

ORF 474: High Frequency Trading High Frequency Volatility

Robert Almgren

March 31, 2020

Suppose that $X(t)$ is a price or log price process, that satisfies

$$dX(t) = \sigma(t) dB(t)$$

where $B(t)$ is a Brownian motion. The volatility process $\sigma(t)$ may be random or nonrandom, but it should be continuous. We observe one realization of $X(t)$ for $0 \leq t \leq T$, for example, one day of intraday tick data. Along with this, of course, will be one realization of $\sigma(t)$. We assume that any drift term is negligible, which is generally adequate for high-frequency data. What can we tell about the volatility?

A fundamental fact about Itô processes says that from observation of $X(t)$ over a very short time interval, we can determine the instantaneous volatility as precisely as we like, given arbitrarily frequent observations. But in fact, we observe $X(t)$ only at discrete times t_0, t_1, \dots, t_N , where we assume $t_0 = 0$ and $t_N = T$, and we denote $x_j = X(t_j)$ (Figure 1). So we will not be able to determine the volatility precisely at each instant.

We set the more modest goal of determining the *total variance*

$$Q = \int_0^T \sigma(t)^2 dt \quad \left(\approx \sigma^2 T \text{ if } \sigma(t) \text{ is constant} \right).$$

from the observations x_0, \dots, x_N . Of course, from this we can determine the average volatility $\langle \sigma^2 \rangle = Q/T$, and if we have enough data points, we can subdivide the interval $[0, T]$ to estimate local values of $\sigma(t)^2$. It does not matter what process governs $\sigma(t)$; this is a statement about the single realization observed on this day.

Thus, let us consider $\sigma(t)$ as a fixed function, and consider the family of random realizations of $X(t)$ as driven by $B(t)$. Then

$$x_j - x_{j-1} = \int_{t_{j-1}}^{t_j} \sigma(t) dB(t)$$

and hence $x_j - x_{j-1}$ is normal with

$$\mathbb{E}(x_j - x_{j-1}) = 0, \quad \text{Var}(x_j - x_{j-1}) = Q_j \equiv \int_{t_{j-1}}^{t_j} \sigma(t)^2 dt.$$

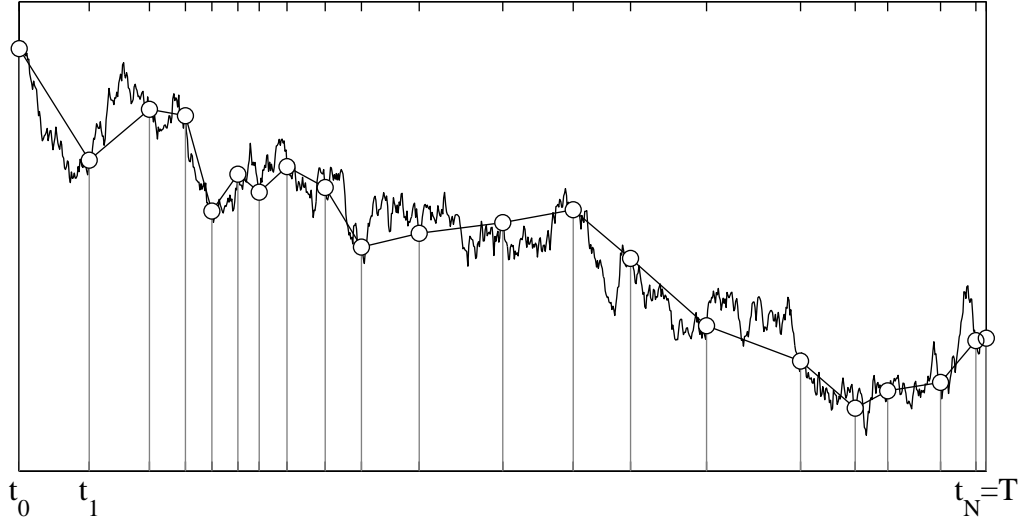


Figure 1: The ideal case: we have exact observations x_j of $X(t)$ at times t_0, \dots, t_N . We can recover the underlying volatility reasonably accurately by taking the realized variance of the observations.

