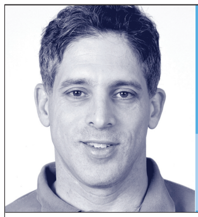

Kent Osband

FINFORMATICS

Rootless Vol

One of the strangest features of Brownian motion is that the parameter that causes all the uncertainty is not supposed to be uncertain...

Even if you started out clueless about the volatility σ, given a good enough measuring stick and fast enough hands you ought to be able to measure it quickly and accurately. In practice you have to settle for wobbly estimation. Less wobbly than estimation of the mean μ. But a whole lot wobblier than you'd like.

Let me quickly review the problem. Suppose we observe Brownian motion for a short period Δt. The observation Δx equals $\mu\Delta t + \sigma\varepsilon\sqrt{\Delta t}$ for some noisy Gaussian ε with mean 0 and variance 1. If we square Δx and divide by Δt we obtain

$$\frac{(\Delta x)^2}{\Delta t} = \mu^2\Delta t + \mu\sigma\varepsilon\sqrt{\Delta t} + \sigma^2\varepsilon^2$$

which has expected value $\sigma^2 + \mu^2\Delta t$ and variance $2\sigma^4 + \mu^2\sigma^2\Delta t$. To refine our estimate further, let's divide the original period into N non-overlapping intervals, not necessarily of equal length, and average the measurements $(\Delta x_i)^2/\Delta t_i$. The average has expectation

$$E\left[\frac{1}{N}\sum_{i=1}^{n}\frac{(\Delta x_i)^2}{\Delta t_i}\right] = \sigma^2 + \frac{\mu^2}{N}\sum_{i=1}^{n}\Delta t_i = \sigma^2 + \mu^2\frac{\Delta t}{N}$$

and variance – thanks to the Brownian assumption of independence across non-overlapping time intervals – of

$$Var\left[\frac{1}{N}\sum_{i=1}^{n}\frac{(\Delta x_i)^2}{\Delta t_i}\right] = \frac{2\sigma^4}{N} + \frac{\mu^2\sigma^2}{N^2}\sum_{i=1}^{n}\Delta t_i = \frac{2\sigma^4}{N} + \mu^2\sigma^2\frac{\Delta t}{N^2}$$

As N approaches infinity, both the bias and the variance vanish, leaving the average infinitesimally close to the true mean σ^2. And this is regardless of the total observation length Δt. So it's child's play for an armchair theorist to get 99 per cent confident within a second of a range less than one per cent wide around the true variance. He just needs to imagine a measly million observations in a second.

You're not impressed? Well, you should be. The empirical rate of drift $\Delta x/\Delta t$ will have mean μ with variance $\sigma^2/\Delta t$. So you have an unbiased mean but the variance approaches zero as Δt shrinks. Subdividing the interval into N parts won't help. Your best estimator always works out to $\Delta x/\Delta t$ with variance $\sigma^2/\Delta t$ so you might as well save the extra slicing.

The only way to identify μ is to wait. And in finance you might have to wait a long while. For example, suppose you proxy the Sharpe ratio μ/σ by the empirical Sharpe $\Delta x/\sigma\Delta t$. The standard deviation of the latter is one. So to get 99 per cent confident that the empirical annualized Sharpe is within 0.5 of its true value you would need over 25 years of data. Relevant data that is, when the drift was just what it is now. Good luck.

Deviously Standard Estimators

In practice σ is indeed much easier to estimate than μ. It is an estimate though and not an asymptotically precise measure. Why not? One explanation which surely is true is that our measuring sticks won't let us. Between discrete quantization of values, bid/ask spreads, and variations in effective trading time, both our $(\Delta x)^2$ and Δt measures have errors that don't scale down proportionally to means. The second explanation, which probably is true too, is that at the core level the underlying processes aren't Brownian: they're serially correlated and riddled with jumps.

