Optimally Thresholded Realized Power Variations for Stochastic Volatility Models with Jumps

José E. Figueroa-López\(^1\)

\(^1\)Department of Mathematics
Washington University

ISI 2015: 60th World Statistics Conference
Rio de Janeiro
July 29, 2015

(Joint work with Jeff Nisen, Cheng Li)
Framework

1. Finite-Jump Activity (FJA) Itô Semimartingales:

\[ dX_t = \gamma_t \, dt + \sigma_t \, dW_t + dJ_t \]

- \( t \to W_t \) is a standard Brownian motion;
- \( t \to J_t := \sum_{j=1}^{N_t} \zeta_j \):
  - \( t \to N_t \) is the counting process of jumps s.t. \( N_t < \infty \), for all \( t > 0 \)
  - \( \{\zeta_j\}_j \) are the jump sizes;
- \( t \to \gamma_t \) and \( t \to \sigma_t \) are the drift and volatility functions;
Framework

1. **Finite-Jump Activity (FJA) Itô Semimartingales:**
   \[ dX_t = \gamma_t dt + \sigma_t dW_t + dJ_t \]
   - \( t \to W_t \) is a standard Brownian motion;
   - \( t \to J_t := \sum_{j=1}^{N_t} \zeta_j \):
     \( t \to N_t \) is the counting process of jumps s.t. \( N_t < \infty \), for all \( t > 0 \)
     \( \{\zeta_j\}_j \) are the jump sizes;
   - \( t \to \gamma_t \) and \( t \to \sigma_t \) are the drift and volatility functions;

2. **FJA Lévy Model:**
   \[ X_t = \gamma t + \sigma W_t + \sum_{j=1}^{N_t} \zeta_j \]
   - \( \{N_t\}_{t \geq 0} \) is a homogeneous Poisson process with jump intensity \( \lambda \);
   - \( \{\zeta_j\}_{j \geq 0} \) are i.i.d. with density \( f_\zeta : \mathbb{R} \to \mathbb{R}_+ \);
   - the triplet \( \{W_t\}, \{N_t\}, \{\zeta_j\} \) are mutually independent.
Statistical Problems

Given a finite discrete record of observations,

\[ X_{t_0}, X_{t_1}, \ldots, X_{t_n}, \quad \pi : 0 = t_0 < t_1 < \cdots < t_n = T, \]

the following problems are of interest under a high-frequency sampling setting (i.e., \( \text{mesh}(\pi) := \max_i \{t_i - t_{i-1}\} \rightarrow 0 \)):

1. Estimating the integrated variance:
   \[
   \bar{\sigma}^2_T := \int_0^T \sigma_t^2 dt.
   \]

2. Estimating the jump features of the process:
   - Jump times: say, \( \{\tau_1 < \tau_2 < \cdots < \tau_{N_T}\} \) if \( N_T \geq 1 \), or \( \emptyset \), otherwise.
   - Corresponding jump sizes: \( \{\zeta_1, \zeta_2, \ldots, \zeta_{N_T}\} \) if \( N_T \geq 1 \), or \( \emptyset \), otherwise.
Two main classes of estimators

1. Multipower Realized Variations (Barndorff-Nielsen and Shephard (2004)):

$$BPV(X)_T := \sum_{i=0}^{n-1} |X_{t_{i+1}} - X_{t_i}| |X_{t_{i+2}} - X_{t_{i+1}}|,$$

$$MPV(X)_T := \sum_{i=0}^{n-k} |X_{t_{i+1}} - X_{t_i}|^{r_1} \ldots |X_{t_{i+k}} - X_{t_{i+k-1}}|^{r_k}, \quad (r_1 + \cdots + r_k = 2).$$


$$TRV(X)[B]_T^\pi := \sum_{i=0}^{n-1} (X_{t_{i+1}} - X_{t_i})^2 1\{|X_{t_{i+1}} - X_{t_i}| \leq B\}, \quad (B \in (0, \infty)).$$
Advantages and Drawbacks of TRV

1. **Pros:**

   - Can exhibit reduced bias for estimating $\bar{\sigma}_T^2$, in the presence of jumps
   - Can be adapted for estimating the process’ jump features:

     \[
     \hat{N}^\pi_B := \sum_{i=0}^{n-1} 1 \{ |X_{t_{i+1}} - X_{t_i}| > B \} \xrightarrow{\text{mesh} \to 0} N_T.
     \]
Advantages and Drawbacks of TRV

1. **Pros:**
   - Can exhibit reduced bias for estimating $\bar{\sigma}_T^2$, in the presence of jumps
   - Can be adapted for estimating the process’ jump features:

   \[ \hat{N}[B]_\pi T := \sum_{i=0}^{n-1} \mathbf{1}_{\{|X_{t_{i+1}} - X_{t_i}| > B\}} \xrightarrow{\text{mesh} \to 0} N_T. \]

2. **Cons:**
   Performance strongly depends on a “good” selection of the threshold $B$;
Advantages and Drawbacks of TRV

① Pros:
• Can exhibit reduced bias for estimating $\bar{\sigma}_T^2$, in the presence of jumps
• Can be adapted for estimating the process’ jump features:

\[ \hat{N}[B]_T^\pi := \sum_{i=0}^{n-1} 1 \left\{ |x_{ti+1} - x_{ti}| > B \right\} \xrightarrow{\text{mesh} \to 0} N_T. \]

② Cons:
Performance strongly depends on a “good” selection of the threshold $B$;
e.g., for a FJA Lévy model and regular sampling ($t_i = ih_n$ with $h_n = T/n$),
F-L & Nisen (SPA, 2013):

(i) $TRV(X)[B_n]_T^\pi$ is consistent for $\bar{\sigma}_T^2$ iff
\[ \frac{B_n}{\sqrt{h_n}} \xrightarrow{n \to \infty} \infty \quad \text{and} \quad B_n \to 0; \]
Advantages and Drawbacks of TRV

① Pros:
- Can exhibit reduced bias for estimating \( \bar{\sigma}_T^2 \), in the presence of jumps
- Can be adapted for estimating the process’ jump features:

\[
\hat{N}[B]_T^\pi := \sum_{i=0}^{n-1} \mathbf{1}\{|x_{t_{i+1}} - x_{t_i}| > B\} \xrightarrow{\text{mesh} \to 0} N_T.
\]

② Cons:
Performance strongly depends on a “good” selection of the threshold \( B \);
e.g., for a FJA Lévy model and regular sampling \((t_i = ih_n \text{ with } h_n = T/n)\),
F-L & Nisen (SPA, 2013):

(i) \( TRV(X)[B_n]_T^\pi \) is consistent for \( \bar{\sigma}_T^2 \) iff \( \frac{B_n}{\sqrt{h_n}} \xrightarrow{n \to \infty} \infty \) and \( B_n \to 0 \);

(ii) \[
\mathbb{E}\left[ TRV(X)[B_n]_T^\pi \right] - \bar{\sigma}_T^2 \sim Th_n \left( \gamma^2 - \lambda \sigma^2 \right) - 2T\sigma \phi\left( \frac{B_n}{\sigma \sqrt{h_n}} \right) \frac{B_n}{\sqrt{h_n}} + \frac{2T\lambda B_n^3 C(f_c)}{3}
\]
Popular thresholds $B$

- **Power Threshold (Mancini (2003))**

$$B_{\alpha,\omega}^{\text{Pow}} := \alpha \text{mesh}(\pi)^{\omega}, \quad \text{for } \alpha > 0 \text{ and } \omega \in (0, 1/2).$$

- **Bonferroni Threshold (Bollerslev et al. (2007) and Gegler & Stadtmüller (2010))**

$$B_{\tilde{\sigma},C}^{\text{BF}} := \tilde{\sigma} \text{mesh}(\pi)^{1/2} \Phi^{-1} \left( 1 - \frac{C \text{mesh}(\pi)}{2} \right), \quad \text{for } C > 0 \text{ and } \tilde{\sigma} > 0.$$
Optimally Thresholded Realized Variations

Aims

1. Introduce an "optimal" selection criterion for the threshold $B$, that minimizes a suitable loss function of estimation.
2. Develop a feasible implementation method for the optimal threshold $B^*$.

