Beyond Black-Scholes

These notes provide an introduction to some of the models that have been proposed as replacements or extensions of the Black-Scholes GBM model for modeling stock prices. Many of these models or variations of them can also be used for pricing commodity and foreign exchange derivatives. These models include local and stochastic volatility models, as well as jump-diffusions and other models with jumps including those built from Levy processes. They are generally used for pricing exotic securities. It is worth emphasizing that the prices of exotics and other non-liquid securities are generally not available in the market-place and so models are needed in order to both price them and calculate their Greeks. This is in contrast to vanilla options where prices are available and easily seen in the market. For these more liquid options, we only need a model (typically Black-Scholes) and the volatility surface to calculate the Greeks.

In addition to describing some of these models, we will also provide an introduction to a commonly used fourier transform method for pricing vanilla options when analytic solutions are not available. We also briefly discuss *model risk*. In particular, we use the example of barrier option pricing to highlight the dangers of using just one model (that has been calibrated to the implied volatility surface) to price exotic derivatives. Instead we argue that it is probably best to price the exotic security using several different but plausible models. One could then take the resulting range of prices as a guide to the true or best price of the security.

1 Local Volatility

The GBM model for stock prices states that

$$dS_t \;=\; \mu S_t\, dt \;+\; \sigma S_t\, dW_t$$

where μ and σ are constants. Moreover, when pricing derivative securities with the cash account as numeraire, we know that $\mu = r - q$ where r is the risk-free interest rate and q is the dividend yield. This means that we have a single free parameter, σ, which we can fit to option prices or, equivalently, the volatility surface. It is not all surprising then that this exercise fails. As we saw before, the volatility surface is never flat so that a constant σ fails to re-produce market prices. This was particularly true after the crash of $'87$ when market participants began to correctly identify that lower strike options should be priced with a higher volatility, i.e. there should be a skew.

After this crash, researchers developed alternative models in an attempt to model the skew. While not the first[1] such model, the local volatility model is probability the simplest extension of Black-Scholes. It assumes that the stock's risk-neutral dynamics satisfy

$$dS_t \;=\; (r-q)S_t\, dt \;+\; \sigma_l(t,S_t)S_t\, dW_t \tag{1}$$

so that the instantaneous volatility, $\sigma_l(t,S_t)$, is now a function of time and stock price. The key result[2] of the local volatility framework is the *Dupire* formula that links the local volatilities, $\sigma_l(t,S_t)$, to the implied volatility surface.

[1] Earlier models included Merton's jump-diffusion model, the CEV model and Heston's stochastic volatility model. Indeed the first two of these models date from the 1970's.

[2] The local volatility framework was developed by Derman and Kani (1994) and in continuous time by Dupire (1994).

Theorem 1 (The Dupire Formula) *Let $C = C(K,T)$ be the price of a call option as a function of strike and time-to-maturity. Then the local volatility function satisfies*

$$\sigma_l^2(T,K) \;=\; \frac{\frac{\partial C}{\partial T} \;+\; (r-q)K\,\frac{\partial C}{\partial K} \;+\; qC}{\frac{K^2}{2}\,\frac{\partial^2 C}{\partial K^2}}. \tag{2}$$

Given the implied volatility surface we can easily compute the corresponding *call option price surface* which is the graph of $C(K,T)$ as a function of K and T. It is then clear from (2) that we need to take first and second derivatives of this latter surface with respect to strike and first derivatives with respect to time-to-maturity in order to compute the local volatilities. Calculating the local volatilities from (2) is therefore difficult and can be unstable as computing derivatives numerically can itself be very unstable. As a result, it is necessary to use a sufficiently smooth Black-Scholes implied volatility[3] surface when calculating local volatilities using (2).

Remark 1 *It is worth emphasizing that the local volatility model (1) with $\sigma_l(\cdot,\cdot)$ computed according to (2) is, by construction, a self-consistent model that is capable of producing the implied volatility surface observed in the market place.*

Local volatility is known to suffer from several weaknesses. For example, it leads to unreasonable skew dynamics and underestimates the volatility of volatility or "vol-of-vol". Moreover the Greeks that are calculated from a local volatility model are generally not consistent with what is observed empirically. Nevertheless, it is an interesting model from a theoretical viewpoint and is often used in practice for pricing barrier[4] options for example.

Gyöngy's Theorem

Gyöngy's Theorem is an important theoretical result that links local volatility models to other diffusion models that are also capable of generating the implied volatility surface. Consider a general n-dimensional Itô process, X_t, satisfying

$$dX_t \;=\; \alpha(t,\omega)\,dt \;+\; \beta(t,\omega)\,dW_t$$

where $\alpha(t,\omega)$ and $\beta(t,\omega)$ are $n\times 1$ and $n\times m$ adapted processes, respectively, and ω is a sample path of the m-dimensional Brownian motion, W_t. Then Gyöngy's Theorem states that there is a Markov process, Y_t, satisfying

$$dY_t \;=\; a(t,Y_t)\,dt \;+\; b(t,Y_t)\,dW_t$$

where X_t and Y_t have the same *marginal* distributions, i.e. X_t and Y_t have the same distribution for each t. Moreover, Y_t can be constructed by setting

$$\begin{aligned} a(t,y) &= \mathrm{E}_0\left[\alpha(t,\omega)|X_t = y\right] \quad \text{and} \\ b(t,y)b(t,y)^T &= \mathrm{E}_0\left[\beta(t,\omega)\beta^T(t,\omega)|X_t = y\right]. \end{aligned}$$

In a financial setting, X_t might represent the true risk-neutral dynamics of a particular security. Then $b(t,y)/y$ represents the local volatility function $\sigma_l(t,\cdot)$ in (1). Because X_t and Y_t have the same marginal distributions then we know (why?) that European option prices can be priced correctly if we assume the price dynamics are given by Y_t. In particular Y_t can produce the correct implied volatility surface. Moreover Gyöngy's Theorem therefore implies that the local volatility model of (1) is in some sense the simplest diffusion model capable of doing this, i.e. reproducing the implied volatility surface. Gyöngy's Theorem has been used recently to develop stochastic-local volatility models as well as approximation techniques for pricing various types of basket options.

[3] It is also possible to write the Dupire formula in terms of the implied volatilities rather than the call option prices. One can then work directly with the implied volatility surface to compute the local volatilities.

[4] While the Black-Scholes GBM framework can be used to barrier options analytically, it is well known that the Black-Scholes model is in fact a truly awful model for barrier options and that it should never be used in practice. As a result more sophisticated models and numerical methods such as PDE or Monte-Carlo methods are used. We will return to barrier options in Section 6 where we will use them to highlight the danger of using just one model to price exotic options.

2 Stochastic Volatility Models

The most well-known stochastic volatility model is due to Heston (1989). It is a two-factor model and assumes separate dynamics for both the stock price and instantaneous volatility. In particular, it assumes

$$dS_t = (r-q)S_t\,dt + \sqrt{\sigma_t}S_t\,dW_t^{(s)} \tag{3}$$
$$d\sigma_t = \kappa\,(\theta-\sigma_t)\,dt + \gamma\sqrt{\sigma_t}\,dW_t^{(vol)} \tag{4}$$

where $W_t^{(s)}$ and $W_t^{(vol)}$ are standard $\mathcal{Q}$-Brownian motions with constant correlation coefficient, ρ.

