Abstract:

We show that it is possible to avoid the discrepancies of continuous path models for stock prices and still be able to hedge options if one uses birth and death models. One needs the stock and another market traded derivative to hedge an option in this setting. However, unlike in continuous models, the number of extra traded derivatives required for hedging does not go up with the introduction of stochastic intensity. We derive the results explicitly when the intensity rate follows a two state Markov process and outline ways of generalizing this to other models for the evolution of the intensity rate. We establish the risk-neutral measure and describe the algorithm for inverting option prices to get values of the unknown parameters in the model. We show that one needs to use filtering equations for updating parameter values and present those equations for the general case.

1. Introduction

A major difficulty on the interface between statistics and finance is that diffusion models are not really valid descriptions of data when it comes to microstructure. Security prices typically move in fixed units like 1/16 or 1/100 of a dollar. This does not present any particular problem when data are observed, say daily. Current technology however permits almost continuous observation, and estimation procedures based on discretely observed diffusions would then require throwing away data so as to fit the model. For some studies of the effect of discreteness in diffusion models, see Ball (1988), Gottlieb and Kalay (1985). The main feature of the difficulty is that the microstructure predicted by diffusion models includes observable quadratic variation (and hence volatilities), whereas this is nowhere nearly true in practice. According to continuous time models, the integrated volatility equals the quadratic variation. Hence, if data are observed continuously, the volatility should be observable. In practice we only observe a sample of the continuous time path. As shown in, for example, Jacod and Shiryaev (2003), the difference between the quadratic variation at discrete and continuous time scales converges to zero as the sampling interval goes to zero. Theorems 5.1 and 5.5 of Jacod and Protter (1998) give the size of the error. Hence, the best possible estimates of integrated volatility should be the observed quadratic variation computed from the highest frequency data obtainable. However, it has been found empirically that there is a bigger bias in the estimate when the sampling interval is quite small. Also, the estimate is not robust to changes in the sampling interval. For references in this area see Andersen et. al. (2001). Thus, although the volatility should be asymptotically observable, this is not true in practice if data are available at very high frequency. This problem is addressed in the context of a continuous model by Ait-Sahalia et. al. (2005). As noted in Dengler and Jarrow (1997), the Black-Scholes formula is based on a derivation where all risk is removable from an option position by a continuously rebalanced delta hedge. However, in practice, delta hedging needs to be augmented with gamma hedging and sometimes even vega hedging, which is ad hoc and inconsistent with the underlying theory.

On the other hand there is a reason why continuous models are used. These are more or less the only models where one can create self financing replicating portfolios for derivative securities. Hence, whereas from a data description point of view, it would make sense to use models with jumps, from a hedging standpoint, these models cannot be used. One of the consequences of this conflict is that statistical information is not used as much as it should be when it comes to valuing derivative securities. Instead there is a substantial reliance on implied quantities, as in Engle and Mustafa (1992). It is shown in Mykland (1996) that this disregard for historical data can lead to mispricing.

It is important to reconcile the needs of statisticians and the needs of option hedgers as it would be desirable to bring as much statistical information as possible to bear on financial modeling and at the same time be able to hedge. This is where birth and death processes become useful. Birth and death processes have the virtue that one can quite successfully set up derivative securities hedging in this model. It is not quite as straightforward as the continuous model. In the latter, in simple cases, options only need to be hedged in the underlying security; in birth and death process models one also needs one market traded derivative security to implement a self financing strategy. However, this is much nicer situation than what is the case in general models with jumps where it is impossible to hedge. In a sense birth and death processes are almost continuous, as one needs to traverse all intermediate states to go from one point to the other. On the other hand, these processes have a microstructure which conflicts much less with the data. In particular, the predictable quadratic variation is not observable.

The paper is organized as follows. Section 2, describes the basic model and presents an Edgeworth expansion for option prices. In Section 4, we introduce stochastic intensity rate and discuss the problems associated with that. As opposed to the continuous model, we do not require extra assets for hedging when we introduce stochastic volatility. This is because the volatility is unobserved. We present a Bayesian solution to the problem and possible generalizations. The prior on the intensity rate process that we study in detail is a two state Poisson jump process. We provide formulas for pricing and hedging options with any given prior on the intensity rate process. In Section 5, we study the performance of the model in some real data applications. We obtain the risk neutral measure by inverting option prices and then compute the posterior under the risk neutral measure. What we are doing is similar to an Empirical Bayes approach where the hyper-parameters are estimated from the data, in
2. The Model

2.1 The Linear Jump Intensity Model

The idea of modeling stock prices by a jump model in which they can go up, go down or stay the same was suggested in Perrakis (1988) to describe thinly traded stocks. The model of pricing and hedging options in birth and death models where the intensity of jumps is proportional to the present stock price (linear jump intensity) is solved in Korn et. al. (1998). We shall call the constant of proportionality rate. This is the discrete state-space version of the popular affine jump diffusion models, for example see Duffie et. al. (2000). It is shown in Sen (2005) that when the tick size $c \to 0$, $S_t \to \exp\left(\int_0^t \rho_s ds\right)S_0$ a.s. where $\rho_t$ is the risk free interest rate. Thus the simple model of jumps of size $\pm c$ with linear jump intensity converges to a deterministic process almost surely, as the jump size goes to zero and the intensity of jumps goes to infinity. Hence this is not a very interesting model. However if the intensity of jumps is allowed to be proportional to the square of the stock price, then such a process can be considered as discretized version of the geometric Brownian motion model for stock prices that corresponds to the popular Black-Scholes theory of asset pricing. This will be called the quadratic jump intensity model. In the following Section this model and the asymptotic result stated above will be formalized. Also, in Korn et. al. (1998) the market has been completed with a very special option, the LEPO-put. We show that one can use any general option to complete the market. In Dengler and Jarrow (1997), the authors consider a birth and death process on log of stock prices, which makes the problem much easier, but contradicts the fact that stock prices move in multiples of the tick size.

