Short and Long Term Smile Effects:
The Binomial Normal Mixture Diffusion Model

ISMA Centre Discussion Papers in Finance 2003-06
First Version: November 2002
This Version: 26 March 2003

Carol Alexander
ISMA Centre, University of Reading, UK

with

George Brintalos
Barclays Capital, London, UK

and

Leonardo Nogueira
ISMA Centre, University of Reading, UK

Copyright 2003 Carol Alexander. All rights reserved.
Abstract

This paper extends the normal mixture diffusion (NMD) local volatility model of Brigo and Mercurio (2000, 2001a,b, 2002) so that it explains both short-term and long-term smile effects. Short-term smile effects are captured by a local volatility model where excess kurtosis in the price density decreases with maturity. This agrees with the ‘stylised facts’ of econometric analysis of ex-post returns of different frequencies and follows from the central limit theorem. We introduce the ‘binomial’ NMD model, so called because it is based on simple and intuitive assumptions that imply that the mixing law for the normal mixture log price density is binomial. This very parsimonious model can easily be calibrated to observed option prices, and it explains the short-term smile effect where leptokurtosis in the log price density decreases rapidly with time. However, smile effects in currency options often persist into fairly long maturities, and to capture at least some part of this it is necessary to introduce uncertainty. Longer-term smile effects that arise from uncertainty in the local volatility surface are modeled by a simple extension of the binomial NMD model. The results are illustrated by calibrating the model to a currency option smile surface.

Author Contacts:
Prof. Carol Alexander (corresponding author)
Chair of Risk Management and Director of Research
ISMA Centre, School of Business, University of Reading, Reading RG6 6BA
c.alexander@ismacentre.rdg.ac.uk

George Brintalos
Quantitative Analyst, Equity Derivatives Group / Product Control
Barclays Capital, 5 The North Colonnade, Canary Wharf, London E14 4BB
george.brintalos@barclayscapital.com

Leonardo Nogueira
ISMA Centre, School of Business, University of Reading, Reading RG6 6BA
l.m.nogueira@reading.ac.uk

JEL Classification: G12, C16

Keywords: Local Volatility, Stochastic Volatility, Smile Consistent Models, Term Structure of Option Prices, Normal Variance Mixtures.

Acknowledgements

Many thanks to Jacques Pézier, Visiting Professor at the ISMA Centre, for many stimulating and thought provoking discussions about volatility, and about this model in particular; also many thanks for valuable comments to Ali Bora Yigitbasioglu of the ISMA Centre; Fabio Mercurio of Bank IMI, Milan; Emmanuel Acar of Bank of America, London; Jessica James of Citibank, London and Hyung-Sok Ahn of Constellation Power, USA. George Brintalos and Leonardo Nogueira wrote the VBA code for calibrating the model parameters as part of their MSc dissertations for the Financial Engineering and Quantitative Analysis program at the ISMA Centre.

This discussion paper is a preliminary version designed to generate ideas and constructive comment. The contents of the paper are presented to the reader in good faith, and neither the author, the ISMA Centre, nor the University, will be held responsible for any losses, financial or otherwise, resulting from actions taken on the basis of its content. Any persons reading the paper are deemed to have accepted this.
1. Introduction

The Black-Scholes (BS) implied volatility smile effect arises from an apparent under-pricing of out-of-the-money (OTM) puts and calls by the Black and Scholes (1973) formula for simple European options under constant volatility. This is because the market does not believe the BS assumption that the price process is a Brownian motion with constant volatility. In particular, if leptokurtosis were present in the price process, the likelihood of large price changes would be higher than that assumed by a Brownian motion with the same volatility. Thus, if option traders believe in a leptokurtic price process, they will place a greater value on OTM puts and calls, giving market prices that are higher than BS model prices. Consequently the BS implied volatility of these options will be higher than the BS implied volatility of ATM options.

Often the smile effect is greatest for near term options but decreases with maturity. In this case the market data corroborate the stylized facts that emerge from econometric research. Examination of historical returns sampled at different frequency has shown that (a) the excess kurtosis estimated from unconditional historical returns densities decreases with sampling frequency\(^1\) and (b) there is strong evidence to support a non-constant conditional volatility model – and this implies leptokurtosis in the unconditional density, even if the conditional densities were normal.\(^2\) Both (a) and (b) indicate that leptokurtosis in the ex-post returns density increases with sampling frequency, but for different reasons.

These findings also concur with the central limit theorem: i.e. that the sum of non-normal variables tends towards a normal variable: if \(X_i\) have independent identical distributions with mean \(0\), variance \(\sigma^2\) and excess kurtosis \(\kappa\), then \(Y = (X_1 + \ldots + X_n)\) has a distribution with mean \(0\), variance \(n\sigma^2\) and kurtosis \(3 + \kappa/n\), so the kurtosis approaches 3 and the excess kurtosis approaches zero as \(n\) increases.

\(^1\) Intra-day and daily returns commonly exhibit highly significant excess kurtosis, but this usually disappears when returns are taken over a month or more (Goodhart and O'Hara, 1997; Gencay et. al., 2001).

\(^2\) The vast literature on estimating conditional densities of returns using models in the generalized autoregressive conditional heteroscedasticity (GARCH) family provides overwhelming evidence for non-constant conditional volatility (Bollerslev, 1986 and 1987; Baillie and Bollerslev, 1989, Bollerslev et. al. 1992 and 1994). Also, conditional heteroscedastic effects become more pronounced as the frequency of returns increases (Baillie and Bollerslev, 1990; Dacorogna et. al., 1998).
Thus, if the market held beliefs about the future evolution of the underlying price based on observations of past returns, a dynamic process with time-varying volatility and possibly also heavy tails in the conditional densities of returns may be used: either or both assumptions would lead to leptokurtosis in the price density that is decreasing with term. Many normal and non-normal GARCH models will capture this, but there are alternatives. For example, one could assume a more general Brownian price process with a time-varying volatility having the property that leptokurtosis in the price density decreases with term whilst average volatility over the longer-term is fairly constant. With such a model the smile effect would decrease with maturity.