Exercise 1 *What sign you would expect ρ to have?*

Remark 2 *The volatility process in (4) is commonly used in interest rate modeling where it is known as the CIR[5] model. It has the property that the process will remain non-negative with probability one. For certain parameter combinations, it will always be strictly positive with probability one.*

Whereas the local volatility model is a complete model, Heston's stochastic volatility model is an incomplete[6] model. This should not be too surprising as there are two sources of uncertainty in the Heston model, $W_t^{(s)}$ and $W_t^{(vol)}$, but only one risky security and so not every security is replicable. Put another way, while the drift in (3) must be $r-q$ under any EMM with the cash account as numeraire, we could use Girsanov's Theorem to change the drift in (4) in infinitely many different ways without changing the drift in (3).

To see this let us first suppose that the P-dynamics of S_t and σ_t satisfy

$$dS_t = \mu_t S_t\,dt + \sqrt{\sigma_t}S_t\,dW_t^{(1)} \tag{5}$$
$$d\sigma_t = \nu_t\,dt + \gamma\sqrt{\sigma_t}\left(\rho dW_t^{(1)} + \sqrt{1-\rho^2}dW_t^{(2)}\right) \tag{6}$$

where μ_t and ν_t are some $\mathcal{F}_t$-adapted processes, and $W_t=(W_t^{(1)},W_t^{(2)})$ is a standard 2-dimensional P-Brownian motion. Let us now define

$$L_t := \exp\left(-\int_0^t \eta_s^T\,dW_s - \frac{1}{2}\int_0^t \eta_s^T\,\eta_s\,ds\right)$$

for $t\in[0,T]$ and where $\eta_t=(\eta_t^{(1)},\eta_t^{(2)})$ is a 2-dimensional adapted process. Then[7] Girsanov's Theorem implies $\widehat{W}_t := W_t + \int_0^t \eta_s ds$ is a standard 2-dimensional Q^η-Brownian motion where $dQ^\eta/dP = L_T$. In particular the Q^η-dynamics of S_t and σ_t satisfy

$$dS_t = \left(\mu_t - \sqrt{\sigma_t}\eta_t^{(1)}\right)S_t\,dt + \sqrt{\sigma_t}S_t\,d\widehat{W}_t^{(1)} \tag{7}$$
$$d\sigma_t = \left(\nu_t - \gamma\sqrt{\sigma_t}\rho\eta_t^{(1)} - \gamma\sqrt{\sigma_t}\sqrt{1-\rho^2}\eta_t^{(2)}\right)dt + \gamma\sqrt{\sigma_t}\left(\rho d\widehat{W}_t^{(1)} + \sqrt{1-\rho^2}d\widehat{W}_t^{(2)}\right). \tag{8}$$

In order for Q^η to be an EMM it is only necessary (why?) that

$$\mu_t - \sqrt{\sigma_t}\eta_t^{(1)} = r-q$$

and we are free to choose $\eta_t^{(2)}$ however we like. This means that there are infinitely many EMM's and so[8] the model is incomplete. In the Heston model we choose $\eta_t^{(2)}$ so that

$$\nu_t - \gamma\sqrt{\sigma_t}\rho\eta_t^{(1)} - \gamma\sqrt{\sigma_t}\sqrt{1-\rho^2}\eta_t^{(2)} = \kappa\,(\theta-\sigma_t)$$

[5] After Cox, Ingersoll and Ross (1985) who used this model for modeling the dynamics of the short interest rate.
[6] Assuming as usual that the stock and the cash-account are the only traded securities.
[7] we are assuming the necessary conditions to ensure that L_t is a martingale. Novikov's condition is sufficient.
[8] But we could make the model complete by introducing into the model another security whose price depends on σ_t.

is satisfied. We therefore recover (3) and (4) once we identify $\widehat{W}_t^{(1)}$ with $W_t^{(s)}$ and $W_t^{(vol)}$ with (via Levy's Theorem) $\rho\widehat{W}_t^{(1)} + \sqrt{1-\rho^2}\widehat{W}_t^{(2)}$. Note that we still have several free parameters which in practice we would determine by calibrating the model to the market prices of European options. This is the typical method of choosing an EMM in incomplete market models.

The pricing PDE that the price, $C(t, S_t, \sigma_t)$, of any derivative security must satisfy in Heston's model is given by

$$\frac{\partial C}{\partial t} + \frac{1}{2}\sigma S^2 \frac{\partial^2 C}{\partial S^2} + \rho\sigma\gamma S \frac{\partial^2 C}{\partial S \partial \sigma} + \frac{1}{2}\gamma^2 \sigma \frac{\partial^2 C}{\partial \sigma^2} + (r-q)S\frac{\partial C}{\partial S} + \kappa\,(\theta - \sigma)\frac{\partial C}{\partial \sigma} = rC. \tag{9}$$

Derivative prices can then be obtained by solving (9) subject to the relevant boundary conditions. Heston succeeded in using transform techniques for solving (9) in the case of European call and put options.

Exercise 2 *(***Feynman-Kac***) Use the Feynman-Kac Theorem to show that (9) is indeed the pricing PDE.* Hint*: Use the fact (why?) that*

$$M(t, S_t, \sigma_t) \;\; := \;\; e^{-rt}\,\mathrm{E}_t^{\mathcal{Q}}\left[e^{-r(T-t)}\;\; \textit{Option Payoff}\right] \;= e^{-rt}\;C(t, S_t, \sigma_t) \tag{10}$$

is a $\mathcal{Q}$-martingale.

Once the parameters of the Heston model have been identified via some calibration algorithm, pricing can be done by either solving (9) numerically or alternatively, solving (10) using Monte-Carlo[9] or transform techniques.

The Characteristic Function of the log-Stock Price in Heston's Model

For example it can be shown that the characteristic function of the *log stock price* at time t, $\phi(u;t)$, satisfies

$$\begin{aligned}
\phi(u;t) &= \mathrm{E}[\exp(iulog(S_t) \mid S_0, \sigma_0)] \\
&= \exp(iu(log(S0) + (r-q)t)) \\
&\quad \times\; \exp(\theta\kappa\gamma^{-2}((\kappa - \rho\gamma ui - d)t - 2\log((1 - g\exp(-dt))/(1-g)))) \\
&\quad \times \exp(\sigma_0^2\gamma^{-2}(\kappa - \rho\gamma ui - d)(1-\exp(-dt))/(1 - g\exp(-dt)))
\end{aligned} \tag{11}$$

where $i = 0 + 1i$ is imaginary and

$$\begin{aligned}
d &= \sqrt{(\rho\gamma ui - \kappa)^2 + \gamma^2(iu + u^2)} \\
g &= (\kappa - \rho\gamma ui - d)/(\kappa - \rho\gamma ui + d).
\end{aligned}$$

While the expression for $\phi(u;t)$ is complicated, it is easy to code and it can then be used to price options as described for example in Section 5.

How Well Does Heston Capture the Skew?

