Why is volatility estimation under microstructure noise so difficult?

Markus Reiß

Institut für Mathematik
Humboldt-Universität zu Berlin
www.mathematik.hu-berlin.de/∼mreiss

Haindorf Workshop, 11 February 2010

Outline

High-frequency data

First “naive” model:

We dispose of asset price quotes observed at high frequency over a period, e.g. log returns of a stock over a day:

$$\boxed{X_{t_i}, \quad 0 = t_0 < \cdots < t_n = 1, \quad \Delta t_i := t_i - t_{i-1} \to 0}.$$

Goal:

Financial mathematics requires $(X_t, t \geqslant 0)$ to be a semi-martingale and its volatility is of key interest. Here we restrict to continuous semi-martingales without drift:

$$\boxed{dX_t = \sigma_t dB_t, \quad t \in [0,1] \quad (B: \text{ Brownian motion})}$$

We want to estimate the *integrated volatility* $IV = \int_0^1 \sigma_t^2 dt$ or the *spot volatility* $(\sigma_t^2, t \in [0,1])$.

A natural estimator

$$dX_t = \sigma_t dB_t, \quad t \in [0,1] \quad (B: \text{ Brownian motion})$$

We know that the quadratic variation (integrated volatility) of X is given as limit (in L^2)

$$IV = \int_0^1 \sigma_t^2 dt = \lim_{\max_i \Delta t_i \to 0} \sum_{i=1}^{n} (X_{t_i} - X_{t_{i-1}})^2.$$

The natural estimator therefore is

$$\widehat{IV}_n := \sum_{i=1}^{n} (X_{t_i} - X_{t_{i-1}})^2$$

Properties of the natural estimator

$$\widehat{IV}_n := \sum_{i=1}^{n} (X_{t_i} - X_{t_{i-1}})^2$$

The estimator is unbiased:

$$\mathbb{E}[\widehat{IV}_n - IV] = \sum_{i=1}^{n} \mathbb{E}\left[\Big(\int_{t_{i-1}}^{t_i} \sigma_s \, dB_s\Big)^2 - \int_{t_{i-1}}^{t_i} \sigma_s^2 ds\right] = 0$$

and has variance (assume $t_i = i/n$ and σ_t determ., cts.):

$$\mathrm{Var}(\widehat{IV}_n) = \sum_{i=1}^{n} \mathrm{Var}((X_{t_i} - X_{t_{i-1}})^2) = \sum_{i=1}^{n} 2\Big(\int_{t_{i-1}}^{t_i} \sigma_s^2 ds\Big)^2$$

$$\approx \sum_{i=1}^{n} 2\sigma_{t_i}^4 (\Delta t_i)^2 \approx \frac{2}{n} \int_0^1 \sigma_t^4 dt.$$

Asymptotics of the natural estimator

$$\widehat{IV}_n := \sum_{i=1}^{n} (X_{t_i} - X_{t_{i-1}})^2$$

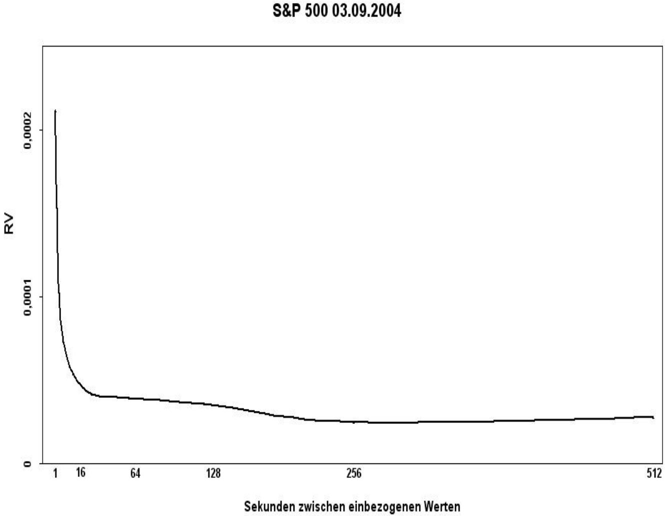
Using a CLT for triangular schemes one can indeed show

$$\sqrt{n}(\widehat{IV}_n - IV) \Rightarrow N\left(0, 2\int_0^1 \sigma_t^4 dt\right)$$

Econometricians call $\int_0^1 \sigma_t^4 dt$ the integrated quarticity IQ.
Note: usual $\sqrt{n}$-rate of convergence.