2.2 The Quadratic Jump Intensity Model

Suppose the stock price $S_t$ is a pure jump process with jumps of size $\pm c$. This implies that the process moves on a grid of resolution $c$ and $N_t = S_t/c$ is a birth and death process. The jumps of the $N_t$ process have random size $Y_t$ which is a binary variable taking values $\pm 1$ and the probability that $Y_t = 1$ is denoted by $p_{t,N}$. Suppose that there is a risk-free interest $\rho_t$ and the intensity of jumps is $N_t^2\lambda_t$ where the rate $\lambda_t$ is a non-negative stochastic process. In order to keep the process away from zero, introduce the condition:

$$\text{When } N_t = 1, Y_t \text{ takes values 0 and 1 with probabilities } p_{t,1} \text{ and } 1 - p_{t,1} \quad (1)$$

We now have the following result, the proof of which is given in Appendix A.

**Proposition 2.2.1.** Let $N_t^{(n)}$ be an integer valued jump process, the jump time $\xi_t^{(n)}$ following a counting process with rate $N_t^{(n)}\lambda_t$ and the random jump size $Y_t^{(n)}$ which is a binary variable taking values $\pm 1$ and probability that $Y_t = 1$ is $p_{t,N_t}$, and satisfying condition (1) and assumptions (2) and (3). Then $X_t^{(n)} = \ln(N_t^{(n)}/n)$ converges in distribution to $X_t$, a continuous Gaussian Martingale with characteristics

$$(f_0^t (\rho_u - \sigma_u^2/2)du, \int_0^t \sigma_u^2 du, 0) \text{ if } p_{t,N_t} = (\rho_t/N_t\sigma_t^2 + 1)/2 \text{ and } p_{t,1} = (\rho_t - \sigma_t^2/2)/(\sigma_t^2 \log(2))$$

The assumptions are:

$$E \sum_{\tau_i \leq t} \left(\frac{1}{N_{\tau_i}}\right)^k \to 0 \text{ as } n \to \infty \quad \text{for all } k \geq 4 \quad (2)$$

$$\int_0^t N_t^2 \sigma^2 du \text{ is finite a.s.} \quad (3)$$

Suppose for each $n$, the stock price $S_t^{(n)} = N_t^{(n)}/n$ where the process $N_t^{(n)}$ is described in Proposition 2.2.1. So the grid size $c$ is $1/n$. Assume initial stock price $S_0^{(n)}$ is the same for all $n$. As $n \to \infty$, by Proposition 2.2.1, the sequence of random processes $X_t^{(n)} = \ln(S_t^{(n)})$ converge in distribution to $X_t$, a continuous Gaussian Martingale with characteristics $(f_0^t (\rho_u - \sigma_u^2/2)du, \int_0^t \sigma_u^2 du, 0)$. Since exp is a continuous function, $S^{(n)} = \exp(X_t^{(n)})$ converge in law to exp($X_t$). The stochastic differential equation of $X$ is:

$$d(X_t) = (\rho_t - 1/2 \sigma_t^2)dt + \sqrt{\sigma_t^2}dW_t$$

where $W_t$ is standard Wiener process. By Ito’s formula,

$$d(S_t) = d(\exp(X_t)) = S_t[(\rho_t - 1/2 \sigma_t^2)dt + \sqrt{\sigma_t^2}dW_t] + \frac{1}{2} S_t \sigma_t^2 dt$$

Thus the limiting distribution is geometric Brownian motion. In Section 3, we consider processes with intensity $\lambda(N_t^2)$ where $\lambda_t$ is a constant. In Section 4, we consider the case of $\lambda_t$ being a stochastic process and $N_t$ is a birth and death process conditional on the $\lambda_t$ process. This can be formally carried out by letting $N_t$ be the integral of the $Y_t$ process with respect to the random measure that has intensity $\lambda_t N_t^2$(see, e.g. Ch. II.1.d(p71-74) of Jacod and Shiryaev (2003)).

2.3 Edgeworth Expansion for Option Prices

We have seen that the distribution of stock price process converges to geometric Brownian motion. Under the assumption of no arbitrage, option prices are expectations of discounted terminal payoffs. It is of interest to see how much these expectations differ from expectation under the Brownian motion model. Let us define

$$X_t^{(n)} = \ln \left( \frac{N_t^{(n)}}{n} \right)$$

$$X_t^{(n)} = X_t^{(n)} - X_t^{(0)} - \int_0^t p_{u,N_u} \ln \left( \frac{1 + \frac{x}{N_u}}{1 + \frac{x}{N_0}} \right) + (1 - p_{u,N_u}) \ln \left( \frac{1 - x}{N_u} \right) N_u^2 \sigma_u^2 du$$

where $p_{t,N_t} = \frac{1}{2} \left[ 1 \left( 1 + \frac{n}{N_t \sigma_t^2} \right) \right]$.

