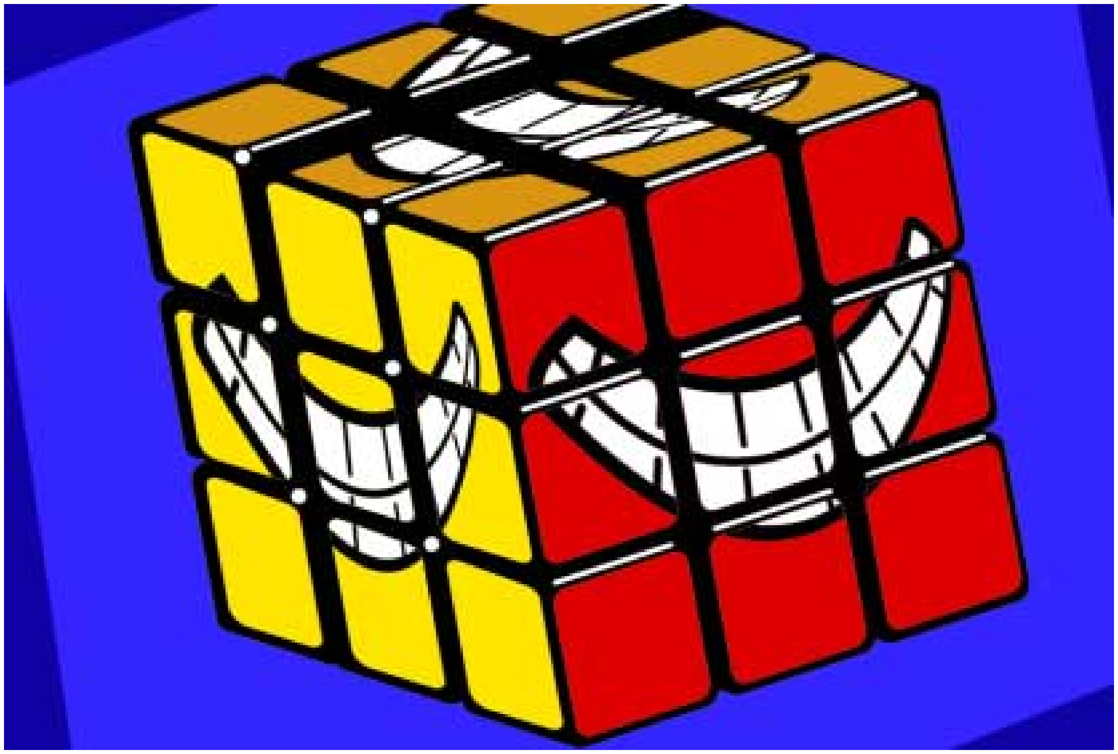
Elie Ayache

Can Anyone Solve The Smile Problem? A Post-Scriptum

A return to issues raised by an article published in January last year



The following article is a transcription of one of my posts (to this day, the longest) on the Wilmott forums. The thread it appeared in, "Of smile and models," is one of the multiple recent threads where worries and questions are starting to emerge, relative to the smile dynamics. The big worry is: "It is one thing to calibrate my smile model to the vanillas right, it is another to get the smile dynamics right, and the prices of path-dependent options right." As this big question was the question that my co-authors and myself had tried to address in our article "Can anyone solve the smile problem?" and the thread was a discussion of issues raised by the article, I thought I would take advantage of my column today and re-edit my post as a post-scriptum to the article.

What is the right smile dynamics?

There is no theoretically right or wrong smile dynamics. The only "right" smile dynamics is the smile dynamics as implied by the market.

(Of course, I am here siding with the view that the market is always right. I will not be debating the issue of whether you should believe your model rather than the market, then trade the model's predictions against the market. I agree, the market cannot always be right. Models are just tools to help you connect the value of some parameters describing the stochastic process with the values of derivative instruments that you cannot price with mental arithmetic, so that you can make bets on the value of the parameters by executing the corresponding trades in the corresponding instruments. Where you should draw the boundary between a) calibrating your model to the market then using it to price some other stuff relative to the initial stuff, and b) imposing the parameters of the model yourself and taking "absolute" positions in the derivative instruments, is up to you and your experience.)