Problem:
Real data do not fit the model...

Applying the natural estimator



Microstructure noise model

Common explication:

Observed prices are not described by semi-martingales because market microstructure effects interfere, like bid-ask spreads, round-off errors, transaction costs...

Model under microstructure noise:

The *efficient price process* X is observed under measurement errors due to microstructure noise:

$$\text{Observations: } Y_i = X_{t_i} + \varepsilon_i, \quad i = 1, \dots, n$$

Simplifying assumptions (here):

(ε_i) are i.i.d., $N(0, \delta^2)$, independent of X, $\delta > 0$ known

$t_i = i/n$ equidistant

σ_t is deterministic and C^α (α-Hölder continuous)

Natural estimator under noise

Recall $Y_i = X_{t_i} + \varepsilon_i$ and consider the bias-corrected estimator

$$\widehat{IV}_n := \sum_{i=1}^{n} \left((Y_i - Y_{i-1})^2 - 2\delta^2 \right)$$

Then:

$$\mathbb{E}[\widehat{IV}_n] = \sum_{i=1}^{n} (\mathbb{E}[(X_{t_i} - X_{t_{i-1}})^2] + \mathbb{E}[(\varepsilon_i - \varepsilon_{i-1})^2] - 2\delta^2) = IV,$$

but:

$$\mathrm{Var}(\widehat{IV}_n) \approx \frac{2}{n} IQ + 6\delta^2 IV + 12\delta^4 n \to \infty.$$

Explication:

In the observed increments $Y_i - Y_{i-1}$ the "signal" $X_{t_i} - X_{t_{i-1}}$ is $O_P(n^{-1/2})$, the noise $\varepsilon_i - \varepsilon_{i-1}$ is $O_P(1)$.

How to cope with microstructure noise?

Common idea:
Smooth out the microstructure noise by averaging, using quadratic forms of the type $\sum_{i,k} w_{ik} Y_i Y_{i-k}$.

Multiscale estimator:
Zhang (2006) takes averages over IV-estimators at different sample frequencies.

Realized kernels:
Barndorff-Nielsen, Hansen, Lunde, Shephard (2008) average auto-covariance-type sums over different localisations.

Pre-averaging:
Jacod, Li, Mykland, Podolskij, Vetter (2008) plug local averages over Y_i into the natural estimator.

Convergence results

All authors obtain a $n^{1/4}$-rate of convergence with asymptotic normality involving integrated quarticity IQ.

A first explanation of the $n^{1/4}$-rate:
As for pre-averaging, we consider

$$\overline{Y}_i := \frac{1}{k_n} \sum_{\ell=1}^{k_n} Y_{i+\ell-k_n/2}$$

For k_n/n small, we have $\overline{Y}_i \approx X_{t_i} + \frac{1}{\sqrt{k_n}} \tilde{\varepsilon}_i$, $\tilde{\varepsilon}_i \sim N(0, \delta^2)$.
Using $\overline{Y}_{ik_n}$, $i = 1, \dots, n/k_n$, instead of Y_i, $i = 1, \dots, n$, in the natural estimator gives a variance of order $\frac{k_n}{n} IQ + \frac{n}{k_n} (\frac{\delta}{\sqrt{k_n}})^4$.
The rate-optimal choice $k_n = \sqrt{n}$ gives the variance $O(n^{-1/2})$.

More thorough analysis

Gloter, Jacod (2001) have proved an LAN-result which shows for the parametric problem that $n^{1/4}$ is the optimal rate. For the benchmark case $\sigma_t^2 \equiv \sigma^2$ constant the optimal asymptotic variance is $8\sigma^3\delta$. None of the discussed methods really achieves this efficiency bound.

Curious questions:

- Why is the $n^{1/4}$-rate optimal?
- Why is not δ^2 the factor in the asymptotic variance?
- Why do we get σ^3, not σ^4 in the asymptotic variance?
- What is efficiency in the nonparametric setup?
- How to construct an efficient nonparametric estimator?

