Optimal Kernel Estimation of Spot Volatility of SDE

José E. Figueroa-López

Department of Mathematics
Washington University in St. Louis

figueroa@math.wustl.edu

(Joint work with Cheng Li from Purdue U.)

INFORMS Applied Probability Conference
Northwestern, Chicago

July 10, 2017
Overview

1. Framework, Estimator, and Assumptions

2. Optimal Bandwidth Selection

3. Optimal Kernel Selection

4. Implementation of the Bandwidth Selection Method

5. Simulation Study

6. Conclusions
1. Framework, Estimator, and Assumptions

2. Optimal Bandwidth Selection

3. Optimal Kernel Selection

4. Implementation of the Bandwidth Selection Method

5. Simulation Study

6. Conclusions
Consider the following Stochastic Differential Equation (SDE):

\[ dX_t = \mu_t \, dt + \sigma_t \, dB_t \]  \hspace{1cm} (1)

where \( B := \{B_t\}_{t \geq 0} \) is a standard Brownian Motion (BM).

In Financial settings, \( X_t \) represents the log return \( \log(S_t) \) of an asset with price process \( \{S_t\}_{t \geq 0} \), while \( \sigma_t \) and \( \mu_t \) represent the spot volatility and mean rate of return at time \( t \);
We revisit the problem of estimating the spot volatility $\sigma_\tau$ at a fixed time $\tau$ based on a discrete record of observations $X_{t_i}$, $i = 0, 2, \ldots, n$. For simplicity, we take a regular sampling scheme: $\Delta := t_i - t_{i-1}$.

We consider the class of kernel estimators:

$$\hat{\sigma}^2_{\tau, n, h} := \sum_{i=1}^{n} K_h(t_{i-1} - \tau)(\Delta_i X)^2,$$

where, as usual,

$$\Delta_i X := X_{t_i} - X_{t_{i-1}}, \quad K_h(x) := \frac{1}{h} K \left( \frac{x}{h} \right)$$

for a suitable bandwidth $h := h_n$ and a kernel function $K : \mathbb{R} \to \mathbb{R}$ such that $\int K(x)dx = 1$. 
Consistency and CLT’s have been established under relatively mild conditions (e.g., Foster and Nelson (1996), Fan and Wang (2008));

Most of these works only assume some asymptotic conditions on the bandwidth \( h \);

However, in finite sample settings, the bandwidth \( h \) significantly affects the performance of the estimator.

In this work, we study the problem of optimal bandwidth and kernel selection, which has received little attention in the literature;

We aim to impose conditions that cover a wide range of models: from the traditional Brownian driven volatility models (Heston, OU, etc) to even those driven by fractional Brownian motions;

The proposed methods should be implementable and computationally efficient as they are meant for high-frequency data;
Comparison to Key Related Works

- Foster and Nelson (1996) assumes that $h = cn^{-1/2}$ and finds the optimal constant $c$; they also conjecture that the exponential kernel $K(x) = \frac{1}{2} e^{-|x|}$ is optimal;

- Kristensen (2010) considers the problem of bandwidth selection but under the following path-wise Hölder condition:

$$
\mathbb{P} - \text{a.e. } \omega : |\sigma_{t+\delta}^2(\omega) - \sigma_t^2(\omega)|^2 = L_t(0; \omega) \delta^{\gamma} + o(\delta^\gamma), \forall t,
$$

where $\gamma \in (0, 2]$;

- Under (■), it proposes the following optimal bandwidth:

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

However, the Assumption (■) is hard to verify with explicit $L_\tau(0) \in (0, \infty)$. 

José E. Figueroa-López (WashU)
Our first assumption is a simplifying non-leverage assumption (also used in Kristensen, 2010):

**Assumption 1**

\((\mu, \sigma)\) is independent of \(B\).

Another assumption is the boundedness of the moments of \(\mu\) and \(\sigma\) up to 4th degree.

**Assumption 2**

There exists \(M_T > 1\) such that \(\mathbb{E}[\mu_t^4 + \sigma_t^4] < M_T\), for all \(0 \leq t \leq T\).
Our Assumptions on the Volatility Process II

The following is the key assumption that we need for our purpose:

Assumption 3 (⋆)

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

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

for some \( \gamma > 0 \) and certain functions \( L : \mathbb{R}_+ \to \mathbb{R}_+ \) and \( C_\gamma : \mathbb{R} \times \mathbb{R} \to \mathbb{R} \) such that \( C_\gamma \) is not identically zero and has the scaling property:

\[
C_\gamma(hr, hs) = h^\gamma C_\gamma(r, s), \quad \text{for } r, s \in \mathbb{R}, h \in \mathbb{R}_+.
\]
Remarks