However, in many markets a persistence of smile effects into 3 months or longer maturity implied volatilities is commonly observed, even though the underlying ex-post returns densities are approximately normal at the 3 month sampling frequency. However, there is no inconsistency here. Price densities will be leptokurtic when ex-post returns are unconditionally normally distributed if there is uncertain volatility in the price process. Therefore smile effects at 3 months or longer maturities can arise from volatility uncertainty. This does not contradict the stylized facts that emerge from econometric analysis. By ‘uncertain volatility’ we do not mean that there is uncertainty in the past, present and future volatility of the price process – the uncertainty is in the minds of traders because the market is uncertain about the future volatility of the price process – and that is why this uncertainty cannot be observed in ex-post returns. If, as is often the case, volatility uncertainty increases with term, then ex-ante price densities could become more leptokurtic as the term increases.

Uncertainty in volatility may be captured by the stochastic volatility models of Heston (1993), Hull and White (1987) and many others. These models do not allow for arbitrage-free pricing because the new uncertainty introduces market incompleteness. In this case option prices will include a risk premium so there is no unique model price if traders have differing attitudes to risk. For liquid options, and short-term at-the-money (ATM) options in particular, risk premia on the buy and sell side are likely to be small and similar; buyers or sellers can close their portfolios quickly if they wish to. But for longer-term OTM options where trading is sparse, the bank that writes the option may include a substantial risk premium in the price, depending on their risk attitude, and so the market incompleteness introduced by uncertain volatility can cause large bid-offer spreads at longer maturities.
Liquidity premia may also contribute to smile effects. Although the market for very ITM options is normally very thin – many investors preferring to trade in the underlying – market prices of these options may also differ from BS prices because of the large cash amounts involved with these transactions. Market prices should also include the bid-offer spread in interest rates, hedging costs (over the longer term) and possibly also credit effects. Thus there are many reasons in addition to volatility uncertainty for market prices of OTM puts and calls to be above the BS constant volatility price, particularly for longer-term options.

In summary, there are two smile effects at play: one that induces a smile that is largest for very short dated options, where the smile and the associated leptokurtosis in the price density decrease with term, and another effect that is due to volatility uncertainty as well as other sources of incompleteness and market imperfections, where leptokurtosis in the price density and the associated smile effect in implied volatilities can increase with term. The prices observed in the market will reflect a mixture of these two effects.

The aim of this paper is to derive a model that captures both short-term and longer-term smile effects. First, an extension of the Normal Mixture Diffusion (NMD) local volatility model of Brigo and Mercurio (2000, 2001a,b, 2002) is used to describe a leptokurtic price process in a complete market setting. The resulting parameterization of local volatility is intuitive, parsimonious and easy to calibrate to the market smile. In this model, arbitrage-free option prices are easily obtained because no market incompleteness is introduced. We give an example of calibrating a simple version of the model, with only three parameters, which is based on an analytic relationship between the local volatility model prices and the BS prices for standard European options. Because there are so few parameters, the model will not fit market prices exactly so its main application should be for hedging liquid options that will be marked-to-market. Option sensitivities for this model are easily calculated as weighted averages of Black-Scholes sensitivities. Having shown that this model explains the short-term component of the smile due to a leptokurtic price process with local volatility, we then distinguish the short-term smile from the longer-term smile due to uncertainty in future volatility, and model these longer-term smile effects by introducing uncertainty to the mixing law. The final model, which may have as few as five parameters, is calibrated to market data on currency option option prices.
The structure of this paper is as follows: section 2 reviews the literature on local volatility models; Section 3 explains the normal mixture diffusion (NMD) local volatility framework introduced by Brigo and Mercurio (2000, 2001a,b, 2002) and Brigo, Mercurio and Sartorelli (2002) where price densities are finite lognormal variance mixtures; Section 4 describes how to extend the NMD model to capture a term structure of kurtosis and the short-term smile effects described above. The simplest parameterization of the extended NMD model has a mixing law at time $n\Delta t$ that is binomial $B(n, \lambda)$ with a fixed $\lambda$. The calibration of the parameters of this model to market data is explained. Section 5 presents the results of calibrating the model to a currency option smile surface, and discusses how much of the observed behaviour of option prices can be attributed to this short-term smile model. Section 6 extends the binomial NMD model to include uncertainty in the value of $\lambda$ and we demonstrate that this model can explain longer maturity smile effects in currency option market data. Section 7 summarizes and concludes.

2. Local Volatility

The deterministic approach to non-constant volatility preserves market completeness by assuming the instantaneous or ‘local’ volatility of the price process is a deterministic function of time and the underlying asset values. Non-parametric local volatilities may be calibrated to current market prices of options using finite difference schemes. The ‘implied tree’ approach was pioneered by Dupire (1994, 1997) and subsequently extended to trinomial trees by Derman, Kani and Chriss (1996).³ One first interpolates and extrapolates the implied volatility surface and then uses a finite difference solution of the Black-Scholes equation to extract local volatilities for each node in the tree. Several refinements of the finite difference schemes employed for the model resolution have been proposed, the most stable of which appear to be the Crank-Nicholson scheme for trinomial lattices used by Andersen and Brotherton-Ratcliffe (1997).