Loss Function

$$Loss_n(B) := \mathbb{E} \sum_{i=1}^{n} \left( \begin{array}{c} \mathbb{1} \left( \left| \Delta_n^i X \right| \leq B, \Delta_n^i N \neq 0 \right) \\ \mathbb{1} \left( \left| \Delta_n^i X \right| > B, \Delta_n^i N = 0 \right) \end{array} \right),$$

where, as usual, $\Delta_n^i X := X_{t_i} - X_{t_{i-1}}$ and $\Delta_n^i N := N_{t_i} - N_{t_{i-1}}$.

Interpretation

$Loss_n(B)$ represents the Total Number of Jump Miss-Classifications:

- fail to identify the occurrence of a jump during $[t_{i-1}, t_i]$.
- to flag that a jump occurred during $[t_{i-1}, t_i)$, when no jump occurred.
Optimally Thresholded Realized Variations

1. Aims
   - Introduce an “optimal" selection criterion for the threshold $B$, that minimizes a suitable loss function of estimation.
   - Develop a feasible implementation method for the optimal threshold $B^\ast$.

2. Loss Function

   $$\text{Loss}_n(B) := \mathbb{E} \sum_{i=1}^{n} \left( 1_{[|\Delta_i^n X| \leq B, \Delta_i^n N \neq 0]} + 1_{[|\Delta_i^n X| > B, \Delta_i^n N = 0]} \right),$$

   where, as usual, $\Delta_i^n X := X_{t_i} - X_{t_{i-1}}$ and $\Delta_i^n N := N_{t_i} - N_{t_{i-1}}$.

3. Interpretation

   $\text{Loss}_n(B)$ represents the Total Number of Jump Miss-Classifications:
   - fail to identify the occurrence of a jump during $[t_{i-1}, t_i]$.
   - to flag that a jump occurred during $[t_{i-1}, t_i)$, when no jump occurred,
Existence and Infill Asymptotic Characterization

Theorem (FL & Nisen (SPA, 2013))

Suppose that $t_{i+1} - t_i =: h_n = T/n$ for any $i$ and $X$ is a Finite Jump Activity Lévy Model with jump density $f_\zeta$, jump intensity $\lambda$, and volatility $\sigma$:
Existence and Infill Asymptotic Characterization

Theorem (FL & Nisen (SPA, 2013))

Suppose that \( t_{i+1} - t_i =: h_n = T / n \) for any \( i \) and \( X \) is a Finite Jump Activity Lévy Model with jump density \( f_\zeta \), jump intensity \( \lambda \), and volatility \( \sigma \):

1. For \( n \) large enough, the loss function \( \text{Loss}_n(B) \) is quasi-convex and, moreover, possesses a unique global minimum \( B_n^* \).

\[
B_n^* = \sqrt{\frac{3}{\pi}} \frac{h_n}{\sqrt{\frac{1}{h_n} \log \left( \frac{1}{h_n} \right)} - \log \left( \frac{\sqrt{2\pi} \lambda}{C(f_\zeta)} \right) \frac{\sigma h_n}{2 n \sqrt{3 \log \left( \frac{1}{h_n} \right)}}} + o(\sqrt{h_n \log \left( \frac{1}{h_n} \right)})
\]
Existence and Infill Asymptotic Characterization

Theorem (FL & Nisen (SPA, 2013))

Suppose that \( t_{i+1} - t_i =: h_n = T / n \) for any \( i \) and \( X \) is a Finite Jump Activity Lévy Model with jump density \( f_\zeta \), jump intensity \( \lambda \), and volatility \( \sigma \):

1. For \( n \) large enough, the loss function \( \text{Loss}_n(B) \) is quasi-convex and, moreover, possesses a unique global minimum \( B^*_n \).

2. As \( n \to \infty \), the optimal threshold sequence \( (B^*_n)_n \) is such that

\[
B^*_n = \sqrt{\frac{3\sigma^2 h_n \log \left( \frac{1}{h_n} \right)}{\sqrt{3 \log(1/h_n)}}} - \frac{\log \left( \sqrt{2\pi \sigma \lambda C(f_\zeta)} \right) \sigma h_n^{1/2}}{\sqrt{3 \log(1/h_n)}} + o \left( \sqrt{\frac{h_n}{\log(1/h_n)}} \right),
\]

where \( C(f_\zeta) = \lim_{\epsilon \to 0} \frac{1}{2\epsilon} \int_{-\epsilon}^{\epsilon} f_\zeta(x) dx \).
Remarks

1. The leading term of the optimal sequence is proportional to Lévy’s modulus of continuity of the Brownian motion:

\[
\limsup_{h \to 0} \frac{1}{\sqrt{2h \log(1/h)}} \sup_{|t-s| < h, s, t \in [0,1]} |W_t - W_s| = 1, \quad \text{a.s.}
\]
Remarks

1. The leading term of the optimal sequence is proportional to Lévy’s modulus of continuity of the Brownian motion:

   \[
   \limsup_{h \to 0} \frac{1}{\sqrt{2h \log(1/h)}} \sup_{|t-s|<h, s,t \in [0,1]} |W_t - W_s| = 1, \quad \text{a.s.}
   \]

2. The threshold sequences

   \[
   B_{n}^{*1} := \sqrt{3\sigma^2 h_n \log \left( \frac{1}{h_n} \right)}, \quad B_{n}^{*2} := B_{n}^{*1} - \frac{\log \left( \sqrt{2\pi} \sigma \lambda C(f_{\zeta}) \right) \sigma h_n^{1/2}}{\sqrt{3 \log(1/h_n)}},
   \]

   are the first and second-order approximations for \( B_{n}^{*} \), and it can be shown that the biases of their corresponding TRV estimators attain the “optimal” rate of \( O(h_n) \) as \( n \to \infty \).
Remarks

1. The leading term of the optimal sequence is proportional to Lévy’s modulus of continuity of the Brownian motion:

\[
\limsup_{h \to 0} \frac{1}{\sqrt{2h \log(1/h)}} \sup_{|t-s| < h, s, t \in [0,1]} |W_t - W_s| = 1, \quad \text{a.s.}
\]

2. The threshold sequences

\[
B^*_n := \sqrt{3\sigma^2 h_n \log \left( \frac{1}{h_n} \right)}, \quad B^*_n := B^*_n - \frac{\log \left( \sqrt{2\pi \sigma \lambda C(f_{\zeta})} \right) \sigma h_n^{1/2}}{\sqrt{3 \log(1/h_n)}},
\]

are the first and second-order approximations for \( B^*_n \), and it can be shown that the biases of their corresponding TRV estimators attain the “optimal” rate of \( O(h_n) \) as \( n \to \infty \).

