Volatility Derivatives and the Implied Volatility Smile

Peter K. Friz
Cambridge University

Judge Business School , 7/10/2005

Joint work with Jim Gatheral & Shalom Benaim

- Assume the risk-neutral stock-process S_t is a continuous martingale and set $x_t = \ln(S_t/S_0)$. Recall that the fair value of a *variance-swap*

$$\mathbb{E}\left[\langle x\rangle_T\right]$$

is given by a strip of European Options

$$\int_0^{S_0} \frac{2dK}{K^2} P(K) + \int_{S_0}^{\infty} \frac{2dK}{K^2} C(K)$$
$$= \int_0^{\infty} dK\ w(K) \begin{cases} P(K) \text{ if } K < S_0 \\ C(K) \text{ if } K > S_0 \end{cases}$$

- Recent work by Carr & Lee (to be reviewed in a little) suggests that

$$\mathbb{E}\left[f\left(\langle x\rangle_T\right)\right] = \int_0^{\infty} dK w_f(K; S_0) \begin{cases} P(K) \text{ if } K < S_0 \\ C(K) \text{ if } K > S_0 \end{cases}$$

- **Theorem** (Structure of the ATM weights): If above holds then

$$w_{ATM} \equiv w_f(S_0; S_0) = 2f'(0)/S_0^2$$

Proof: Assume S_t follows an arbitrary stochastic volatility model possibly with non-zero correlation ρ. Dynamics:

$$\begin{aligned} dS_t &= S_t\sqrt{v_t}dW_t^1 \\ dv_t &= a(v_t)dt + b\left(v_t\right)\left[\rho dW_t^1 + \sqrt{1-\rho^2}dW_t^2\right] \\ dz_t &= v_t dt, \qquad z_0 = 0. \end{aligned}$$

As 3-dimensional diffusion its generator reads

$$\mathcal{L} = \frac{1}{2}S^2 v\partial_{SS} + (a\partial_v + \frac{1}{2}b^2\partial_{vv} + \rho S\sqrt{v}\partial_{vS}) + v\partial_z.$$

Denote $g_K(S)$ the payoff of out-of-money puts resp. calls struck at K. Solving the usual PDE may be written as

$P\left(K;T,S_0,v_0\right) = \left[e^{\mathcal{L}\,T}g_K\right]\left(S_0,v_0\right)$ and similar for C.

Similar to pricing of Asian options,

$$\begin{aligned} \mathbb{E}\left[f\left(\langle x\rangle_T\right)\right] &= \left[e^{\mathcal{L}\,T}f\right]\left(S_0,v_0\right) \\ &= \left[e^{\mathcal{L}\,T}f\right]\left(v_0\right). \end{aligned}$$

Thus

$$\left[e^{\mathcal{L}\,T}f\right]\left(v_0\right) = \int_0^\infty dK w_f(K;S_0)\left[e^{\mathcal{L}\,T}g_K\right]\left(S_0,v_0\right).$$

Take derivatives w.r.t. T and evaluate at $T = 0$,

$$[\mathcal{L}\ f]\,(v_0) = \int_0^\infty dK w_f(K; S_0)\,[\mathcal{L}\ g_K]\,(S_0, v_0).$$

Note that none of the payoffs under consideration (f, g_K) depends on instantaneous variance v. Thus

$$\begin{aligned} v_0 f'(z_0) &= \int_0^\infty dK w_f(K; S_0) \frac{1}{2} S_0^2 v_0 \partial_{SS} g_K(S_0) \\ v_0 f'(0) &= \int_0^\infty dK w_f(K; S_0) \frac{1}{2} S_0^2 v_0 v_0 \delta_K(S_0) \\ &= w_f(S_0; S_0) \frac{1}{2} S_0^2 v_0. \end{aligned}$$

QED

- **Applications:**
  Variance Swap $\Rightarrow w_{ATM} = 2/S_0^2$
  Volatility Swap $\Rightarrow w_{ATM} = +\infty$
  Variance Call $\Rightarrow$
  $w_{ATM} \equiv 0$ for all positive (total var.) strikes !!!