An interesting question that arises is whether or not Heston's model accurately represents the dynamics of stock prices. This question is often reduced in practice to the less demanding question of how well the Heston model captures the volatility skew. By "capturing" the skew we have in mind the following: once the Heston model has been calibrated, then European option prices can be computed using numerical techniques such as Monte-Carlo, PDE or transform methods. The resulting option prices can then be used to determine the corresponding Black-Scholes implied volatilities. These volatilities can then be graphed to create the *model's* implied volatility surface which can then be compared to the *market's* implied volatility surface.

[9] Special care must be taken when simulating the dynamics in (3) and (4) as standard Euler-type simulation schemes do not always converge well.

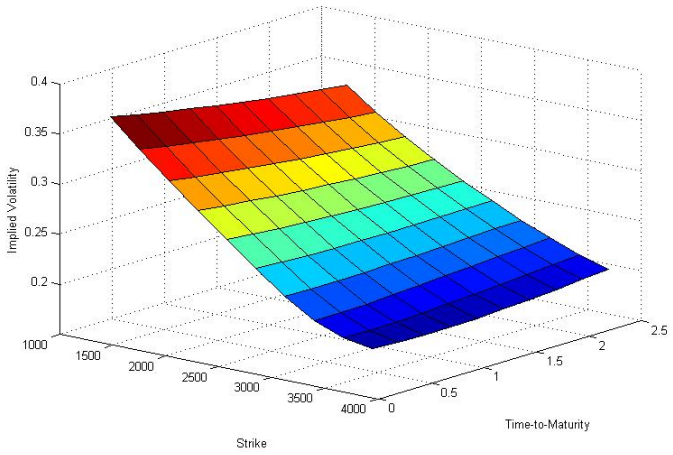


Figure 1: An Implied Volatility Surface under Heston's Stochastic Volatility Model

Figure 1 displays the implied volatility surface for the following choice of parameters: $r = .03$, $q = 0$, $\sigma_0 = \sqrt{.0654}$, $\gamma = .2928$, $\rho = -.7571$, $\kappa = .6067$ and $\theta = .0707$. Perhaps the most noticeable feature of this surface is the persistence of the skew for long-dated options. Indeed the Heston model generally captures longer-dated skew quite well but it typically struggles to capture the near term skew, particularly when the latter is very steep. The problem with a steep short-term skew is that any diffusion model will struggle to capture it as there is not enough time available for the stock price to diffuse sufficiently far from its current level. In order to solve this problem jumps are needed.

Note that some instruments can be priced analytically in Heston's model. For example, the price of the continuous-time version of a variance swap has a closed-form solution as the following example demonstrates.

Example 1 (Variance-swaps in Heston's model)

Let V_T denote the total variance in the Heston model from $t = 0$ to $t = T$. Then we know from (4) that

$$\begin{aligned} \mathrm{E}_0^{\mathcal{Q}}[V_T] &= \mathrm{E}_0^{\mathcal{Q}}\left[\int_0^T \sigma_t \, dt\right] \\ &= \int_0^T \mathrm{E}_0^{\mathcal{Q}}[\sigma_t] \, dt \qquad (12) \\ &= \int_0^T \left(e^{-\kappa t}(\sigma_0 - \theta) + \theta\right) \, dt \qquad (13) \\ &= \frac{1 - e^{-\kappa T}}{\kappa}(\sigma_0 - \theta) \; + \; \theta T \end{aligned}$$

so that the annualized variance is given by

$$\frac{\mathrm{E}_0^{\mathcal{Q}}[V_T]}{T} = \frac{1 - e^{-\kappa T}}{\kappa T}(\sigma_0 - \theta) \; + \; \theta.$$

∎

Exercise 3 *In the previous example, justify the step where we went from (12) to (13).* Hint*: Take expectations in (4) and then use the martingale property of stochastic integrals to eliminate the last term. You can then obtain a simple ODE for* $\mathrm{E}_0^{\mathcal{Q}}[\sigma_t]$.

Note that the fair price of a variance-swap in Heston's model does not depend on γ or ρ. This should not be too surprising.

---

3 Jump-Diffusion Models

Jump diffusion models for security pricing were first introduced by Merton (1975). Before we describe his model we first need some definitions.

Definition 1 *We say that J_t is a* **pure jump process** *if it is constant between jumps and is adapted and right-continuous.*

Typical examples of pure jump processes are Poisson processes and *compound* Poisson processes as described in the next example.

Example 2 (A Compound Poisson Process) Let N_t be a Poisson process with intensity λ. Then

$$X_t \;:=\; \sum_{i=0}^{N_t} Y_i$$

where the Y_i's are IID random variables is a compound Poisson process. It is easy to check that $\mathrm{E}[X_t] = \lambda t \mathrm{E}[Y]$ and that $\mathrm{Var}(Y_t) = \lambda t \mathrm{E}[Y^2]$. ∎

Definition 2 *A* **jump-diffusion** *or* **jump** *process is a process of the form*

$$\begin{aligned} X_t \;&=\; X_0 \;+\; \int_0^t \gamma_s \; dW_s \;+\; \int_0^t \theta_s \; ds \;+\; J_t \\ &=:\; X_t^c \;+\; J_t \end{aligned} \tag{14}$$

where J_t is a pure jump process and X_t^c is the continuous *part of X_t. Of course, γ_s and θ_s are adapted processes and W_s is a Brownian motion.*

Definition 3 *The* **stochastic integral** *of the process Φ_t with respect to a jump-diffusion, X_t, is*

$$\int_0^t \phi_s \; dX_s \;=\; \int_0^t \phi_s \gamma_s \; dW_s \;+\; \int_0^t \phi_s \theta_s \; ds \;+\; \sum_{0<s\le t} \phi_s \Delta J_s. \tag{15}$$

When we work with jump processes we will often want to insist that $\phi(s)$ be left-continuous or (almost equivalently), *predictable*. This is particularly true for financial applications where ϕ_s can be interpreted as a trading strategy. However, (15) is still defined if we only assume Φ_s is adapted.

Example 3 (Merton's Jump-Diffusion Model)

The first jump diffusion model to be proposed as a model of stock price dynamics was Merton's jump diffusion. This model assumes that the time t stock price, S_t, satisfies

$$S_t \;=\; S_0 \, e^{(\mu - \sigma^2/2)t + \sigma W_t} \prod_{i=1}^{N_t} Y_i \tag{16}$$

where N_t is a Poisson process with mean arrival rate λ, and the Y_i's are IID log-normal random variables with $\mu_y := \mathrm{E}[Y_i]$ for all i. The Poisson process and Brownian motions are independent processes. Each Y_i represents

the magnitude of the i^{th} jump. In between jumps, the stock price behaves like a regular GBM. If the dynamics in (16) are under an EMM, $\mathcal{Q}$, then μ, λ and the mean jump size are constrained in such a way that the $\mathcal{Q}$-expected rate of return must equal $r-q$. To be specific, if (16) reflects the risk-neutral dynamics of the stock price under an EMM, $\mathcal{Q}$, note that

$$\begin{aligned}
\mathrm{E}_0^{\mathcal{Q}}[S_t] &= S_0 e^{\mu t}\, \mathrm{E}_0^{\mathcal{Q}}\left[\prod_{i=1}^{N_t} Y_i\right] \\
&= S_0 e^{\mu t}\, \mathrm{E}_0^{\mathcal{Q}}\left[\mathrm{E}_0^{\mathcal{Q}}\left[\prod_{i=1}^{N_t} Y_i \mid N_t\right]\right] \\
&= S_0 e^{\mu t}\, \mathrm{E}_0^{\mathcal{Q}}\left[\mu_y^{N_t}\right] \\
&= S_0 e^{\mu t} \sum_{i=0}^{\infty} e^{-\lambda t}\frac{(\lambda t)^i}{i!}\mu_y^i \\
&= S_0 e^{\mu t + \lambda t(\mu_y - 1)}.
\end{aligned}$$