Direct calibration with non-parametric local volatilities will provide an exact fit to the current market data, but the local volatilities (and therefore also the model hedge ratios) are typically very sensitive to the interpolation and extrapolation methods used, particularly for the wings of the implied volatility smile and for longer maturity options. If the smile is very pronounced, local volatilities may become negative, necessitating the use of some ad hoc procedures. Moreover, with incomplete and/or stale option price data, the calibrated local volatility surfaces will be excessively ‘spikey’ and consequently can give large variations in delta from day to day. For this
reason, regularization methods have been used by Avellaneda et. al. (1997), and Bouchouev and Isakov (1997, 1999) amongst others, to obtain the smoothest possible fit to the interpolated implied volatility surface. But still, with such an exact fit, the local volatility surface may jump considerably over time and the approach provides no foresight of future movements in the local volatility surface. These local volatilities are unlikely to perform well in out-of-sample tests and the direct calibration approach may be of limited practical use for hedging purposes.

Much research now focuses on the use of a parametric form for local volatilities. Pioneered by Cox and Ross (1976), in the parametric approach a functional form for the local volatility is chosen and the parameters are calibrated using only the available and reliable market prices. Typically, parameterized local volatilities will be smoother and more stable over time than those obtained by direct calibration. Although the model prices based on a parameterized local volatility surface will not exactly match the current market prices, these local volatilities could be more useful for hedging. If the calibration is sufficiently robust, we should gain some idea of the likely movements in the local volatility surface over time, and could modify hedge ratios accordingly.4

Recently a number of parametric and semi-parametric forms for local volatility have been proposed in the literature, including: simple polynomials (Dumas et. al.,1998); cubic splines (Coleman et al., 1999); hyperbolic trigonometric functions (Brown and Randall, 1999); Hermite polynomials (McIntyre, 2001); and piecewise quadratic forms (Beaglehole and Chebanier, 2002) amongst many others. Whilst these all represent useful developments in the local volatility literature, the problem for practitioners is now how to choose the ‘best’ functional form for their purposes. It is not simply a question of choosing a functional form for local volatility that is flexible enough to provide a good fit to the observed smiles or skews in option market data, nor simply a question of ensuring that the parameterization is sufficiently parsimonious to be calibrated with accuracy. An important factor when choosing a functional form for local volatility is that it should reflect what we believe about the underlying price process: what is the risk neutral density of a price process with a given functional form for the local volatility, and does this density have appealing properties?

3 See also Breeden and Litzenberger (1978), Rubinstein (1994).
4 In the presence of the smile the option delta is $\DeltaBS + \text{vega } \partial \sigma/\partial S$ where $\DeltaBS$ is the Black-Scholes delta. Hence when hedging options it is important to have some idea of the movements in the smile as the underlying changes.
Whilst the price process can always be specified as a Brownian with local volatility given by the calibrated chosen parametric form, it is difficult– if not impossible – to analyse the properties of the risk neutral price density. In general there will be no tractable functional form for the risk neutral density and, furthermore, model prices and hedge ratios may not be given in closed form.

However, Ritchey (1990) and Melick and Thomas (1997) introduced a finite normal variance mixture model for pricing options, and the finite normal mixture framework has since found many applications in finance – see Bingham and Keisel (2002) for a survey. One particularly important application is that the local volatility can be linked to an analytically tractable price density, a result that was first stated by Brigo and Mercurio (2000). Consequently Brigo and Mercurio (2001a) prove that if the risk neutral price density is a lognormal variance mixture, with the volatility in each lognormal density being a deterministic function of time, then the price process will be a Brownian motion with a local volatility given by the volatility of a lognormal variance mixture.\footnote{The condition for their result is that the mixing law for the risk neutral density must be finite and discrete.} Brigo, Mercurio and Sartorelli (2002) then extended this result to general lognormal mixtures, appropriate when there is skew in the log price density, and also prove the converse result, that if the local volatility is given as a weighted sum of average volatilities of deterministic volatility processes, the log price density will be a normal mixture. They have named this model the ‘Normal Mixture Diffusion’ (NMD) model.

The NMD model can been as an extension of the Black-Scholes model where the volatility is not constant, but instead there are a finite number of continuous and bounded deterministic volatility processes; and price densities are not lognormal, but lognormal mixtures with a fixed mixing law. It is therefore not a stochastic volatility model, but a local volatility model with some very tractable properties. In particular, risk neutral option prices are just a weighted average of Black-Scholes prices and option sensitivities will also be weighted averages of Black-Scholes sensitivities. Moreover, the model will fit market prices almost exactly when several parameterized volatility processes are assumed; in this way it has important applications to pricing path dependent options. On the other hand, with an elementary parameterization the calibration is more likely to be robust, so a more parsimonious parameterization of the NMD model holds many attractions for hedging purposes. By linking local volatility to the price process, the NMD model has revived the literature on local volatility models and this paper now continues this line of research.
3. The NMD Model

Suppose that there are a fixed, finite number of continuous and bounded deterministic volatility processes \( \sigma_1(t), \sigma_2(t), \ldots, \sigma_m(t) \). At any point in time \( t \) the \( t \)-period volatility associated with the \( i \)-th instantaneous volatility is denoted \( \sqrt{v_i(t)} \) where

\[
v_i(t) = \int_0^t \sigma_i^2(s) ds
\]

Denote by \( X(t) \) the log price of an asset, such as an equity or index or exchange rate. So \( X(t) \) is a random variable with some probability measure which, without loss of generality, can be assumed to be the risk neutral measure.\(^6\)

