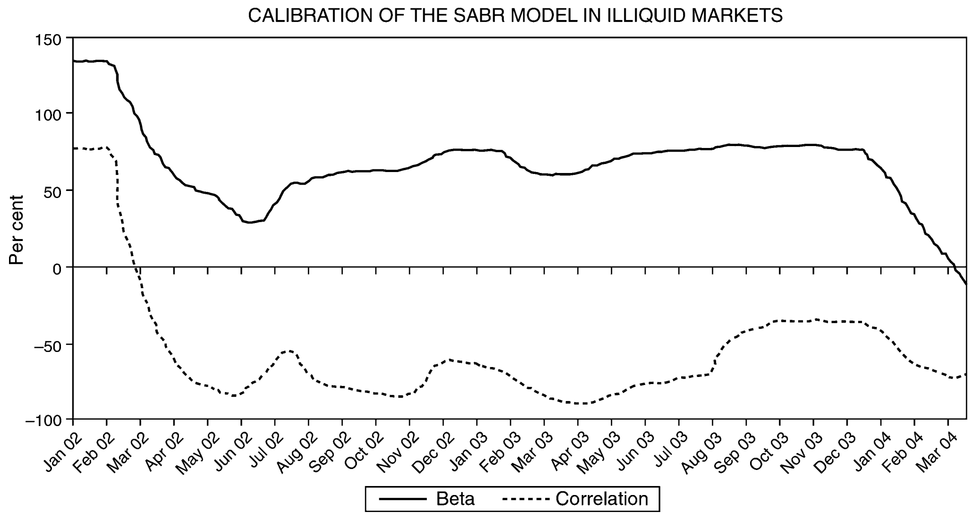

Figure 2. The evolution of the estimate of β for the March 04 contract.

An analysis of the March 05 contract will be included later, in a further analysis, where we will see that there is an argument – as in Hagan *et al.* (2002) – to simply choose a value of β, and stick with it for the entire life of the contract.

The α Parameter

This parameter is calibrated to the level of at-the-money volatility. There is perhaps a common perception amongst some market participants that it is the at-the-money volatility. However, what one rather does is retool the SABR model to have the at-the-money volatility as an input and that the correct value of α be calculated internally. Thus, as the at-the-money volatility has a term structure and changes frequently, so too does the value of α, albeit 'invisibly'.

To obtain α from σ_{atm}, we invert (16). Doing so, we easily see that α is a root of the cubic

$$\frac{(1-\beta)^2\tau}{24F^{2-2\beta}}\alpha^3+\frac{\rho\beta\nu\tau}{4F^{1-\beta}}\alpha^2+\left(1+\frac{2-3\rho^2}{24}\nu^2\tau\right)\alpha-\sigma_{\text{atm}}F^{1-\beta}=0 \qquad (22)$$

where we are assuming that we have already solved for ρ and ν. For typical parameter inputs, this cubic has only one real root, but it is perfectly possible for it to have three real roots, in which case we seek the smallest positive root.[4] One wants a rapid algorithm to find α to double precision, as then when in code finding the skew volatility for an option which is in fact at-the-money, one recovers the at-the-money volatility exactly. We use the Tartaglia method (as published by Cardano in the 16th century!) to find the desired real root. For this, we use the implementation and code in Press *et al.* (1992, Section 5.6). See Weisstein (1999–2004) for a synopsis of the history of these root-finding methods.

Having now reformulated the option skew with σ_{atm} and not α as input (in other words, α is not a constant), the skew volatility is in fact invariant under $\frac{X}{F}$, in other words $\sigma(X, F)=\sigma(X/F, 1)$. Thus we can perform calibration on relative strikes rather than absolute strikes. This is very convenient, as the trader should think in terms of relative strikes.

