Volatility Estimation with High-Frequency Data: Three Approaches and Three Horizons

E.J. Hannan Lecture
Canberra-July 2009

Eric RENAULT
(joint with Per A. MYKLAND and Lan ZHANG)

INTRODUCTION:

Benefits and limits of High-Frequency data

- Goal = estimation of continuous time models of asset prices popular in mathematical finance
- This talk: no jumps $\Rightarrow$ price processes consistent with no-arbitrage are Brownian semi-martingales
- Univariate setting for log-price process X with drift μ and volatility σ (defined from Brownian Motion W) :

$$dX_t = \mu_t dt + \sigma_t dW_t$$

OUTLINE

1. A general and non-technical overview: Three main issues
2. Technical developments : Three approaches and three horizons

Three main issues:

1. Higher frequency of sampling the log-price process $\rightarrow$ useless to estimate its drift $\rightarrow$ focus on volatility.
2. Market frictions are non-negligible $\rightarrow$ the "efficient" semi-martingale is not directly observed.
3. Non-parametric estimation of instantaneous (spot) volatility = more involved than integrated volatility estimation over a day.

Issue 1: Higher frequency of sampling = useless to estimate the drift

- Simplest possible example:

Geometric Brownian Motion (GBM)

$$dX_t = \mu dt + \sigma dW_t$$

$$\text{Equi-spaced obervations } X_{t_i}, \text{i} = 0,1,\ldots\text{n, over } [0,T]:$$

$$t_i = i\Delta t, i = 0,1,\ldots n, \Delta t = \frac{T}{n}, t_0 = 0, t_n = T.$$

$\Rightarrow$ Parametric model with unknown parameters θ =(μ,σ) for (continuously compounded) rates of returns:

$$X_{t_i} - X_{t_{i-1}}, i = 1,\ldots, n.$$

$$\Delta X_{t_{i+1}} = X_{t_{i+1}} - X_{t_i}, i = 0,1,\ldots,n-1, i.i.d. \approx N(\mu\Delta t, \sigma^2 \Delta t)$$

Maximum Likelihood Estimators (MLE):

$$\hat{\mu}_n = \frac{1}{n\Delta t}\sum_{i=0}^{n-1}\Delta X_{t_{i+1}} = \frac{X_T - X_0}{T} = \text{does not depend on n!!!}$$

$$\hat{\sigma}^2_{\text{n,MLE}} = \frac{1}{n\Delta t}\sum_{i=0}^{n-1}(\Delta X_{t_{i+1}} - \hat{\mu}_n \Delta t)^2$$

Cost: No way to improve the estimation of the drift μ by increasing the number n of intraday observations → critical empirical issue of the estimation of expected returns : Merton (1980)

1st Benefit: No need to estimate the mean to estimate the volatility

- (Asymptotically) Efficient Estimator of σ (as if μ=0): $\hat{\sigma}_n^2 = \frac{1}{n\Delta t}\sum_{i=0}^{n-1}(\Delta X_{t_{i+1}})^2 = \text{efficient for } \sigma^2!$

Since MLE: $\hat{\sigma}_{n,MLE}^2 = \frac{1}{n\Delta t}\sum_{i=0}^{n-1}(\Delta X_{t_{i+1}} - \hat{\mu}_n \Delta t)^2 = \hat{\sigma}_n^2 - \hat{\mu}_n^2 \Delta t$

$$\Rightarrow \sqrt{n}(\hat{\sigma}_{n,MLE}^2 - \hat{\sigma}_n^2) = -\hat{\mu}_n^2 \frac{T}{\sqrt{n}} \xrightarrow{P} 0 \; (n \to \infty, T fixed).$$

Efficiency bound:

$$\sqrt{n}(\hat{\sigma}_n^2 - \sigma^2) \approx \sigma^2 \sqrt{2}\, \frac{\chi^2(n) - n}{\sqrt{2n}} \approx N(0, 2\sigma^4)$$

2nd Benefit: Control variable (as if μ=0)

- When we have i.i.d. observations of Z and we know that E(Z)=0, the (asymptotically)efficient estimator of Var(Z) is not in general the simple sample counterpart.
- Asymptotically efficient estimator of Var(Z):

$$\frac{1}{n}\sum_{i=1}^{n} Z_i^2 - \hat{b}_n \frac{1}{n}\sum_{i=1}^{n} Z_i$$

$$\text{with } b = \frac{\text{cov}[Z^2, Z]}{Var(Z)} = \frac{E(Z^3)}{Var(Z)}$$

$\Rightarrow$ Room for improvement (by control variable Z) if and only if: $E(Z^3) \neq 0$

$$\Delta X_{t_{i+1}} = \int_{t_i}^{t_{i+1}} \mu_s ds + \int_{t_i}^{t_{i+1}} \sigma_s dW_s$$

Control variable (as if $\mu_s \equiv 0$)

$\Rightarrow$ Provides an estimator of volatility better than $\sum_{i=0}^{n-1} (\Delta X_{t_{i+1}})^2$

if and only if $E\left[(\Delta X_{t_{i+1}})^3\right] \neq 0$ that is : $E\left[(W_{t_{i+1}} - W_{t_i})^3\right] \neq 0$

$\rightarrow$ May be the case when sampling times are endogenous, that is correlated with prices

$\rightarrow$ Economic motivation: random transaction dates or dates of quote changes = informative about arrival of news ~ correlated to shocks in prices

2^{nd} Issue: Market frictions are non-negligible $\rightarrow$ the "efficient" semi-martingale is not directly observed.