Thus $(x_j - x_{j-1})^2$ has a χ^2 distribution with one degree of freedom and

$$\mathbb{E}((x_j - x_{j-1})^2) = Q_j, \quad \text{Var}((x_j - x_{j-1})^2) = 2Q_j^2.$$

Note that if $\sigma(t)$ is approximately constant, then $Q_j \approx \sigma^2(t_j - t_{j-1})$.

Estimation of volatility If $\sigma(t)$ is constant, then $Q_j = \sigma^2(t_j - t_{j-1})$ and the best estimator for σ based on a single interval is

$$\hat{\sigma}_j^2 = \frac{(x_j - x_{j-1})^2}{t_j - t_{j-1}},$$

having

$$\mathbb{E}(\hat{\sigma}_j^2) = \sigma^2 \quad \text{and} \quad \text{Var}(\hat{\sigma}_j^2) = 2\sigma^4.$$

The variance of this estimator is independent of the time interval. Thus, if we wanted to combine these interval estimators into a single estimator for the overall (constant) volatility parameter, we should average them with equal weight as

$$\hat{\sigma}^2 = \frac{1}{N} \sum_{j=1}^N \frac{(x_j - x_{j-1})^2}{t_j - t_{j-1}}.$$

We do not weight them by $t_j - t_{j-1}$. This estimator has

$$\mathbb{E}(\hat{\sigma}^2) = \sigma^2 \quad \text{and} \quad \text{Var}(\hat{\sigma}^2) = \frac{2\sigma^4}{N}.$$

It is unbiased, and its error is the smallest possible.

Estimation of realized variance If we do not assume $\sigma(t)$ constant, then we may directly estimate the total variance Q as the *realized variance*

$$\text{RV} = \sum_{j=1}^N (x_j - x_{j-1})^2.$$

We do not divide by N or $t_j - t_{j-1}$. This estimator has

$$\mathbb{E}(\text{RV}) = \sum_{j=1}^N Q_j = Q, \quad \text{Var}(\text{RV}) = 2 \sum_{j=1}^N Q_j^2.$$

If N is large, if the $\{t_j\}_{j=1}^N$ are reasonably well spaced, and if the volatility is not too far from constant, then we can write

$$\text{RV} = Q + \xi, \quad \text{with } \text{Var}(\xi) \sim \left(\frac{\sigma^4 T^2}{N} \right),$$

This justifies using RV as an approximation for Q in any particular realization of $x(t)$, regardless of the volatility process.

This estimator and the one above differ only in whether or not they weight $(x_j - x_{j-1})^2$ by $t_j - t_{j-1}$. Which choice is better? If we assume that $\sigma(t)$ is constant, then realized variance also gives us an estimator

$$\hat{\sigma}_{\text{RV}}^2 = \frac{\text{RV}}{T}$$

with

$$\mathbb{E}(\hat{\sigma}_{\text{RV}}^2) = \frac{Q}{T} = \sigma^2, \quad \text{Var}(\hat{\sigma}_{\text{RV}}^2) = \frac{2}{T^2} \sum_{j=1}^N Q_j^2 = \frac{2\sigma^4}{T^2} \sum_{j=1}^N (t_j - t_{j-1})^2.$$

We compare this estimator to the one above by calculating

$$\frac{\text{Var}(\hat{\sigma}_{\text{RV}}^2)}{\text{Var}(\hat{\sigma}^2)} = \frac{N}{T^2} \sum_{j=1}^N (t_j - t_{j-1})^2 = \frac{\frac{1}{N} \sum (t_j - t_{j-1})^2}{\left(\frac{1}{N} \sum (t_j - t_{j-1}) \right)^2}.$$

This is the ratio of the average of the square of the time intervals to the square of the average time interval. Hence it is always ≥ 1 , and equal to one only if all the time intervals

are equally spaced so that $t_j - t_{j-1} = T/N$. If volatility is constant, then realized variance always gives a *worse* estimator of the constant value σ . However, it is more flexible if volatility varies in some unknown way (as it usually does) and is preferred for that reason.

For example, a typical application would be to compute a total volatility number for one day of trading, without trying to model the intraday volatility profile. It would be sufficient to sample at say 5-minute intervals, and simply add the square price changes to get the daily number.

