Short-Time Asymptotic Methods
In Financial Mathematics

José E. Figueroa-López
Department of Mathematics and Statistics
Washington University in St. Louis

Department Of Statistics and Probability, MSU
November 29, 2018

Based on joint works with
R. Gong, C. Mancini, J. Nisen, C. Houdré, and S. Ólafsson
Research supported by NSF grants: DMS-1613016 & DMS-1561141
Outline

1. Overview Of Three Applications
   - Inference Of Stochastic Processes Using High-Frequency Data
   - Model Selection & Calibration via Short-Time Option Prices
   - Change-Point Detection In Continuous-Time

2. A Closer Look At The Problem Of Optimal Threshold Selection

3. A Closer Look At Optimal Kernel Estimation of Spot Volatility
Overview Of Three Applications

Inference Of Stochastic Processes Using High-Frequency Data
Model Selection & Calibration via Short-Time Option Prices
Change-Point Detection In Continuous-Time

A Closer Look At The Problem Of Optimal Threshold Selection

A Closer Look At Optimal Kernel Estimation of Spot Volatility
Outline

1. Overview Of Three Applications
   Inference Of Stochastic Processes Using High-Frequency Data
   Model Selection & Calibration via Short-Time Option Prices
   Change-Point Detection In Continuous-Time

2. A Closer Look At The Problem Of Optimal Threshold Selection

3. A Closer Look At Optimal Kernel Estimation of Spot Volatility
In econometrics, an important problem is the estimation of the integrated variance,

\[ IV_T := \int_0^T \sigma_s^2 ds, \quad \text{(fixed } T) , \]

for a semimartingale model

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

(\( W \) Standard Brownian Motion, \( J \) pure-jump process),

based on a record \( X_{t_1}, \ldots, X_{t_n} \) of discrete observations of \( X \) when

\[ \max_i (t_i - t_{i-1}) \to 0 \]

(high-frequency or infill estimation).
Overview Of Three Applications

Inference Of Stochastic Processes Using High-Frequency Data

Key Estimators

1. Realized Quadratic Variation:

\[
QV_n := \sum_{i=0}^{n-1} (X_{t_{i+1}} - X_{t_i})^2 \xrightarrow{\text{mesh} \to 0} \int_0^T \sigma_s^2 ds + \sum_{t \leq T} (\Delta J_t)^2
\]

2. Bipower Realized Variations (Barndorff-Nielsen and Shephard):

\[
BPV_n := \sum_{i=0}^{n-2} \left| X_{t_{i+1}} - X_{t_i} \right| \left| X_{t_{i+2}} - X_{t_{i+1}} \right| \xrightarrow{\text{mesh} \to 0} \frac{2}{\pi} \int_0^T \sigma_s^2 ds
\]

3. Truncated Realized Variations (Mancini):

\[
TRV_n(\varepsilon_n) := \sum_{i=0}^{n-1} (X_{t_{i+1}} - X_{t_i})^2 1_{\left\{ \left| X_{t_{i+1}} - X_{t_i} \right| \leq \varepsilon_n \right\}} \xrightarrow{\text{mesh} \to 0} \int_0^T \sigma_s^2 ds
\]

J.E. Figueroa-López (WUSTL)

Short-Time Asymptotics in Financial Mathematics

Dept. Of Statistics & Probability, MSU
Some Questions

- What are the infill asymptotic properties of the estimators? Asymptotic bias, variance, CLT, etc… Is it better to use BPV or TRV?

- How do you calibrate or tune up the truncation parameter $\varepsilon_n$?

- How do you optimally estimate the spot volatility $\sigma_t$?

- How do you account for potential observation errors (called microstructure noise)?
Outline

1. Overview Of Three Applications
   - Inference Of Stochastic Processes Using High-Frequency Data
   - Model Selection & Calibration via Short-Time Option Prices
   - Change-Point Detection In Continuous-Time

2. A Closer Look At The Problem Of Optimal Threshold Selection

3. A Closer Look At Optimal Kernel Estimation of Spot Volatility
Basic Financial Terminology

- One of the main problems of financial mathematics is to characterize sensible prices of options.
- The most important type of options are European call and European put option.
- A call (put) is a contract in which the holder of the option has the right to buy (sell) one share of an underlying stock for a predetermined price $K$ (strike) at some specified future time $t$ (maturity). In exchange, the holder pays a premium $\Pi_0$ at time 0.
- At maturity $t$, the holder’s payoff is $(S_t - K)_+ ((K - S_t)_+)$ for call (put) option, where $S_t$ is price of stock per share.
- If $K > S_0$ ($K < S_0$), we say the Call is Out-The-Money (In-The-Money), while if $K = S_0$ we say the option is At-The-Money.
Carr and Wu ’03 studied the asymptotic behavior of European option prices as time-to-maturity $t$ shrinks to 0.

Argued prices are sharply different in short-time for a Purely Continuous, a Pure-Jump, or a Mixed (combination of both) Model.

The following rough asymptotics for option prices were suggested:

<table>
<thead>
<tr>
<th>Model</th>
<th>Out-of-the-money (OTM)</th>
<th>At-the-money (ATM)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Purely Continuous (PC)</td>
<td>$O(e^{-c/t})$, $c &gt; 0$</td>
<td>$O(\sqrt{t})$</td>
</tr>
<tr>
<td>Pure Jump (PJ)</td>
<td>$O(t)$</td>
<td>$O(t^p)$, $p \in (0, 1)$</td>
</tr>
<tr>
<td>Mixed Model (M)</td>
<td>$O(t)$</td>
<td>$O(t^p)$, $p \in (0, 1/2)$</td>
</tr>
</tbody>
</table>

The above have been formalized and extended to higher order in several papers (Forde & Jaquier ’09; Tankov ’11; F-L & Forde ’12; F-L, Gong, & Houdré ’12 and ‘16; F-L & Ólafsson ’16).
Black-Scholes Implied Volatility Asymptotics

- The Black-Scholes (BS) price $\Pi^{BS}$ is the only sensible option premium in a market where the stock price process is a geometric B.M.:

$$S_t = S_0 e^{\sigma W_t + \mu t}.$$

- The Implied Volatility $\hat{\sigma}$ of an option price $\Pi^*$ is the value of the BM’s standard deviation $\sigma$ needed for the BS price $\Pi^{BS}(\sigma)$ to coincide with $\Pi^*$.