- Back to the simplest example of GBM
- Now we only observe Y, that is efficient price X plus microstructure noise ε $\rightarrow$ parametric model with unknown parameters $\theta = (\sigma, a)$ (immaterial to assume $\mu=0$):

$$Y_{t_i} = X_{t_i} + \varepsilon_{t_i}, i = 0,1,..,n.$$

$$dX_t = \sigma dW_t, (\varepsilon_t) \text{ i.i.d.} \approx N(0, a^2),$$

$$(\varepsilon_t) \text{ and } (X_t) \text{ independent.}$$

$$\Delta Y_{t_{i+1}} = \Delta X_{t_{i+1}} + \Delta \varepsilon_{t_{i+1}}$$

$$\Rightarrow \hat{\sigma}_n^2 \Delta t = \frac{1}{n}\sum_{i=0}^{n-1} (\Delta Y_{t_{i+1}})^2 = \frac{1}{n}\sum_{i=0}^{n-1} (\Delta X_{t_{i+1}})^2 + \frac{1}{n}\sum_{i=0}^{n-1} (\Delta \varepsilon_{t_{i+1}})^2 + \frac{2}{n}\sum_{i=0}^{n-1} (\Delta X_{t_{i+1}})(\Delta \varepsilon_{t_{i+1}})$$

$$\Rightarrow \hat{\sigma}_n^2 \Delta t = 2a^2 + O_P\left(\frac{1}{\sqrt{n}}\right) \text{since } E\left[(\Delta X_{t_{i+1}})^2\right] = \sigma^2 \Delta t = O(1/n)$$

Conclusion: Root-n consistent estimator for a (measure of market lack of efficiency)

BUT no such thing for efficient volatility σ

→ Efficiency bound (MLE) for σ:

$$n^{1/4}(\hat{\sigma}_{n,MLE}^2 - \sigma^2) \xrightarrow{d} N(0, \frac{8\sigma^3 a}{\sqrt{T}})$$

Rate of convergence divided by two except if a=0 (no noise) or T→∞ (long time span)

Intuition: $\Delta t \to 0 \Rightarrow \Delta Y = \Delta X + \Delta \varepsilon$

$$\text{with } \mathrm{Var}(\Delta \mathrm{X}) << \mathrm{Var}(\Delta \varepsilon)$$

Far from being negligible, the noise is indeed dominant

$\Rightarrow$ need to "sacrifice" some observations to get rid of the noise

$\Rightarrow$ The effective number of observations for volatility estimation = only square-root n

1st Consequence:

When time-varying volatility $\rightarrow$ For estimation of DAILY VOLATILITY:

Even more difficult (albeit possible) to keep the efficient rate of convergence : $n^{1/4}$

$\rightarrow$ 3 (rate) efficient approaches (among others):

(i) Flat-top realized kernels

(ii) Multi-scale realized volatility

(iii) Realized volatility after pre-averaging

$\rightarrow$ And other rate-suboptimal (but simpler) approaches:

(i) Non flat-top kernels

(ii) Two-scale realized volatility

2nd Consequence:

- Since, for the purpose of estimation of daily volatility, we may have rates of convergence as slow as (optimal) n exponent 1/4
- Or (suboptimal) n exponent 1/5, 1/6,…
- Relevant to take advantage of observations on previous days
- High volatility persistence $\Rightarrow$ volatility measurements on previous days allow to improve current daily volatility measurement
- Multi-days horizon

3rd Issue: Instantaneous volatility

- With constant volatility σ (GBM), equivalent to estimate spot volatility σ or "volatility" over a time interval [0,t] (like a day), i.e. more properly integrated variance:

$$\int_0^t \sigma_s^2 ds = t\sigma^2 \text{ for GBM}$$

- With time-varying volatility, the above rates of convergence : root-n without noise, smaller (sub) optimal rates with noise can be maintained only for integrated variance.
- Rate of convergence still to be divided by an additional factor two when it goes to spot volatility $\rightarrow$ like estimating a density function instead of a cumulative distribution function.

OUTLINE (for technical developments)

1. Estimation of daily volatility without microstructure noise, but possibly endogenous sampling times:

"Realized volatility when sampling times can be endogenous" (2008)

By Yingying Li, Per Mykland, Eric Renault, Lan Zhang and Xinghua Zheng

2. The three efficient approaches for estimation of daily volatility with noisy observations:

1st approach:

"Efficient estimation of stochastic volatility using noisy observations: a multi-scale approach" in Bernoulli (2006)

By Lan Zhang.

2nd approach:

"Microstructure noise in the continuous case: The pre-averaging approach" in Stochastic processes and their applications (2009)

By Jean Jacod, Yingying Li, Per Mykland, Mark Podolskij and Mathias Vetter.