White noise observation model

We prefer the continuous white noise observation model:

$$dY_t = X_t dt + \varepsilon\, dW_t, \quad t \in [0,1], \quad \varepsilon = \frac{\delta}{\sqrt{n}} \to 0$$

This means that for any function $\varphi \in L^2$ we observe

$$\langle \varphi, dY \rangle := \int_0^1 \varphi(t) dY_t = \int_0^1 \varphi(t) X_t dt + \varepsilon \int_0^1 \varphi(t)\, dW_t.$$

Here, W denotes another Brownian motion (independent of B!).
The analogue in the regression model is observing

$$\frac{1}{n}\sum_{i=1}^n \varphi(i/n) Y_i = \frac{1}{n}\sum_{i=1}^n \varphi(i/n) X_{i/n} + \frac{1}{n}\sum_{i=1}^n \varphi(i/n)\varepsilon_i.$$

Note: $\varepsilon \int_0^1 \varphi(t)\, dW_t \sim N(0, \delta^2 n^{-1} \int \varphi^2)$,
$\frac{1}{n}\sum_{i=1}^n \varphi(i/n)\varepsilon_i \sim N(0, \delta^2 n^{-2} \sum_{i=1}^n \varphi(i/n)^2)$.

Mathematical result

Regression model:

$$Y_i = X_{i/n} + \varepsilon_i, \quad i = 1, \dots, n, \quad \varepsilon_i \sim N(0, \delta^2) \text{ i.i.d.}$$

White noise model:

$$dY_t = X_t dt + \varepsilon\, dW_t, \quad t \in [0,1], \quad \varepsilon = \delta/\sqrt{n}$$

Theorem.(MR)
Both observation models are asymptotically statistically equivalent for $n \to \infty$ provided $\sigma^2 \in C^\alpha$ for some $\alpha > 0$.
Conclusion:
All inference procedures from one model have the same asymptotic properties in the other model.
Remark:
The models have a Bayesian flavour in the sense that the spot volatility σ_t^2 generates a random price process X_t which is the function observed under noise.

Observation laws

White noise model:

$$dY_t = X_t dt + \varepsilon\, dW_t, \quad t \in [0,1], \quad \varepsilon = \delta/\sqrt{n}$$

For any (weight) function φ with $\|\varphi\|_{L^2} = 1$ we observe

$$\begin{aligned}\int_0^1 \varphi(t) dY_t &= \int_0^1 \varphi(t) X_t\, dt + \varepsilon \int_0^1 \varphi(t)\, dW_t \\ &= \Phi(1)X_1 - \Phi(0)X_0 - \int_0^1 \Phi(t)\sigma_t\, dB_t + \varepsilon \int_0^1 \varphi(t)\, dW_t \\ &= \Big(\int_0^1 \Phi^2(t)\sigma_t^2\, dt + \varepsilon^2\Big)^{1/2} \zeta_\varphi\end{aligned}$$

with the antiderivative Φ of φ ($\Phi'(t) = \varphi(t)$, $\Phi(1) = 0$) and $\zeta_\varphi \sim N(0,1)$.
If $\sigma_t = const.$ on the support of φ, the "information load" is $\int \Phi^2$.

Maximal local information load

Which weight function φ_k with $\|\varphi_k\|_{L^2} = 1$ maximizes the information load $\int \Phi_k^2$ localized to a block $[kh, (k+1)h]$, i.e. under the constraint $\operatorname{supp}(\varphi_k) = \operatorname{supp}(\Phi_k) = [kh, (k+1)h]$?
Lagrange theory yields:

$$\varphi_k(t) = \sqrt{2}h^{-1/2} \cos(\pi(t - kh)/h) \mathbf{1}_{[kh,(k+1)h]}(t)$$

Compare with the standard choice in preaveraging, where $\tilde{\varphi}_k(t) = h^{-1/2}(\mathbf{1}_{[kh,(k+1/2)h]}(t) - \mathbf{1}_{[(k+1/2)h,(k+1)h]}(t))$.
For $k = 0, \ldots, h^{-1} - 1$ we have the observations

$$y_k := \int \varphi_k(t)\, dY_t = \left(h^2 \pi^{-2} \bar{\sigma}_k^2 + \varepsilon^2\right)^{1/2} \zeta_k$$

with $\zeta_k \sim N(0,1)$ i.i.d. and
$\bar{\sigma}_k^2 = 2h^{-1} \int_{kh}^{(k+1)h} \sigma_t^2 \sin^2(\pi(t - kh)/h)\, dt \approx \sigma_{kh}^2$.