3. They both provide “blueprints” for devising good threshold sequences!
A Feasible Implementation Algorithm Based on $B_n^{*1}$

Key Issue: The threshold $B^{*1}$ would allow us to find an (approximately) "optimal" estimate $\hat{\sigma}^2$ for $\sigma^2$ of the form

$$\hat{\sigma}^2 := \frac{1}{T} TRV(X)[B^{*1}]_n$$
A Feasible Implementation Algorithm Based on $B_{n}^{*1}$

1. **Key Issue**: The threshold $B_{n}^{*1}$ would allow us to find an (approximately) “optimal” estimate $\hat{\sigma}^2$ for $\sigma^2$ of the form

$$\hat{\sigma}^2 := \frac{1}{T} TRV(X)[B_{n}^{*1}]_n$$

but $B_{n}^{*1} := B_{n}^{*1}(\sigma^2)$ depends precisely on $\sigma^2$. 

---

Figueroa-López, José (Purdue, Statistics)
A Feasible Implementation Algorithm Based on $B_{n}^{*1}$

1. **Key Issue**: The threshold $B_{n}^{*1}$ would allow us to find an (approximately) “optimal” estimate $\hat{\sigma}^2$ for $\sigma^2$ of the form

$$\hat{\sigma}^2 := \frac{1}{T} TRV(X)[B_{n}^{*1}]_n$$

but $B_{n}^{*1} := B_{n}^{*1}(\sigma^2)$ depends precisely on $\sigma^2$.

2. The previous issue suggests a “fixed-point" type of implementation:
A Feasible Implementation Algorithm Based on $B^*_n$\(^1\)

① **Key Issue**: The threshold $B^*_1$ would allow us to find an (approximately) “optimal” estimate $\hat{\sigma}^2$ for $\sigma^2$ of the form

$$\hat{\sigma}^2 := \frac{1}{T} TRV(X)[B^*_1]_n$$

but $B^*_1 := B^*_1(\sigma^2)$ depends precisely on $\sigma^2$.

② The previous issue suggests a “fixed-point” type of implementation:

(i) Get a “rough” estimate of $\sigma^2$ via, e.g., the QV:

$$\hat{\sigma}^2_{n,0} := \frac{1}{T} \text{QV}_T := \frac{1}{T} \sum_{i=1}^{n} |X_{t_i} - X_{t_{i-1}}|^2$$

(ii) Use $\hat{\sigma}^2_{n,0}$ to estimate the optimal threshold $\hat{B}^*_n,0 := (3 \hat{\sigma}^2_{n,0} h_n \log(1/h_n))^{1/2}$

(iii) Refine $\hat{\sigma}^2_{n,0}$ using thresholding,

$$\hat{\sigma}^2_{n,1} = \frac{1}{T} \sum_{i=1}^{n} |X_{t_i} - X_{t_{i-1}}|^2 1[|X_{t_i} - X_{t_{i-1}}| \leq \hat{B}^*_n,0]$$

(iv) Iterate Steps (ii) and (iii):

$$\hat{\sigma}^2_{n,0} \rightarrow \hat{B}^*_n,0 \rightarrow \hat{\sigma}^2_{n,1} \rightarrow \hat{B}^*_n,1 \rightarrow \hat{\sigma}^2_{n,2} \rightarrow \cdots$$
### A numerical illustration I

<table>
<thead>
<tr>
<th>Method</th>
<th>$\hat{B}_{n,k_n}^*$</th>
<th>$\hat{S}_{n,k_n}^*$</th>
<th>$\hat{\text{Loss}}$</th>
<th>$\hat{S}_{\text{Loss}}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\hat{B}_{n,k_n}^*$</td>
<td>0.5004</td>
<td>0.0186</td>
<td>0.2232</td>
<td>0.4706</td>
</tr>
<tr>
<td>$\text{Pow}_{\omega=0.495;\alpha=1}$</td>
<td>0.4407</td>
<td>0.0142</td>
<td>13.5302</td>
<td>3.6392</td>
</tr>
<tr>
<td>$BF$</td>
<td>0.4917</td>
<td>0.0193</td>
<td>1.180</td>
<td>1.0775</td>
</tr>
</tbody>
</table>

**Table:** Finite-sample performance of the threshold realized variation (TRV) estimators based on 5,000 sample paths of the Kou model:

$$f_{\text{Kou}}(x) = \frac{p}{\alpha_+} e^{-x/\alpha_+} 1_{[x \geq 0]} + \frac{(1-p)}{\alpha_-} e^{-|x|/\alpha_-} 1_{[x < 0]}.$$  

*Loss* represents the total number of Jump Misclassification Errors, while $\overline{TRV}$, $\overline{Loss}$, $S_{TRV}$, and $S_{Loss}$ denote the corresponding sample means and standard deviations, respectively.
A numerical illustration II

<table>
<thead>
<tr>
<th>Method</th>
<th>TRV</th>
<th>S_{TRV}</th>
<th>Loss</th>
<th>S_{Loss}</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\hat{B}^*_{n,k_n}$</td>
<td>0.4039</td>
<td>0.0028</td>
<td>139.6776</td>
<td>12.2193</td>
</tr>
<tr>
<td>Pow_{ω=0.495;α=1}</td>
<td>0.3767</td>
<td>0.0019</td>
<td>230.0170</td>
<td>15.0308</td>
</tr>
<tr>
<td>BF</td>
<td>0.6495</td>
<td>0.0315</td>
<td>375.5850</td>
<td>24.3999</td>
</tr>
</tbody>
</table>

Table: Finite-sample performance of the threshold realized variation (TRV) estimators based on $K = 5,000$ sample paths of the Kou model:

$$ f_\zeta(x) = \frac{p}{\alpha_+} e^{-x/\alpha_+} 1_{[x \geq 0]} + \frac{q}{\alpha_-} e^{-|x|/\alpha_-} 1_{[x < 0]}.$$

Loss represents the total number of Jump Misclassification Errors, while TRV, Loss, $S_{TRV}$, and $S_{Loss}$ denote the corresponding sample means and standard deviations, respectively.
Additive Processes and general sampling schemes

1. The model:

\[ X_s := \int_0^s \gamma(u)du + \int_0^s \sigma(u)dW_u + \sum_{j=1}^{N_s} \zeta_j =: X_s^c + J_s, \]

where \( \zeta_j \) i.i.d. \( f_\zeta \) and \( (N_s)_{s \geq 0} \sim \text{Poiss} (\{\lambda(s)\}_{s \geq 0}) \), independent of \( W \), and deterministic smooth functions \( \sigma, \lambda : [0, \infty) \to \mathbb{R}_+ \) and \( \gamma : [0, \infty) \to \mathbb{R} \) with \( \sigma \) and \( \lambda \) bounded away from 0.
Additive Processes and general sampling schemes

1. The model:

\[ X_s := \int_0^s \gamma(u)du + \int_0^s \sigma(u)dW_u + \sum_{j=1}^{N_s} \zeta_j =: X_s^c + J_s, \]

where \( \zeta_j \) i.i.d. \( f_\zeta \) and \( (N_s)_{s \geq 0} \sim \text{Poiss}(\{\lambda(s)\}_{s \geq 0}) \), independent of \( W \), and deterministic smooth functions \( \sigma, \lambda : [0, \infty) \to \mathbb{R}_+ \) and \( \gamma : [0, \infty) \to \mathbb{R} \) with \( \sigma \) and \( \lambda \) bounded away from 0.