If $\mathcal{Q}$ is an EMM (with the cash account as numeraire) then the expected growth rate under $\mathcal{Q}$ must be $r-q$. This implies

$$\mu + \lambda(\mu_y - 1) = r - q \tag{17}$$

which is an equation in three unknowns. It therefore has infinitely many solutions and so we can conclude from the Second Fundamental Theorem of Asset Pricing that Merton's model is incomplete. Indeed this is true of almost all jump-diffusion models. We would like to be able to price European options in Merton's model and there are several ways to do this including Monte-Carlo simulation and Laplace or Fourier transform methods. We can also price these options, however, by expressing them as an infinitely weighted sum of Black-Scholes options prices. To see this, note that *conditional* on $N_T = n$ we can write

$$\begin{aligned}
S_T &= S_0 e^{(\mu-\sigma^2/2)T \,+\, \sigma W_T \,+\, \sum_{i=1}^{n} Z_i} \\
&=^{\text{dist}} S_0 e^{(\mu-\sigma^2/2)T \,+\, n\mu_z \,+\, \sqrt{\sigma^2 + n\sigma_z^2/T}\, W_T} \\
&= S_0 e^{\left(\mu + \frac{2n\mu_z + n\sigma_z^2}{2T} - \widehat{\sigma}_n^2/2\right)T \,+\, \widehat{\sigma}_n W_T}
\end{aligned} \tag{18}$$

where $Z_i := \log(Y_i) \sim N(\mu_z, \sigma_z^2)$ are IID, "$=^{\text{dist}}$" denotes "equal in distribution" and $\widehat{\sigma}_n := \sqrt{\sigma^2 + n\sigma_z^2/T}$. Conditional on $N_T = n$, we therefore see that the risk-neutral drift of S_t in (18) is given by

$$\begin{aligned}
\mu + \frac{2n\mu_z + n\sigma_z^2}{2T} &= r \;-\; q \;-\; \lambda(\mu_y - 1) \;+\; \frac{2n\mu_z + n\sigma_z^2}{2T} && (19)\\
&= r \;-\; q \;-\; \lambda(e^{\mu_z + \sigma_z^2/2} - 1) \;+\; n(\mu_z + \sigma_z^2/2)/T && (20)\\
&= r - \widehat{q}_n && (21)
\end{aligned}$$

where $\widehat{q}_n := q \;+\; \lambda(e^{\mu_z+\sigma_z^2/2} - 1) \;-\; n(\mu_z + \sigma_z^2/2)/T$. Note that (19) follows from (17) and (20) follows from the observation that $\mu_y = e^{\mu_z + \sigma_z^2/2}$.

We are now in a position to derive an expression for European call options in Merton's jump-diffusion model. We obtain

$$\begin{aligned}
\mathrm{E}_0^{\mathcal{Q}}[e^{-rT}(S_T - K)^+] &= \sum_{n=0}^{\infty} e^{-\lambda T}\frac{(\lambda T)^n}{n!}\mathrm{E}_0^{\mathcal{Q}}[e^{-rT}(S_T - K)^+ | N_T = n] \\
&= \sum_{n=0}^{\infty} e^{-\lambda T}\frac{(\lambda T)^n}{n!} C_{bs}(S_0, K, r, \widehat{q}_n, \widehat{\sigma}_n, T)
\end{aligned} \tag{22}$$

where C_{bs} is the usual Black-Scholes call option price and (22) follows from the log-normality of S_T conditional on $N_T = n$ and (21).

An interesting question to consider is how well Merton's jump-diffusion model can replicate the implied volatility surfaces that are typically observed in the market. Note that in contrast to the geometric Brownian motion

(GBM), we have five parameters[10], σ, λ, μ, μ_y and σ_y^2 and just one equation to satisfy, namely (17). We therefore have much more flexibility than GBM which can only achieve constant implied volatility surfaces.

Figure 2 displays the implied volatility surface under a Merton jump-diffusion model when $\sigma = 20\%$, $r = 2\%$, $q = 1\%$, $\lambda = 10\%$, $\mu_z = -.05$ and $\sigma_z = \sqrt{0.1}$. While other shapes are also possible by varying these

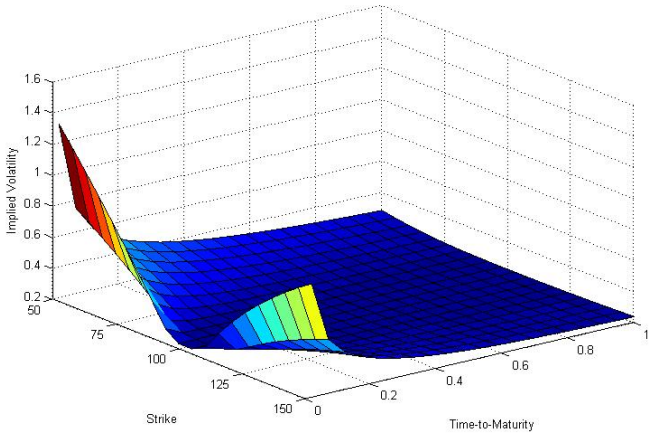


Figure 2: An Implied Volatility Surface under Merton's Jump Diffusion Model

parameters, Figure 2 demonstrates one of the principal weaknesses of Merton's jump-diffusion model, namely the rapid flattening of the volatility surface as time-to-maturity increases. For very short time-to-maturities, however, the model has no difficulty with producing a steep volatility skew. This is in contrast to stochastic volatility models which do not allow jumps. ∎

In recent years many other jump-diffusion models have been proposed and analyzed. For example, Duffie, Pan and Singleton (1998) introduced the class of *affine jump-diffusion processes*. The affine property of these properties has made them very amenable to pricing via transform methods. More recently, Kou (2002) developed the double-exponential jump-diffusion model where the jump-sizes have a double exponential distribution. The memoryless property of the exponential distribution affords this model considerable tractability. While introducing a new model is a fairly simple task, demonstrating that the model is *tractable* and of practical use is the real challenge.

The stochastic calculus we have studied for diffusion processes can be extended for jump-diffusions and indeed other stochastic processes, including Levy processes and more generally, *semi-martingales*[11]. We will go no further than stating Itô's Lemma for jump-diffusions.

