Optimal Kernel Estimation of Spot Volatility

José E. Figueroa-López

Department of Mathematics
Washington University in St. Louis

figueroa@math.wustl.edu

(Joint work with Cheng Li from Purdue)

12th German Probability and Statistics Days
Bochum, Germany

March 1-4, 2016

Overview

1. Framework and Estimator
2. Optimal Bandwidth Selection
3. Optimal Kernel Selection
4. Implementation of the Bandwidth Selection Method
5. Simulation Study
6. Conclusions

1 Framework and Estimator

2 Optimal Bandwidth Selection

3 Optimal Kernel Selection

4 Implementation of the Bandwidth Selection Method

5 Simulation Study

6 Conclusions

The Model and the Statistical Problem I

- Consider the following Stochastic Differential Equation (SDE):

$$dX_t = \mu_t dt + \sigma_t dB_t \tag{1}$$

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

- In Financial settings, $X_t = \log(S_t)$, where S_t is the price of an asset, and the return of the asset is

$$\frac{S_{t+\Delta t} - S_t}{S_t} \approx \log(S_{t+\Delta t}) - \log(S_t) = \mu_t \Delta t + \sigma_t (B_{t+\Delta t} - B_t).$$

- **Data**: equally spaced observations:

$$\text{We observe } X_{t_i} \text{ at time } t_i := t_{i,n} = iT/n, 0 \leq i \leq n.$$

- **Notations**:

$$\text{Data increments: } \Delta_i X := \Delta_i^n X := X_{t_i} - X_{t_{i-1}},$$
$$\text{Time increments: } \Delta := \Delta_n = T/n.$$

The Model and the Statistical Problem II

- **The problem**:

  To estimate the volatility σ_τ for a fixed τ based on all observations X_{t_i}.

- **Why do spot volatility estimation matter**?
  1. Spot volatility σ_τ measures the (local) risk of an asset.
  2. Volatility is crucial in asset allocation, option pricing, risk management,...
  3. Spot volatility allows to detect patterns of volatility within a certain period of time.

The Idea of Kernel Estimator I

- Start from the Integrated Variance or Quadratic Variation:

$$IV_t = [X]_t = \int_0^t \sigma_s^2 ds.$$

- The standard estimator of IV is the Realized Quadratic Variation:

$$RV_{t,n} := \widehat{[X]}_{t,n} := \sum_{i:t_i \leq t} (\Delta_i^n X)^2 \to \int_0^t \sigma_s^2 ds, \quad n \to \infty.$$

- From IV to Spot Vol at τ: consider the Kernel Weighted Volatility:

$$KV_h(\tau) = \int_0^T K_h(s-\tau)\sigma_s^2 ds = \int_0^T K_h(s-\tau) d[X]_s \to \sigma_\tau^2, \quad h \to 0.$$

Here, weights are given by the scaled kernel function $K_h(x) := K(x/h)/h$ such that $\int K(x)dx = 1$.
K is the kernel function and h is the bandwidth.

The Idea of Kernel Estimator II

- It is then natural to estimate the Kernel Weighted Volatility by the following Kernel Weighted Realized Variation, for given h:

$$\hat{\sigma}^2_{\tau,n,h} := \int_0^T K_h(s-\tau)d[\widehat{X}]_{s,n} = \sum_{i=1}^{n} K_h(t_{i-1}-\tau)(\Delta_i^n X)^2$$
$$\to KV_h(\tau), \quad n \to \infty$$

- For a fixed n, we cannot set h arbitrarily small.
- Instead, we should carefully select $h = h_n$.
- Define the MSE of the kernel estimator as

$$MSE_n(h) = \mathbb{E}[(\hat{\sigma}^2_\tau - \sigma^2_\tau)^2] = \mathbb{E}\left[\left(\sum_{i=1}^{n} K_h(t_{i-1}-\tau)(\Delta_i X)^2 - \sigma^2_\tau\right)^2\right].$$

The key problem is:

Choose $h = h_n$ to minimize MSE_n.

