An arbitrage-free method for smile extrapolation

Shalom Benaim, Matthew Dodgson and Dherminder Kainth
Royal Bank of Scotland

A robust method for pricing options at strikes where there is not an observed price is a vital tool for the pricing, hedging, and risk management of derivatives. All institutions that trade derivatives will have an approach to this task. Typical examples might be a simple interpolation scheme across implied volatilities, or the use of a model-based formula optimized to fit observed prices. It is our view that while these methods work well for *interpolation* across actively traded strikes, they often break down when used for *extrapolation*. We introduce in this paper a technique for smile extrapolation that is robust, simple, fast and offers control on the form of the tails in the distribution. Using this method allows distribution-sensitive products such as CMS rates or inverse-FX options to be priced consistently with the smile of traded vanilla options. The resulting arbitrage-free distributions are also key to the copula-based pricing of multi-asset products such as spread options, quantos and hybrids. Our approach fixes several problems currently seen in today's stressed markets, e.g. with CMS rates.

There are two distinct areas where smile extrapolations can be a problem: when an institution has legacy trades on its books that are struck far from today's at-the-money strike (a common situation in the current market) and in the pricing of more exotic products that depend on the use of an entire risk-neutral probability density function (PDF) of the underlying. Examples of the second area include:

- Pricing derivatives with non-standard payoffs by replication with a portfolio of vanilla options over all strikes. Calculating the convexity correction for constant maturity swap (CMS) rates is a particularly topical case.
- Pricing options that depend on multiple underlyings, such as hybrids, using copulas. This method requires the entire marginal cumulative distributions (CDFs).
- Generating the local volatility surface from a set of observed prices. This requires inter- and extrapolation of prices both in the strike dimension (considered in this article) as well as in the time dimension.

The most important feature of a smile extrapolation method is that it should deliver arbitrage-free prices for the vanilla options, i.e., the option prices must be convex functions of strike, and stay within certain bounds. In addition, the extrapolation method should ideally have the following properties:

1. It should reprice all *observed* vanilla options correctly.
2. The PDF, CDF and vanilla option prices should be easy to compute.
3. The method should not generate unrealistically fat tails, and if possible, it should allow us to control how fat the tails are.
4. It should be robust and flexible enough to use with a wide variety of different implied volatility surfaces.
5. It should be easy and fast to initialize for a given smile.

The above points are not so straightforward to satisfy. Working directly with the PDF or the CDF makes it difficult to satisfy the first property on the list above as the conditions that we impose (on option prices) are on the integral of the CDF. Therefore the traditional method concentrates on extrapolating the price; it is common practice to do this via the extrapolation of implied volatility. We can name two such commonly used methods which do not satisfy our above wish list: The first is to use a simple interpolation within the region of observed prices, and just set the implied volatility to be a constant outside of this region. This method is flawed as it introduces unstable behaviour at the boundary between the smile and the flat volatility, and we will have unrealistically narrow tails at extreme strikes. The second is to fit a parametric form for the implied volatility derived from a model, such as the SABR formula[1] (see [5]). There are two problems with this method: it gives us little control over the distribution; indeed this approach often leads to excessively fat tails. This can lead to risk neutral distributions that have unrealistically large probabilities of extreme movements, and have moment explosions that lead to infinite prices, even for simple products like Libor in arrears and FX quantos (see e.g., [1]). Furthermore, in the case of the SABR formula, which is the result of an asymptotic expansion about a simple stochastic volatility model, the expansion becomes less accurate at strikes away from the money, and often leads to concave option prices, or equivalently negative PDFs, even at modestly low strikes.[2]

We introduce a method that allows us to control the extrapolation of option prices outside of a core region of market observability. A crucial point of our method is the use of calls/puts as the basis of our extrapolation. This allows us to control the convexity so as to keep the set of prices over all strikes arbitrage free, while retaining some control on the fatness of the tail at extreme strikes. In the rest of this article we describe our method in more detail, and give examples where the method is applied to a range of different pricing problems.

The tails as parametrised option prices

It is well known that, given the prices of options on a particular random variable (such as a stock price at a given maturity) for all strikes, we can deduce the probability distribution of this random variable by differentiating the option prices. In particular, the cumulative distribution function and the density are respectively the first and second derivatives of (undiscounted) put prices with respect to strike.

