Mixed Log-Normal Volatility Model

Richard White*

First version: 2 August 2012; this version†August 17, 2012

Abstract

We present two versions of a Mixed Log-Normal Volatility Model. The first is applicable to produce a single, arbitrage free, volatility smile (i.e. it produces a single terminal Probability Density Function (PDF) for some expiry) but not a consistent volatility surface. The second version can produce an arbitrage free implied volatility surface and the corresponding local volatility surface via simple analytic expressions. These surfaces are useful for testing other numerical methods, in particular methods to calculate the fair value of variance swaps as the true value is known from the parameters of the model via another simple expression.

1 Single Horizon Model

Let the terminal stock price S_T be given by

$$S_T = F_T e^X \tag{1}$$

where F_T is the forward and X is a random variable with density $\rho(x)$ given by a mixture of normal distributions:

$$\rho(x) = \sum_{i=1}^{N} w_i \phi\left[x; \left(\mu_i - \frac{\sigma_i^2}{2}\right)T, \sigma_i^2 T\right] \qquad \sum_{i=1}^{N} w_i = 1 \qquad w_i \geq 0\,\forall\, i \tag{2}$$

where $\phi(x;\mu,\sigma^2)$ denotes a normal distribution with mean μ and variance σ^2. In order that $\mathbb{E}[S_T] = F_T$, we we require $\mathbb{E}[e^X] = 1$, hence

$$\sum_{i=1}^{N} w_i e^{\mu_i T} = 1 \tag{3}$$

Except in the special case of $\mu_i = 0\,\forall\, i$, equation 3 can only be satisfied at one expiry T - this is why this setup can only produce a single smile not a full volatility surface.

The price of a European call is given by

$$\begin{aligned} C(T,k) &= P(0,T)\int_{-\infty}^{\infty} (F_T e^x - k)^+ \rho(x) dx \\ &= P(0,T)\sum_{i=1}^{N} w_i B(F_i^*, k, T, \sigma_i) \end{aligned} \tag{4}$$

*richard@opengamma.com

†Version 1.1

where $F_i^* = F_T e^{\mu_i T}$ and $B(\cdot)$ is the Black formula for the forward price

$$B(F, k, T, \sigma) = F\Phi(d_1) - k\Phi(d_2)$$
$$d_1 = \frac{\ln\left(\frac{F}{K}\right) + \frac{\sigma^2 T}{2}}{\sigma\sqrt{T}} \qquad d_2 = d_1 - \sigma\sqrt{T} \tag{5}$$

Where $\Phi(\cdot)$ is the cumulative Normal distribution. Hence options prices in this model are given simply by a weighted sum of Black prices. Implied volatilities are found by numerically inverting (root finding) the Black formula in the the usual way. Clearly Greeks are also given by the weighted sum of Black greeks.

1.1 Calibration

If the drifts μ are all set to zero, then the model has $2N-1$ free parameters (one free parameter is absorbed by the weights constraint $\sum_{i=1}^N w_i = 1$) - the N volatilities and the $N-1$ parameters that form the weights. Call these $N-1$ parameters the vector $\mathbf{y} \in \mathbb{R}^{N-1}$, then we require a mapping

$$\mathbf{y} \rightarrow \mathbf{w} \quad \text{such that} \quad \sum_{i=1}^N w_i = 1 \quad \text{and} \quad w_i \geq 0 \quad \forall\, i \tag{6}$$

Various non-linear mappings (normally trigonometric functions) exist and are discussed in the appendix. Calibration can then proceed by finding the $2N-1$ parameters that minimised the squared difference between the model and market prices.[1]

In the case of non-zero drifts, there are $3N-2$ free parameters. Let $a_i \equiv w_i e^{\mu_i T}$, then the second constraint is $\sum_{i=1}^N a_i = 1 \quad a_i \geq 0 \quad \forall\, i$, which is exactly the same form as the weight constraint. Having found the a_i from an additional $N-1$ free parameters, the drifts are $\mu_i = \frac{1}{T}\ln\left(\frac{a_i}{w_i}\right)$. So once again calibration can then proceed by finding the $3N-2$ parameters that minimised the squared difference between the model and market prices.