Let $\mathcal{C}$ be the class of functions $g$ that satisfy the following: (i) $f | g(x) | dx < \infty$, uniformly in $C$, and $\{\sum x^2 E g(x), g \in C\}$ is uniformly integrable (here, $\hat{g}$ is the Fourier transform of $g$, which must exist for each $g \in C$); or (ii) $g(x) = f(x \cdot x_1)$, with $\sum z^2 \hat{f}, \hat{f}$ and $f''$ bounded, uniformly in $\mathcal{C}$, and with
\{f'' : g \in C\} equicontinuous almost everywhere (under Lebesgue measure). Under assumptions (I1) and (I2) stated in Appendix B, for any \( g \in C \),

\[
E_g(X_T^{(n)}) = E_g(N(0, \lambda T)) + o(1/n)
\]

The details are given in Appendix B. Thus, if the payoff function has some smoothness properties, then the expectation under the birth and death model is very close to that under the geometric Brownian motion model. However, most traded options do not have these smoothness properties. That is why we study these deviations computationally in section 5.

3. Pricing and Hedging of options when intensity rate is constant

For any \( \lambda \), let \( P_\lambda \) be the measure associated with a birth-death process with event rate \( \lambda N_i^2 \) and probability of birth \( p_i N_i = (1 + p_i/\lambda N_i)/2 \). We have seen that as the tick size goes to zero, this process converges to geometric Brownian motion and if the payoff function has some smoothness properties, then option prices converge to those given by the Black-Scholes model. We now restrict our attention to the case of fixed tick-size. Let us first consider the case when \( \lambda \) is constant. In section ?? we explore conditions under which the risk-neutral distribution is a birth and death process and conditions for the market to be complete. In section 3.1 we derive the hedging strategy in this market.

3.1 Hedging

We have shown that the market is complete when we add a market traded derivative security. So we can hedge an option by trading the stock, the bond and another option. In this Section we obtain the hedge ratios by forming a self-financing risk-less portfolio with the stock, the bond and two options.

Let \( F_2(x, t), F_3(x, t) \) be the prices of two options \( O_2 \) and \( O_3 \) at time \( t \) when the price of the stock is \( c_x \). Let \( F_1(x, t) = c_x \) be the price of the stock and \( F_0(x, t) = B_0 \exp \left( -\int_0^t \rho_s ds \right) \) be the price of the bond. Note that \( N_i \) is the difference of two counting processes: \( N_{1t}, \) the number of births and \( N_{2t}, \) the number of deaths. Assume \( F_i \) are continuous in both arguments and the following partial derivatives exist.

\[
\frac{\partial F_i}{\partial t} = \alpha_F, \quad \frac{\partial F_i}{\partial N_1} = \beta_F, \quad \frac{\partial F_i}{\partial N_2} = \gamma_F.
\]

We shall construct a self financing risk-less portfolio

\[
V(t) = \sum_{i=0}^3 \phi^{(i)}(t) F_i(x, t)
\]

Let \( u^{(i)}(t) = \frac{\phi^{(i)}(t) F_i(x, t)}{V(t)} \) be the proportion of wealth invested in asset \( i \).

\[
\sum_{i=0}^3 u^{(i)} = 1
\]

Since \( V_t \) is self financing,

\[
dV(t) = \sum_{i=0}^3 u^{(i)}(t) dF_i(x, t)
\]

\[
= \sum_{i=0}^3 u^{(i)}(t) \frac{dF_i(x, t)}{F_i(x, t)}
\]

\[
= u^{(0)}(t) \rho dt + u^{(1)}(t) \frac{dN_{1t}}{\lambda} - dN_{2t}
\]

\[
+ \sum_{i=2}^3 u^{(i)}(t) (\alpha_F(x, t) dt + \beta_F(x, t) dN_{1t} + \gamma_F(x, t) dN_{2t})
\]

\( V_t \) is risk-less implies

\[
u^{(1)}(t) \frac{1}{\gamma} + \sum_{i=2}^3 u^{(i)}(t) \beta_F(x, t) = 0
\]

\[
u^{(1)}(t) + \sum_{i=2}^3 u^{(i)}(t) \gamma_F(x, t) = 0
\]

The no arbitrage assumption implies

\[
u^{(0)}(t) \rho_t + \sum_{i=2}^3 u^{(i)}(t) \alpha_F(x, t) = \rho_t
\]

Solving equations (4)-(7), we get the hedge ratios as:

\[
u^{(2)} = \left[ (1 - \frac{\alpha_F^3}{\rho} - x \beta_F) - \frac{\gamma_F + \beta_F}{\gamma_F + \beta_F} \right]^{-1}
\]

\[
u^{(3)} = \left[ (1 - \frac{\alpha_F^3}{\rho} - x \beta_F) - \frac{\gamma_F + \beta_F}{\gamma_F + \beta_F} \right]^{-1}
\]

\[
u^{(0)} = -\frac{1}{\rho_t} \left[ u^{(2)} \alpha_F + u^{(3)} \alpha_F \right]
\]

\[
u^{(1)} = -x(u^{(2)} \beta_F + u^{(3)} \beta_F)
\]

A replicating portfolio for option \( O_2 \) can be formed by investing \( \phi^{(2)} \phi^{(2)} \) units of asset \( i \), that is, \( u^{(0)} F^{(3)}/u^{(3)} F^{(0)} \) units of Bond, \( u^{(1)} F^{(3)}/u^{(3)} F^{(1)} \) units of Stock and \( u^{(2)} F^{(3)}/u^{(3)} F^{(2)} \) of option \( O_2 \).

4. Stochastic Intensity Rate

We now introduce stochastic intensity rate. The intensity rate of jump processes is analogous to volatility in continuous models. Both theoretical and empirical considerations support the need for stochastic volatility. Asset returns have been modeled as continuous processes with stochastic volatility as in Hull and White (1987), Naik (1993) or as jump processes with stochastic volatility as in Bates (1996), Duffie et al. (2000).