So what do we do? The standard way is to measure a lot of $(\Delta x)^2/\Delta t$ values for N small equal-length intervals Δt and average them together. To get fancier, subtract off the sample mean before squaring and multiply the average by $\frac{N}{N-1}$. To estimate σ take the square root of the estimate for σ^2.

Ideally you want to take lots of these estimates, down to the limits of granularity, and update them every new measurement. This leads to the standard EWA (exponentially weighted average) method of estimating volatility. Suppose we seed a variance estimator $\hat{\sigma}^2$ in some fashion Defining $y \equiv \frac{\Delta x - \mu\Delta t}{\sqrt{\Delta t}}$ and using the lag operator L to denote the lagged value of a variable, let us update the variance estimator as follows:

$$\hat{\sigma}^2 = \lambda(y^2 - L\hat{\sigma}^2) + L\hat{\sigma}^2$$
$$\hat{\sigma} = \sqrt{\hat{\sigma}^2}$$

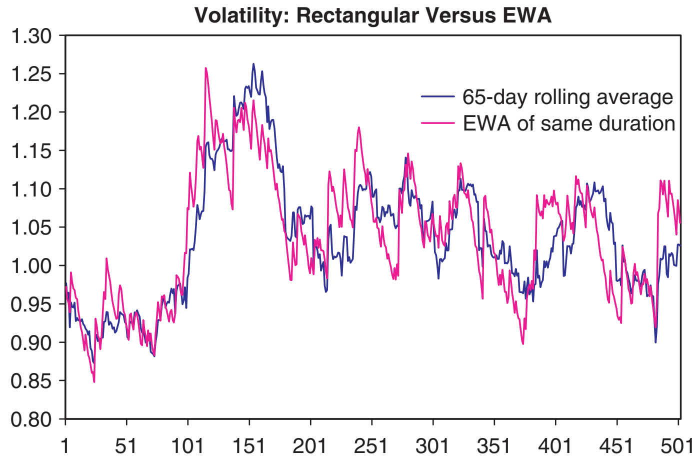

Figure 1:

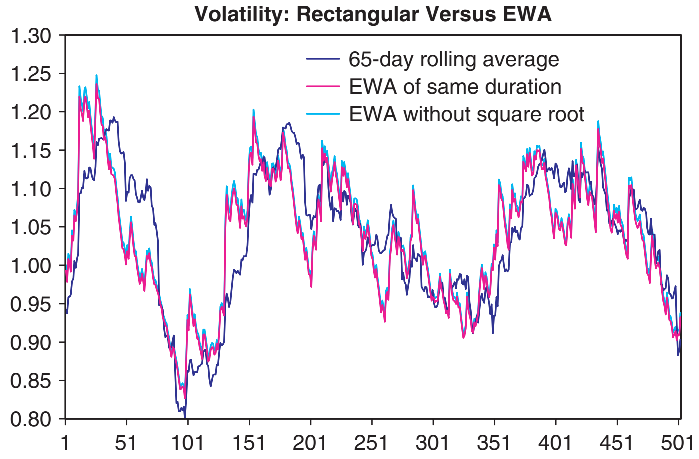

Figure 2:

for some nudge factor $\lambda < 1$. This implies weights of $\lambda(1-\lambda)^n$ on an observation lagged n periods, which approximates exponential decay if stretched ad infinitum.

The higher λ, the more you weight the most recent observations. Note that the duration D satisfies the recursion

$$D = \lambda(0 - (D+1)) + D + 1$$

which implies $\lambda = \frac{1}{D+1}$. A rectangular (equally weighted) series of N observations would have duration $D = \frac{N-1}{2}$. The two series will have equal durations if $\lambda = \frac{2}{N+1}$. Above I have graphed Monte Carlo simulations of volatility estimation for a standard Gaussian series. The rectangular and EWA estimators, each with 65 days' duration, track each other reasonably well though not perfectly.