1.2 Variation of the Model

The constraint of equation 3 can be rewritten as

$$\sum_{i=1}^N w_i \hat{F}_i = 1 \qquad \text{where} \quad \hat{F}_i = \frac{F_i^*}{F_T} \tag{7}$$

and the call price as

$$C(T, k) = P(0, T) F_T \sum_{i=1}^N w_i B(\hat{F}_i, x, T, \sigma_i) \tag{8}$$

where $x = \frac{k}{F_T}$ is the moneyness. If the relative forward terms, $\hat{F}_i$, are treated as constant across expiries, then equation 8 gives European option prices for all T and k while preserving the forward. However the interpretation is thus: at time zero, the stock price, S_0, jumps to one of N levels $\hat{F}_i S_0$, it then follows a geometrical Brownian motion with volatility σ_i such that

$$\mathbb{E}[S_t|\text{state}_i] = \hat{F}_i F_t$$
$$\text{and} \quad \mathbb{E}[S_t] = \sum_{i=1}^N w_i \mathbb{E}[S_t|\text{state}_i] = F_t \tag{9}$$

[1] It is best to do this between market and model implied volatilities or prices weighted by vega.

This jump at time zero is somewhat unphysical and means it is not possible to delta hedge.

2 Multi Horizon Model

We make small change from equation 1 and write the stock price as

$$S_t = F_t e^{X_t - \Omega_t} \tag{10}$$

where Ω_t is the *drift compensator*, which ensures that $\mathbb{E}[S_t] = F_t \; \forall t$. Its value is given by

$$\Omega_t = \ln\left[\sum_{i=1}^{N} w_i e^{\mu_i t}\right] \tag{11}$$

Writing the relative forward, $\hat{F}_i(t)$ as

$$\hat{F}_i(t) \equiv e^{\mu_i t - \Omega_t} \equiv \frac{e^{\mu_i t}}{\sum_{i=1}^{N} w_i e^{\mu_i t}} \tag{12}$$

then the call price is again given by

$$C(T,k) = P(0,T) F_T \sum_{i=1}^{N} w_i B(\hat{F}_i(T), x, T, \sigma_i) \tag{13}$$

however the interpretation is different: at time zero the stock goes into one of N states, but the stock price does not jump, instead it follows a geometrical Brownian motion with volatility σ_i such that

$$\begin{aligned} \mathbb{E}[S_t|\text{state}_i] &= F_t e^{\mu_i t - \Omega_t} \\ \text{and} \quad \mathbb{E}[S_t] = \sum_{i=1}^{N} w_i \mathbb{E}[S_t|\text{state}_i] &= F_t \end{aligned} \tag{14}$$

Since we can find an arbitrage free price at any expiry, T, and any strike, k, we can find the corresponding implied volatility, hence we have a complete volatility surface.

2.0.1 Implied Volatility for $T \to 0$

For an ATM option we have

$$C(T,F) = F\left(2\Phi\left(\frac{\sigma\sqrt{T}}{2}\right) - 1\right) \tag{15}$$

so for small times[2] (or zero drifts) the ATM implied volatility is

$$\sigma(T,F) = \frac{2}{\sqrt{T}}\Phi^{-1}\left(\sum_{i=1}^{N} w_i \Phi\left(\frac{\sigma_i\sqrt{T}}{2}\right)\right) \tag{16}$$

[2]small times mean when diffusion dominates, i.e. $\sigma\sqrt{t} \gg \mu t \to t \ll \left(\frac{\sigma}{\mu}\right)^2$, and since σ is nearly always numerically greater than μ, this means any time significantly less than a year.