- It is not hard to see that $\gamma > 0$ and $C_\gamma(r, s; t) := L(t)C_\gamma(r, s)$ are uniquely defined. Furthermore, $C_\gamma(\cdot, \cdot)$ is non-negative definite; i.e.,

$$\int \int K(r)K(s)C_\gamma(r, s)drds \geq 0, \quad \forall K : \mathbb{R} \to \mathbb{R}.$$

- The condition $(\star)$ imposes is a local scaling property on the covariance function of the variance process and not on its paths;
- Though a covariance condition is more desirable than a pathwise condition, how restrictive is this assumption in practice?
- We will see that it is not and is satisfied by most of the volatility models in the literature.
Examples I: Deterministic Volatility Processes

The first class of volatility processes that satisfy the Key Assumption (⋆),

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

is a differentiable deterministic volatility:

**Proposition 1**

Let \( f(t) \), \( 0 \leq t \leq T \), be differentiable at \( t \) and \( f'(t) \neq 0 \). Then, the squared volatility process \( V_t = \sigma_t^2 = f(t) \) satisfies (⋆) with

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

Proposition 2

Suppose that \( V_t = \sigma^2(t) \) 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, the Key Assumption (\(*\)) is satisfied with

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

Lemma 1

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, both $Y^H_t$ and $\exp(Y^H_t)$ satisfy the key Assumption ($\star$) with $\gamma = 2H \in (1, 2)$ and

$$C_\gamma(r, s) := \mathbb{E}[B^H_r B^H_s] = \frac{1}{2}(r^{2H} + s^{2H} - |r - s|^{2H}), \quad r, s \in \mathbb{R}.$$
Conditions on the Kernel Function

Assumption 4

Given $\gamma$ and $C_\gamma$ as defined in our key Assumption ($\star$), we suppose that the kernel function $K : \mathbb{R} \to \mathbb{R}$ satisfies the following:

1. $\int K(x) \, dx = 1$,
2. $K(x)$ is piece-wise continuously differentiable,
3. $\int |K(x)||x|^\gamma \, dx < \infty$, $K(x)x^{\gamma+1} \to 0$, $|x| \to \infty$,
4. $\int\int K(x)K(y)C_\gamma(x,y) \, dx \, dy > 0$.

- The condition (4) above does not put substantial restriction on $K$ since $C_\gamma$ is already integrally non-negative definite.
- In the case of Brownian driven volatilities (when $C_\gamma(r, s) := (|r| \wedge |s|)1_{\{rs \geq 0\}}$), condition (4) holds for any nonzero $K$;
1. Framework, Estimator, and Assumptions

2. Optimal Bandwidth Selection

3. Optimal Kernel Selection

4. Implementation of the Bandwidth Selection Method

5. Simulation Study

6. Conclusions
Define the MSE and IMSE of the kernel estimator as

\[ \text{MSE}_n(h; \tau) = \mathbb{E}[(\hat{\sigma}_\tau^2 - \sigma_\tau^2)^2] \]

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

The key problem is:

Choose bandwidth \( h = h_n \) and kernel \( K \) to minimize \( \text{MSE}_n \) or \( \text{IMSE}_n \).

Remarks:

- The first error leads to an optimal local bandwidth selection (i.e., one depending on \( \tau \)) that is more desirable but is harder to implement;
- We propose to solve the problem in two-steps:
  1. first minimizing MSE in \( h \) for a fixed \( K \);
  2. then, minimizing the resulting minimal MSE in \( K \);
Main Result

**Theorem 1 (F-L & Li 2017)**

For $\mu$ and $\sigma$ satisfying Assumptions 1, 2 and 3 and a kernel function $K$ satisfying Assumption 4,

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

with an analogous asymptotic expansion for IMSE, 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 $\int_0^T L(\tau) d\tau$. 
Proposition 3 (F-L & Li 2017)

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

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

while the resulting minima value of the approximated MSE is given by

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

- When $\sigma_t^2 = f(t)$ is deterministic and smooth ($\gamma = 2$, $L(t) = (f'(t))^2$, and $C_\gamma(r, s) = rs$),

$$h_{n,T}^{opt} = n^{-\frac{1}{3}} \left( \frac{\text{ Tf}(t)^2 \| K \|_1}{f'(\tau)^2 \kappa^2(K)} \right)^{\frac{1}{3}}, \quad \kappa(K) := \int K(x) dx \neq 0;$$

$$\text{MSE}_n^{opt} = \frac{3}{2} n^{-\frac{1}{3}} \left( 2 \text{ Tf}(t)^2 \| K \|_1 \right)^{\frac{2}{3}} \left( 2f'(t)^2 \kappa^2(K) \right)^{\frac{1}{3}} + o(n^{-\frac{1}{3}});$$