Theorem 2 (Itô's Lemma for 1-Dimensional Jump-Diffusions)

Let X_t be a jump process and $f(x)$ a function for which $f'(x)$ and $f''(x)$ are defined and continuous. Then

$$f(X_t) \;=\; f(X_0) \;+\; \int_0^t f'(X_s)\, dX_s^c \;+\; \frac{1}{2}\int_0^t f''(X_s)\, dX_s^c dX_s^c \;+\; \sum_{0<s\le T} [f(X_s) - f(X_{s-})] \qquad (23)$$

where X_t^c denotes the continuous, i.e. non-jump, component of X_t, the summation in (23) is over the jump times of the process and X_{s-} is the value of X immediately before a jump at time s.

[10] We could use σ_z and $\sigma - z^2$ in place of μ_y and σ_y^2.
[11] We will not define semi-martingales in these notes.

Example 4 (Merton's Jump Diffusion Model)

Let $X_t := \log(S_t)$ where S_t is given by (16). Then $X_t = \log(S_0) + (\mu - \sigma^2/2)t + \sigma W_t + \sum_{i=1}^{N_t} \log(Y_i)$ and so

$$dX_t = (\mu - \sigma^2/2)\, dt + \sigma\, dW_t + \log(Y_i)\, dN_t.$$

Applying Itô's Lemma to $S_t = e^{X_t}$ we recover the dynamics for S_t and obtain[12]

$$\begin{aligned} dS_t &= S_t(\mu - \sigma^2/2)\, dt + S_t\sigma\, dW_t + \frac{1}{2}S_t\sigma^2\, dt + S_t(Y_t - 1)\, dN_t \\ &= S_t(\mu + (\mu_y - 1)\lambda)\, dt + S_t\sigma\, dW_t + [S_t(Y_t - 1)\, dN_t - S_t(\mu_y - 1)\lambda\, dt]. \end{aligned} \tag{24}$$

Referring to (24) note that the dW_t term and the term in the square brackets are both martingales. Therefore if (24) describes the risk-neutral dynamics of S_t is must be the case that the drift term equals $S_t(r - q)\, dt$ and so we obtain (17) once again. ∎

Example 5 (Shreve E.G. 11.5.2: Geometric Poisson Process)

Suppose S_t satisfies

$$S_t := S_0 \exp\left(N_t \log(\sigma + 1) - \lambda\sigma t\right) = S_0 e^{-\lambda\sigma t}(\sigma + 1)^{N_t}$$

where $\sigma > -1$ and N_t is again a Poisson process. Note that if $\sigma > 0$ then this process only jumps up and drifts down between jumps. If $-1 < \sigma < 0$ then the process only jumps down and drifts up between jumps. Let's apply Itô's Lemma to show that S_t is a martingale.

Define $X_t = N_t \log(\sigma + 1) - \lambda\sigma t$ so that $S_t = S_0 f(X_t)$ where $f(x) = \exp(x)$. Itô's Lemma now implies

$$\begin{aligned} S_t &= S_0 f(X_t) \\ &= S_0 f(X_0) - \lambda\sigma S_0 \int_0^t f'(X_u)\, du + S_0 \sum_{0<u\le t} [f(X_u) - f(X_{u-})] \\ &= S_0 - \lambda\sigma \int_0^t S_u\, du + \sum_{0<u\le t} [S_u - S_{u-}] \\ &= S_0 - \lambda\sigma \int_0^t S_u\, du + \sum_{0<u\le t} \sigma S_{u-} \\ &= S_0 - \lambda\sigma \int_0^t S_{u-}\, du + \int_0^t \sigma S_{u-}\, dN_u \\ &= S_0 + \sigma \int_0^t S_{u-}\, (dN_u - \lambda\, du). \end{aligned}$$

where we have used the fact that $S_u = (\sigma + 1)S_{u-}$ if a jump takes place at time u and where S_{u-} is the value of S immediately before that jump. Note that because we were able to write the jump in S at time u in terms of S_{u-}, we can write the SDE in differential form

$$dS_t = -\lambda\sigma S_t\, dt + \sigma S_{t-}\, dN_t = \sigma S_t\, dX_t$$

where $X_t := N_t - \lambda t$ is clearly a martingale. ∎

It is worth emphasizing that jump-models are almost invariably incomplete models with infinitely many EMM's. A particular EMM, $\mathcal{Q}$, is typically chosen (possibly from a parameterized subset of the possible EMMs) by some calibration algorithm.

[12] We could of course have written the dynamics for S_t directly using (16) but then we wouldn't have been able to practice using Itô's Lemma for jump-diffusions!

4 Levy Processes

Another class of models that has become more popular[13] in recent years is the class of *exponential Levy processes* and *time-changed exponential Levy processes*.

Definition 4 *A Levy process is any continuous-time process with stationary*[14] *and independent increments.*

The two most common examples of Levy processes are Brownian motion and the Poisson process. A Levy process with jumps can be of *infinite activity* so that it jumps infinitely often[15] in any finite time interval, or of *finite activity* so that it makes only finitely many jumps in any finite time interval. The most important result[16] for Levy processes is the famous *Levy-Khintchine formula* which describes the characteristic function of any Levy Process.

Theorem 3 *(***Levy-Khintchine Formula***)*

If X_t is a Levy process then its characteristic function satisfies

$$\mathbf{E}_0\left[e^{iuX_t}\right] = e^{tf(u)} \quad \textit{where} \tag{25}$$

$$f(u) = ibu - \frac{1}{2}au^2 + \int_{\mathbf{R}\setminus\{0\}} \left[e^{iuy} - 1 - iuy\,\mathbf{1}_{\{0<|y|<1\}}(y)\right]\,\nu(dy) \tag{26}$$

for some real b, real $a > 0$ and $\nu(\cdot)$ a Levy measure on $\mathbf{R}\setminus\{0\}$ so that $\int_{\mathbf{R}\setminus\{0\}} \min(1,\,|y|^2)\,\nu(dy) < \infty$.

The constant b plays the role of the drift and a the volatility of the Brownian motion. In fact if $\nu \equiv 0$, then X_t is simply a Brownian motion with drift b and volatility a, i.e. $X_t = bt + aW_t$ where W_t is a standard Brownian motion. The Levy measure $\nu(\cdot)$ controls the jumps of the Levy process.

Definition 5 *An exponential Levy process, S_t, satisfies $S_t = \exp(X_t)$ where X_t is a Levy process.*

The Merton jump-diffusion model is an example of an exponential[17] Levy process of finite activity. One of the weaknesses of an exponential Levy process is that it does not capture "volatility clustering", i.e. the tendency of high volatility periods to be followed by periods of high volatility, and low volatility periods to be followed by periods of low volatility. This is due to the stationary and independent increments assumption. Levy models with stochastic time, however, can capture volatility clustering.

Definition 6 *A* **subordinator** *is a non-negative almost surely non-decreasing Levy process.*

A subordinator can be used to change the "clock speed" or "calendar speed". More generally, if y_t is a positive process, then we can define our stochastic clock, Y_t, as

$$Y_t := \int_0^t y_s\,ds. \tag{27}$$

Using Y_t to measure time instead of the usual t, we can then model a security price process, S_t, as an exponential time-changed Levy process. We can write the $\mathcal{Q}$-dynamics for S_t as

$$S_t = S_0 e^{(r-q)t}\frac{e^{X_{Y_t}}}{\mathrm{E}_0^{\mathcal{Q}}\left[e^{X_{Y_t}}\,|y_0\right]} \tag{28}$$

[13] This class is popular in the academic literature at any rate. Some of these models are now being implemented by information providers such as Bloomberg, however, and internally in some financial institutions.