In the limit of $T \to 0$ this becomes[3]

$$\sigma(0, F) = \sum_{i=1}^{N} w_i \sigma_i \tag{17}$$

This should be very accurate for expiries less than a few days.

2.1 Local Volatility

Defining the fractional call price[Whi12b] as

$$\hat{C}(T, x) = \frac{C(T, xF_T)}{P(0, T)F_T} \tag{18}$$

we have

$$\hat{C}(T, x) = \sum_{i=1}^{N} w_i B(\hat{F}_i(T), x, T, \sigma_i) \tag{19}$$

The local volatility can be found by a slight modification of the Dupire formula [Dup94, Whi12b]

$$\hat{\sigma}(T, x)^2 = 2\frac{\frac{\partial \hat{C}}{\partial T}}{x^2 \frac{\partial^2 \hat{C}}{\partial x^2}} \tag{20}$$

The local volatility given here is parameterised by moneyness, x - the local volatility parameterised by strike, k, is given by $\sigma(T, k) = \hat{\sigma}\left(T, \frac{k}{F_T}\right)$.

The term in the denominator of equation 20, $\frac{\partial^2 \hat{C}}{\partial x^2}$, is the (modified) dual gamma. It is given by

$$\begin{aligned} \frac{\partial^2 \hat{C}}{\partial x^2} &\equiv \Gamma_x = \sum_{i=1}^{N} w_i \Gamma_x(\hat{F}_i(T), x, T, \sigma_i) \\ \text{where} \quad \Gamma_x(F, x, T, \sigma) &= \frac{\phi(d_2)}{x\sigma\sqrt{T}} = \frac{F\phi(d_1)}{x^2\sigma\sqrt{T}} \end{aligned} \tag{21}$$

$\phi(\cdot)$ is the Normal PDF.

The numerator is the modified theta[4], it is given by

$$\begin{aligned} \frac{\partial \hat{C}}{\partial T} \equiv \Theta &= \sum_{i=1}^{N} w_i \left[\Theta(\hat{F}_i(T), x, T, \sigma_i) + \Delta(\hat{F}_i(T), x, T, \sigma_i)\frac{\partial \hat{F}_i(T)}{\partial T}\right] \\ \text{where} \quad \Theta(F, x, T, \sigma) &= \frac{F\phi(d_1)\sigma}{2\sqrt{T}}, \qquad \Delta(F, x, T, \sigma) = \Phi(d_1) \\ \text{and} \quad \frac{\partial \hat{F}_i(T)}{\partial T} &= \hat{F}_i(T)\left(\mu_i - \mu_T^*\right) \quad \text{where} \quad \mu_T^* = \frac{\sum_{j=1}^{N} w_j \mu_j e^{\mu_j T}}{\sum_{j=1}^{N} w_j e^{\mu_j T}} \end{aligned} \tag{22}$$

[3]We have used the well know result that $\Phi(x) \approx \frac{1}{2} + \frac{x}{\sqrt{2\pi}}$ for $x \to 0$

[4]theta is normally defined as an options time decay - the change in price with respect to calendar time. The theta here is the sensitivity of (fractional) option price to expiry.

Putting this all together, we have an expression for the local volatility

$$\begin{aligned}\hat{\sigma}(T,x)^2 &= \frac{\sum_{i=1}^N w_i \hat{F}_i(T)\left[\phi(d_{1,i})\sigma_i + 2\sqrt{T}\Phi(d_{1,i})(\mu_i - \mu_T^*)\right]}{\sum_{i=1}^N w_i \hat{F}_i(T)\frac{\phi(d_{1,i})}{\sigma_i}} \\ \text{where} \quad d_{1,i} &= \frac{\ln\left(\frac{\hat{F}_i(T)}{x}\right) + \frac{\sigma_i^2 T}{2}}{\sigma_i\sqrt{T}}\end{aligned} \tag{23}$$

It should be noted that switching from call delta, $\Phi(d_{1,i})$, to put delta, $-\Phi(-d_{1,i}) \equiv \Phi(d_{1,i}) - 1$, gives the same local volatility (as it must) by virtue of the definition of μ_T^*, and one should use the one for out of the money options (i.e. call for strike greater than forward).