- **Theorem** (Carr/Lee): Assume independence between W and v. With $p(\lambda) = 1/2 \pm \sqrt{1/4 + 2\lambda}$ have

$$\mathbb{E}[e^{\lambda \langle x \rangle_T}] = \mathbb{E}[e^{p(\lambda) x_T}].$$

Proof: Elementary computation when $\langle x \rangle_T \equiv \sigma_{BS}^2 T$. Then average over different realizations of $\langle x \rangle_T$. QED

- Assume a Laplace representation for f

$$f(y) = \frac{1}{2\pi i} \int_{a-i\infty}^{a+i\infty} F(\lambda)\, e^{\lambda\, y}\, d\lambda.$$

This yields

$$\mathbb{E}\left[f(\langle x \rangle_T)\right] = \frac{1}{2\pi i} \int_{a-i\infty}^{a+i\infty} F(\lambda)\, \mathbb{E}\left[e^{\lambda \langle x \rangle_T}\right]\, d\lambda.$$

On the other hand, the power-options with value

$$\mathbb{E}\left[e^{\lambda \langle x \rangle_T}\right]$$

expand naturally in terms of put/call options.

- Using put-call symmetry and working with dimensionless quantities

$$\begin{aligned} k &= \log(K/S_0) \\ c(k) &= C(K)/K, \text{ etc.} \end{aligned}$$

find

$$\mathbb{E}\left[f(\langle x\rangle_T)\right] = \int_0^\infty dk\ c(k)\ w(k)$$

with (formal) weights $w(k)$ equal to

$$4\,e^{k/2}\left(\frac{1}{2\pi i}\int F(\lambda)\,\lambda\,\cosh[k\sqrt{1/4+2\lambda}]\,d\lambda\right).$$

- For comparison, the variance swap reads

$$\mathbb{E}\left[\langle x\rangle_T\right] = \int_0^\infty dk\ c(k)\ 4e^{k/2}\cosh\left[k/2\right].$$

- Apply this to the *Volatility Swap*

$$f(\langle x\rangle_T) = \sqrt{\langle x\rangle_T}.$$

Brute force computation (Mathematica) $\Rightarrow$

$$\begin{aligned} w_f(k) &= \left(\sqrt{\frac{\pi}{2}} I_1(k/2) + \sqrt{2\pi}\delta_0(k)\ \right) e^{k/2} \\ &\approx \sqrt{2\pi}\delta_0(k), \qquad I_1 \text{ ... mod. Bessel Fct.} \end{aligned}$$

Simple understanding of approximation via standard expansion of B.S. formula,

$$c(0) \sim \frac{1}{\sqrt{2\pi}}\sqrt{\sigma^2 T} = \frac{1}{\sqrt{2\pi}}\sqrt{\langle x \rangle_T}.$$

Proof (no brute force): W.l.o.g. $T \equiv 1$. Ansatz:

$$\int_0^\infty dk\ \left(\sqrt{\frac{\pi}{2}} F'(k/2) + \sqrt{2\pi}\delta(k)\ \right) e^{k/2} c_{BS}(k;\sigma) = \sigma$$

for some F to be found with $F(0) = 1$. This last choice and the constant $\sqrt{\frac{\pi}{2}}$ are motivated by the following integration by parts argument,

$$\begin{aligned} &\int_0^\infty dk\ \left(\sqrt{\frac{\pi}{2}} F'(k/2) + \sqrt{2\pi}\delta(k)\ \right) e^{k/2} c_{BS}(k;\sigma) \\ &= -\sqrt{2\pi}\left(\int_0^\infty dk F(k/2) \frac{\partial}{\partial k}\left[e^{k/2} c_{BS}(k;\sigma)\right]\right). \end{aligned}$$

Our statement is equivalent to saying that the above strip of European has Black–Scholes-vega equal to one. Using the well-known BS-vega for $c_{BS}(k;\sigma) = e^{-k}C_{BS}(k;\sigma)$ with

$$d_1 = \frac{-k+\sigma^2/2}{\sigma} = \frac{-k+v/2}{\sqrt{v}}$$

find

$$\begin{aligned}
& -\sqrt{2\pi}\left(\int_0^\infty dk F\left(k/2\right)\frac{\partial}{\partial k}\left[e^{k/2}\frac{\partial}{\partial\sigma}c_{BS}(k;\sigma)\right]\right) \\
= \;& -\sqrt{2\pi}\left(\int_0^\infty dk F\left(k/2\right)\frac{\partial}{\partial k}\left[e^{-k/2}\frac{1}{\sqrt{2\pi}}e^{-d_1^2/2}\right]\right) \\
= \;& \frac{e^{-v/8}}{v}\int_0^\infty dk\ F\left(k/2\right)\ k\ e^{-k^2/(2v)} \equiv 1\ \forall v
\end{aligned}$$