Comparison to Key Related Works I

- Fan and Wang (2008) gives consistency and CLT results but did not consider the problem of bandwidth and kernel selection
- Kristensen (2010) considers the problem bandwidth selection but under the following relatively strong path-wise assumption:

$$\mathbb{P}-a.e.\ \omega : |\sigma^2_{t+\delta}(\omega) - \sigma^2_t(\omega)|^2 = L_t(\delta;\omega)\delta^\gamma + o_P(\delta^\upsilon),\ \forall t,\ (\delta \to 0), \quad (\blacksquare)$$

where $\gamma \in (0,2]$ and $\delta \to L_t(\delta)$ is a slowly varying (random) function at 0 and $t \to L_t(0) := \lim_{\delta \to 0^+} L_t(\delta)$ is continuous.

- Then, based on a non-leverage assumption, argues that σ can be assumed deterministic
- Under the latter assumption, proposes

$$h^{opt}_{n,\tau} = n^{-\frac{1}{\gamma+1}} \left(\frac{2T\sigma_t^4 \|K^2\|_1}{\gamma L_\tau(0)\kappa^2(K)} \right)^{\frac{1}{\gamma+1}}, \quad \kappa(K) := \int K(x)x dx \neq 0;$$

Assumption ($\blacksquare$) is hard to be verified with explicit $L_t(0) \in (0,\infty)$.

Assumptions on the Volatility Process I

- Our first assumption is a simplifying non-leverage assumption (also used in Kristensen, 2010):

Assumption 1

(μ, σ) is independent of B.

- Another assumption is the boundedness of the moments of μ and σ 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 $0 \leq t \leq T$.

Assumptions on the Volatility Process II

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

Assumption 3 ($\star$)

Suppose that the variance process $V := \{V_t = \sigma_t^2 : t \geq 0\}$ *satisfies*

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

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_γ *is not identically zero and has the scaling property:*

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

Hereafter, we denote $C_\gamma(r,s;t) = L(t)C_\gamma(r,s)$*, which can be interpreted as the "local covariance function" of the variance process.* V

Assumptions on the Volatility Process III

- Two useful propositions directly follow:

Proposition 1

Under the Key Assumption ($\star$), the following two assertions hold:

- *The γ and $C_\gamma(r,s;t) := L(t)C_\gamma(r,s)$ are unique.*
- *$C_\gamma(\cdot,\cdot)$ is integrally non-negative definite. That is, for any integrable $K:\mathbb{R}\to\mathbb{R}$, we have*

$$\iint K(x)K(y)C_\gamma(x,y)dxdy \geq 0.$$

Remark: In that case that

$$\iint K(x)K(y)C_\gamma(x,y)dxdy > 0,$$

for any integrable nonzero $K:\mathbb{R}\to\mathbb{R}$, we say that $C_\gamma(x,y)$ is integrally positive definite.

Deterministic Volatility Processes

The first class of volatility processes that satisfy the Key Assumption ($\star$),

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

is a deterministic and smooth volatility:

Lemma 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 ($\star$) with

$$\gamma = 2, \quad L(t) = (f'(t))^2, \quad C_\gamma(r,s) = rs.$$

BM Type Volatility Processes

- We consider a standard SDE driven by BM.
- In Financial Mathematics, this is one of the most important way to generalize Black-Scholes model to non-constant volatility.

Lemma 2

Consider a stochastic process $V_t = \sigma^2(t)$ *that satisfies*

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

where W *is a standard Brownian Motion. Assume that* $\mathbb{E}\left[f^2(t)\right] < M$ *for* $t \in [0, T]$ *and* $\mathbb{E}\left[g^2(t)\right]$ *is continuous for* $t \in [0, T]$*. Then, the Key Assumption* ($\star$) *is satisfied with* $\gamma = 1$, $L(t) = \mathbb{E}[g^2(t)]$, *and the function*

$$C_\gamma(r, s) := \min\{|r|, |s|\}1_{\{rs \geq 0\}} = \mathbb{E}[(W_{t+r} - W_t)(W_{t+s} - W_t)],$$

which is furthermore integrally positive definite.

Heston Model

- One of the most widely used stochastic volatility models in Finance is

$$\begin{aligned} dX_t &= \mu_t dt + \sqrt{V_t} dB_t, \\ dV_t &= \kappa(\theta - V_t)dt + \xi\sqrt{V_t} dW_t, \end{aligned} \tag{2}$$

  where parameters are restricted to $2\kappa\theta > \xi^2$, s.t. $V_t = \sigma_t^2$ is always positive.