High-low data Garman and Klass [1980] point out the possible improvement of volatility estimation using high-low data. Suppose that in addition to the point observations $x_j = X(t_j)$, we also have the interval maxima and minima

$$H_j = \sup_{t_{j-1} \leq t \leq t_j} X(t) \quad \text{and} \quad L_j = \inf_{t_{j-1} \leq t \leq t_j} X(t).$$

Then an improved estimator for the realized variance is

$$\text{RV}_{\text{GK}} = \sum_{j=1}^N \left(\frac{1}{2} (H_j - L_j)^2 - (2 \log 2 - 1) (x_j - x_{j-1})^2 \right).$$

The variance of this estimator is approximately 1/8 of that of the simple estimator. In their paper, Garman and Klass are interested in using daily open-close-high-low data, and so they explicitly consider the overnight gap (in our setup there is no gap since x_j is the close of the j th interval and the open of the $(j + 1)$ st.) They also consider the effect of discrete sampling on the underestimation of the high-low spread.

Microstructure Noise

Real intraday data is not as clean as a simple Brownian motion model, on a fine time scale (Figure 2):

1. Prices are restricted to discrete levels. Trade prices typically occur at quote levels and thus are also quantized, although sometimes prices can be reported as averages of several prints. Modeling of this effect is complicated since successive errors are not independent.
2. Trade prices typically move back and forth between the bid and offer levels; this is called “bid-ask bounce.” This source of noise generally has some positive serial correlation though only a small amount. This noise is the easiest to model, as in the Roll model for trade prices, but it is also easy to remove by using the bid-ask midpoint as the “true” price. Typically, in order to preserve the role of trade arrival rate, one would use the quote midpoint in effect at the time of each trade, assuming that the relative timing of quotes and trades is sufficiently reliable.

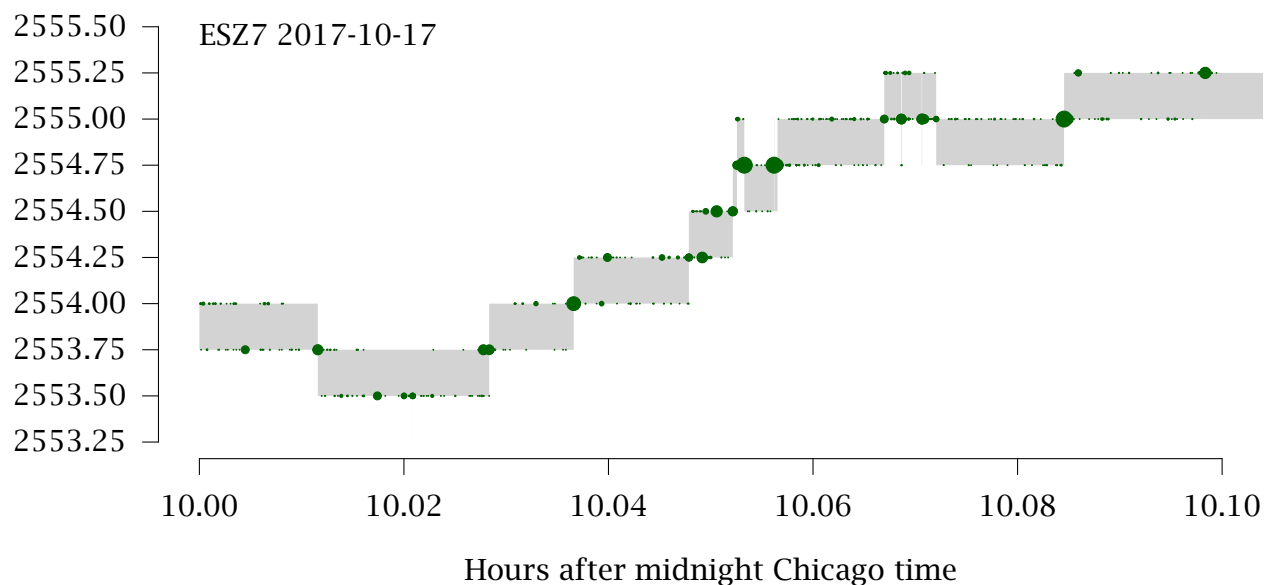


Figure 2: Six minutes of intraday tick data (the S&P500 futures ESZ7 on Oct 17, 2017). The gray region is the best bid and offer; for this large-tick product the spread is almost always equal to the tick size 0.25. Dots are trades, with size proportional to trade size.