2. Optimal Threshold Problem

Given a sampling scheme \( \pi : 0 = t_0 < \cdots < t_n = T \), determine the vector \( \bar{B}^{\pi, \ast} = (B_{t_1}^{\pi, \ast}, \ldots, B_{t_n}^{\pi, \ast}) \) that minimizes the problem

\[
\inf_{\bar{B}=(B_{t_1}, \ldots, B_{t_n}) \in \mathbb{R}_+^n} \mathbb{E} \sum_{i=1}^n \left( 1 \left[ |X_{t_i} - X_{t_{i-1}}| > B_{t_i}, N_{t_i} - N_{t_{i-1}} = 0 \right] + 1 \left[ |X_{t_i} - X_{t_{i-1}}| \leq B_{t_i}, N_{t_i} - N_{t_{i-1}} \neq 0 \right] \right)
\]
Additive Processes and general sampling schemes

1. The model:

\[ X_s := \int_0^s \gamma(u) \, du + \int_0^s \sigma(u) \, dW_u + \sum_{j=1}^{N_s} \zeta_j =: X_s^c + J_s, \]

where \( \zeta_j \overset{\text{i.i.d.}}{\sim} f_\zeta \) and \((N_s)_{s \geq 0} \sim \text{Poiss} \left( \{\lambda(s)\} \right)_{s \geq 0}\), independent of \( W \), and deterministic smooth functions \( \sigma, \lambda : [0, \infty) \to \mathbb{R}_+ \) and \( \gamma : [0, \infty) \to \mathbb{R} \) with \( \sigma \) and \( \lambda \) bounded away from 0.

2. Optimal Threshold Problem

Given a sampling scheme \( \pi : 0 = t_0 < \cdots < t_n = T \), determine the vector \( \bar{B}^{\pi,*} = (B_{t_1}^{\pi,*}, \ldots, B_{t_n}^{\pi,*}) \) that minimizes the problem

\[
\inf_{B=(B_{t_1},\ldots,B_{t_n}) \in \mathbb{R}^m_+} \mathbb{E} \sum_{i=1}^n \left( \mathbb{I}[|X_{t_i} - X_{t_{i-1}}| > B_{t_i}, N_{t_i} - N_{t_{i-1}} = 0] + \mathbb{I}[|X_{t_i} - X_{t_{i-1}}| \leq B_{t_i}, N_{t_i} - N_{t_{i-1}} \neq 0] \right)
\]

\[
= \sum_{i=1}^n \inf_{B_{t_i}} \{ \mathbb{P}(|\Delta_i X| > B_{t_i}, \Delta_i N = 0) + \mathbb{P}(|\Delta_i X| \leq B_{t_i}, \Delta_i N \neq 0) \},
\]
Well-posedness and Asymptotic Characterization

Theorem (FL & Nisen, 2014)

For any fixed $T > 0$, there exists $h_0 := h_0(T) > 0$ such that, for each $t \in [0, T]$ and $h \in (0, h_0]$, the function

$$L_{t,h}(B) := \mathbb{P}(|X_{t+h} - X_t| > B, N_{t+h} - N_t = 0) + \mathbb{P}(|X_{t+h} - X_t| \leq B, N_{t+h} - N_t \neq 0),$$

is quasi-convex and possesses a unique global minimum, $B_{t,h}^*$, such that, as $h \to 0$,

$$B_{t,h}^* = \sqrt{3 \sigma^2(t) h \log \left( \frac{1}{h} \right)} - \frac{\log(\sqrt{2\pi\sigma(t)\lambda_t C(f_\zeta)}\sigma(t)h^{1/2}}{\sqrt{3 \log(1/h)}} + o \left( \sqrt{\frac{h}{\log(1/h)}} \right).$$
Well-posedness and Asymptotic Characterization

Theorem (FL & Nisen, 2014)

For any fixed $T > 0$, there exists $h_0 := h_0(T) > 0$ such that, for each $t \in [0, T]$ and $h \in (0, h_0]$, the function

$$L_{t,h}(B) := \mathbb{P}(|X_{t+h} - X_t| > B, N_{t+h} - N_t = 0)$$

$$+ \mathbb{P}(|X_{t+h} - X_t| \leq B, N_{t+h} - N_t \neq 0),$$

is quasi-convex and possesses a unique global minimum, $B_{t,h}^*$, such that, as $h \to 0$,

$$B_{t,h}^* = \sqrt{3\sigma^2(t)h\log \left(\frac{1}{h}\right)} - \frac{\log(\sqrt{2\pi\sigma(t)\lambda_t C(f_\zeta)})\sigma(t)h^{1/2}}{\sqrt{3\log(1/h)}} + o \left(\sqrt{\frac{h}{\log(1/h)}}\right).$$

Again, the leading term $B_{t,h}^{*1} = \sqrt{3\sigma^2(t)h\log \left(\frac{1}{h}\right)}$ provides a blueprint to devise a good threshold parameter.
Spot Volatility Estimation via Kernel Methods

Notation: \( h_i = t_i - t_{i-1} \), \( \overline{K}_\theta(t) = \frac{1}{\theta} K \left( \frac{t}{\theta} \right) \) (Kernel), \( \theta = \text{Bandwidth} \)

Algorithm based on Kernel estimation (Fan & Wang(2008), Kristensen(2010))
Spot Volatility Estimation via Kernel Methods

Notation: \( h_i = t_i - t_{i-1} \), \( \overline{K}_\theta(t) = \frac{1}{\theta} K \left( \frac{t}{\theta} \right) \) (Kernel), \( \theta = \) Bandwidth

Algorithm based on Kernel estimation (Fan & Wang(2008), Kristensen(2010))

1. Get a “rough” estimate for \( t \rightarrow \sigma^2_t \) via the Kernel estimator:

\[
\hat{\sigma}^2_0(t_i) := \sum_{j=-\ell}^{\ell} |\triangle i+j X|^2 \overline{K}_\theta(t_i - t_{i+j}),
\]

with \( \theta \) chosen by a cross-validation type method (cf. Kristensen(2010)).
Spot Volatility Estimation via Kernel Methods

Notation: $h_i = t_i - t_{i-1}$, $\overline{K}_\theta(t) = \frac{1}{\theta} K \left( \frac{t}{\theta} \right)$ (Kernel), $\theta = \text{Bandwidth}$

Algorithm based on Kernel estimation (Fan & Wang(2008), Kristensen(2010))

1. Get a “rough” estimate for $t \rightarrow \sigma^2_t$ via the Kernel estimator:

$$\hat{\sigma}^2_0(t_i) := \sum_{j=-\ell}^{\ell} |\triangle_{i+j} X|^2 \overline{K}_\theta(t_i - t_{i+j}),$$

with $\theta$ chosen by a cross-validation type method (cf. Kristensen(2010)).