Assume that the dynamics of the log price process follow a diffusion with local volatility:

\[
dX = \mu dt + \sigma(X, t) dB
\]

where \( B \) is a Brownian motion, and \( \mu \) is a constant. Furthermore, assume that \( X \) has a normal mixture risk neutral density at every time \( t \) given by:

\[
f_i(x) = \sum_{i=1}^{m} \lambda_i \phi(x; \mu t, \nu_i(t)) = \sum_{i=1}^{m} \lambda_i \phi_{i,j}(x)
\]

where \( \phi \) denotes the normal density function and \( \sum_{i=1}^{m} \lambda_i = 1 \). Then Brigo and Mercurio (2001a) prove that the local volatility in (1) that is consistent with the price densities (2) is given by:

\[
\sigma(x, t) = \sqrt{\sum_{i=1}^{m} \lambda^*_i \sigma_i^2(t)}
\]

where

\[
\lambda^*_i = \lambda_i \phi_{i,j}(x) / f_i(x)
\]

Since market completeness is preserved in this framework, arbitrage-free pricing of standard European options on \( X \) is straightforward. Absence of arbitrage implies that the option price is the discounted expectation of the pay-off under the risk neutral density (2). The simple form of this

---

\(^6\) Note that the use of the logarithm means that results here will be presented in terms of normal mixture log price densities and arithmetic diffusions rather than lognormal mixture price densities and geometric diffusions. However the mixing law is invariant, since when \( X = \ln S \) the price density \( g(s) = (1/s) f(lns) \) so if \( f(.) \) is a normal mixture with mixing law \( [\lambda_i, \ldots, \lambda_m] \), then \( g(.) \) will be a lognormal mixture with the same mixing law.
density allows this expectation to be expressed as a weighted average of expectations under normal densities. That is, the normal mixture option price will be a weighted average of Black-Scholes option prices. For a European option with strike $K$ and maturity $T$ the normal mixture option price is given by:

$$NM(x, K, T) = \sum_{i=1}^{m} \lambda_i BS(x, K, T, \eta_i(T))$$

(4)

where $\eta_i(T) = \sqrt{\nu_i(T)}$. Note that hedging with European options is also straightforward in the NMD framework, since the expression (4) allows a simple representation of the option sensitivities in terms of the Black-Scholes sensitivities.

The basis of the model calibration is to minimize some distance metric between model prices and current market prices of simple European options. But before this can be done it is first necessary to fix the number of instantaneous volatility processes $m$ and parameterize the volatilities $\eta_i(t)$ for $i = 1, 2, ..., m$. In the absence of further model structure these decisions are fairly arbitrary. One factor to take into account is the large number of model parameters: In addition to $m - 1$ weights $\lambda_1, \lambda_2, ..., \lambda_{m-1}$ [with $\lambda_m = 1 - (\lambda_1 + \lambda_2 + ... + \lambda_{m-1})$] each of the $m$ volatility processes could have many parameters. Therefore, to reduce the number of parameters, Brigo and Mercurio (2000) have suggested setting $m = 2$ or $3$, and assuming that $\eta_i(t) = c_i$ (a constant) for all $t$. In that case, the densities $f_i(x)$ will be independent of $t$, but this is inconsistent with a term structure of kurtosis observed in the market. If the number of normal densities in the mixture does not vary over time, and neither do their weights in the mixture, the only way to model variation in kurtosis over time would be to build this into the parameterization of the volatilities.

4. Extending the NMD Model to the Short-Term Smile

To calibrate the NMD model using (4) one has first to make some assumptions about the number $m$ of normal densities in the mixture and the behaviour of the average volatilities $\eta_i(t)$ for $i = 1, 2, ..., m$. In the absence of further structure, these assumptions are quite arbitrary. We now propose some additional, fundamental structure for the NMD model that will determine both the number of volatility processes and the behaviour of the average volatilities. By restricting the values for each volatility process on any time interval $\Delta t$, and linking the mixing law to the values of the volatility processes, we derive a parsimonious parameterization of the NMD model that captures
the term structure of kurtosis in price densities, and the short-term smile effect, in a tractable manner. We make two assumptions:

*Assumption 1: The deterministic volatility processes \( \sigma_1(t), \sigma_2(t), \ldots, \sigma_m(t) \) are piecewise constant over a certain time interval \( \Delta t \) and each takes a value from the set \([s_1, \ldots, s_d]\) in every time interval, with \( d \leq m \).

This assumption implies a constant, bounded volatility over each interval \( \Delta t \). It is not a very restricting assumption, since the time interval can be made as short as you wish.\(^7\) The number \( m \) of volatility processes in the model and the average volatility are determined by the maximum maturity of the options on \( X \) that are to be priced and/or hedged and the length of the basic time interval \( \Delta t \).\(^8\) For example, suppose that there are only two possible volatility values, so the value of each of the volatility processes can be either high \((s_1 = \sigma_H)\) or low \((s_2 = \sigma_L)\) in each time interval of length \( \Delta t \). If the maximum option maturity is \( N\Delta t \) then \( m = 2^N \). More generally, with \( d \) distinct volatility values, \( m = d^N \).

Figure 1 depicts this assumption when there are only two volatility values \((\sigma_H \text{ and } \sigma_L)\) and the maximum maturity is \( 3\Delta t \). In this case there are \( m = 8 \) distinct deterministic volatility processes, marked in different colors on the figure. In general, the assumption of a finite number of possible volatility values in each time interval \( \Delta t \) allows one to enumerate also the number of distinct total variances over each interval \( n\Delta t \). For example, when there are only two possible volatility values (as in figure 1) the number of distinct values for \( v(n\Delta t) \) is \( n + 1 \) [for \( n = 1, 2, \ldots, N \)]. There are only four distinct values for \( v(3\Delta t) \), only three distinct values for \( v(2\Delta t) \), and only two distinct values for \( v(\Delta t) \). These values are shown in the three centre columns of Table 1.