As mentioned above we do not, in practice, have this information. Only options for a finite set of strikes are traded and therefore have observable prices. However, if we can interpolate and extrapolate the known *prices* to the entire positive real line (for a positive random variable like a stock) we can use this to derive the full distribution. By construction, the probability distribution derived in this way will match all our observed market prices. The interpolation of option prices is relatively straightforward. For example, we could use a cubic spline, either on call prices or implied volatilities, but this leads to problems as the option prices computed in this way will not necessarily be convex in strike, leading to arbitrage and negative densities. A better approach is a convexity-preserving spline. Alternatively, we can fit a parametric form to the implied volatility smile. Popular choices for the parametric form are the SABR formula or the SVI parametric form[3] (see[4]).

Extrapolation outside the range of observable prices, however, is more difficult. We may be tempted to simply use the same functional form used for interpolation. However, this is problematic, since there is no guarantee that this functional form will lead to arbitrage free prices for very large and small strikes. The SABR formula discussed above, for example, is based on an asymptotic expansion that breaks down for extreme strikes, violating Lee's bounds for the limiting implied volatility [6]. This means that the risk neutral distribution implied by the SABR formula will not be arbitrage free and will have negative densities. Furthermore, blindly using the functional form used for interpolation can lead to very fat tails that imply, for example, that most of the value of even a simple vanilla product comes from extremely large movements in spot.

It is for these reasons that we propose to separate the interpolation and extrapolation methods. Briefly, the

[1] This is an asymptotic expansion for the implied volatility in a particular stochastic volatility model. The justification for using the SABR formula to describe smiles is that it is derived from an arbitrage-free option-pricing model, so the interpolated option prices should also be arbitrage free. However, this is not the case when the asymptotic expansion loses accuracy. The SABR formula is now widely used by market participants, despite the problems that can occur because of the breakdown of the expansion far enough away from the money.

[2] One might argue that the use of a better description of the SABR model, or indeed another model, might solve these problems. Even so, the tail methodology that we introduce here is a more general approach, that allows for a greater control of extrapolations.

[3] 'Stochastic Volatility Inspired' i.e. based on the behaviour of the smile in the Heston stochastic volatility model.

method works as follows: We define a core region of observability, where we use any standard smile interpolation method. Outside of this region we extrapolate by using a simple analytic formula for the option prices. In the low strike extrapolation we use a formula for put prices that will vanish at zero strike, and that remains convex. In the high strike extrapolation we use a formula for call prices that will approach zero at very large strikes, and that remains convex. We parametrize each of these formulas so that we can match the option price as well as its first two derivatives at the corresponding boundary with the core region. We are also able to retain a measure of control over the form of the tails at extreme strikes. A further consequence of our method is that the first moment of the distribution will be exactly equal to the forward value implied by put-call parity.[4]

Constructing the tails

We will assume that we can interpolate observed prices so that we have a parametrisation for the prices over a finite interval $[K^-, K^+]$ with $0 < K^- < K^+ < \infty$. In practice the choice of K^- and K^+ may come from specific knowledge of the range of market quoted prices, or from a more general system that defines the range where the core region will lie, e.g. with the use of Black-Scholes deltas. The interpolation in the core region can be done using the SABR formula as described above, or otherwise. We now need to extrapolate these prices so that we have option prices for all positive strikes. It is more convenient to use put prices for the range $(0, K^-)$ and call prices for the range (K^+, ∞). Extrapolation is by its nature relatively arbitrary; clearly there is not a unique extrapolating function. However, our methodology depends on the following features:

- We require a functional form for our extrapolation, which is continuous, twice differentiable, and tends to 0 as strike tends to 0 or infinity.
- The functional form should have at least three parameters that we can choose to ensure that prices are twice continuously differentiable at the boundaries.[5] Ideally the boundary conditions will lead to linear equations in these parameters.

As an example, we have found the following functional form to be useful for the extrapolation of put prices for a variety of underlyings. Let

$$P(K) = K^{\mu} \exp(a + bK + cK^2). \tag{1}$$

We fix $\mu > 1$, which ensures that the price is zero at zero strike, and there is no probability for the underlying to be zero at maturity.[6] Alternatively, we can choose μ to reflect our view of the fatness of the tail of the risk neutral distribution. It is easy to check that this extrapolation generates a distribution where the m-th negative moment is finite for $m < 1 - \mu$ and infinite for $m > 1 - \mu$. While there is no general reason for a negative moment to exist, there are cases such as the inverse FX option described below where we do expect a certain negative moment to remain finite.