- The volatility process has the so-called mean reverting property;
- The parameters:
  - θ is the long-run variance level;
  - κ is the is the speed at which V_t reverts to θ;
  - ξ is the volatility of volatility.
- By the previous lemma, Assumption 3 is satisfied with

$$\gamma = 1, \quad L(t) = \xi^2 \mathbb{E}(V_t), \quad C_\gamma(r,s) = \min\{|r|,|s|\} 1_{\{rs \geq 0\}}.$$

fBM Type Volatility

Lemma 3

Consider a process $\{Y_t^H\}_{t\geq 0}$ *that satisfies*

$$Y_t^H = \int_{-\infty}^{t} f(u) dB_u^H,$$

where $\{B_u^H\}_{u\in\mathbb{R}}$ *is a (two-sided) fractional Brownian motion with Hurst parameter* $H \in (\frac{1}{2}, 1)$ *and* $f(u)$ *is a continuous deterministic function s.t.*

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(u)f(v)|u-v|^{2H-2} du dv < \infty.$$

Then, the process Y_t^H *and* $\exp(Y_t^H)$ *satisfies the key Assumption* ($\star$) *with* $\gamma = 2H \in (1,2)$ *and*

$$C_\gamma(r,s) := \mathbb{E}[B_r^H B_s^H] = \frac{1}{2}(|r|^{2H} + |s|^{2H} - |r-s|^{2H}), \quad r,s \in \mathbb{R}.$$

Conditions on Kernel Function

Assumption 4

Given γ and C_γ 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) $\iint K(x)K(y)C_\gamma(x,y)dxdy > 0$.

- Notice that the condition (4) above does not put substantial restriction on K since C_γ is already integrally non-negative definite.

1 Framework and Estimator

2 Optimal Bandwidth Selection

3 Optimal Kernel Selection

4 Implementation of the Bandwidth Selection Method

5 Simulation Study

6 Conclusions

Approximation of MSE

The idea is to approximate MSE by leading order infinitesimals:

$$\begin{aligned}
\text{MSE} &= \mathbb{E}\left[\left(\sum_{i=1}^{n} K_h(t_{i-1}-\tau)(\Delta_i X)^2 - \sigma_\tau^2\right)^2\right] \\
&= 2\sum_{i=1}^{n} K_h^2(t_{i-1}-\tau)\mathbb{E}\left[\left(\int_{t_{i-1}}^{t_i} \sigma_t^2 dt\right)^2\right] \\
&\quad + \sum_{i=1}^{n}\sum_{j=1}^{n} K_h(t_{i-1}-\tau)K_h(t_{j-1}-\tau) \\
&\qquad \times \int_{t_{i-1}}^{t_i}\int_{t_{j-1}}^{t_j} \mathbb{E}[(\sigma_t^2-\sigma_\tau^2)(\sigma_s^2-\sigma_\tau^2)]dtds \\
&\quad + o(\frac{\Delta}{h}) + o(h^\gamma)
\end{aligned}$$

Approximation of MSE

We have the following first order approximation of the MSE:

Theorem 1 (F-L & Li 2016)

For μ and σ satisfying Assumptions 1, 2 and 3 and a kernel function K satisfying Assumption 4, we have, for any $\tau \in (0, T)$,

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

where $C(x, y; t) = L(t)C_\gamma(x, y)$ is the "local covariance function" of $V_t = \sigma^2_t$ as defined by the Key Assumption ($\star$).

Approximation of the Optimal Bandwidth I

- Consider the approximated MSE of the kernel estimator:

$$\mathrm{MSE}^{a}_{\tau,n}(h) := 2\frac{\Delta}{h}\mathbb{E}[\sigma^4_\tau]\|K^2\|_1 + h^\gamma L(\tau)\iint K(x)K(y)C_\gamma(x,y)dxdy$$

- Note that $2\mathbb{E}[\sigma^4_\tau]\|K^2\|_1 > 0$.
- Furthermore, by the previous assumptions, $L(\tau)\iint K(x)K(y)C_\gamma(x,y)dxdy > 0$.
- Taking the derivative of $\mathrm{MSE}^{a}_{\tau,n}(h)$ w.r.t. h, we then have the following:

Approximation of the Optimal Bandwidth II

Proposition 2

The approximated optimal bandwidth, denoted by $h_n^{a,opt}$*, which, by definition, minimizes the approximated* $MSE_n^a(h)$*, is given by*