3rd approach:

"Designing realized kernels to measure ex-post variation of equity prices in the presence of noise"

in Econometrica (2009)

By Ole Barndorff-Nielsen, Peter Hansen, Asgar Lunde and Neil Shephard

3. The multi-day horizon:

"In-sample Asymptotics and Across-sample Efficiency Gains for High Frequency Data Statistics" (2009)

By Eric Ghysels, Per Mykland and Eric Renault

4.The instantaneous horizon:

"Aggregated and Instantaneous Volatility: Connections and Comparisons" (2009)
By Per Mykland, Eric Renault and Lan Zhang.

See also:

(i) "Nonparametric Filtering of the Realized Spot Volatility: A Kernel-based approach" (2009) in Econometric Theory
By Dennis Kristensen.

(ii) "Nonparametric stochastic volatility" (2008)
By Federico Bandi and Roberto Reno

1. Estimation of daily volatility without microstructure noise, but possibly endogenous sampling times:

- (Immaterial) local martingale assumption:

$$X_t = X_0 + \int_0^t \sigma_s dW_s$$

- Quadratic variation relative to a grid G:

$$G = \{0 = t_0 < t_1 < \ldots < t_n = T\},$$

$$Max(t_i - t_{i-1}) \xrightarrow{P}_{n=\infty} 0$$

$$\Rightarrow [X, X]_t^G = \sum_{t_{i+1} \leq t} (\Delta X_{t_{i+1}})^2 \xrightarrow{P}_{n=\infty} \int_0^t \sigma_s^2 ds$$

Estimation Error

$$[X, X]_t^G - \int_0^t \sigma_s^2 ds$$

$$\approx M_t = \sum_{t_{i+1} \le t} (X_{t_{i+1}} - X_{t_i})^2 + (X_t - X_{t^*})^2 - \int_0^t \sigma_s^2 ds$$

= local martingale:

$$M_t = 2 \sum_{t_{i+1} \le t} \int_{t_i}^{t_{i+1}} (X_s - X_{t_i}) dX_s + 2 \int_{t^*}^{t} (X_s - X_{t^*}) dX_s$$

since by Ito's lemma:

$$d(X_t - X_{t_i})^2 = 2(X_t - X_{t_i}) dX_t + \sigma_t^2 dt$$

Key properties of estimation error M:

- Define Quarticity: $[X,X,X,X]_t^G = \sum_{t_{i+1}\le t}(X_{t_{i+1}} - X_{t_i})^4 + (X_t - X_{t^*})^4$
- And Tricity: $[X,X,X]_t^G = \sum_{t_{i+1}\le t}(X_{t_{i+1}} - X_{t_i})^3 + (X_t - X_{t^*})^3$
- If

$$Max|t_{i+1} - t_i| = o_P(n^{-\frac{2}{3}+\varepsilon}) \text{ for some } \varepsilon > 0:$$

$$P\lim_{n=\infty}[\sqrt{n}M, \sqrt{n}M]_t = P\lim_{n=\infty}\frac{2}{3}n[X,X,X,X]_t^G = \frac{2}{3}\int_0^t u_s ds$$

$$P\lim_{n=\infty}[\sqrt{n}M, X]_t = P\lim_{n=\infty}\frac{2}{3}\sqrt{n}[X,X,X]_t^G = \frac{2}{3}\int_0^t v_s ds$$

(Asymptotic) control variable:

- Orthogonal decomposition of the estimation error: $\sqrt{n}M_t = \tilde{M}_t^{(n)} + \int_0^t g_s dX_s$ with $P\lim[\tilde{M}^{(n)}, X]_t = 0$

$$\Leftrightarrow \int_0^t g_s \sigma_s^2 ds = P\lim_{n=\infty} [\sqrt{n}M, X]_t = \frac{2}{3}\int_0^t v_s ds$$

$$\Leftrightarrow g_t = \frac{2v_t}{3\sigma_t^2} \rightarrow \frac{v_t}{\sigma_t^2} \cong \frac{E(Z^3)}{E(Z^2)} = \frac{Cov(Z^2, Z)}{Var(Z)} \text{ (control variable)}$$

- Consequence:

$$\sqrt{n}M_t \xrightarrow{d} \frac{2}{3}\int_0^t \frac{v_s}{\sigma_s^2} dX_s + \int_0^t \sqrt{\frac{2}{3}u_s - \frac{4}{9}\frac{v_s^2}{\sigma_s^2}} dB_s$$

Where B is a Brownian motion independent of the underlying σ-field

CONCLUSIONS:

$$\sqrt{n}\left\{[X,X]_t^G - \int_0^t \sigma_s^2 ds\right\} \xrightarrow{d} \underbrace{\frac{2}{3}\int_0^t \frac{v_s}{\sigma_s^2} dX_s}_{\text{asymptotic bias}} + \left[\int_0^t \left(\frac{2}{3}u_s - \frac{4}{9}\frac{v_s^2}{\sigma_s^2}\right) ds\right]^{1/2} N(0,1)$$