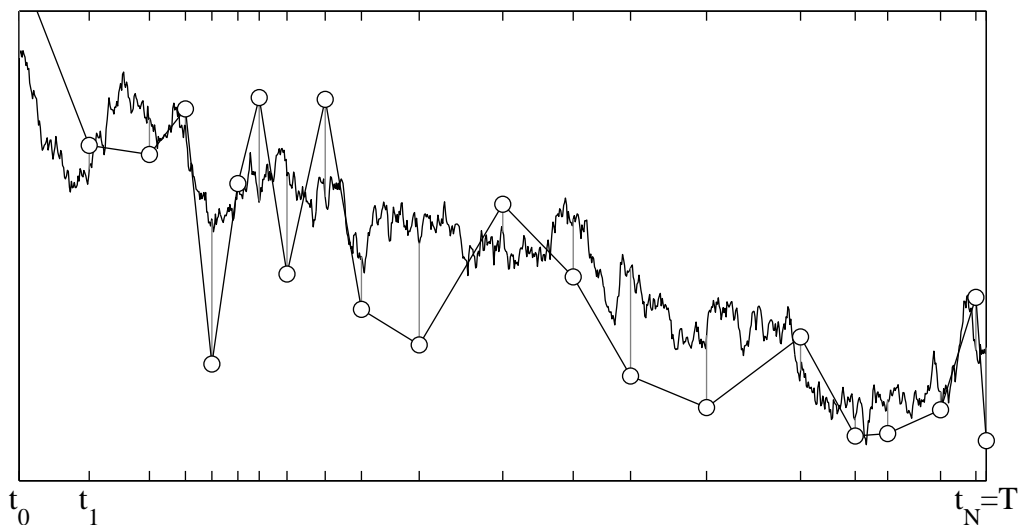


Figure 3: Observed values y_j are the Brownian motion plus noise. The realized variance has a large component from the noise, added to the volatility of the underlying process.

Microstructure noise The easiest way to incorporate microstructure noise is a state space model. We suppose that the observed values are not x_j but

$$y_j = x_j + \epsilon_j$$

where the ϵ_j are noise terms (Figure 3). The simplest assumption is that the ϵ_j are i.i.d. with mean zero and $\text{Var}(\epsilon_j) = \eta^2$, and that the ϵ_j are independent of $x(t)$. It is not difficult to extend the model to permit serial autocorrelation in the ϵ_j .

Clearly, this noise model is very different from real microstructure. It might be a reasonable way to incorporate bid-ask bounce, but we have pointed out that that noise is easy to remove by using the midpoint price. Discretization errors are very different from this. But this model is simple enough to allow us to develop some candidate improved versions of the realized variance estimator and to calculate their theoretical properties. Of course the final test is how well these estimators perform on real data.

Applying RV to the observed values y_j , we get

$$\begin{aligned} \text{RV} &= \sum_{j=1}^N (y_j - y_{j-1})^2 \\ &= \sum_{j=1}^N (x_j - x_{j-1})^2 + 2 \sum_{j=1}^N (x_j - x_{j-1})(\epsilon_j - \epsilon_{j-1}) + \sum_{j=1}^N (\epsilon_j - \epsilon_{j-1})^2. \end{aligned}$$

We analyse the three terms as follows.

1. The first term is $Q + \xi_1$, with $\mathbb{E}(\xi_1) = 0$ and $\text{Var}(\xi_1) \sim (\sigma^4 T^2 / N)$.
2. In the second term, since the ϵ_j are independent of the x_j , $\text{Var}((x_j - x_{j-1})(\epsilon_j - \epsilon_{j-1})) = \mathbb{E}((x_j - x_{j-1})^2 (\epsilon_j - \epsilon_{j-1})^2) = Q_j \cdot 2\eta^2$. Thus it is ξ_2 , with $\mathbb{E}(\xi_2) = 0$ and $\text{Var}(\xi_2) \sim (Q\eta^2)$. (Successive terms are not quite independent, but the dependence is short range.)
3. In the third term, each $(\epsilon_j - \epsilon_{j-1})^2$ has mean $2\eta^2$ and variance $\sim \eta^4$. Thus the sum is $2N\eta^2 + \xi_3$, with $\mathbb{E}(\xi_3) = 0$ and $\text{Var}(\xi_3) \sim (N\eta^4)$. (A more precise calculation, shown below, gives $\text{Var}(\xi_3) = 12N\eta^4$.)

