Supplement to "Optimal Kernel Estimation of Spot Volatility of Stochastic Differential Equations"

José E. Figueroa-López*
Department of Mathematics, Washington University in St. Louis
and
Cheng Li
Department of Statistics, Purdue University

February 16, 2018

Abstract

We provide the proofs of some key lemmas and propositions that were omitted in the article Figueroa-López & Li (2018).

Keywords: Spot volatility estimation; Kernel estimation; Bandwidth selection; Kernel function selection; vol vol estimation.

1 Notation and Assumptions

In this section, we recall the terminology in Figueroa-López & Li (2018)for future reference. We study the problem of estimating the spot volatility σ_τ of the stochastic differential

*The first author's research is partially supported by the NSF grants DMS-1149692 and DMS-1613016.

equation (SDE)

$$dX_t = \mu_t dt + \sigma_t dB_t. \tag{1.1}$$

We refer to Figueroa-López & Li (2018)for further assumptions on the coefficient processes of the previous SDE. We suppose that we observe the log price process X at the times $t_i := t_{i,n} := i\Delta_n$, $0 \le i \le n$, where $\Delta_n := T/n$. We will use $\Delta_i^n X := \Delta X_{t_{i-1}} := X_{t_i} - X_{t_{i-1}}$ to denote the increments of log prices and $\Delta_n = T/n$ to denote the time increments. The key estimator is

$$\hat{\sigma}^2_{\tau,n,h} := \sum_{i=1}^{n} K_h(t_{i-1} - \tau)(\Delta_i^n X)^2, \tag{1.2}$$

where $K_h(x) = K(x/h)/h$.

For future reference we recall the following assumption.

Assumption 1. (μ, σ) *is independent of* B *and there exists* $M_T > 1$ *such that* $\mathbb{E}[\mu_t^4 + \sigma_t^4] < M_T$*, for all* $0 \le t \le T$.

Note that this assumption implies $\mathbb{E}[|\mu_t|] < M_T$, $\mathbb{E}[\mu_t^2] < M_T$, and $\mathbb{E}[\sigma_t^2] < M_T$, for all $t \in [0, T]$. We will use the notation M_T later.

The following is is the most important assumption:

Assumption 2. *Suppose that for* $\gamma > 0$ *and certain functions* $L : \mathbb{R}_+ \to \mathbb{R}_+$, $C_\gamma : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$*, such that* C_γ *is not identically zero and*

$$C_\gamma(hr, hs) = h^\gamma C_\gamma(r, s), \quad \textit{for } r, s \in \mathbb{R}, h \in \mathbb{R}_+, \tag{1.3}$$

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

$$\mathbb{E}[(V_{t+r} - V_t)(V_{t+s} - V_t)] = L(t)C_\gamma(r, s) + o((r^2 + s^2)^{\gamma/2}), \quad r, s \to 0. \tag{1.4}$$

Hereafter, we will also denote $C(r, s; t) = L(t)C_\gamma(r, s)$.

The above condition is a weaker version of the following, which is satisfied by, BM type volatility processes:

$$\begin{aligned} \mathbb{E}[(V_{t+h}-V_t)^2] =& L(t)h + O(h^{\alpha+\beta}), \quad t>0, h\to 0, \\ |\mathbb{E}[(V_{t+h}-V_t)(V_t-V_{t-s})]| \leq& A(t)h^{\alpha}s^{\beta}, \quad h>0,\ t>s>0. \end{aligned} \tag{1.5}$$

where $A:\mathbb{R}_+\to\mathbb{R}_+$ and $L:\mathbb{R}_+\to\mathbb{R}_+$ are locally bounded functions and $\alpha,\beta\geq 0$ are reals such that $\alpha+\beta>1$.

Finally, we imposed the following conditions on the kernel functions.

Assumption 3. *Given $\gamma>0$ and C_γ as defined in Assumption 2, we assume that the kernel function $K:\mathbb{R}\to\mathbb{R}$ satisfies the following conditions:*

(1) $\int K(x)dx = 1$;

(2) K is Lipschitz and piecewise C^1 on its support (A,B), where $-\infty\leq A<0<B\leq\infty$;

(3) (i) $\int|K(x)||x|^{\gamma}dx<\infty$; (ii) $K(x)x^{\gamma+1}\to 0$, as $|x|\to\infty$; (iii) $\int|K'(x)|dx<\infty$, (iv) $V_{-\infty}^{\infty}(K')<\infty$, where $V_{-\infty}^{\infty}(\cdot)$ is the total variation;

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

2 Proofs of Section 2 in Figueroa-López & Li (2018)

Proposition 2.1. *Under Assumption 2, both $C_\gamma(\cdot,\cdot;t)$ and $C_\gamma(\cdot,\cdot)$ are integrally non-negative definite. That is,*

$$\iint K(x)K(y)C(x,y)dxdy\geq 0, \tag{2.1}$$

for all functions $K:\mathbb{R}\to\mathbb{R}$ for which the integral therein is well-defined. Furthermore, both γ and $C_\gamma(r,s;t)$ are unique.

Proof. To prove the first part of the result, we write (1.4) as $\mathbb{E}[(V_{t+r} - V_t)(V_{t+s} - V_t)] = C_\gamma(r, s; t) + D(r, s; t)$, where $D(r, s; t) = o((r^2 + s^2)^{\gamma/2})$, as $r, s \to 0$. We first show that C_γ is non-negative definite. Indeed, for $n \in \mathbb{N}$, $(x_1, ..., x_n) \in \mathbb{R}^n$, $(c_1, ..., c_n) \in \mathbb{R}^n - \{0\}$ and $h \in \mathbb{R}_+$, we have

$$\sum_{i=1}^n \sum_{j=1}^n c_i c_j C_\gamma(x_i, x_j; t) = h^{-\gamma} \sum_{i=1}^n \sum_{j=1}^n c_i c_j C_\gamma(hx_i, hx_j; t)$$

$$= h^{-\gamma} \sum_{i=1}^n \sum_{j=1}^n c_i c_j \mathbb{E}[(V_{t+hx_i} - V_t)(V_{t+hx_j} - V_t)] - h^{-\gamma} \sum_{i=1}^n \sum_{j=1}^n c_i c_j D(hx_i, hx_j; t)$$

$$= h^{-\gamma} \mathbb{E}\left[\left(\sum_{i=1}^n c_i (V_{t+hx_i} - V_t)\right)^2\right] - h^{-\gamma} \sum_{i=1}^n \sum_{j=1}^n c_i c_j D(hx_i, hx_j; t).$$

On the right-hand side of the previous equation, we let $h \to 0_+$ and we have that the first term is always non-negative, while the second term converges to zero. This shows the non-negative definiteness of C_γ. The integral non-negative definiteness follows then, since the Riemann integration is defined to be the limit of finite sum, which is always non-negative.

We now prove the uniqueness of γ. Suppose there are γ, γ' such that $\gamma' > \gamma > 0$, and correspondingly, C_γ and $C'_{\gamma'}$, that satisfies (1.4). Since C_γ is non-zero, there exists $r, s \in \mathbb{R}$, such that $C_\gamma(r, s; t) \neq 0$. Then,

$$\begin{aligned} \mathbb{E}[(V_{t+rh} - V_t)(V_{t+sh} - V_t)] &= h^\gamma C_\gamma(r, s; t) + o(h^\gamma (r^2 + s^2)^{\gamma/2}) \\ &= h^{\gamma'} C'_{\gamma'}(r, s; t) + o(h^{\gamma'} (r^2 + s^2)^{\gamma/2}), \quad h \to 0. \end{aligned}$$

Note now that in the right two parts, all the terms are $o(h^\gamma)$ except $h^\gamma C_\gamma(r, s; t)$. Since we have assumed that $C_\gamma(r, s; t) \neq 0$, this is impossible. Therefore, $\gamma = \gamma'$ and, thus, γ must be unique. Now with the same γ, suppose at some r, s, we have $C_\gamma(r, s) \neq C'_\gamma(r, s)$. Then, a similar argument shows that this leads to a contradiction. This proves the uniqueness of γ and C_γ. □

Proposition 2.2. *Suppose the squared volatility process is given by a deterministic function* $f(t) = \sigma_t^2$, $0 \le t \le T$, *such that, for some* $m \ge 1$, f *is* m^{th}*-times differentiable at* $\tau \in (0, T)$, $f^{(i)}(\tau) = 0$, *for* $1 \le i < m$, *and* $f^{(m)}(\tau) \neq 0$. *Then,* f *satisfies Assumption 2 with* $\gamma = 2m$ *and* $C_{2m}(r, s) := r^m s^m$.

Proof. Using Taylor expansion, we have

$$f(\tau + s) - f(\tau) = \sum_{j=1}^{m} \frac{f^{(j)}(\tau)}{j!} s^j + O(s^{m+1}) = \frac{f^{(m)}(\tau)}{m!} s^m + O(s^{m+1}).$$

Therefore, since $f(\cdot)$ is non-random, we have

$$\mathbb{E}[f(\tau + r) - f(\tau)][f(\tau + s) - f(\tau)] = \left(\frac{f^{(m)}(\tau)}{m!}\right)^2 r^m s^m + o((r^2 + s^2)^m),$$

and, thus, the Assumption 2 is satisfied with $C_{2m}(r, s) = r^m s^m$. □

Proposition 2.3. *Consider a complete filtered probability space* $(\Omega, \mathscr{F}, \mathbb{F} = \{\mathscr{F}_t\}_{t \ge 0}, \mathbb{P})$ *and an Itô process* $V_t = \sigma^2(t, \omega)$ *that satisfies the SDE*

$$dV_t = f(t, \omega)dt + g(t, \omega)dW_t, \quad t \in [0, T], \tag{2.2}$$

where $\{W_t\}_{t \ge 0}$ *is a standard Wiener process adapted to* $\mathbb{F}$. *Assume that* $f(t, \omega)$ *and* $g(t, \omega)$ *are adapted and progressively measurable with respect to* $\mathbb{F}$, $\mathbb{E}\left[f^2(t, \omega)\right] < M$, *for* $t \in [0, T]$, *and* $\mathbb{E}\left[g^2(t, \omega)\right]$ *is continuous for* $t \in [0, T]$. *Then, Assumption 2 is satisfied with* $\gamma = 1$, $C_1(r, s) = \min\{|r|, |s|\} 1_{\{rs \ge 0\}}$, *and* $L(t) = \mathbb{E}[g^2(t, \omega)]$. *Furthermore,* $C_1(r, s)$ *is an integrable positive definite function; i.e., we have strict inequality in (2.1) for all* $K : \mathbb{R} \to \mathbb{R}$ *such that* $\int |K(x)| dx > 0$.

