Volatility estimation from high-frequency observations with irregular errors – Concepts and consequences

or *Improved volatility estimation based on limit order books*

Markus Bibinger, joint work with Moritz Jirak and Markus Reiß

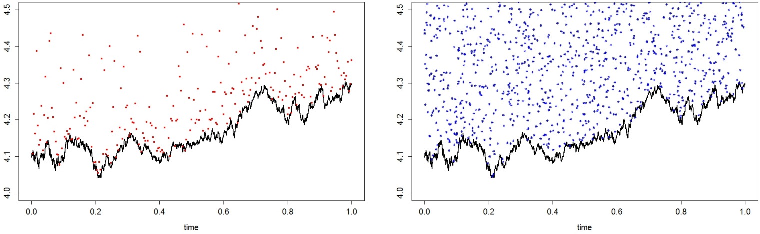
Outline

1. Model & Motivation
2. Volatility estimation based on local order statistics
3. Law of local minima and upper bound
4. Lower bound
5. Application

Outline

1. Model & Motivation
2. Volatility estimation based on local order statistics
3. Law of local minima and upper bound
4. Lower bound
5. Application

Statistical model



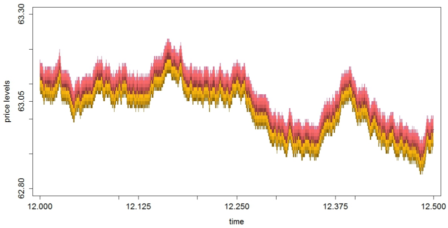
Continuous Itô semi-martingale $X_t = X_0 + \int_0^t a_s\,ds + \int_0^t \sigma_s\,dW_s$, $t \in [0,1]$, $(T_j, \mathcal{Y}_j)$ obs. of Poisson point process on $[0,1]\times\mathbb{R}$ with intensity measure

$$\Lambda(A) = \int_0^1 \int_{\mathbb{R}} \mathbb{1}_A(t,y)\lambda_{t,y}\,dt\,dy, \quad \lambda_{t,y} = n\lambda\mathbb{1}(y \geq X_t).$$

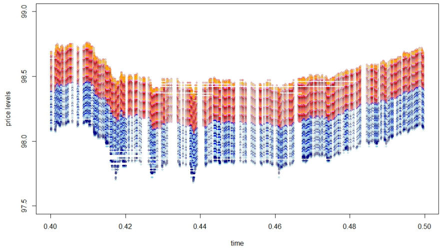
Connatural discrete-time model:

$$Y_i = X_{t_i^n} + \varepsilon_i,\ i = 0,\ldots,n,\ \varepsilon_i \geq 0,\ \varepsilon_i \overset{iid}{\sim} F_\lambda(x) = \lambda x(1+\mathcal{O}(1)).$$

Intra-day order book price dynamics



Order price levels for Facebook asset, 12:00 - 12:30, June 2nd 2014, levels 1-5, bid-ask spread colored in dark red.



Order levels 1-30 and arrivals for AAPL, 12:00 - 12:45, July 28th 2014.

Data by LOBSTER.

Groundwork & Contribution

Principal objective: Recovery of quadratic variation of stochastic boundary X_t.

Main application: Estimation of *integrated volatility* for portfolio and risk management.

Groundwork on volatility estimation:

- For discrete observations $X_{i/n}, i = 0, \ldots, n$, the *realized volatility* satisfies
$$n^{\frac{1}{2}}\left(\sum_{i=1}^{n}\left(X_{\frac{i}{n}} - X_{\frac{(i-1)}{n}}\right)^2 - \int_0^1 \sigma_s^2\,ds\right) \rightsquigarrow \mathrm{N}\left(0, 2\int_0^1 \sigma_s^4\,ds\right),$$
and is *asymptotically efficient*.

- However, the direct observation model does not accurately fit high-frequency data.

- Prominent *microstructure noise model*: $Y_i = X_{t_i^n} + \varepsilon_i$ with ε_i i.i.d., $\mathbb{E}[\varepsilon_i] = 0$.

Groundwork & Contribution

Principal objective: Recovery of quadratic variation of stochastic boundary X_t.

Main application: Estimation of *integrated volatility* for portfolio and risk management.