2.1.1 Calendar Arbitrage

While the denominator of equation 20 is always positive, the numerator is not. If the numerator does become negative, the local volatility does not exist at that point and the model is not valid for times at and beyond that point. The condition

$$\sum_{i=1}^N w_i \hat{F}_i(T)\left[\phi(d_{1,i})\sigma_i + 2\sqrt{T}\Phi(d_{1,i})(\mu_i - \mu_T^*)\right] \geq 0 \quad \forall T, x \tag{24}$$

cannot be satisfied for arbitrary $\boldsymbol{\mu}$ - this limits its applications as care must be taken in the choice of parameters.

2.1.2 Regularisation

At extreme strikes, the values $|d_{1,i}|$ can be quite large and thus the $\phi(d_{1,i})$ terms are very small - this division of two very small numbers can give spurious results (and not-a-number in extreme cases). If j is the index of the smallest value of $|d_{1,i}|$ and we dividend top and bottom by $\phi(d_{1,j})$, we rewrite the expression as as

$$\hat{\sigma}(T,x)^2 = \frac{\sum_{i=1}^N w_i \hat{F}_i(T)\left[\exp\left(\frac{d_{1,j}^2 - d_{1,i}^2}{2}\right)\sigma_i + 2\sqrt{2\pi T}\exp\left(\frac{d_{1,j}^2}{2}\right)\Phi(d_{1,i})(\mu_i - \mu_T^*)\right]}{\sum_{i=1}^N w_i \frac{\hat{F}_i(T)}{\sigma_i}\exp\left(\frac{d_{1,j}^2 - d_{1,i}^2}{2}\right)} \tag{25}$$

The term $\exp\left(\frac{d_{1,j}^2}{2}\right)\Phi(d_{1,i})$ can still be problematic, however for very large negative $d_{1,i}$ we can use the approximation

$$\Phi(x) \approx \frac{\phi(x)}{\sqrt{1+x^2}} \quad \text{for} \quad x \ll 0 \tag{26}$$

and write the term as

$$\exp\left(\frac{d_{1,j}^2}{2}\right)\Phi(d_{1,i}) \approx \frac{\exp\left(\frac{d_{1,j}^2 - d_{1,i}^2}{2}\right)}{\sqrt{2\pi}\sqrt{1+d_{1,i}^2}} \tag{27}$$

2.1.3 Local Volatility for $T \to 0$

In the limit of $T \to 0$, the ATM (i.e. $x = 1$) local volatility is given by

$$\hat{\sigma}(0,1)^2 = \frac{\sum_{i=1}^N w_i \sigma_i}{\sum_{i=1}^N \frac{w_i}{\sigma_i}} \tag{28}$$

If we define $y = \frac{\ln(x)}{\sqrt{T}}$ then the local volatility for $T \to 0$ is

$$\hat{\sigma}(T,x)^2 = \frac{\sum_{i=1}^{N} w_i \exp\left(-\frac{y^2}{2\sigma_i^2}\right)\sigma_i}{\sum_{i=1}^{N} w_i \exp\left(-\frac{y^2}{2\sigma_i^2}\right)\frac{1}{\sigma_i}} \tag{29}$$

That is, it is a function of y only. For $T < T_0$ where T_0 is some small expiry cut-off[5], we have

$$\hat{\sigma}(T,x) = \hat{\sigma}\left(T_0, x^{\sqrt{\frac{T_0}{T}}}\right) \tag{30}$$

In this way we can extrapolate the local volatility surface to $T = 0$ without numerical problems.