The second assumption concerns the mixing law \([\lambda_1, \lambda_2, \ldots, \lambda_m]\) for the normal densities \( \phi_1(x), \phi_2(x), \ldots, \phi_m(x) \) in the densities \( f_t(x) \) of \( X \):

---

\(^7\) In fact, its choice should depend on the kurtosis term structure of \( X \): If daily or intra-day changes in \( X \) have high excess kurtosis but this excess kurtosis disappears after a week or so, \( \Delta t \) might be taken to be 1 day. But if excess kurtosis is still strong in weekly data, \( \Delta t \) could be 1 week.

\(^8\) If the maximum maturity of options on the underlying is 6-months (26 weeks) and the basic time interval \( \Delta t = 1 \text{ week} \), there will be \( m = 2^{26} \) different volatility processes in the NMD model. Note that, even in this most simple case, \( m \) is much greater than the two or three distinct volatility processes that were previously used by Brigo and Mercurio (2000, 2001a,b, 2002) when calibrating the NMD model.
Assumption 2: For each $i$, set $\lambda_i = \theta_1^{n_{1,i}} \theta_2^{n_{2,i}} \ldots \theta_d^{n_{d,i}}$ where $n_{i,j}$ is the number of intervals $\Delta t$ in which the $i$th volatility process takes value $s_j$. The new model parameters are such that $\theta_1 + \theta_2 + \ldots + \theta_d = 1$.

To motivate this assumption, consider again the case $d = 2$ and the 3-period model depicted in Figure 1. For brevity, write $\theta_1 = \lambda$ and $\theta_2 = 1-\lambda$, so the risk neutral density $f_t(x)$ of $X$ at any point in time $t$ has only three parameters: $\lambda$, $\sigma_H$ and $\sigma_L$. The mixing law $\lambda_1, \lambda_2, \ldots, \lambda_8$ applied to the eight volatility processes $\sigma_1(t), \sigma_2(t), \ldots, \sigma_8(t)$ is given in the right hand column of Table 1. Although at any point in time $t$ the density $f_t(x)$ will be a weighted sum of the same number of normal densities, they are not all distinct. For example, in the 3-period process depicted in Figure 1, there are only four distinct normal densities in the density at time $3 \Delta t$, only three distinct normal densities in the density at time $2 \Delta t$, and only two distinct normal densities in the density at time $\Delta t$. The effective weights on these densities in the mixing law are obtained by summing the $\lambda_i$ in the right hand column of Table 1 that are relevant to each variance.

In this way we see that assumption 2 implies that the mixing law at maturity $n \Delta t$ is the well-known binomial density $B(n, \lambda)$. Specifically, denote by $v_{j}(n \Delta t)$ the $n + 1$ distinct variances in the density at time $n \Delta t$. Then for $j = 1, 2, \ldots, n + 1$ we have:

$$v_{j}(n \Delta t; \sigma_H, \sigma_L) = (n - j + 1) \sigma_H^2 + (j - 1)\sigma_L^2$$  \hspace{1cm} (5)

with corresponding weight in the normal mixture:

$$w_{j}(n \Delta t; \lambda) = \frac{n!}{(j - 1)!(n - j + 1)!} \lambda^{n-j+1} (1 - \lambda)^{j-1}$$  \hspace{1cm} (6)

The binomial NMD option price for an option of maturity $n \Delta t$ with strike $K_{n,s}$ is therefore:

$$NM(x, K_{n,s}, n \Delta t; \lambda, \sigma_H, \sigma_L) = \sum_{j=1}^{n+1} w_{j}(n \Delta t; \lambda, \sigma_H, \sigma_L) BS(x, K_{n,s}, n \Delta t, \eta_{j}(n \Delta t; \sigma_H, \sigma_L), n \Delta t, \sigma_H, \sigma_L)$$  \hspace{1cm} (7)

where $\eta_{j}(n \Delta t; \sigma_H, \sigma_L) = \sqrt{v_{j}(n \Delta t; \sigma_H, \sigma_L)/n}$, and $v_{j}(n \Delta t; \sigma_H, \sigma_L)$ and $w_{j}(n \Delta t; \lambda)$ are given by (5) and (6).

The argument above can be extended to more than two volatility values in a straightforward manner. In general the mixing law, which determines the effective weight on each distinct normal density at time $n \Delta t$ is the multinomial density, from the expansion of $(\theta_1 + \theta_2 + \ldots + \theta_d)^n$. Again
the number of normal densities in the mixture will increase with maturity, giving the short-term smile effect that we seek. However, the binomial, and more generally the multinomial NMD model has no volatility term structure, as can easily be verified. For example, in the trinomial model the variance at $\Delta t$ is:

$$\theta_1 \sigma_H^2 + \theta_2 \sigma_M^2 + \theta_3 \sigma_L^2$$

which is the same as the variance at $2\Delta t$:

$$\theta_1 \sigma_H^2 + \theta_2 \sigma_M^2 + \theta_3 \sigma_L^2 + \theta_1 \theta_2 (\sigma_H^2 + \sigma_M^2) + \theta_1 \theta_3 (\sigma_H^2 + \sigma_L^2) + \theta_2 \theta_3 (\sigma_M^2 + \sigma_L^2)$$

The extension of the NMD model that we have described here is, therefore, a pure smile effect model. However, to model the volatility term structure, time variation into the possible values $[s_1, \ldots, s_d]$ of the volatility processes could be introduced.