Groundwork on volatility estimation:

- For discrete observations $X_{i/n}, i = 0, \ldots, n$, the *realized volatility* is *asymptotically efficient*.
- However, the direct observation model does not accurately fit high-frequency data.
- Prominent *microstructure noise model*: $Y_i = X_{t_i^n} + \varepsilon_i$ with ε_i i.i.d., $\mathbb{E}[\varepsilon_i] = 0$. Efficient estimator by Bibinger et al (2014) for $\varepsilon_i \stackrel{iid}{\sim} N(0, \eta^2)$ satisfies
  $n^{\frac{1}{4}} \left(\hat{IV} - \int_0^1 \sigma_s^2 \, ds \right) \rightsquigarrow N\left(0, 8\eta \int_0^1 \sigma_s^3 \, ds\right)$.

Groundwork & Contribution

Principal objective: Recovery of quadratic variation of stochastic boundary X_t.

Main application: Estimation of *integrated volatility* for portfolio and risk management.

Groundwork on volatility estimation:

- For discrete observations $X_{i/n}, i = 0, \ldots, n$, the *realized volatility* is *asymptotically efficient*.
- However, the direct observation model does not accurately fit high-frequency data.
- Prominent *microstructure noise model*: $Y_i = X_{t_i^n} + \varepsilon_i$ with ε_i i.i.d., $\mathbb{E}[\varepsilon_i] = 0$. Efficient estimator by Bibinger et al (2014).
- Apply estimators to *which* time series of prices (micro prices, traded prices, etc.)?

Outline

1. Model & Motivation
2. Volatility estimation based on local order statistics
3. Law of local minima and upper bound
4. Lower bound
5. Application

Construction of an estimator

Partition the unit interval into $h_n^{-1} \in \mathbb{N}$ equi-spaced bins $\mathcal{T}_k^n = [kh_n, (k+1)h_n), k = 0, \ldots, h_n^{-1} - 1$, $nh_n \in \mathbb{N}$, $h_n \to 0$.
Parametric estimation theory motivates bin-wise minima

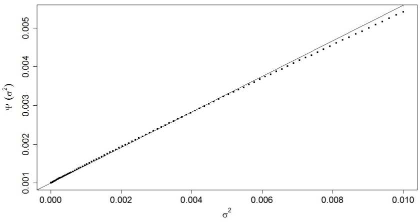
$$m_{n,k} = \min_{i \in \mathcal{I}_k^n} Y_i \ , \ \ \mathcal{I}_k^n = \{kh_n n, kh_n n + 1, \ldots, (k+1)h_n n - 1\},$$

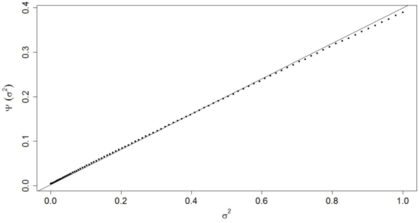
$$m_{n,k} = \min_{T_j \in \mathcal{T}_k^n} \mathcal{Y}_j \, .$$

as estimators of X_{kh_n}. $\mathbb{V}\mathrm{ar}(\min_{i \in \mathcal{I}_k^n} \varepsilon_i) \propto (n\lambda h_n)^{-2}$, so locally constant signal approximation $X_t = X_{kh_n} + \mathcal{O}_{\mathbb{P}}(h_n^{1/2})$ on $\mathcal{T}_k^n$ is only admissible when $h_n^{1/2} = o((n\lambda h_n)^{-1})$.
Optimal rate attained when

$$h_n = \mathcal{K}^{\frac{2}{3}} (n\lambda)^{-\frac{2}{3}}, \mathcal{K} > 0, \ nh_n \propto n^{\frac{1}{3}} \lambda^{-\frac{2}{3}} \, .$$

The function Ψ





Introduce for $X_t = X_{kh_n} + \sigma_{kh_n} \int_{kh_n}^{t} dW_s$ on $\mathcal{T}_k^n$ in PPP-model:

$$\Psi\left(\sigma^2_{kh_n}\right) = h_n^{-1}\mathbb{E}\Big[\big(m_{n,k} - m_{n,k-1}\big)^2\Big], k = 1, \ldots, h_n^{-1} - 1,$$

an invertible function, MC approximation above for $\mathcal{K} = 32$.

$$\Psi^{-1}\left(\sum_{k=(l-1)r_n^{-1}/2+1}^{lr_n^{-1}/2} \big(m_{n,2k} - m_{n,2k-1}\big)^2 2h_n^{-1} r_n\right) \approx \sigma^2_{lr_n^{-1}h_n},$$

where $r_n^{-1}h_n$ is a coarse grid size with $r_n h_n^{-1}, r_n^{-1} \in 2\mathbb{N}$.