2. Get an initial estimate for the leading term of the optimal threshold:

$$\hat{B}^{*1}_0(t_i) := \sqrt{3 \hat{\sigma}^2_0(t_i) h_i \log (1/h_i)}$$
Spot Volatility Estimation via Kernel Methods

Notation: \( h_i = t_i - t_{i-1} \), \( \overline{K}_\theta(t) = \frac{1}{\theta} K \left( \frac{t}{\theta} \right) \) (Kernel), \( \theta = \) Bandwidth

Algorithm based on Kernel estimation (Fan & Wang(2008), Kristensen(2010))

1. Get a “rough” estimate for \( t \rightarrow \sigma^2_t \) via the Kernel estimator:

\[
\hat{\sigma}^2_0(t_i) := \sum_{j=-\ell}^{\ell} |\triangle_{i+j}X|^2 \overline{K}_\theta(t_i - t_{i+j}),
\]

with \( \theta \) chosen by a cross-validation type method (cf. Kristensen(2010)).

2. Get an initial estimate for the leading term of the optimal threshold:

\[
\hat{B}^*_1(t_i) := \sqrt{3\hat{\sigma}^2_0(t_i)h_i \log \left( \frac{1}{h_i} \right)}
\]

3. Refine the estimate \( \hat{\sigma}^2_0(t_i) \) using thresholding:

\[
\hat{\sigma}^2_1(t_i) := \sum_{j=-\ell}^{\ell} |\triangle_{i+j}X|^2 \overline{K}_\theta(t_i - t_{i+j}) 1[|\triangle_{i+j}X| \leq \hat{B}^*_1(t_{i+j})]
\]
Spot Volatility Estimation via Kernel Methods

Notation: \( h_i = t_i - t_{i-1} \), \( \overline{K}_\theta(t) = \frac{1}{\theta} K \left( \frac{t}{\theta} \right) \) (Kernel), \( \theta = \text{Bandwidth} \)

Algorithm based on Kernel estimation (Fan & Wang(2008), Kristensen(2010))

1. Get a “rough” estimate for \( t \rightarrow \sigma_i^2 \) via the Kernel estimator:

\[
\hat{\sigma}_0^2(t_i) := \sum_{j=-\ell}^{\ell} |\Delta_i + jX|^2 \overline{K}_\theta(t_i - t_{i+j}),
\]

with \( \theta \) chosen by a cross-validation type method (cf. Kristensen(2010)).

2. Get an initial estimate for the leading term of the optimal threshold:

\[
\hat{B}_{0}^{*1}(t_i) := \sqrt{3\hat{\sigma}_0^2(t_i)h_i \log (1/h_i)}
\]

3. Refine the estimate \( \hat{\sigma}_0^2(t_i) \) using thresholding:

\[
\hat{\sigma}_1^2(t_i) := \sum_{j=-\ell}^{\ell} |\Delta_i + jX|^2 \overline{K}_\theta(t_i - t_{i+j}) 1_{|\Delta_i + jX|\leq \hat{B}_{0}^{*1}(t_{i+j})}
\]

4. Iterate Steps 2 and 3:

\[
\hat{\sigma}_0^2(\cdot) \rightarrow \hat{B}_{0}^{*1}(\cdot) \rightarrow \hat{\sigma}_1^2(\cdot) \rightarrow \hat{B}_{1}^{*1}(\cdot) \rightarrow \hat{\sigma}_2^2(\cdot) \rightarrow \cdots
\]

Figure: Estimation of Spot Volatility using Adaptive Kernel Weighted Realized Volatility. (A) The initial estimates. (B) Intermediate estimates. Parameters: $\gamma(t) = 0.1t$, $\sigma(t) = 4.5t \sin(2\pi e^t)^2 + 0.2$, $\lambda(t) = 25(e^{3t} - 1)$, $\zeta_i \overset{i.i.d.}{\sim} \mathcal{N}(\mu = 0.025, \delta = 0.025)$.

(C) Terminal Estimates

(D) Estimation Variability

Figure: Estimation of Spot Volatility using Adaptive Kernel Weighted Realized Volatility.

(C) The terminal estimates. (D) Estimation variability, based on 100 generated sample paths, for the Quadratic Kernel based estimator. Parameters: \( \gamma(t) = 0.1t \), \( \sigma(t) = 4.5t \sin(2\pi e^{t^2})^2 + 0.2 \), \( \lambda(t) = 25(e^{3t} - 1) \), \( \zeta_i \overset{\text{i.i.d.}}{\sim} \mathcal{N}(\mu = 0.025, \delta = 0.025) \).
Stochastic Volatility Models with FJA

1 Motivation: In financial application, one usually encounters models like

\[ X_s := \int_0^s \gamma_u du + \int_0^s \sigma_u dW_u + \sum_{j=1}^{N_s} \zeta_j =: X_s^c + J_s, \]

where \( \{\sigma_t\}_{t \geq 0} \) itself is erratic and not smooth.

2 Prototypical Example: Mean-reverting square-root process (CIR Model):

\[ d\sigma_t^2 = \kappa (\alpha - \sigma_t^2) dt + \beta \sigma_t dW^{(\sigma)}_t, \quad (2\kappa \alpha - \beta^2 > 0, \ Cov(dW^{(\sigma)}_t, dW_t) = \rho dt) \]
Stochastic Volatility Models with FJA

1. **Motivation:** In financial application, one usually encounters models like

\[ X_s := \int_0^s \gamma_u du + \int_0^s \sigma_u dW_u + \sum_{j=1}^{N_s} \zeta_j =: X_s^c + J_s, \]

where \( \{\sigma_t\}_{t \geq 0} \) itself is erratic and not smooth.

2. **Prototypical Example:** Mean-reverting square-root process (CIR Model):

\[ d\sigma_t^2 = \kappa (\alpha - \sigma_t^2) \, dt + \beta \sigma_t \, dW_t^{(\sigma)}, \quad (2\kappa \alpha - \beta^2 > 0, \ \text{Cov}(dW_t^{(\sigma)}, dW_t) = \rho \, dt) \]

3. **Pitfalls of the Estimation Methods for \( \sigma \):**
   - Spot volatility estimation has received comparatively much less attention than integrated variance estimation;
   - Few bandwidth selection methods (e.g., Kristensen(2010))
A Proposed Bandwidth Selection Method

1. Technical Assumption (Kristensen, 2010):

\[ |\sigma^2_{t+\delta} - \sigma^2_t| = L_t(\delta)\delta^\nu + o_P(\delta^\nu), \quad \forall t, \quad a.s., \quad (\delta \to 0), \quad (\star) \]

where \( \nu \in (0,1] \) and \( \delta \to L_t(\delta) \) is slowly varying (random) function at 0 and \( t \to L_t(0) := \lim_{\delta \to 0^+} L_t(\delta) \) is continuous.
A Proposed Bandwidth Selection Method

1. Technical Assumption (Kristensen, 2010):

\[ |\sigma^2_{t+\delta} - \sigma^2_t| = L_t(\delta)\delta^\nu + o_P(\delta^\nu), \quad \forall t, \quad a.s., \quad (\delta \to 0), \quad (\star) \]

where \( \nu \in (0, 1] \) and \( \delta \to L_t(\delta) \) is slowly varying (random) function at 0 and \( t \to L_t(0) := \lim_{\delta \to 0^+} L_t(\delta) \) is continuous.