5. Application to Currency Options

The binomial NMD model has been applied to Euro – USdollar options of all available strikes and maturities on 28/06/2002. Thirty-three prices at three different maturities (of two, six and eleven weeks) were available at eleven strikes (between 90 and 110). The spot rate was 99.07, the interbank rate for the Euro was taken as 3.86% and that for the US dollar was 1.71%.\(^9\) We chose to minimize the following objective function:\(^10\)

$$\sum_{i=1}^{k} \sum_{i=1}^{11} \gamma_{ij} (\sigma_{nm,ij} - \sigma_{m,ij})^2$$

where the gamma $\gamma_{ij}$ denotes the gamma of an option with the $i^{th}$ strike and $j^{th}$ maturity, $\sigma_{nm,ij}$ denotes the Black-Scholes implied volatility of the normal mixture model price that is obtained by backing out the implied volatility from the normal mixture model price, and $\sigma_{m,ij}$ denotes the Black-Scholes implied volatility of the market price of the option. In the following we take $\Delta t = 1$ week and $k = \text{either} 1 \text{ or } 3$. That is, we shall use (8) to calibrate to each single smile at a fixed maturity, as well as to the whole smile surface.

The optimization problem belongs to the category of non-linear, multi-dimensional constrained minimisation. Such problems are difficult because of the need to keep the solution within a boundary which is determined by constraints. Several algorithms were tested which were either

\(^9\) Note that we have assumed constant interest rates at this point.

\(^10\) The weighting of the squared volatility difference by gamma has the effect of giving more weight to more certain option prices. The gamma is greatest for at-the-money short-dated options, and decreases with both maturity and moneyness, but of course, other calibration objectives could be used.
too slow, or lacked robustness. The fastest and most reliable algorithm that has been tested with a variety of data sets appears to be the Downhill Simplex Method.\textsuperscript{11}

First consider the calibration of $\sigma_H$, $\sigma_L$ and $\lambda$ to all options of a fixed maturity, first calibrating a mixture of 3 normal densities to the smile at 2 weeks, then calibrating a mixture of 7 normal densities to the smile at 6 weeks, and finally calibrating a mixture of 12 normal densities to the smile at 11 weeks. The results are shown in Table 2. We see that the log price density calibrated on the 6 week smile is not the same as the density that is inferred at 6 weeks from calibration to the 2-week smile. The same remark applies to the log price density at 11 weeks. When longer-term smiles are inferred from the 2-week smile parameters, the excess kurtosis in the log price densities decreases with maturity: this is exactly the short term smile effect that, by the central limit theorem, decreases fairly rapidly with maturity.

[Table 2 and Figure 2 here]

The excess kurtosis that can attributed to the short-term smile effect (that is, the excess kurtosis inferred from the 2-week price density parameters) is 1.50 at 6 weeks and 0.82 at 11 weeks. By the 11 week maturity (with 12 normal densities in the mixture) the price density inferred from the 2-week smile parameters is near to normal. On the other hand, the excess kurtosis in the smile, including the longer-term smile effects, may be estimated by calibrating parameters directly on the individual smiles. Although a model is used, the fitted smile fits the observed market implied volatilities very closely, as shown in Figure 2. Thus it seems reasonable to infer that the model implied excess kurtosis is an accurate estimate of the total excess kurtosis in the price density. From Table 2, these estimates are 6.74 at 6 weeks and 4.05 at 11 weeks. We conclude that only a

\textsuperscript{11} Powell’s Set Direction Method transforms the problem to one dimension and then applies successive line minimisations to find local minima. The problem with this method is that it is extremely slow. Moreover the method requires as inputs a set of directions where it will search for the minima. If the set of directions defined is not a “longsighted” one, then the algorithm does not converge. The Broyden-Fletcher-Goldfarb-Schanno (BFGS) algorithm is a variable metric or “quasi-Newton” method that calculates the partial derivatives with respect to all the variables, looks for the “steepest” gradient and then employ a line minimisation to that direction. This method very often gave infeasible solutions, i.e. $\lambda$ greater than 1 or negative volatilities.
small part of the smile at longer maturities than 2 weeks can be explained by the binomial NMD model.\footnote{Attempts to calibrate the two volatility state model to all thirty-three option prices simultaneously [that is, with $k = 3$ in (8)] led to disappointing results. There was too much excess kurtosis in the fitted smile at 2 weeks (9.53), and too little excess kurtosis at 11 weeks (1.78).}

6. Modelling Longer-term Smile Effects with Stochastic $\lambda$

Above we have noted a limitation of the binomial NMD local volatility model: the log price density converges too rapidly to a normal density as the maturity increases. In this section we introduce a volatility uncertainty to the model by making $\lambda$ stochastic. That is, we introduce an uncertainty on which local volatility surface will apply to the price process, and we show that this construction will prevent the price density from converging to a normal density as the maturity increases.

Assume that $\lambda$ is a Bernoulli variable with probability $p$ on $\lambda_H$ and probability $(1-p)$ on $\lambda_L$. Thus at time $t = 0$, the trader perceives two possible local volatility surfaces; a high volatility surface with probability $p$ and a low volatility surface with probability $(1-p)$. This additional uncertainty introduces a market incompleteness, so there is no unique ‘risk neutral’ option price. However, for liquid options the price differences arising from differences in risk premia should be small, so market prices should be close to the ‘risk neutral’ option prices, which in this model are given by

$$p \, \text{NM}(x, K_{n,s}, n\Delta t; \lambda_H, \sigma_H, \sigma_L) + (1-p) \, \text{NM}(x, K_{n,s}, n\Delta t; \lambda_L, \sigma_H, \sigma_L)$$

with $\text{NM}(x, K_{n,s}, n\Delta t; \lambda, \sigma_H, \sigma_L)$ given by (5) – (7). As before, the model calibration will equate these to the observed market prices, using the favoured calibration objective. The option sensitivities will be simple weighted averages of the Black-Scholes sensitivities, but it is important to note that there will be residual hedging uncertainty because of the market incompleteness arising from the uncertainty surrounding the local volatility surface.