- The total asymptotic MSE is always:

$$\frac{2}{3}\int_0^t u_s ds = P\lim_{n=\infty} \frac{2}{3} n[X,X,X,X]_t^G$$

$\rightarrow$ The asymptotic bias can be estimated to get a new estimator with asymptotic variance reduced by control variable:

$$[X,X]_t^G - \frac{2}{3\sqrt{n}} \overbrace{\int_0^t \frac{v_s}{\sigma_s^2} dX_s}^{estimated}$$

Tricity and random sampling times:

- Control variable = reduces the variance

$$\Leftrightarrow 0 \neq P\lim_{n=\infty}[\sqrt{n}M, X]_t = P\lim_{n=\infty} \frac{2}{3}\sqrt{n}\sum_{t_{i+1}\leq t}(X_{t_{i+1}} - X_{t_i})^3$$

This would be zero if, given the sampling times:

$$X_{t_{i+1}} - X_{t_i} \approx N[0, \sigma_{t_i}^2(t_{i+1} - t_i)]$$

Non-zero takes sampling times endogenous because correlated to the price process $\rightarrow$ example of times generated by price hitting a barrier (more generally structural ACD of *Renault, Van der Heijden, Werker* (2009), correlated Brownian motion hitting a random barrier)

Example of times generated by price hitting a barrier

- Times are defined recursively by:

$$t_0 = 0, t_{i+1} = Min\{t \geq t_i, \sqrt{n}(X_t - X_{t_i}) \in \{a, -b\}\}, a, b > 0$$

$$\Leftrightarrow X_{t_{i+1}} - X_{t_i} = n^{-1/2} Z_i, \text{ with } Z_1, Z_2, \ldots i.i.d. \in \{a, -b\}.$$

- Assume GBM with μ=0 and σ=1.
- Then: P[Z=a] = b/(a+b) and:

$$E(Z) = 0 \text{ but } E(Z^3) \neq 0 \text{ if } a \neq b$$

$$v_t \equiv \frac{E(Z^3)}{\sqrt{(ab)^3}}.$$

Quarticity and irregular sampling times

- Heuristically, in the exogenous time case:

$$n\sum_{t_{i+1}\le t}(X_{t_{i+1}}-X_{t_i})^4 \approx n\sum_{t_{i+1}\le t}\sigma_{t_i}^4(t_{i+1}-t_i)^2[N(0,1)]^4$$

$$\Rightarrow P\lim_{n=\infty}\frac{2}{3}n[X,X,X,X]_t^G = 2T\int_0^t \sigma_s^4 H'(s)ds$$

$$\text{with } H(t) = P\lim_{n=\infty}\frac{n}{T}\sum_{t_{i+1}\le t}(t_{i+1}-t_i)^2$$

H(t) = Asymptotic Quadratic Variation in Time (AQVT) (Mykland and Zhang (2006))

$$\Delta t = \frac{T}{n} \Rightarrow H(t) = Lim_{n=\infty}\frac{n}{T}\times\frac{t}{\Delta t}\times\left(\frac{T}{n}\right)^2 = t$$

More generally, AQVT H(t) is non-decreasing and runs quickly when the sampling is slower than normal

$\Rightarrow$ With exogenous sampling times:

$$\sqrt{\frac{n}{T}}\left\{[X,X]_t^G - \int_0^t \sigma_s^2 ds\right\} \xrightarrow{d} N\left(0, 2\int_0^t \sigma_s^4 H'(s) ds\right)$$

$\Rightarrow$ More weight on long time intervals

Conclusion: 3 ways to build a confidence interval:

(i) With AQVT

(ii)With quarticity [X,X,X,X] (If random times)

(iii)Taking control variable into account

2. The three efficient approaches for estimation of daily volatility with noisy observations

- 2.1. Three simple intuitions:
- 1st intuition: With HF data, the noise will swamp the true QV:

$$Y_t = X_t + \varepsilon_t, dX_t = (\mu_t dt) + \sigma_t dW_t,$$

$$\varepsilon \text{ and X independent}, \varepsilon \text{ i.i.d.}, E(\varepsilon) = 0,$$

$$\frac{1}{n}\sum(Y_{t_{i+1}} - Y_{t_i})^2 = \frac{1}{n}\sum(X_{t_{i+1}} - X_{t_i})^2 + \frac{1}{n}\sum(\varepsilon_{t_{i+1}} - \varepsilon_{t_i})^2 + \frac{2}{n}\sum(X_{t_{i+1}} - X_{t_i})(\varepsilon_{t_{i+1}} - \varepsilon_{t_i})$$

$$\frac{[Y,Y]_t}{n} = \frac{[X,X]_t}{n} + 2Var(\varepsilon) + O_P(1/\sqrt{n}) \text{ ("Bias" in QV(Y)} = 2nVar(\varepsilon))$$

(1/2n)QV(Y) = root-n consistent estimator of the variance of the noise($\rightarrow$bias correction)

2^{nd} intuition: Sampling sparsely may be an answer BUT you should also:

1. Averaging across the low-frequency grids in order to use all available data
2. Bias-correcting the averaged estimator

→ *"A Tale of Two Time Scales: Determining Integrated Volatility with Noisy High-Frequency Data"*, 2005, JASA

By Lan Zhang, Per A. Mykland and Yacine Ait-Sahalia

1. Averaging across the low-frequency grids

$Grid\ \mathrm{G}: t_i, i = 0, ,\dots n,$

Low frequency subgrids (frequency / K) :

$t_1, t_{K+1}, \dots$

$t_2, t_{K+2}, \dots$

....

