# Volatility estimation under microstructure noise and Le Cam theory

Markus Reiß

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

Workshop Le Mans, 22 March 2011

# Outline

# High-frequency data

**Observations:**
We dispose of asset prices observed over a period,
e.g. log prices of a stock over a day:

$$\boxed{X_{t_i}, \quad 0 = t_0 < \cdots < t_n = 1}$$

Often, tick data is available at time intervals $\Delta t_i \leqslant 1$ sec. over a day, resulting in 30,000 observations per day and more.

**Mathematical model:**
The axiom of *no arbitrage* requires $(X_t,\ t \geqslant 0)$ to be a *semi-martingale*.

# High-frequency data

Observations:

We dispose of asset prices observed over a period, e.g. log prices of a stock over a day:

$$\boxed{X_{t_i}, \quad 0 = t_0 < \cdots < t_n = 1}$$

Often, tick data is available at time intervals $\Delta t_i \leqslant 1$ sec. over a day, resulting in 30,000 observations per day and more.

Mathematical model:

The axiom of *no arbitrage* requires $(X_t,\ t \geqslant 0)$ to be a *semi-martingale*.

# Main objective

Mathematical finance:

The quadratic variation or *integrated volatility* (IV) of $X$ is of key interest for pricing, hedging, risk management etc.

$$IV_t := [X, X]_t = \lim_{\Delta t_i \to 0} \sum_{t_i \leqslant t} (X_{t_i} - X_{t_{i-1}})^2$$

Principal objective:

Use HF data to determine realized volatility (ex post).

# Assumptions (here)

Continuous martingale model (no jumps):

$$X_t = X_0 + \int_0^t \sigma_s \, dB_s, \quad B: \text{ Brownian motion}$$

$\sigma_s$: unknown *spot volatility* function

Observation design:
Equidistant time points $t_i = i/n$ and $n \to \infty$.

# Asymptotics of realized variance

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

Using a CLT for triangular schemes one can show

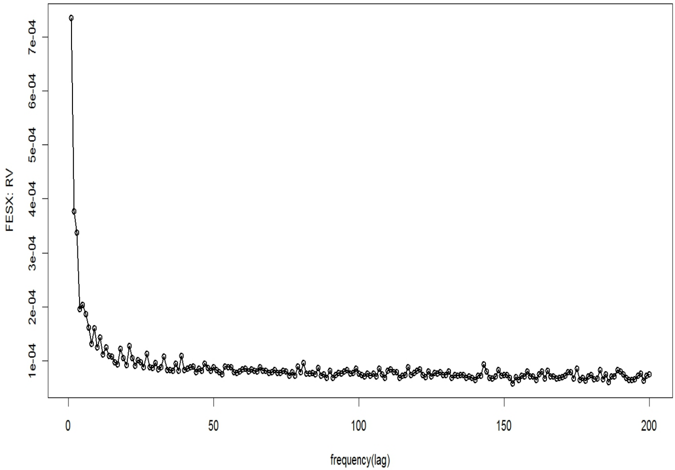
$$\sqrt{n}(\widehat{IV}_n - IV) \Rightarrow N\Big(0, 2\int_0^1 \sigma_t^4 dt\Big)$$

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 realized variance

Euro Stoxx 50 future, 10 January 2008, expiry 06/2008



# 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:

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

## Simplifying assumptions (only here!):

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

# 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 (only here!):

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

$\sigma^2$ deterministic

# Realized variance under noise

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

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

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:

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

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 $O_P(1)$.

# One way out

Local averaging:

Consider blocks $[kh,(k+1)h]$ on $[0,1]$ and take averages:

$$\overline{\Delta Y}_k^h = \frac{2}{nh}\sum_{t_i=kh}^{(k+1/2)h} Y_i - \frac{2}{nh}\sum_{t_i=(k+1/2)h}^{(k+1)h} Y_i$$

Then $\overline{\Delta Y}_k^h$ is centred with $\operatorname{Var}(\overline{\Delta Y}_k^h) \approx \frac{h}{12}\sigma^2(kh) + \frac{4}{nh}\delta^2$.