Consider the case where the unobserved intensity rate \( \lambda_t \) is a stochastic process. We first put a two state Markov model prior on \( \lambda_t \). This assumes that stock price movements fluctuate between low and high intensity rate regimes. This is the approach in Naik (1993). In Section ?? we provide formulas for pricing and hedging options with any given prior on the intensity rate process. For example, alternatively we can consider cases where the intensity rate follows a diffusion as in Hull and White (1987).

Suppose there is an unobserved state process \( \theta_t \) which takes 2 values, say 0 and 1. The transition matrix is \( Q \). When \( \theta_t = i, \lambda_t = \lambda_i \). Jump process associated with \( \theta_t \) is \( \zeta_t \).

Let us denote by \( \{ \tilde{G}_t \} \) the complete filtration \( \sigma(S_u, \lambda_u, 0 \leq u \leq t) \) and by \( P \) the probability measure on \( \{ \tilde{G}_t \} \) associated with the process \( \{ S_t, \lambda_t \} \).
4.1 Hedging

Let us assume that the risk-neutral measure is the measure associated with a birth-death process with jump intensity $\lambda_i N_i^2$ and probability of birth $p_i = (1 + \lambda_i N_i/p_1)$ where $\lambda_i$ is a Markov process with state space $\{\lambda_0, \lambda_1\}$. The model is incomplete and we need to introduce a market traded option to complete the model. In order to determine the risk-neutral parameters, we have to equate the observed price of a market traded option to its expected price under the model. We get two different values of the expected price under the two values of $\theta(0)$. The $\theta$ process is unobserved. So there is no way of determining which value to use. We cannot invent an option to get $\theta(0)$ either, because it takes two discrete values and does not vary over a continuum.

We need to introduce $\pi_i(t) = P(\theta_i = i | F_t)$ where $F_t = \sigma(S_u, 0 \leq u \leq t)$

Let $F_i(x, t) := E(X|G_t)$ when the stock price is $cx$ and $\lambda_i = \lambda_0$.

Let $G(x, t) := E(X|F_t) = \pi_0(t)F_0(x, t) + \pi_1(t)F_1(x, t)$

As shown in Snyder (1973), under any $P \in \mathcal{P}$ the $\pi_{1t}$ process evolves as:

$$d\pi_{1t} = a(t)dt + b(t, 1)dN_{1t} + b(t, 2)dN_{2t}$$

(8)

where $a(t)$ and $b(t, i)$ are $F_t$ adapted processes.

**PROPOSITION 4.1.1.**

$$dG/G = \alpha_F dt + \beta_F dN_{1t} + \gamma_F dN_{2t}$$

where

$$\alpha_F(x, t) = \pi_0 \frac{\partial \alpha_0}{\partial x} + \pi_1 \frac{\partial \alpha_1}{\partial x} + a(t)(F_0 - F_1)$$

$$\beta_F = \pi_0 \alpha_F F_0 + \pi_1 \alpha_F F_1 + b(t, 1)(F_0 + b(t, 1) F_0(1 + \beta_F))$$

$$\gamma_F = \pi_0 \gamma_F F_0 + \pi_1 \gamma_F F_1 + b(t, 1)(F_0 + b(t, 1) F_0 - F_1(1 + \gamma_F))$$

Proof.

$$dG = \pi_0 dF_0 + \pi_1 dF_1 + d\pi_0 F_0 + d\pi_1 F_1 + \text{covariance term}$$

$$= \pi_0 dF_0 + \pi_1 dF_1 + [(F_0 - F_1)d\pi_0 + d[F_0 - F_1], \pi_0]$$

$$= -(\frac{\partial \alpha_0}{\partial t} + \frac{\partial \alpha_1}{\partial t})dt$$

$$+ (\frac{\partial \alpha_F}{\partial x} F_0 + \frac{\partial \alpha_F}{\partial x} F_1)dN_{1t}$$

$$+ (\pi_0 \gamma_F F_0 + \pi_1 \gamma_F F_1)dN_{2t}$$

$$+ (F_0 - F_1)(a(t)dt + b(t, 1)dN_{1t} + b(t, 1)dN_{2t})$$

$$+ (\beta_F F_0 - \beta_F F_1)b(t, 1)dN_{1t}$$

$$+ (\gamma_F F_0 - \gamma_F F_1)b(t, 1)dN_{2t}$$

$$= -\alpha dt + \beta dN_{1t} + \gamma dN_{2t}$$

We can hedge an option in the same way as in the fixed $\lambda$ case with $\alpha, \beta, \gamma$ replaced by $\alpha, \beta, \gamma$. We still need one market traded option and the stock to hedge an option. But to get the hedge ratios, we need $\pi(t), a(t), b(t)$.

### 4.2 Inference on the $\pi_t$ process

Inference on the parameter $\pi(t)$ can be done in the risk-neutral setting by inverting options at each point of time. Besides involving huge amount of computations this also has the following theoretical drawbacks:

- The hedge ratios involve $a(t), b(t)$. This implies that we need them to be predictable. But if we have to invert an option to get them, then we need to observe the price at time $t$ to infer $\pi_t$ and from there to get $a_t$ and $b_t$. So they are no more predictable.

- Options are not as frequently traded as stocks and hence option prices are not as reliable as stock prices. Thus, inferring $\pi(t)$ at each time point $t$ by inverting options will give incorrect prices and lead to arbitrage.

So we base our updates only on the stock prices. Inference on the initial parameter $\pi(0)$, the transition rates $q_{01}, q_{10}$ and the jump rates $\lambda_0, \lambda_1$ is done under the risk-neutral measure based on stock and option prices. However subsequent inference on the process $\pi_t$ is done based only on the stock price process using Bayesian filtering equations.