Proof. We consider $h > 0$ in what follows, while the case of $h < 0$ is similar. Using the boundedness of $\mathbb{E}[f^2(t,\omega)]$ and continuity of $\mathbb{E}[g^2(t,\omega)]$, we have

$$\begin{aligned}\mathbb{E}[(V_{t+h}-V_t)^2] &= \mathbb{E}\left[\int_t^{t+h} f(s,\omega)ds\right]^2 + \mathbb{E}\left[\int_t^{t+h} g^2(s,\omega)ds\right] \\ &\quad + 2\mathbb{E}\left[\int_t^{t+h} f(s,\omega)ds \int_t^{t+h} g(s,\omega)dB_s\right] \\ &= h\mathbb{E}[g^2(t,\omega)] + o(h), \quad h \to 0,\end{aligned}$$

where in the last equality we used that

$$\begin{aligned}&\mathbb{E}\left[\int_t^{t+h} f(s,\omega)ds\right]^2 \leq h\mathbb{E}\left[\int_t^{t+h} f^2(s,\omega)ds\right] \\ &= h\int_t^{t+h} \mathbb{E}[f^2(s,\omega)]ds = O(h^2), \\ &\mathbb{E}\left[\int_t^{t+h} f(s,\omega)ds \int_t^{t+h} g(s,\omega)dB_s\right] \\ &\leq \sqrt{\mathbb{E}\left[\int_t^{t+h} f(s,\omega)ds\right]^2 \mathbb{E}\left[\int_t^{t+h} g(s,\omega)dB_s\right]^2} = O(h^{3/2}), \\ &h^{-1}\mathbb{E}\left[\int_t^{t+h} g^2(s,\omega)ds\right] - \mathbb{E}\left[g^2(t,\omega)\right] = o(1),\end{aligned}$$

for $h \to 0_+$. Now, for $r > 0$ and $t > s > 0$, we have

$$
\begin{aligned}
&|\mathbb{E}[(V_{t+r} - V_t)(V_t - V_{t-s})]| \\
&= \left|\mathbb{E}\left[\int_t^{t+r} f(s,\omega)ds + \int_t^{t+r} g(s,\omega)dB_s\right]\left[\int_{t-s}^t f(s,\omega)ds + \int_{t-s}^t g(s,\omega)dB_s\right]\right| \\
&\leq \left|\mathbb{E}\int_t^{t+r} f(s,\omega)ds \int_{t-s}^t f(s,\omega)ds\right| + \left|\mathbb{E}\int_t^{t+r} g(s,\omega)dB_s \int_{t-s}^t f(s,\omega)ds\right| \\
&\quad + \left|\mathbb{E}\int_t^{t+r} f(s,\omega)ds \int_{t-s}^t g(s,\omega)dB_s\right| + \left|\mathbb{E}\int_t^{t+r} g(s,\omega)dB_s \int_{t-s}^t g(s,\omega)dB_s\right| \\
&\leq \sqrt{\mathbb{E}\left[\int_t^{t+r} f(s,\omega)ds\right]^2 \mathbb{E}\left[\int_{t-s}^t f(s,\omega)ds\right]^2} \\
&\quad + \sqrt{\mathbb{E}\left[\int_t^{t+r} f(s,\omega)ds\right]^2 \mathbb{E}\left[\int_{t-s}^t g(s,\omega)dB_s\right]^2} \\
&\leq A_1 rs + A_2 r\sqrt{s} \leq Ar\sqrt{s},
\end{aligned}
$$

for some constant A_1, A_2 and A. Note that A can be made uniform over $t \in (0,T)$ due to boundedness of $\mathbb{E}[f^2(t,\omega)]$ and continuity of $\mathbb{E}[g^2(t,\omega)]$. Finally, for $r > s > 0$, we have

$$
\begin{aligned}
&\mathbb{E}[(V_{t+r} - V_t)(V_{t+s} - V_t)] \\
&= \mathbb{E}[(V_{t+s} - V_t)^2] + \mathbb{E}[(V_{t+r} - V_{t+s})(V_{t+s} - V_t)] \\
&= s\mathbb{E}[g^2(t,\omega)] + o(s) + O((r-s)\sqrt{s}) = s\mathbb{E}[g^2(t,\omega)] + o(s) + O(r\sqrt{s}) \\
&= s\mathbb{E}[g^2(t,\omega)] + o((r^2+s^2)^{1/2}).
\end{aligned}
$$

Similar arguments can be applied for $r < s < 0$, while the case of $r < 0 < s$ can be proved by noticing that $r\sqrt{s} = o((r^2+s^2)^{1/2})$. Therefore, in summary, we have proved that Assumption 2 hold true with $\gamma = 1$ and $C_1(r,s) = \min\{|r|,|s|\}1_{\{rs\geq 0\}}$ and $L(t) = \mathbb{E}[g^2(t,\omega)]$.

It remains to prove that C_1 is positive definite. To that end, note that

$$
\begin{aligned}
&\iint K(r)K(s)\min\{|r|,|s|\}1_{\{rs\geq 0\}}drds \\
&= \int_0^\infty\int_0^\infty K(r)K(s)\min\{r,s\}drds + \int_0^\infty\int_0^\infty K(-r)K(-s)\min\{r,s\}drds \\
&= \int_0^\infty\int_0^\infty [K(r)K(s)+K(-r)K(-s)]\int_0^\infty 1_{\{t\leq r\}}1_{\{t\leq s\}}dtdrds \\
&= \int_0^\infty\left[\int_0^\infty K(r)1_{\{t\leq r\}}dr\right]^2 dt + \int_0^\infty\left[\int_0^\infty K(-r)1_{\{t\leq r\}}dr\right]^2 dt,
\end{aligned}
$$

which is positive as long as $\int |K(x)|dx > 0$. □

Proposition 2.4. *Consider a filtered probability space $(\Omega,\mathscr{F},\mathbb{F}=\{\mathscr{F}_t\}_{t\geq 0},P)$ and a process $Y^{(H)}=\{Y_t^{(H)}\}_{t\geq 0}$ that satisfies*

$$Y_t^{(H)} = \int_{-\infty}^t f(u)dB_u^{(H)}, \quad t\geq 0,$$

where $\{B_t^{(H)}\}_{t\in\mathbb{R}}$ is a (two-sided) fBM with Hurst parameter $H\in(\frac{1}{2},1)$ and $f(\cdot)$ is a continuous function that satisfies

$$\iint |f(u)f(v)||u-v|^{2H-2}dudv < \infty. \tag{2.3}$$

Then, the processes $Y^{(H)}$ and $\{\exp(Y_t^{(H)})\}_{t\geq 0}$ satisfy Assumption 2 with $\gamma = 2H\in(1,2)$ and C_γ given by

$$C_\gamma(r,s) = \frac{1}{2}(|r|^\gamma + |s|^\gamma - |r-s|^\gamma). \tag{2.4}$$

Proof. For easiness of notation we write B^H and Y^H instead of $B^{(H)}$ and $Y^{(H)}$. Pipiras and Taqqu (2000) gave the following result:

$$\mathbb{E}\left[\int g_1(u)dB_u^H\int g_2(u)dB_u^H\right] = H(2H-1)\iint g_1(u)g_2(v)|u-v|^{2H-2}dudv, \tag{2.5}$$

where g_1 and g_2 are assumed to be real valued function satisfying the integrability condition (2.3). We first use this results with $g_1(u) = 1_{[t,t+r]}(u), g_2(u) = 1_{[t,t+s]}(u)$, where $r, s \in \mathbb{R}$. We consider $r = s > 0$ first and we have

$$\mathbb{E}\left[\left(\int_t^{t+r} dB_u^H\right)^2\right] = H(2H-1)\int_t^{t+r}\int_t^{t+r}|u-v|^{2H-2}dudv$$
$$= 2H(2H-1)\int_t^{t+r}\left(\int_t^u (u-v)^{2H-2}dv\right)du = r^{2H}.$$

For the case of $r > 0 > s$, we have

$$\mathbb{E}\left[\left(\int_t^{t+s} dB_u^H\right)\left(\int_t^{t+r} dB_u^H\right)\right] = -H(2H-1)\int_{t+s}^{t}\int_t^{t+r}|u-v|^{2H-2}dudv$$
$$= -H\int_{t+s}^{t}|t+r-v|^{2H-1} - |t-v|^{2H-1}dv = \frac{1}{2}(|r|^{2H} + |s|^{2H} - |r-s|^{2H}).$$

These two results can be combined to be

$$\mathbb{E}\left[\left(\int_t^{t+s} dB_u^H\right)\left(\int_t^{t+r} dB_u^H\right)\right] = \frac{1}{2}(|r|^{2H} + |s|^{2H} - |r-s|^{2H}) =: C_{2H}(r,s;t),$$

for all $r, s \in \mathbb{R}$. Next, we first assume that $f(t) > 0$ and prove the case of $r, s > 0$. Other cases can be proved similarly. Since f is assumed to be continues, for any $\epsilon \in (0, f(t))$, let $\delta = \delta_{\varepsilon,t} > 0$, such that $\forall h \in (0,\delta)$, $|f(t+h) - f(t)| < \epsilon$. Then, we have the following upper bound, for any $0 < r, s < \delta$:

$$\mathbb{E}\left[\left(\int_t^{t+r} f(u)dB_u^H\right)\left(\int_t^{t+s} f(u)dB_u^H\right)\right]$$
$$= \mathbb{E}\left[\left(\int f(u)1_{[t,t+r]}dB_u^H\right)\left(\int f(u)1_{[t,t+s]}dB_u^H\right)\right]$$
$$= H(2H-1)\int_t^{t+r}\int_t^{t+s} f(u)f(v)|u-v|^{2H-2}dudv$$
$$\leq (f(t)+\epsilon)^2 H(2H-1)\int_t^{t+r}\int_t^{t+s}|u-v|^{2H-2}dudv$$
$$= (f(t)+\epsilon)^2 C(r,s;t).$$

A similar lower bound holds for $0 < r, s < \delta$ as the follows:

$$\mathbb{E}\left[\left(\int_t^{t+r} f(u)dB_u^H\right)\left(\int_t^{t+s} f(u)dB_u^H\right)\right] \geq (f(t)-\epsilon)^2 C(r,s;t).$$

These two equations lead to the following result:

$$\lim_{r,s\to 0_+} C^{-1}(r,s;t)\mathbb{E}\left[\left(\int_t^{t+r} f(u)dB_u^H\right)\left(\int_t^{t+s} f(u)dB_u^H\right)\right] = f^2(t). \tag{2.6}$$

The case of $r, s \to 0_-$ and $f(t) \leq 0$ can be deduced similarly. This proves that the Assmption 2 is satisfied with $\gamma = 2H$ and C_γ given by (2.4). The case of $\exp(Y_t^H)$ follows from Proposition 2.5. $\square$

Proposition 2.5. *Assume that* $(Z_t)_{t\geq 0}$ *is a Gaussian process that satisfies Assumption 2 uniformly over* $(0,T)$,[1] *with* $\gamma^{(Z)} \in [1,2)$, $L(\cdot)$, *and* $C_\gamma^{(Z)}(\cdot,\cdot)$ *defined as in* (1.4). *For each fixed* $\tau \in (0,T)$ *and a function* $f \in C^2(\mathbb{R})$, *further assume the following:*

(a) $\mathbb{E}[(Z_{\tau+r} - Z_\tau)Z_\tau] = O(|r|)$, $\mathbb{E}[Z_{\tau+r}] - \mathbb{E}[Z_\tau] = O(|r|)$, *as* $r \to 0$.