Result:

For $\widetilde{IV}_{n,h} := \sum_{k=0}^{h^{-1}-1} 12\big((\overline{\Delta Y}_k^h)^2 - \frac{4}{nh}\delta^2\big)$ we obtain

$$\operatorname{Var}(\widetilde{IV}_{n,h}) \sim hIQ + \delta^2(nh)^{-1}IV + \delta^4 n^{-2}h^{-3}$$

Rate-optimal tuning yields

$$h = h(n) \sim n^{-1/2}, \quad \operatorname{Var}(\widetilde{IV}_{n,h(n)})^{1/2} \sim n^{-1/4}$$

# How to cope with microstructure noise?

**Common idea:**

Smooth out the microstructure noise by averaging, using weighted 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 $k$.

**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 $Y_i - Y_{i-1}$ into realized volatility.

# Known results

Each method attains a $n^{1/4}$-rate of convergence with asymptotic normality involving integrated quarticity IQ (and IV).

Gloter, Jacod (2001) prove an LAN-result with $n^{1/4}$-rate. For $\sigma_t^2 \equiv \sigma^2$ constant the optimal asymptotic variance is $8\sigma^3\delta$. The nonparametric methods are designed to come close to this efficient asymptotic variance.

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

# Known results

Each method attains a $n^{1/4}$-rate of convergence with asymptotic normality involving integrated quarticity IQ (and IV).

Gloter, Jacod (2001) prove an LAN-result with $n^{1/4}$-rate. For $\sigma_t^2 \equiv \sigma^2$ constant the optimal asymptotic variance is $8\sigma^3\delta$. The nonparametric methods are designed to come close to this efficient asymptotic variance.

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

# Main result

HF-data experiment $\mathcal{E}_0(n, \delta, \alpha, R, \underline{\sigma}^2)$: set $X_t = \int_0^t \sigma_s \, dB_s$,

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

Gaussian shift experiment $\mathcal{G}_0(n, \delta, \alpha, R, \underline{\sigma}^2)$:

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

**Theorem.**

For $\|\sigma^2\|_{C^\alpha} \leqslant R$, $\alpha > \frac{1+\sqrt{5}}{4} \approx 0.81$, and $\sigma^2(t) \geqslant \underline{\sigma}^2 > 0$ the experiments $\mathcal{E}_0$ and $\mathcal{G}_0$ are asymptotically equivalent:

$$\lim_{n \to \infty} \Delta\Big(\mathcal{E}_0(n, \delta, \alpha, R, \underline{\sigma}^2), \mathcal{G}_0(n, \delta, \alpha, R, \underline{\sigma}^2)\Big) = 0.$$

# Consequences

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

Spot volatility:

Estimating $\sigma_t^2$ nonparametrically is possible with the classical nonparametric rates, but in terms of the noise level $\delta^{1/2} n^{-1/4}$ instead of $\delta n^{-1/2}$, cf. Munk, Schmidt-Hieber (2009).

Integrated volatility and other moments:

$T_p(\sigma) = \int_0^1 \sigma_t^p dt$, $p \geqslant 1$, corresponds to the functional $T_p(f) = 2^{-p} \int_0^1 f(x)^{2p} dx$. We have $T_p'(f) = \langle p2^{-p+1} f(x)^{2p-1}, \bullet \rangle$ and obtain (Ibragimov, Khasminskii 1991)

$$\delta^{-1/2} n^{1/4} (\widehat{T_p} - T_p) \Rightarrow N(0, \int T'(f)^2) = N(0, 2p^2 \int_0^1 \sigma_t^{2p-1} dt).$$

In particular: $n^{1/4}(\widehat{IV}_n - IV) \Rightarrow N(0, 8\delta \int_0^1 \sigma_t^3\, dt) \rightsquigarrow$ No IQ !

# First equivalence result

Regression experiment $\mathcal{E}_0(n, \delta, \alpha, R, \underline{\sigma}^2)$:

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

White noise experiment $\mathcal{E}_1(n, \delta, \alpha, R, \underline{\sigma}^2)$:

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

**Theorem.**
Both experiments are asymptotically equivalent as $n \to \infty$ for $\delta, \alpha, R > 0$, $\underline{\sigma}^2 \geqslant 0$.

**Remark:**
The proof relies on Hellinger bounds for cylindrical Gaussian measures and constructions from Carter (2006), MR (2008).
A pathwise approach using Brown, Low (1996) is not applicable due to $X_t \notin C^{1/2}$.

# Blockwise constant observation model

White noise experiment $\mathcal{E}_1(n, \delta, \alpha, R, \underline{\sigma}^2)$:

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

Blockwise constant white noise experiment $\mathcal{E}_2(n, \delta, h, \alpha, R, \underline{\sigma}^2)$:

$$dY_t^h = X_t^h dt + \frac{\delta}{\sqrt{n}} dW_t, \quad t \in [0, 1], \quad X_t^h = \int_0^t \sigma_{\lfloor s/h \rfloor h} dB_s$$

Theorem.