It should be remarked that in an illiquid market such as the one we are considering, even the at-the-money volatility can be a tenuous input. This is because of the mark-to-market mechanism employed by the exchange. Even if new options trade, if they are not near the money, the mark-to-market at-the-money volatility will not be altered. Therefore, the sophisticated model user may, on a day by day basis, wish to modify the at-the-money volatility input to this model, in order to attempt to infer – from the away from the money options traded – at what level at-the-money options would have traded if they had indeed done so. Of course, even this is subject to error, not only because it is outright speculative, but because the volatilities that were indeed dealt could be parts of packages, as we will see later.

Calibration to Existing Market Data

The calibration procedure is as follows: we fit β using the log–log plot. According to Hagan *et al.* (2002), it may be appropriate to fit this parameter in advance, and never change it. We will return to this point later.

The values of ρ and v need to be fitted. As already discussed, the value of F and the value of σ_{atm} are inputs, and given these and the values of ρ and v, α is no longer a required input parameter.

It is possible to simply specify a discrete skew (input by the dealer) and find the SABR model which best fits it. But we can be more ambitious, and ask ourselves to find the SABR model which best fits given traded data, independently of any dealer input as to the skew. Thus, we will not *a priori* have a discrete skew to which we calibrate the SABR model; rather, we seek that SABR model which provides a best fit to the traded data.

As already discussed, we fix in advance the value of β. Then, for any input pair (ρ, v), we determine an error expression $\mathrm{err}_{\rho,v}$, which is a measure of the distance from (optimally re-mapped) traded volatilities to the skew implied by these parameters. The trades that have been observed in the market may be weighted for age, for example, by using an exponential decay factor: the further in the past the trade is, the less contribution it makes to the optimisation.

Then, we seek the minimum of these error expressions $\mathrm{err}_{\rho,v}$ among all pairs (ρ, v), for which we use the Nelder–Mead simplex search. See Press *et al.* (1992, Section 10,4). Also see Spradlin (2003) which we have used as a guide for implementing this algorithm in two dimensions. The Nelder–Mead algorithm is a non-analytic search method that is very robust. Note that the error expression is essentially non-differentiable because it implicitly involves the root of a cubic (in other words, the differentiation involved would be horrendous). Thus for the second procedure we use a non-differentiable approach, for the first, all analytic procedures are available. In this application both optimisation procedures are extremely rapid.

The Nelder–Mead method needs to have error traps built in: that $-1 \leqslant \rho \leqslant 1$ and $v>0$. The method might 'stray' out of these bounds in the initial stages (of expansion, for example). The first condition is achieved by collaring ρ within $[-1, 1]$ while the second condition is ensured by bounding v below by 0.01, which suffices.

For any input pair (ρ, v), the mapping procedure for historical trades will be as follows.

Single Trades

A single trade for Q-many options made at date t_1 trades at $\sigma_{\text{tr}} := \sigma_{\text{tr}}(F(t_1), X, t_1)$. If the trade had been done on the skew, it would have been done at a volatility of

$$\sigma_{\text{mod}} := \sigma(F(t_1), X, t_1, \sigma_{\text{atm}}(t_1), \rho, v) \tag{23}$$

The contribution to the error term $\text{err}_{\rho,v}$ (modulo weighting for age) will be deemed to be

$$f = |Q|\mathcal{V}\left(\sigma^{\text{mod}}\right)[\sigma_{\text{tr}} - \sigma_{\text{mod}}]^2 \tag{24}$$

where the symbol $\mathcal{V}$ refers to the vega of the option with that strike.[5] This is a sensible modelling method as the importance of the volatility parameter fit is proportional to the vega of the option. Naturally, then, greater weight is given to near the money options. This is most suitable in markets where options, when dealt, are typically near the money.

Trade Sets

An issue which often arises, is that certain strategies (e.g. bull or bear spreads, butterflies, condors) trade for a pair, triple or quadruple of volatilities which may appear off market. In reality it is the price of the strategy that is trading and so a relevant set of volatilities may be found which is closer to the market than may at first appear.