In other words, smile dynamics is a "price," just as implied volatility is a price, or the implied yield curve is a price, or the implied credit spread curve is a price. And you have to handle it the same way you usually handle a price that the market offers you. You bet against it, or you use it in a strategy to help determine the price of other things. And what market prices can help us imply the smile dynamics? (In other words, what current spot market information can help us infer the market prediction of the future evolution of the smile?) Answer: the spot market prices of the barrier options, or the spot prices of the forward starting options, or similar path-dependent objects.

Why not the vanillas?

As explained in the article "Can anyone solve the smile problem?" the spot market prices of the vanillas, even a continuum of such vanillas $C(K, T)$, just map the probabilities of traveling from current spot and current time to the point located at (K, T). Let us call this transition probability $P(S_0, t_0, K, T)$. The vanillas give you no information about the path. In other words, you have no idea about forward probabilities such as $P(S, t, K, T)$. As you observe the world from (S_0, t_0), you have no idea what the probability of ending up at (K, T) may be, conditionally on finding yourself sitting at spot S, at a future time t (which is prior to T, of course). As the probabilities of ending up at (K, T) starting your trip from (S, t) map into the vanilla options prices, all we are saying is that, given the spot prices of the vanillas $C(S_0, t_0, K, T)$, you have no idea what their future prices $C(S, t, K, T)$ will be. One more time: Given the spot implied volatility smile, you have no idea what the future implied volatility smile will be. You have no idea of the smile dynamics. The only constraint the vanilla prices $C(S_0, t_0, K, T)$, or equivalently the spot transition probabilities $P(S_0, t_0, K, T)$, impose on the future transition probabilities $P(S, t, K, T)$ is the law of compound probabilities. The probability of moving from (t_0, S_0) to (t_1, S_1), and the probability of moving from (t_1, S_1) to (t_2, S_2), should be linked to the probability of moving straight from (t_0, S_0) to (t_2, S_2). But there is complete underdetermination of the second, given the first and the third. Two models can perfectly agree on both $P(S_0, t_0, S_1, t_1)$ and $P(S_0, t_0, S_2, t_2)$ and disagree on $P(S_1, t_1, S_2, t_2)$. For instance, the underlying can suddenly jump in between the future dates t_1 and t_2 and hit a barrier, in one of the two models, and not jump in the other. Barrier option prices will then be different.

Sounds shocking?

Is not the purpose of option pricing models and option pricing theory precisely to give us the evolution of prices of options at future times and different states of the world? Indeed so. But this is precisely where model-dependency kicks in. While they can be made to agree on $C(S_0, t_0, K, T)$, or the spot implied volatility surface, different models of stochastic behavior of the underlying, can very well disagree on $C(S, t, K, T)$, or smile dynamics. So we might as well elevate ourselves for a while and take a completely model independent (or completely non-parametric) view of option pricing.

A mathematical expectation, however, is not a price. People may be risk-averse or risk-lovers and not expect the utility of expectation to be equal to the expectation of utilities. While all people agree on the price of the underlying, they may not agree on the prices of lotteries written on that underlying, i.e. derivative instruments

Fully non-parametric option pricing model

Let us imagine a discrete space and time grid for simplicity. We start out at (S_0, t_0). A fully non-parametric option pricing model is a model where it is completely free what the future transition probabilities $P(S_i, t_j, S_l, t_m)$ should be. $((i, j, l, m)$ are our discrete indices.) In other words, a model where you can assign any matrix of transition probabilities from states (S_i) at time t_j to states (S_l) at the following time t_{j+1}. (Compounding of probability will apply of course.) If your time horizon is discretized into N time slices and your space grid into M grid points, you model will have $N \times M \times M$ parameters, more or less. (In fact, $N \times M \times (M-1)$, as probabilities sum to one). NOT MENTIONING that the price of the underlying, S, may not be the only state variable. Other state variables may come into play, for instance instantaneous volatility, or instantaneous interest rate, or instantaneous hazard rate, or the indicator indicating whether the underlying name is in default or not, etc.. In other words, the matrices 7 $P(S_i, t_j, S_l, t_{j+1})$ might as well have to be indexed by $(v, r, h, \ldots)$ the n-tuple of extra state variables that we assume are discretized too: $P[v, r, h, \ldots](S_i, t_j, S_l, t_{j+1})$, and the probabilities of moving between different n-tuples $(v, r, h, \ldots)$ will have to be given too. The number of parameters will grow accordingly.