For the total variance of RV, we take the three terms as independent; they are not independent but the dependences are not too large. Thus the total variance is $\sim (\sigma^4 T^2 / N) + Q\eta^2 + N\eta^4$. For large N , the second term is dominated by the third and can be neglected. The total variance is then $\sim (\sigma^4 T^2 / N) + N\eta^4 = N(\eta^4 + \sigma^4 \tau^2)$, where again $\tau = T/N$ is a typical time step $t_j - t_{j-1}$. Thus $\sigma^4 \tau^2 = (\sigma \sqrt{\tau})^4$ is the fourth power of a typical change due to volatility over one time step, which is added to the fourth power of the microstructure noise on each observation.

Thus the total expression is

$$\text{RV} = Q + 2N\eta^2 + \xi \quad \text{with } \text{Var}(\xi) \sim (N(\eta^4 + \sigma^4 \tau^2)).$$

Realized variance applied to noisy data is biased, and the size of the bias grows linearly with N for fixed T : for large N (fine time resolution), the computed value will be dominated by the microstructure noise rather than by the true variance of the underlying process. The typical error in the observation is $\sim \sqrt{N}$ times the square of the uncertainty per observation. If we sample on a very fine time scale, we will have $\sigma^2 \tau \ll \eta^2$ and the total variance $\approx N\eta^4$.

(To see the “more precise” observation alluded to above, consider

$$S = \sum_{j=1}^N (\epsilon_j - \epsilon_{j-1})^2 = \sum_{j=1}^N (\epsilon_j^2 - 2\epsilon_j \epsilon_{j-1} + \epsilon_{j-1}^2).$$

Clearly $\mathbb{E}(S) = 2N\eta^2$, and then

$$\begin{aligned} (S - 2N\eta^2)^2 &= \left[\sum_{j=1}^n ((\epsilon_j^2 - \eta^2) + (\epsilon_{j-1}^2 - \eta^2) - 2\epsilon_j \epsilon_{j-1}) \right]^2 \\ &= \sum_{i,j=1}^N \left[(\epsilon_i^2 - \eta^2)(\epsilon_j^2 - \eta^2) + (\epsilon_{i-1}^2 - \eta^2)(\epsilon_{j-1}^2 - \eta^2) + 4\epsilon_i \epsilon_{i-1} \epsilon_j \epsilon_{j-1} \right. \\ &\quad \left. + 2(\epsilon_i^2 - \eta^2)(\epsilon_{j-1}^2 - \eta^2) - 4(\epsilon_i^2 - \eta^2)\epsilon_j \epsilon_{j-1} - 4(\epsilon_{i-1}^2 - \eta^2)\epsilon_j \epsilon_{j-1} \right]. \end{aligned}$$

On taking the expectation of the last expression, all the terms vanish except the ones in the first row for $i = j$, and the first term in the second row for $i = j - 1$. Then since $\mathbb{E}(\epsilon_j^2) = \eta^2$ and $\mathbb{E}((\epsilon_j^2 - \eta^2)^2) = 2\eta^4$, we get finally $\text{Var}(S) = \mathbb{E}((S - 2N\eta^2)^2) = 12N\eta^4$.

Subsampling One way to reduce the effect of the microstructure noise is to sample at a lower frequency than every tick. For any $k \geq 1$, consider the modified version of RV

$$\begin{aligned} \text{RV}(k) &= \sum_{j=1}^{m(k)} (\mathcal{Y}_{jk} - \mathcal{Y}_{(j-1)k})^2 \\ &= (\mathcal{Y}_k - \mathcal{Y}_0)^2 + (\mathcal{Y}_{2k} - \mathcal{Y}_k)^2 + \cdots + (\mathcal{Y}_{mk} - \mathcal{Y}_{(m-1)k})^2 \end{aligned}$$