[14] X_t is a stationary process if the distribution of $X_{t+h} - X_t$ does not depend on t.

[15] Most of the jumps, however, will be very small.

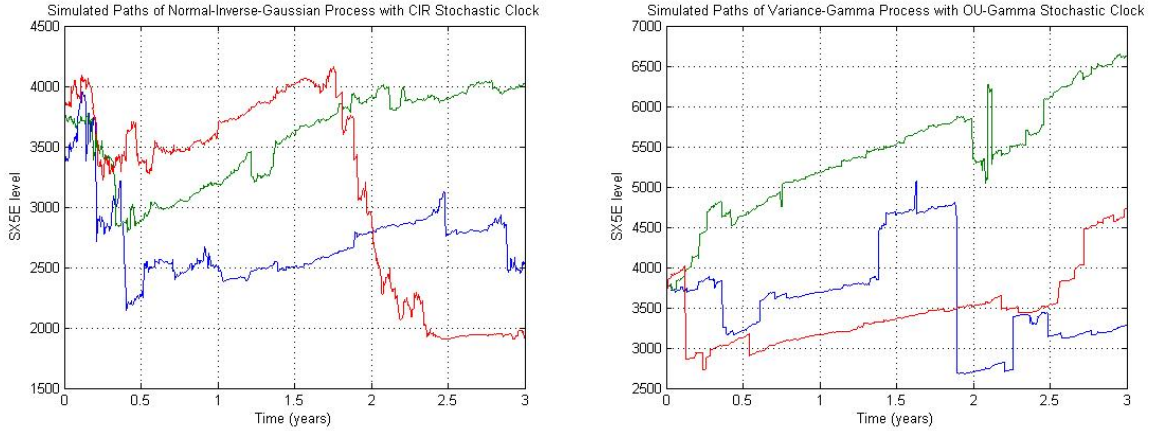
[16] We will have no need of this result beyond the fact that the characteristic functions of Levy processes are well understood. We state it here only because of it's general importance to the study of Levy processes.

[17] We take exponentials to guarantee the process remains non-negative thereby reflecting the limited liability of a shareholder.

where r is the constant risk-free rate, q the constant dividend yield and X_t the underlying Levy process. Note that stochastic volatility, i.e. volatility clustering, is captured using the stochastic time clock, Y_t. For example between times t and $t + \Delta t$ the clock of the Levy process, X_t, will have moved approximately $y_t \times \Delta t$ units. So when y_t is large the stochastic clock will have moved further. The increments of S_t will generally be more volatile as a result. Similarly when y_t is small, the stock price will display lower levels of volatility.

Remark 1: Note that if the subordinator, Y_t, is a jump process, then S_t can jump even if the process X_t cannot jump itself. (If Y_t is defined as an integral as in (27) then it will never jump.)

Remark 2: Note that in (28) the dynamics of S_t are $\mathcal{Q}$-dynamics corresponding to the cash account as numeraire. Why is this the case? (This is typical of how incomplete markets are often modeled: we directly specify the $\mathcal{Q}$-dynamics so that martingale pricing holds by construction. The other free parameters of the process are then chosen by a calibration algorithm.)



(a) NIG Process with CIR Clock

(b) Variance-Gamma Process with OU-Gamma Clock

Figure 3: Sample Paths of Some Time-Changed Exponential Levy Processes

Figures 3(a) and 3(b) display three samples paths each for two time-changed exponential Levy processes. Figure 3(a) uses a *normal-inverse Gaussian* Levy process with a *CIR* subordinator. Figure 3(b) uses a *variance-gamma* Levy process[18] with a *Gamma - OU* subordinator. Note how these paths differ from those generated by a geometric Brownian motion. The GBM paths of course cannot jump and do not display volatility clustering.

Tractability of Exponential Levy Processes with Stochastic Clock

Many Levy processes, including the time-changed normal-inverse-Gaussian and variance-gamma processes of Figures 3(a) and 3(b), respectively, are quite straightforward to simulate. However, the characteristic functions of the log-stock price are often available in closed form, even when the Levy process has a stochastic clock. In fact we have

$$\begin{aligned}
\phi(u,t) &:= \mathrm{E}_0^{\mathcal{Q}}\left[e^{iu\ \log(S_t)} \mid S_0, y_0\right] \\
&= \mathrm{E}_0^{\mathcal{Q}}\left[e^{iu\ \left[\log(S_0)+(r-q)t+X_{Y_t}-\log\left(\mathrm{E}_0^{\mathcal{Q}}\left[e^{X_{Y_t}}\right]\right)\right]} \mid S_0, y_0\right] \\
&= e^{iu\ [\log(S_0)+(r-q)t]}\ \frac{\mathrm{E}_0^{\mathcal{Q}}\left[e^{iuX_{Y_t}}\right]}{\mathrm{E}_0^{\mathcal{Q}}\left[e^{X_{Y_t}}\right]^{iu}}.
\end{aligned} \tag{29}$$

[18] See Cont and Tankov's book or the article "A Perfect Calibration! Now What?" by Schoutens, Simons and Tistaert (Wilmott, 2007) for further details on these processes.

Now define the *characteristic exponent* of the Levy process as

$$\psi_x(u) := \log \mathrm{E}_0^{\mathcal{Q}} \left[e^{iuX_1} \right]. \tag{30}$$

and note that

$$\begin{aligned} \mathrm{E}_0^{\mathcal{Q}} \left[e^{iuX_{Y_t}} \right] &= \mathrm{E}_0^{\mathcal{Q}} \left[\mathrm{E}_0^{\mathcal{Q}} \left[e^{iuX_{Y_t}} \mid Y_t \right] \right] \\ &= \mathrm{E}_0^{\mathcal{Q}} \left[\mathrm{E}_0^{\mathcal{Q}} \left[e^{iuX_1} \right]^{Y_t} \right] \\ &= \mathrm{E}_0^{\mathcal{Q}} \left[e^{\psi_x(u) Y_t} \right]. \end{aligned} \tag{31}$$

If we substitute (31) into both the numerator and denominator (with $u = -i$) of (29) we then obtain

$$\phi(u,t) = e^{iu((r-q)t + \log(S_0))} \, \frac{\varphi(-i\psi_x(u); t, y_0)}{\varphi(-i\psi_x(-i); t, y_0)^{iu}} \tag{32}$$

where $\varphi(u; t, y_0)$ is the characteristic function of Y_t given y_0. It is therefore the case that if we know the characteristic function of the integrated subordinator, Y_t, and the characteristic function of the Levy process, X_t, then we also know the characteristic function of the log-stock price. We can use this characteristic function to calculate option prices as discussed in the Section 5.

---

5 A Fourier Transform Method for Pricing European Options

Fourier and Laplace transform methods have proved particularly useful for pricing derivative securities when closed-form solutions are not available. Transform methods can be used to solve the pricing PDE or to compute the risk-neutral expected value directly. We briefly describe a technique[19] based on the Fourier transform that takes the latter approach. It requires the characteristic function of the log-stock price, which we calculated for time-changed exponential Levy processes in (32) and which we provided for the Heston model in (11).