Option pricing models (local volatility, Heston, Merton, Bates, Pan Duffie Singleton, etc.) are just particular parametrizations of this fully non-parametric picture!

Risk-neutral vs. Real probability

This is not all. Indeed, I have said nothing about whether the probabilities in question are risk-neutral or real. Let us assume they are real. (What else?). Then, given the payoff of my particular derivative instrument (and this payoff may be anything, for instance the instrument can pay off differently in different states of volatility, or knock-out at certain barriers, or it can pay off differently in case of default, for instance a convertible bond can be restructured in case of default, or it can be knocked-out, etc.), given the payoff, I can certainly compute the discounted mathematical expectation of that payoff, through my very complex chain of transition matrices, and transitions between n-tuples indexing my transition matrices.

A mathematical expectation, however, is not a price. People may be risk-averse or risk-lovers and not expect the utility of expectation to be equal to

I'd rather hedge my instrument optimally not perfectly, with a small number of hedging instruments, and control the standard deviation of the P&L by minimizing it and estimating this minimum, than entertain the illusion that the standard deviation is zero

the expectation of utilities. While all people agree on the price of the underlying, they may not agree on the prices of lotteries written on that underlying, i.e. derivative instruments. As a matter of fact, different people with different utility functions can produce different families of prices of derivative instruments, under the sole constraint that the family of prices that any such agent is producing may not generate internal arbitrage opportunities. (In other words, my option prices may turn out completely different from yours, yet you cannot arbitrage me by simultaneously selling and buying from me, on the prices I produce. You may think you can arbitrage me, of course. According to your model, that is.)

There is a powerful theorem which says that a sufficient and necessary condition for the family of derivative instruments prices produced by any one agent not to generate internal arbitrage opportunities is that those prices may be formally written as the discounted expectation of the payoff under some probability. This guarantees positivity and linearity (which are the minimum requirements). There are of course, many such expectation operators. As many as there are different combinations of choices of transition matrices between our states of the world. Or as many as there are changes of probability measure, which can map the real probabilities into some other measure.

Since your derivative instrument prices are being represented as the discounted mathematical expectation of their payoff under the probability measure corresponding to your family of prices by the theorem above (assuming your prices do not violate non arbitrage), it all looks as if you were risk-neutral under this probability measure, as far as the prices of lotteries written on the underlying are concerned. This is the reason why probability measures that people use to express option prices as expectations of payoff are called "risk-neutral probability measures."

Why do we need the real probabilities then?

In order to hedge. You need them as soon as you start worrying about the unfolding of time and P&L. Only when you start worrying about connecting the evolution of the option price you are producing (or believing, or picking, or trading, etc.) to the evolution of the price of the underlying (or, for that matter, to the evolution of the price of any other instrument), in order to somehow compensate the one with the other, or balance the P&L of the one with the P&L from the other, you will start worrying about the real price dynamics, or the real probability. Real P&L and price movements take place in the real world, under the real probability. The risk-neutral probability, which is just a pricing operator, can only provide you with a formula such as $V(S, v, r, h, \ldots, t)$, linking the value of your derivative instrument with the value of the state variables, through time. The delta $\partial V/\partial S$ will only hedge you against continuous movement of the underlying S. It won't hedge you against jumps in S, or against changes of volatility, or default, or restructuring, or extraordinary dividends, etc.

Granted, you can use additional hedging instruments and construct a portfolio such that the other partial derivatives of your risk-neutral pricing operator, $\partial V/\partial v$, $\partial V/\partial r$, ..., are immunized. And then risk-neutral pricing will again coincide with perfect hedging and the detour in the real probability measure will not be needed. But I claim you shouldn't complete the market this way, right from the start. I'd rather hedge my instrument optimally not perfectly, with a small number of hedging instruments, and control the standard deviation of the P&L by minimizing it and estimating this minimum, than entertain the illusion that the standard deviation is zero. Optimal hedging (as opposed to perfect) in incomplete markets, and control of the standard deviation, ensure that the hedging is robust. Also it is always a precious piece of information to know how an additional hedging instrument can improve your hedging strategy. If you start out with (what you think is) a perfect hedging strategy, this information becomes irrelevant.