Table 3 reports the results of calibrating this stochastic $\lambda$ model to all 33 currency option implied volatilities simultaneously. At each maturity the log price densities are a weighted average of two binomial NMD log price densities, one fitted with $\lambda_H$, $\sigma_H$, and $\sigma_L$ and the other fitted with $\lambda_L$, $\sigma_H$, and $\sigma_L$. These two log price densities have quite different volatilities. For example, at
2 weeks the annual volatility of the density calibrated with $\lambda_H$ is 26.87%, and the annual volatility of the density calibrated with $\lambda_L$ density is 9.25%. The difference in standard deviations of the log price densities, which becomes more pronounced with maturity, means that when the two densities are mixed there is considerable excess kurtosis. Even at the longer maturities where the excess kurtosis in each of the binomial NMD models has virtually disappeared (so we are taking a mixture of two almost normal densities), because of the difference in their standard deviations, there is still a substantial excess kurtosis in the mixture log price density.

The excess kurtosis estimated from the stochastic $\lambda$ model compares well with the estimated excess kurtosis from the smile that was evaluated in Table 2 by fitting a deterministic $\lambda$ model to each maturity separately. Table 3 gives an estimated excess kurtosis with the stochastic $\lambda$ binomial NMD model of 3.74 at 6 weeks and 3.63 at 11 weeks. Figure 3 compares the market smiles and the model smiles that are fitted with the stochastic $\lambda$ model at each maturity. Clearly the model does not explain all the observed excess kurtosis in this smile – stochastic interest rates, liquidity, and market imperfections also affect currency option prices at longer maturities – but the entire fitted smile surface is based on only five parameters, and should therefore be relatively stable over time.

[Figure 3 here]

7. Summary and Conclusions

This paper began by placing additional structure on the NMD model of Brigo and Mercurio (2000, 2001a,b, 2002) and Brigo, Mercurio and Sartorelli (2002) by assuming the finite number of deterministic volatility processes are piecewise constant over time and are limited to a small number of values. This assumption induces an elegant and parsimonious parameterization where the mixing law for the lognormal mixture price density is multinomial and the term structure for the excess kurtosis in the risk neutral price density decreases with maturity. Consequently only short-term smile effects are explained. However, marked smile effects that do not decrease with maturity are often observed in implied volatilities, and in order to model a longer-term smile effect within the framework, we have introduced uncertainty over the local volatility surface by making the weights in mixing law stochastic.
A simple binomial NMD model has been implemented in this paper, where the mixing law coefficient $\lambda$ is (a) fixed and (b) stochastic with $\lambda$ being a simple Bernoulli variable. In case (a) the very parsimonious model has just three parameters but it only explains a short-term smile effect; in case (b) the five-parameter explains the persistence of smile effects into longer maturities.

The model is parsimonious, intuitive, and easy to calibrate to the observed market data. However, it does not provide an exact fit to market prices and, unless the values of the volatility processes are allowed to vary over time, there will be no volatility term structure in the model. However, by capturing both short and longer term smile effects with very few parameters, the new approach to smile modelling that has been introduced in this paper should have great potential for hedging. Current research is now focusing on extending the model in several ways: in particular to introduce a volatility term structure, and to capture a more realistic volatility uncertainty (so that $\lambda$ is not completely determined by a single ‘coin flip’ at time $t = 0$, but instead $\lambda$ is binomially distributed with a fixed parameter $p$ over the whole surface). Note that no more parameters need to be introduced in this second extension. The main application of these models should be their hedging performance, and this is the main focus of our current research.
References


Figure 1: Volatility Tree \(d = 2\) and \(N = 3\)
Figure 2: Market Smile and Fitted Smile at 6 Week Maturity

Deterministic Lambda Model

Table 1: Volatilities and Mixing Law \([d = 2 \text{ and } N = 3]\)

<table>
<thead>
<tr>
<th>Volatility</th>
<th>(v_i(\Delta t))</th>
<th>(v_i(2\Delta t))</th>
<th>(v_i(3\Delta t))</th>
<th>(\lambda_i)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(\sigma_1)</td>
<td>(\sigma_H^2)</td>
<td>(2\sigma_H^2)</td>
<td>(3\sigma_H^2)</td>
<td>(\lambda^3)</td>
</tr>
<tr>
<td>(\sigma_2)</td>
<td>(\sigma_H^2)</td>
<td>(2\sigma_H^2)</td>
<td>(2\sigma_H^2 + \sigma_L^2)</td>
<td>(\lambda^2(1-\lambda))</td>
</tr>
<tr>
<td>(\sigma_3)</td>
<td>(\sigma_H^2)</td>
<td>(\sigma_H^2 + \sigma_L^2)</td>
<td>(2\sigma_H^2 + \sigma_L^2)</td>
<td>(\lambda^2(1-\lambda))</td>
</tr>
<tr>
<td>(\sigma_4)</td>
<td>(\sigma_H^2)</td>
<td>(\sigma_H^2 + \sigma_L^2)</td>
<td>(\sigma_H^2 + 2\sigma_L^2)</td>
<td>(\lambda(1-\lambda)^2)</td>
</tr>
<tr>
<td>(\sigma_5)</td>
<td>(\sigma_L^2)</td>
<td>(\sigma_H^2 + \sigma_L^2)</td>
<td>(2\sigma_H^2 + \sigma_L^2)</td>
<td>(\lambda^2(1-\lambda))</td>
</tr>
<tr>
<td>(\sigma_6)</td>
<td>(\sigma_L^2)</td>
<td>(\sigma_H^2 + \sigma_L^2)</td>
<td>(\sigma_H^2 + 2\sigma_L^2)</td>
<td>(\lambda(1-\lambda)^2)</td>
</tr>
<tr>
<td>(\sigma_7)</td>
<td>(\sigma_L^2)</td>
<td>(2\sigma_L^2)</td>
<td>(\sigma_H^2 + 2\sigma_L^2)</td>
<td>(\lambda(1-\lambda)^2)</td>
</tr>
<tr>
<td>(\sigma_8)</td>
<td>(\sigma_L^2)</td>
<td>(2\sigma_L^2)</td>
<td>(3\sigma_L^2)</td>
<td>((1-\lambda)^3)</td>
</tr>
</tbody>
</table>
Figure 3: Market Smiles and Fitted Smiles with the Stochastic Lambda Model
### Table 2: Comparison of Log Price Densities Inferred from 2 Week Fitted Smile and Fitted to Each Smile