2. Under (\( \star \)), Kristensen(2010) proposes the following (asymptotically) "optimal" bandwidth:

\[ bw_{opt,t} = n^{-\frac{1}{2\nu+1}} \left( \frac{\sigma^4_t \|K\|^2}{\nu L^2_t(0)} \right)^{\frac{1}{2\nu+1}} \quad (\star\star) \]
A Proposed Bandwidth Selection Method

Technical Assumption (Kristensen, 2010):

\[ |\sigma^2_{t+\delta} - \sigma^2_t| = L_t(\delta)\delta^\nu + o_P(\delta^\nu), \quad \forall t, \quad a.s., \quad (\delta \to 0), \quad (\ast) \]

where \( \nu \in (0, 1] \) and \( \delta \to L_t(\delta) \) is slowly varying (random) function at 0 and \( t \to L_t(0) := \lim_{\delta \to 0^+} L_t(\delta) \) is continuous.

Under (\ast), Kristensen(2010) proposes the following (asymptotically) "optimal" bandwidth:

\[ bw^{loc}_{opt,t} = n^{-\frac{1}{2\nu + 1}} \left( \frac{\sigma^4_t \| K \|_2^2}{\nu L_t^2(0)} \right)^{\frac{1}{2\nu + 1}} \quad (\ast\ast) \]

Pitfall: In general, it is hard to check (\ast) with explicit constants \( \nu \) and \( L_t(0) \neq 0 \).
**Heuristic alternative approach**

1. **Idea:** Replace the "path-wise holder continuity assumption" with:

\[
\mathbb{E} \left[ (\sigma_{t+\delta}^2 - \sigma_t^2)^2 \middle| F_t \right] = L_t^2(0) \delta^{2\upsilon} + o_P(\delta^{\upsilon}), \quad \text{a.s.} \quad (\delta \to 0), \quad (***)
\]

for some positive adapted \( \{L_t(0)\}_{t \geq 0} \).

2. Then, use (**):

\[
bw_{opt,t}^{loc} = n^{-\frac{1}{2\upsilon+1}} \left( \frac{\sigma_t^4 \|K\|_2^2}{\upsilon L_t^2(0)} \right)^{\frac{1}{2\upsilon+1}} \quad (**)\]
Heuristic alternative approach

1. **Idea:** Replace the "path-wise holder continuity assumption" with:

\[
\mathbb{E} \left[ (\sigma_{t+\delta}^2 - \sigma_t^2)^2 \bigg| \mathcal{F}_t \right] = L_t^2(0)\delta^{2\upsilon} + o_P(\delta^\upsilon), \quad \text{a.s.} \quad (\delta \to 0), \quad (\star \star \star)
\]

for some positive adapted \( \{L_t(0)\}_{t \geq 0} \).

2. Then, use (\star \star):

\[
bw_{opt,t}^{loc} = n^{-\frac{1}{2\upsilon+1}} \left( \frac{\sigma_t^4 \| K \|_2^2}{\upsilon L_t^2(0)} \right)^{\frac{1}{2\upsilon+1}}
\]

3. **Example:** For the CIR model \( d\sigma_t^2 = \kappa (\alpha - \sigma_t^2) \, dt + \beta \sigma_t \, dW_t^{(\sigma)} \), it turns out that

\[
\mathbb{E} \left( (\sigma_{t+\delta}^2 - \sigma_t^2)^2 \bigg| \mathcal{F}_t \right) = \beta^2 \sigma_t^2 \delta + o(\delta).
\]

Hence, (\star \star \star) holds with \( \upsilon = \frac{1}{2} \) and \( L_t^2(0) = \beta^2 \sigma_t^2 \).

4. This in turn suggests the following local bandwidth selection method:

\[
bw_{opt,t}^{loc} = n^{-1/2} \left( \frac{2\sigma_t^2 \| K \|_2^2}{\beta^2} \right)^{1/2}
\]
Estimation Method for the CIR model (No Jumps)

Get a "rough" estimate of $\sigma^2_t$; e.g., using Alvarez et al. (2010):

$$\hat{\sigma}^2_0(t_i) = \frac{QV_{t_i+\sqrt{h_i}} - QV_{t_i}}{\sqrt{h_i}} = \frac{1}{\sqrt{h_i}} \sum_{j: t_j \in (t_i, t_i + \sqrt{h_i}]} (\Delta^n_j X)^2.$$
Estimation Method for the CIR model (No Jumps)

1. Get a “rough” estimate of $\sigma_t^2$; e.g., using Alvarez et al. (2010):
\[
\hat{\sigma}_0^2(t_i) = \frac{QV_{t_i + \sqrt{h_i}} - QV_{t_i}}{\sqrt{h_i}} = \frac{1}{\sqrt{h_i}} \sum_{j : t_j \in (t_i, t_i + \sqrt{h_i})} (\Delta_j^n X)^2.
\]

2. Estimate the vol vol $\beta$ using the realized variation of $\{\hat{\sigma}_0(t_i)\}_{i=1,...,n}$ since, for the CIR model, $\langle \sigma, \sigma \rangle_t = \frac{\beta^2}{4}$; denote such an estimate by $\hat{\beta}$;
Estimation Method for the CIR model (No Jumps)

1. Get a “rough” estimate of $\sigma^2_t$; e.g., using Alvarez et al. (2010):

$$\hat{\sigma}^2_0(t_i) = \frac{QV_{t_i+\sqrt{h_i}} - QV_{t_i}}{\sqrt{h_i}} = \frac{1}{\sqrt{h_i}} \sum_{j:t_j \in (t_i, t_i+\sqrt{h_i}]} (\Delta^n_j X)^2.$$

2. Estimate the vol vol $\beta$ using the realized variation of $\{\hat{\sigma}_0(t_i)\}_{i=1,\ldots,n}$ since, for the CIR model, $\langle \sigma, \sigma \rangle_t = \frac{\beta^2}{4}$; denote such an estimate by $\hat{\beta}$.