The top value $m(k)$ of the sum is the largest j for which $jk \leq N$. That is,

$$mk \leq N < (m+1)k \quad \implies \quad m(k) = \left\lfloor \frac{N}{k} \right\rfloor$$

($\lfloor \cdot \rfloor$ is the “floor” function). This is simply RV computed on a grid of size $m(k) \approx N/k$ rather than N , and hence

$$\text{RV}(k) \sim Q + \frac{2N}{k} \eta^2 + \xi.$$

The first term is the desired value Q . The second term is the bias in the estimator, which is reduced by taking k large (sampling infrequently). The third term is the statistical error in the estimator, whose size is

$$\text{Var}(\xi) \sim \left(\frac{N}{k} (\eta^4 + \sigma^4(k\tau)^2) \right) \sim \left(\frac{N\eta^4}{k} + \frac{\sigma^4 T^2 k}{N} \right).$$

Note that the time step appearing is $k\tau = kT/N$, the time interval on which we sample.

Typically people choose k large enough so that volatility dominates microstructure noise. Making the second term in the variance larger than the first requires

$$\frac{\sigma^4 T^2 k}{N} \gg \frac{N\eta^4}{k} \iff k \gg \frac{\eta^2}{\sigma^2 \tau}$$

If you are willing to make k that large, then you can ignore microstructure noise completely and just use realized variance; this is why people commonly use *ad hoc* sampling intervals of one minute or five minutes, for stocks that trade several times per second. The disadvantage of that approach is that you throw out almost all of your data.

A more systematic approach is to choose k so that the bias in the estimator is roughly the same size as the RMS size of the random error:

$$\frac{N}{k} \eta^2 \sim \sigma^2 T \sqrt{\frac{k}{N}} \iff k \sim N \left(\frac{\eta^2}{\sigma^2 T} \right)^{2/3} \sim N^{1/3} \left(\frac{\eta^2}{\sigma^2 \tau} \right)^{2/3}$$

Here $\eta^2/\sigma^2 \tau$ is the inverse of the signal-to-noise ratio: the noise introduced by the artificial microstructure on each observation, compared to the effect of volatility on one average interval. Bandi and Russell [2006] give a fuller discussion.

Improved Estimators

Besides the rather crude subsampling described above, some more sophisticated techniques are available.

As a preliminary, note that one can equally well start the above sum at an arbitrary index h rather than 0. For a given k , and for any h with $0 \leq h < k$, set

$$\begin{aligned} \text{RV}(k, h) &= \sum_{j=1}^{m(k, h)} (\mathcal{Y}_{h+jk} - \mathcal{Y}_{h+(j-1)k})^2 \\ &= (\mathcal{Y}_{h+k} - \mathcal{Y}_h)^2 + (\mathcal{Y}_{h+2k} - \mathcal{Y}_{h+k})^2 + \cdots + (\mathcal{Y}_{h+mk} - \mathcal{Y}_{h+(m-1)k})^2 \end{aligned}$$

Here, $m(k, h)$ is the largest value of j so that $h + jk \leq N$; as above,

$$h + mk \leq N < h + (m+1)k \implies m(k, h) = \left\lfloor \frac{N-h}{k} \right\rfloor.$$

This estimator has exactly the same properties as $RV(k)$. For any value of h , this estimator has expected value $Q + (2N/k)\eta^2$, and variance k/N .

Each estimator $RV(k, h)$ uses only a fraction $1/k$ of the data. But we may average together *all* the possible offsets h to write

$$\begin{aligned} R(k) &= \frac{1}{k} \sum_{h=0}^{k-1} RV(k, h) \\ &= \frac{1}{k} \left((y_k - y_0)^2 + (y_{k+1} - y_1)^2 + \cdots + (y_N - y_{N-k})^2 \right) \end{aligned}$$

Since this is an average of k terms each with the same mean, it again has mean $Q + (2N/k)\eta^2$. The factor $1/k$ accounts for the fact that the terms in the sum overlap by a factor of k .

The variance of $R(k)$ is lower than the variance of each $RV(k, h)$ due to the averaging, but the exact amount of reduction is not easy to calculate. If the quantities $RV(k, h)$ were *independent* for different values of h , then the reduction would be a factor of k , but they are strongly dependent since they are differences on overlapping intervals. Certainly this is better than $RV(k, h)$ for any single value of h since it uses all the data, and we shall take the variance reduction as an unspecified benefit.