A very simple regression model

From the observations for $k = 0, \dots, h^{-1} - 1$

$$y_k := \int_0^1 \varphi_k(t)\, dY_t = \left(h^2\pi^{-2}\bar{\sigma}_k^2 + \varepsilon^2\right)^{1/2}\zeta_k$$

we obtain by log-transformation the regression model

$$z_k := \log(y_k^2 h^{-2}\pi^2) - \mathbb{E}[\log(\zeta_k^2)] = \log\left(\bar{\sigma}_k^2 + \varepsilon^2 h^{-2}\pi^2\right) + \eta_k$$

with $\eta_k = \log(\zeta_k^2) - \mathbb{E}[\log(\zeta_k^2)] \sim IID(0,2)$.

A second explanation of the $n^{1/4}$-rate:

The block length h influences our information on σ_t^2 in (z_k): the linearized signal-to-noise ratio is of order $h(1 + h^4/\varepsilon^4)$ which is minimal for $h \sim \varepsilon \sim n^{-1/2}$. Choosing $h = h_0\varepsilon$ with some constant $h_0 > 0$, we obtain $\varepsilon^{-1} \sim \sqrt{n}$ observations

$$z_k = \log\left(\bar{\sigma}_k^2 + h_0^{-2}\pi^2\right) + \eta_k.$$

Theoretical framework

Theorem.(MR)
Let $h = h_0\varepsilon$ and $\sigma^2 \in C^\alpha$ for some $\alpha > 1/2$, $\sigma^2 > 0$. Then observing (y_k) is asymptotically statistically equivalent to the Gaussian regression model

$$z_k = \frac{1}{\sqrt{2}} \log\left(\sigma_{kh}^2 + h_0^{-2}\pi^2\right) + \zeta_k, \quad k = 0, \ldots, h^{-1} - 1$$

with $\zeta_k \sim N(0,1)$ i.i.d. and to the Gaussian shift model

$$dZ_t = \frac{1}{\sqrt{2h_0}} \log\left(\sigma_t^2 + h_0^{-2}\pi^2\right) dt + \varepsilon^{1/2} dW_t, \quad t \in [0,1]$$

with an independent Brownian motion W and $\varepsilon^{1/2} = \delta^{1/2} n^{-1/4}$.

First consequences

Corollary.
The microstructure noise model is asymptotically more informative than observing (y_k), (z_k) resp. (Z_t) above.

Corollary.
We can estimate the spot volatility $(\sigma_t^2, t \in [0,1])$ nonparametrically with rate $n^{-\alpha/(4\alpha+2)}$ for $\sigma_t^2 \in C^\alpha$, $\alpha > 1/2$: apply a standard kernel estimator to (z_k) or directly to $(h^2\pi^{-2}y_k^2)$ which has the right mean, but is heteroskedastic.

How much do we lose?

Benchmark case: $\sigma_t^2 \equiv \sigma^2$ constant
For the optimal oracle choice $h_0 = \sqrt{3}\pi\sigma^{-1}$ the Fisher information from observing (z_k) or (Z_t) is $\frac{\sqrt{27}}{32\pi}\sigma^{-3}\delta n^{-1/2}$ such that an optimal estimator only based on theses observations has a by 50% larger standard deviation than the parametrically optimal Gloter/Jacod-estimator (a non-explicit MLE).

Surprising fact:
We only lose information on σ_t^2 by a (small) constant factor. The benefit is an easy nonparametric regression model with i.i.d. noise.

Why is most information already stored in the (y_k)?