As shown in Elliott et. al. (1995), the posterior of $\pi_j(t)$ is given by:

$$\pi_j(t) = \pi_j(0) + \int_0^t \sum_i q_{ij} \pi_i(u)du + \int_0^t \pi_j(u)(\lambda(u) - \lambda_j)N_u^2du$$

$$+ \sum_{0 < u < t} \pi_j(u-) \left( \sum_{i \neq j} \pi_i(u)\lambda_i p_\lambda(S_u - \rightarrow S_u) - 1 \right)$$

where $\lambda(t) = \sum_i \pi_i(t)\lambda_i$.

Thus, $a_j(u) = \sum_i q_{ij} \pi_i(u) + \int_0^u \pi_j(u)(\lambda(u) - \lambda_j)N_u^2du$

and $b_j(u) = \pi_j(u-) \left( \sum_{i \neq j} \pi_i(u)\lambda_i p_\lambda(S_u - \rightarrow S_u) - 1 \right)$

This is the conditional law of the unobserved rate process given the path up to time $t$ of the observed stock price process as in a hidden Markov model.

5. Real Data Applications

5.1 Description

Data on stock price and option price was obtained from the option-metrics database for three stocks Ford, ABMD and IBM. The stock data is transaction by transaction. The format of the raw stock data is: SYMBOL, DATE, TIME, PRICE, SIZE, G127, CORR, COND, EX. After filtering for after hour and international market trading, the data is on tradings in NASDAQ regular hours. The prices are divided by the tick size to obtain integers. We use part of the data as training sample and the rest as test sample. The option data is daily best bid and ask prices. We preprocess the data to remove volume zero and symbols not equal to F, IBM or ABMD. The processed data has DATE, CALL/PUT, EXPIRATION, BEST-BID, BEST-ASK, STRIKE.

We estimate risk-neutral parameters by inverting option prices in training sample and use these estimates to predict prices of options in the test sample. The 3 stocks provide some variety. The Ford stock is a little old when the tick size used to be $1/16$ while the others have tick size $1/100$. The ABMD data is more thinly traded than the other two. Table 1 presents the dates, sample size, observed range of option prices and the average bid-ask spread. The figures quoted are in multiples of the tick size.
5.2 Constant Intensity Rate

The predicted option price from the model is computed at seven points during the day at the first times that a trade takes place on the stock at every hour. The minimum, maximum and average of these 7 predicted values will be denoted by min, max and avg respectively. The observed interval is (bid, ask) and mid:==(ask+bid)/2. For discussion on the issue of using the midpoint when data is recorded as discrete bid-ask quotes, refer to Hasbrouk (1999).

The performance of a method is good if the predicted interval (min, max) is close to the observed interval (bid, ask). We use the following measure of distance between the 2 intervals to summarize the results:

\[
a = \text{Average of } [(\text{mid-max})_+ + (\text{min-mid})_+]\]

This measures how far the bid-ask midpoint is from the predicted interval. Table II presents \(\lambda\), average length of predicted interval and the distance measure \(a\) for the 2 methods and various datasets. Here \(\lambda\) is the volatility for Black-Scholes model and the intensity for the birth-death model. It is observed that the 2 models provide very similar results. The Black-Scholes model gives better prediction for the IBM data. The gain in using the birth-death model is most in the AMBD data possibly because it is thinly traded. There is some gain in the Ford data possibly because of the large tick-size. In both of these cases once we fix the parameter we are considering one fixed model. We get the intervals as the maximum and minimum of the option prices as the stock price varies over the day. So these are not any prediction intervals, but just arise because of inter-day variation in the predicted option price.

5.3 Stochastic Intensity Rate

For stochastic intensity rate, we have to find the price of options given 5 parameters: the two intensity parameters \(\lambda_0, \lambda_1\), the transition matrix of the intensity process that is determined by \(q_{01}, q_{10}\), and \(\pi\) the probability that the intensity process is in state \(\lambda_0\) at time 0. The objective is to find the parameter set that minimizes the root mean square error between the bid-ask-midpoint and the daily average of the predicted option price, for all options in the training sample. We followed a diagonally scaled steepest descent algorithm with central difference approximation to the differential. The starting values of \(\lambda_0, \lambda_1\) are taken to be equal to the value of the estimator \(\hat{\lambda}\) obtained in the constant intensity model. The starting values of \(q_{01}, q_{10}\) are obtained by a hidden Markov model approach using an iterative method that has 2 steps. Let the underlying state process be \(\eta_t\), that is the intensity is \(\lambda_t\) when \(\eta_t\) is \(i\). In one step, the MLE of the parameters is obtained given the \(\eta_t\) process. For details on the method for obtaining the MLE in this setting, refer to Elliott et. al. (1995). Given the latent process the MLE’s are obtained from the following equations:

\[
\hat{q}_{ij} = \frac{\text{Number of times latent process jumps from state } i \text{ to state } j \text{ by latent process in state } j}{\text{Time spent by latent process in state } j}
\]

\[
\sum_{j \in \text{State space}} A^{m}_{j,j+1} / \lambda_m + A^{m}_{j,j-1} / \lambda_m - j^2 T^{m}_{j} = 0
\]

where \(A^{m}_{ij} = \text{Number of times observed process jumps from state } i \text{ to state } j \text{ when latent process is in state } m \) and \(T^{m}_{j} = \text{Time during which latent process is in state } m \) and observed process is in state \(j\).

In the next step, for each \(t\) when there is a jump in the stock price, we assign \(\eta_t\) to that \(i\) which maximizes the probability of an event. When this method converges, we do a finite search on the parameter \(\pi\). Then we perform the minimization algorithm to get the risk neutral parameters.

