VOLATILITY FORECASTING IN A TICK-DATA MODEL

L. C. G. Rogers

University of Bath

Summary. In the Black-Scholes paradigm, the variance of the change in log price during a time interval is proportional to the length t of the time interval, but this appears not to hold in practice, as is evidenced by implied volatility smile effects. In this paper, we find how the variance depends on t in a tick data model first proposed in [1].

1. Introduction. The simple model of an asset price process which is the key to the success of the Black-Scholes approach assumes that the price S_t at time t can be expressed as $\exp(X_t)$, where X is a Brownian motion with constant drift and constant volatility. A consequence of this is that if we consider the sequence $X_{n\delta} - X_{(n-1)\delta}$, $n = 1, \ldots, N$ of log-price changes over intervals of fixed length $\delta > 0$, then we see a sequence of independent Gaussian random variables with common mean and common variance, and we can estimate the common variance $\sigma^2\delta$ by the sample variance in the usual way. Dividing by δ therefore gives us an estimate of σ^2, which (taking account of sample fluctuations) should not depend on the choice of δ - but in practice it does. As the value of δ increases, we see that the estimates tend to settle down, but for small δ (of the order of a day or less) the estimates seem to be badly out of line. Given these empirical observations, we may not feel too confident about estimating σ^2, nor about forecasting volatility of log-price changes over coming time periods. Of course, if we are interested in a particular time interval (say, the time to expiry of an option), we can estimate using this time interval as the value of δ, but this is only a response to the problem, not a solution to it.

The viewpoint taken here is that this problem is due to a failure of the underlying asset model, and various adjustments of the model will never address the basic issue. The basic issue is that the price data simply do not look like a diffusion, at least on a small time scale; trades happen one at a time, and even 'the price' at some time between trades is a concept that needs careful definition. Aggregating over a longer timescale, the diffusion approximation looks much more appropriate, but on shorter timescales we have to deal with quite different models, which acknowledge the discrete nature of the price data.

In this paper, we will consider a class of tick-data models introduced in Rogers & Zane [1], and will derive an expression for

$$v(t) \equiv \mathrm{var}(\log(S_t/S_0)) \tag{1}$$

in this context. Under certain natural assumptions, we find that there exist positive constants σ and b such that for times which are reasonably large compared to the inter-event times of the tick data

$$v(t) \sim \sigma^2 t + b. \tag{2}$$

Research supported in part by EPSRC grant GR/J97281

Section 2 reviews the modelling framework of [1] in the special case of a single asset, and Section 3 derives the functional form of v. Section 4 concludes.

2. The modelling framework. The approach of [1] is to model the tick data itself. An event in the tick record of the trading of some asset consists of three numbers: the time at which the event happened, the price at which the asset traded, and an amount of the asset which changed hands. The assumptions of [1] are that the amounts traded at different events are IID (independent, identically-distributed), and that there is some underlying 'notional' price process z with stationary increments such that the log price y_i at which the asset traded at event time τ_i is expressed as

$$y_i = z(\tau_i) + \varepsilon_i, \tag{3}$$

Here, the noise terms ε_i are independent conditional on $\{\tau_i, a_i; i \in \mathbb{Z}\}$, where a_i denotes the amount traded at the ith event, and the distribution of ε_i depends only on a_i. The rationale for this assumption is that an agent may be prepared to trade at an anomalous price as a way of gaining information about market sentiment, or as a way of generating interest; but he is unlikely to be willing to trade a *large* amount at an anomalous price. In short, large trades are likely to be more keenly priced than small ones. This modelling structure permits such an effect. Of course, we could for simplicity assume that the ε_i were independent with a common distribution.