3. Estimate the optimal local bandwidth based on $\hat{\sigma}^2_0$ and $\hat{\beta}$:

$$\hat{\theta}_0(t_i) := n^{-1/2} \left( \frac{2\hat{\sigma}^2_0(t_i)\|K\|^2_2}{\hat{\beta}^2} \right)^{1/2}$$
Estimation Method for the CIR model (No Jumps)

1. Get a “rough” estimate of $\sigma^2_t$; e.g., using Alvarez et al. (2010):

$$\hat{\sigma}^2_0(t_i) = \frac{QV_{t_i+\sqrt{h_i}} - QV_{t_i}}{\sqrt{h_i}} = \frac{1}{\sqrt{h_i}} \sum_{j: t_j \in (t_i, t_i+\sqrt{h_i})} (\Delta^n_j X)^2.$$ 

2. Estimate the vol vol $\beta$ using the realized variation of $\{\hat{\sigma}_0(t_i)\}_{i=1,...,n}$ since, for the CIR model, $\langle \sigma, \sigma \rangle_t = \frac{\beta^2}{4}$; denote such an estimate by $\hat{\beta}$;

3. Estimate the optimal local bandwidth based on $\hat{\sigma}^2_0$ and $\hat{\beta}$:

$$\hat{\theta}_0(t_i) := n^{-1/2} \left( \frac{2\hat{\sigma}^2_0(t_i) \| K \|_2^2}{\hat{\beta}^2} \right)^{1/2}$$

4. Apply kernel estimation with the estimated optimal local bandwidth:

$$\hat{\sigma}^2_1(t_i) := \sum_{j=-\ell}^{\ell} |\Delta_{i+j} X|^2 K_{\hat{\theta}_0(t_i)}(t_i - t_{i+j})$$
Estimation Method for the CIR model (No Jumps)

1. Get a “rough” estimate of $\sigma_t^2$; e.g., using Alvarez et al. (2010):

$$\hat{\sigma}_0^2(t_i) = \frac{QV_{t_i+\sqrt{h_i}} - QV_{t_i}}{\sqrt{h_i}} = \frac{1}{\sqrt{h_i}} \sum_{j: t_j \in (t_i, t_i+\sqrt{h_i})} (\Delta_j X)^2.$$ 

2. Estimate the vol vol $\beta$ using the realized variation of $\{\hat{\sigma}_0(t_i)\}_{i=1,\ldots,n}$ since, for the CIR model, $\langle \sigma, \sigma \rangle_t = \frac{\beta^2}{4}$; denote such an estimate by $\hat{\beta}$.

3. Estimate the optimal local bandwidth based on $\hat{\sigma}_0^2$ and $\hat{\beta}$:

$$\hat{\theta}_0(t_i) := n^{-1/2} \left( \frac{2\hat{\sigma}_0^2(t_i) \|K\|_2^2}{\beta^2} \right)^{1/2}$$

4. Apply kernel estimation with the estimated optimal local bandwidth:

$$\hat{\sigma}_1^2(t_i) := \sum_{j=-\ell}^{\ell} |\triangle_{i+j} X|^2 K_{\hat{\theta}_0(t_i)}(t_i - t_{i+j})$$

5. Iterate steps 3 and 4: $\hat{\sigma}_0^2(\cdot) \rightarrow \hat{\theta}_0(\cdot) \rightarrow \hat{\sigma}_1^2(\cdot) \rightarrow \hat{\theta}_1(\cdot) \rightarrow \hat{\sigma}_2^2(\cdot) \rightarrow \cdots$
Numerical Illustration (no jumps): Uniform Kernel

Model:
CIR stochastic volatility
\[ d\sigma_t^2 = \kappa (\alpha - \sigma_t^2) \, dt + \beta \sigma_t dW_t^{(\sigma)}. \]

Parameters:
\[ \gamma_t \equiv 0.05, \lambda = 0, \zeta_i \sim \mathcal{N}(0, 0.3), \kappa = 5, \alpha = 0.04, \beta = 0.5, \]
\[ \text{Cov}(dW_t, dW_t^{(\sigma)}) = \rho dt \]

Regular Sampling Scheme:
T=1/12 (one month) and \( h_n = 5 \) min

Monte Carlo results of \( \text{MSE} = \sum_{i=1}^{n} (\hat{\sigma}^2(t_i) - \sigma^2(t_i))^2 \) based on 500 runs

<table>
<thead>
<tr>
<th>Method</th>
<th>( \rho = -0.5 )</th>
<th>( \rho = 0 )</th>
<th>( \rho = 0.5 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alvarez et al. Method</td>
<td>0.0741</td>
<td>0.0756</td>
<td>0.0766</td>
</tr>
<tr>
<td>Iterated Kernel Est. with opt. loc. bw</td>
<td>0.0437</td>
<td>0.0441</td>
<td>0.0445</td>
</tr>
<tr>
<td>Oracle(^1) Kernel Est. with opt. loc. bw</td>
<td>0.0463</td>
<td>0.0466</td>
<td>0.0472</td>
</tr>
</tbody>
</table>

\(^1\)Using true parameters values for \( \gamma \) and \( \sigma \)
Get a “rough” estimate of \( t \rightarrow \sigma_t^2 \) by 
\[
\tilde{\sigma}^2(t_i) = \frac{1}{\sqrt{h_i}} \sum_{j: t_j \in (t_i, t_i + \sqrt{h_i}]} \left( \Delta_j^n X \right)^2.
\]
Outlined of Estimation Method (With Jumps)

1. Get a “rough” estimate of $t \rightarrow \sigma_t^2$ by $$\tilde{\sigma}^2(t_i) = \frac{1}{\sqrt{h_i}} \sum_{j: t_j \in (t_i, t_i + \sqrt{h_i})} \left( \Delta_j^n X \right)^2.$$  

2. Using $\tilde{\sigma}^2$, estimate the optimal threshold $\tilde{B}(t_i) := \left[ 3\tilde{\sigma}^2(t_i) h_i \ln(1/h_i) \right]^{1/2}$;
Outlined of Estimation Method (With Jumps)

1. Get a “rough” estimate of $t \rightarrow \sigma_t^2$ by $
\tilde{\sigma}_t^2 = \frac{1}{\sqrt{h_i}} \sum_{j : t_j \in (t_i, t_i + \sqrt{h_i})} (\Delta_j^n X)^2.

2. Using $\tilde{\sigma}^2$, estimate the optimal threshold $\tilde{B}(t_i) := \left[3\tilde{\sigma}^2(t_i)h_i \ln(1/h_i)\right]^{1/2};$

3. Refine $\tilde{\sigma}^2(t_i)$ using thresholding,

$$
\hat{\sigma}_0^2(t_i) = \frac{1}{\sqrt{h_i}} \sum_{j : t_j \in (t_i, t_i + \sqrt{h_i})} (\Delta_j^n X)^2 1_{[|\Delta_j X| \leq \tilde{B}(t_i)]}
$$
Outlined of Estimation Method (With Jumps)

1. Get a “rough” estimate of $t \rightarrow \sigma_t^2$ by $	ilde{\sigma}^2(t_i) = \frac{1}{\sqrt{h_i}} \sum_{j: t_j \in (t_i, t_i + \sqrt{h_i})} (\Delta_j^n X)^2$.
2. Using $\tilde{\sigma}^2$, estimate the optimal threshold $\tilde{B}(t_i) := [3\tilde{\sigma}^2(t_i)h_i \ln(1/h_i)]^{1/2}$;
3. Refine $\tilde{\sigma}^2(t_i)$ using thresholding,

$$\hat{\sigma}^2_0(t_i) = \frac{1}{\sqrt{h_i}} \sum_{j: t_j \in (t_i, t_i + \sqrt{h_i})} (\Delta_j^n X)^2 1_{|\Delta_j X| \leq \tilde{B}(t_i)}$$

4. Estimate $\beta$ using the realized variation of $\{\hat{\sigma}_0(t_i)\}_{i=1,\ldots,n} \Rightarrow \hat{\beta}$. 
_Outlined of Estimation Method (With Jumps)_