The case of constant σ_t^2

$$dY_t = \sigma B_t dt + \varepsilon\, dW_t, \quad t \in [0,1].$$

Natural idea:

Diagonalize the covariance operator of B_t.
ONB of eigenfunctions: $\psi_j(x) = \sqrt{2}\sin((j-1/2)\pi t)$, $j \geqslant 1$
(Karhunen-Loève basis)

Equivalent observations:

$$y_j^{\psi} := \int_0^1 \psi_j(t)\, dY_t \sim N\Big(0, \frac{\sigma^2}{(j-1/2)^2\pi^2} + \varepsilon^2\Big)$$

independently for $j \geqslant 1$.

How to estimate constant σ^2

The MLE for σ^2 at *frequency* j is given by

$$\hat{\sigma}_j^2 = (j - 1/2)^2\pi^2((y_j^\psi)^2 - \varepsilon^2)$$

and satisfies

$$\mathbb{E}[\hat{\sigma}_j^2] = \sigma^2, \quad \operatorname{Var}(\hat{\sigma}_j^2) = 2(\sigma^2 + \varepsilon^2(j - 1/2)^2\pi^2)^2.$$

Consider an average over the independent $\hat{\sigma}_j^2$:

$$\hat{\sigma}^2 = \sum_{j\geqslant 1} w_j\hat{\sigma}_j^2, \quad \sum_j w_j = 1.$$

Then $\hat{\sigma}^2$ is unbiased with variance $\sum_j w_j^2 \operatorname{Var}(\hat{\sigma}_j^2)$. Optimal (oracle) weights $w_j \sim 1/\operatorname{Var}(\hat{\sigma}_j^2)$ yield the MLE with

$$\operatorname{Var}(\hat{\sigma}^2) = \frac{2}{\sum_{j\geqslant 1}(\sigma^2+\varepsilon^2(j-1/2)^2\pi^2)^{-2}} \approx \frac{2\varepsilon}{\int_0^\infty(\sigma^2+x^2\pi^2)^{-2}dx} = 8\sigma^3\varepsilon.$$

A third explanation of the asymptotics: constant σ

Recall that we have independent estimators $\hat{\sigma}_j^2$ with

$$\mathbb{E}[\hat{\sigma}_j^2] = \sigma^2, \quad \mathrm{Var}(\hat{\sigma}_j^2) = 2(\sigma^2 + \varepsilon^2 (j - 1/2)^2 \pi^2)^2.$$

For frequencies $j \lesssim \varepsilon^{-1} \sim \sqrt{n}/\delta$ the variance is bounded, for $j \gg \varepsilon^{-1} \sim \sqrt{n}/\delta$ the overwhelming part is just noise. Averaging over the $\hat{\sigma}_j^2$, $j = 1, \ldots, \varepsilon^{-1}$, gives an unbiased estimator with variance of order $\varepsilon \sim \delta n^{-1/2}$.

If the weights w_j are not data-driven, i.e. dependent on the true σ or an estimator of it, then the asymptotic variance involves powers σ^4, σ^2. Using estimates of the optimal weights $w_j \sim (\sigma^2 + \varepsilon^2 (j - 1/2)^2 \pi^2)^{-2}$ yields the value $8\sigma^3$.

Putting everything together

$$dY_t = X_t dt + \varepsilon\, dW_t, \quad t \in [0,1], \quad \varepsilon = \delta/\sqrt{n}$$

Findings:

- The system (φ_k) of cosine functions is localized on blocks $[kh, (k+1)h]$.
- The spot volatility $\sigma_t^2 \in C^\alpha$ is well approximated by a blockwise constant function if $h \ll \varepsilon^{1/2\alpha}$.
- For constant volatility full information is spread over different frequencies j.

Time-frequency analysis

Consider the orthonormal weight functions

$$\varphi_{jk}(t) = \sqrt{2}h^{-1/2}\cos(j\pi(t-kh)/h)\mathbf{1}_{[kh,(k+1)h]}(t)$$

for blocks $k = 0, \dots, h^{-1} - 1$ and frequencies $j \geqslant 1$. They give rise to the observations

$$\boxed{y_{jk} := \langle \varphi_{jk}, dY \rangle = \left(h^2\pi^{-2}j^{-2}\sigma_{kh}^2 + \varepsilon^2\right)^{1/2}\zeta_{jk}}$$

with $\zeta_{jk} \sim N(0,1)$ i.i.d. (for blockwise constant volatility).