- In particular, one can improve the rate of convergence $n^{-\frac{1}{3}}$ of the MSE by choosing $K$ such that $\kappa(K) = 0$;
Examples II

- For B.M.-driven volatilities \( d\sigma_t^2 = f(t)dt + g(t)dW_t \),

\[
h_{n,\tau}^{a, opt} = n^{-\frac{1}{2}} \left[ \frac{2T \mathbb{E}[\sigma_\tau^4]\|K\|_1}{\mathbb{E}[g^2(\tau)] \kappa_{BM}(K)} \right]^{\frac{1}{2}},
\]

\[
\kappa_{BM}(K) = \int_0^\infty \int_0^\infty [K(x)K(y) + K(-x)K(-y)] \min(x, y) \, dx \, dy,
\]

where the latter is always positive (regardless \( K \neq 0 \));

- The rate of convergence of the MSE, \( n^{-1/2} \), cannot be improved regardless of the choice of \( K \);
In the case that we consider the Integrated MSE (IMSE)

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

the optimal (uniform) bandwidth takes the form:

$$h_{n, opt} = n^{-\frac{1}{\gamma+1}} \left[ \frac{2T \int_0^T \mathbb{E}[\sigma_\tau^4] d\tau \| K_2 \|_1}{\gamma \int_0^T L(\tau) d\tau \int \int K(x)K(y)C_\gamma(x,y) dx dy} \right]^\frac{1}{\gamma+1}$$
1 Framework, Estimator, and Assumptions

2 Optimal Bandwidth Selection

3 Optimal Kernel Selection

4 Implementation of the Bandwidth Selection Method

5 Simulation Study

6 Conclusions
Recall that the approximated optimal MSE takes the form:

\[
\text{MSE}_{n, \text{opt}}^a(K) = n^{-\frac{\gamma}{1+\gamma}} \left( 1 + \frac{1}{\gamma} \right) \left( 2 T \mathbb{E}[\sigma_\tau^4] \int K^2(x) dx \right)^{\frac{\gamma}{1+\gamma}} \times \left( \gamma L(\tau) \int \int K(x)K(y)C_\gamma(x,y) dxdy \right)^{\frac{1}{1+\gamma}}.
\]

This leads to consider the following calculus of variation problem:

\[
\min_K \left( \int K^2(x) dx \right)^\gamma \int \int K(x)K(y)C_\gamma(x,y) dxdy,
\]

subject to the restriction \( \int K(x) dx = 1 \).
Theorem 2 (F-L & Li 2016)

The optimal kernel function is the exponential kernel:

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

Remark: 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

\[ \frac{\text{MSE}_{n}^{a\cdot\text{opt}}(K_{\text{exp}})}{\text{MSE}_{n}^{a\cdot\text{opt}}(K_0)} = 0.86; \quad \frac{\text{MSE}_{n}^{a\cdot\text{opt}}(K_{\text{exp}})}{\text{MSE}_{n}^{a\cdot\text{opt}}(K_2)} = 0.93; \]
1. Framework, Estimator, and Assumptions

2. Optimal Bandwidth Selection

3. Optimal Kernel Selection

4. Implementation of the Bandwidth Selection Method

5. Simulation Study

6. Conclusions
Generally, there are two types of bandwidth selection methods.

**Cross-validation methods:**
- Advantages: yield relatively good results for a wide range of volatility processes (regardless of $\gamma$);
- Disadvantages: time consuming, hard to implement.

**Plug-in type methods:**
- Advantages: usually faster and have better accuracy.
- Disadvantages: less general (i.e., it is tailored to some specific $\gamma$).

In Kristensen (2010), a leave-one-out cross validation method is proposed.

In this work, we consider a plug-in type estimation.
The idea of the plug-in method is to estimate all the parameters encountered in the explicit approximated optimal bandwidth; consider the BM type volatility processes: 

\[ dV_t = f(t)dt + g(t)dW_t. \]

The approximated (uniform) optimal bandwidth is then given by

\[
h_{n, \text{opt}} = \left[ \frac{2T \int_0^T \mathbb{E}[\sigma_t^4]dt \int K^2(x)dx}{n \int_0^T L(t)dt \int \int K(x)K(y)C_1(x, y)dxdy} \right]^{1/2}.
\]

We need to estimate \( \int_0^T \mathbb{E}[\sigma_t^4]dt \) and \( \int_0^T L(t)dt = \int_0^T \mathbb{E}[g^2(t)]dt \).