(b) $\mathbb{E}[(f'(Z_\tau))^4] < \infty$, $\mathbb{E}[\sup_{t\in(\tau-\epsilon,\tau+\epsilon)}(f''(Z_t))^4] < \infty$ *for some* $\epsilon > 0$.

Then, the process $V_t := f(Z_t)$, $t \geq 0$, *satisfies Assumption 2 with* $\gamma^{(V)} = \gamma$ *and* $C_\gamma^{(V)} = \mathbb{E}[(f'(Z_t))^2]C_\gamma^{(Z)}$.

Proof. To begin with, since we assume that Assumption 2 is satisfied uniformly over $(0,T)$ and $\sup_{t\in(0,T)}|L(t)| < \infty$, we can use Kolmogorov-Čentsov continuity theorem to conclude that there is a continuous modification of Z and, thus, hereafter, we assume that $\{Z_t\}_{t\in[0,T]}$

[1] The Assumption 2 is satisfied uniformly over $(0,T)$ if $\sup_{\tau\in(0,T)}(r^2 + s^2)^{-\gamma/2}\,|\mathbb{E}[(V_{\tau+r}-V_\tau)(V_{\tau+s}-V_\tau)] - L(\tau)C_\gamma(r,s)| \to 0$, as $r, s \to 0$, and, also, $\sup_{\tau\in(0,T)}|L(\tau)| < \infty$. This implies the existence of a positive constant C such that $\mathbb{E}[(Z_t - Z_s)^2] \leq C|t-s|^\gamma$, for all $t, s \in (0,T)$.

is a continuous process[2]. Next, by Taylor's expansion, there exists $\theta(\tau, r) \in (\min(\tau, \tau + r), \max(\tau, \tau + r))$ such that, a.s.,

$$f(Z_{\tau+r}) - f(Z_\tau) = f'(Z_\tau)(Z_{\tau+r} - Z_\tau) + f''(Z_{\theta(\tau,r)})(Z_{\tau+r} - Z_\tau)^2.$$

Thus, we have the following decomposition

$$\begin{aligned}
&(f(Z_{\tau+r}) - f(Z_\tau))(f(Z_{\tau+s}) - f(Z_\tau)) \\
&= (f'(Z_\tau))^2(Z_{\tau+r} - Z_\tau)(Z_{\tau+s} - Z_\tau) \\
&\quad + f'(Z_\tau)f''(Z_{\theta(\tau,s)})(Z_{\tau+r} - Z_\tau)(Z_{\tau+s} - Z_\tau)^2 \\
&\quad + f'(Z_\tau)f''(Z_{\theta(\tau,r)})(Z_{\tau+s} - Z_\tau)(Z_{\tau+r} - Z_\tau)^2 \\
&\quad + f''(Z_{\theta(\tau,r)})f''(Z_{\theta(\tau,s)})(Z_{\tau+r} - Z_\tau)^2(Z_{\tau+s} - Z_\tau)^2.
\end{aligned} \tag{2.7}$$

Except for the first term, all other terms are of higher order. As an example, take the second term and note that

$$\begin{aligned}
&\mathbb{E}|f'(Z_\tau)f''(Z_{\theta(\tau,r)})(Z_{\tau+r} - Z_\tau)(Z_{\tau+s} - Z_\tau)^2| \\
&\le \left(\mathbb{E}[(f'(Z_\tau))^4]\mathbb{E}[(f''(Z_{\theta(\tau,s)}))^4]\mathbb{E}[(Z_{\tau+r} - Z_\tau)^4]\mathbb{E}[(Z_{\tau+s} - Z_\tau)^8]\right)^{1/4} \\
&= O((r^2 + s^2)^{3\gamma/4}),
\end{aligned}$$

where the last equality uses (a) and the normality of Z. Indeed, if we define $m_t = \mathbb{E}[Z_t]$ and $z_t = Z_t - m_t$, we have

$$\begin{aligned}
&\mathbb{E}[(Z_{\tau+r} - Z_\tau)^4] \\
&= \mathbb{E}[(z_{\tau+r} - z_\tau)^4] + 6\mathbb{E}[(z_{\tau+r} - z_\tau)^2](m_{\tau+r} - m_\tau)^2 + (m_{\tau+r} - m_\tau)^4 \\
&= \mathbb{E}[(z_{\tau+r} - z_\tau)^4] + o((r^2 + s^2)^\gamma).
\end{aligned}$$

[2]Indeed, for any $0 < s < t < T$, we have $\mathbb{E}[(Z_t - Z_s)^{2k}] = (2k-1)!!(\mathbb{E}[(Z_t - Z_s)^2])^k \le C|t-s|^{k\gamma}$, for some constant C, independent of s and t. Then, we can conclude that there exists a modification of Z that is Hölder continuous of order $(k\gamma - 1)/2k$ and, thus, of any order less than $\gamma/2$.

We proceed to consider the first term of (2.7). With similar argument as the above, we can assume, without loss of generality, that Z has zero mean. Next, since $(Z_\tau, Z_{\tau+r}, Z_{\tau+s})$ are jointly Gaussian, we can define two independent standard normal variables $X(\tau, r, s)$ and $Y(\tau, r, s)$ that are also independent of Z_τ such that

$$\begin{aligned} Z_{\tau+r} - Z_\tau &= a_1 Z_\tau + a_2 X(\tau, r, s) + a_3 Y(\tau, r, s), \\ Z_{\tau+s} - Z_\tau &= b_1 Z_\tau + b_2 X(\tau, r, s) + b_3 Y(\tau, r, s), \end{aligned}$$

for some constants a_i and b_i, $i = 1, 2, 3$, depending on τ, r, and s. Furthermore, a_1 and b_1 are such that

$$a_1 = \frac{\mathbb{E}[(Z_{\tau+r} - Z_\tau)Z_\tau]}{\mathbb{E}[Z_\tau^2]} = O(|r|), \quad b_1 = \frac{\mathbb{E}[(Z_{\tau+s} - Z_\tau)Z_\tau]}{\mathbb{E}[Z_\tau^2]} = O(|s|).$$

Now, since Z satisfies Assumption 2 and $O(|rs|) = o((r^2 + s^2)^{\gamma/2})$, we have

$$\begin{aligned} &a_2 b_2 \mathbb{E}[X(\tau, r, s)^2] + a_3 b_3 \mathbb{E}[Y(\tau, r, s)^2] \\ &= \mathbb{E}[(Z_{\tau+r} - Z_\tau)(Z_{\tau+s} - Z_\tau)] - a_1 b_1 \mathbb{E}[Z_\tau^2] \\ &= C_\gamma(r, s) + o((r^2 + s^2)^{\gamma/2}). \end{aligned}$$

Finally,

$$\begin{aligned} &\mathbb{E}[(f'(Z_\tau))^2 (Z_{\tau+r} - Z_\tau)(Z_{\tau+s} - Z_\tau)] \\ &= a_1 b_1 \mathbb{E}[(f'(Z_\tau))^2 Z_\tau^2] + \mathbb{E}[(f'(Z_\tau))^2](a_2 b_2 \mathbb{E}[X(\tau, r, s)^2] + a_3 b_3 \mathbb{E}[Y(\tau, r, s)^2]) \\ &= \mathbb{E}[(f'(Z_\tau))^2] C_\gamma(r, s) + o((r^2 + s^2)^{\gamma/2}), \end{aligned}$$

and we conclude. □

3 Key Lemmas

The following two key lemmas are used throughout the paper.

Lemma 3.1. *For $\gamma > 0$, assume the following for a function $f : \mathbb{R}^m \to \mathbb{R}$ and functions $K_i : \mathbb{R} \to \mathbb{R}$, $1 \le i \le m$:*

(i) $f(\tau+s_1, ..., \tau+s_m) - f(\tau, ..., \tau) = C_\gamma(s_1, ..., s_m; \tau) + o((s_1^2 + .. + s_m^2)^{\gamma/2})$, as $(s_1, ..., s_m) \to 0$ for any given $\tau \in (0, T)$, where $C_\gamma : \mathbb{R}^m \times [0, T] \to \mathbb{R}$ is a function such that

$$C_\gamma(hs_1, ..., hs_m; \tau) = h^\gamma C_\gamma(s_1, ..., s_m; \tau), \quad s_1, ..., s_m \in \mathbb{R}, h > 0, \tau \in (0, T).$$

(ii) $f \in C([0, T]^m)$.

(iii) For $1 \le i \le m$, K_i satisfies Conditions (2) and (3) of Assumption 3 with a support (A_i, B_i).