Multiscale sampling Zhang et al. [2005] propose the following improvement. Suppose we compute the estimator $R(k)$ for two different values k_1 and k_2 , on the same set of data. We have

$$\begin{aligned} R(k_1) &\sim Q + \frac{2N}{k_1} \eta^2 + \langle \text{mean-zero noise} \rangle \\ R(k_2) &\sim Q + \frac{2N}{k_2} \eta^2 + \langle \text{mean-zero noise} \rangle. \end{aligned}$$

We then consider the linear combination

$$R(k_1, k_2) = a_1 R(k_1) + a_2 R(k_2).$$

We choose the coefficients a_1 and a_2 to cancel the unwanted bias:

$$a_1 + a_2 = 1 \quad \text{and} \quad a_1 \frac{2N}{k_1} + a_2 \frac{2N}{k_2} = 0$$

which gives

$$a_1 = \frac{k_1}{k_1 - k_2}, \quad a_2 = -\frac{k_2}{k_1 - k_2}.$$

Then $R(k_1, k_2)$ is an *unbiased* estimator of Q for any k_1, k_2 .

It is natural to take one of k_1, k_2 as small as possible, say $k_2 = 1$. This gives the estimator

$$\begin{aligned} \text{ZMA}(k) &= \frac{1}{k-1} \left(k R(k) - R(1) \right) \\ &= \frac{1}{k-1} \left((\mathcal{Y}_k - \mathcal{Y}_0)^2 + (\mathcal{Y}_{k+1} - \mathcal{Y}_1)^2 + (\mathcal{Y}_{k+2} - \mathcal{Y}_2)^2 + \dots \right. \\ &\quad \left. - (\mathcal{Y}_1 - \mathcal{Y}_0)^2 - (\mathcal{Y}_2 - \mathcal{Y}_1)^2 - (\mathcal{Y}_3 - \mathcal{Y}_2)^2 - \dots \right). \end{aligned}$$

Based on the relative values of the process volatility σ and the noise variance η^2 , Zhang et al. [2005] calculate an optimal value for k to minimize the variance of $\text{ZMA}(k)$. In practice, it usually works equally well to take an *ad hoc* value determined by experiment.

Direct correction Zhou [1996] in effect proposes the following approach, based on estimating the microstructure noise and subtracting it: It is quite easy to see that

$$\sum_{j=1}^{N-1} (\mathcal{Y}_j - \mathcal{Y}_{j-1})(\mathcal{Y}_{j+1} - \mathcal{Y}_j) \approx -N\eta^2$$

(substitute $\mathcal{Y}_j = x_j + \epsilon_j$, take expectation, and use independence). In this calculation, unlike the ones above, the independence of the $\{\epsilon_j\}$ is crucial: serial correlation introduces significant corrections.

Therefore, we can remove the unwanted bias term $2N\eta^2$ by adding twice the above expression to the realized variance. This gives

$$Z = \sum_{j=1}^{N-1} \left((\mathcal{Y}_j - \mathcal{Y}_{j-1})^2 + 2(\mathcal{Y}_j - \mathcal{Y}_{j-1})(\mathcal{Y}_{j+1} - \mathcal{Y}_j) \right),$$

an unbiased estimator of Q . We emphasize that this is accurate only if the noise terms ϵ_j have zero serial correlation.

As above, one can evaluate on a grid with step k with offset h :

$$Z(k, h) = \sum_{j=1}^{m(k, h)} \left[(\mathcal{Y}_{h+jk} - \mathcal{Y}_{h+(j-1)k})^2 + 2(\mathcal{Y}_{h+jk} - \mathcal{Y}_{h+(j-1)k})(\mathcal{Y}_{h+(j+1)k} - \mathcal{Y}_{h+jk}) \right].$$

Note that the correction term is also evaluated using steps of k rather than 1. This makes the assumption of uncorrelation much more realistic: k need only be a few times the correlation length, usually very short. Zhou [1996] calculates the variance of this estimator and proposes optimal values for the grid size k .