A power-expansion shows that

$$F\left(k\right) = \sum_{m=0}^{\infty}\frac{\left(k^2/4\right)^m}{\left(m!\right)^2}.$$

We recognize F as Bessel function $I_0(k)$ and finally recall

$$I_1(k) = \partial_k I_0(k).$$

QED

- Apply this to *Variance Calls* with payoff, say $K = 0.04 = (20\%)^2 \times 1$ year.

$$f(\langle x \rangle_T) = (\langle x \rangle_T - K)^+ .$$

Well-known Laplace-transform of f is

$$F(\lambda) = e^{-\lambda K} / \lambda^2.$$

Now the brute-force computation ... but defining integral for w does not converge!

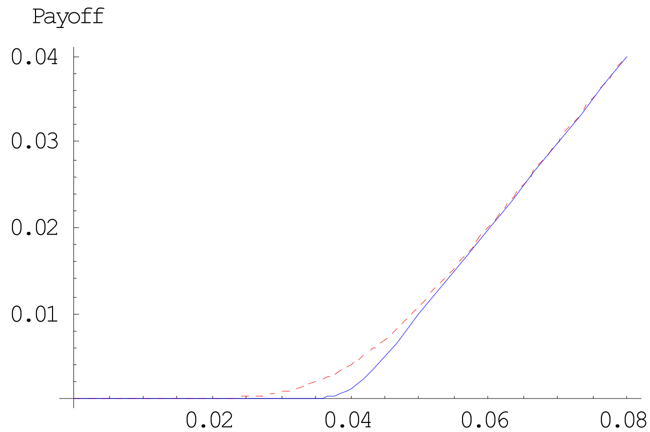

Figure 1:

- To see what is going on (and to guarantee converge) use convolution of payoff with Gauss-kernel of std.dev. $\sqrt{h}$ for $h = 10^{-4}$ (dashed line) and 10^{-5} (solid) respectively. This introduces an exponential damping factor and the weights

$$w(k) = w_{call}(k, K; h)$$

are given by

$$\frac{4\, e^{k/2}}{2\pi i} \int_{a-i\infty}^{a+i\infty} e^{h\lambda^2/2}\, \frac{e^{-\lambda K}}{\lambda} \cosh\left[k\sqrt{1/4 + 2\lambda}\right] d\lambda.$$

- Still numerically hard since $h \ll 1$ but scaling in λ does the trick and we can plot for any $h > 0$.

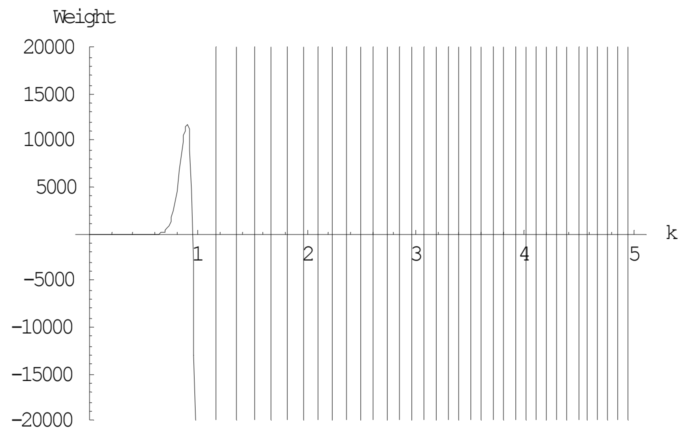

Figure 2: For $h = 10^{-4}$.

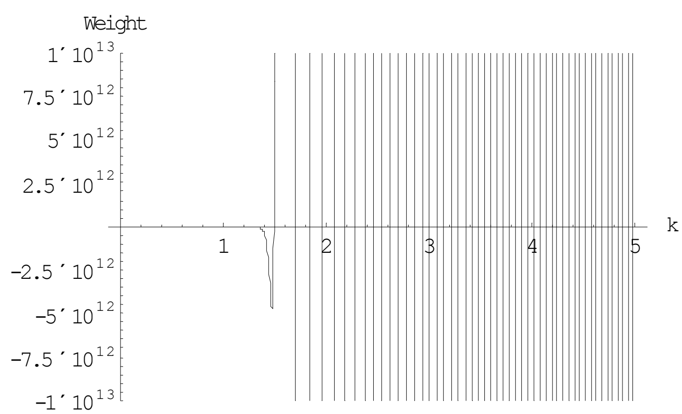

Figure 3: For $h = 10^{-5}$

- Since $w_{call}(k, K; h)$ is now defined by absolutely convergent integrals have no problem differentiating under the integral. Fix h and set

$$
\begin{aligned}
&\tilde{w}(k, K) \\
= &\ \frac{2\pi i}{4e^{k/2}} w_{call} \\
= &\ \int_{a-i\infty}^{a+i\infty} e^{h\lambda^2/2} \frac{e^{-\lambda K}}{\lambda} \cosh\left[k\sqrt{1/4 + 2\lambda}\right] d\lambda.
\end{aligned}
$$