The estimator based on local minima

$$\widetilde{IV}_n^{h_n,r_n} = \sum_{l=1}^{r_n h_n^{-1}} \Psi^{-1}\Bigg(\sum_{k=(l-1)r_n^{-1}/2+1}^{l r_n^{-1}/2} \big(m_{n,2k} - m_{n,2k-1}\big)^2 \, 2h_n^{-1} r_n \Bigg) h_n r_n^{-1} .$$

In the regression-type model

$$\Psi_n\big(\sigma^2_{kh_n}\big) = h_n^{-1}\mathbb{E}\Big[\big(m_{n,k} - m_{n,k-1}\big)^2\Big], k = 1, \ldots, h_n^{-1} - 1 ,$$

with a sequence $\Psi_n \to \Psi$. Estimator

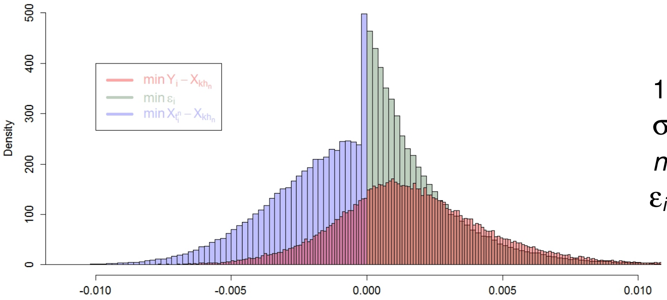
$$\widehat{IV}_n^{h_n,r_n} = \sum_{l=1}^{r_n h_n^{-1}} \Psi_n^{-1}\Bigg(\sum_{k=(l-1)r_n^{-1}/2+1}^{l r_n^{-1}/2} \big(m_{n,2k} - m_{n,2k-1}\big)^2 \, 2h_n^{-1} r_n \Bigg) h_n r_n^{-1} .$$

For $\sigma_t = \sigma =$ const., use $\widehat{IV}_n^{h_n,h_n}$. For Lipschitz σ_t balance approximation error $r_n^{-1} h_n$ with second order term on each coarse interval of order $r_n \Rightarrow r_n \propto h_n^{1/2} = (n\lambda)^{-1/3}$.

Outline

1. Model & Motivation
2. Volatility estimation based on local order statistics
3. **Law of local minima and upper bound**
4. Lower bound
5. Application

The law of bin-wise minima



10^5 bins, $\sigma = 1$, nh_n=100, $\varepsilon_i \sim \text{Exp}(5)$

Rewrite $m_{n,k} - m_{n,k-1} = \mathcal{R}_{n,k} - \mathcal{L}_{n,k}$, where $\mathcal{R}_{n,k} = m_{n,k} - X_{kh_n}$, $\mathcal{L}_{n,k} = m_{n,k-1} - X_{kh_n}$. For $X_t = X_{kh_n} + \sigma \int_{kh_n}^t dW_s$, invoke time-reversibility to see that $\mathcal{R}_{n,k}, \mathcal{L}_{n,k}$, $k = (l-1)r_n^{-1}+1, \ldots, lr_n^{-1}$, are all identically distributed and independent, except $\mathcal{R}_{n,k}$ and $\mathcal{L}_{n,k+1}$. Infer $\Psi(\sigma_{kh_n}^2)h_n = 2\,\mathbb{V}\text{ar}(\mathcal{R}_{n,k})$, and similarly for Ψ_n.