One approach to option pricing, or in other words, to producing the pricing operator of derivative instruments, or the risk-neutral probability, is the following. In the real probability and the real dynamics (as given by my complex combination of matrices of real transition probabilities above), work out, by stochastic control, some self-financing dynamic strategy involving, say, the underlying, which will optimally hedge the real P&L of my derivative instrument, in some sense of optimality. For instance, you can pick as criterion the dynamic hedging strategy which will make you break-even on average, in the real world, and guarantee that the standard deviation of your P&L is minimum. It is then up to you to quote as price for your derivative instrument, the initial cost of this self-financing strategy. We can show that this "initial cost of optimal dynamic hedging strategy" has the properties of a discounted expectation operator, therefore can act as pricing operator, or risk-neutral probability.

Black-Scholes or local volatility

Like I said, different option pricing models are just different parametrizations of the overly non-parametric picture above. If you assume diffusion for the underlying dynamics, as characterized by a diffusion coefficient $\sigma(S, t)$, and no other source of uncertainty, then the underdetermination of transition probabilities mentioned earlier disappears. The forward Kolmogorov equation, governing the probabilities $P(S, t, K, T)$, has $\sigma(S, t)$ as sole coefficient. Given the family of values $P(S_0, t_0, K, T)$, or in other words, the vanilla option prices, one can invert the Kolmogorov equation and map $\sigma(K, T)$ – or $\sigma(S, t)$ after relabelling – completely. (This is the essence of Dupire's formula.) The knowledge of $\sigma(S, t)$ then completes the knowledge of the underlying dynamics (because we have assumed diffusion, and nothing else). This, in turn, determines completely the whole future, the whole future transition probabilities, and the smile dynamics, thereby imposing the prices of what other derivative instruments are left, namely the exotics.

All we are saying is that local volatility happens to parameterize the overly non-parametric picture above with just the right number of parameters $\sigma(S_i, t_j)$ for the knowledge of $C(K_i, T_j)$ to be sufficient for determining the value of those parameters. You can see now what complete misrepresentation this is, to think that local volatility is the one non-parametric model!

There is additionally an important gain. The optimal dynamic, self-financing hedging strategy mentioned above happens to be perfect under the diffusion assumption. You can always perfectly replicate any derivative instrument with the help of the underlying alone! Although the family of prices of derivative instruments that other agents are producing may not generate internal arbitrage opportunities, those agents have no choice now but to agree with the family of prices you are producing with your optimal (in fact, perfect) hedging strategy. Otherwise, you would arbitrage them against the underlying! However, if hedging were forbidden in the Black-Scholes world, then families would multiply again! Even though the real volatility of the underlying should be known to be σ, nothing would stop me indeed from quoting non-arbitrage option prices using a different value σ' in the BS formula! And there will be as many different price families as there are different values σ'

How do we become model-independent again?

Going back to the full non-parametric picture described above – and this means that the underlying dynamics can be any wild Markovian process you may think of: jump-diffusion, stochastic volatility, etc. – we would achieve model-independence if we had as many option prices available in the market as we had parameters (for instance, $N \times M \times M$ in the case where the underlying is the sole state variable), to help us calibrate our parameters. And the options will have to be independent of each other. And by that I mean that it would not help to add options that you can replicate, in a model-independent way, with options you have already picked. If the vanillas are available, it would not help you adding call spreads, or butterflies, to the calibration set. (So to answer the frequently asked question about the American digital, or one-touch, being statically perfectly replicable by vanillas, therefore dependent only on the spot vanilla smile and no smile dynamics, we should go check whether the static replication argument that is being invoked doesn't itself depend on some hidden diffusion assumption... In other words, adding the American digitals, or general-

And this is not the end of the story. For even if such a huge number of options and options prices were available, we would still be model-dependent to the extent that we would only be calibrating risk-neutral transition probability matrices from those prices!