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 \( \int_0^T g^2(t)dt \);

\( \int_0^T \sigma_t^4 dt \) can be estimated by Realized Quarticity:

\[
\hat{IQ} = (3\Delta)^{-1} \sum_{i=1}^{n} (\Delta_i X)^4.
\]
Two-time Scale Realized Volatility of Volatility (TSRVV)

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

- 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 TSRVV estimator:

$$
\hat{\text{IVV}}_{T}^{(\text{tsrvv})} = \frac{1}{k} \sum_{i=b}^{n-k-b} (\hat{\sigma}^2_{t_{i+k}} - \hat{\sigma}^2_{t_i})^2 - \frac{n - k + 1}{nk} \sum_{i=b+k}^{n-k-b} (\hat{\sigma}^2_{t_{i+1}} - \hat{\sigma}^2_{t_i})^2.
$$
Theorem 3 (Consistency of TSRVV, F-L & Li 2017)

For any fixed \( t_b \in (0, T/2) \), the TSRVV is a consistent estimator of \( \int_{t_b}^{T-t_b} g_t^2 \, dt \).

The convergence rate is given by \( O_P \left( \frac{n^{1/4}}{k^{1/2}} \right) + O_P \left( \sqrt{\frac{k}{n}} \right) \).
Iterative Plug-in Bandwidth Selection

- 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:** \( \Delta_1^n X = X_{t1} - X_{t0}, \ldots, \Delta_n^n X = X_{tn} - X_{tn-1} \);

  Set an initial value of \( h \);

  **while** Stopping criteria not met **do**

  - Get \( \hat{\sigma}_t^2 \) for all \( 0 \leq i \leq n \) based on previous bandwidth \( h \);
  - Estimate the vol vol \( \langle \sigma^2, \sigma^2 \rangle \) using the new estimation of spot volatility;
  - Update the approximated optimal bandwidth \( h \);

  **end**

- In our simulations, two iterations are typically enough for satisfactory result, even with bad initial guess.
1. Framework, Estimator, and Assumptions

2. Optimal Bandwidth Selection

3. Optimal Kernel Selection

4. Implementation of the Bandwidth Selection Method

5. Simulation Study

6. Conclusions
Simulation Study

- We consider Heston model

\[ dX_t = \mu_t dt + \sqrt{V_t} dB_t, \]
\[ dV_t = \kappa (\theta - V_t) dt + \xi \sqrt{V_t} dW_t, \tag{2} \]

with the following parameter settings:

1. 5 or 21 trading days, 1 or 5 minute data, 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 show the sample Mean of the Average Squared Error of the estimators based on 2000 simulations:

\[ MASE := \frac{1}{n - 2l + 1} \sum_{i=l}^{n-l} (\hat{\sigma}_{t_i}^2 - \sigma_{t_i}^2)^2, \quad l = 0.1n. \]
## 5 Days Data

<table>
<thead>
<tr>
<th>nData/h</th>
<th>$\rho$</th>
<th>$MASE_{PI}$</th>
<th>$MASE_{CV}$</th>
<th>$MASE_{oracle}$</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>$MASE_{PI}$</th>
<th>$MASE_{CV}$</th>
<th>$MASE_{oracle}$</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-09</td>
<td>1.6011E-09</td>
<td>1.5509E-09</td>
</tr>
</tbody>
</table>
Estimation of Volatility of Volatility

<table>
<thead>
<tr>
<th>nData/h</th>
<th>ρ</th>
<th>ξ</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)
Compare Different Kernels

- 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>2.5974E-05</td>
<td>2.8721E-05</td>
<td>2.6441E-05</td>
<td>2.7085E-05</td>
</tr>
<tr>
<td>5 days</td>
<td>-0.5</td>
<td>2.5233E-05</td>
<td>2.8252E-05</td>
<td>2.5759E-05</td>
<td>2.6490E-05</td>
</tr>
<tr>
<td>21 days</td>
<td>0</td>
<td>2.3406E-05</td>
<td>2.8047E-05</td>
<td>2.4988E-05</td>
<td>2.5914E-05</td>
</tr>
<tr>
<td>21 days</td>
<td>-0.5</td>
<td>2.3692E-05</td>
<td>2.8603E-05</td>
<td>2.5248E-05</td>
<td>2.6173E-05</td>
</tr>
</tbody>
</table>

**Table:** Comparison of Different Kernel Functions (5 min data, 2000 sample paths)
1. Framework, Estimator, and Assumptions
2. Optimal Bandwidth Selection
3. Optimal Kernel Selection
4. Implementation of the Bandwidth Selection Method
5. Simulation Study
6. Conclusions
Conclusions

1. An optimal bandwidth selection method is put forward under a new assumption 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.
Fan and Wang.
Spot volatility estimation for high-frequency data.  

Foster and Nelson.
Continuous record asymptotics for rolling sample variance estimators.  

Kristensen.
Nonparametric filtering of the realized spot volatility: A kernel-based approach.  

Zhang, Mykland, and Aït-Sahalia.
A tale of two time scales.  