To achieve this, we first determine the price of the strategy, implemented at time t. This is given by

$$P = \sum_{i=1}^{n} Q_i \eta_i \left[F(t)N\left(\eta_i \text{d}_1^{\text{i}}\right) - X_i N\left(\eta_i \text{d}_2^{\text{i}}\right)\right] \tag{25}$$

$$\text{d}_{1,2}^{\text{i}} = \frac{\ln \frac{\text{F(t)}}{\text{X}_\text{i}} \pm \frac{1}{2}\sigma(\text{t}, \text{X}_\text{i})^2 \tau}{\sigma(\text{t}, \text{X}_\text{i})\sqrt{\tau}} \tag{26}$$

$$\tau = T - t \tag{27}$$

where T is the expiry date of the options and time is measured in years, $\eta_i = \pm 1$ for a call/put, Q_i is the number of options traded at the *ith* strike as part of the strategy, and $\sigma(t, X_i)$ is the quoted volatility for the *ith* strike.

We would like to re-map this to the identical strategy, with the same price, but booked at different volatilities. These volatilities are found to be as close as possible to the volatilities on the skew curve. Thus, we would like to minimize

$$f\left(\sigma_1^{\text{fit}}, \sigma_2^{\text{fit}}, \ldots, \sigma_n^{\text{fit}}\right) = \sum_{i=1}^{n} |Q_i|\mathcal{V}\left(\sigma_i^{\text{mod}}\right)\left(\sigma_i^{\text{fit}} - \sigma_i^{\text{mod}}\right)^2 \tag{28}$$

where 'fit' denotes fitted volatilities and 'mod' denotes volatilities from the SABR model, subject to

$$P = g\left(\sigma_1^{\text{fit}}, \sigma_2^{\text{fit}}, \ldots, \sigma_n^{\text{fit}}\right) := \sum_{i=1}^{n} Q_i \eta_i \left[F(t) N\left(\eta_i \mathrm{d}_1^{\mathrm{i}}\right) - X N\left(\eta_i \mathrm{d}_2^{\mathrm{i}}\right)\right] \tag{29}$$

$$d_{1,2}^{i} = \frac{\ln \frac{F(t)}{X_i} \pm \frac{1}{2} \sigma_i^{\text{fit}} 2\tau}{\sigma_i^{\text{fit}} \sqrt{\tau}} \tag{30}$$

$$\tau = T - t \tag{31}$$

Once done, the value of f will contribute to $\mathrm{err}_{\rho,\nu}$. This minimization is done using the method of Lagrange multipliers. In order to minimize

$$f\left(\sigma_1^{\text{fit}}, \sigma_2^{\text{fit}}, \ldots, \sigma_n^{\text{fit}}\right)$$

subject to

$$g\left(\sigma_1^{\text{fit}}, \sigma_2^{\text{fit}}, \ldots, \sigma_n^{\text{fit}}\right) = P,$$

we solve the simultaneous set of equations

$$\nabla f = \lambda \nabla g \tag{32}$$

$$g\left(\sigma_1^{\text{fit}}, \sigma_2^{\text{fit}}, \ldots, \sigma_n^{\text{fit}}\right) = P \tag{33}$$

which easily simplifies to

$$2|Q_i| \mathcal{V}\left(\sigma_i^{\text{mod}}\right)\left(\sigma_i^{\text{fit}} - \sigma_i^{\text{mod}}\right) - \lambda Q_i \mathcal{V}\left(\sigma_i^{\text{fit}}\right) = 0 \ (1 \le i \le n) \tag{34}$$

$$g\left(\sigma_1^{\text{fit}}, \sigma_2^{\text{fit}}, \ldots, \sigma_n^{\text{fit}}\right) - P = 0 \tag{35}$$

Let $\mathcal{V}_2(\sigma_i) = \frac{\partial^2 V_i}{\partial \sigma_i^2}$ be the volga of the *ith* option. Note that

$$\mathcal{V}(\sigma_i) = F \sqrt{\tau} N'\left(\mathrm{d}_1^{\mathrm{i}}\right) \tag{36}$$

$$\mathcal{V}_2(\sigma_i) = F \sqrt{\tau} N'\left(\mathrm{d}_1^{\mathrm{i}}\right) \frac{\mathrm{d}_1^i \mathrm{d}_2^i}{\sigma_i} \tag{37}$$