It remains to understand how the process of times τ_i of events is generated. The model is based on a Markov process X which is stationary and ergodic with invariant distribution π. Independent of X we take a standard Poisson counting process $\tilde{N}$, and consider the counting process

$$N_t \equiv \tilde{N}\Big(\int_0^t f(X_s)ds\Big),$$

where f is a positive function on the statespace of X. As is explained in [1], it is possible to build in a deterministic dependence on time to model the observed pattern of intra-day activity, but for simplicity we shall assume that all such effects have been corrected for. Even when this is done, though, there are still irregularities in the trading of different shares, with periods of heightened activity interspersed with quieter periods, and these do not happen in any predictable pattern. So some sort of stochastic intensity for the event process seems unavoidable; moreover, when we realise that a deterministic intensity would imply that changes in log-prices of *different* assets would be uncorrelated, a stochastic intensity model is more or less forced on us.

The paper [1] presents a few very simple examples, and discusses estimation procedures for them, so we will say no more about that here. Instead we turn to the functional form of v implied by this modelling framework.

3. The functional form of v. The first step in finding the form of v is to determine the meaning of S_t in the expression (1). Since the price jumps discretely, we propose to take as the price at time t the price at the last time prior to t that the asset was traded; if

$$T_t \equiv \sup\{\tau_n : \tau_n \le t\} \equiv \tau_{\nu(t)},$$

then we define $\log S_t \equiv y_{\nu(t)}$. It is of course perfectly possible that for $t > 0$ we may have $T_t = T_0$; this is equivalent to the statement that there is no event in the interval $(0, t]$. We have to bear this possibility in mind. It follows from (3) that

$$\log(\frac{S_t}{S_0}) = z(T_t) - z(T_0) + \varepsilon_{\nu(t)} - \varepsilon_{\nu(0)}, \tag{4}$$

so that

$$E\Big[\log(\frac{S_t}{S_0}) \Big] = E[\, z(T_t) - z(T_0)\,] \tag{5}$$
$$= \mu E[\, T_t - T_0\,]. \tag{6}$$

Here, we have used the assumption that z has stationary increments, which implies in particular that for some μ

$$E[\, z(t) - z(s)\,] = \mu(t - s)$$

for all s, t. Rather remarkably, the expression (6) simplifies. Indeed, because the underlying Markov process X is assumed to be stationary, T_t is the same in distribution as $T_0 + t$, so we have more simply that

$$E\Big[\log(\frac{S_t}{S_0}) \Big] = \mu t. \tag{7}$$

We may similarly analyse the second moment of the change in log price over the interval $(0, t]$:

$$E\Big[\{ \log(\frac{S_t}{S_0})\}^2 \Big] = E[\, (z(T_t) - z(T_0))^2\,] + E[\, (\varepsilon_{\nu(t)} - \varepsilon_{\nu(0)})^2\,]$$
$$= E[\, \mathrm{var}(z(T_t) - z(T_0))\,] + \mu^2 E[\, (T_t - T_0)^2\,] + 2\mathrm{var}(\varepsilon) P[\, T_t > T_0\,], \tag{8}$$

which we understand by noting that if $T_t = T_0$ then $\varepsilon_{\nu(t)} - \varepsilon_{\nu(0)} = 0$, whereas if $T_t > T_0$ then the difference of the ε terms in (4) is the difference of two (conditionally) independent variables both with the same marginal distribution. In general, no simplification of (8) is possible without further explicit information concerning the underlying probabilistic structure. In particular, the term $E[\, (T_t - T_0)^2\,]$ does not reduce simply, and the term

$$P[\, T_t > T_0\,] = 1 - E \exp(-\int_0^t f(X_s)\, ds) \tag{9}$$

cannot be simplified further without knowledge of the process X (and perhaps not even then!) Nevertheless, if we were to assume that *the increments of the notional price process z are uncorrelated* (which would be the case if we took z to be a Brownian motion with constant volatility and drift), then we can simplify

$$E[\, \mathrm{var}(z(T_t) - z(T_0))\,] = \sigma^2 E[\, T_t - T_0\,]$$
$$= \sigma^2 t. \tag{10}$$