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

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

$$MSE_n^{a,opt} = n^{-\gamma/(1+\gamma)} \left(1 + \frac{1}{\gamma} \right) \left(2T\mathbb{E}[\sigma_\tau^4] \| K^2 \|_1 \right)^{\gamma/(1+\gamma)} \times \left(\gamma L(\tau) \iint K(x)K(y)C_\gamma(x,y)dxdy \right)^{1/(1+\gamma)}$$

Examples

- When $\sigma_t^2 = f(t)$ is deterministic and smooth ($\gamma = 2$, $L(t) = (f'(t))^2$, and $C_\gamma(r,s) = rs$), we recover Kristensen (2010):

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

  If $\int K(x)xdx = 0$, one can improve the rate of convergence of the MSE;

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

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

$$\kappa_{BM}(K) = \int_0^\infty \int_0^\infty \left[K(x)K(y) + K(-x)K(-y)\right] \min(x,y) dxdy,$$

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

- The rate of convergence of the MSE, $n^{-1/2}$, cannot be improved;

Integrated MSE

- In the case that we consider the Integrated MSE (IMSE)

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

the optimal (uniform) bandwidth takes the form:

$$h_n^{a,opt} = n^{-1/(\gamma+1)} \left[\frac{2T \int_0^T \mathbb{E}[\sigma_\tau^4]d\tau \|K^2\|_1}{\gamma \int_0^T L(\tau)d\tau \iint K(x)K(y)C_\gamma(x,y)dxdy} \right]^{1/(\gamma+1)}.$$

Equivalence of the Approximated Optimal Bandwidth I

- The questions now are:
  - whether our approximated optimal bandwidth yields good approximation for the true optimal bandwidth.
  - whether the kernel estimator with such approximated optimal bandwidth performs good.
- Formally, consider
  - the true MSE: $MSE_n(h) = \mathbb{E}[(\hat{\sigma}^2_{\tau,n,h} - \sigma^2_\tau)^2]$.
  - A true optimal bandwidth $h_n^{opt} = \arg\min_{h \in \mathbb{R}_+} MSE_n(h)$
  - The approximated optimal bandwidth: $h_n^{a,opt}$.

Equivalence of the Approximated Optimal Bandwidth II

Theorem 2 (F-L & Li 2016)

For μ and σ satisfying Assumptions 1, 2, 3 and a kernel function K satisfying Assumption 4, we have

$$h_n^{a,opt} = h_n^{opt} + o(h_n^{opt}),$$
$$MSE_n(h_n^{a,opt}) = MSE_n(h_n^{opt}) + o(MSE_n(h_n^{opt})).$$

1 Framework and Estimator

2 Optimal Bandwidth Selection

3 Optimal Kernel Selection

4 Implementation of the Bandwidth Selection Method

5 Simulation Study

6 Conclusions

BM Volatility Case I

- Recall the approximated optimal MSE:

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

- In the BM case, if we try to minimize this approximated optimal $\mathrm{MSE}^{a,opt}(K)$ over K, we need to minimize the following objective function $I(K)$:

$$I(K) = \int K^2(x)dx \int_0^\infty \int_0^\infty [K(x)K(y)+K(-x)K(-y)] \min(x,y)dxdy,$$

with the restriction $\int K(x)dx = 1$.

BM Volatility Case II

- By solving the optimization problem, we obtain the following theorem:

Theorem 3 (F-L & Li 2016)

With all previous assumptions, in the case of BM type volatility, the optimal kernel function is given by $K^{opt}(x) = \frac{1}{2}e^{-|x|}, x \in \mathbb{R}$.

- We now do some calculations and demonstrate to what extent the exponential kernel decrease the MSE.

Example 3.1

As seen before, $MSE_n^{a,opt}(K) = C\sqrt{I(K)}$ *width a constant* C *that does not depend on* K. *Then, we can see that* $\sqrt{I(\frac{1}{2}e^{-|x|})} = \frac{1}{4} = 0.25$. *Some simple calculations show that* $\sqrt{I(\frac{1}{2}1_{\{|x|<1\}})} = \frac{1}{2\sqrt{3}} \approx 0.2887$, $\sqrt{I(|1-x|1_{\{|x|<1\}})} = \frac{1}{\sqrt{15}} \approx 0.2582$.