2.2 Variance Swaps

The pricing of variance swaps from implied and local volatility surfaces is discussed here [Whi12a]. When the underlying is a pure diffusion process (i.e. no jumps), the expected value of the annualised realised variance is given by [DDKZ99, Gat06]

$$EV(T) = -\frac{2}{T}\mathbb{E}\left[\ln\left(\frac{S_T}{F_T}\right)\right] \tag{31}$$

For the mixed log-normal model this becomes

$$\begin{aligned} EV(T) &= -\frac{2}{T}(\mathbb{E}[X_T] - \Omega_T) \\ &= -\frac{2}{T}\left(\sum_{i=1}^{N} w_i\left(\mu_i - \frac{\sigma_i^2}{2}\right)T - \ln\left[\sum_{i=1}^{N} w_i e^{\mu_i T}\right]\right) \\ &= 2\left(\frac{1}{T}\ln\left[\sum_{i=1}^{N} w_i e^{\mu_i T}\right] - \sum_{i=1}^{N} w_i\left(\mu_i - \frac{\sigma_i^2}{2}\right)\right) \end{aligned} \tag{32}$$

In the limit of $T \to 0$ (or $\mu_i = 0\ \forall\, i$) this becomes simply

$$EV(T \to 0) = \sum_{i=1}^{N} w_i \sigma_i^2 \tag{33}$$

So we have a very simple formula for the exact value of a variance swap under a mixed log-normal model. The volatility surfaces (implied and local) that this model produces can be treated as extraneously given, and use in numerical procedures to calculate the value of variance swaps - thus providing an excellent test of the numerical procedures.

2.3 Examples

Figures 1 and 2 show the implied and local volatility surfaces from a model with $N = 3$, $\mathbf{w} = \{0.7, 0.25, 0.05\}$, $\boldsymbol{\sigma} = \{0.3, 0.6, 1.0\}$ and $\boldsymbol{\mu} = \{0.0, 0.3, -0.5\}$. The spot is 100 and drift is 5%.

[5] $T_0 = 10^{-4}$ is general suitable

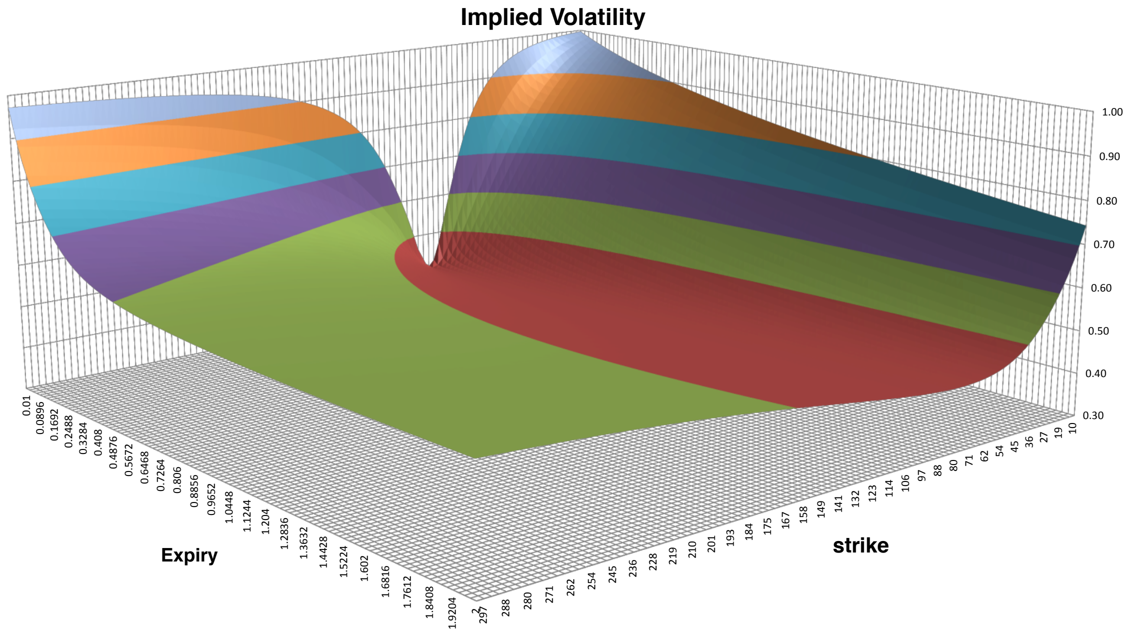

Figure 1: The Implied Volatility Surface from a mixed log-normal model.