For the ABMD and Ford datasets, the RMSE of prediction obtained from the constant intensity method is less than the bid-ask spread. This means that the constant intensity model attains the lower bound on the possible quadratic calibration error as referred to the intrinsic error in Cont (2005). Since the data is observed with an error, that is only as bid-ask quotes, we cannot hope to achieve a level of error less than the precision of the observed data. However for the IBM data there is scope for improvement.

Table III shows the values of \(\lambda_0, \lambda_1\) and the RMS error for the evolution of the algorithm. The initial values of \(q_{01}, q_{10}\) and \(\pi\) are 8.64e-02,1.2126 and 0.5 and they do not change in the significant digits over the evolution of the algorithm.

The algorithm converges with the value of RMSE equal to 29.6799. It is seen that we do achieve substantial improvement in the RMSE by using the stochastic intensity rate model over the birth death model with constant intensity (RMSE=52.7729). However, the bias and smile that were observed in the constant volatility model still remain for the stochastic volatility model. Also, this is an ill-posed problem in the sense that the solutions are not unique and depend on the starting values of the parameters. There is still scope for improvement since the best RMSE we can hope to attain is the daily bid-ask spread which is 5 in this case.

6. Conclusion

The birth and death model is different from general models with jumps in the capacity that one can set up derivative security hedging in this model. It also takes into account the fact that stock prices move in multiples of the tick. Both Edgeworth expansion results and study of real data show that the birth and death model with quadratic intensity rate produces option prices that are very similar to the Black-Scholes model. However, this model removes the inconsistency between theory and practice that the continuous path models run into. In a way, this study explains why the industry practice of using the Black-Scholes model works in practice inspite of being adhoc and contradicting the underlying theory. Though the pricing from this model is very similar to that from Black-Scholes model, the hedging is significantly different. One needs not only the stock and bond, but another market traded option to hedge. Also, the introduction of stochastic volatility does not necessitate the introduction of another derivative security for hedging purposes.
APPENDIX

A Proof of Proposition 2.2.1

Define $X^n(\tau) = \ln(N^n(\tau))$ and

$X^n_{\text{sd}} = X^n - \int_0^\tau [\rho_t \log(1 + \frac{1}{N^n_t}) + (1 - \rho_t) \log(1 - \frac{1}{N^n_t})]N^n_t^2 \sigma^2_t du - X^n_0$

where $p_t,N_t = \frac{\rho_t}{1 + \frac{1}{N^n_t}}$

LEMMA 1. For $t > 0$,

$\left[ p_t,N_t \log(1 + \frac{1}{N^n_t}) + (1 - p_t,N_t) \log(1 - \frac{1}{N^n_t}) \right] N^n_t^2 \sigma^2_t \to \rho_t - \frac{\sigma_t^2}{2}$

Proof. $\left[ p_t,N_t \log(1 + \frac{1}{N^n_t}) + (1 - p_t,N_t) \log(1 - \frac{1}{N^n_t}) \right] N^n_t^2 \sigma^2_t$

$= (p_t,N_t - 1) N^n_t^2 \sigma^2_t \sum_{k=1}^{\infty} \frac{1}{(2k-1)N^n_{k+1}^2} - N^n_t^2 \sigma^2_t \sum_{k=1}^{\infty} \frac{1}{k N^n_k^2}$

$= \rho_t - \frac{\sigma_t^2}{2} + p_t \sum_{k=1}^{\infty} \frac{1}{(2k+1)N^n_{k+1}^2} - \frac{\sigma_t^2}{2} \sum_{k=1}^{\infty} \frac{1}{(2k+1)N^n_k^2}$

$\to \rho_t - \frac{\sigma_t^2}{2}$

LEMMA 2. $X^n_{\text{sd}}$ is a local martingale.

Proof. Let $dN_{1t} = -\alpha_t dt + dN_t$ and $dN_{2t} = -I_{\{t = 1\}} dN_t$ with $N_0 = 0$ and $\alpha_t$ is a 2-dimensional process with $dN_t = dN_{1t} + dN_{2t}$ and intensity $(\lambda_1 := \sigma^2_t p_t,N_t^2, \lambda_2 := \sigma^2_t (1 - p_t,N_t))$.

$Q_t = N_t - \int_0^t \alpha_t N^2_s ds$ is the compensated process associated with $N_t$ and hence by Thm 8 of Bremaud(1981) $X^n_{\text{sd}} = \int_0^t \left[ \rho_t \log(1 + \frac{1}{N^n_s}) + (1 - \rho_t) \log(1 - \frac{1}{N^n_s}) \right] N^n_s^2 \sigma^2_s dQ_s$ is a local martingale.

$\int_0^t | \log(1 + \frac{1}{N^n_s}) | N^n_s^2 \sigma^2_s dt$ and $\int_0^t | \log(1 - \frac{1}{N^n_s}) | N^n_s^2 \sigma^2_s dt$ are finite a.s.

LEMMA 3. $\frac{d}{dt} < X^n_{\text{sd}} X^n_{\text{sd}} > t = [p_t,N_t \log(1 + \frac{1}{N^n_t})^2$ + $(1 - p_t,N_t) \log(1 - \frac{1}{N^n_t})^2] N^n_t^2 \sigma^2_t$

An expansion similar to the one in Lemma 1 shows that this quantity converges in probability to $\sigma_t^2$ for all $t > 0$.