An advantage of this form of tail is that the condition for matching the price and its first two derivatives at K_- is a set of *linear* equations in the parameters a, b and c. We can then solve rather than optimize for a, b, and c to give an exact match at the boundary. The setting up of the tail is therefore fast and stable. We should be clear at this stage that the tail form (1) is *not guaranteed* always to be convex, and thus arbitrage free. In extreme cases we sometimes find a region in the tail with negative density. We therefore note the following points:

- The convexity of the tails always needs to be checked.
- If extrapolating the SABR formula, it may be legitimate to move K_- closer to the forward to recover an arbitrage-free tail.

[4]One can easilty prove that the distribution is correctly normalized, $\int p(x)dx = 1$. Less obviously, one can also show that we have $\int xp(x)dx = F$ where $F = [C(K) - P(K)]/D - K$ is the forward satisfying put-call parity (D is the discount factor).

[5]This is so that the CDF, $D^{-1}dP/dK$, and the PDF, $D^{-1}d^2P/dK^2$, are both continuous. Strictly speaking, we only need the option price to be continuously differentiable, as the density may have steps and still be arbitrage free.

[6]A "natural" choice of the exponent is $\mu = d\ln P/d\ln K|_{K=K^-}$. This makes sense from a theoretical point of view as this derivative should converge to μ as strike goes to zero.

- Alternatively, in the rare cases where the tail extrapolation has a negative density we can simply change the functional form, either by choosing a different value of μ in (1), or by exploring other forms. For example, the problem may be solved by choosing a tail form where the CDF approaches a finite value as $K \to 0$, which corresponds to a delta function in the density at zero, this is described below in the example on a defaultable stock.
- We have not found an example that we couldn't fix.

For the extrapolation of call prices for large strikes, we might use the functional form

$$C(K) = K^{-\nu} \exp(a + b/K + c/K^2). \tag{2}$$

We fix $\nu > 0$ to ensure that the call price approaches zero at large enough strikes. Our choice of ν controls the fatness of the tail; the m-th moment will be finite if $m < \nu - 1$ and infinite if $m > \nu - 1$.

Because these functional forms are relatively simple, we can differentiate them analytically and so we can efficiently compute the distribution and density functions. We can of course use other functional forms, with lighter tails, say, instead of the ones above, using the same ideas.[7]

In Figures 1–3 we illustrate our extrapolation method at low strikes. We take our example from the rates markets, where the use of SABR formulas for smiles is ubiquitous. Figure 1 shows the smile for a 10-year expiry EUR caplet with a forward of 5%. The smile is characterized by a SABR formula, with different extrapolations below $K_- = 3.5\%$; we plot the original SABR extrapolation, as well as three examples of formula (1) with $\mu = 1.5$, 2.5 and 4.5. Note the large range of implied volatilities that can result from our choice of tail exponent. This should be an explicit input in a pricing system.

Figure 2 shows that the SABR extrapolation clearly leads to an aribtrageable smile: the negative slope in the CDF below 0.4% means that option prices are concave at these low strikes. The three cases that use (1) are arbitrage-free extrapolations. This figure also illustrates a constraint on the form of the tails: the integral of the CDF from zero to K_- is fixed by the put price at this strike. Figure 3 shows the negative density at low strikes for the SABR formula. For the three tail extrapolations we always have a "bimodal" distribution, which is related to the skew in the original smile.[8] Whether or not the density goes to zero at $K \to 0$ or diverges in this limit depends on the value of μ. However, our earlier requirement that $\mu > 1$ ensures that integration over the density will be finite.

Example pricing problems

In the rest of this paper we use the above described methods in important examples over a range of asset classes.

The low-strike digital option

For our first application, consider Figure 2. This figure represents the (undiscounted) price of a digital floorlet, that pays a notional amount only if the Euribor fixes below a certain strike. If we were to use the SABR formula (the black curve) to price such a digital option, we might end up buying a digital floorlet at a strike $K = 0.25\%$ and selling a digital floorlet at a lower price with a strike $K = 0.5\%$. The net payoff for us will be always negative (i.e., we pay) or zero, but we have paid a positive amount for the privilege. We can remove this channel for throwing away money by using a controlled tail, as shown in the same figure.