Let α be a positive[20] constant. Then the price of a European call option, $C(K,T)$, with strike K and time-to-maturity T can be shown to satisfy

$$C(K,T) = \frac{e^{-\alpha \log(K)}}{\pi} \int_0^\infty e^{-iv \log(K)} \rho(v) \, dv \tag{33}$$

where

$$\begin{aligned} \rho(v) &:= \frac{e^{-rT} \, \mathrm{E}_0 \left[e^{i(v-(\alpha+1)i) \log(S_T)} \right]}{\alpha^2 + \alpha - v^2 + i(2\alpha+1)v} \\ &= \frac{e^{-rT} \, \phi(v - (\alpha+1)i, T)}{\alpha^2 + \alpha - v^2 + i(2\alpha+1)v} \end{aligned}$$

$i = 0 + i$ is imaginary, and $\phi(\cdot; T)$ is the characteristic function of the log-stock price, $\log(S_T)$. The option price in (33) can be found using the usual Fourier inversion techniques. If many options with different strikes but the same time-to-maturity need to be priced, then the Fast Fourier transform can be used. Alternatively[21], if just a single option price is required a standard numerical[22] integration of the right-hand-side of (33) can be performed. Indeed the implied volatility surface in Figure 1 was constructed by using (33) to compute call option prices in the Heston model.

[19] Developed in "Option Valuation Using the Fast Fourier Transform" by Carr and Madan in the *Journal of Computational Finance* (1998).

[20] Values of $\alpha = .75$ have been recommended in the literature but depending on the application at hand, a different value may be required. Note that (33) is valid for any positive α so any difficulties that might arise with α are due to the difficulties that arise with the numerical inversion of the right-hand-side of (33).

[21] There is still some ongoing debate as to what are the most efficient methods for evaluating (33).

[22] Note that some software packages / languages such as *Matlab* can handle complex calculations whereas others such as *VBA* require additional functions.

6 Barrier Options and Model Risk

Barrier options are options whose payoff depends in some way on whether or not a particular barrier has been crossed before the option expires. A barrier can be (i) a put or a call (ii) a knock-in or knock-out and (iii) digital[23] or vanilla resulting in eight different payoff combinations. (Of course this number increases if you also include early exercise possibilities and multiple barriers.) For example, a knockout put option with strike K, barrier B and maturity T has a payoff given by

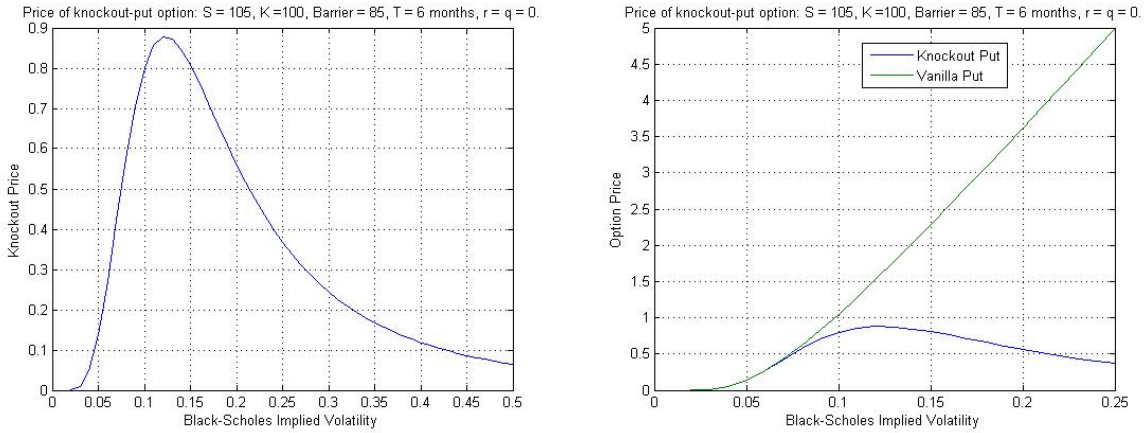
$$\text{Knock-Out Put Payoff } = \max\left(0,\ (K - S_T)\ 1_{\{S_t \geq B \text{ for all } t \in [0,T]\}}\right).$$

Similarly a digital down-and-in call option with strike K, barrier B and maturity T has a payoff given by

$$\text{Digital Down-and-In Call } = \max\left(0,\ 1_{\{\min_{t\in[0,T]} S_t \leq B\}} \times 1_{\{S_T \geq K\}}\right).$$

Knock-in options can be priced from knock-out options and vice versa since a knock-in plus a knock-out with the same strike is equal to the vanilla option or digital with the same strike. Analytic solutions can be computed for European barrier options in the Black-Scholes framework where the underlying security follows a geometric Brownian motion. We will not bother to derive or present these solutions here, however, since they are of little use in practice. As mentioned earlier, this is because the Black-Scholes model is not a good approximation to real markets and the model is no longer used in practice to calculate barrier prices.

We can still use it to develop some intuition, however, and for this we will concentrate on knock-out put options which are traded quite frequently in practice. Figure 4(a) shows the Black-Scholes price of a knock-out put option as a function of the implied volatility, σ. The barrier is $\$85$, the strike is $\$100$, current stock price is $\$105$, time-to-maturity is six months and $r = q = 0$. The price of the same knock-out option is plotted (on a different scale) alongside the price of a vanilla put option in Figure 4(b).



(a) Price of Knockout Put

(b) Knockout Put v. Vanilla Put

Figure 4: Knockout Put Options in Black-Scholes

It is clear that the knock-out put option is always cheaper than the corresponding vanilla put option. For low values of σ, however, the prices almost coincide. This is to be expected as decreasing σ decreases the chances of the knock-out option actually being knocked out. While the vanilla option is unambiguously increasing in σ the

[23] A digital call option pays \$1 if the underlying security expires above the strike and 0 otherwise. A digital put pays \$1 if the security expire below the strike. By "vanilla" option we mean the same payoff as a European call or put option.

same is not true for the knock-out option. Beyond a certain level, which depends on the strike and other parameters, the value of the knock-out put option decreases as σ increases.

Exercise 4 *What do you think would happen to the value of the knock-out put option as $\sigma \to \infty$?*

As mentioned earlier, the Black-Scholes model is not a good model for pricing barrier options. After all, what value of σ should be used: $\sigma(T, K)$, $\sigma(T, B)$, some function of the two or some other value entirely? Besides, we know the Black-Scholes model is wrong to begin with, so it makes no sense to use it for pricing exotic derivatives. Moreover, the Greeks provided by Black-Scholes would also be very problematic. Suppose, for example that barrier options were liquid and market prices were readily available. Then we might want to consider backing out an implied volatility, σ_b say, for a given barrier option and then using this σ_b to compute the Greeks for the barrier option. But already this leads to problems. It is clear from Figure 4(a) that there may not be a unique σ_b. There may in fact be no solution or there may be two possible values, $\sigma_{1,b} < \sigma_{2,b}$. Which value would we take? If we took $\sigma_{1,b}$ then we would see a positive vega but if we took $\sigma_{2,b}$ then our vega would be negative! Similarly, we might see a delta that has the opposite sign to the true delta. The Black-Scholes model then is particularly inappropriate for handling barrier options and derivative securities in general.