This EWA updating rule is the basic workhorse of most industrial-strength vol estimation. It's easy to understand and simple to apply. However, there are a few problems with it.

- It's annoying to keep squaring and taking the square root each time, when in most cases the new square root doesn't differ that much from the old.
- The updating rule tends to overreact to dirty data. A 10 standard deviation outlier will typically more than double the vol estimate, when on balance it would be better to ignore it.
- It's hard to justify this rule theoretically.

Recall from previous Finformatics columns that an EWA can be viewed as a reduced form of the appropriate Bayesian updating rule if both observations and beliefs are normally distributed with fixed variances. But the various y^2 aren't normally distributed. They're distributed much closer to chi-squared, which is much higher-skew and higher-kurtosis even before you add dirty data. An optimal updating rule ought to trim the nudge factor λ as y^2 rises. No wonder the measure is so sensitive to outliers.

EWA Updating Without Square Roots

In searching for alternative vol estimators, let's start by getting rid of the square root. Combining the two updating parts into one and applying some algebra:

$$\hat{\sigma} = \sqrt{\hat{\sigma}^2} = \sqrt{\lambda(y^2 - L\hat{\sigma}^2) + L\hat{\sigma}^2} = L\hat{\sigma}\sqrt{1 + \lambda\left(\frac{y^2}{L\hat{\sigma}^2} - 1\right)}$$
$$\cong L\hat{\sigma}\left(1 + 1/2\lambda\left(\frac{y^2}{L\hat{\sigma}^2} - 1\right)\right) = L\hat{\sigma} + 1/2\lambda\left(\frac{y^2}{L\hat{\sigma}} - L\hat{\sigma}\right)$$

where the approximation follows from the first-order Taylor expansion $\sqrt{1+\delta} \cong 1 + 1/2\delta$.

Below I've graphed another Monte Carlo simulation. You'll see that the two EWA versions are very close. There's no need to take the square root.

However, the Taylor approximation doesn't reduce the sensitivity to large outliers. Indeed it slightly exacerbates it. For example, if the estimated variance doubles, the square root rises 42 per cent whereas the linear approximation would up the volatility estimate 50 per cent.

Taking a second-order Taylor approximation runs the risk of generating negative vols. A Pade(1,1) approximation to the square root works better. It approximates $\sqrt{1+\delta}$ by $1 + \frac{2\delta}{4+\delta}$, eg 1.4 for $\sqrt{2}$. It also has the mathematically false but empirically appealing implication that the vol can never triple. Hence for an $\hat{\sigma}$ estimator much more robust to extremes than the standard EWA, update as:

$$\hat{\sigma} = \left(1 + \frac{2\delta}{4+\delta}\right)L\hat{\sigma} \text{ for } \delta = \lambda\left(\frac{y^2}{L\hat{\sigma}} - 1\right)$$

Vol Estimation Using Absolute Deviations

The previous estimators remain sensitive to outliers, outliers that the observations y^2 are all too prone to create. Can we not transform the

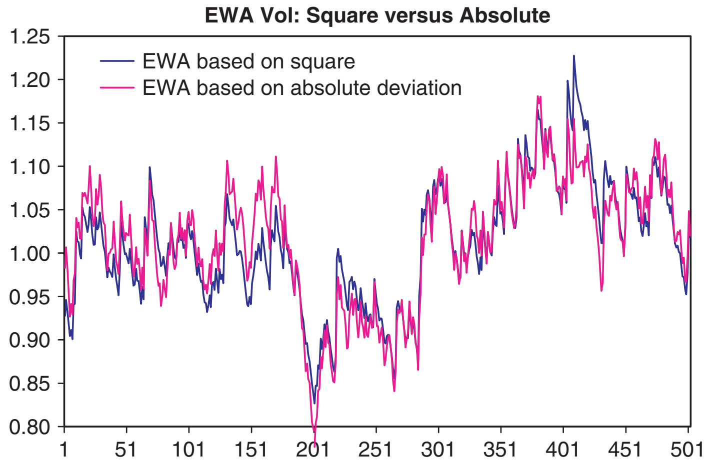

Figure 3:

observations to make them more normal? An obvious candidate is to take the square root of the square, that is, the absolute value of y. If y is Gaussian, $|y|$ will have skew 0.995 and kurtosis 0.87, whereas y^2 has a skew of 2.83 and kurtosis of 12.