Under these assumptions, we may combine and find

$$\text{(11)} \qquad \mathrm{var}\left(\log(\frac{S_t}{S_0})\right) = \sigma^2 t + \mu^2 \mathrm{var}(T_t - T_0) + 2\mathrm{var}(\varepsilon) P[\,T_t > T_0\,].$$

While the exact form of the different terms in (11) may not be explicitly calculable except in a few special cases, the asymptotics of (11) are not hard to understand. The term $\mathrm{var}(T_t - T_0)$ is bounded above by $4ET_0^2$, and tends to zero as $t \downarrow 0$. Assuming that the Markov process X satisfies some mixing condition, we will have for large enough t that

$$\mathrm{var}(T_t - T_0) \doteqdot 2\mathrm{var}T_0.$$

The term $P[\,T_t > T_0\,]$ is increasing in t, bounded by 1, and behaves as $Ef(X_0)$ as $t \downarrow 0$. For times which are large compared to the mean time between trades, this probability will be essentially 1. So except for thinly-traded shares viewed over quite short time intervals, we may safely take the probability to be 1, which justifies the form (2) asserted earlier for the variance of the log price.

4. Discussion and conclusions. We have shown how a natural model for tick data leads us to the functional form

$$\sigma(t) \sim \sqrt{\sigma^2 + b/t}$$

for the 'volatility' $\sigma(t)$ over a time period of length t. This appears to be consistent with observed non-Black-Scholes behaviour of share prices in various ways. Firstly, implied volatility typically decreases with time to expiry, and the 'volatility' in this model displays this feature. Secondly, log returns look more nearly Gaussian over longer time periods, and we may see this reflected here in that if we assume the notional price is a Brownian motion with constant volatility and drift, then the log return is a sum of a Gaussian part (the increment of z) and two noise terms with common variance. For small times, the noise terms dominate, but as the time interval increases, the variance of $z(t)$ increases while the variance of the two noise terms remains constant; it follows that the distribution will look more nearly Gaussian for longer time periods, but could be very different for short time periods. Thirdly, there is the empirical result of Roll [2] who studies the direction of successive price jumps in tick data, and finds that the next price change is much more likely to be in the opposite direction from the one just seen; this is easily explained by a model in which there is some notional underlying price, and observed prices are noisy observations of it.

Given tick data on some asset, the ideal would be to fit the entire Markovian intensity structure of Section 2, though this may not always be easy. However, in terms of forecasting volatility, if we accept the modelling assumptions which led to (2), this level of fitting is not needed. We could form estimates $\hat{\sigma}(\delta_i)$ of the variance of $\log(S(\delta_i)/S(0))$ for a range of time intervals δ_i (for example, hourly, daily, weekly and monthly) and then fit the functional form (2) to the estimates, a linear regression problem. Of course, we would want to be confident that all of the time intervals δ_i chosen were long enough for negligible probability of no event in such an interval; but if that is not satisfied, how are we going to be able to form the estimator $\hat{\sigma}(\delta_i)$?! In this way, we are able to extract *more* information from the record of tick-by-tick data than would have been possible had we imposed the log-Brownian model on that data. It seems likely that tick data *should* tell us much more than just a record of end-of-day prices, but until we have suitable models of tick data, we cannot hope to extract this additional information.

References

[1] ROGERS, L. C. G., & ZANE, O.. Designing and estimating models of high-frequency data. University of Bath preprint, 1998.

[2] ROLL, R.. A simple implicit measure of the effective bid-ask spread in an efficient market. *J. Finance* **39**, 1127–1139, 1984.

Financial Studies Group

Department of Mathematical Sciences

University of Bath

Bath BA2 7AY

phone: +44 1225 826224

fax: +44 1225 826492

email: lcgr@maths.bath.ac.uk

web: http://www2.bath.ac.uk/ maslcgr/home.html