- The asymptotic behavior of option prices translates into asymptotics for the IV:

<table>
<thead>
<tr>
<th>Model</th>
<th>Out-of-the-money (OTM)</th>
<th>At-the-money (ATM)</th>
</tr>
</thead>
<tbody>
<tr>
<td>PC</td>
<td>$\hat{\sigma}_t \rightarrow d \neq 0$</td>
<td>$\hat{\sigma}_t \rightarrow \sigma_0$, spot vol.</td>
</tr>
<tr>
<td>PJ</td>
<td>$\hat{\sigma}_t \sim \frac{\kappa}{\sqrt{2t \log(1/t)}} \rightarrow \infty$</td>
<td>$\hat{\sigma}_t = O(t^p)$, $p \in (0, \frac{1}{2}]$</td>
</tr>
<tr>
<td>M</td>
<td>$\hat{\sigma}_t \sim \frac{\kappa}{\sqrt{2t \log(1/t)}} \rightarrow \infty$</td>
<td>$\hat{\sigma}_t \rightarrow \sigma_0$, spot vol.</td>
</tr>
</tbody>
</table>
Some Implications For The IV Smile

- The IV smile is a graph of the IV’s of options with the same maturity but different strikes against their strikes $K$ or log-moneyness $\kappa = \ln(K/S_0)$.
- The property that $\hat{\sigma}_t \nearrow \infty$ for OTM options while $\hat{\sigma}_t \to \sigma_0$ for ATM options means that, under the presence of jumps, we expect very pronounced IV “smiles” for short maturity options.
- This behavior is consistent with market-quotes as shown by the following IV smiles on Jan. 15, ’14 (mat. from 8 days to 3 months).
Outline

1. Overview Of Three Applications
   - Inference Of Stochastic Processes Using High-Frequency Data
   - Model Selection & Calibration via Short-Time Option Prices
   - Change-Point Detection In Continuous-Time

2. A Closer Look At The Problem Of Optimal Threshold Selection

3. A Closer Look At Optimal Kernel Estimation of Spot Volatility
Change-Point Detection

- **Change-point detection** is the problem of detecting, with as little delay as possible, a change in the statistical properties of a process that is being observed discretely or continuously in time.

- First began to emerge in quality control applications in the 1930’s.

- Applications in various branches of science and engineering: Signal processing, statistical surveillance, climate monitoring, cybersecurity, ...

- Applications in finance:
  1. Active risk management (e.g., shifts in the parameters of credit risk models or the expected performance of an investment)
  2. Actuarial science (e.g., changes in mortality rates);
  3. Algorithmic trading strategies.
Lorden’s Change-Point Detection Problem

- Sequentially observe a process \((X_t)_{t \in \mathcal{T}}\), whose statistical properties change abruptly at some unknown nonrandom point in time \(\tau \in \mathcal{T} \cup \{\infty\}\).
- A sequential detection procedure is a stopping time \(T\), w.r.t. the observed process \((X_t)_{t \in \mathcal{T}}\), at which a “change” is declared.
- The design of an optimal detection procedure aims at minimizing the detection delay \((T - \tau) \mathbf{1}_{\{T > \tau\}}\), while controlling the false alarm “rate” \(P_\infty [T < t], \forall t)\):
Lorden’s Change-Point Detection Problem

- Sequentially observe a process \((X_t)_{t \in \mathcal{T}}\), whose statistical properties change abruptly at some unknown nonrandom point in time \(\tau \in \mathcal{T} \cup \{\infty\}\).
- A sequential detection procedure is a stopping time \(T\), w.r.t. the observed process \((X_t)_{t \in \mathcal{T}}\), at which a “change” is declared.
- The design of an optimal detection procedure aims at minimizing the detection delay \((T - \tau) \mathbf{1}_{\{T > \tau\}}\), while controlling the false alarm “rate” \(\mathbb{P}_\infty[\tau < t], \forall t\):

\[
\inf_{T: \mathbb{E}_\infty[T] \geq \gamma} \sup_{\tau \in \mathcal{T}} \text{ess sup}_\omega \mathbb{E}_\tau \left[ (T - \tau)_+ \ \big| \mathcal{F}_\tau \right]
\]

(Lorden’s worst-worst case approach)
Moustakides, ’86: When $X_1, X_2, \ldots$ are i.i.d. with densities $f$ and $g$ before and after the change point $\tau$, the optimal stopping time is the CUSUM Rule:

$$T_h = \inf\{t : y_t \geq h\}, \quad y_t := u_t - \inf_{s \leq t} u_s,$$

**Drawup Process**

$$u_t := \sum_{i=0}^{t} \log \frac{g(X_i)}{f(X_i)}$$

**Log-Likelihood Ratio Process**
Change-Point Detection In Continuous Time

- By shifting Lorden’s optimality criterion from $\mathbb{E}_\tau [(T - \tau)_+ | \mathcal{F}_\tau]$ to $\mathbb{E}_\tau [(u_T - u_\tau) \mathbf{1}_{\{T \geq \tau\}} | \mathcal{F}_\tau]$, optimality of the CUSUM rule in continuous time has been established for arbitrary processes with continuous paths (Chronopoulou & Fellouris ’13)

- Results for processes with discontinuities are rare, except in special cases; e.g., changes in Cox processes (El Karoui et al. ’15)

- F-L & Ólafsson ’18+: The CUSUM stopping time $T_{ch}$ is optimal to detect the change point $\tau$ from a Lévy process to another:

$$dX_t = dX_0 t \mathbf{1}_{\{t < \tau\}} + dX_1 t \mathbf{1}_{\{t \geq \tau\}},$$

where $(X_0 t)_{t \geq 0}$ and $(X_1 t)_{t \geq 0}$ are two Lévy processes with different Lévy triplets so that their laws are mutually equivalent.
Change-Point Detection In Continuous Time

- By shifting Lorden’s optimality criterion from $\mathbb{E}_\tau [(T - \tau)^+ | \mathcal{F}_\tau]$ to $\mathbb{E}_\tau [(u_T - u_\tau)1_{\{T \geq \tau\}} | \mathcal{F}_\tau]$, optimality of the CUSUM rule in continuous time has been established for arbitrary processes with continuous paths (Chronopoulou & Fellouris ’13).

- Results for processes with discontinuities are rare, except in special cases; e.g., changes in Cox processes (El Karoui et al. ’15).
Change-Point Detection In Continuous Time

- By shifting Lorden’s optimality criterion from $E_{\tau} \left[ (T - \tau)^+ \mid \mathcal{F}_{\tau} \right]$ to $E_{\tau} \left[ (u_T - u_\tau) \mathbf{1}_{\{T \geq \tau\}} \mid \mathcal{F}_{\tau} \right]$, optimality of the CUSUM rule in continuous time has been established for arbitrary processes with continuous paths (Chronopoulou & Fellouris ’13).

- Results for processes with discontinuities are rare, except in special cases; e.g., changes in Cox processes (El Karoui et al. ’15).

- F-L & Ólafsson ’18+: The CUSUM stopping time $T^c_h$ is optimal to detect the change point $\tau$ from a Lévy process to another:

$$dX_t = dX^0_t \mathbf{1}_{\{t < \tau\}} + dX^1_t \mathbf{1}_{\{t \geq \tau\}},$$

where $(X^0_t)_{t \geq 0}$ and $(X^1_t)_{t \geq 0}$ are two Lévy process with different Lévy triplets so that their laws are mutually equivalent.
Approach of Proof

The proof of the theorem consists of two main steps and uses again short-time asymptotic methods:

1. For $\Delta > 0$ show that a discretized version of the CUSUM stopping time,

$$T^c_h(\Delta) := \inf\{k\Delta \geq 0 : y_{k\Delta} \geq h\},$$

solves a hybrid problem where the change-point $\tau$ is restricted to values in the discrete set $(k\Delta)_{k \geq 0}$.