Let

$$D_1(f) := \sum_{i_1,...,i_m=1}^{n} \left[\prod_{j=1}^{m} K_{jh}(t_{i_j-1} - \tau) \right] \int_{t_{i_1-1}}^{t_{i_1}} ... \int_{t_{i_m-1}}^{t_{i_m}} f(s_1, ..., s_m) ds_1 ... ds_m$$
$$- \int_{[0,T]^m} \left[\prod_{j=1}^{m} K_{jh}(s_j - \tau) \right] f(s_1, ..., s_m) ds_1 ... ds_m,$$

where $K_{ih}(x) := K_i(x/h)/h$. Then, for each $\tau \in (0, T)$, we have the following:

$$D_1(f) = \frac{1}{2} f(\tau, ..., \tau) \left[\prod_{i=1}^{m} \int K_i(x) dx \right] \sum_{i=1}^{m} \frac{K_i(A_i^+) - K_i(B_i^-)}{\int K_i(x) dx} \frac{\Delta}{h} + o\left(\frac{\Delta}{h} \right).$$

as $h \to 0$ and $\Delta/h \to 0$. If, furthermore, the condition (i) above is satisfied uniformly over $\tau \in (0, T)$, then the approximation above is also uniform over $\tau \in (0, T)$.

Proof. We consider the case that $m = 1$, $K \in C((A, B))$, and K is piecewise C^1 in (A, B), where $A < 0 < B$. In the whole proof, all the summations are taken under the additional constrain that $(\frac{t_{i-1}-\tau}{h}, \frac{t_i-\tau}{h}) \in (A, B)$. Note that this constraint introduces an additional term of order $o(\frac{\Delta}{h})$ (this is because the term at the right boundary is sometimes excluded). We first assume that $K \in C^1((A,B))$, even though the same arguments apply for piecewise C^1 functions. Let us start by noting that

$$\begin{aligned}
D_1 &= \sum_{i=1}^{n} \left[K_h(t_{i-1}-\tau) \int_{t_{i-1}}^{t_i} f(t)dt - \int_{t_{i-1}}^{t_i} K_h(t-\tau) f(t)dt \right] \\
&= \frac{1}{h} \sum_{i=1}^{n} \int_{t_{i-1}}^{t_i} [K(\frac{t_{i-1}-\tau}{h}) - K(\frac{t-\tau}{h})] f(t)dt \\
&= \frac{1}{h} \sum_{i=1}^{n} \int_{t_{i-1}}^{t_i} K'(\frac{s_t-\tau}{h}) \frac{t_{i-1}-t}{h} f(t)dt \\
&= \frac{1}{h} \sum_{i=1}^{n} \int_{t_{i-1}}^{t_i} K'(\frac{t_{i-1}-\tau}{h}) \frac{t_{i-1}-t}{h} f(t)dt \\
&\quad + \frac{1}{h} \sum_{i=1}^{n} \int_{t_{i-1}}^{t_i} [K'(\frac{s_t-\tau}{h}) - K'(\frac{t_{i-1}-\tau}{h})] \frac{t_{i-1}-t}{h} f(t)dt \\
&=: D_{11} + D_{12},
\end{aligned}$$

for some $s_t \in (t_{i-1}, t)$. Above, D_{12} can be controlled as the following:

$$\begin{aligned}
|D_{12}| &\leq \frac{M_f \Delta^2}{2h^2} \sum_{i=1}^{n} \max_{t \in [t_{i-1}, t_i]} |K'(\frac{t_{i-1}-\tau}{h}) - K'(\frac{t-\tau}{h})| \\
&\leq \frac{M_f \Delta^2}{2h^2} V_{-\infty}^{\infty}(K') = O(\frac{\Delta^2}{h^2}),
\end{aligned}$$

where $M_f := \max_{t \in [0,T]} |f(t)|$. Next, we consider D_{11}. Indeed, for any $\delta \in (0, \min(T-\tau, \tau))$, we have

$$\begin{aligned}
D_{11} &= \frac{1}{h} \left(\sum_{|t_i-\tau|<\delta} + \sum_{\delta \leq |t_i-\tau| \leq T} \right) K'(\frac{t_{i-1}-\tau}{h}) \int_{t_{i-1}}^{t_i} \frac{t_{i-1}-t}{h} f(t)dt \\
&\triangleq D_{111} + D_{112}.
\end{aligned} \tag{3.1}$$

The term D_{112} above can be controlled as the following:

$$\begin{aligned}|D_{112}| &\leq \frac{1}{h}\sum_{\delta\leq|t_i-\tau|\leq T}|K'(\frac{t_{i-1}-\tau}{h})|\int_{t_{i-1}}^{t_i}\frac{|t_{i-1}-t|}{h}|f(t)|dt \\ &\leq \frac{M_f\Delta}{2h}\sum_{|t_i-\tau|\geq\delta}\frac{\Delta}{h}|K'(\frac{t_{i-1}-\tau}{h})| = o(\frac{\Delta}{h}),\end{aligned} \tag{3.2}$$

since, due to the absolute integrability of $K'(\cdot)$, for some $s_{i-1}\in(t_{i-1},t_i)$,

$$\begin{aligned}&\sum_{|t_i-\tau|\geq\delta}\frac{\Delta}{h}|K'(\frac{t_{i-1}-\tau}{h})| \\ &\leq \sum_{|t_i-\tau|\geq\delta}\int_{(t_{i-1}-\tau)/h}^{(t_i-\tau)/h}|K'(t)|dt + \frac{\Delta}{h}\sum_{|t_i-\tau|\geq\delta}|K'(\frac{t_{i-1}-\tau}{h}) - K'(\frac{s_{i-1}-\tau}{h})| \\ &\leq \left(\int_{\delta/h}^{+\infty} + \int_{-\infty}^{-\delta/h}\right)|K'(t)|dt + \frac{\Delta}{h}V_{-\infty}^{\infty}(K') = o(1).\end{aligned}$$

To control the term D_{111}, let us remark that, by assumption, there exists $\delta_0\in(0,\min(T-\tau,\tau))$, such that for all $\delta\in(0,\delta_0)$ and $|t-\tau|<\delta$, we have $|f(t)-f(\tau)|\leq\epsilon(\delta)$. Then, when $\Delta<\delta_0$ and fixing $\epsilon(\delta)=A\delta^\gamma$,

$$\begin{aligned}D_{111} &= \frac{1}{h}\left(\sum_{|t_i-\tau|<\delta,K'\leq 0} + \sum_{|t_i-\tau|<\delta,K'>0}\right)K'(\frac{t_{i-1}-\tau}{h})\int_{t_{i-1}}^{t_i}\frac{t_{i-1}-t}{h}f(t)dt \\ &\leq \frac{1}{h}\sum_{|t_i-\tau|<\delta,K'\leq 0}K'(\frac{t_{i-1}-\tau}{h})\int_{t_{i-1}}^{t_i}\frac{t_{i-1}-t}{h}(f(\tau)+\epsilon(\delta))dt \\ &\quad + \frac{1}{h}\sum_{|t_i-\tau|<\delta,K'>0}K'(\frac{t_{i-1}-\tau}{h})\int_{t_{i-1}}^{t_i}\frac{t_{i-1}-t}{h}(f(\tau)-\epsilon(\delta))dt \\ &= -\frac{f(\tau)\Delta}{2h}\sum_{|t_i-\tau|<\delta}\frac{\Delta}{h}K'(\frac{t_{i-1}-\tau}{h}) + \frac{\epsilon(\delta)\Delta}{2h}\sum_{|t_i-\tau|<\delta}\frac{\Delta}{h}|K'(\frac{t_{i-1}-\tau}{h})| \triangleq \overline{D}_{111}.\end{aligned}$$

Similarly, the lower bound can be written as

$$D_{111} \geq -\frac{f(\tau)\Delta}{2h}\sum_{|t_i-\tau|<\delta}\frac{\Delta}{h}K'(\frac{t_{i-1}-\tau}{h}) - \frac{\epsilon(\delta)\Delta}{2h}\sum_{|t_i-\tau|<\delta}\frac{\Delta}{h}|K'(\frac{t_{i-1}-\tau}{h})| \triangleq \underline{D}_{111}.$$

Now we can set $\delta = \sqrt{h}$ and we assume that $\delta < \delta_0$. In the following, all limits are taken when $h \to 0$, $\frac{\Delta}{h} \to 0$ and $\frac{\delta}{h} \to \infty$.

Firstly we consider $\sum_{|t_i-\tau|<\delta} \frac{\Delta}{h} K'(\frac{t_{i-1}-\tau}{h})$. Indeed, there exists $s_{i-1} \in (t_{i-1}, t_i)$, such that $\int_{(t_{i-1}-\tau)/h}^{(t_i-\tau)/h} K'(x)dx = \frac{\Delta}{h} K'(\frac{s_{i-1}-\tau}{h})$. Then, we have

$$\begin{aligned}
&\sum_{|t_i-\tau|<\delta} \frac{\Delta}{h} K'(\frac{t_{i-1}-\tau}{h}) \\
&= \sum_{|t_i-\tau|<\delta} \int_{(t_{i-1}-\tau)/h}^{(t_i-\tau)/h} K'(x)dx + \frac{\Delta}{h} \sum_{|t_i-\tau|<\delta} \left[K'(\frac{t_{i-1}-\tau}{h}) - K'(\frac{s_{i-1}-\tau}{h}) \right] \\
&= \int_{(\delta^-/h, \delta^+/h)\cap(A,B)} K'(x)dx + \frac{\Delta}{h} \sum_{|t_i-\tau|<\delta} \left[K'(\frac{t_{i-1}-\tau}{h}) - K'(\frac{s_{i-1}-\tau}{h}) \right] \\
&= (K(B-) - K(A+)) + o(1).
\end{aligned}$$

since we have

$$\frac{\Delta}{h} \sum_{|t_i-\tau|<\delta} \left[K'(\frac{t_{i-1}-\tau}{h}) - K'(\frac{s_{i-1}-\tau}{h}) \right] \le \frac{\Delta}{h} V_{-\infty}^{\infty}(K') = O(\frac{\Delta}{h}).$$

Here we define $\delta^+ = \max\{t_i - \tau : t_i < \tau + \delta\}, \delta^- = \min\{t_{i-1} - \tau : t_i > \tau - \delta\}$. Note that since $\frac{\Delta}{h} \to 0$, we have $\frac{\delta^+}{h} \to +\infty$, $\frac{\delta^-}{h} \to -\infty$, so we have $\int_{(\delta^-/h, \delta^+/h)\cap(A,B)} K'(x)dx \to K(B-) - K(A+)$. Here we notice that in the case that K' is not continuous in some intervals, the constant $V_{-\infty}^{\infty}(K')$ is replaced by$V_{-\infty}^{\infty}(K') + 2pL$.