While the digital gives a very graphic example of how the use of SABR formulas can lead to arbitrageable prices, the same is true for plain vanilla options at low strikes. When there is a region of concavity in the vanilla option prices as a function of strike, it is easy to construct portfolios of options with payoffs that are positive or zero in all future scenarios, but which will have a negative value today using this smile. Again, using a controlled tail extrapolation removes this arbitrage.

---

[7]Other choices of tails include $P(K) = aK^{\mu}/(K-b)^c$ and $C(K) = e^{a-bK+c/K^{\nu}}$. The crucial point is that there should be simple solutions to matching at the boundary; this can be ensured with a quasilinear form in the parameters.

[8]The bimodal distribution with a second peak in the density at low strikes seems to be neccessary to fit the market smile in an arbitrage-free manner. Note that the CEV model that underlies the SABR model has this bimodality for all values of β between 0 and 1.

The Inverse FX option

Consider an FX market where options are actively traded on the rate X between a foreign and a domestic currency. Say we want to price an inverse FX option, which is defined by the payoff in the domestic curency of $(X_T^{-1} - K^{-1})_+$ at an expiry T. A possible pricing method would be to replicate this payoff using standard option payoffs over different strikes. This is equivalent to integrating the payoff multiplied by the risk-neutral probability density of the FX rate.

Clearly the relevant integral will be dominated by low strikes, and so the smile extrapolation used is crucial to the pricing. In fact, the use of SABR formulas is extremely dangerous here: negative densities at low strikes can give completely nonsensical results. Even when the densities remain positive, the form of the density $p(x)$ can be such that the integral of $p(x)/x$ diverges as the lower integration limit approaches zero.

We have priced inverse FX options with our tail methodology. We take the USDJPY market, and characterize the smiles over a core region with SABR formulas. We can then take different values of the lower tail exponent μ, for which we show the smiles in Figure 4. Note that, while these smiles are identical down to a strike of 50, which covers 96% of the distribution, there is a significant range of prices for an inverse FX option struck at the forward, see Table 1. Note also that the SABR formula cannot be used to price this product as it has negative densities at small strikes. The important point is that we are now able to tune the lower tail, to which the inverse option is so sensitive, with the choice of μ, thereby giving control of the pricing to the trader.

This example product can be described as a regular FX option in the foreign market, quantoed into the domestic currency. If we want to generalize to FX options between two different foreign currencies, quantoed into the domestic currency, we can do this in a smile-consistent manner with the use of a copula on two risk-neutral distributions (because the cross FX can be written in terms of the two other FX rates, $X_{23} = X_{21}/X_{31}$). This is another case where our tail methodology is extremely useful. We do not describe this in detail here, as it is rather similar to the calculation of spread option prices using a copula, which is described below.

CMS products

Many products in the interest-rate derivative market depend on the value of a swap rate at a future time. For instance, a CMS swap will exchange a series of payments of a swap rate corresponding to a fixed length of swap for a series of floating-rate payments plus a spread. There is also an active market in CMS options. The forward value of a CMS payment (or an option) should be calculated using the market smiles for vanilla swaptions (which are far more liquid than CMS products). This can be shown from a replication argument, or equivalently, through a measure change from the swap measure. The difference between the forward CMS rate and the present forward value of the relevant swap rate is the CMS convexity correction. It can be shown that the convexity correction is related to the second moment of the swap rate in the swap measure[7].

The use of SABR formulas to calculate CMS rates is extremely problematic. The problem is that for large enough expiries *the second moment in the SABR-formula distribution appears to diverge*. Any implementation of CMS through replication using SABR will give a result that strongly depends on the upper strike limit used for replication. On the other hand, using a high-strike extrapolation as in (2) allows us to control the value of the CMS convexity correction through the choice of the exponent ν. In Table 2 we show the results for the forward CMS rate for different choices of the exponent ν. We consider a 10-year EUR swap rate fixing in 20 years. Note that there is no applicable SABR price, as at this expiry the integral is not converging. As long as $\nu > 1$ we will get a converging integral using the tail in (2), and tuning ν gives different CMS rates over a range of about 40 bps. Where there are observed prices for CMS, the value of ν can be chosen to best match them.