For convenience, put

$$L_i = 2|Q_i| \mathcal{V}\left(\sigma_i^{\text{mod}}\right) \tag{38}$$

$$K_i = L_i \sigma_i^{\text{mod}} \tag{39}$$

We have a system of n+1 non-linear equations in (σ_1^{fit}, σ_2^{fit}, $\ldots$, σ_n^{fit}, and λ) and so we use the multi-dimensional Newton–Raphson method for this part of the problem. (See, for example, Press *et al.*, 1992, Section 9.6.) By the 'economic' nature of the problem it is fairly clear that the zero is unique and that pathology will not arise in the use of the Newton–Raphson method. Let $\underline{x}$=(σ_1, σ_2, …, σ_i, λ) be the unknown and required vector, where we have dropped the superscript 'fit'. The iteration is

$$\underline{x}_0 = \begin{bmatrix} \sigma_1^{\text{mod}} \\ \sigma_1^{\text{mod}} \\ \vdots \\ \sigma_n^{\text{mod}} \\ 0 \end{bmatrix} \tag{40}$$

$$\underline{x}_{m+1} = \underline{x}_m - J^{-1}\underline{F} \tag{41}$$

$$\underline{F} = \begin{bmatrix} L_1\sigma_1 - K_1 - \lambda Q_1 \mathcal{V}(\sigma_1) \\ L_2\sigma_2 - K_2 - \lambda Q_2 \mathcal{V}(\sigma_2) \\ \vdots \\ L_n\sigma_n - K_n - \lambda Q_n \mathcal{V}(\sigma_n) \\ g(\sigma_1, \sigma_2, \ldots, \sigma_n) - P \end{bmatrix} \tag{42}$$

$$[J]_{ij} = \begin{cases} L_i - \lambda Q_i \mathcal{V}_2(\sigma_i) & \text{if} \quad i=j\le n \\ 0 & \text{if} \quad i\ne j\le n \\ -Q_i\mathcal{V}(\sigma_i) & \text{if} \quad i\le n, j=n+1 \\ Q_j\mathcal{V}(\sigma_j) & \text{if} \quad i=n+1, j\le n \\ 0 & \text{if} \quad i=j=n+1 \end{cases} \tag{43}$$

Here J is the Jacobean, the matrix of partial derivatives: $[J]_{ij} = \frac{\partial F_i}{\partial x_j}$. The inverse of J is found via the LU decomposition (Press *et al.*, 1992, Section 2.3). Convergence is very rapid.[6] Thus, for any strategy booked, equivalent volatilities can be found which are most compatible with the SABR model selected.

The error for the input pair (ρ, v) is the sum $\text{err}_{\rho,v}$ of the above errors, possibly weighted for age. We then seek, amongst all ρ and v, the minimum of these expressions, using the two-dimensional Nelder–Mead algorithm.

As one can see in Figure 3 – this result is typical – the choice of parameters is fairly robust, with the minimum found at the bottom of a shallow valley. The skew one obtains is in Figure 4.

As pointed out in Hagan *et al.* (2002), the idea is that the parameter selection change infrequently (perhaps only once or twice a month) whereas the input values of F and σ_{atm} change as frequently as they are observed. This is in order to ensure hedge efficiency.