Since the mean of a Gaussian $|y|$ will be $\sigma\sqrt{\frac{2}{\pi}} \cong 0.7979\sigma$, we need to multiply the observed $|y|$ by 1.2533 to produce an unbiased estimator of σ. This suggests an EWA of the form:

$$\sigma_{ABS} = L\sigma_{ABS} + \lambda(1.2533\,|y| - L\sigma_{ABS})$$

Below is a Monte Carlo simulation for two EWAs, each with 65 days' duration but one based on squares and the other on absolute deviations. They track each other closely except for large deviations where the absolute deviation-based vol is more stable. This is exactly what you'd expect. A four-standard deviation event will up σ_{ABS} by a fraction $(1.2533 * 4 - 1)\lambda \cong 4\lambda$ but up $\hat{\sigma}$ by $(4 * 4 - 1)\lambda = 15\lambda$.

To make the absolute deviation-based EWA more robust to large outliers, one can cap $|y|$ at say 4 times the current standard deviation, creating something like this:

$$\sigma_{ABS} = \left(1 + \lambda \min\left(4, \frac{1.2533\,|y|}{L\sigma_{ABS}} - 1\right)\right) L\sigma_{ABS}$$

Biases from Using EWAs

Let me caution that σ_{ABS} is an unbiased estimator of vol only when deviations are normal. For fat-tailed distributions it will understate vol. How much? To get a better sense of the size of the effect, let's assume the probability distribution can be approximated by a mixture of two normal densities centered on μ. Note this rules out skewness. The first density, which applies ν times as frequently as the other, is assumed to have variance $(1-\eta)\sigma^2$ for some η between zero and one. For the total variance to equal σ^2, the other density must have variance $(1+\nu\eta)\sigma^2$. To generate a kurtosis of K, we must have

$$\begin{aligned} 3 + K &= \frac{3(1-\eta)^2\nu + 3(1+\nu\eta)^2}{\nu+1} \\ &= \frac{3\nu - 6\eta\nu + \eta^2\nu + 3 + 6\nu\eta + 3\nu^2\eta^2}{\nu+1} \\ &= 3 + 3\nu\eta^2 \end{aligned}$$

Hence $\nu\delta^2 = 1/3K$, implying

$$\begin{aligned} E[1.2533\,|y|] &= \frac{\nu\sqrt{1-\eta} + \sqrt{1+\nu\eta}}{1+\nu}\sigma \\ &\cong \frac{\nu(1 - 1/2\eta - 1/8\eta^2) + (1 + 1/2\nu\eta - 1/8\nu^2\eta^2)}{1+\nu} \\ &= 1 - \frac{1}{8}\nu\eta^2 \\ &= 1 - \frac{K}{24} \simeq 1 - 0.0417K \end{aligned}$$

So the understatement grows with kurtosis.

The square-based $\hat{\sigma}$ is not an unbiased estimator of σ either, only for a different reason. Given any unbiased but wobbly estimator s of σ, s^2 must tend to overstate σ^2 while $1/s$ will tend to understate $1/\sigma$. This follows from Jensen's inequality, a notion known to every options trader as "positive convexity is good''. If you take second-order Taylor approximations and work thru the implications, you will find that $\hat{\sigma}$ tends to overstate σ by a factor of $\left(\frac{1+1/2K}{4}\right)\lambda$. At least I hope you will.