1. Get a “rough” estimate of $t \to \sigma^2_t$ by 
   $$\tilde{\sigma}^2(t_i) = \frac{1}{\sqrt{h_i}} \sum_{j: t_j \in (t_i, t_i + \sqrt{h_i})} (\Delta^n_j X)^2.$$ 
2. Using $\tilde{\sigma}^2$, estimate the optimal threshold 
   $$\tilde{B}(t_i) := \left[ 3\tilde{\sigma}^2(t_i) h_i \ln(1/h_i) \right]^{1/2};$$ 
3. Refine $\tilde{\sigma}^2(t_i)$ using thresholding, 
   $$\hat{\sigma}^2_0(t_i) = \frac{1}{\sqrt{h_i}} \sum_{j: t_j \in (t_i, t_i + \sqrt{h_i})} (\Delta^n_j X)^2 1[|\Delta_j X| \leq \tilde{B}(t_i)]$$ 
4. Estimate $\beta$ using the realized variation of 
   $$\{\hat{\sigma}_0(t_i)\}_{i=1,\ldots,n} \Rightarrow \hat{\beta}.$$ 
5. Refine the estimates of the optimal threshold and local bandwidth: 
   $$\hat{B}_0(t_i) := \left( 3\hat{\sigma}^2_0(t_i) h_i \ln \frac{1}{h_i} \right)^{1/2}, \quad \hat{\theta}_0(t_i) := n^{-1/2} \left( \frac{2\hat{\sigma}^2_0(t_i) \| K \|_2^2}{\hat{\beta}^2} \right)^{1/2}$$
Outlined of Estimation Method (With Jumps)

1. Get a “rough” estimate of $t \rightarrow \sigma_t^2$ by $\tilde{\sigma}^2(t_i) = \frac{1}{\sqrt{h_i}} \sum_{j: t_j \in (t_i, t_{i+\sqrt{h_i}}]} \left( \Delta_j^n X \right)^2$.
2. Using $\tilde{\sigma}^2$, estimate the optimal threshold $\tilde{B}(t_i) := \left[ 3\tilde{\sigma}^2(t_i) h_i \ln(1/h_i) \right]^{1/2}$;
3. Refine $\tilde{\sigma}^2(t_i)$ using thresholding,
   \[ \hat{\sigma}_0^2(t_i) = \frac{1}{\sqrt{h_i}} \sum_{j: t_j \in (t_i, t_{i+\sqrt{h_i}}]} \left( \Delta_j^n X \right)^2 1_{|\Delta_j X| \leq \tilde{B}(t_i)} \]
4. Estimate $\beta$ using the realized variation of $\{\hat{\sigma}_0(t_i)\}_{i=1}^{n} \Rightarrow \hat{\beta}$.
5. Refine the estimates of the optimal threshold and local bandwidth:
   \[ \hat{B}_0(t_i) := \left( 3\hat{\sigma}_0^2(t_i) h_i \ln \frac{1}{h_i} \right)^{1/2}, \quad \hat{\theta}_0(t_i) := n^{-1/2} \left( \frac{2\hat{\sigma}_0^2(t_i) \|K\|_2^2}{\hat{\beta}^2} \right)^{1/2} \]
6. Based on $\hat{\theta}_0(t_i)$ and $\hat{B}_0(t_i)$, refine the estimate $\hat{\sigma}_0^2(t_i)$:
   \[ \hat{\sigma}_1^2(t_i) := \sum_{j=-\ell}^{\ell} |\Delta_{i+j} X|^2 K_{\hat{\theta}_0(t_i)}(t_i - t_{i+j}) 1_{|\Delta_j X| \leq \hat{B}_0(t_i)} \]
Numerical Illustration (with jumps): Uniform Kernel

Model:
Normal jump sizes and CIR stochastic volatility.

Parameters:
\( \gamma_t \equiv 0.05, \lambda = 120, \zeta_i \sim \mathcal{N}(0, 0.3), \kappa = 5, \alpha = 0.04, \beta = 0.5. \)

Regular Sampling Scheme:
\( T = 1/12 \) (one month) and \( h_n = 5 \) min

Monte Carlo results of \( \text{MSE} = \sum_{i=1}^{n} \left( \hat{\sigma}^2(t_i) - \sigma^2(t_i) \right)^2 \) based on 500 runs

<table>
<thead>
<tr>
<th>Method</th>
<th>MSE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alvarez et al. Method</td>
<td>2.3854</td>
</tr>
<tr>
<td>Alvarez et al. Method with thresholding</td>
<td>0.8602</td>
</tr>
<tr>
<td>Kernel Est. with thresholding and opt. loc. bw selection</td>
<td>0.0751</td>
</tr>
<tr>
<td>Twice Kernel Est. with thresholding and opt. loc. bw selection</td>
<td>0.0652</td>
</tr>
<tr>
<td>Oracle(^2) Kernel Est. with thresholding and loc. bw selection</td>
<td>0.0533</td>
</tr>
</tbody>
</table>

\(^2\)Using true parameters values for \( \gamma \) and \( \sigma \)
Main Contributions

1. Introduced an objective threshold selection procedure based on statistically optimal criteria.
2. Developed the infill asymptotic characterization of the optimal threshold.
3. Proposed an iterative algorithm to find the optimal threshold sequence.
4. Proposed extensions to more general stochastic models, which allows time-varying stochastic volatility and jump intensity.
Ongoing and Future Work

1. Connection between the optimal threshold $B^*_{1}$ and the data-based threshold $\hat{B}^*_{1}$ obtained by the iterative method
2. Implementation based on the second-order approximation $B^*_{2}$ of $B^*$
3. Extensions to infinite-jump activity processes
4. Incorporation of a microstructure noise component
5. Extensions to other loss functions
6. Extensions to other estimation problems where parameter tuning is needed.
For Further Reading I

Figueroa-López & Nisen.  
Optimally Thresholded Realized Power Variations for Lévy Jump Diffusion Models  

Figueroa-López & Nisen.  
Figure: Classical Path of $X_t = \gamma t + \sigma W_t + \sum_{i=1}^{N_t} \zeta_i$; times and sizes of consecutive jumps are denoted by $\tau_1 < \cdots < \tau_n$ and $\zeta_1, \ldots, \zeta_n$, respectively.
**Figure:** Boxplots for Volatility Estimation: Based on 1,000 sample paths. Parameters:

\[ \sigma = 0.35, \lambda = 197.55, \zeta = D \mathcal{N}(0, 0.037^2), \quad T = 6\text{-months}, \text{ sampling frequency} = 5\text{-min}. \]
Figure: Boxplots for Volatility Estimators: Based on 1,000 sample paths. Parameters: $\sigma = 0.45$, $\lambda = 32.84$, $\zeta = D\ t\text{-student}(3 \text{ d.f.})$, $T = 1\text{-year}$, $h_n = 15\text{-min.}$
Figure: Boxplots for rate estimators $\hat{\lambda} := \frac{\hat{N}[B]}{T}$: Based on 1,000 sample paths.
Parameters: $\sigma = 0.45, \lambda = 32.84, \zeta =_{D} t$-student(3 d.f.), $T = 1$-year, $h_n = 15$-min.