Spread options

Another common exotic rates derivative is the CMS spread option. A structured note might pay a coupon proportional to $(S_1 - S_2 - K)_+$, where S_1 and S_2 are swap rates of different tenors. This allows the buyer of such a note to bet on the future shape of the yield curve. Such products can be priced by constructing a bivariate risk-neutral distribtion for the two rates using the marginal distributions extracted from the swaption smiles connected by some choice of copula, as described in [2].

Unfortunately, it is almost impossible to follow this methododology with the use of SABR formulas: the copula will need well-defined monotonically increasing marginal distribution functions. On the other hand, we have successfully implemented CMS spread option pricing with our tail extrapolations (for reasons of space, we do not show results here). The pricing will strongly depend on the parameter choices for the copula, e.g., correlation in a Gaussian copula.

A defaultable stock

To illustrate the flexibility of our method, we now consider the problem of characterizing the smile for options on a stock, where the market sees a high probability of default before the option expiry. We can extend the lower-strike extrapolation (1) to allow for a finite probability q for the stock to have zero value, using the form:

$$P(K) = qK + K^{\mu} \exp(a + bK + cK^2). \tag{3}$$

The value of q could be determined by looking at the survival probabilities implied by the CDS market.

Another generalization could be a situation where the market did not believe the stock could be at any intermediate value between zero and K^* (e.g., if the immediate demise of the company is expected if the stock falls below K^*). The implications of assuming such a "default corridor" have been discussed in [3]. In this case we could match this expectation with the form

$$P(K) = qK + \tilde{K}^{\mu} \exp(a + b\tilde{K} + c\tilde{K}^2), \tag{4}$$

with $\tilde{K} = (K - K^*)_+$. In Figures 5 and 6 we show an example using this form. The example shows the distribution implied from 9-month expiry options on RBS in November 2008, where our extrapolation allows for a 7% probability that the stock will be worthless before the option expiry, along with a vanishing probability for the stock price to lie between 0 and 50. Therefore the CDF (Figure 5) is exactly 7% over the strike range $K < 50$, while the PDF (Figure 6) is zero in this range and has a δ-function spike at zero.[9] We believe that this generalized form of extrapolation is extremely relevant in the current market.

Local volatility

Finally we mention that our tail methodology is a useful tool when pricing path-dependent derivatives with a local volatility model. A complete description of the local volatility function requires the specification of the implied volatility surface at all strikes. Therefore, extrapolation beyond quoted option prices is crucial. Our extrapolation method enables the construction of a consistent local volatility surface over the entire range of strikes.

Conclusions

In this article we have introduced a simple, fast and robust methodology for the extrapolation of option prices to strikes outside of a core region of market observability. The main attraction of the method is that it ensures an arbitrage-free set of prices, while allowing a measure of control on the asymptotic behaviour of the risk-neutral distributions at extreme strikes. We have seen how we can tune the price of an exotic instrument such as the inverse FX option, or a CMS rate, by tuning a parameter such as the limiting power-law exponent used here. This should allow for the consistent pricing and risk management of such products. The method is general enough to be applied to many asset classes, as shown in the range of examples we have presented. By choosing different tail forms, the method is flexible enough to recover the different distributions expected for different underlyings.

References

[1] **Andersen L and Piterbarg V, 2006**

[9] The reason for the "wobbles" in Figure 6 in the core region is that a convexity preserving spline method was used to interpolate the option prices in the region of market obervability.

Moment explosions in stochastic volatility models
Finance and Stochastics 11 (1), pages 1432-1122

[2] **Berrahoui M, 2004**
Pricing CMS Spread Options and Digital CMS Spread Options with Smile
Wilmott magazine May 2004. pages 63–69

[3] **Carr P and Wu L, 2008**
A simple robust link between American puts and credit protection
Global Derivatives Trading and Risk Management Conference, Paris.

[4] **Gatheral J, 2006**
The Volatility Surface, A Practitioner's guide
John Wiley

[5] **Hagan P, Kumar D, Lesniewski A and Woodward D, 2002**
Managing Smile Risk
Wilmott magazine 3, pages 84–108

[6] **Lee R, 2004**
The Moment Formula for Implied Volatility at Extreme Strikes
Mathematical Finance 14 (3), pages 469–480

[7] **Pelsser A, 2003**
Mathematical foundation of convexity correction
Quantitative Finance 3, pages 59–65.