Vol Estimation Using Signs

For even more robust estimation we can simply monitor whether a given observation exceeds a specified fraction of the current vol estimate. For example, with normality y has even odds of falling within 0.6745 standard deviations of the mean. Suppose we use the updating rule

$$\sigma_{SIGN} = (1 + 0.72\lambda * Sign(|y| - 0.6944L\sigma_{SIGN}))L\sigma_{SIGN}$$

I scaled λ by the factor 0.72 to ensure that the standard deviation of changes was comparable to previous measures.

As constructed, the nudge has a sub-normal kurtosis of −2. A variant that has zero mean and zero kurtosis under normality of y is

$$\sigma_{SIGN} = (1 + 0.72\lambda * if(|y| > 1.25L\sigma_{SIGN}, +0.79. - 0.21))\,L\sigma_{SIGN}$$

Here is a chart of Monte Carlo simulations under normality. The sign measure tracks better than you might expect. It's noisier with respect to small deviations and less responsive to regime change but on the other hand more robust to big outliers.

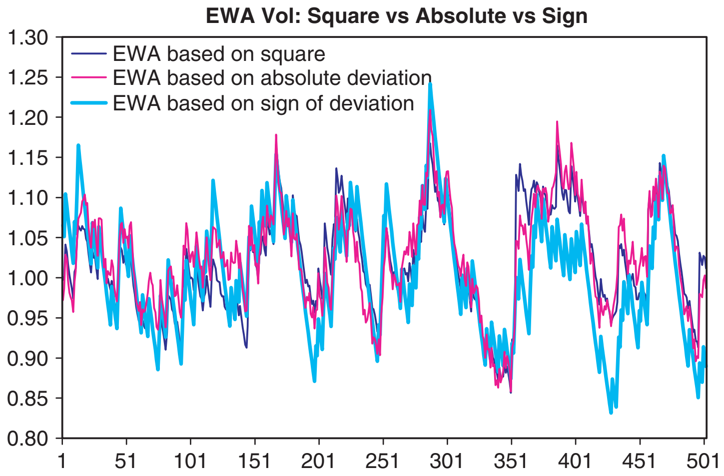

Figure 4:

Rootless Range Estimators of Vol

These aren't the only ways to estimate volatility without taking square roots. One appealing method is to use $\frac{1}{2\sqrt{\ln 2}} \cong 60\%$ of the day's trading range. It's much more precise than the daily close though you may need to calibrate the multiplier to mitigate real-word distortions. A measure based on 84% of the range less 39% of the absolute return is even more precise though again you'll need to calibrate. In Monte Carlo simulations I have found it more precise than the better-known Garman-Klass estimators using square roots.

Regime Switching Across Vol Estimators

It's rarely clear which type of vol estimator we should choose or at what duration. Fortunately we don't need to. We can take a belief-weighted average of vol predictors to use as a consensus, and set up regime-switching models to update beliefs. It's not perfect because we won't know the meta-parameters that define the switching rates. Fortunately it needn't be perfect to be useful. A simple average of short-term and long-term EWAs of the same type often outperforms substantially more complicated GARCH models. By tossing in more types and allow the averages to dynamically update, you can improve on this. However you should also allow persistently weaker models to prune themselves out.

Regime switching methods can also be useful in calibrating transforms of vol. One such transform is the inverse vol. It tells us how much of an asset or portfolio we need to generate unit volatility. Only not quite. By Jensen's equality again, an asset sized by its estimated inverse vol will tend to average a vol greater than one. Concretely,

$$E\left[\frac{y}{\sigma_{est}}\right] = E\left[\frac{y}{\sigma(1 \pm \varepsilon)}\right] \cong E\left[\frac{y}{\sigma}\right] \cdot E\left[1 \mp \varepsilon + \varepsilon^2\right] = 1 + var\,[\varepsilon]$$

Instead of estimating variance from first principles it may be substantially easier to decide the correction empirically, again using an overlay of EWAs.