Combining previous equations together, we get

$$\overline{D}_{111}, \underline{D}_{111}, D_{111} = \frac{(K(A+) - K(B-))f(\tau)}{2} \frac{\Delta}{h} + o(\frac{\Delta}{h}),$$

and thus we have the first order approximation of D_1 as the following

$$D_1 = \frac{(K(A+) - K(B-))f(\tau)}{2} \frac{\Delta}{h} + o(\frac{\Delta}{h}).$$

From the previous proof, we observe that such a first order approximation is uniform for $\tau \in (0, T)$. The general case can be proved similarly. □

Remark 3.1. *By modifying the proof above, it is possible to show that when f is bounded on $[0,T]$, we have that $D_1(f) = O(\Delta/h)$. Indeed, it suffices to take $\delta = h$ in (3.1) and control D_{112} as in (3.2), while using the boundedness of f to control D_{111}.*

Lemma 3.2. *For $\gamma > 0$, assume the following for a function $f : \mathbb{R}^m \to \mathbb{R}$ and a function $K : \mathbb{R} \to \mathbb{R}$:*

(i) f satisfies the conditions (i) and (ii) of Lemma 3.1,

(ii) K satisfies the conditions (2) and (3) of Assumption 3 with a support (A,B).

Let

$$D_2(f) := \int_{[0,T]^m} \prod_{j=1}^{m} K_h(t_j - \tau) f(t_1, ..., t_m) dt_1 ... dt_m - f(\tau, ..., \tau) \left(\int K(x) dx \right)^m .$$

Then, for all $\tau \in (0,T)$, we have:

$$D_2(f) = h^\gamma \int K(t_1)...K(t_m) C_\gamma(t_1, ..., t_m; \tau) dt_1 \dots dt_m + o(h^\gamma), \qquad (h \to 0).$$

The result remains the same if the integration domain of the first term in $D_2(f)$ is $\mathbb{R}^m$, instead of $[0,T]^m$. Furthermore, if the condition (i) of Lemma 3.1 is satisfied uniformly over $\tau \in (0,T)$, the approximation above holds true uniformly over $\tau \in (0,T)$.

Proof. We will only prove the case of $m = 2$ for simplicity of notation. The proof of general case is quite clear from the proof below. For any $\epsilon \in (0, \min(\tau, T - \tau))$, define $S_\epsilon = \{|t - \tau|, |s - \tau| \le \epsilon\}$ and we divide the integration into two parts:

$$\begin{aligned} &\int_0^T \int_0^T K_h(t-\tau) K_h(s-\tau) f(t,s) dt ds \\ &= \left(\int_{S_\epsilon} + \int_{[0,T]^2 - S_\epsilon} \right) K\left(\frac{t-\tau}{h}\right) K\left(\frac{s-\tau}{h}\right) \frac{1}{h^2} f(t,s) dt ds \triangleq A_\epsilon + B_\epsilon. \end{aligned}$$

Then, we have

$$|B_\epsilon| \le 2M_f \int |K(t)|dt \int_{|s|>\epsilon/h} |K(s)|ds.$$

As to A_ϵ, for all $\delta > 0$, there exists $\epsilon > 0$ small enough, s.t. $\forall |t-\tau|, |s-\tau| < \epsilon$

$$C(t,s;\tau) - \delta(t^2+s^2)^{\gamma/2} \le f(\tau+t,\tau+s) - f(\tau,\tau) \le C(t,s;\tau) + \delta(t^2+s^2)^{\gamma/2}.$$

Note that if we assume that $f(\tau+t,\tau+s) - f(\tau,\tau) = C(t,s;\tau) + o((t^2+s^2)^{\gamma/2})$ uniformly over $\tau \in (0,T)$, then for $\delta > 0$, the ϵ can be picked such that the above holds for all $\tau \in (0,T)$. With this set up, we can get upper bound of $A_\epsilon - f(t)$ as the following, where we define $P = \{K(\frac{t-\tau}{h})K(\frac{s-\tau}{h}) \geq 0\}$:

$$\begin{aligned}
&A_\epsilon - f(\tau,\tau)\left(\int K(s)ds\right)^2 \\
=& \int_{S_\epsilon} K(\frac{t-\tau}{h})K(\frac{s-\tau}{h})\frac{1}{h^2}(f(t,s)-f(\tau,\tau))dtds \\
&- f(\tau,\tau)\int_{\mathbb{R}^2 - S_\epsilon} K(\frac{t-\tau}{h})K(\frac{s-\tau}{h})\frac{1}{h^2}dtds \\
=& \left(\int_{S_\epsilon \cap P} + \int_{S_\epsilon \cap P^C}\right) K(\frac{t-\tau}{h})K(\frac{s-\tau}{h})\frac{1}{h^2}(f(t,s)-f(\tau,\tau))dtds \\
&- f(\tau,\tau)\int_{\mathbb{R}^2 - S_\epsilon} K(\frac{t-\tau}{h})K(\frac{s-\tau}{h})\frac{1}{h^2}dtds \\
\leq& \int_{S_\epsilon \cap P} K(\frac{t-\tau}{h})K(\frac{s-\tau}{h})\frac{1}{h^2}[C(t-\tau, s-\tau;\tau) + \delta(t^2+s^2)^{\gamma/2}]dtds \\
&+ \int_{S_\epsilon \cap P^C} K(\frac{t-\tau}{h})K(\frac{s-\tau}{h})\frac{1}{h^2}[C(t-\tau, s-\tau;\tau) - \delta(t^2+s^2)^{\gamma/2}]dtds \\
&+ 2f(\tau,\tau)\int |K(t)|dt \int_{|s|>\epsilon/h} |K(s)|ds \\
=& h^\gamma \int_{-\epsilon/h}^{\epsilon/h}\int_{-\epsilon/h}^{\epsilon/h} K(t)K(s)C(t,s;\tau)dtds \\
&+ h^\gamma \delta \int_{-\epsilon/h}^{\epsilon/h}\int_{-\epsilon/h}^{\epsilon/h} |K(t)K(s)|(t^2+s^2)^{\gamma/2}dtds \\
&+ 2f(\tau,\tau)\int |K(t)|dt \int_{|s|>\epsilon/h} |K(s)|ds.
\end{aligned}$$

Similarly, the lower bound is the following:

$$\begin{aligned}
A_\epsilon - f(\tau,\tau)\left(\int K(s)ds\right)^2 \geq\ & h^\gamma \int_{-\epsilon/h}^{\epsilon/h}\int_{-\epsilon/h}^{\epsilon/h} K(t)K(s)C(t,s;\tau)dtds \\
&- h^\gamma \delta \int_{-\epsilon/h}^{\epsilon/h}\int_{-\epsilon/h}^{\epsilon/h} |K(t)K(s)|(t^2+s^2)^{\gamma/2}dtds \\
&- 2f(\tau,\tau)\int |K(t)|dt \int_{|s|>\epsilon/h} |K(s)|ds.
\end{aligned}$$

For any δ, we can find satisfactory ϵ and we fix this two numbers. Then we let $h \to 0$. By l'Hopital rule, we have:

$$\lim_{h\to 0} \frac{\int_{\epsilon/h}^{\infty} K(s)ds}{h^\gamma} = \lim_{h\to 0} \frac{\epsilon h^{-2} K(\frac{\epsilon}{h})}{\gamma h^{\gamma-1}} = \lim_{h\to 0} \frac{\epsilon K(\frac{\epsilon}{h})}{\gamma h^{\gamma+1}} = \lim_{x\to\infty} CK(x)x^{\gamma+1} \to 0.$$

Therefore, we have

$$\begin{aligned} &h^{-\gamma}\left|A_\epsilon + B_\epsilon - f(\tau,\tau)\left(\int K(s)ds\right)^2 - h^\gamma \int_{-\epsilon/h}^{\epsilon/h}\int_{-\epsilon/h}^{\epsilon/h} K(t)K(s)C(t,s;\tau)dtds\right| \\ \leq\; &\delta \iint |K(t)K(s)|(t^2+s^2)^{\gamma/2}dtds + h^{-\gamma}|M_f + 2f(\tau,\tau)| \int |K(t)|dt \int_{|s|>\epsilon/h} |K(s)|ds. \end{aligned}$$

Now let $h \to 0$, and notice that δ is arbitrary, we have

$$D_2(f) = h^\gamma \iint K(t)K(s)C(t,s;\tau)dtds + o(h^\gamma).$$

□

4 Optimal Kernel Selection for a Deterministic Volatility

As explained in Figueroa-López & Li (2018), the problem of finding the optimal kernel of order p reduces to minimizing the following functional

$$I_p(K) = \left(\int_0^\infty K^2(x)dx\right)^p \int_0^\infty K(x)x^p dx, \tag{4.1}$$

subject to the constrain $\int_0^\infty K(x)dx = 1/2$. For such a problem, we further limit ourself to kernel function with support $[0,1]$ and use calculus of variation to derive the optimal kernel function. Indeed, for any continuous function $\eta : [0,1] \to \mathbb{R}$ such that $\int_0^1 \eta(x)dx = 0$

and a real number ϵ, we consider

$$I_p(K+\epsilon\eta) = \left(\int_0^\infty (K(x)+\epsilon\eta(x))^2 dx\right)^p \int_0^\infty (K(x)+\epsilon\eta(x))x^p dx.$$

Next, in order to find a local minimum point of $I(K)$, we take the derivative of I with respect to ϵ to get

$$\begin{aligned}\frac{\partial I_p}{\partial \epsilon}\bigg|_{\epsilon=0} &= 2p\left(\int_0^\infty K^2(x)dx\right)^{p-1}\int_0^\infty K(x)\eta(x)dx\int_0^\infty K(x)x^p dx \\ &+ \left(\int_0^\infty K^2(x)dx\right)^p \int_0^\infty \eta(x)x^p dx.\end{aligned}$$

Then, we solve $\frac{\partial I_p}{\partial \epsilon}\Big|_{\epsilon=0} = 0$, which is equivalent to solve

$$2p\int_0^\infty K(x)\eta(x)dx\int_0^\infty K(x)x^p dx + \int_0^\infty K^2(x)dx\int_0^\infty \eta(x)x^p dx = 0.$$

In order for the above to hold for any η satisfying the stated properties, K needs to have the form $K(x) = a(1-bx^p)$ for $a, b > 0$. By plugging such a K in, we get

$$a\int_0^\infty \eta(x)x^p dx\left(-2pb\times\left(\frac{1}{p+1}-\frac{b}{2p+1}\right)+1-\frac{2b}{p+1}+\frac{b^2}{2p+1}\right) = 0.$$