Complexity and Fast Estimation I

- Recall that

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

- In general, to find $\hat{\sigma}_{t_i}$ for one fixed i, we need a time complexity of $O(n)$ or $O(nh)$ for a kernel function with unbounded or bounded support, respectively.
- So, to find $\hat{\sigma}_{t_i}$ for all $0 \leq i \leq n$, we need a time complexity of $O(n^2)$ or $O(n^2 h)$, respectively.
- The best time complexity of $O(n^{2-1/(\gamma+1)})$, with optimal bandwidth, is not be efficient for high frequency data
- But, the rate can be improved for uniform and exponential kernels using a type of fast implementation algorithm.

Complexity and Fast Estimation II

- For the uniform kernel, we have

$$\hat{\sigma}^2_{t_{i+1},unif} = \hat{\sigma}^2_{t_i,unif} + \frac{1}{2h}(\Delta_k X)^2 - \frac{1}{2h}(\Delta_j X)^2$$

  where k and j satisfies

$$k = \min\{l : t_{l-1} > t_i + h\}, \quad j = \max\{l : t_{l-1} \geq t_i - h\}.$$

- The time complexity to compute $\hat{\sigma}_{t_i}$ for all $0 \leq i \leq n$ is $O(n)$ vs. $O(n^{2-1/(\gamma+1)})$

Complexity and Fast Estimation III

- For exponential kernel, we use the following notation:

$$\hat{\sigma}^2_{t_i,-} = \sum_{j<i} K_h(t_{j-1} - t_i)(\Delta_j X)^2,$$
$$\hat{\sigma}^2_{t_i,+} = \sum_{j\geq i+1} K_h(t_{j-1} - t_i)(\Delta_j X)^2$$

- Note that $\hat{\sigma}^2_{t_i,exp} = \sum_j K_h(t_{j-1} - t_i)(\Delta_j X)^2$ can be written as

$$\hat{\sigma}^2_{t_i,exp} = \hat{\sigma}^2_{t_i,+} + \hat{\sigma}^2_{t_i,-} + K_h(h)(\Delta_i X)^2$$

- We can then use the following recurrent algorithm:

$$\begin{aligned}\hat{\sigma}^2_{t_{i+1},-} &= e^{-\Delta/h}\left[\hat{\sigma}^2_{t_i,-} + K_h(h)(\Delta_i X)^2\right] \\ \hat{\sigma}^2_{t_{i+1},+} &= e^{\Delta/h}\left[\hat{\sigma}^2_{t_i,+} - K_h(0)(\Delta_{i+1} X)^2\right]\end{aligned} \tag{3}$$

- The time complexity is again $O(n)$.

1 Framework and Estimator

2 Optimal Bandwidth Selection

3 Optimal Kernel Selection

4 Implementation of the Bandwidth Selection Method

5 Simulation Study

6 Conclusions

Practical Bandwidth Selection Methods

- Generally, two types of bandwidth selection methods: plug-in type and cross-validation methods.
- Cross-validation methods:
  1. Advantages: a general method, yields satisfactory accuracy.
  2. Disadvantages: time consuming, hard to find initial value and implement the involved optimization algorithm.
- Plug-in type methods:
  1. Advantages: always faster and usually better accuracy.
  2. Disadvantages: less general.

Cross-Validation Bandwidth Selection Methods

- In Kristensen (2010), a leave-one-out cross validation method is proposed.
- Recall that, therein, σ is assumed to be deterministic and satisfy a Hölder continuity condition.
- The idea is to consider the Integrated Squared Error: $ISE(h) = \int_{T_l}^{T_r} (\sigma_t^2 - \hat{\sigma}_{t,h}^2)^2 dt$, for $0 < T_l < T_r < T$.
- An estimator of ISE is $\widehat{ISE}(h) = \sum_{i=1}^{n} 1_{\{T_l < t_{i-1} < T_r\}} \left[\frac{(\Delta_i X)^2}{\Delta} - \hat{\sigma}_{-i,t_i}^2 \right]^2$, where $\hat{\sigma}_{-i,t_i}^2 = \sum_{j \neq i} K_h(t_{i-1} - t_i)(\Delta_j X)^2$ is the leave-one-out estimator.
- Now find $h^{opt} = \text{argmin}_{h \in \mathbb{R}_+} \widehat{ISE}(h)$.
- This result in a numerical optimization problem.