$t_K, t_{2K}, \dots\dots$

Averaging across these subgrids :

$$[Y, Y]_t^{(n,K)} = \frac{1}{K} \sum_{t_i \le t, i \ge K} (Y_{t_i} - Y_{t_{i-K}})^2 \rightarrow Bias = 2\frac{n-K+1}{K} Var(\varepsilon)$$

2. Bias-correcting the averaged estimator

(i)Bias correction for the noise:

$$[Y,Y]_t^{(n,K)} - \frac{n-K+1}{nK}[Y,Y]_t^{(n,1)}$$

(ii) Making the averaged estimator consistent

while effective number of observations in a low-frequency grid is (n/K)→ Rate of convergence:

$$(n/K)^{1/2} \Rightarrow \frac{K}{n} \rightarrow_{n=\infty} 0$$

$$\text{but also}: K \rightarrow_{n=\infty} \infty \text{ (for variance of the noise)}$$

(iii)Finite sample bias-correction:

$$[X,X]_t^{TSRV} = \left(1 - \frac{n-K+1}{nK}\right)^{-1} \left\{[Y,Y]_t^{(n,K)} - \frac{n-K+1}{nK}[Y,Y]_t^{(n,1)}\right\}$$

3rd intuition: The Two Time Scales approach yields a suboptimal rate of convergence

- Effective number of observations: $\bar{n} = \frac{n}{K}$
- We have to balance:

(i) Rate of convergence of the no-noise part: $\sqrt{\bar{n}}$

(ii) Rate of convergence of the noise part: $\sqrt{\frac{K}{\bar{n}}}$

- Optimal balance: $K = \bar{n}^2 = \left(\frac{n}{K}\right)^2 \Rightarrow K = cn^{2/3}$
- Best possible rate of convergence:

$$\sqrt{\frac{n}{K}} = n^{1/6} << n^{1/4} = \text{"efficient" rate}$$

2.2. The three efficient approaches:

- 1st efficient approach: Multi Time Scales:

$\Rightarrow$ Generalizing the TSRV by considering M time scales:

$$[X,X]_t^{MSRV} = \sum_{j=1}^{M} \alpha_j [Y,Y]_t^{(n,K_j)}$$

$$\sum_{j=1}^{M} \alpha_j \frac{n+1-K_j}{K_j} = 0 \text{ (removing bias due to noise)}$$

$$\sum_{j=1}^{M} \alpha_j = 1 \text{ (no bias in the QV(X) part)}$$

$$w.l.o.g.\ K_j = j = 1,2,\ldots,M$$

$$\Rightarrow \text{effective number of observations}: \bar{n} = \frac{n}{M}$$

Optimal balance:

- Rate of convergence of the no-noise part: $\sqrt{\bar{n}}$

- Rate of convergence of the noise part: $\dfrac{M}{\sqrt{\bar{n}}}$ instead of $\dfrac{\sqrt{K}}{\sqrt{\bar{n}}}$

$\Rightarrow$ Optimal balance: $M = \bar{n} = \dfrac{n}{M} \Rightarrow M = n^{1/2}$

$\Rightarrow$ Best possible rate of convergence:

$$\sqrt{\frac{n}{M}} = n^{1/4}$$ = efficient rate of convergence!

2^{nd} efficient approach: Pre-averaging

- Revisiting the multi-scale approach in the simplest case Δt=(T/n) $\quad Y_{t_i} \to Y_i, \Delta t \to \Delta_n$

$\Rightarrow$ Replacing the i-th observation by the local average:

$$\bar{Y}_i^n = \frac{1}{k_n} \sum_{j=k_n/2}^{k_n-1} [Y_{i+j} - Y_{i+k_n-j-1}] \Rightarrow \sum_{i=0}^{[T/\Delta_n]-k_n+1} (\bar{Y}_i^n)^2 \approx MSRV$$

$$\text{with scales}: 2j - k_n + 1,\ j = \frac{k_n}{2}, \ldots, k_n - 1,$$

$$\Rightarrow \text{Optimal } k_n = \frac{\theta}{\sqrt{\Delta t}} \propto \sqrt{n}.$$

Bias correction for the noise:

$$[X,X]_T^{preav} = \frac{12\sqrt{\Delta t}}{\theta}\left\{\sum_{i=0}^{[T/\Delta_n]-k_n+1}(\overline{Y}_i^{\,n})^2 - \frac{\sqrt{\Delta t}}{2\theta}\sum_{i=1}^{[T/\Delta_n]}(Y_i - Y_{i-1})^2\right\}$$

R1. Second term = immaterial for asymptotic distribution (once noise has been removed).

R2. Equivalence between:

Multi-scale RV $\Leftrightarrow$ Averaging QVs computed on $\neq$ scales

Pre-averaging $\Leftrightarrow$ Averaging returns on $\neq$ horizons before computing QV.