Solving such an equation yields a unique solution $b = 1$ and, by solving $\int_0^1 K(x)dx = 1/2$, we can get $a = \frac{2p}{p+1}$. Therefore, we get a local minimum point of (4.1) as the following:

$$K_p(x) = \frac{p+1}{2p}(1-|x|^p)1_{[-1,1]}.$$

As stated in Figueroa-López & Li (2018), there is still a problem for such a kernel since it may the case the resulting kernel does not satisfy all the conditions of a higher order kernel. Take $p = 4$ as an example. Although K_4 minimizes (4.1), it does not satisfy

$\int_{-1}^{1} K(x)x^2dx = 0$. Therefore, we propose to consider instead the following optimization problem:

$$\text{minimize } I_{2q}(K) = \left(\int_0^1 K^2(x)dx\right)^{2q} \int_0^1 K(x)x^{2q}dx,$$
$$\text{subject to } \int_0^1 K(x)x^{2r}dx = 0, \quad \text{for } 0 < r < q, \quad \text{and } \int_0^1 K(x)dx = \frac{1}{2}.$$

To solve such a problem, we consider the Lagrangian

$$\begin{aligned} I_{2q}^c(K) = & \left(\int_0^1 K^2(x)dx\right)^{2q} \int_0^1 K(x)x^{2q}dx \\ & + \sum_{i=1}^{q-1} \lambda_i \left(\int_0^1 K^2(x)dx\right)^{2q} \int_0^1 K(x)x^{2i}dx \\ & + \lambda_0 \left(\int_0^1 K^2(x)dx\right)^{2q} \left(\int_0^1 K(x)dx - \frac{1}{2}\right), \end{aligned} \tag{4.2}$$

where λ_i are Lagrangian multipliers. In order to solve such an optimization problem, we set $\left.\frac{\partial I_{2q}^c(K+\epsilon\eta)}{\partial\epsilon}\right|_{\epsilon=0} = 0$ and $\frac{\partial I_{2q}^c(K+\epsilon\eta)}{\partial\lambda_i} = 0$. After some simplifications, these yield the system of equations:

$$4q \int_0^1 K(x)\eta(x)dx \int_0^1 K(x)x^{2q}dx + \int_0^1 K^2(x)dx \int_0^1 \eta(x) \left(x^{2q} + \sum_{i=0}^{q-1} \lambda_i x^{2i}\right) dx = 0,$$
$$\int_0^1 K(x)x^{2r}dx = 0, \quad \text{for } 0 < r < q, \qquad \int_0^1 K(x)dx = \frac{1}{2}.$$

Therefore, K needs to take the form

$$K(x) = a\left(x^{2q} + \sum_{i=0}^{q-1} \lambda_i x^{2i}\right),$$

where a and $\lambda_0, \ldots, \lambda_{q-1}$ satisfies the equations:

$$0 = (4q+1)a\left(\frac{1}{4q+1} + \sum_{i=0}^{q-1} \lambda_i \frac{1}{2(q+i)+1}\right) + \frac{\lambda_0}{2},$$

$$0 = \frac{1}{2(q+r)+1} + \sum_{i=0}^{q-1} \lambda_i \frac{1}{2(i+r)+1}, \quad 0 < r < q,$$

$$\frac{1}{2} = a\left(\frac{1}{2q+1} + \sum_{i=0}^{q-1} \lambda_i \frac{1}{2i+1}\right).$$

5 Consistency of the TSRVV estimator

The following Two-time Scale Realized Volatility of Volatility estimator was introduced in Figueroa-López & Li (2018):

$$\widehat{IVV}_{(\mathrm{tsrvv})} = \frac{1}{k}\sum_{i=b}^{n-k-b} (\Delta_i^{(k)}\hat{\sigma}^2)^2 - \frac{n-k+1}{nk}\sum_{i=b+k-1}^{n-k-b} (\Delta_i \hat{\sigma}^2)^2, \tag{5.1}$$

where $\Delta_i \hat{\sigma}^2 = \hat{\sigma}^2_{r,t_{i+1}} - \hat{\sigma}^2_{l,t_i}$, $\Delta_i^{(k)}\hat{\sigma}^2 = \hat{\sigma}^2_{r,t_{i+k}} - \hat{\sigma}^2_{l,t_i}$, and

$$\hat{\sigma}^2_{l,t_i} = \frac{\sum_{j>i} K_h(t_{j-1}-t_i)(\Delta_j^n X)^2}{\Delta \sum_{j>i} K_h(t_{j-1}-\tau)}, \quad \hat{\sigma}^2_{r,t_i} = \frac{\sum_{j\le i} K_h(t_{j-1}-t_i)(\Delta_j^n X)^2}{\Delta \sum_{j\le i} K_h(t_{j-1}-\tau)}. \tag{5.2}$$

The following result shows the consistency of (5.1) and shed some light on the rate of convergence.

Theorem 5.1. *Fix a $t_b \in (0, T/2)$. Then, for the model (1.1) with μ and σ satisfying Assumption 1 and σ being a squared integrable Itô process as in Eq.* (2.2) *(thus satisfying Assumption 2), and a kernel function K satisfying Assumption 3,* (5.1) *is a consistent estimator of $\int_{t_b}^{T-t_b} g_t^2 dt$ with $b = t_b/\Delta$. Furthermore, the convergence rate is given by $O_p(\frac{n^{1/4}}{k^{1/2}}) + O_p(\sqrt{\frac{k}{n}})$ and, thus, k can be chosen to be of the form $Cn^{3/4}$ so that to attain the 'optimal' convergence rate $n^{-1/8}$.*

Proof. The t_b here is basically to rule our boundary effects and for brevity of notation, we will write $t_b = 0$ and assume we have a left side estimator near $t = 0$ and a right side estimator near $T = t$, with the same convergence rate. Define the error terms from the left and right side estimators as $l_i = \hat{\sigma}^2_{l,t_i} - \sigma^2_{t_i}$ and $r_i = \hat{\sigma}^2_{r,t_i} - \sigma^2_{t_i}$, respectively. We will consider the following slightly different estimator:

$$\widehat{IVV}_T^{(\text{tsrvv})} = \frac{1}{k}\sum_{i=0}^{n-k}(\Delta_i^{(k)}\hat{\sigma}^2)^2 - \frac{1}{k}\sum_{i=0}^{n-1}(\Delta_i\hat{\sigma}^2)^2. \tag{5.3}$$

In terms of the error terms r_i and l_i, this can be written as

$$\begin{aligned}\widehat{IVV}_T^{(\text{tsrvv})} = \frac{1}{k}\Bigg[&\sum_{i=0}^{n-k}(\Delta_i^{(k)}\sigma^2)^2 - \sum_{i=0}^{n-1}(\Delta_i\sigma^2)^2 + 2\sum_{i=n-k+1}^{n-1}\sigma^2_{t_i}l_i - 2\sum_{i=1}^{k-1}\sigma^2_{t_i}r_i \\ &+ 2\sum_{i=k}^{n-1}(\sigma^2_{t_i} - \sigma^2_{t_{i-k+1}})r_{i+1} - 2\sum_{i=0}^{n-k}(\sigma^2_{t_{i+k}} - \sigma^2_{t_{i+1}})l_i + 2\sum_{i=0}^{k-1}\sigma^2_{t_i}r_{i+1} \\ &-2\sum_{i=n-k+1}^{n-1}\sigma^2_{t_{i+1}}l_i - \sum_{i=n-k+1}^{n-1}l_i^2 - \sum_{i=1}^{k-1}r_i^2 - 2\sum_{i=0}^{n-k}l_ir_{i+k} + 2\sum_{i=0}^{n-1}l_ir_{i+1}\Bigg].\end{aligned}$$

Now, for each pair of similar terms, we consider the convergence rate of only one of them. The others have the same convergence rate. Indeed, we have

$$\begin{aligned}&\mathbb{E}\left|\sum_{i=k}^{n-1}(\sigma^2_{t_i} - \sigma^2_{t_{i-k+1}})r_{i+1}\right| \le \sqrt{\sum_{i=0}^{n-k}\mathbb{E}[(\sigma^2_{t_{i+k}} - \sigma^2_{t_{i+1}})^2]\sum_{i=0}^{n-k}\mathbb{E}(r^2_{i+1})} = O(k^{\frac{1}{2}}n^{\frac{1}{4}}), \\ &\mathbb{E}\left|\sum_{i=n-k+1}^{n-1}\sigma^2_{t_i}l_i\right| \le \sqrt{\sum_{i=n-k+1}^{n-1}\mathbb{E}(\sigma^4_{t_i})\sum_{i=n-k+1}^{n-1}\mathbb{E}(l_i^2)} = O(\frac{k}{n^{1/4}}), \\ &\mathbb{E}\sum_{i=n-k+1}^{n-1}l_i^2 = O(\frac{k}{\sqrt{n}}), \quad \mathbb{E}\left|\sum_{i=0}^{n-k}l_ir_{i+k}\right| \le \sqrt{\sum_{i=0}^{n-k}\mathbb{E}(l_i^2)\sum_{i=0}^{n-k}\mathbb{E}(r^2_{i+k})} = O(\sqrt{n}).\end{aligned}$$

Similarly, we can see that the difference between (5.1) and (5.3) is $O_p(\Delta)$. Putting all these together, we get

$$\text{TSRVV} - \sum_{i=0}^{n-k} (\Delta_i^{(k)} \sigma^2)^2 - \sum_{i=0}^{n-1} (\Delta_i \sigma^2)^2 = O_p(\frac{n^{1/4}}{k^{1/2}}). \tag{5.4}$$

On the other hand, with similar assumptions and proofs as Theorem 2 and 3 of Zhang et al. (2005), we have the following:

$$\frac{1}{k} \left[\sum_{i=0}^{n-k} (\Delta_i^{(k)} \sigma^2)^2 - \sum_{i=0}^{n-1} (\Delta_i \sigma^2)^2 \right] - \int_0^T g^2(t) dt = O_p(\sqrt{\frac{k}{n}}). \tag{5.5}$$

Therefore, we have

$$\text{TSRVV} - \int_0^T g^2(t) dt = O_p(\frac{n^{1/4}}{k^{1/2}}) + O_p(\sqrt{\frac{k}{n}}),$$

which implies the consistency and also yields that the optimal k is given by $Cn^{3/4}$, in which case the convergence rate is $n^{-1/8}$. □

6 Central Limit Theorems

In what follows, we are going to assume that the relevant processes (such as σ, μ, and the coefficients driving the dynamics of σ) are bounded. This can be justified by localization as in Jacod and Shiryaev Jacod & Shiryaev (2003), Section 5.4, p. 549.