Efficient estimator

$$y_{jk} := \langle \varphi_{jk}, dY \rangle = \Big(h^2\pi^{-2}j^{-2}\sigma_{kh}^2 + \varepsilon^2\Big)^{1/2}\zeta_{jk}$$

Estimate integrated volatility IV by summing frequency-weighted y_{jk}^2 over blocks:

$$\widehat{IV}_\varepsilon := \sum_{k=0}^{h^{-1}-1} h \sum_{j=1}^{J} w_j(\tilde\sigma_\varepsilon^2(kh))h^{-2}j^2\pi^2(y_{jk}^2 - \varepsilon^2)$$

We use optimal weights depending on a nonparametric pilot estimator $\tilde\sigma_t^2$ of the spot volatility σ_t^2 (see above).
Choosing h slightly larger than ε and the maximal frequency slowly increasing, we obtain efficiency.

Efficient estimator

$$\widehat{IV}_\varepsilon := \sum_{k=0}^{h^{-1}-1} h \sum_{j=1}^{J} w_j(\tilde{\sigma}_\varepsilon^2(kh))h^{-2}j^2\pi^2(y_{jk}^2-\varepsilon^2)$$

Theorem. (MR)

Let $\sigma_t^2 \in C^\alpha$, $\alpha > 1/2$, $\sigma_t^2 > 0$, $h = \varepsilon \log(\varepsilon^{-1})$, $J \gg \log(\varepsilon^{-1})$. Then

$$\varepsilon^{-1/2}(\widehat{IV}_\varepsilon - IV) \Rightarrow N\Big(0, 8\int_0^1 \sigma_t^3\,dt\Big).$$

The estimator achieves the optimal asymptotic variance.

Remark.

The optimal variance does *not* involve integrated quarticity IQ, but $\int \sigma^3$. This can only be achieved by a data-driven local weighting scheme.

Statistical equivalence

Using the (y_{jk}) and applying the theory by Grama, Nussbaum (2002), a localisation scheme and fine bounds on Hellinger distances for cylindrical Gaussian measures, a complete asymptotic equivalence result can be established.

Theorem.(MR)
Let $\sigma_t^2 \in C^\alpha$, $\alpha > (1+\sqrt{5})/4$, $\sigma_t^2 > 0$. Then observing

$$Y_i = X_{i/n} + \varepsilon_i, \quad \varepsilon_i \sim N(0, \delta^2) \text{ i.i.d.}, \quad i = 1, \dots, n,$$

with $X_t = \int_0^t \sigma_s dB_s$ is asymptotically equivalent for $n \to \infty$ to observing

$$\boxed{dY_t = \sqrt{2\sigma_t}dt + \delta^{1/2} n^{-1/4} dW_t, \quad t \in [0,1]}$$

Conclusion

- Smoothing is needed for volatility estimation under microstructure noise.
- Block size/spectral cut-off explains the rate $n^{1/4}$.
- A very simple reduced model increases standard deviation by only 50%.
- Rate-optimal results are easy, e.g. nonparametric estimation of the spot volatility.
- Standard quadratic form estimators for IV are not efficient.
- Efficient estimator uses time-frequency analysis and locally adaptive weights.
- Asymptotic equivalence permits to transfer all other kinds of inference on σ_t^2, e.g. testing.

THANK YOU FOR YOUR ATTENTION!

Conclusion

- Smoothing is needed for volatility estimation under microstructure noise.
- Block size/spectral cut-off explains the rate $n^{1/4}$.
- A very simple reduced model increases standard deviation by only 50%.
- Rate-optimal results are easy, e.g. nonparametric estimation of the spot volatility.
- Standard quadratic form estimators for IV are not efficient.
- Efficient estimator uses time-frequency analysis and locally adaptive weights.
- Asymptotic equivalence permits to transfer all other kinds of inference on σ_t^2, e.g. testing.

THANK YOU FOR YOUR ATTENTION!