Connection to Brownian excursion areas

Proposition *Choose* $h_n = \mathcal{K}^{2/3}(n\lambda)^{-2/3}$. *Consider* $t \in \mathcal{T}_k^n$ *and* $X_t = X_{kh_n} + \int_{kh_n}^t \sigma\, dW_s, t \in \mathcal{T}_k^n$. *Then in the PPP-model for all* $x \in \mathbb{R}$:

$$\mathbb{P}\Big(h_n^{-1/2}\mathcal{R}_{n,k} > x\sigma\Big) = \mathbb{E}\Big[\exp\Big(-\mathcal{K}\sigma\int_0^1 (x+W_t)_+ dt\Big)\Big].$$

In the regression-type model for all $x \in \mathbb{R}$:

$$\lim_{n\to\infty}\mathbb{P}\Big(h_n^{-1/2}\mathcal{R}_{n,k} > x\sigma\Big) = \mathbb{E}\Big[\exp\Big(-\mathcal{K}\sigma\int_0^1 (x+W_t)_+\, dt\Big)\Big].$$

Laplace transforms of such expressions can be obtained via *Feynman–Kac formula*. It draws a connection between measures on the path space to parabolic PDEs as the transitional probability p for Brownian motion obeys

$$\partial_t p(t,x;s,y) = \tfrac{1}{2}\partial_x^2 p(t,x;s,y)\,,$$
$$-\partial_s p(t,x;s,y) = \tfrac{1}{2}\partial_y^2 p(t,x;s,y)\,.$$

Result with Feynman–Kac formula

The Laplace transform (in t) of

$$\mathbb{E}\Big[\exp\Big(-\sqrt{2}\theta\int_0^t (W_s+x)_+\,ds\Big)\Big],\ \theta\in\mathbb{R},$$

is derived as solution of

$$\frac{d^2\zeta}{dx^2}=2s\zeta-2\theta^{2/3},x<0,\ \frac{d^2\zeta}{dx^2}=2(\sqrt{2}\theta x+s)\zeta-2\theta^{2/3},x>0.$$

With the *Airy function* Ai and *Scorer function* Gi

$$\mathrm{Ai}(x)=\pi^{-1}\int_0^\infty \cos\left(t^3/3+xt\right)dt,\ \mathrm{Gi}(x)=\pi^{-1}\int_0^\infty \sin\left(t^3/3+xt\right)dt\ :$$

$$\mathbb{E}\left[\int_0^\infty \exp\left(-st-\sqrt{2}\theta\int_0^t (W_s+x)_+\,ds\right)dt\right]=\theta^{-\frac{2}{3}}\zeta_s(x,\theta),$$

$$\zeta_{s,+}(x,\theta)=A_s\,\mathrm{Ai}\left(\sqrt{2}\theta^{\frac{1}{3}}x+\theta^{-\frac{2}{3}}s\right)+\pi\mathrm{Gi}\left(\sqrt{2}\theta^{\frac{1}{3}}x+\theta^{-\frac{2}{3}}s\right)$$

$$\zeta_{s,-}(x,\theta)=B_s\exp\left(\sqrt{2s}x\right)+s^{-1}\theta^{2/3}.$$

Convergence rate of the estimator

Assumption *The drift a_s is bounded and Borel-measurable, the volatility σ_t is a Lipschitz function, $\sigma_t > 0$. The constant $\mathcal{K}$ of h_n is chosen sufficiently large.*

Theorem *For $h_n = \mathcal{K}^{2/3}(n\lambda)^{-2/3}$ and $r_n = \kappa n^{-1/3}$, $\kappa > 0$, the estimator based on the PPP-model satisfies*

$$\left(\widetilde{IV}_n^{h_n,r_n} - \int_0^1 \sigma_s^2\,ds\right) = \mathcal{O}_{\mathbb{P}}\left(n^{-\frac{1}{3}}\right).$$

Corollary *The estimator based on the regression-type model satisfies*

$$\left(\widehat{IV}_n^{h_n,r_n} - \int_0^1 \sigma_s^2\,ds\right) = \mathcal{O}_{\mathbb{P}}\left(n^{-\frac{1}{3}}\right).$$

Outline

The lower bound for the minimax rate

We show for the PPP-model that in the parametric experiment, $X_t = \sigma\, dW_t$, $t \in [0,1]$, $\sigma > 0$ unknown, the optimal rate of convergence is $n^{-1/3}$ in minimax sense. This serves a fortiori as a lower bound for the general nonparametric case.

