1 Notation

- 1-second last mid price was observed for 1-month
- Daily volatility is being estimated
- Sampling period : K (60 * 60 * 24 = 86400)
- Sequence length : n (86400 * 30 = 2592000)
- time : $0 < t < n$
- log-price : x_t
- log-return : r_t

2 Jump adjust

Removes jumps and stitch the log-pric

2.1 Truncation

Sub-Gaussian concentration condition was used to determine threshold parameter τ for each day.

$$\tau = \sigma\sqrt{K} \sim \sqrt{\sum_{t}^{t+K} r_t^2}$$

Truncation function with parameter τ was applied to the log return.

$$\Psi_\tau(x) = \operatorname{sgn}(x)\min(|x|, \tau)$$

2.2 Truncation with MLE

Huber loss of truncated series was minimized with L1 regularization, using daily volatility as prior.

- Daily volatility : $\sigma_t^2 = \frac{1}{n-K}\sum_{t=1}^{n-K}\left(\frac{1}{K}\sum_{i=t}^{t+K} r_i\right)^2$
- Huber loss : $l_{\tau'}(x) = \operatorname{sgn}(x)\min(\frac{x^2}{2}, \tau'|x| - \frac{\tau'^2}{2})$
  with $\tau' = \sigma\sqrt{K} \sim \sqrt{\sum_{t}^{t+K} r_t^2}$
- Truncation : $\Psi_\tau(x) = \operatorname{sgn}(x)\min(|x|, \tau)$
- Optimal truncation threshold $\hat{\tau} = \arg\min\left(\sum_{t}^{t_K} \tau l_{\tau'}(\sigma_t - \Psi(r_t))\right)$

2.3 Benchmark

Does not remove jump noise for benchmark purpose

3 Microstructure noise adjust

Removes high-frequency component

3.1 Pre-Averaging

Hourly rolling average

3.2 Fracdiff

Ensures stationarity while minimizng information loss.

3.3 Padé transform

Models jumps as Lorentz peak of oscillator resonance.

4 Volatility estimation

Estimate volatility, for given raw log-price and denoised log-price.

4.1 RV

4.2 PRV

4.3 TRV

5 Evaluation

Compare methods based on information measure

5.1 KLD of Conditional distribution

Measures the amount of information in time Filtration $\mathcal{F}_t$.

$$\mathrm{KL}(P|P_0)$$

where $P = P(\sigma_t|\sigma_t - 1) = P(\sigma_t|\mathcal{F}_t)P(\sigma_{t-1})$ and $P_0 = P(\sigma_t)P(\sigma_t) = P(\sigma_t)P(\sigma_{t-1})$. P is estimated using kernel density estimation.

5.2 Self-similarity dimension

. Fractal dimension is used to measure information amount, using box-counting method.

5.3 Minkowski Measure

For given fractal dimension, measures persistency.