2. Let $\Delta \to 0$ and

   (i) show that $T^c_h(\Delta) \to T^c_h$,

   (ii) use a limiting procedure to establish the optimality of $T^c_h$ for the continuous-time change-detection problem.
Outline

1. Overview Of Three Applications
   - Inference Of Stochastic Processes Using High-Frequency Data
   - Model Selection & Calibration via Short-Time Option Prices
   - Change-Point Detection In Continuous-Time

2. A Closer Look At The Problem Of Optimal Threshold Selection

3. A Closer Look At Optimal Kernel Estimation of Spot Volatility
Given discrete observations $X_{t_1}, \ldots, X_{t_n}$ of a continuous semimartingale

$$dX_t = a_t dt + \sigma_t dW_t + dJ_t,$$

where $J_t$ is a pure-jump process, we consider the problem of estimating the integrated variance,

$$IV_T = \int_0^T \sigma_s^2 ds,$$

in a high-frequency and fixed time horizon sampling setting:

$$h := t_i - t_{i-1} \to 0, \quad t_n = T.$$
One of the most used estimator is the Truncated Realized Variations (TRV), proposed by Mancini ’03:

$$TRV_n(\varepsilon) := \sum_{i=0}^{n-1} (X_{t_{i+1}} - X_{t_i})^2 \mathbf{1}_{\{|X_{t_{i+1}} - X_{t_i}| \leq \varepsilon_n\}},$$

where $\varepsilon_n$, called the threshold, is a parameter to be tuned by the statistician.

**Problem**: How do we choose the threshold parameter $\varepsilon_n$?
Illustration I: Log-normal Merton Model

\[ X_t = at + \sigma W_t + \sum_{i: \tau_i \leq t} \zeta_i, \quad \zeta_i \sim \text{i.i.d. } N(\mu_{jmp}, \sigma_{jmp}^2), \quad \{\tau_i\}_{i \geq 1} \sim \text{Poisson}(\lambda) \]
Illustration I: Log-normal Merton Model

\[ X_t = at + \sigma W_t + \sum_{i: \tau_i \leq t} \zeta_i, \quad \zeta_i \overset{\text{i.i.d.}}{\sim} N(\mu_{\text{jmp}}, \sigma^2_{\text{jmp}}), \quad \{\tau_i\}_{i \geq 1} \sim \text{Poisson}(\lambda) \]

Figure: (left) 5-min Merton observations with \( \sigma = 0.4, \sigma_{\text{jmp}} = 3\sqrt{h}, \mu_{\text{jmp}} = 0, \lambda = 200; \) (right) TRV against its truncation level.
Illustration II: Log-normal Merton Model

Log-Normal Merton Model

Performace of Truncated Realized Variations

Figure: (left) 5 minute Merton observations with $\sigma = 0.2$, $\sigma_{jmp} = 1.5\sqrt{h}$, $\mu_{jmp} = 0$, $\lambda = 1000$; (right) TRV performance wrt the truncation level
Approach To Select The Threshold

1. Fix a sensible metric of the estimation error (typically, the Mean-Square Error);
2. Show the existence of a unique optimal threshold $\varepsilon_n^*$ that minimizes the chosen metric;
3. Analyze the (infill) asymptotic behavior $\varepsilon_n^*$ (when $n \to \infty$) to
   - determine its dependence on the underlying parameters of the model
   - Devise a plug-in type calibration of $\varepsilon_n^*$ by estimating those parameters (if possible).
Via Expected number of jump misclassifications

Notation:

\[ \Delta_n^i X := X_{t_i} - X_{t_{i-1}} \]

\[ \Delta_n^i N := \# \text{ of jumps during } (t_{i-1}, t_i] \]

Estimation Error: (F-L & Nisen '13)

\[ \text{Loss}_n(\varepsilon) := E_n \sum_{i=1}^{\Delta_n^i N} \left(1 \left[ |\Delta_n^i X| > \varepsilon, \Delta_n^i N = 0 \right] + 1 \left[ |\Delta_n^i X| \leq \varepsilon, \Delta_n^i N \neq 0 \right]\right) . \]

Underlying Principle

The estimation error of the TRV must strongly depend on the ability of the threshold to detect jumps.
Via Expected number of jump misclassifications

1. Notation:

\[ \Delta_i^n X := X_{t_i} - X_{t_{i-1}} \]

\[ \Delta_i^n N := \# \text{ of jumps during } (t_{i-1}, t_i] \]
Via Expected number of jump misclassifications

1. Notation:

\[ \Delta_i^n X := X_{t_i} - X_{t_{i-1}} \]
\[ \Delta_i^n N := \# \text{ of jumps during } (t_{i-1}, t_i] \]

2. Estimation Error: (F-L & Nisen ’13)

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

3. Underlying Principle

The estimation error of the TRV must strongly depend on the ability of the threshold to detect jumps.
Existence and Infill Asymptotic Characterization

Theorem (FL & Nisen, SPA 2013)

Let $X$ be a Finite Jump Activity Lévy process

$$X_t = at + \sigma W_t + \sum_{i=1}^{N_t} \xi_i.$$
Existence and Infill Asymptotic Characterization

Theorem (FL & Nisen, SPA 2013)

Let $X$ be a Finite Jump Activity Lévy process

$$X_t = a t + \sigma W_t + \sum_{i=1}^{N_t} \xi_i.$$ 

1. The loss function $\text{Loss}_n(\varepsilon)$ is quasi-convex with a unique global minimum $\varepsilon^*_n$. 

As $n \to \infty$, the optimal threshold sequence $(\varepsilon^*_n)_n$ is such that 

$$\varepsilon^*_n = \sqrt{\frac{3 \sigma^2}{h_n \log(1/h_n)}} + h.o.t.$$

where hereafter h.o.t. refers to 'higher order terms'.
Existence and Infill Asymptotic Characterization

Theorem (FL & Nisen, SPA 2013)

Let $X$ be a Finite Jump Activity Lévy process

$$X_t = at + \sigma W_t + \sum_{i=1}^{N_t} \xi_i.$$ 

1. The loss function $\text{Loss}_n(\varepsilon)$ is quasi-convex with a unique global minimum $\varepsilon_n^*$. 

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

$$\varepsilon_n^* = \sqrt{3\sigma^2 h_n \log \left( \frac{1}{h_n} \right)} + \text{h.o.t.},$$

where hereafter h.o.t. refers to ‘higher order terms’.
Remarks

1. Why $\sqrt{h \log(1/h)}$?
   This is proportional to modulus of continuity of the B.M.:
   \[
   \limsup_{h \to 0} \frac{1}{\sqrt{2h \log(1/h)}} \sup_{s,t \in [0,1]: |t-s| < h} |W_t - W_s| = 1.
   \]

2. Practically,
   \[
   \varepsilon_n^* := \sqrt{3\sigma^2 h_n \log \left( \frac{1}{h_n} \right)}
   \]
   provides us with a “blueprint” for devising threshold sequences with good estimation properties!
A Feasible Implementation based on $\varepsilon^*_n$

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

$$\hat{\sigma}^2_{n,0} := \frac{1}{T} \sum_{i=1}^{n} |X_{t_i} - X_{t_{i-1}}|^2$$
A Feasible Implementation based on $\varepsilon_n^*$