Exercise 5 *How do you think the price of the knock-out put option will behave as the skew increases? (This question is somewhat vague and the answer will depend on what the base level is from which the skew is increasing.)*

Barrier Price Depends on the Joint Distributions

We saw in an earlier lecture that the price of a barrier option is not solely determined by the marginal distributions of the underlying stock price. Figure 5 emphasizes this point. It displays the price of a down-and-out barrier call option as a function of the barrier level for several different[24] models. In each case the strike was set equal to the initial level of the underlying security. The parameter values for the various models were calibrated to the implied volatility surface of the Eurostoxx 50 Index on October 7^{th}, 2003. As a result, their marginal distributions coincide, at least to the extent that the calibration errors are small.

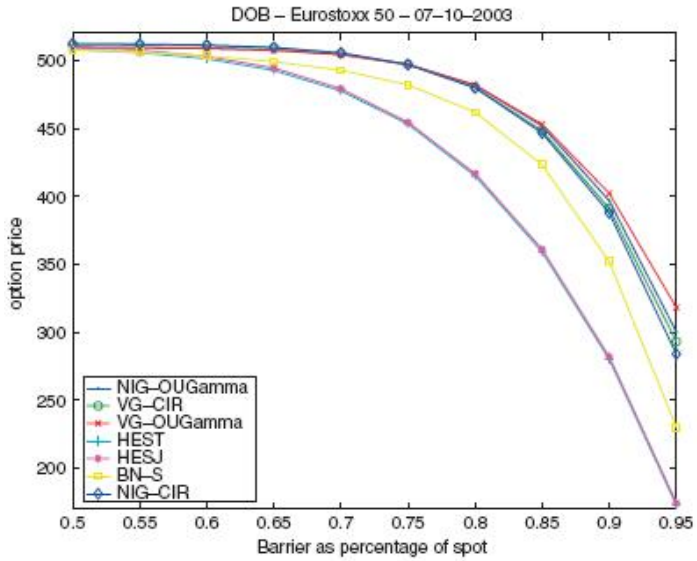


Figure 5: Down-and-out option prices for various pricing models.

[24]This example is taken taken from the article "A Perfect Calibration! Now What?" by Schoutens, Simons and Tistaert (Wilmott, 2007).

Yet it is clear that the different models result in very different barrier prices. The question therefore arises as to what model one should use in practice. This is a difficult question to answer. Perhaps the best[25] solution is to price the barrier using several different models that (a) have been calibrated to the market prices of liquid securities and (b) have reasonable dynamics that make sense from a modeling viewpoint. The minimum and maximum of these prices could be taken as the bid-offer prices if they are not too far apart. If they are far apart, then they simply provide guidance on where the fair price might be.

In the case of barrier options in the equity derivatives market, many participants appear to use local volatility models but with various ad-hoc adjustments included. These adjustments are intended to account for the possibility of jumps and other hedging difficulties that might arise. Barrier options are considerably more liquid in the foreign exchange markets, however, and so there has been a greater emphasis there on developing models that can accurately reproduce the market prices of vanilla and barrier options. This has led, for example, to the development of stochastic-local volatility models and it is still a topic of ongoing research.

---

7 Calibration

One has to be very careful when calibrating models to market prices. In general, and this is certainly the case with equity derivatives markets, the most liquid instruments are vanilla call and put options. Assuming that we have a volatility surface available[26] to us then, at the very least, we can certainly calibrate our model to this surface. An obvious but potentially hazardous solution would be to solve

$$\min_{\gamma} \sum_{i=1}^{N} \omega_i \ \left(\mathsf{ModelPrice}_i(\gamma) - \mathsf{MarketPrice}_i\right)^2 \tag{34}$$

where $\mathsf{ModelPrice}_i$ and $\mathsf{MarketPrice}_i$ are the model and market prices, respectively, of the i^{th} option used in the calibration. The ω_i's are fixed weights that we choose to reflect either the importance or accuracy of the i^{th} observation and γ is the vector of model parameters. There are many problems with performing a calibration in this manner:

1. In general (34) is a non-linear and non-convex optimization problem. As such it may be difficult to solve as there may be many local minima. It is possible that the optimization routine could get "stuck" in one of these local minima and not discover other local minima that could result in a much smaller objective function.

2. Even if there is only one local minimum, there may be "valleys" containing the local minimum in which the objective function is more or less flat. It is then possible that the optimization routine will terminate at different points in the valley even when given the same input, i.e. market option prices. This is clearly unsatisfactory but it becomes unacceptable, for example, in the following circumstances:

   Day 1: Model is calibrated to option prices and the resulting calibrated model is then used to price some path dependent security. Let P_1 be the price of this path dependent security.

   Day 2: Market environment is unchanged. (We are ignoring the fact that there is 1 day less to maturity which we assume has an insignificant impact on prices.) In particular, the volatility surface has not changed from Day 1. The calibration routine is rerun, the path dependent option is priced again and this time its price is P_2. But now P_2 is very different from P_1, despite the fact that the market hasn't changed. What has happened?

[25] These observations hold true for exotic securities in general.

[26] Since there will only be options with finitely many strikes and maturities, it will be necessary to use interpolation and extrapolation methods to construct the volatility surface from the implied volatilities of these options. In fact it is probably a good idea to also use variance-swaps to construct the volatility surface since the prices of variance swaps will depend in part on the prices of deep out-of-the money options. Market prices will generally not exist for these options and variance-swap prices can therefore help us to determine implied volatility levels in these (deep OTM) regions.

3. This is related to the previous point and explains why P_2 can be so different from P_1. Recall that the implied volatility surface only determines the marginal distributions of the stock prices at different times, T. It tells you nothing about the joint distributions of the stock price at different times. Therefore, when we are calibrating to the volatility surface, we are only determining the model dynamics up-to the marginal distributions. All of the parameter combinations in the "valley" might result in the very similar marginal distributions, but they will often result in very different joint distributions. This was demonstrated in Figure 5 when we discussed the pricing of barrier options. (In the latter case, however, different models rather than different parameter values within the same model were responsible for generating different joint distributions.)

There are several methods that can be used to resolve the problems listed above. If possible, it is a good idea to include in the calibration exercise some path-dependent securities. This will decrease the likelihood of the troublesome valleys existing. Unfortunately, liquid, path-dependent option prices may not be available in the market place and so this solution may not be feasible.

Regardless of whether or not we can include path-dependent options in the calibration, it is generally a good idea to try and "convexify" the optimization problem in (34). This can be done by adding an additional (convex) term to the objective function in (34). This term would penalize you the further the selected process is from some pre-specified "base" process. There are many ways in which this could be implemented.

Another possibility is to include a penalty term in the objective function that penalizes deviations in the parameter values from the calibrated parameter values of the previous day. By making this penalty term large, the calibration algorithm will only choose a new set of parameters when they perform significantly better than the previous day's parameters.