Theorem

For any sequence of estimators $\hat{\sigma}_n^2$ of $\sigma^2 \in (0,\infty)$ from the parametric PPP-model for each $\sigma_0^2 > 0$, the local minimax lower bound is

$$\exists \delta > 0: \liminf_{n\to\infty} \inf_{\hat{\sigma}_n} \max_{\sigma^2 \in \{\sigma_0^2, \sigma_0^2 + \delta n^{-1/3}\}} \mathbb{P}_{\sigma^2}\left(|\hat{\sigma}_n^2 - \sigma^2| \geq \delta n^{-1/3}\right) > 0,$$

where the infimum extends over all estimators $\hat{\sigma}_n$ based on the PPP-model with $\lambda = 1$ and $X_t = \sigma W_t$. The law of the latter is denoted by $\mathbb{P}_{\sigma^2}$.

Sketch of information-theoretic proof

First reduction: Decompose in sum of two independent PPPs:

$$\mathrm{PPP}_{\Lambda} = \mathrm{PPP}_{\Lambda_r} + \mathrm{PPP}_{\Lambda_s} \text{ with } \Lambda = \Lambda_r + \Lambda_s\,,$$
$$\lambda_r(t,y) = n\Big(\big(b^{-1}(y - X_t)_+\big)^2 \wedge 1\Big), b > 0, \lambda_s = \lambda - \lambda_r\,.$$

Provide more information by $(T_j^s, X_{T_j^s})_{j\geq 1}$ instead of $(T_j^s, Y_j^s)_{j\geq 1}$.
Second reduction: Conditional on (T_j^s) observations $(T_i^r, \mathcal{Y}_i^r)$, $T_i^r \in [T_{j-1}^s, T_j^s)$ form for each j independent PPPs on $[0, T_j^s - T_{j-1}^s]$ with intensities

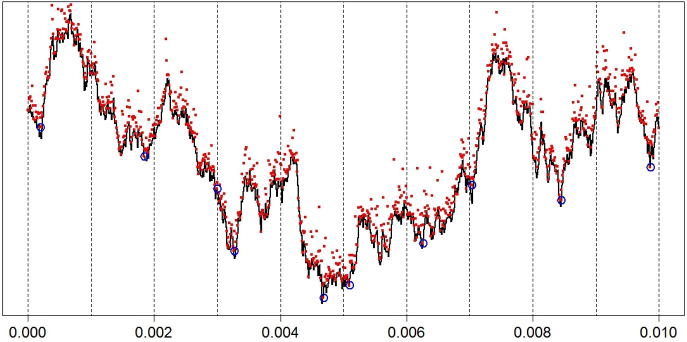
$$\lambda^j(t,y) = n\Big(b^{-1}\Big(y - \sigma B_t^{0, T_j^s - T_{j-1}^s}\Big)_+ \wedge 1\Big)\,,$$

with a Brownian bridge denoted by $B^{0,T}$.
For this more informative experiment standard bounds for the Hellinger distance imply the Theorem, $b \propto n^{-1/3}$.

Outline

Example for one realised path

Regression model, $\sigma = 1$, exponential errors with $\lambda^{-1} = 0.005$, $n = 100000$, $h_n = 0.001$, $nh_n = 100$.



First order approximation of Ψ

When $h_n = \mathcal{K}^{2/3}\lambda^{-2/3}n^{-2/3}$ and $\lambda^{-1} \ll \sigma_t$ for all t, or $\mathcal{K}$ large, the local minima are predominantly determined by $\min_{i\in\mathcal{I}_k^n} X_{t_i^n}$.
High signal-to-noise ratios found for high-frequency data in empirical studies.
Joint law of end-point of $Z = \int \sigma dW$ and its minimum:

$$\mathbb{P}\Big(\min_{0\le s\le t} Z_s < m, Z_t \ge w\Big) = \int_{-\infty}^{2m-w} (2\pi\sigma^2)^{-1/2}\exp\{-1/(2\sigma^2)z^2\}dz\,,$$

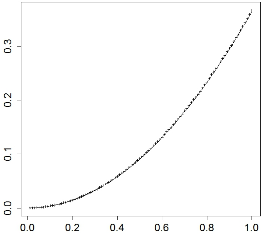
$$g(m,w) = \tfrac{2(w-2m)}{\sigma^3\sqrt{2\pi}}\exp\{-1/(2\sigma^2)(2m-w)^2\}, m\in(-\infty,0], w\in[m,\infty)\,.$$