Then $\tilde{w}$ satisfies a heat equation (with killing) on the domain $[0, \infty) \times [0, \infty)$

$$
\begin{aligned}
\frac{\partial^2}{\partial k^2} \tilde{w} &= \ldots \\
&= \frac{1}{4} \tilde{w} - 2 \frac{\partial}{\partial K} \tilde{w}.
\end{aligned}
$$

but the "time-derivative" $\frac{\partial}{\partial K}$ has the wrong sign. (Try to solve heat equations against the natural time direction ...)

Mathematically, such equations have exponential blow-up of all Fourier-modes and the oscillations seen before become understandable.

Dealing with Ill-posedness:

- Recall our standing assumptions (in particular $\rho = 0$) and that normalized BS-prices only depend only log-strike k and total variance $y = \sigma^2 T$.

- Let g denote the law of total variance. As observed by Hull/White

$$(*) \quad c(k) = \int_0^\infty c_{BS}(k, y)\, dg(y) \;=: [\mathfrak{L} g(.)](k).$$

(∼ convolution using a BS-integral kernel).

- Suffices to invert this relation such as to find the law g.Then simply price

$$\mathbb{E}\left[f(\langle x \rangle_T)\right] = \int_0^\infty f(y) dg(y).$$

- How to invert $(*)$? Similar to getting sharp pictures from Hubble ... (again, an ill-posed problem).

After a discretization can write

$$c_i = \sum_{j=1}^{\#var} A_{ij}\, g_j, \qquad i = 1, ..., \#c.$$

Naiv guess $\#var = \#c\#$ and invert matrix (badly conditioned). Ad hoc, stability can be obtained by $\#var \gg \#c$ and use Moore-Penrose's pseudo-inverse to find the least-square solution

$$\mathbf{g} = \mathbf{A}^{\mathbf{T}}(\mathbf{A}\,\mathbf{A}^{\mathbf{T}})^{-1}\,\mathbf{c} =: \mathbf{M}\,\mathbf{c}$$

NB: Matrix-inversion takes place in the "smaller" dimension. This punishes (unwanted oscillations) in the probability vector g.

- **Example** (Heston with BCC parameters): From CIR bond pricing formula obtain explicit c.f. of total variance. Now can price variance calls using the Carr/Madan techniques in quasi-closed form.

  Alternatively, we compute call-price (also in quasi-closed form) with BCC parameters but zero correlation and use the machinery above.

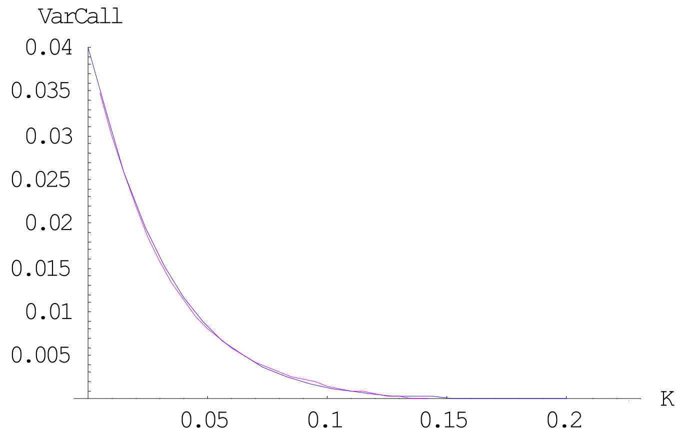

Figure 4: Blue line from quasi-closed form based on CIR formula. Red line from Moore-Penrose approach.