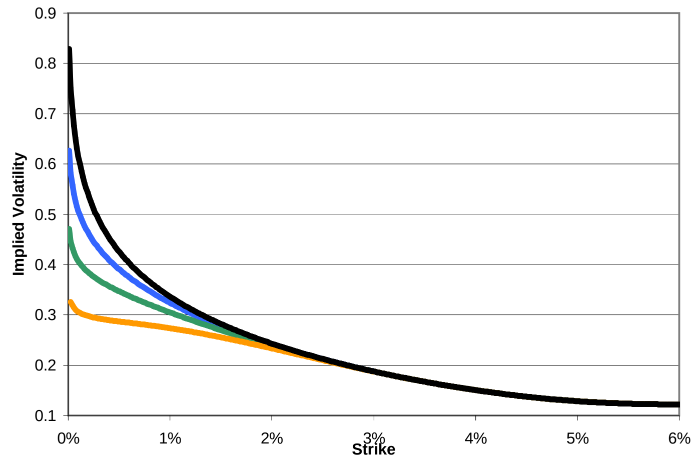

Figure 1: Caplet smile (EUR, 10 year expiry, forward=5%) from the SABR formula (black), and with our extrapolations at low strikes with $\nu = 1.5$ (blue), 2.5 (green), and 4.5 (orange).

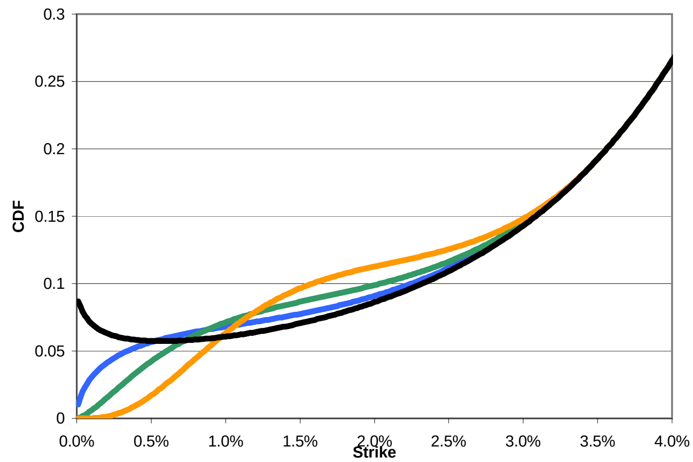

Figure 2: Implied cumulative distribution of forward rate (EUR, 10 year expiry, forward=5%) from the SABR formula (black), and with our extrapolations at low strikes with $\nu = 1.5$ (blue), 2.5 (green), and 4.5 (orange).

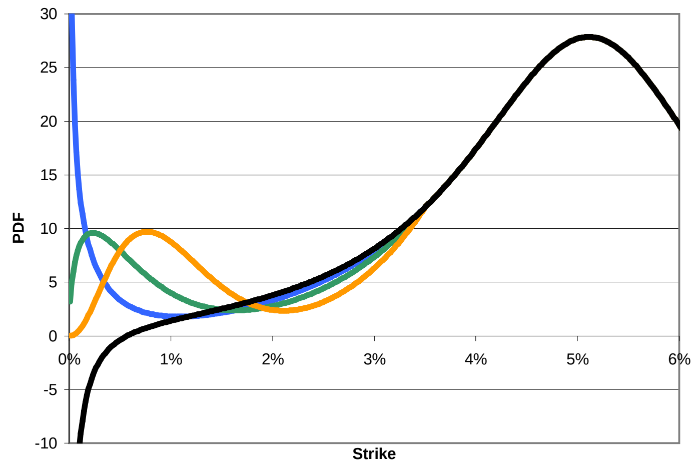

Figure 3: Implied probability density of forward rate (EUR, 10 year expiry, forward=5%) from the SABR formula (black), and with our extrapolations at low strikes with $\nu = 1.5$ (blue), 2.5 (green), and 4.5 (orange).