Asymptotic distribution:

$$\left(\frac{n}{T}\right)^{1/4}\left\{[X,X]_T^{preav}-\int_0^T\sigma_s^2 ds\right\}\xrightarrow{d} N\left(0,\int_0^T\gamma_s^2 ds\right)$$

$$with, denoting\ a^2=Var(\varepsilon):$$

$$\gamma_t^2=4.12^2\left(\frac{151}{80640}\theta\sigma_t^4+\frac{1}{48}\frac{a^2\sigma_t^2}{\theta}+\frac{1}{6}\frac{a^4}{\theta^3}\right).$$

R1. For GBM, optimal $\theta \approx 4.777.(a/\sigma)$

$\Rightarrow$ Variance $\approx 8.545 a\sigma^3 \approx efficiency\ bound\ 8a\sigma^3$

R2.Time varying σ (and possibly a) $\Rightarrow$ optimal θ should be time varying and computed from estimates of instantaneous volatility (section 4)

3rd efficient approach: Realized kernel

- Pre-averaging $\Leftrightarrow$ Averaging returns on $\neq$ horizons before computing QV(Δt=T/n):

$$\sum_{i=0}^{[T/\Delta_n]-k_n+1} (\overline{Y}_i^{\,n})^2 = \sum_i \left[\frac{1}{k_n} \sum_j (Y_{i+j} - Y_{i+k_n-j-1}) \right]^2$$

$\rightarrow$ Belongs to the class of flat top kernels:

$$K(Y) = \gamma_0(Y) + \sum_{h=1}^{H} k\left(\frac{h-1}{H}\right)\left\{\gamma_h(Y) + \gamma_{-h}(Y)\right\}$$

$$\gamma_h(Y) = \sum_i (Y_i - Y_{i-1})(Y_{i-h} - Y_{i-h-1})$$

$$k(0) = 1, k(1) = 0, H \propto k_n \propto \sqrt{n}$$

R. Computing realized kernels $\Rightarrow$ Issue of border terms $\rightarrow$ Possible solution = "jittering"$\Leftrightarrow$ redefining the price measurement at the two endpoints to be an average of m distinct observations

No "edge effect" in the asymptotic variance

$\Leftrightarrow$m $\rightarrow \infty$

Otherwise: additional terms in the asymptotic variance which are attached at each end point (no integrals on the time interval corresponding to a local martingale)

Asymptotic distribution (with m→∞)

- 1st case: k'(0)≠0 or k'(1)≠0:

$$H = cn^{2/3} \Rightarrow n^{1/6}\left\{K(Y) - \int_0^T \sigma_s^2 ds\right\} \xrightarrow{d} N\left(0, g_1(c)T\int_0^T \sigma_s^4 ds\right)$$

- 2nd case: k'(0) = k'(1) = 0:

$$H = cn^{1/2} \Rightarrow n^{1/4}\left\{K(Y) - \int_0^T \sigma_s^2 ds\right\} \xrightarrow{d} N\left(0, g_2(c)T\int_0^T \sigma_s^4 ds\right)$$

R. Takes k'(0)=k'(1)=0 to get the efficient rate of convergence

Example 1: k(x) = 1-x:

$$K(Y)=\gamma_0(Y)+\sum_{h=1}^{H}\left(1-\frac{h-1}{H}\right)\left\{\gamma_h(Y)+\gamma_{-h}(Y)\right\}$$

$\rightarrow$ Bartlett kernel (~ Newey-West HAC estimator)

$\rightarrow$ Asymptotic distribution = TSRV (suboptimal rate)

$\rightarrow$ Intuition:

$$(X_{t_i}-X_{t_{i-K}})^2=\left(\sum_{j=i-K}^{i-1}\Delta X_{t_{i+1}}\right)^2=\sum_{j=i-K}^{i-1}(\Delta X_{t_{i+1}})^2+2\sum_{i-1\geq j>l\geq i-K}(\Delta X_{t_j})(\Delta X_{t_l})$$

$$\Rightarrow [X,X]_T^{(n,K)}-[X,X]_T^{(n,1)}=2\sum_{i=1}^{n-1}(\Delta X_{t_i})\sum_{j=1}^{K\wedge i}\left(1-\frac{j}{K}\right)(\Delta X_{t_{i-j}})+O_P(K/n)$$

$\rightarrow$ Edge effects encapsulated in O(K/n)

Example 2: $k(x) = 1 - 3x^2 + 2x^3$

$\rightarrow$ Asymptotic distribution = MSRV (efficient rate)

GENERAL CONCLUSION:

3 almost equivalent approaches to accommodate noise:

Averaging QVs

$\Leftrightarrow$ Pre-Averaging returns to compute QVs

$\Leftrightarrow$ Kernel-averaging auto-covariances.

3. The multi-day horizon:

3.1. Motivation:

- RV-type objects are interpreted as estimators for day t (time interval]t-1,t]) of population quantities: $$[X,X]_t, [X,X,X]_t, [X,X,X,X]_t$$
- Also bipower variation, leverage effect, etc.
- Consistent estimators for a number n of intraday observations going to infinity.
- However, due to market frictions, either not wise to use too high-frequency data or noise-correction procedures making smaller the effective sample size (always ~ averaging)

Volatility persistence

- In-sample RV objects on day t (in-fill asymptotics about n) can be complemented with observations on prior intervals. Hence benefit from across-sample observations (time series dynamics τ = t,t-1,…,t-H).
- Across-sample efficiency gains : persistence from one day to the other must be sufficiently high to be non-negligible in spite of infill consistency of estimators $\Rightarrow$ asymptotic view on persistence.