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

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

(ii) Use $\hat{\sigma}_{n,0}^2$ to estimate the optimal threshold

$$\hat{\varepsilon}_{n,0}^* := (3 \hat{\sigma}_{n,0}^2 h_n \log(1/h_n))^{1/2}$$
A Feasible Implementation based on $\varepsilon^*_n$

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

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

(ii) Use $\hat{\sigma}^2_{n,0}$ to estimate the optimal threshold

$$\hat{\varepsilon}^*_{n,0} := \left( 3 \hat{\sigma}^2_{n,0} h_n \log(1/h_n) \right)^{1/2}$$

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

$$\hat{\sigma}^2_{n,1} = \frac{1}{T} \sum_{i=1}^{n} |X_{t_i} - X_{t_{i-1}}|^2 1[|X_{t_i} - X_{t_{i-1}}| \leq \hat{\varepsilon}^*_{n,0}]$$
A Feasible Implementation based on $\varepsilon_n^1$

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

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

(ii) Use $\hat{\sigma}_{n,0}^2$ to estimate the optimal threshold

$$\hat{\varepsilon}_{n,0}^* := \left(3 \hat{\sigma}_{n,0}^2 h_n \log(1/h_n)\right)^{1/2}$$

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

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

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

$$\hat{\sigma}_{n,0}^2 \rightarrow \hat{\varepsilon}_{n,0}^* \rightarrow \hat{\sigma}_{n,1}^2 \rightarrow \hat{\varepsilon}_{n,1}^* \rightarrow \hat{\sigma}_{n,2}^2 \rightarrow \cdots \rightarrow \hat{\sigma}_{n,\infty}^2$$
Illustration I

**Figure:** (left) Merton Model with $\sigma = 0.4$, $\sigma_{jmp} = 3\sqrt{h}$, $\mu_{jmp} = 0$, $\lambda = 200$; (right) TRV performance wrt the truncation level. Red dot is $\hat{\sigma}_{n,1} = 0.409$, while purple dot is the limiting estimator $\hat{\sigma}_{n,\infty} = 0.405$.
Figure: (left) Merton Model with $\sigma = 0.2$, $\sigma_{jmp} = 1.5\sqrt{h}$, $\mu_{jmp} = 0$, $\lambda = 1000$; (right) TRV performance wrt the truncation level. Red dot is $\hat{\sigma}_{n,1} = 0.336$, while purple dot is the limiting estimator $\hat{\sigma}_{n,\infty} = 0.215$.
We now propose a second approach in which we aim to control the conditional MSE:

$$\text{MSE}_c(\varepsilon) := \mathbb{E}\left[ (\text{TRV}_n(\varepsilon) - \int_0^T \sigma^2_s ds)^2 \bigg| \sigma, \mathcal{J} \right].$$

Assumptions:

- $\sigma_t > 0$, $\forall t$,
- $\sigma$ and $\mathcal{J}$ are independent of $\mathcal{W}$. 

J.E. Figueroa-López (WUSTL)  
Short-Time Asymptotics in Financial Mathematics  
Dept. Of Statistics & Probability, MSU
We now propose a second approach in which we aim to control the conditional MSE:

$$MSE_c(\varepsilon) := \mathbb{E} \left[ \left( TRV_n(\varepsilon) - \int_0^T \sigma_s^2 ds \right)^2 \left| \sigma, J \right. \right].$$

Assumptions: $\sigma_t > 0$, $\forall t$, and $\sigma$ and $J$ are independent of $W$. 

### Short-Time Asymptotics in Financial Mathematics

Dept. Of Statistics & Probability, MSU 31 / 63
We now propose a second approach in which we aim to control the conditional MSE:

$$MSE_c(\varepsilon) := \mathbb{E} \left[ \left( TRV_n(\varepsilon) - \int_0^T \sigma_s^2 ds \right)^2 \bigg| \sigma, J \right].$$

Assumptions:

\( \sigma_t > 0, \ \forall \ t, \) and \( \sigma \) and \( J \) are independent of \( W \).
Suppose that $\sigma_t \equiv \sigma$ is constant and $J$ is an arbitrary finite jump activity process. Then, as $n \to \infty$, the optimal threshold $\varepsilon_{n}^{**}$ is such that

$$
\varepsilon_{n}^{**} \sim \sqrt{2\sigma^2 h_n \log \left( \frac{1}{h_n} \right)}
$$

For time-varying volatilities $t \to \sigma_t$, we have:

$$
\varepsilon_{n}^{**} \sim \sqrt{2\bar{\sigma}^2 h_n \ln(1/h_n)}, \quad n \to \infty,
$$

where

$$
\bar{\sigma} := \max_{s \in [0, T]} \sigma_s.
$$
A Feasible Implementation of $\varepsilon_{n}^{**}$

(i) Get a “rough” estimate of $\sigma^2$ via the realized QV:

$$\tilde{\sigma}^2_{n,0} := \frac{1}{T} \sum_{i=1}^{n} |X_{t_i} - X_{t_{i-1}}|^2$$
A Feasible Implementation of $\varepsilon_{n}^{**}$

(i) Get a “rough” estimate of $\sigma^2$ via the realized QV:

$$\tilde{\sigma}_{n,0}^2 := \frac{1}{T} \sum_{i=1}^{n} |X_{ti} - X_{ti-1}|^2$$

(ii) Use $\tilde{\sigma}_{n,0}^2$ to estimate the optimal threshold

$$\hat{\varepsilon}_{n,0} := \left(2 \tilde{\sigma}_{n,0}^2 h_n \log(1/h_n) \right)^{1/2}$$
A Feasible Implementation of $\varepsilon_{n}^{**}$

(i) Get a “rough” estimate of $\sigma^2$ via the realized QV:

$$\tilde{\sigma}_{n,0}^2 := \frac{1}{T} \sum_{i=1}^{n} |X_{t_i} - X_{t_i-1}|^2$$

(ii) Use $\tilde{\sigma}_{n,0}^2$ to estimate the optimal threshold

$$\hat{\varepsilon}_{n,0}^{**} := \left(2 \tilde{\sigma}_{n,0}^2 h_n \log(1/h_n)\right)^{1/2}$$

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

$$\tilde{\sigma}_{n,1}^2 = \frac{1}{T} \sum_{i=1}^{n} |X_{t_i} - X_{t_i-1}|^2 \mathbf{1}[|X_{t_i} - X_{t_i-1}| \leq \hat{\varepsilon}_{n,0}^{**}]$$
A Feasible Implementation of $\varepsilon^{**}_n$

(i) Get a “rough” estimate of $\sigma^2$ via the realized QV:

$$\tilde{\sigma}^2_{n,0} \coloneqq \frac{1}{T} \sum_{i=1}^{n} |X_{t_i} - X_{t_{i-1}}|^2$$

(ii) Use $\tilde{\sigma}^2_{n,0}$ to estimate the optimal threshold

$$\hat{\varepsilon}^{**}_{n,0} \coloneqq \left( 2 \tilde{\sigma}^2_{n,0} h_n \log(1/h_n) \right)^{1/2}$$

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

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

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

$$\tilde{\sigma}^2_{n,0} \to \hat{\varepsilon}^{**}_{n,0} \to \tilde{\sigma}^2_{n,1} \to \hat{\varepsilon}^{**}_{n,1} \to \tilde{\sigma}^2_{n,2} \to \cdots \to \tilde{\sigma}^2_{n,\infty}$$
Illustration II. Continued...