Let $M^n_{\text{sd}} = \sum_{\tau_i < t} | X^n_{\text{sd}} X^n_{\text{sd}} > \tau_i$ if Assumption 2 holds, $[M^n_{\text{sd}}, M^n_{\text{sd}}]_t = \sum_{\tau_i \leq t} | \Delta X^n_{\text{sd}} |^4 \to 0$

and the result follows

B Edgeworth Expansion

We shall use the following results from Mykland (1995):

PROPOSITION 2.0.1. Let $(\epsilon_t^n)_{0 \leq t \leq T}$ be a triangular array of zero mean cadlag martingales. Assume conditions (I1)-

(13) and that $g^{ij}$ is well defined. Then

$P(\epsilon_t^n T_t^n \leq x) = \Phi(x, \kappa) + r_n(I(\kappa) + o_2(\kappa))$

This is Theorem 1 of Mykland (1995). The conditions are:

(11) For each $i$, there are $\kappa_i, \kappa_i < \kappa_i < \kappa_i$ so that

$r_n^{-1}(\epsilon_t^n, \epsilon_t^n) \leq 1 - o(\kappa_i)$

is uniformly integrable

(12) For the same $\kappa_i, \kappa_i$,

$P_b(\kappa_i \leq \epsilon_t^n, \epsilon_t^n \leq \kappa_i) = 1 - o(\kappa_i)$

for each $i$

(13) For each $i$,

$E(\epsilon_t^n, \epsilon_t^n, \epsilon_t^n, \epsilon_t^n | T_t^n) = O(c^n_\kappa R^n)$

holds uniformly in the class $C$ described in Section 2.3 of functions $g$. One can take

$g^{ij} = E_{\text{sd}} \left( \epsilon_t^n, \epsilon_t^n, \epsilon_t^n, \epsilon_t^n, T_t^n \right) = \Phi(\kappa) + r_n(I(\kappa) + o_2(\kappa))$

provided the right hand side is well defined. The subscript "sd" means asymptotically. $J(g)$ is then given by

$J(g) = \frac{1}{2} E_{\text{sd}} g^{ij}(Z) \frac{\partial}{\partial \alpha_i} \frac{\partial}{\partial \alpha_j} g(Z)$

We shall also need the following part of Proposition 3 of Mykland (1995):

PROPOSITION 2.0.2. Suppose that condition (13) of Theorem 2.0.1 holds. Then conditions (H1) and (H2) on $(\epsilon_t^n, \epsilon_t^n)_{0 \leq t \leq T}$ are equivalent to the same conditions imposed on either $(\epsilon_t^n, \epsilon_t^n)_{0 \leq t \leq T}$ or $< \epsilon_t^n, \epsilon_t^n > T_t^n$. Suppose furthermore that $< \epsilon_t^n, \epsilon_t^n, \epsilon_t^n, \epsilon_t^n > u_{T_t^n} / c^u T_t^n$ converges in probability to a constant for each $u \in [0,1]$, with $< \epsilon_t^n, \epsilon_t^n, \epsilon_t^n, \epsilon_t^n > u_{T_t^n} / c^u T_t^n \rightarrow \eta_t^{u}$. Then, if $\kappa_j$ is non-singular, the following is valid:

$\frac{\epsilon_t^n, \epsilon_t^n, \epsilon_t^n, \epsilon_t^n}{c_n T_t^n} \rightarrow \eta_t^{u}$

where $U_n = O_p(1)$ and asymptotically has zero expectation given $c_n T_t^n$.

Let us consider $\xi_t^n = \sqrt{n} X^n_{\text{sd}}(n)$ where $X^n_{\text{sd}}$ is defined in Appendix A. Let $\kappa = X_T, T_T^n = T, c_n = n, r_n = 1/n$. We shall suppress the $i$ in the notation, because we are in the univariate case. Then

$[\epsilon^n, \epsilon^n, \epsilon^n, \epsilon^n] = \int_0^T \left[ \left( 1 + \frac{\lambda_t}{N_{t-1}} \right)^4 d\xi_t^n$
and (I2), then for any \( H \) hence, if there exist constants \( \ell_1 \), \( \ell_2 \), then

\[
E[\ell_1, \ell_2, \ell_3]_{T_n} = n^2 \int_0^T E \left[ \left( \frac{1}{N_u} \right)^4 N_u^2 \lambda \right] du + o(1)
\]

\[
= \int_0^T E \left( \frac{1}{s_u^2} \right) \lambda du + o(1)
\]

\[
= O(c_n^2 n^2)
\]

So condition (I3) is satisfied.

\[
\frac{\ell_1, \ell_2, \ell_3 > \sigma T_n}{c_n^2 r_n} = \frac{n^{3/2}}{\sqrt{n}} \int_0^T E \left[ \left( \ln \left( 1 + \frac{Y_u}{N_u} \right) \right)^3 \right] H_u \bigg| N_u^2 \lambda du
\]

\[
= \frac{1}{n} (\rho - \frac{3 \lambda}{\pi}) \int_0^T \frac{1}{s_u^2} du + o(1/n)
\]

This implies \( \eta \) of Proposition 2.0.2 is 0. Hence we can replace \( (\ell_1, \ell_2, \ell_3) \), by \( < \ell_1, \ell_2 > T_n \) in the definition of \( a \).