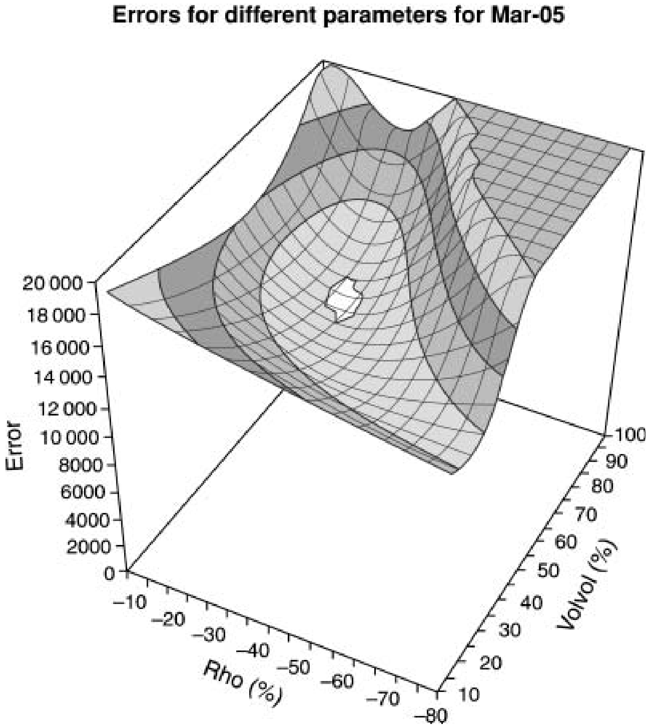

Figure 3. The error quantities for ρ and v.

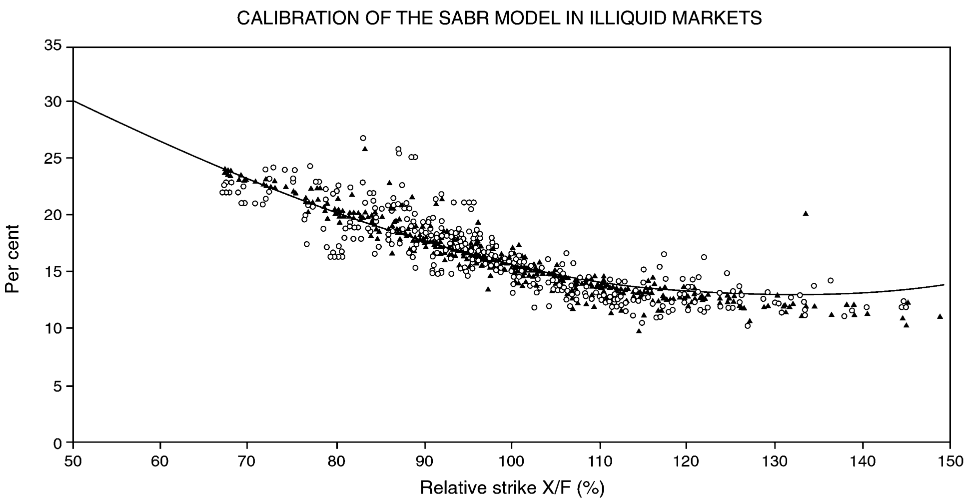

Figure 4. The SABR model for March 2005 expiry, with traded (quoted) volatilities (○ ○), and with strategies recalibrated to a fitted skew (▲▲▲), and the fitted skew itself (——). Historical trades have simply been shifted by the difference in the then and current at-the-money volatility; this is simply for graphical purposes. No ‘sticky’ rules are assumed in this analysis.

We now consider the evolution of parameters for the Mar 05 contract. Once again, the features that occur are typical. As has been mentioned, it can be argued that the use of these algorithms to find the best parameters should not be undertaken too frequently.

We performed two analyses: in the first instance, the finding of the β parameter every day, using the exponentially weighted regression methodology mentioned, and then finding the consequential ρ and ν. The evolution is shown in Figure 5. In the second instance, we fixed by economic considerations a value of $\beta=70\%$ throughout, and again found the consequential ρ and ν. The evolution is shown in Figure 6.

