Intraday Volatility Part 1
& Back to Trade and Quote Data

Merrell Hora

Courant Institute of Mathematical Sciences

Market Microstructure
Spring 2022

Intraday Volatility - Applications

The trading cost of executing a large order of size Q is $TC = \pm\{B/S\}(p_{ave} - p_0)/p_0$ where

- p_0 is the mid-quote just prior to the start of execution, and
- p_{ave} is the notional weighted average execution price
- $p_{ave} = \sum_{i=1}^{N} p_i q_i / Q$, where $Q = \sum_{i=1}^{N} q_i$
- TC is also referred to as an order's slippage to arrival.

Forecasts of TC are used widely in automated quoting (market making) and pricing trades, e.g. block trades. Primary factors in TC models include

- Bid-ask spread
- Q as a percentage of volume
- Volatility over the horizon of the trade

Intraday Volatility - Applications

A main driver of execution algorithm performance is spread capture (as TC depends on spread.) Spread capture occurs when limit orders are filled passively without improving the best bid or offer.

Factors driving spread capture

- The ability to obtain order book queue priority
- Anticipating short-term volatility

Given a time horizon T, volatility bands, $\sigma\sqrt{T}$, can help to optimize limit price selection.

[Yes, a graphic would be nice here.]

Intraday Volatility - Theoretical Background

Main reference: Hansen and Lunde (2006) (H-L)

Assume the efficient and unobservable log price process p_t^* follows a continuous semi-martingale

$$p_t^* = p_0^* + \int_0^t \mu_\tau d\tau + \int_0^t \sigma_\tau dW\tau \qquad (1)$$

where W is a standard Wiener process, μ_τ and σ_τ are well behaved (càdlàg, etc) drift and volatility processes.

Observable prices are given by p_t, so the noise process is

$$u_t \equiv p_t - p_t^*$$

u_t is hypothesized to arise from actual market frictions such as price discreteness and the so-called bid-ask bounce. Thus, u_t is referred to as *microstructure noise*.

Intraday Volatility - Theoretical Background

We are interested in estimators of the integrated variance (IV)

$$IV(a,b) \equiv \int_a^b \sigma_\tau^2 d\tau$$

Continuity of p_t^* implies that IV is equal to the quadratic variation (QV) of p^*, so we can turn to estimators of QV.

QV is the stochastic limit of the following sequence defined on the partition of the interval $[a,b]$ into m subintervals

$$a = t_{0,m} < t_{1,m} < \cdots < t_{m,m} = b.$$

The quadratic variation of p_t^* on the partition with m subintervals:

$$QV_m^*(a,b) \equiv \sum_{i=1}^{m} (p_{t_i}^* - p_{t_{i-1}}^*)^2$$

Theoretical Background

Of course, p^* is not observable (and neither is σ), so we turn to the observable p_t, which leads to the realized variance

$$RV_m(a,b) \equiv \sum_{i=1}^{m} (p_{t_i} - p_{t_{i-1}})^2$$

which introduces bias and statistical inconsistency ($RV_m(a,b)$ does not converge in probability to $IV(a,b)$ as $m \to \infty$.).

To analyze the bias and convergence properties of RV_m, we define the following

$$y^*_{i,m} \equiv p^*(t_{i,m}) - p^*(t_{i-1,m})$$
$$y_{i,m} \equiv p(t_{i,m}) - p(t_{i-1,m})$$
$$e_{i,m} \equiv u(t_{i,m}) - u(t_{i-1,m})$$

Sampling Schemes

Starting with the complete record of trades and quotes, we have to choose how to reduce, or *sample*, these to a single price series $\{p\}$.

Given observed (bid, ask, mid, or trade) prices at times $t_0 < \cdots < t_N$, there are 2 standard methods of calendar time sampling (CTS) to construct artificial prices at time $\tau \in [t_j, t_{j+1})$

- $p(\tau) \equiv p_{t_j}$ (previous tick)
- $\tilde{p}(\tau) \equiv p_{t_j} + \frac{\tau - t_j}{t_{j+1} - t_j}(p_{t_{j+1}} - p_{t_j})$ (linear interpolation.)

Tick time sampling (TTS) refers to the selection of prices as a function of tick times (essentially index value when sorted by event time.). Usually TTS selects trades at regular index increments, such as every *kth* price.

Sampling Schemes

Linear interpolation has the property that for a fixed sample of size N, $RV_m(a, b) \xrightarrow{p} 0$ as $m \to \infty$.

That is, RV_m computed from linearly interpolated prices converges in probability to 0 as the sampling frequency goes to infinity.

Hence, linearly interpolated prices are not appropriate for purposes of estimating high-frequency volatility (yet they are still used in places.)

The Bias of Realized Variance

Assumption 1: The noise process u is covariance stationary with mean 0 and autocovariance function given by $\pi(s) \equiv E[u(t)u(t+s)]$.

Theorem 1 (H-L) Given (1) and assumption 1

$$E[RV(m) - IV] = 2\rho_m + 2m\left[\pi(0) + \pi\left(\frac{b-a}{m}\right)\right], \qquad (2)$$

where $\rho_m \equiv E(\sum_{i=1}^{m} y^*_{i,m} e_{i,m})$.