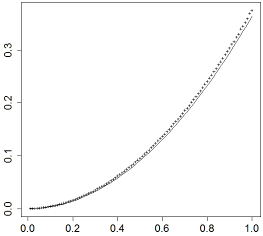
This yields for $k = \lfloor th_n^{-1}\rfloor$, $m_{n,k} = \min_{i\in\mathcal{I}_k^n} X_{t_i^n}$:

$$\lim_{n\to\infty} h_n^{-1/2}\mathbb{E}[\mathcal{L}_{n,k}] = -\sqrt{(2/\pi)}\sigma_t\,, \lim_{n\to\infty} h_n^{-1}\mathbb{E}[\mathcal{L}_{n,k}^2] = \sigma_t^2\,,$$

$$\lim_{n\to\infty} h_n^{-1}\mathbb{E}[\mathcal{L}_{n,k+1}\mathcal{R}_{n,k}] = \tfrac{1}{2}\sigma_t^2, \Rightarrow \Psi(\sigma_t^2) = 2\sigma_t^2\frac{(\pi-2)}{\pi}\,.$$

Accuracy of first-order approximation





Comparison of $\sum_{k=1}^{h_n^{-1}/2}(m_{n,2k}-m_{n,2k-1})^2$ and $\sigma^2(\pi-2)/\pi$ for $\sigma\in[0,1]$, $n=100000, nh_n=100$; $\lambda^{-1}=0.005$ (top) and $\lambda^{-1}=0.05$ (bottom).

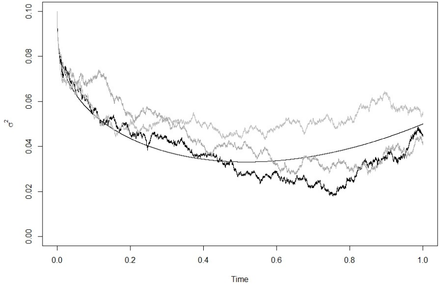
Conclude simple estimator

$$\widehat{IV}_{n,app}^{h_n}=\frac{\pi}{\pi-2}\sum_{k=1}^{h_n^{-1}/2}(m_{n,2k}-m_{n,2k-1})^2.$$

Monte Carlo simulations: Setup

Simulate $Y_i = X_{i/n} + \varepsilon_i, i = 0, \ldots, n,$.

1. $\sigma_t^2 = 0.1\left(1 - 0.4\sin\left(\frac{3}{4}\pi t\right)\right),\; t \in [0,1]$, drift $a = 0.1$.
2. $\sigma_t^2 = \left(\int_0^t c \cdot \rho\, dW_s + \int_0^t \sqrt{1-\rho^2} \cdot c\, dW_s^{\perp}\right) \cdot \tilde{\sigma}_t,$
   $W^{\perp}$ Brownian motion independent of W, $c = 0.05$, $\rho = 0.5$, $a = 0.1$ and seasonality function $\tilde{\sigma}_t = 0.1\left(1 - t^{\frac{1}{3}} + 0.5 \cdot t^2\right)$.



Monte Carlo simulations: Results

$n = 100000$		$\varepsilon_i \overset{iid}{\sim} \text{Exp}(\lambda = 2000)$			$\varepsilon_i \overset{iid}{\sim} \text{Exp}(\lambda = 200)$		
h_n^{-1}	nh_n	Bias $n^{1/3}$	Var $n^{2/3}$	Bias^2%MSE	Bias $n^{1/3}$	Var $n^{2/3}$	Bias^2%MSE
500	200	0.0015	0.1031	0.00	0.0332	0.1171	0.93
1000	100	0.0068	0.0542	0.08	0.0407	0.0518	3.10
10000	10	0.1393	0.0060	76.40	0.8688	0.0085	98.88
$n = 10000$		$\varepsilon_i \overset{iid}{\sim} \text{Exp}(\lambda = 2000)$			$\varepsilon_i \overset{iid}{\sim} \text{Exp}(\lambda = 200)$		
h_n^{-1}	nh_n	Bias $n^{1/3}$	Var $n^{2/3}$	Bias^2%MSE	Bias $n^{1/3}$	Var $n^{2/3}$	Bias^2%MSE
100	100	0.0020	0.1227	0.00	0.0179	0.1200	0.26
500	20	0.0183	0.0219	1.35	0.0555	0.0239	11.43
1000	10	0.0392	0.0116	11.75	0.1401	0.0138	58.72

Results for $\hat{IV}_{n,app}^{h_n}$ in setup 1. Bias rescaled with $n^{1/3}$, variance with factor $n^{2/3}$. Based on *first order approximation*.
Results for setup 2 equally well, $\hat{IV}_{n,app}^{h_n}$ comes close but does not attain minimal MSE because of the bias.