Local-to-unity asymptotics:

- Correlation today/yesterday arbitrarily close to 1 as n $\to \infty$.
- Standard local-to-unity asymptotics (Bobkoski (1983),Chan and Wei (1987), Phillips (1987)) is used to better approximate the finite sample behavior of estimators in AR models with roots near the unit circle: AR coefficients indistinguishable from 1 when time series length T$\to \infty$.
- For us: Correlation coefficient (no AR model) indistinguishable from 1 when effective number of intraday data goes to infinity (n $\to \infty$).

Conditional filters

- We consider path-dependent or conditional filters. Hence the weights are going to be time-varying.
- Tightly connected to Kalman filtering with heteroskedasticity but model-free.
- Filtering based on prediction errors, rather than linear combinations of past and present realized volatilities.

3.2. Improving the estimation using prior day information

- Focus on the example of estimation of the quadratic variation:
- $$QV_t = P\lim_{n=\infty} \sum_{t-1<t_i<t_{i+1}\le t} (X_{t_{i+1}} - X_{t_i})^2 (= \int_{t-1}^{t} \sigma_s^2 ds \text{ if no jumps})$$
- Conditional variance is highly persistent stationary process $\rightarrow$ high correlation coefficient:

$$\varphi_0 = \frac{Cov[QV_t, QV_{t-1}]}{Var(QV_t)}$$

- In general, we could take advantage of more than one lag.
- We have a daily estimator : $P\lim_{n=\infty} \hat{V}_t = QV_t$

Population characteristics of the time series of daily estimators:

$$\sigma^2 = E(QV_t) = E(\hat{V}_t)$$

$$\varphi = \frac{Cov[\hat{V}_t, \hat{V}_{t-1}]}{Var(\hat{V}_t)}$$

Estimation of σ and φ is from time series (daily) information: $\hat{V}_\tau, \tau = 1,..,T$

Estimation error immaterial if $\sqrt{T} >> n^\alpha$

Note that the notation α is used to accommodate ≠ volatility estimators

$$n^{\alpha/2} \frac{\left(\hat{V}_t - QV_t\right)}{\sqrt{2Q_t}} \xrightarrow{d} N(0,1) \text{ since :}$$

$$\varphi - \varphi_0 \approx \frac{2\varphi}{n^\alpha} \frac{E(Q_t)}{Var(QV_t)}$$

Time-series estimation error = immaterial

$\Rightarrow$ we can compute the best linear forecast of $\hat{V}_t$ given $\hat{V}_{t-1}$

$$\hat{V}_{t|t-1} = \varphi \hat{V}_{t-1} + (1-\varphi)\sigma^2$$

Goal = to combine the two measurements for day t (current + forecast from yesterday) to define a new estimator for day t:

$$\tilde{V}_t(\omega_t) = (1-\omega_t)\hat{V}_t + \omega_t \hat{V}_{t|t-1}$$

The more persistent the volatility process, the more informative the forecast from yesterday is about current volatility and larger the weight ω should be. The weight ω depend on t $\rightarrow$ its computation will be volatility path dependent.

Conditioning on the volatility path, we choose the weight to minimize the (asymptotic) conditional MSE:

$$E_c[\widetilde{V}_t(\omega_t) - QV_t]^2 = E_c[\hat{V}_t - QV_t - \omega_t(\hat{V}_t - \hat{V}_{t|t-1})]^2$$

Conditional to volatility path, $E_c[\hat{V}_t - \hat{V}_{t|t-1}] \neq 0.$

Optimal weight (control variables with bias, Glynn and Iglehart,1998)

$$\omega_t^* = \frac{Cov_c[\hat{V}_t, \hat{V}_t - \hat{V}_{t|t-1}]}{Var_c[\hat{V}_t - \hat{V}_{t|t-1}] + (E_c[\hat{V}_t - \hat{V}_{t|t-1}])^2}.$$

$$\omega_t^* = \frac{2Q_t}{2Q_t + 2\varphi^2 Q_{t-1} + n^{2\alpha}[QV_t - \varphi QV_{t-1} - (1-\varphi)\sigma^2]^2}$$

Properties of optimal weights:

- In order to estimate volatility on day t, we give a non-zero weight ω^* to prediction error from volatility information on day (t-1).
- This weight increases as the relative size of the variance 2Q(t)/n of V(t) is large in comparison to both:

(i) The variance 2Q(t-1)/n of V(t-1)

(ii) The squared conditional bias of forecast

- For a given non-zero conditional bias, the optimal weight ω^* goes to zero when n goes to infinity.

Local to unity:

- There likely is a sensible trade-off between:

(i) Forecasting squared bias

(ii) Variance of our intraday estimator

- This trade-off can be rationalized by considering the correlations coefficients φ as functions of n and assuming that:

$$1-\varphi_0^2(n) = O[\varphi_0(n) - \varphi(n)]$$

$\rightarrow$ Measure of the quality of our intraday information.