This follows from

$$RV_m = \sum_{i=1}^{m} y^*_{i,j} + 2\sum_{i=1}^{m} e_{i,m} y^*_{i,m} + \sum_{i=1}^{m} e^2_{i,m}$$

Thus, the bias is a function of the correlation between the efficient returns and noise, and autocovariances of the noise.

Volatility Signature Plots

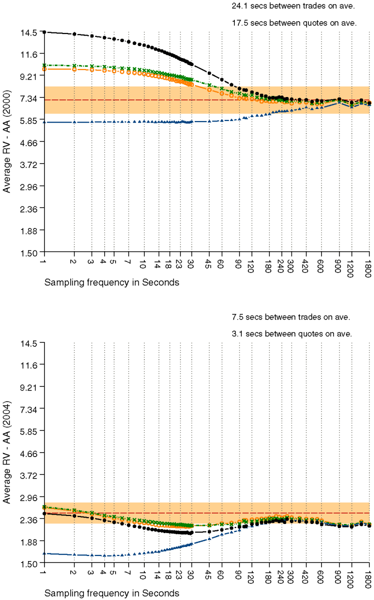
A volatility signature plot displays sample averages of estimates of *RV* as a function of the sampling frequency *m*

$$\overline{RV}_m \equiv n^{-1} \sum_{t=1}^{n} RV_{t,m}$$

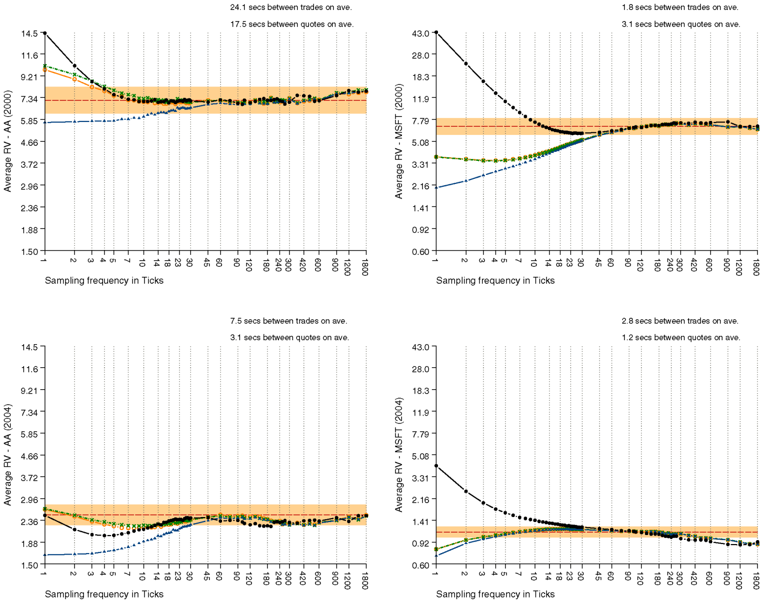
H-L provide numerous volatility signature plots for Alcoa (AA) and Microsoft (MSFT) using different *RV* estimators, sampling methods, and samples of bid, ask, mid and trade prices from 2000 and 2004.

The following slides show their baseline results using the naive estimator compared with a robust estimator (horizontal red line.)

Volatility Signature Plots - Time Sampling



Volatility Signature Plots - Tick Sampling



Observations

- Low to middle sampling frequencies appear approximately unbiased.
- Bias increases with sampling frequency and is pronounced at ultra-high levels.
- Mid-quotes (blue diamond) behave differently, with $\overline{RV}$ decreasing as sampling frequency increases.
- Microstructure noise is negatively correlated with efficient returns.

The last point follows because the second term in the bias expression (2) is always nonnegative. H-L note that quote revisions are asynchronous, and that upward revisions in the efficient price may affect the ask price first, causing the spread to widen by only half a tick.

The Realized Kernel Estimator

Barndorff, Nielsen, Hansen, Lunde, and Shepard (2008) propose the following estimator that exhibits a number of improvements (see Hautsch 8.1.2.)

$$RV_K \equiv \hat{\gamma}_0 + \sum_{h=1}^{H} k\left(\frac{h-1}{H}\right)(\hat{\gamma}_h + \hat{\gamma}_{-h}) \tag{3}$$

where $\hat{\gamma}_h \equiv \sum_{i=1}^{N} y_i y_{i+h}$, the *hth* sample autocovariance and $k(\cdot)$ is a kernel function. The Tukey-Hanning$_2$ kernel is recommended

$$k(x) = \sin^2(\pi/2(1-x)^2)$$

with bandwidth $H = 5.74\zeta\sqrt{n}$ and ζ being rather tedious to estimate, but a good exercise in sampling and computing *RVs* (next homework.)

References

Hautsch Ch.8

Hansen, Peter and Asger Lunde, "Realized Variance and Market Microstructure Noise," *Journal of Business & Economic Statistics*, 2006, Vol. 24, No. 2, p. 127-161.

Barndorff-Nielsen, O. E., P. Hansen, Al Lunde and N. Shephard, "Realized kernels in practice: trades and quotes," *The Econometrics Journal*, 2009, Vol. 12, No. 3, pp. C1-C32.