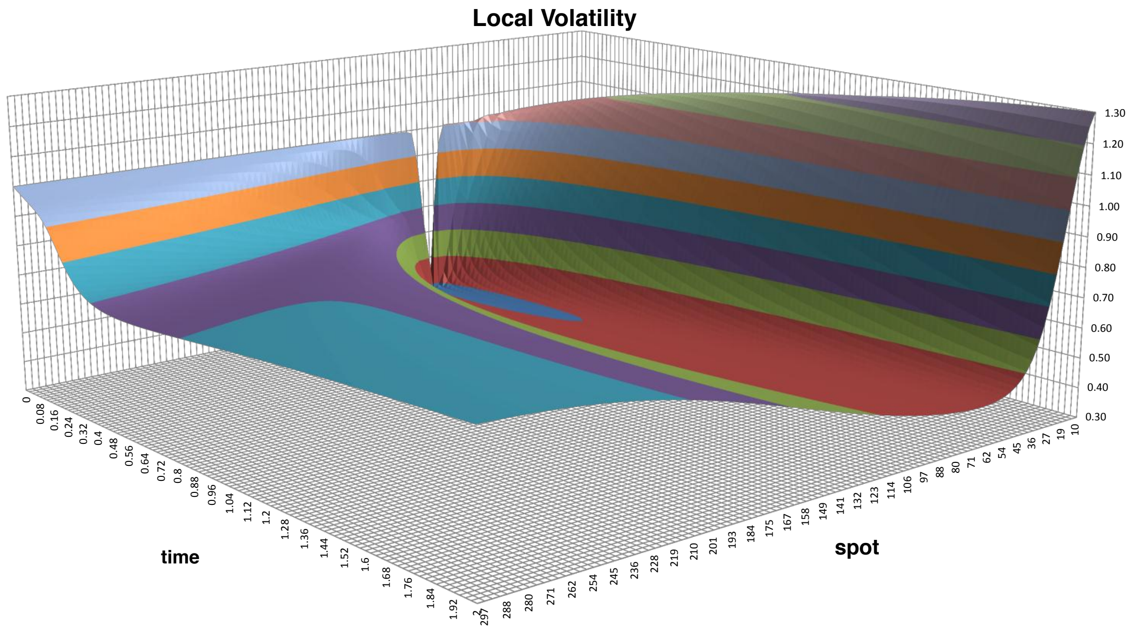

Figure 2: The LocalVolatility Surface from a mixed log-normal model.

A Sum to a Fixed Number Constraint

The constraint

$$\sum_{i=1}^{N} \alpha_i^2 = C \quad C > 0 \tag{34}$$

where C is some constant, occurs frequently We can rescale to $C = 1$ without any loss of generality. The constraint

$$\sum_{i=1}^{N} w_i = 1 \quad w_i \geq 0 \,\forall\, i \tag{35}$$

is equivalent by setting $w_i = \alpha_i^2$. For $N = 2$ the obvious solution is

$$\begin{aligned} \alpha_1 &= \cos\theta \\ \alpha_2 &= \sin\theta \end{aligned} \tag{36}$$

where the angle θ is unconstrained[6]. The vector $\boldsymbol{\alpha} = (\alpha_1, \alpha_2)^T$, lies on the unit circle. This can be extended to higher dimensions by describing a point on the N-dimensional unit hypersphere using $N-1$ angles. One such scheme [Reb04] is to build the N scheme from the $N-1$ scheme. If the vector $\boldsymbol{\alpha}^{N-1}$ satisfies equation 34 then