Assume $\alpha > 1/2, R, \underline{\sigma}^2 > 0$ and $h^\alpha = o(n^{-1/4})$. Then $\mathcal{E}_1$ and $\mathcal{E}_2$ are asymptotically equivalent for $n \to \infty$.

Remark:

This reflects the idea that a regular function $\sigma_t^2$ can be considered as constant on blocks $[kh, (k+1)h)$, $0 \leqslant k < h^{-1}$.

# Observation laws

$$dY_t^h = X_t^h dt + \frac{\delta}{\sqrt{n}} dW_t, \quad t \in [0,1], \quad X_t^h = \int_0^t \sigma_{\lfloor s/h \rfloor h} dB_s$$

For local orthonormal weight functions $\varphi_{jk}$ (and $\Phi'_{jk} = \varphi_{jk}$) with $\mathrm{supp}(\varphi_{jk}), \mathrm{supp}(\Phi_{jk}) \subseteq [kh, (k+1)h]$ and $(\Phi_{jk})$ orthogonal we observe

$$\begin{aligned} y_{jk} := \int \varphi_{jk}(t) dY_t^h &= \int \varphi_{jk}(t) X_t^h \, dt + \frac{\delta}{\sqrt{n}} \int \varphi_{jk}(t) \, dW_t \\ &= -\int \Phi_{jk}(t) \sigma_{kh} \, dB_t + \frac{\delta}{\sqrt{n}} \int \varphi_{jk}(t) \, dW_t \\ &= \Big(\sigma_{kh}^2 \int \Phi_{jk}^2(t) \, dt + \frac{\delta^2}{n}\Big)^{1/2} \zeta_{jk} \quad \zeta_{jk} \sim N(0,1) \text{i.i.d.} \end{aligned}$$

# Time-frequency analysis

A feasible basis (inspired by locally constant MLE):
Consider the orthonormal trigonometric functions

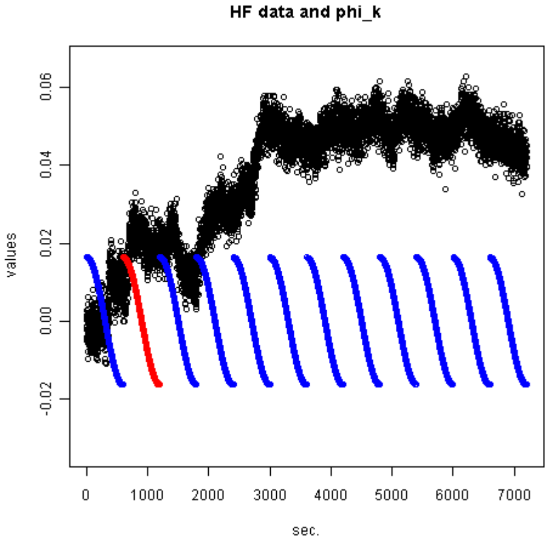
$$\varphi_{jk}(t) = \sqrt{2}h^{-1/2}\cos(j\pi(t-kh)/h)\mathbf{1}_{[kh,(k+1)h]}(t)$$

for blocks $k = 0, \ldots, h^{-1}-1$ and frequencies $j \geqslant 1$. Then $(\varphi_{jk})$ and $\Phi_{jk}$ are orthogonal and on each block $k$ we can observe

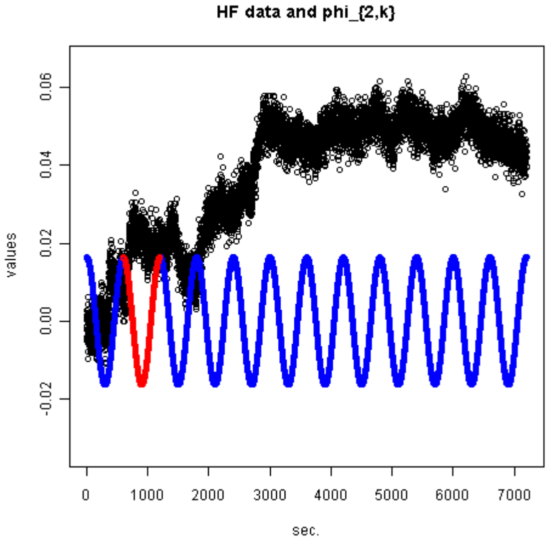
$$y_{jk} := \int \varphi_{jk}(t)dY_t^h = \left(h^2\pi^{-2}j^{-2}\sigma_{kh}^2 + \frac{\delta^2}{n}\right)^{1/2}\zeta_{jk}$$

with $\zeta_{jk} \sim N(0,1)$ i.i.d.
Note: $(\varphi_{jk})_{jk} \cup (\mathbf{1}([kh,(k+1)h]))_k$ is ONB in $L^2([0,1])$.