Theorem 6.1. *For the model (1.1) with adapted cádlág μ and σ, and a kernel function K satisfying Assumption 3, we have, for any $\tau \in (0, T)$,*

$$\left(\frac{\Delta}{h} \right)^{-1/2} \left[\sum_{i=1}^{n} K_h(t_{i-1} - \tau)(\Delta_i X)^2 - \int_0^T K_h(t - \tau) \sigma_t^2 dt \right] \to_D \delta_1 N(0, 1), \tag{6.1}$$

where $\delta_1^2 = 2\sigma_\tau^4 \int K^2(x) dx$.

Proof. Let

$$A_n = \left(\frac{\Delta}{h}\right)^{-1/2} \left[\sum_{i=1}^{n} K_h(t_{i-1}-\tau)(\Delta_i X)^2 - \int_0^T K_h(t-\tau)\sigma_t^2 dt\right].$$

Let us start with the approximations:

$$\sum_{i=1}^{n} K_h(t_{i-1}-\tau)(\Delta_i X)^2 = \sum_{i=1}^{n} K_h(t_{i-1}-\tau)\left(\int_{t_{i-1}}^{t_i} \sigma_s dB_s\right)^2 + O_p(\Delta^{1/2})$$
$$\int_0^T K_h(t-\tau)\sigma_t^2 dt = \sum_{i=1}^{n} K_h(t_{i-1}-\tau)\int_{t_{i-1}}^{t_i} \sigma_t^2 dt + o_p\left((\Delta/h)^{1/2}\right).$$

The first approximation above follows from the fact that $\int_{t_{i-1}}^{t_i} \sigma_s dB_s = O_P(\Delta^{1/2})$ and $\Delta \sum_{i=1}^{n} |K_h(t_{i-1}-\tau)| \to \int_0^T |K(x)|dx$, while the second one follows from the proof of Lemma 3.1 and the fact that σ is bounded (see also Remark 3.1). For an alternative proof see Lemma A.1 in Mancini et al. (2004).

We can then write:

$$\begin{aligned} A_n &= \left(\frac{\Delta}{h}\right)^{-1/2} \sum_{i=1}^{n} K_h(t_{i-1}-\tau)\left\{\left(\int_{t_{i-1}}^{t_i} \sigma_s dB_s\right)^2 - \int_{t_{i-1}}^{t_i} \sigma_t^2 dt\right\} + o_P(1) \\ &=: S_n + o_P(1). \end{aligned}$$

Clearly, S_n can be written as a sum $\sum_{i=1}^{n} \alpha_{n,i}$ of martingale differences relative to $\{\mathcal{F}_{n,i} := \mathcal{F}_{t_i}\}_{i=1,\ldots,n}$ with

$$\alpha_{n,i} := \left(\frac{\Delta}{h}\right)^{-1/2} K_h(t_{i-1}-\tau)\left\{\left(\int_{t_{i-1}}^{t_i} \sigma_s dB_s\right)^2 - \int_{t_{i-1}}^{t_i} \sigma_t^2 dt\right\}.$$

To obtained the CLT, we first need to show the following (see Theorem IX.7.28 in Jacod & Shiryaev (2003)):

$$A_n := \sum_{i=1}^{n} \mathbb{E}[\alpha_{n,i}^2 | \mathcal{F}_{n,i-1}] \xrightarrow{P} 2\sigma_\tau^4 \|K\|_2^2. \tag{6.2}$$

First note that, by Itô's lemma,

$$A_n = 4\left(\frac{\Delta}{h}\right)^{-1} \sum_{i=1}^{n} K_h^2(t_{i-1}-\tau) \int_{t_{i-1}}^{t_i} \mathbb{E}\left[\left(\int_{t_{i-1}}^{s} \sigma_u dB_u\right)^2 \sigma_s^2 \middle| \mathcal{F}_{n,i-1}\right] ds.$$

By the Cauchy-Schwarz and the BDG inequalities,

$$\begin{aligned}
&\mathbb{E}\left[\left(\int_{t_{i-1}}^{s} \sigma_u dB_u\right)^2 \left(\sigma_s^2 - \sigma_{t_{i-1}}^2\right) \middle| \mathcal{F}_{n,i-1}\right]^2 \\
&\le \mathbb{E}\left[\left(\int_{t_{i-1}}^{s} \sigma_u dB_u\right)^4 \middle| \mathcal{F}_{n,i-1}\right] \mathbb{E}\left[\left(\sigma_s^2 - \sigma_{t_{i-1}}^2\right)^2 \middle| \mathcal{F}_{n,i-1}\right] \\
&\le C\mathbb{E}\left[\left(\int_{t_{i-1}}^{s} \sigma_u^2 du\right)^2 \middle| \mathcal{F}_{n,i-1}\right] \mathbb{E}\left[\left(\sigma_s^2 - \sigma_{t_{i-1}}^2\right)^2 \middle| \mathcal{F}_{n,i-1}\right] \\
&= O_P(\Delta^{2+\gamma}),
\end{aligned}$$

uniformly on i, due to Assumption 2.[3] Therefore,

$$\begin{aligned}
&\left(\frac{\Delta}{h}\right)^{-1} \sum_{i=1}^{n} K_h^2(t_{i-1}-\tau) \int_{t_{i-1}}^{t_i} \mathbb{E}\left[\left(\int_{t_{i-1}}^{s} \sigma_u dB_u\right)^2 (\sigma_s^2 - \sigma_{t_{i-1}}^2) \middle| \mathcal{F}_{n,i-1}\right] ds \\
&= O_P(\Delta^{(2+\gamma)/2}) h \sum_{i=1}^{n} K_h^2(t_{i-1}-\tau) \\
&= O_P(\Delta^{\gamma/2}),
\end{aligned}$$

[3] We probably need that $L(t)$ is bounded and that the term $o((r^2+s^2)^{\gamma/2})$ in (1.4) is uniform in t.

since $\Delta h \sum_{i=1}^n K_h^2(t_{i-1}-\tau) \to \|K\|_2^2$. We then have that:

$$\begin{aligned}
A_n &= 4\left(\frac{\Delta}{h}\right)^{-1} \sum_{i=1}^n K_h^2(t_{i-1}-\tau) \int_{t_{i-1}}^{t_i} \sigma_{t_{i-1}}^2 \mathbb{E}\left[\left(\int_{t_{i-1}}^s \sigma_u dB_u\right)^2 \middle| \mathcal{F}_{n,i-1}\right] ds + o_P(1) \\
&= 4\left(\frac{\Delta}{h}\right)^{-1} \sum_{i=1}^n K_h^2(t_{i-1}-\tau) \int_{t_{i-1}}^{t_i} \sigma_{t_{i-1}}^2 \mathbb{E}\left[\int_{t_{i-1}}^s \sigma_u^2 du \middle| \mathcal{F}_{n,i-1}\right] ds + o_P(1) \\
&= 4\left(\frac{\Delta}{h}\right)^{-1} \sum_{i=1}^n K_h^2(t_{i-1}-\tau)\sigma_{t_{i-1}}^4 \int_{t_{i-1}}^{t_i} (s-t_{i-1}) ds + o_P(1) \\
&= 2h\Delta \sum_{i=1}^n K_h^2(t_{i-1}-\tau)\sigma_{t_{i-1}}^4 + o_P(1) \\
&\xrightarrow{P} 2\sigma_\tau^4 \|K\|_2^2.
\end{aligned}$$

The following is the final identity needed to conclude the CLT:

$$\sum_{i=1}^n \mathbb{E}[\alpha_{n,i}^4 | \mathcal{F}_{n,i-1}] \xrightarrow{P} 0,$$

for which it suffices to show that

$$\begin{aligned}
T_{1n} &:= \left(\frac{\Delta}{h}\right)^{-2} \sum_{i=1}^n K_h^4(t_{i-1}-\tau) \mathbb{E}\left[\left(\int_{t_{i-1}}^{t_i} \sigma_s dB_s\right)^8 \middle| \mathcal{F}_{n,i-1}\right] \xrightarrow{P} 0, \\
T_{2n} &:= \left(\frac{\Delta}{h}\right)^{-2} \sum_{i=1}^n K_h^4(t_{i-1}-\tau) \mathbb{E}\left[\left(\int_{t_{i-1}}^{t_i} \sigma_s^2 ds\right)^4 \middle| \mathcal{F}_{n,i-1}\right] \xrightarrow{P} 0.
\end{aligned}$$

By BDG inequality, for some constant $C < \infty$,

$$T_{1n} \leq C T_{2n} = O_P(\Delta^2) h^2 \sum_{i=1}^n K_h^4(t_{i-1}-\tau) = O_P\left(\frac{\Delta}{h}\right),$$

since $\Delta h^3 \sum_{i=1}^n K_h^4(t_{i-1}-\tau) \to \int K^4(x)dx$. Finally, we can apply the CLT for martingale differences as in, e.g., Theorem IX.7.28 in Jacod & Shiryaev (2003). □

Proof Of Corollary 6.1. We show the result in the first setting (1) of Theorem 6.2 (the second can be handled similarly). Let U_n and V_n be the first and second terms of the decomposition

$$\begin{aligned}\hat{\sigma}_\tau^2 - \sigma_\tau^2 = \sum_{i=1}^{n} K_h(t_{i-1}-\tau)(\Delta_i X)^2 - \int_0^T K_h(t-\tau)\sigma_t^2 dt \\ + \int_0^T K_h(t-\tau)(\sigma_t^2 - \sigma_\tau^2)dt + o_p(h^\gamma),\end{aligned} \tag{6.3}$$

Let us start by noting that

$$\begin{aligned}\mathbb{E}\left[e^{iuh^{-\gamma/2}(U_n+V_n)}\right] &= \mathbb{E}\left[\mathbb{E}\left[e^{iuh^{-\gamma/2}(U_n+V_n)}\middle| \mathcal{F}(\sigma_s : s \le T)\right]\right] \\ &= \mathbb{E}\left[e^{iuh^{-\gamma/2}V_n}\mathbb{E}\left[e^{iuh^{-\gamma/2}U_n}\middle| \mathcal{F}(\sigma_s : s \le T)\right]\right].\end{aligned}$$