Log-Normal Merton Model

Performace of Truncated Realized Variations

Figure: (left) Merton Model with $\lambda = 1000$. Red dot is $\hat{\sigma}_{n,1} = 0.336$, while purple dot is the limiting $\hat{\sigma}_{n,k} = 0.215$. Orange square is $\tilde{\sigma}_{n,1} = 0.225$, while brown square is the limiting estimator $\tilde{\sigma}_{n,\infty} = 0.199$
### Monte Carlo Simulations

<table>
<thead>
<tr>
<th>Estimator</th>
<th>$\hat{\sigma}$</th>
<th>std($\hat{\sigma}$)</th>
<th>$\bar{\varepsilon}$</th>
<th>$\bar{N}$ (# of iterations)</th>
</tr>
</thead>
<tbody>
<tr>
<td>RQV</td>
<td>0.3921</td>
<td>0.0279</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\hat{\sigma}_{n,1}$</td>
<td>0.29618</td>
<td>0.02148</td>
<td>0.01523</td>
<td>1</td>
</tr>
<tr>
<td>$\hat{\sigma}_{n,\infty}$</td>
<td>0.23</td>
<td>0.0108</td>
<td>0.00892</td>
<td>5.86</td>
</tr>
<tr>
<td>$\tilde{\sigma}_{n,1}$</td>
<td>0.265</td>
<td>0.0163</td>
<td>0.0124</td>
<td>1</td>
</tr>
<tr>
<td>$\tilde{\sigma}_{n,\infty}$</td>
<td>0.211</td>
<td>0.00588</td>
<td>0.00671</td>
<td>5.10</td>
</tr>
<tr>
<td>BPV</td>
<td>0.2664</td>
<td>0.0129</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Table:** Estimation of the volatility $\sigma = 0.2$ for a log-normal Merton model based on 10000 simulations of 5-minute observations over a 1 month time horizon. The jump parameters are $\lambda = 1000$, $\sigma^{\text{Jmp}} = 1.5\sqrt{h}$ and $\mu^{\text{Jmp}} = 0$. Loss is the number of jump misclassifications and $N$ is the number of iterations. bar is used to denote average.
Main Breakthroughs

1. Introduced for the first time an objective threshold selection procedure based on arguments of statistical optimality.
2. Developed the infill asymptotic characterization of the optimal thresholds.
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 Research

1. In principle, we can apply the constant-volatility method for varying volatility $t \rightarrow \sigma_t$ by localization; i.e., applying it to periods where $\sigma$ is approximately constant.
In principle, we can apply the constant-volatility method for varying volatility $t \rightarrow \sigma_t$ by localization; i.e., applying it to periods where $\sigma$ is approximately constant.

This could also help us to estimate $\bar{\sigma} = \max_{s \in [0, T]} \sigma_s$ and, then, apply the asymptotic equivalence

$$\varepsilon_{n}^{**} \sim \sqrt{2\bar{\sigma}^2 h_n \ln(1/h_n)}.$$
Ongoing and Future Research

3. It can be shown that for a Lévy jump process $J$:

- If $J$ has finite jump activity, $\bar{\varepsilon}_n \sim \sqrt{\frac{2}{\sigma^2}} h_n \log \left(\frac{1}{h_n}\right)$

- But, surprisingly, if $J$ is a $Y$-stable Lévy process (IA), $\bar{\varepsilon}_n \sim \sqrt{\left(2 - Y\right) \sigma^2} h_n \log \left(\frac{1}{h_n}\right)$

Thus the higher the jump activity is, the lower the optimal threshold has to be to discard the higher noise represented by the jumps.

Problem: Can we generalize it to Lévy processes with stable-like jumps?
Ongoing and Future Research

It can be shown that for a Lévy jump process $J$:

- if $J$ has finite jump activity,

$$\tilde{\varepsilon}_n^* \sim \sqrt{2\sigma^2 h_n \log \left( \frac{1}{h_n} \right)}$$

- But, surprisingly, if $J$ is a $Y$-stable Lévy process (IA),

$$\tilde{\varepsilon}_n^* \sim \sqrt{2\sigma^2 h_n \log \left( \frac{1}{h_n} \right)}$$

Thus the higher the jump activity is, the lower the optimal threshold has to be to discard the higher noise represented by the jumps.

Problem: Can we generalize it to Lévy processes with stable like jumps?
It can be shown that for a Lévy jump process $J$:

- if $J$ has finite jump activity,
  \[ \bar{\varepsilon}_n^* \sim \sqrt{2\sigma^2 h_n \log \left( \frac{1}{h_n} \right)} \]

- But, surprisingly, if $J$ is a $Y$-stable Lévy process (IA),
  \[ \bar{\varepsilon}_n^* \sim \sqrt{(2 - Y)\sigma^2 h_n \log \left( \frac{1}{h_n} \right)} \]

Thus the higher the jump activity is, the lower the optimal threshold has to be to discard the higher noise represented by the jumps.
It can be shown that for a Lévy jump process $J$:

- if $J$ has finite jump activity,
  \[
  \bar{\varepsilon}_n^* \sim \sqrt{2\sigma^2 h_n \log\left(\frac{1}{h_n}\right)}
  \]

- But, surprisingly, if $J$ is a $Y$-stable Lévy process (IA),
  \[
  \bar{\varepsilon}_n^* \sim \sqrt{(2 - Y)\sigma^2 h_n \log\left(\frac{1}{h_n}\right)}
  \]

Thus the higher the jump activity is, the lower the optimal threshold has to be to discard the higher noise represented by the jumps.

Problem: Can we generalize it to Lévy processes with stable like jumps?
Outline

1. Overview Of Three Applications
   - Inference Of Stochastic Processes Using High-Frequency Data
   - Model Selection & Calibration via Short-Time Option Prices
   - Change-Point Detection In Continuous-Time

2. A Closer Look At The Problem Of Optimal Threshold Selection

3. A Closer Look At Optimal Kernel Estimation of Spot Volatility
The Problem

- We consider the SDE
  \[ dX_t = \mu_t dt + \sigma_t dB_t \quad (1) \]
  where \( B := \{B_t\}_{t \geq 0} \) is a standard Brownian Motion (BM).
- Want to estimate the spot volatility \( \sigma_t \) at a fixed time \( t \in (0, T) \) based on a discrete observations \( X_{t_i} \), where \( t_i = i\Delta = Ti/n, i = 0, 1, \ldots, n \).
- We consider the class of kernel estimators:
  \[ \hat{\sigma}_{t,n,h}^2 := \sum_{i=1}^{n} K_h(t_{i-1} - t)(\Delta_i X)^2, \]
  where \( \Delta_i X := X_{t_i} - X_{t_{i-1}} \) and \( K_h(x) := \frac{1}{h} K \left( \frac{x}{h} \right) \);
- Problem: What are the optimal bandwidth \( h \) and kernel \( K \)?
Literature Review