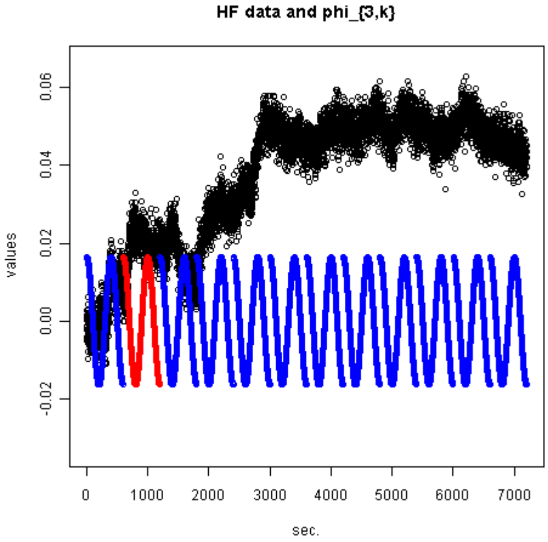
# Optimal weight functions on blocks



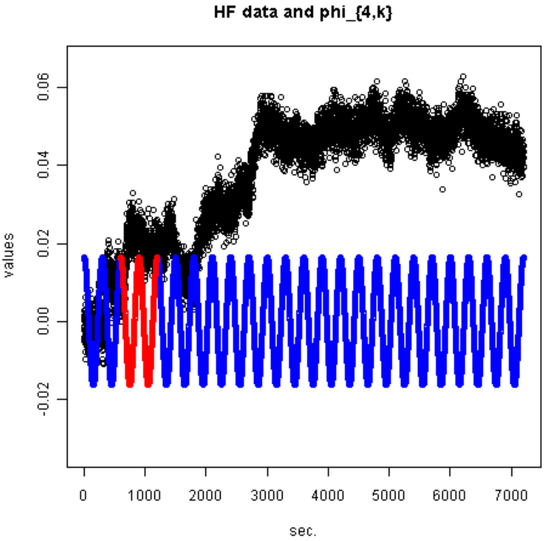
# Optimal weight functions on blocks



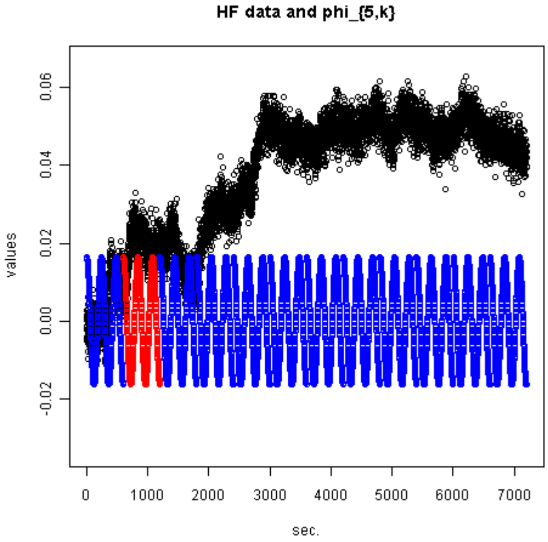
# Optimal weight functions on blocks



# Optimal weight functions on blocks



# Optimal weight functions on blocks



# Local nonparametric likelihood theory

1. For $\alpha > (1+\sqrt{5})/4$ and $hn^{1/2} \to \infty$ the $(y_{jk})$ are *asymptotically sufficient* in $\mathcal{E}_2$: we may neglect the observations

$$y_k^0 := \int \mathbf{1}_{[kh,(k+1)h]}(t)\, dY_t^h,$$

   which are heavily dependent on $(y_{jk})$.
   $\mapsto$ consider only the law of the diffusion bridges!
2. The $(y_{jk})$ are independent Gaussian with unknown scale parameter. The *local nonparametric likelihood theory* by Grama, Nussbaum (2002) and fine bounds on Gaussian Hellinger distances yield a *local asymptotic equivalence* result.
3. A *variance-stabilising transform* and a (not completely standard) *globalisation* procedure complete the proof.