And of course for any particular k one can again average over h :

$$\begin{aligned}
 Z(k) &= \frac{1}{k} \sum_{h=0}^k Z(k, h) \\
 &= \frac{1}{k} \left[(y_k - y_0)^2 + (y_{k+1} - y_1)^2 + \cdots + (y_N - y_{N-k})^2 \right] \\
 &\quad + \frac{2}{k} \left[(y_k - y_0)(y_{2k} - y_k) + (y_{k+1} - y_1)(y_{2k+1} - y_{k+1}) + \right. \\
 &\quad \left. \cdots + (y_{N-k} - y_{N-2k})(y_N - y_{N-k}) \right]
 \end{aligned}$$

(taking some liberties with the handling of the endpoints). The only remaining variable is the grid step size k . One can either use a theoretically optimal value or (better) test the estimator empirically to find reasonable values.

Empirical tests

Of course the real test is to try these estimators on real price data. The problem is how to tell whether the computed volatility value is correct or not, since there is no “official” value; also the underlying process is not strictly a Brownian motion so its volatility may not be well defined. Here are two references:

- Gatheral and Oomen [2010] implement the “zero-intelligence market” model of Smith et al. [2003]. This model has somewhat realistic microstructure dynamics on short time scales, but acts as a pure random walk on larger time scales. They test a variety of simple and sophisticated volatility estimators and make several specific recommendations.
- Shang [2018] tests a number of estimators against real market data, using the principle that observed price changes over finite intervals, divided by the square root of the realized variance on each interval, should form a unit normal population. He concludes that the best combination of accuracy, robustness, and simplicity is to sample the quote midpoint price at each time of a trade whose price is different than the preceding trade.

As mentioned above, if quote data is available the best price values to use are the bid-offer midpoint, evaluated at the times of the trades. This eliminates bid-ask bounce, and it preserves the important role of trade arrival rate, which would be lost if one evaluated quote values on an arbitrary time grid. Quote data might not be available either because it is too large (an order of magnitude larger than trade data for US equities) or because the exchange does publish it (Hasbrouck [2004] has an example from the futures pits).

One approach is to compute the estimated volatility across a range of parameters, most importantly the timestep lag k . Since this parameter is arbitrary, the computed value is meaningful if and only if it is constant across some range of parameter values.

Another approach is to find two different versions of the same asset, for example the S&P500 futures contract trading on CME, and compare it to the SPY ETF. Since these are essentially the same thing with different microstructure noise, a good estimator should give the same value of the volatility (Hasbrouck [2003] has a detailed discussion of the relationship). It is surprisingly difficult to find an estimator that gives good results. Another possible source of related assets would be a dual-listed Canadian stock, for example, RIM on TSX vs RIMM on Nasdaq (one must account for the intraday foreign exchange fluctuations).

References

- Federico M. Bandi and Jeffrey R. Russell. Separating microstructure noise from volatility. *J. Financial Econ.*, 79(3):655–692, 2006.
- Mark B. Garman and Michael J. Klass. On the estimation of security price volatilities from historical data. *J. Business*, 53(1):67–78, January 1980.
- Jim Gatheral and Roel C. A. Oomen. Zero-intelligence realized variance estimation. *Finance Stoch.*, 14(2):249–283, 2010.
- Joel Hasbrouck. Intraday price formation in U.S. equity index markets. *J. Finance*, 58(6): 2375–2399, December 2003.
- Joel Hasbrouck. Liquidity in the futures pits: Inferring market dynamics from incomplete data. *J. Fin. Quant. Anal.*, 39(2):305–326, June 2004.
- Linwei Shang. Intraday realized variance. Technical report, Quantitative Brokers, April 2018.
- Eric Smith, J. Doyne Farmer, László Gillemot, and Supriya Krishnamurthy. Statistical theory of the continuous double auction. *Quant. Finance*, 3:481–514, 2003.
- Lan Zhang, Per A. Mykland, and Yacine Aït-Sahalia. A tale of two time scales: Determining integrated volatility with noisy high-frequency data. *J. Amer. Statist. Assoc.*, 100(472): 1394–1411, 2005.
- Bin Zhou. High-frequency data and volatility in foreign-exchange rates. *J. Bus. Econ. Statist.*, 14(1):45–52, 1996.