- Foster and Nelson (1996) assumes that $h_n = c \Delta_n^{1/2}$ and finds the optimal constant $c$; argues that $K(x) = \frac{1}{2} e^{-|x|}$ is optimal;

- Kristensen (2010) considers the problem of optimal bandwidth selection under a path-wise Hölder condition of the form:

$$P-a.e. \omega : |\sigma_{t+\delta}^2(\omega) - \sigma_t^2(\omega)|^2 = L_t(\delta; \omega) \delta^\gamma + o(\delta^\gamma), \forall t,$$  \hspace{1cm} (**) 

where $\gamma \in (0, 2]$ and $\delta \to L_t(\delta)$ slowly varying at 0;

- Under (**), he proposes the following optimal bandwidth:

$$h_{n,t}^{opt} = n^{-\frac{1}{\gamma+1}} \left( \frac{2 T \sigma_t^4 \| K^2 \|_1}{\gamma L_t(0^+)} \right)^{\frac{1}{\gamma+1}};$$

- However, (**) is hard to verify with $L_t(0^+) \in (0, \infty)$. 
Key Assumption on the Volatility Process

Assumption (⋆)

The variance process $V := \{V_t = \sigma_t^2 : t \geq 0\}$ satisfies

$$
\mathbb{E}[(V_{t+r} - V_t)(V_{t+s} - V_t)] = L(t) C_\gamma(r, s) + o((r^2 + s^2)^{\gamma/2}), \quad r, s \to 0,
$$

for some $\gamma > 0$ and certain locally bounded function $L : \mathbb{R}_+ \to \mathbb{R}_+$ and a function $C_\gamma : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$ that is not identically zero and satisfies the scaling property:

$$C_\gamma(hr, hs) = h^\gamma C_\gamma(r, s), \quad \text{for } r, s \in \mathbb{R}, \ h \in \mathbb{R}_+.$$
Examples I: Differentiable Volatility Processes

Proposition

If $V_t = \sigma_t^2 = f(t)$ for a differentiable function $f$ s.t. $f'(t) \neq 0$, then

$$\mathbb{E}[(V_{t+r} - V_t)(V_{t+s} - V_t)] = \mathbb{E}[(f(t + r) - f(t))(f(t + s) - f(t))]$$

$$= \mathbb{E}[(f'(t)r + o(r)) \times (f'(t)s + o(s))]$$

$$= \mathbb{E}[(f'(t))^2]rs + o((r^2 + s^2)^{\gamma/2}),$$

and ($\star$) holds with

$$L(t) = \mathbb{E}[(f'(t))^2], \quad C_\gamma(r, s) = rs, \quad \gamma = 2.$$
Example II: BM Type Volatility Processes

Proposition

Suppose that $V_t = \sigma_t^2$ satisfies the SDE:

$$dV_t = f(t)dt + g(t)dW_t, \quad t \in [0, T],$$

where $W$ is a standard Brownian Motion. Then,

$$\mathbb{E}[(V_{t+r} - V_t)(V_{t+s} - V_t)] = \mathbb{E}\left[ \int_t^{t+r} g(u) dW_u \int_t^{t+s} g(u) dW_u \right] + \text{h.o.t.} \quad (\star)$$

$$= \mathbb{E}\left[ g^2(t)(W_{t+r} - W_t)(W_{t+s} - W_t) \right] + \text{h.o.t.}$$

$$= L(t) C_{\gamma}(r, s) + \text{h.o.t.},$$

with

$$L(t) = \mathbb{E}[g^2(t)], \quad C_{\gamma}(r, s) := \min\{|r|, |s|\} 1_{\{rs \geq 0\}}, \quad \gamma = 1.$$
Example III: fBM Type Volatility

Proposition

Consider a process \( \{ Y^H_t \} \) \( t \geq 0 \) that satisfies

\[
Y^H_t = \int_{-\infty}^{t} f(u) dB^H_u,
\]

where \( \{ B^H_u \} \) \( u \in \mathbb{R} \) is a (two-sided) fractional Brownian motion with Hurst parameter \( H \in (\frac{1}{2}, 1) \).

Then, for any \( C^2 \)-function \( g \), \( V_t = g(Y^H_t) \) satisfies the key Assumption (*) with \( \gamma = 2H \in (1, 2) \) and

\[
C_{\gamma}(r,s) := \mathbb{E}[B^H_r B^H_s] = \frac{1}{2}(|r|^{2H} + |s|^{2H} - |r-s|^{2H}), \quad r, s \in \mathbb{R}.
\]
Main Result

Theorem (F-L & Li 2017)

Under the Key Assumption and other mild conditions, 

\[ \text{MSE}_{\tau, n}(h) = \mathbb{E} \left[ \left( \hat{\sigma}_{\tau, n, h}^2 - \sigma_{\tau}^2 \right)^2 \right] = 2 \frac{\Delta h}{h} \mathbb{E}[\sigma_{\tau}^4] \int K^2(x) dx + h^\gamma L(\tau) \int \int K(x) K(y) C_{\gamma}(x, y) dxdy + o \left( \frac{\Delta h}{h} \right) + o(h^\gamma); \]

with an analogous asymptotic expansion for \( \text{IMSE}_n(h) = \int_0^T \mathbb{E}[(\hat{\sigma}_{\tau}^2 - \sigma_{\tau}^2)^2] d\tau \), but replacing \( \mathbb{E}[\sigma_{\tau}^4] \) and \( L(\tau) \) with its integrated versions \( \int_0^T \mathbb{E}[\sigma_{\tau}^4] d\tau \) and \( \int_0^T L(\tau) d\tau \).
Approximation of the Optimal Local Bandwidth

Proposition (F-L & Li 2017)

The approximated local optimal bandwidth, which, by definition, minimizes the leading order terms of the MSE, is given by

\[ h_{n, \tau}^{a, opt} = n^{-\frac{1}{\gamma+1}} \left[ \frac{2 \hat{T} \mathbb{E} \left[ \sigma_\tau^4 \right] \|K^2\|_1}{\gamma L(\tau) \int \int K(x)K(y)C_\gamma(x, y) \, dx \, dy} \right]^{\frac{1}{\gamma+1}}, \]

resulting in an approximation of the minima MSE of the form:

\[ MSE_{n, \tau}^{a, opt} = n^{-\frac{1}{\gamma+1}} \left( 1 + \frac{1}{\gamma} \right) \left( 2 \hat{T} \mathbb{E} \left[ \sigma_\tau^4 \right] \|K^2\|_1 \right)^{\frac{\gamma}{\gamma+1}} \times \left( \gamma L(\tau) \int \int K(x)K(y)C_\gamma(x, y) \, dx \, dy \right)^{\frac{1}{\gamma+1}}. \]
In the case that we consider the Integrated MSE (IMSE)

\[ \text{IMSE}_n(h) = \int_0^T \mathbb{E}[(\hat{\sigma}_{\tau,n,h}^2 - \sigma_{\tau}^2)^2] d\tau \]

the optimal (uniform) bandwidth takes the form:

\[
h_n^{a,\text{opt}} = n^{-\frac{1}{\gamma+1}} \left[ \frac{2T \int_0^T \mathbb{E}[\sigma_{\tau}^4] d\tau \| K^2 \|_1}{\gamma \int_0^T L(\tau) d\tau \int \int K(x)K(y)C_{\gamma}(x,y) dx dy} \right]^{\frac{1}{\gamma+1}}
\]
Optimal Kernel Selection