The reason for the discrepancy on the exotics is not that the models slightly differ on the vanillas and these little differences explode on the exotics! The reason is that the models massively differ on the exotics to begin with

ly the barrier options, will help expand our calibration set beyond the vanillas, for we cannot replicate them with the vanillas in a model-independent fashion).

This is a huge number of options we are talking about. For instance, when the underlying is the sole state variable (and this is hardly realistic), the knowledge of prices of, say, all knock-out call options $C(K, B, T)$ (where K is the strike, B the barrier level) is in theory sufficient to calibrate a fully non-parametric model. Now think that the whole point is precisely to price those barrier options, so how can we possibly require that their prices be fully available?! (In a way, it seems no model can price the barrier options in a model-independent fashion (!). Only the market can.)

And this is not the end of the story. For even if such a huge number of options and options prices were available, we would still be model-dependent to the extent that we would only be calibrating risk-neutral transition probability matrices from those prices! We would still have to "independently depend" on some model of utility function, or some hedging rationale, etc., to access the real world of hedging and P&L. Admittedly we wouldn't need that extra step if the sole purpose or our calibrated model is the pricing of some other stuff relative to the given stuff.

Other models (Heston, Merton, Heston with jumps, etc.)

So what do other smile models achieve, and how do they differ from each other with respect to smile dynamics? To repeat, different option pricing models are just different parametrizations of the overly non-parametric picture above. Stochastic volatility models offer a parametrization of different nature to local volatility.

Let us first assume, therefore, two state variables (S, ν), without first specifying the meaning of ν. And suppose ν is discretized on a grid with P points. You can think of ν as some abstract index, or a regime, indexing what used to be the situation "where the underlying was the sole state variable." At each time period t_k, we now need a transition probability matrix to govern the transition from (S_i, ν_j) to $(S_{i'}, \nu_{j'})$ at time t_{k+1}. We are talking about a number of entries of order $M \times P \times M \times P$ for that matrix! Therefore, there are now something like $N \times M \times M \times P \times P$ parameters in total.

We said that, for fixed ν, the underlying can follow any dynamics such that "the underlying is the sole state variable." (This is almost a tautology). Imagine, for instance, that the underlying diffuses with coefficient $\sigma(S, t)$ in each of the regimes ν. To be exact, we should index the local volatility functions $\sigma(S, t)$ prevailing in each regime with the regime label, and write $\sigma(\nu, S, t)$ in all rigour. Assuming diffusion in each regime reduces the number of parameters to $N \times M$ in each regime, as we've already seen. However, we still have $P \times P$ different ways of assigning transition probabilities from the collection of regimes (ν_j) at time t_k to the collection $(\nu_{j'})$ at time t_{k+1}. Perhaps v can be assumed to be diffusing in turn, therefore further reducing the degrees of freedom. Let us call $\nu(\nu, S, t)$ the diffusion coefficient of ν. And perhaps the diffusion of ν is correlated with the diffusion of S, through $\rho(\nu, S, t)$. And perhaps v should be mean-reverting in order that it doesn't blow up. And while we are at it, why not go all the way and assume that:

- $\sigma(\nu, S, t)$ is constant in each regime ν, why not propose $\sigma(\nu) = \nu$; in other words the states ν, or the regimes ν, are just states of instantaneous volatility of the underlying.
- ν, ρ, mean-reversion, etc., are constant across the whole state space (S, ν).

This is exactly the description of a stochastic volatility model à la Heston.

We can try and rehearse the steps of parametrization that helped us reduce the number of parameters from the wild $N \times M \times M \times P \times P$ picture down to the four or five parameters characteristic of Heston. First, the diffusion assumption was in itself a restraining parametrization. We then assumed diffusion for both the state variables S and ν. And finally the coefficients of this double diffusion process were assumed to be constant. People who think the vanillas are enough to calibrate the Heston model, should by contrast think of a situation where all the coefficients of Heston (vol of vol, correlation, mean reversion, long-term vol, etc.) are made function of the two state variables. Not forgetting that there is a double diffusion assumption at work here! Indeed, volatility can jump too, not in a uniform or orderly fashion, but correlated with jumps in the underlying, etc. etc.