Monte Carlo simulations: Results

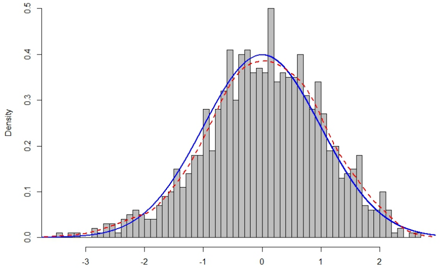
$n = 100000$		$\varepsilon_i \overset{iid}{\sim} \text{Exp}(\lambda = 2000)$			$\varepsilon_i \overset{iid}{\sim} \text{Exp}(\lambda = 200)$		
h_n^{-1}	r_n^{-1}	Bias $n^{1/3}$	Var $n^{2/3}$	Bias^2%MSE	Bias $n^{1/3}$	Var $n^{2/3}$	Bias^2%MSE
500	100	-0.0069	0.1051	0.04	-0.0194	0.1049	0.36
1000	100	0.0095	0.0528	0.17	-0.0148	0.0571	0.38
10000	100	0.0064	0.0056	0.72	0.0066	0.0077	0.56
$n = 10000$		$\varepsilon_i \overset{iid}{\sim} \text{Exp}(\lambda = 2000)$			$\varepsilon_i \overset{iid}{\sim} \text{Exp}(\lambda = 200)$		
h_n^{-1}	r_n^{-1}	Bias $n^{1/3}$	Var $n^{2/3}$	Bias^2%MSE	Bias $n^{1/3}$	Var $n^{2/3}$	Bias^2%MSE
100	20	-0.0078	0.1168	0.05	0.0029	0.1199	0.00
500	100	-0.0136	0.0237	0.77	-0.0008	0.0257	0.00
1000	100	0.0086	0.0125	0.59	0.0109	0.0135	0.87

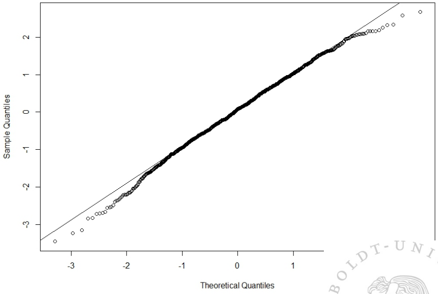
Results for $\hat{IV}_n^{h_n,r_n}$ in setup 1. Bias rescaled with $n^{1/3}$, variance with factor $n^{2/3}$. Results for setup 2 equally well, also different error distributions considered in report. Above choice of r_n slightly better than $r_n = \sqrt{h_n}$.

Monte Carlo approximation of Ψ_n employed.

Monte Carlo empirical distribution

Parametric setup $\sigma = 1$. Empirical distribution of $(\widehat{\mathbb{V}\mathrm{ar}}(\hat{\sigma}_n^2))^{-1/2}\,\hat{\sigma}_n^2$ for 1000 iterations, , $\lambda^{-1} = 0.005$, $n = 100000, h_n = 1000$.





Summary & Outlook

Summary:

- *Improved* volatility estimation based on order prices.
- We obtain $n^{1/3}$ as *optimal convergence rate*.
- First applications promising.
- First data applications using orders and traded prices support the idea of the *same latent efficient price* and its volatility recovered with methods based on microstructure noise model with centred and one-sided errors, respectively.

Summary & Outlook

Outlook:

- Strive for stable CLT in model where the volatility is a semi-martingale.
- Design an explicit feasible estimator.
- Application: Validate model using all available information from bids, asks, trades.

Literature

Bibinger, M., Jirak, M., Reiß, M., (2014).
Improved volatility estimation based on limit order books.
arXiv:1408.3768

Bibinger, M., Jirak, M., Reiß, M., (2014).
Applying volatility estimators based on limit order books.
technical report, available under `www.mathematik.hu-berlin.de/for1735/Publ/application.pdf`.

Thank you for your attention!