The leading order term of the MSE when $h$ is optimal is

$$\text{MSE}_{n, \text{opt}}(K) = n^{-\frac{\gamma}{t+\gamma}} \left(1 + \frac{1}{\gamma}\right) \left(2T \mathbb{E}[\sigma_\tau^4] \int K^2(x)dx\right)^{\frac{\gamma}{t+\gamma}}$$

$$\times \left(\gamma L(\tau) \iint K(x)K(y)C_\gamma(x, y)dxdy\right)^{\frac{1}{1+\gamma}}.$$  

To choose $K$, it is then natural to consider the calculus of variation problem:

$$\min_K \left(\int K^2(x)dx\right)^{\gamma} \iint K(x)K(y)C_\gamma(x, y)dxdy,$$

subject to the restriction $\int K(x)dx = 1.$
BM Volatility Case

**Theorem (F-L & Li 2017)**

For Brownian-driven volatilities, when $C_\gamma(r, s) := (|r| \wedge |s|)1_{\{rs \geq 0\}}$, the optimal kernel function is the exponential kernel:

$$K_{\text{exp}}(x) = \frac{1}{2} e^{-|x|}, \quad x \in \mathbb{R}.$$ 

**Remark:** Other two common kernels are the uniform $K_0(x) = \frac{1}{2} 1_{\{|x|<1\}}$ and the Epanechnikov $K_2(x) = \frac{3}{4}(1 - x^2)1_{\{|x|<1\}}$ kernels; As it turns out

$$\frac{\text{MSE}_{n, \text{opt}}^a(K_{\text{exp}})}{\text{MSE}_{n, \text{opt}}^a(K_0)} = 0.86; \quad \frac{\text{MSE}_{n, \text{opt}}^a(K_{\text{exp}})}{\text{MSE}_{n, \text{opt}}^a(K_2)} = 0.93;$$
Plug-In Type Bandwidth Selection Methods I

- We propose a plug-in type of implementation;
- Consider the BM type volatility processes: \( dV_t = f(t)dt + g(t)dW_t \).
- The optimal bandwidth that minimizes the approximate IMSE is given by
  \[
  h_{n,\text{opt}}^a = \left[ \frac{2T \int_0^T \mathbb{E}[\sigma_t^4]dt \int K^2(x)dx}{n \int_0^T L(t)dt \int \int K(x)K(y)C_1(x,y)dx\,dy} \right]^{1/2}.
  \]
- We need to estimate \( \int_0^T \mathbb{E}[\sigma_t^4]dt \) and \( \int_0^T L(t)dt = \int_0^T \mathbb{E}[g^2(t)]dt \), the expected quarticity and integrated vol vol.
- Given that we have at hand only one realization of \( X \), it is natural to estimate these two quantities with \( \int_0^T \sigma_t^4 dt \) and \( \int_0^T g^2(t)dt \);
Plug-In Type Bandwidth Selection Methods II

- $\int_0^T \sigma_t^4 dt$ can be estimated by the Realized Quarticity:
  \[ \hat{IQ} = (3\Delta)^{-1} \sum_{i=1}^n (\Delta_i X)^4. \]

- Estimation of $\int_0^T g^2(t) dt$, which is just the quadratic variation of $\sigma^2$, $\langle \sigma^2, \sigma^2 \rangle_T$, is more involved.

- $\langle \sigma^2, \sigma^2 \rangle_T = \int_0^T g^2(t) dt$ is sometimes called \textit{volatility of volatility} or simply \textit{vol vol}.
Zhang et al. (2005) proposed a Two-time Scale Realized Volatility (TSRV) estimator of the quadratic variation $\langle Y, Y \rangle_T$ of a process $Y$ in the presence of market “micro-structure” noise:

$$\text{TSRV} = \frac{1}{k} \sum_{i=0}^{n-k} (Y_{t_i+k} - Y_{t_i})^2 - \frac{n - k + 1}{nk} \sum_{i=0}^{n-1} (Y_{t_{i+1}} - Y_{t_i})^2.$$

Inspired by this, we propose the following estimator:

$$\widehat{IVV}^{(tsrvv)}_T = \frac{1}{k} \sum_{i=b}^{n-k-b} (\hat{\sigma}_{t_{i+k}}^2 - \hat{\sigma}_{t_i}^2)^2 - \frac{n - k + 1}{nk} \sum_{i=b+k-1}^{n-k-b} (\hat{\sigma}_{t_{i+1}}^2 - \hat{\sigma}_{t_i}^2)^2.$$
The TSRVV involves the estimation of spot volatility, which we do not know in advance, so it is natural to consider the following iterative algorithm:

**The Iterative Plug-in Bandwidth Selection Algorithm:**

- **Data:** \( \Delta_n^1 X = X_{t_1} - X_{t_0}, \ldots, \Delta_n^n X = X_{t_n} - X_{t_{n-1}} \);
- Set an initial value of \( h \); 
- **while** Stopping criteria not met **do**
  - Estimate \( \hat{\sigma}_{t_i}^2 \) for all \( 0 \leq i \leq n \) based on the bandwidth \( h \);
  - Estimate the integrated vol vol \( \langle \sigma^2, \sigma^2 \rangle \) using the TSRVV;
  - Update the approximated optimal bandwidth \( h \);
- **end**

In our simulations, two iterations are typically enough for satisfactory result, even with bad initial guess.
Simulation Study

- We consider Heston model
  \[
  dX_t = \mu_t \, dt + \sqrt{V_t} \, dB_t, \\
  dV_t = \kappa (\theta - V_t) \, dt + \xi \sqrt{V_t} \, dW_t,
  \]
  with the following parameter settings:
  1. \( T = 5 \) or 21 days, \( \Delta = 1 \) or 5 minute, 6.5 trading hours.
  2. \( \mu_t = \frac{1}{2} - \frac{1}{2} V_t \), \( \sigma_0 = 0.2 \), \( \kappa = 5 \), \( \theta = 0.04 \), \( \xi = 0.5 \).
  3. The leverage is taken to be 0 and \(-0.5\).

- Except when we compare different kernel functions, we use the exponential kernel function.