<table>
<thead>
<tr>
<th></th>
<th>2 weeks</th>
<th></th>
<th>6 weeks</th>
<th></th>
<th>11 weeks</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>$\lambda$</td>
<td>7.89%</td>
<td>3.48%</td>
<td>7.89%</td>
<td>5.66%</td>
<td>7.89%</td>
<td></td>
</tr>
<tr>
<td>$\sigma_H$</td>
<td>41.24%</td>
<td>78.60%</td>
<td>41.24%</td>
<td>68.62%</td>
<td>41.24%</td>
<td></td>
</tr>
<tr>
<td>$\sigma_L$</td>
<td>11.01%</td>
<td>9.59%</td>
<td>11.01%</td>
<td>3.99%</td>
<td>11.01%</td>
<td></td>
</tr>
<tr>
<td></td>
<td>weight</td>
<td>variance</td>
<td>weight</td>
<td>variance</td>
<td>weight</td>
<td>variance</td>
</tr>
<tr>
<td>w1</td>
<td>0.006225</td>
<td>0.170058</td>
<td>0.000000</td>
<td>0.617773</td>
<td>0.000000</td>
<td>0.170058</td>
</tr>
<tr>
<td>w2</td>
<td>0.145353</td>
<td>0.091088</td>
<td>0.000000</td>
<td>0.516342</td>
<td>0.000017</td>
<td>0.143734</td>
</tr>
<tr>
<td>w3</td>
<td>0.848422</td>
<td>0.012118</td>
<td>0.000021</td>
<td>0.414912</td>
<td>0.000493</td>
<td>0.117411</td>
</tr>
<tr>
<td>w4</td>
<td>0.000760</td>
<td>0.313482</td>
<td>0.007677</td>
<td>0.091088</td>
<td>0.000000</td>
<td>0.342873</td>
</tr>
<tr>
<td>w5</td>
<td>0.015799</td>
<td>0.212051</td>
<td>0.067218</td>
<td>0.064764</td>
<td>0.000000</td>
<td>0.300213</td>
</tr>
<tr>
<td>w6</td>
<td>0.175075</td>
<td>0.110621</td>
<td>0.313883</td>
<td>0.038441</td>
<td>0.000011</td>
<td>0.257553</td>
</tr>
<tr>
<td>w7</td>
<td>0.808345</td>
<td>0.009190</td>
<td>0.610711</td>
<td>0.012118</td>
<td>0.000189</td>
<td>0.214893</td>
</tr>
<tr>
<td>w8</td>
<td>0.002254</td>
<td>0.172233</td>
<td>0.007195</td>
<td>0.069550</td>
<td>0.001878</td>
<td>0.129573</td>
</tr>
<tr>
<td>w9</td>
<td>0.104315</td>
<td>0.086913</td>
<td>0.163414</td>
<td>0.040834</td>
<td>0.347690</td>
<td>0.044253</td>
</tr>
<tr>
<td>w10</td>
<td>0.526763</td>
<td>0.001593</td>
<td>0.404917</td>
<td>0.012118</td>
<td>0.526763</td>
<td>0.001593</td>
</tr>
<tr>
<td>Annual Vol</td>
<td>15.68%</td>
<td>17.43%</td>
<td>15.68%</td>
<td>16.78%</td>
<td>15.68%</td>
<td></td>
</tr>
<tr>
<td>XS</td>
<td>4.50</td>
<td>6.74</td>
<td>1.50</td>
<td>4.05</td>
<td>0.82</td>
<td></td>
</tr>
<tr>
<td>Kurtosis</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table 3: Stochastic Lambda Model Parameter Estimates

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Maturity (weeks)</th>
<th>Deterministic λ models</th>
<th>Stochastic λ model</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>λ</td>
<td>Volatility</td>
<td>XS kurtosis</td>
</tr>
<tr>
<td>p 28.12%</td>
<td>2</td>
<td>84.71%</td>
<td>26.87%</td>
</tr>
<tr>
<td>σ_H 28.98%</td>
<td>2</td>
<td>2.45%</td>
<td>9.25%</td>
</tr>
<tr>
<td>σ_L 8.17%</td>
<td>6</td>
<td>84.71%</td>
<td>26.87%</td>
</tr>
<tr>
<td>λ_H 84.71%</td>
<td>6</td>
<td>2.45%</td>
<td>9.25%</td>
</tr>
<tr>
<td>λ_L 2.45%</td>
<td>11</td>
<td>84.71%</td>
<td>26.87%</td>
</tr>
<tr>
<td></td>
<td>11</td>
<td>2.45%</td>
<td>9.25%</td>
</tr>
</tbody>
</table>