# Efficient estimator

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

Estimate integrated volatility IV by summing frequency-weighted $y_{jk}^2$ over blocks (local-likelihood approach):

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

We use optimal weights $w_j$ depending on a nonparametric pilot estimator $\tilde{\sigma}_n(t)^2$ of the spot volatility $\sigma_t^2$.

# Efficient estimator

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

**Theorem.**
Let $\sigma_t^2 \in C^\alpha$, $\alpha > 1/2$, $\sigma_t^2 > 0$, $h = \frac{\delta \log(n)}{\sqrt{n}}$, $J \gg \log(n)$. Then

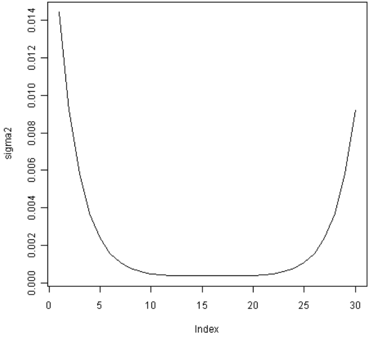
$$n^{1/4}(\widehat{IV}_n - IV) \Rightarrow N\Big(0, 8\delta \int_0^1 \sigma_t^3 \, dt\Big).$$

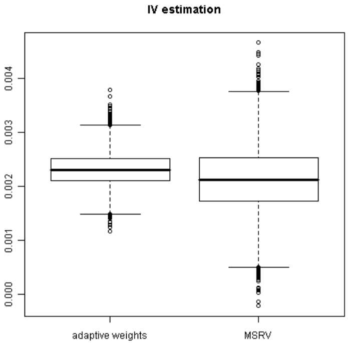
The estimator achieves the optimal asymptotic variance.

**Remark.**
The optimal variance can only be achieved by a data-driven local weighting scheme.

# Applying the estimator

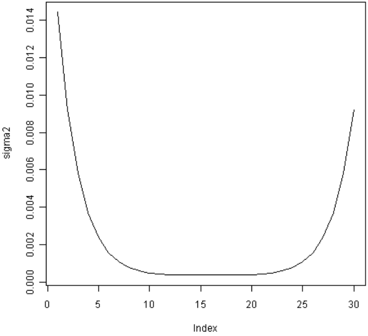
$$n = 30,000; \delta = 0.01; \sigma(t) = 0.02 + 0.1(1 - 2t)^4$$

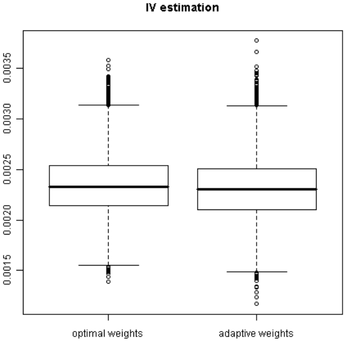


# Applying the estimator

MC simulations: $\text{Var}(\widehat{IV}_n)^{1/2} = 1.07 * (8\delta \int \sigma^3 / \sqrt{n})^{1/2}$
optimal weights: 1.03 (lower bound for global tuning: 1.19)



# Conclusion

- Smoothing is needed for volatility estimation under microstructure noise; efficient noise level is $\delta^{1/2} n^{-1/4}$.
- Standard quadratic form estimators for IV are not efficient.
- Efficient estimator uses time-frequency analysis (local Karhunen-Loève bases without linear function).
- Asymptotic equivalence permits to transfer all kinds of inference (tests, confidence) on $\sigma_t^2$ and its functionals.
- *Open questions:*
  Critical regularity; stochastic volatility; covariation

THANK YOU FOR YOUR ATTENTION!

# Conclusion

- Smoothing is needed for volatility estimation under microstructure noise; efficient noise level is $\delta^{1/2} n^{-1/4}$.
- Standard quadratic form estimators for IV are not efficient.
- Efficient estimator uses time-frequency analysis (local Karhunen-Loève bases without linear function).
- Asymptotic equivalence permits to transfer all kinds of inference (tests, confidence) on $\sigma_t^2$ and its functionals.
- *Open questions:*
  Critical regularity; stochastic volatility; covariation

THANK YOU FOR YOUR ATTENTION!