When you calibrate a stochastic volatility model such as Heston, or a jump-diffusion model such as Merton, or some model combining the two features, such as Bates or Pan Duffie Singleton, etc., to the vanillas, all you are doing in effect is finding the value of the few constant parameters which are left free (vol of vol, correlation, intensity of jumps, moments of the distribution of jump sizes, etc.), through some search procedure which minimizes the distance between the model vanilla prices and the market vanilla prices. But you have to keep in mind that the huge parametrization step, from the wild non-parametric picture to the few parameters that are left for calibration, is completely model-dependent! When you first select Heston, or Merton, or Bates, etc., then turn to calibration, you have already committed yourself to a huge prescription as to how the conditional transition matrices should behave.

So what happens in practice?

In practice, you calibrate a parsimonious model (say Heston with five parameters) to a large number of vanilla options. Your minimization procedure converges to a local minimum. The vanillas are not exactly matched, of course, given the small number of parameters compared to the number of options. You calibrate another parsimonious model, say Pan Duffie Singleton, and fit the vanillas almost as well, only with a different distribution of calibration errors. You subsequently price the exotics and observe large differences between the models!

The reason for the discrepancy on the exotics is not that the models slightly differ on the vanillas and these little differences explode on the exotics! The reason is that the models massively differ on the exotics *to begin with*, in other words they are structurally different from each other as far as the smile dynamics, or conditional transition matrices, are concerned, and the calibration procedure is only succeeding in getting them this much in agreement on the vanillas! You could alternatively make them agree on the exotics by calibrating them to the exotics, but then they will diverge elsewhere. The unstated reason why people calibrate them to the vanillas, is that analytical or semi-analytical formulae are only available for the vanillas!

Recall that the ideal solution is to adopt the fully non-parametric model and require as many option market prices to calibrate it against as there are degrees of freedom. This is unrealistic, so the optimal solution should lie somewhere in between

The situation is even worse than you think. You may start selecting a complex model like Pan Duffie Singleton, which includes stochastic volatility correlated with the underlying, jumps in the underlying, jumps in volatility correlated with jumps in the underlying. Different empirical studies seem to suggest that this class of models are best for explaining the option smiles. You may then calibrate the model to the vanillas through the usual minimization procedure. As the vanilla prices are empirical (i.e. they contain noise and are not nearly as neat as the prices produced by some theoretical model), chances are the cost function will admit of several local minima. In other words, if you initialize your procedure with different guesses on the model parameters, chances are you will find a different solution. Now each of the two solutions fits the vanilla smile almost as satisfactorily as the other, yet with a different distribution of calibration errors (for we are talking about two different local minima). And chances are that these two solutions will yield completely different smile dynamics, or in other words, different exotic prices. The article "Can anyone solve the smile problem? " describes, with numerical detail, one such predicament. And it will not help at all, but only make things worse, to try to find the global minimum, as some have suggested, for you would only be calibrating noise, mindlessly of the exotics!

Can anyone solve the smile problem, then?

It seems there is no way but to find a model that you may calibrate to both the vanillas and the exotics. Recall that the ideal solution is to adopt the fully non-parametric model and require as many option market prices to calibrate it against as there are degrees of freedom. This is unrealistic, so the optimal solution should lie somewhere in between. And there is no reason why this practical model should follow along the lines of Heston, or Merton, or any similar parametric structure. Since the analytical solvability of the model is given up anyway (as far as the exotics are concerned), we might as well chose a structure which is best adapted to the given problem. We believe the regime-switching model provides such a platform.

Finally, with respect to the question of agreement of the models on the hedging strategies, recall that the delta V is only a partial derivative and far from being a hedge. Hedges have to be worked out in incomplete markets in some sense of optimality of hedging. "Can anyone solve the smile problem?" shows that models that disagree on the smile dynamics while agreeing on the spot implied volatility surface will in general disagree on the exotics *and* the hedging strategies (in the full sense of hedging in incomplete markets). Even two instances of the same model, reflecting two local minima of the vanilla calibration procedure, will differ on the hedging strategy. Two models (or two local minima of the same model) can even agree on both the vanilla smile and the deltas of the vanillas (typically when the models are homogeneous – by Euler's theorem), yet disagree on the hedging strategy!