$$\begin{gathered} \alpha_i^N = \begin{cases} c_1 & \text{for } i = i \\ s_1 \alpha_i^{N-1} & \text{for } i > 1 \end{cases} \\ \text{where} \quad s_i \equiv \sin\theta_i \;\&\; c_i \equiv \cos\theta_i \end{gathered} \tag{37}$$

will also satisfy the equation. Starting from the $N = 2$ case, the general expression is

$$\alpha_i = \left(\prod_{k=1}^{i-2} s_k \right) \times \begin{cases} c_i & \text{for } i < N \\ s_{N-1} & \text{for } i = N-1 \end{cases} \tag{38}$$

where the product term is set to 1 if it contains no terms (i.e. $i < 3$).

This does not evenly distribute the number of trigonometric terms among the αs - α_N has $N-1$ terms while α_1 has only a single cosine. This means that a uniformly distributed set of points in the $N-1$ hypercube, will map to a set of points heavily concentrated around the poles of the hypersphere.[7] The effect grows with N and can affect search algorithms.

In order to democratise the distribution of trigonometric terms, we prepose the following scheme. Firstly assume that we have two schemes of size N_1 and N_2 with vectors $\boldsymbol{\alpha}^{N_1}$ and $\boldsymbol{\alpha}^{N_2}$ both satisfying equation 34. The scheme for $N = N_1 + N_2$ is then

$$\alpha_i = \begin{cases} s_{N-1} \alpha_i^{N_1} & \text{for } i \leq N_1 \\ c_{N-1} \alpha_{i-N_1}^{N_2} & \text{for } i > N_1 \end{cases} \tag{39}$$

and it is easy to verify that this also satisfies equation 34. So by partitioning N into two near equal sized parts N_1 and N_2 and repeating until $N = 1$, we can build up the scheme starting

[6] of course only some 2π range is unique.

[7] This is the opposite of what happens when a global is projected onto a plane (i.e. a map), making Greenland disproportionally large.

from the trivial $\alpha_1^{N=1} = 1$. For the case of $N = 7$ the following scheme results:

$$
\begin{aligned}
\alpha_1 &= s_1 s_2 s_4 \\
\alpha_2 &= s_1 s_2 c_4 \\
\alpha_3 &= s_1 c_2 s_5 \\
\alpha_4 &= s_1 c_2 c_5 \\
\alpha_5 &= c_1 s_3 s_6 \\
\alpha_6 &= c_1 s_3 c_6 \\
\alpha_7 &= c_1 c_3 \, 1
\end{aligned} \tag{40}
$$

We have included the multiply by 1 term in α_7 for clarity. The maximum number of trigonometric terms for any α is $\lceil \log_2 N \rceil$ and the minimum is one less than this. Also the number of multiplication has been reduced from $\frac{N(N-1)}{2}$ to $N \log_2 N$.

References

[DDKZ99] Demeterfi, Derman, Kamal, and Zou. More than you ever wanted to know about volatility swaps. Technical report, Goldman Sachs Quantitative Strategies Research Notes, 1999.

[Dup94] Bruno Dupire. Pricing with a smile. *Risk*, 7(1):18–20, 1994. Reprinted in Derivative Pricing:The Classical Collection, Risk Books (2004).

[Gat06] J. Gatheral. *The Volatility Surface: a Practitioner's Guide*. Finance Series. Wiley, 2006.

[Reb04] Riccardo Rebonato. *Volatility and Correlation*. John Wiley & Sons, 2nd edition, 2004.

[Whi12a] Richard White. Equity variance swap with dividends. Technical report, OpenGamma working paper, 2012. `http://developers.opengamma.com/quantitative-research/Equity-Variance-Swaps-with-Dividends-OpenGamma.pdf`.

[Whi12b] Richard White. Local volatility. OG notes, OpenGamma, 2012. `http://developers.opengamma.com/quantitative-research/Local-Volatility-OpenGamma.pdf`.