Heston/CIR dynamics for instantaneous variance

$$dv_t = -\lambda(v_t - \bar{v})dt + \eta\sqrt{v_t}dB_t$$

and consider variance calls with payoffs

$$\mathbb{E}\left[(\langle x\rangle_T - K)^+\right] = \mathbb{E}\left[\left(\int_0^T v_t dt - K\right)^+\right]$$

with BCC-fit: $v_0 = \bar{v} = 0.04$, $\lambda = 1.15$, $\eta = 0.39$

- The quality of the Moore-Penrose inversion deteriorates for high vvol $\eta \sim 1$ and high variances-strikes K which arise naturally in the pricing of *Capped Variance Swaps*.

- To get stable result in such regimes follow the standard procedure in the theory of ill-posed problems. Minimize

$$
\begin{aligned}
g \;\mapsto\; O(g) = & \sum_{j=1}^{n_{Strikes}} \left| \left(\sum_{i=1}^{n_{Var}} g_i c_{BS}(k_j, y_i) \right) - c(k_j) \right|^2 \\
& + \epsilon_1 \times \sum_{i=1}^{n} g_i^2 \\
& + \epsilon_2 \times d(p, g)
\end{aligned}
$$

  The Moore-Penrose inversion is a special case of the above with $\epsilon_2 = 0, \epsilon_1 \to 0$. The very last term with ϵ_2 weight serves as a distance between g and some a-priori probability vector p. An example for a distance d between two probability vectors is given by

$$
\sum p_i \ln \left[p_i / q_i \right].
$$

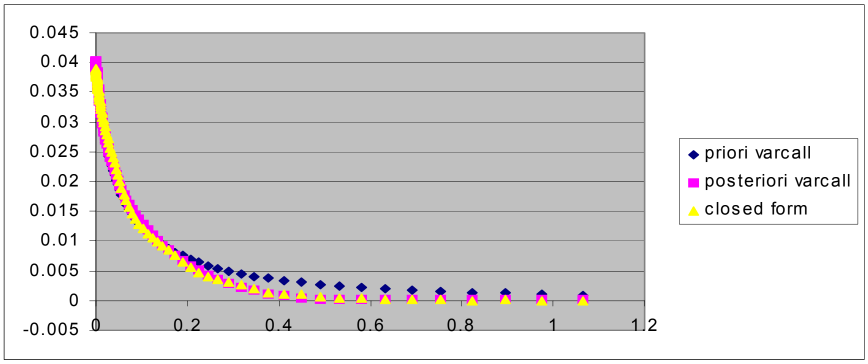

Figure 5: Heston with modified BCC parameter; $\eta = 1$ instead of 0.39

How to choose a good a priori law p? Take a 2 parameter family of laws (e.g. log-normal) and match variance and volatiliy swap.

- Recap: for var-swaps and zero-correlation vol-swaps

$$\begin{aligned} \mathbb{E}\langle x\rangle_T &= \int_0^\infty dk\ c(k)\ 4e^{k/2}\cosh\left[k/2\right] \\ \mathbb{E}\sqrt{\langle x\rangle_T} &= \int_0^\infty dk\ c(k)\ \left(\sqrt{\frac{\pi}{2}}I_1(\frac{k}{2}) + \sqrt{2\pi}\delta_0\left(k\right)\ \right)e^{\frac{k}{2}} \end{aligned}$$

- Matytsin showed: if $z \equiv d_2\left(k\right) = \frac{-k-\sigma^2(k)T/2}{\sigma(k)\sqrt{T}}$ then

$$\mathbb{E}\left[\langle x\rangle_T\right] = \int_{-\infty}^{\infty} dz N'\left(z\right)\sigma^2_{implied}\left(k\left(z\right)\right)$$

Var-swap price directly from the smile!

- Similar for vol-swap?

- Observe how d_2 comes into play although our setup is far more general than Black-Scholes.

- To understand $d_2(k)$ need to understand $\sigma_{implied}(k)$.

- Fundamental result by Roger Lee

$$\begin{aligned} \beta &\equiv \limsup_{k \to \infty} \frac{\sigma^2_{implied}(k)}{k} \in [0,2] \\ p &\equiv \sup\left\{q : ES_T^{1+q} < \infty\right\} \end{aligned}$$

  then $\beta = 2 - 4\left(\sqrt{p^2+p} - p\right)$ resp. 0 if $p = \infty$. (Similar formula for $k \to -\infty$)

- Interpretation: implied (variance) smile is asymptotically linear with slope β

- Too naiv: in Merton's jump model we have $p = \infty$ but smile is certainly not flat!

- [F Benaim] Implied variance in Merton is linear with logarithmic correction. Proof based on saddle point approximation.

- Other results: Hagan's SABR formula is wrong in the wings ...