Plug-In Type Bandwidth Selection Methods I

- The idea of plug-in bandwidth selection is to use the explicit approximated optimal bandwidth we obtained and estimate all the parameters encountered.
- We consider the BM type volatility processes:
  $dV_t = f(t)dt + g(t)dW_t$,
  in which case $\gamma = 1$ and $C_1(r,s) = \min\{|r|,|s|\}\mathbf{1}_{\{rs>0\}}$.
- The approximated optimal bandwidth is then given by

$$h_n^{a,opt} = \left[\frac{2T\mathbb{E}[\sigma_\tau^4]\int K^2(x)dx}{nL(\tau)\iint K(x)K(y)C_1(x,y)dxdy}\right]^{1/2}.$$

- **We need to estimate $\mathbb{E}[\sigma_\tau^4]$ and $L(\tau) = \mathbb{E}[g^2(\tau)]$.**
- **Or we can estimate $\int_0^T \mathbb{E}[\sigma_\tau^4]d\tau$ and $\int_0^T L(\tau)d\tau = \int_0^T \mathbb{E}[g^2(\tau)]d\tau$ in the IMSE or stationary cases.**

Plug-In Type Bandwidth Selection Methods II

- Recall the dynamic of Heston Model:

$$\begin{aligned} dX_t &= \mu_t dt + \sqrt{V_t} dB_t, \\ dV_t &= \kappa(\theta - V_t)dt + \xi\sqrt{V_t} dW_t. \end{aligned}$$

- If we assume a Heston Model, the optimal bandwidth becomes

$$h_n^{a,opt} = \left[\frac{2T\mathbb{E}[\sigma_\tau^4] \int K^2(x)dx}{n\xi^2\mathbb{E}[\sigma_\tau^2] \iint K(x)K(y)C_1(x,y)dxdy} \right]^{1/2}.$$

- Although $\mathbb{E}[\sigma_\tau^4]$ and $\mathbb{E}[\sigma_\tau^2]$ are not available, it is natural to use $\hat{\sigma}_\tau^4$ and $\hat{\sigma}_\tau^2$ as their estimators.
- Since $[V]_T = \xi^2[X]_T$, an estimator of ξ is given by

$$\hat{\xi}^2 = \frac{\sum_{i=1}^{[n/K]} \left(\hat{\sigma}^2_{t_{Ki}} - \hat{\sigma}^2_{t_{K(i-1)}} \right)^2}{\sum_{i=1}^{n} (\Delta_i^n X)^2},$$

where K can be chosen to be $K \approx n^{2/3}$, for example.

Plug-In Type Bandwidth Selection Methods III

- If we assume a general Itô processes (say, $dV_t = f(t)dt + g(t)dW_t$), the Integrated Volatility of Volatility $IVV = \int_0^T L(t)dt = \int_0^T \mathbb{E}[g^2(t)]dt$ can be estimated by

$$\widehat{IVV} = \sum_{i=1}^{[n/K]} \left(\hat{\sigma}^2_{t_{Ki}} - \hat{\sigma}^2_{t_{K(i-1)}} \right)^2, \tag{4}$$

for some suitable K. Therefore, the final bandwidth can be given by:

$$h_n^{a,opt} = \left[\frac{2T\widehat{IV^2}_\tau \int K^2(x)dx}{n\widehat{IVV} \iint K(x)K(y)C_1(x,y)dxdy} \right]^{1/2},$$

where $\widehat{IV^2} = \frac{1}{3\Delta}\sum_{i=1}^n (\Delta_i X)^4$ is the realized quartic variation and $\widehat{IVV}$ is given by (4).

Plug-In Type Bandwidth Selection Methods IV

- The two estimators of the optimal bandwidth involves the estimation of spot volatility, which we do not know in advance, so we consider the following iterative algorithm:
- **The Iterative Plug-in Bandwidth Selection Algorithm:**
  Data: $\Delta_1^n X = X_\Delta - X_0, ..., \Delta_n^n X = X_{n\Delta} - X_{(n-1)\Delta}$
  Set initial values for h or h_{t_i} ;
  while *Stopping criteria not met* **do**
  Find a estimation of spot volatility on all the grid points:
  $\hat{\sigma}^2_{t_i}$ based on previous bandwidth h or h_{t_i} ;
  Update the bandwidth h or h_{t_i}
  based on the new estimation of spot volatility.
  end
- In reality, one or two iterations are enough for satisfactory result, even with bad initial guess.
- More iterations do not generally improve the result.