Different feasible weighting schemes:

- Optimal: $(\omega_u^*)^{-1} = 2 + \frac{n(\hat{V}_t - \hat{V}_{t|t-1})^2}{Q_t} + \frac{Q_{t-1} - Q_t}{Q_t}$
- "Kalman": $(\omega_K^*)^{-1} = 2 + \frac{nVar(\hat{V}_t - \hat{V}_{t|t-1})}{Q_t} + \frac{Q_{t-1} - Q_t}{Q_t}$
- Unconditional : $(\omega^{unc})^{-1} = 2 + \frac{nVar[\hat{V}_t - \hat{V}_{t|t-1}]}{E(Q_t)} > 2$

(Meddahi,2002)

With correction for heteroskedasticity: $(\omega^{unc-hc})^{-1} = .. + \frac{Q_{t-1} - Q_t}{Q_t}$

- Rule of thumb : $\omega^* = 1/2$

(Andreou and Ghysels,2002)

Summary of simulation/empirical results

- Conditional approach = significantly more efficient
- Bias = very small (3%)
- Optimal weights as large as 25% in average for simulated data with 10 minutes intervals (with 20% MSE improvement) on simulated data
- GM stock over 3 years 2000-2002 using 5-minutes returns $\rightarrow$ Optimal weight: 15% in average

4. The instantaneous horizon:

- Start from an estimator of integrated variance:

$$\hat{V}_{T_1,t} \xrightarrow{P} \int_{T_1}^{t} \sigma_s^2 ds$$

- Two natural estimators of spot volatility:

(i) The fixed reference estimator:

$$\hat{\sigma}_t^2 = \frac{1}{h}\left(\hat{V}_{T_1,t} - \hat{V}_{T_1,t-h}\right), T_1 < t-h < t.$$

(ii) The internal information estimator:

$$\hat{\sigma}_t^2 = \frac{1}{h}\hat{V}_{t-h,t}$$

$\rightarrow$ Difference due to edge effects (only in case of microstructure noise correction).

$\rightarrow$ Fixed reference in this talk.

$$\hat{V}_{T_1,t} = [X,X]_{T_1,t} + n^{-\alpha} M_t^{(n)} + n^{-\alpha} Z_t^{(n)} + negligible$$

$$\alpha = \text{convergence rate } (1/2, 1/4, 1/6, ..)$$

$$M_t^{(n)} = martingale$$

$$Z_t^{(n)} = "edge\ effect"$$

$$M_t^n \xrightarrow{d} M_t = \int_{T_1}^t f_s dW_s^* + M_t^*$$

W* = Brownian motion independent of the underlying σ-field ***F***

M* = martingale adapted to σ-field ***F***

(M*≠0 if endogenous times)

f = adapted to σ-field ***F***

R1. Edge effect at point t = asymptotically normal, independent of other limits (including edge effects) at other points).

R2 . Edge effect at fixed origin point T1 = immaterial

KEY DECOMPOSITION:

$$\hat{\sigma}_t^2 - \sigma_t^2 = \underbrace{\frac{n^{-\alpha}}{h}\left(M_t^{(n)} - M_{t-h}^{(n)} + Z_t^{(n)} - Z_{t-h}^{(n)} + o_P(n^{-\alpha/2})\right)}_{\text{"variance.term"}}$$

$$\underbrace{-\frac{1}{h}\int_{t-h}^{t}(\sigma_u^2 - \sigma_t^2)du}_{\text{"bias.term"}}$$

Under the assumption that volatility is a semimartingale

- Bias term ~ another variance term

$$-\frac{1}{h}\int_{t-h}^{t}(\sigma_u^2-\sigma_t^2)du \approx \frac{1}{h}\int_{t-h}^{t}(t-u)d\sigma_u^2$$

$$\to \text{Quadratic variation}: \frac{1}{h^2}\int_{t-h}^{t}(t-u)^2 d[\sigma^2,\sigma^2]_u \approx \frac{h}{3}\frac{\partial[\sigma^2,\sigma^2]_t}{\partial t} = O_P(h)$$

Conclusion: the squared bias term is order h to balance with the variance term of order : $n^{-\alpha}/h$

$\Rightarrow$ Optimal balance: $h = cn^{-\alpha/2}$

We deduce that up to edge effects:

$$n^{\alpha/2}(\hat{\sigma}_t^2 - \sigma_t^2) \xrightarrow{d} N(0, \eta_t^2) \text{ with:}$$

$$\eta_t^2 = \frac{1}{c}\left(f_t^2 + \frac{\partial [M^*, M^*]_t}{\partial t} \right) + \frac{c}{3} \frac{\partial [\sigma^2, \sigma^2]_t}{\partial t}$$

$\rightarrow$ Possible to optimize on c

Key result: Going from estimation of integrated volatility to estimation of spot volatility, the rate of convergence is divided by TWO.

WARNING: In principle, anything can happen with edge effects (case-by-case study)

$\rightarrow$ negligible with TSRV