- We will estimate the sample Mean of the Average Squared Error of the estimators based on 2000 simulations:
  \[
  MASE := \frac{1}{n - 2 \ell + 1} \sum_{i=\ell}^{n-\ell} (\hat{\sigma}_{ti}^2 - \sigma_{ti}^2)^2, \quad \ell = 0.1n.
  \]
### Plug-In v.s. Cross-Validation

#### 5 Days Data

<table>
<thead>
<tr>
<th>nData/h</th>
<th>$\rho$</th>
<th>$MASE_{Pl}$</th>
<th>$MASE_{CV}$</th>
<th>$MASE_{\text{oracle}}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>12</td>
<td>0</td>
<td>1.0796E-07</td>
<td>1.3386E-07</td>
<td>9.1266E-08</td>
</tr>
<tr>
<td>60</td>
<td>0</td>
<td>7.1439E-09</td>
<td>8.0542E-09</td>
<td>6.7286E-09</td>
</tr>
<tr>
<td>12</td>
<td>-0.5</td>
<td>1.0296E-07</td>
<td>1.4180E-07</td>
<td>9.2620E-08</td>
</tr>
<tr>
<td>60</td>
<td>-0.5</td>
<td>7.3872E-09</td>
<td>8.2567E-09</td>
<td>6.9356E-09</td>
</tr>
</tbody>
</table>

#### 21 Days Data

<table>
<thead>
<tr>
<th>nData/h</th>
<th>$\rho$</th>
<th>$MASE_{Pl}$</th>
<th>$MASE_{CV}$</th>
<th>$MASE_{\text{oracle}}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>12</td>
<td>0</td>
<td>1.9088E-08</td>
<td>2.1221E-08</td>
<td>1.8265E-08</td>
</tr>
<tr>
<td>60</td>
<td>0</td>
<td>1.7064E-09</td>
<td>1.6868E-09</td>
<td>1.5984E-09</td>
</tr>
<tr>
<td>12</td>
<td>-0.5</td>
<td>1.9039E-08</td>
<td>1.9495E-08</td>
<td>1.7587E-08</td>
</tr>
<tr>
<td>60</td>
<td>-0.5</td>
<td>1.6652E-08</td>
<td>1.6011E-08</td>
<td>1.5509E-08</td>
</tr>
</tbody>
</table>
## Estimation of Volatility of Volatility

<table>
<thead>
<tr>
<th>nData/h</th>
<th>$\rho$</th>
<th>$\xi$</th>
<th>Bias</th>
<th>Std</th>
<th>$\sqrt{MSE}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>12</td>
<td>0</td>
<td>0.2</td>
<td>-0.0006</td>
<td>0.0990</td>
<td>0.0990</td>
</tr>
<tr>
<td>12</td>
<td>0</td>
<td>0.5</td>
<td>-0.0584</td>
<td>0.1979</td>
<td>0.2063</td>
</tr>
<tr>
<td>60</td>
<td>0</td>
<td>0.2</td>
<td>-0.0122</td>
<td>0.0772</td>
<td>0.0782</td>
</tr>
<tr>
<td>60</td>
<td>0</td>
<td>0.5</td>
<td>-0.0411</td>
<td>0.1549</td>
<td>0.1603</td>
</tr>
<tr>
<td>12</td>
<td>-0.5</td>
<td>0.2</td>
<td>-0.0002</td>
<td>0.0987</td>
<td>0.0987</td>
</tr>
<tr>
<td>12</td>
<td>-0.5</td>
<td>0.5</td>
<td>-0.0571</td>
<td>0.1984</td>
<td>0.2065</td>
</tr>
<tr>
<td>60</td>
<td>-0.5</td>
<td>0.2</td>
<td>-0.0138</td>
<td>0.0779</td>
<td>0.0791</td>
</tr>
<tr>
<td>60</td>
<td>-0.5</td>
<td>0.5</td>
<td>-0.0443</td>
<td>0.1551</td>
<td>0.1613</td>
</tr>
</tbody>
</table>

**Table:** Estimation of Volatility of Volatility by TSRVV (1 month data, 10000 sample paths)
• We consider four different kernels:

\[ K_{\text{exp}}(x) = \frac{1}{2} e^{-|x|}, \quad K_0(x) = \frac{1}{2} 1_{\{|x|<1\}} \]

\[ K_1(x) = |1 - x| 1_{\{|x|<1\}}, \quad K_2(x) = \frac{3}{4} (1 - x^2) 1_{\{|x|<1\}} \]

<table>
<thead>
<tr>
<th>length</th>
<th>( \rho )</th>
<th>( K_{\text{exp}} )</th>
<th>( K_0 )</th>
<th>( K_1 )</th>
<th>( K_2 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>5 days</td>
<td>0</td>
<td>2.597E-05</td>
<td>2.872E-05</td>
<td>2.644E-05</td>
<td>2.708E-05</td>
</tr>
<tr>
<td>5 days</td>
<td>-0.5</td>
<td>2.523E-05</td>
<td>2.825E-05</td>
<td>2.575E-05</td>
<td>2.649E-05</td>
</tr>
<tr>
<td>21 days</td>
<td>0</td>
<td>2.340E-05</td>
<td>2.804E-05</td>
<td>2.498E-05</td>
<td>2.591E-05</td>
</tr>
<tr>
<td>21 days</td>
<td>-0.5</td>
<td>2.369E-05</td>
<td>2.860E-05</td>
<td>2.524E-05</td>
<td>2.617E-05</td>
</tr>
</tbody>
</table>

**Table:** Comparison of Different Kernel Functions (5 min data, 2000 sample paths)
Conclusions

1. An optimal bandwidth selection method is proposed under a mild scaling condition on the local behavior of the covariance function of the variance process.

2. The considered framework covers a wide range of models including volatility models driven by BM and fBM.

3. The problem of optimal kernel selection is also considered: it is shown that an exponential kernel is the optimal kernel function for B.M.-driven volatility models.

4. Fast iterated plug-in type algorithms are also devised as a way to implement the proposed optimal selection methods.
Future Work - Market Micro-Structure Noise

- Assume the log price with additive noise is observed:

\[
dX_t = \mu_t dt + \sigma_t dB_t, \\
Y_t = X_t + \varepsilon_t,
\]

where \( \{\varepsilon_t\}_t \) are iid with mean zero and variance \( \omega^2 \).

- A natural way to handle such a case is to replace the quadratic variation

\[
[X, X]_t = \int_0^t \sigma_s^2 ds
\]

below

\[
\bar{\sigma}_t^2 = \int_0^t K_h(s - t) d[X, X]_s,
\]

with an integrated variance estimator \([\widehat{X}, X]_t\) that is robust against microstructure noise.
Further Reading I

- J.E. Figueroa-López and S. Ólafsson, Change-point detection for Levy processes. To Appear in Annals of Applied Probability, 2018+
Further Reading II

J.E. Figueroa-López and S. Ólafsson,
Short-time expansions for close-to-the-money options under a Lévy jump model with stochastic volatility.

J.E. Figueroa-López and S. Ólafsson,
Short-time asymptotics for the implied volatility skew under a stochastic volatility model with Lévy jumps.
Finance & Stochastics 20(4), 973-1020, 2016b.

J.E. Figueroa-López, R. Gong, and C. Houdré,
Third-Order Short-Time Expansions for Close-to-the-Money Option Prices Under the CGMY Model.
To Appear in Journal of Applied Mathematical Finance, 2018+
Optimally Thresholded Realized Power Variations for Lévy Jump Diffusion Models.

J.E. Figueroa-López & C. Mancini.
Optimum thresholding using mean and conditional mean square error.

Optimal iterative threshold-kernel estimation of jump diffusion processes.
In preparation, 2018.