1 Framework and Estimator

2 Optimal Bandwidth Selection

3 Optimal Kernel Selection

4 Implementation of the Bandwidth Selection Method

5 Simulation Study

6 Conclusions

Simulation Study I

- We consider Heston model with the following parameter:
  1. 1 year time period, 252 trading days, 6.5 hours each data, minute data.
  2. $\mu_t \equiv 0.05, \sigma_0 = 0.2, \kappa = 5, \theta = 0.2, \xi = 0.5$.
  3. The leverage coefficient $dB_t \cdot dW_t = \rho dt$ is taken to be either $\rho = 0$ or $\rho = -0.5$.
- First, we compare different kernel functions. We consider five different kernels:

$$K_1(x) = \frac{1}{2}e^{-|x|}, \quad K_2(x) = \frac{|x|}{x^4 + 2x^2 + 1}, \quad K_3(x) = \frac{1}{2}1_{\{|x|<1\}}$$

$$K_4(x) = |1 - x|1_{\{|x|<1\}}, \quad K_5(x) = \frac{3}{4}(1 - x^2)1_{\{|x|<1\}}$$

- The first kernel is the optimal kernel we obtained previously.
- The second one is another kernel with infinite domain.

Simulation Study II

- The third, fourth and fifth kernel are finite domain kernel with different order of polynomial.
- The fifth kernel is the so called Epanechnikov kernel that Kristensen (2010) claims to be optimal in the deterministic case.
- In the following table, we only estimate the volatility at $t = \frac{3}{4}$. We see that exponential kernel outperforms other kernel functions.

Kernel	MSE($\rho = 0$)	MSE($\rho = -0.5$)
K_1	1.1000e-05	1.1569e-05
K_2	1.1976e-05	1.3064e-05
K_3	1.2242e-05	1.3501e-05
K_4	1.1154e-05	1.2037e-05
K_5	1.1420e-05	1.2427e-05

Simulation Study III

- Now we compare the following bandwidth (bw) selection methods for an exponential kernel function:
  1. Cross-validation method proposed by Kristensen.
  2. Optimal uniform bandwidth selection with 0, 1 and 2 iterations.
  3. Optimal spot bandwidth selection for 0, 1 and 2 iterations.
- In the following, we consider the approximated integrated MSE $\sum_i \left(\hat{\sigma}^2_{t_i} - \sigma^2_{t_i}\right)^2$ for each method:

Method	MSE($\rho = 0$)	MSE($\rho = -0.5$)
Cross-Validation	0.4465	0.4196
Uniform bw h, no iteration	0.4567	0.4342
Uniform bw h, 1 iteration	0.4571	0.4339
Uniform bw h, 2 iteration	0.4575	0.4342
Spot bw h, no iteration	0.4439	0.4230
Spot bw h, 1 iteration	0.4422	0.4211
Spot bw h, 2 iteration	0.4423	0.4211

1 Framework and Estimator

2 Optimal Bandwidth Selection

3 Optimal Kernel Selection

4 Implementation of the Bandwidth Selection Method

5 Simulation Study

6 Conclusions

Conclusions

1. A new theory for optimal bandwidth selection is put forward under a key assumption on the local covariance of the variance process
2. The considered framework covers several key models in the literature including volatility models driven by BM and fBM
3. Our results show that earlier optimal bandwidth selection methods, obtained under a deterministic volatility assumption, are not valid in a stochastic volatility setting
4. 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
5. Fast iterated plug-in type algorithms are also devised as a way to implement the proposed optimal selection methods

References I

Fan and Wang.
Spot volatility estimation for high-frequency data.
Statistics and its Interface, 1(2):279–288, 2008.

Figueroa-López and Nisen.
Optimally thresholded realized power variations for lévy jump diffusion models.
Stochastic Processes and their Applications, 123(7):2648–2677, 2013.

Kristensen.
Nonparametric filtering of the realized spot volatility: A kernel-based approach.
Econometric Theory, 26(01):60–93, 2010.

Rosenblatt et al.
Remarks on some nonparametric estimates of a density function.
The Annals of Mathematical Statistics, 27(3):832–837, 1956.

References II

Zhang, Mykland, and Aït-Sahalia.
A tale of two time scales.
Journal of the American Statistical Association, 100(472), 2005.