Interestingly, in the second instance, the parameters ρ and ν only change infrequently.[7] This again may be a feature that favours the choice of a single β which,

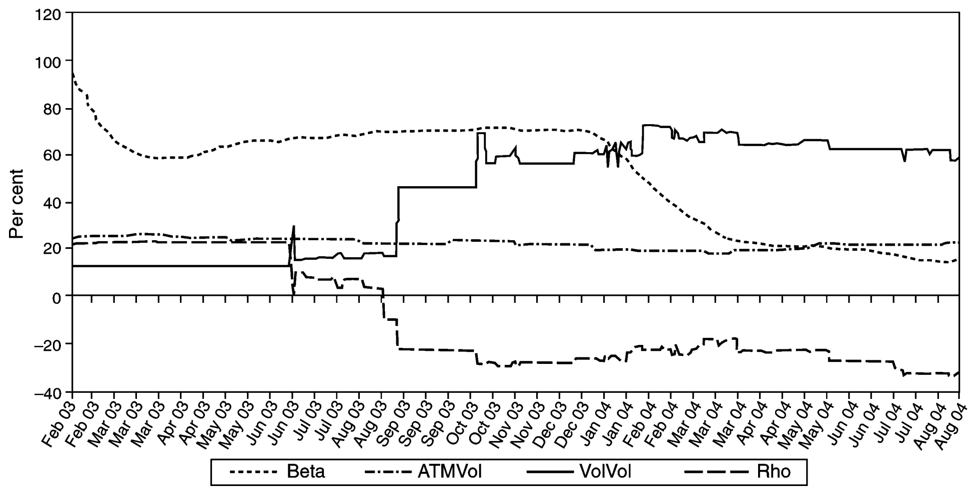

Figure 5. The SABR model for March 2005 expiry, with estimated β, and the estimated ρ and ν.

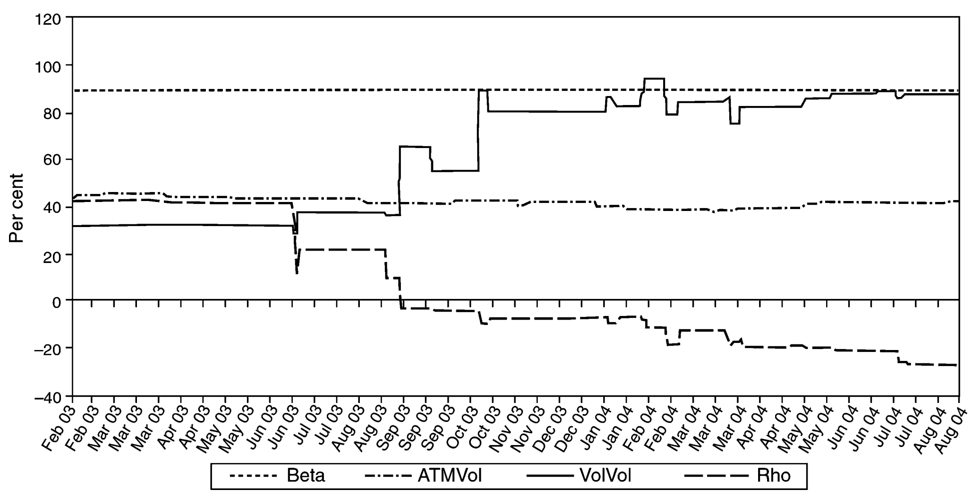

Figure 6. The SABR model for March 2005 expiry, with constant β, and the estimated ρ and ν.

in the absence of extraordinary events, remains constant throughout the life of the contract: parameters remaining unchanged implies, as seen in Hagan *et al*. (2002), lower hedging costs.

Acknowledgment

Thanks to Adam Myers of ABSA Bank, Leon Sanderson of Nedcor Bank, Arthur Phillips of RiskWorX, and Patrick Hagan of Bloomberg's.

Notes

[1] Some care needs to be taken with machine precision issues here. One can have that $z \approx 0$ and $\chi(z)=0$ to double precision. This needs to be trapped, and the limit result invoked, again putting $\frac{z}{\chi(z)}=1$.

[2] Although significantly, not all. Real Africa Durolink, a smaller bank, but major player in the equity derivatives market, failed within days of the introduction of the skew, as they were completely unprepared for the dramatic impact the new methodology would have on their margin requirements. See West (2005, Section 13.4).