\[
\frac{\ell_1, \ell_2, \ell_3 > T_n}{n} = \frac{\ell_1}{T_n} E \left[ \left( \ln \left( 1 + \frac{Y_u}{N_u} \right) \right)^2 \right] H_u \bigg| N_u^2 \lambda du
\]

\[
= \frac{\ell_1}{T_n} \left( p_{u,N_u} \left( \ln \left( 1 + \frac{Y_u}{N_u} \right) \right)^2 \lambda N_u \right)
\]

\[
+ \frac{\ell_2}{T_n} \left( 1 - p_{u,N_u} \right) \left( \ln \left( 1 + \frac{Y_u}{N_u} \right) \right)^2 \lambda N_u \lambda du
\]

\[+ \text{smaller order terms}
\]

\[
= \lambda T + \left( \frac{1}{2} + \frac{1}{3} - \frac{4}{3} \right) \lambda \int_0^T \frac{1}{s_u^2} du + \text{smaller order terms}
\]

\[
= \frac{11}{12} \lambda + \rho \int_0^T \frac{1}{s_u^2} du + o(1/n)
\]

\[
\eta = E_{aw} \left( \frac{< \ell_1, \ell_2 > T_n}{n} - \lambda T \right) + o(1/n)
\]

Hence, if there exist constants \( \ell_1, \ell_2 \) satisfying assumptions (I1) and (I2), then for any \( g \in C \),

\[
Eg(X_T^{(n)}) = Eg(N(0, \lambda T)) + o(1/n)
\]

Table 1: Results for Constant Intensity Rate

<table>
<thead>
<tr>
<th>Model</th>
<th>Sample</th>
<th>Data</th>
<th>( \lambda )</th>
<th>Length</th>
<th>Distance</th>
</tr>
</thead>
<tbody>
<tr>
<td>BS</td>
<td>Training</td>
<td>IBM</td>
<td>8.0e-07</td>
<td>99.180</td>
<td>10.603</td>
</tr>
<tr>
<td>BS</td>
<td>Test</td>
<td>IBM</td>
<td>8.0e-07</td>
<td>52.199</td>
<td>19.481</td>
</tr>
<tr>
<td>BD</td>
<td>Training</td>
<td>IBM</td>
<td>1.1e-06</td>
<td>124.402</td>
<td>10.111</td>
</tr>
<tr>
<td>BD</td>
<td>Test</td>
<td>IBM</td>
<td>1.1e-06</td>
<td>81.048</td>
<td>20.502</td>
</tr>
<tr>
<td>BS</td>
<td>Training</td>
<td>ABMD</td>
<td>6.5e-06</td>
<td>9.637</td>
<td>1.031</td>
</tr>
<tr>
<td>BS</td>
<td>Test</td>
<td>ABMD</td>
<td>6.5e-06</td>
<td>12.768</td>
<td>3.873</td>
</tr>
<tr>
<td>BD</td>
<td>Training</td>
<td>ABMD</td>
<td>7.5e-06</td>
<td>17.540</td>
<td>0.973</td>
</tr>
<tr>
<td>BD</td>
<td>Test</td>
<td>ABMD</td>
<td>7.5e-06</td>
<td>19.466</td>
<td>2.431</td>
</tr>
<tr>
<td>BS</td>
<td>Training</td>
<td>F</td>
<td>9.0e-07</td>
<td>6.096</td>
<td>0.297</td>
</tr>
<tr>
<td>BS</td>
<td>Test</td>
<td>F</td>
<td>9.0e-07</td>
<td>5.919</td>
<td>0.609</td>
</tr>
<tr>
<td>BD</td>
<td>Training</td>
<td>F</td>
<td>1.3e-06</td>
<td>6.791</td>
<td>0.225</td>
</tr>
<tr>
<td>BD</td>
<td>Test</td>
<td>F</td>
<td>1.3e-06</td>
<td>6.578</td>
<td>0.586</td>
</tr>
</tbody>
</table>

Table 2: Evolution of algorithm

<table>
<thead>
<tr>
<th>Data</th>
<th>( \lambda_0 )</th>
<th>( \lambda_1 )</th>
<th>RMSE</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.100000e-06</td>
<td>1.100000e-06</td>
<td>52.7742</td>
<td></td>
</tr>
<tr>
<td>1.065162e-06</td>
<td>1.594270e-07</td>
<td>39.5022</td>
<td></td>
</tr>
<tr>
<td>1.043488e-06</td>
<td>8.320846e-08</td>
<td>38.2857</td>
<td></td>
</tr>
<tr>
<td>1.042039e-06</td>
<td>8.326884e-08</td>
<td>29.6799</td>
<td></td>
</tr>
</tbody>
</table>

Table 3: Description of Data

<table>
<thead>
<tr>
<th>Data</th>
<th>Date</th>
<th>Sample Size</th>
<th>Option Price</th>
<th>Bid-ask spread</th>
<th>Trades per day</th>
</tr>
</thead>
<tbody>
<tr>
<td>Field Test</td>
<td>Dec 5-31, 00</td>
<td>578</td>
<td>0-256</td>
<td>2</td>
<td>1675</td>
</tr>
<tr>
<td>ABMD Training</td>
<td>Feb 3, 03</td>
<td>12</td>
<td>0-160</td>
<td>25</td>
<td>400</td>
</tr>
<tr>
<td>ABMD Test</td>
<td>Feb 4-28, 03</td>
<td>33</td>
<td>0-250</td>
<td>25</td>
<td>400</td>
</tr>
<tr>
<td>IBM Training</td>
<td>June 3, 02</td>
<td>92</td>
<td>0-3700</td>
<td>5</td>
<td>4270</td>
</tr>
<tr>
<td>IBM Test</td>
<td>June 4-12, 02</td>
<td>620</td>
<td>0-8000</td>
<td>5</td>
<td>4270</td>
</tr>
</tbody>
</table>

All figures are in multiples of the tick size. Sample size denotes the number of different options that are traded: trading date, put/call, strike, expiration combinations. The last column gives the average number of trades per day for that stock over a year.

References


Perrakis, S. (1988). Preference-free option prices when the stock returns can go up, go down, or stay the same. Advances in Futures and options 3, 209–235.