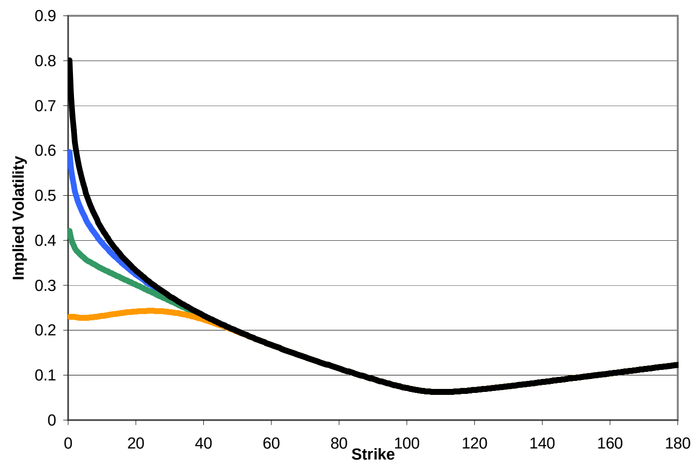

Figure 4: FX smile (USDJPY, expiry=5yr, Forward=97.4) from the SABR formula (black) and our extrapolations with $\nu = 2.33$ (blue), 5 (green), and 15 (orange). with the SABR formula .

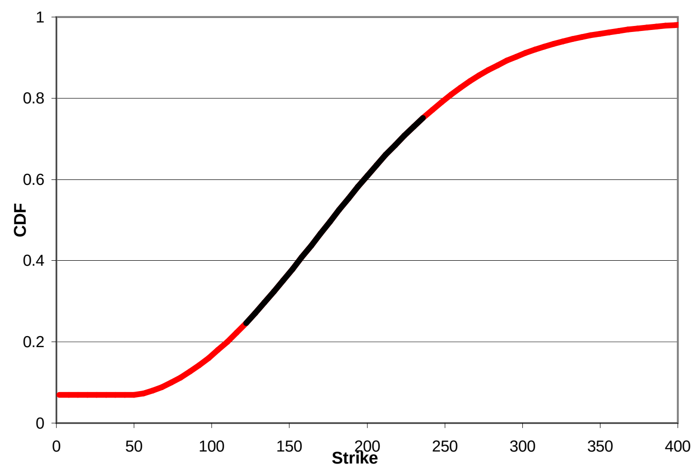

Figure 5: Implied cumulative distribution from options on a defaultable equity (black) with extrapolations (red) that assume a 7% probability of default before expiry, and zero probability for stock to lie below 50.

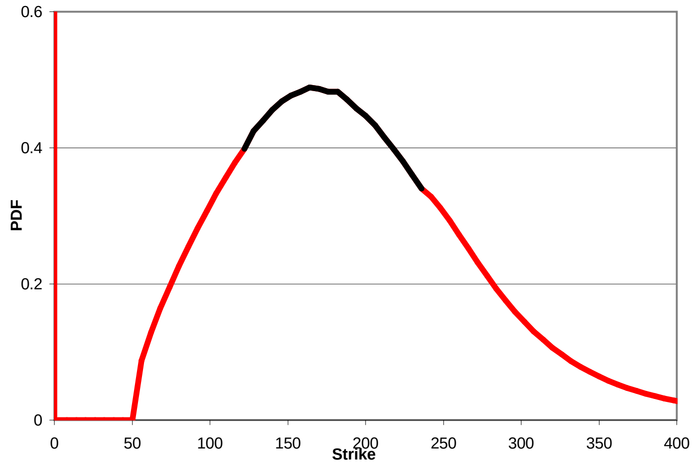

Figure 6: Implied probability density from options on a defaultable equity (black) with extrapolations (red) that assume a 7% probability of default before expiry, and zero probability for stock to lie below 50.

Extrapolation	Price[yen]
SABR	N/A
μ=2.33	4424
μ=3	2446
μ=4	1900
μ=5	1698
μ=15	1306

Table 1: Inverse FX option price on USDJPY, K=Forward, N=1 million yen.

Extrapolation	CMS rate
SABR	N/A
ν=1.25	3.44%
ν=2	3.35%
ν=3	3.24%
ν=5	3.13%
ν=10	3.10%

Table 2: CMS rate for 20yr expiry, 10yr tenor, underlying swap rate = 2.71%, CCY=EUR.