[3] The only difference here is that we do not make the assumption of zero means, which we do when using returns to calculate volatilities. The implementation is elementary.

[4] When there are three real roots, they are of the order of -1000, 1 and +1000. So we take the root of order 1.

[5] Of course, some experimentation with the choice of the weight determined by the quantum is necessary. One could choose Q^2 for example, or indeed any positive weight. There will be no requirement for any smoothness of the weight in what follows.

[6] Note that the matrix will almost certainly not be of size greater than 5×5.

[7] Of course, to very high precision they are always changing. Here we mean that they are unchanged up to the (fairly high) precision that we chose in the Nelder–Mead algorithm.

References

Derman, E. and Kani, I. (1994) Riding on a smile, *Risk*, 7(2).

Derman, E. and Kani, I. (1998) Stochastic implied trees: arbitrage pricing with stochastic term and strike structure of volatility, *International Journal of Theory and Applications in Finance*, 1, pp. 61–110.

Derman, E., Kani, I. and Chriss, N. (1996) Implied trinomial trees of the volatility smile, *Journal of Derivatives*, 4(Summer).

Dupire, B. (1994) Pricing with a smile, *Risk*, 7(1).

Dupire, B. (1997) Pricing and hedging with smiles, in: M. Dempster & S. Pliska (Eds), *Mathematics of Derivative Securities*, pp. 103–111 (Cambridge: Cambridge University Press).

Fouque, J.-P., Papanicolaou, G. and Sircar, K. R. (2000) *Derivatives in Financial Markets with Stochastic Volatility* (Cambridge: Cambridge University Press).

Hagan, P. S., Kumar, D., Lesniewski, A. S. and Woodward, D. E. (2002) Managing smile risk, *WILMOTT Magazine*, September, pp. 84–108, *http://www.wilmott.com/pdfs/021118_smile.pdf

Heston, S. L. (1993) A closed form solution for options with stochastic volatility with applications to bond and currency options, *Review of Financial Studies*, 6(2).

Hull, J. and White, A. (1987) The pricing of options on assets with stochastic volatilities, *Journal of Finance*, 42, pp. 281–300.

Lewis, A. (2000) *Option Valuation under Stochastic Volatility: with Mathematica Code* (Finance Press).

Lipton, A. (2003) *Exotic Options: The Cutting-Edge Collection* (Risk Books).

Press, W. H., Teukolsky, S. A., Vetterling, W. T. and Flannery, B. P. (1992) *Numerical Recipes in Fortran 77: The Art of Scientific Computing*, 2nd edn (Cambridge University Press).

SAFEX, (2004) *http://www.safex.co.za

Spradlin, G. (2003) The Nelder-Mead method in two dimensions, *http://www.adeptscience.co.uk/maplearticles/f1198.html

Weisstein, E. W. (1999–2004) Cubic formula, *From MathWorld–A Wolfram Web Resource*, *http://mathworld.wolfram.com/CubicFormula.html

West, G. (2005) The Mathematics of South African Financial Markets and Instruments, Lecture notes, Honours in Mathematics of Finance, University of the Witwatersrand, Johannesburg *http://www.cam.wits.ac.za/mfinance/graeme/safm.pdf

Appendix: Another Calibration Method

Here is an alternative calibration method: for every single trade or strategy, calculate the premium that the trade or strategy was done at, as before. Now, for any given input pair (ρ, v), the contribution of that trade or strategy to $\mathrm{err}_{\rho,v}$, is simply the absolute value of the difference between this true premium and the premium at which the trade or strategy would have been done if it had been done on the SABR skew with that ρ and v, possibly weighted for age.

The best choice (ρ, v) is found using the Nelder–Mead algorithm, as before.

The Newton–Raphson algorithm does not feature here per se. The only place that it will enter the analysis is if we wish to perform an analysis of the type as in Figure 4: all strategies are 'rebooked' at the equivalent volatilities which are closest to the model which has already been derived.