From Theorem 6.1,

$$\mathbb{E}\left[e^{iuh^{-\gamma/2}U_n}\middle| \mathcal{F}(\sigma_s : s \le T)\right] \to e^{-u^2\sigma_\tau^4\int K^2(x)dx},$$

so it suffices to show that

$$\mathbb{E}\left[e^{iuh^{-\gamma/2}V_n - u^2\sigma_\tau^4\int K^2(x)dx}\right] \to \mathbb{E}\left[e^{-\frac{u^2}{2}\left(\delta_1^2+\delta_2^2\right)}\right].$$

For this, first note that, since $\sigma_{\tau-\sqrt{h}} \to \sigma_\tau$, a.s., and, σ is bounded (by virtue of localization), we have

$$\left|\mathbb{E}\left[e^{iuh^{-\gamma/2}V_n - u^2\sigma_\tau^4\int K^2(x)dx}\right] - \mathbb{E}\left[e^{iuh^{-\gamma/2}V_n - u^2\sigma_{\tau-\sqrt{h}}^4\int K^2(x)dx}\right]\right| \to 0.$$

Finally,

$$\mathbb{E}\left[e^{iuh^{-\gamma/2}V_n - u^2\sigma_{\tau-\sqrt{h}}^4\int K^2(x)dx}\right] \to \mathbb{E}\left[e^{-\frac{u^2}{2}\left(\delta_1^2+\delta_2^2\right)}\right],$$

along the same arguments as those used in the proof of Theorem 6.2. □

Theorem 6.2. *Consider the model (1.1) with a cádlag process μ and an Itô process σ given by $\sigma_t^2 = \sigma_0^2 + \int_0^t f_s ds + \int_0^t g_s dW_s$ where W is a Brownian motion such that $\mathbb{E}(dB_t \cdot dW_t) = \rho dt$ and $\{f_t\}_{t\geq 0}$ and $\{g_t\}_{t\geq 0}$ are adapted cádlág processes. Let K be a kernel function satisfying Assumption 3 and, in addition, $K(x) = 0$ for all $x < 0$. Then,*

$$\Delta^{-\frac{\gamma}{2(1+\gamma)}} \left(\hat{\sigma}_\tau^2 - \sigma_\tau^2\right) \to_D \sqrt{\delta_1^2 + \delta_2^2}\, \bar{\xi}.$$

Proof. By virtue of localization (as in Jacod and Shiryaev Jacod & Shiryaev (2003), Section 5.4, p. 549), we can (and will) assume that the relevant processes (such as σ, μ, and the coefficients driving the dynamics of σ) are bounded. We again consider the following decomposition (6.3) and called the first and second terms on the right-hand side $A_{1,n}$ and $A_{2,n}$, respectively. As stated in the theorem, we take $\Delta = h^2$, in which case, the two terms attained the optimal rate $h^{1/2}$. Let us start with the approximations:

$$\sum_{i=1}^n K_h(t_{i-1} - \tau)(\Delta_i X)^2 = \sum_{i=1}^n K_h(t_{i-1} - \tau) \left(\int_{t_{i-1}}^{t_i} \sigma_s dB_s\right)^2 + O_p(\Delta^{1/2})$$
$$\int_0^T K_h(t - \tau)\sigma_t^2 dt = \sum_{i=1}^n K_h(t_{i-1} - \tau) \int_{t_{i-1}}^{t_i} \sigma_t^2 dt + O_p(h).$$

The first approximation above follows from the fact that $\int_{t_{i-1}}^{t_i} \sigma_s dB_s = O_P(\Delta^{1/2})$ and $\Delta \sum_{i=1}^n |K_h(t_{i-1} - \tau)| \to \int_0^T |K(x)|dx$, while the second one follows from the fact that $t \to \sigma_t$ is bounded on $[0, T]$ together with Lemma 3.1 and Remark 3.1 (see also Lemma A.1 in Mancini et al. (2004) for a related result under stronger conditions). We can then write:

$$\begin{aligned} h^{-1/2} A_{1,n} &= h^{-1/2} \sum_{i=1}^n K_h(t_{i-1} - \tau) \left\{ \left(\int_{t_{i-1}}^{t_i} \sigma_s dB_s\right)^2 - \int_{t_{i-1}}^{t_i} \sigma_t^2 dt \right\} + o_P(1) \\ &=: \sum_{i=1}^n \alpha_{n,i} + o_P(1). \end{aligned}$$

For $A_{2,n}$, by similar arguments as those used in the proof of Theorem **??**, we have

$$h^{-1/2}A_{2,n} = h^{-1/2}g_{t_{j-1}}\int_{t_{j-1}}^{T} L\left(\frac{t-\tau}{h}\right)dW_t + o_P(1) =: \sum_{i=1}^{n}\beta_{n,i} + o_P(1),$$

where $t_{j-1} = \min\{t_i : \tau \le t_i\}$. Next, we consider the following sum of martingale differences relative to $\{\mathcal{F}_{n,i} := \mathcal{F}_{t_i}\}_{i=1,\dots,n}$:

$$S_n = \sum_{i=1}^{n}\xi_{n,i} = \sum_{i=1}^{n}\left(\alpha_{n,i} + \beta_{n,i}\right).$$

To apply the CLT for martingale differences (see Theorem IX.7.28 in Jacod and Shiryaev), we first need to show that:

$$\sum_{i=1}^{n}\mathbb{E}[\xi_{n,i}^2|\mathcal{F}_{n,i-1}] \xrightarrow{P} 2\sigma_\tau^4\|K\|_2^2 + g_\tau^2\int\int K(x)K(y)C(x,y)dxdy.$$

To this end, we prove that

$$B_n := \sum_{i=1}^{n}\mathbb{E}[\alpha_{n,i}^2|\mathcal{F}_{n,i-1}] \xrightarrow{P} 2\sigma_\tau^4\|K\|_2^2 \tag{6.4}$$

$$C_n := \sum_{i=1}^{n}\mathbb{E}[\beta_{n,i}^2|\mathcal{F}_{n,i-1}] \xrightarrow{P} g_\tau^2\int\int K(x)K(y)C(x,y)dxdy \tag{6.5}$$

$$D_n := \sum_{i=1}^{n}\mathbb{E}[\alpha_{n,i}\beta_{n,i}|\mathcal{F}_{n,i-1}] \xrightarrow{P} 0. \tag{6.6}$$

The proof of (6.4) is embedded in the proof of Theorem 6.1. For (6.5), note that

$$C_n = h^{-1}g_{t_{j-1}}^2\int_{t_{j-1}}^{T} L^2\left(\frac{t-\tau}{h}\right)dt \to g_\tau^2\int_0^\infty L^2(s)ds,$$

and it is easy to see that $\int_0^\infty L^2(s)ds = \int\int K(x)K(y)C(x,y)dxdy$. It remains to show (6.6). To this end, note that, in terms of $U_{is} := \int_{t_{i-1}}^{s}\sigma_u dB_u$, $\mathbb{E}[\alpha_{n,i}\beta_{n,i}|\mathcal{F}_{n,i-1}]$ can be written as

$$
\begin{aligned}
&2h^{-1}g_{t_{j-1}}\rho K_h(t_{i-1}-\tau)g_{t_{j-1}}\mathbb{E}\left[\int_{t_{i-1}}^{t_i} U_{is}\sigma_s dB_s \int_{t_{i-1}}^{t_i} L\left(\frac{s-\tau}{h}\right) dW_s|\mathcal{F}_{n,i-1}\right]\\
&= 2h^{-1}g_{t_{j-1}}\rho K_h(t_{i-1}-\tau)\mathbb{E}\left[\int_{t_{i-1}}^{t_i} U_{is}\sigma_s L\left(\frac{s-\tau}{h}\right) ds|\mathcal{F}_{n,i-1}\right]\\
&= 2h^{-1}g_{t_{j-1}}\rho K_h(t_{i-1}-\tau)\mathbb{E}\left[\int_{t_{i-1}}^{t_i} U_{is}(\sigma_s-\sigma_{t_{i-1}}) L\left(\frac{s-\tau}{h}\right) ds|\mathcal{F}_{n,i-1}\right],
\end{aligned}
$$

which, by Cauchy-Schwarz inequality, can be shown to be $O_P(h)$, uniformly in i. Thus, since $\Delta\sum_{i=j}^n |K_h(t_{i-1}-\tau)| \to \int |K(x)|dx$, as $n\to\infty$,

$$
\begin{aligned}
D_n &\le 2h^{-1}g_{t_{j-1}}O_P(\Delta)\rho\sum_{i=j}^n |K_h(t_{i-1}-\tau)| \int_{t_{i-1}}^{t_i}\left|L\left(\frac{s-\tau}{h}\right)\right| ds\\
&\le 2h^{-1}g_{t_{j-1}}O_P(\Delta^2)\rho\sum_{i=j}^n |K_h(t_{i-1}-\tau)| = O_P(\Delta/h) = O_P(h).
\end{aligned}
$$

The final identity needed to conclude the CLT is $\sum_{i=1}^n \mathbb{E}[\xi_{n,i}^4|\mathcal{F}_{n,i-1}] \stackrel{P}{\longrightarrow} 0$, for which it suffices to show that

$$
\begin{aligned}
T_{1n} &:= h^{-2}\sum_{i=1}^n K_h^4(t_{i-1}-\tau)\mathbb{E}\left[\left(\int_{t_{i-1}}^{t_i}\sigma_s dB_s\right)^8\middle|\mathcal{F}_{n,i-1}\right] \stackrel{P}{\longrightarrow} 0,\\
T_{2n} &:= h^{-2}\sum_{i=1}^n K_h^4(t_{i-1}-\tau)\mathbb{E}\left[\left(\int_{t_{i-1}}^{t_i}\sigma_s^2 ds\right)^4\middle|\mathcal{F}_{n,i-1}\right] \stackrel{P}{\longrightarrow} 0,\\
T_{3n} &:= h^{-2}g_{t_{j-1}}^4\sum_{i=j}^n \mathbb{E}\left[\left(\int_{t_{i-1}}^{t_i} L\left(\frac{t-\tau}{h}\right) dW_t\right)^4\middle|\mathcal{F}_{n,i-1}\right] \stackrel{P}{\longrightarrow} 0.
\end{aligned}
$$

The previous limits can be shown by applying BDG inequality and using the fact that σ is bounded. □

7 Equivalence of the Approximated Optimal Bandwidth

In this subsection, we compare the performance of the approximated optimal bandwidth to that of the true optimal bandwidth, whenever it exists. In what follows, h_n^* stands for the "the true" optimal bandwidth, which is defined to "minimize" the actual MSE of the kernel estimator, $MSE_n(h) = \mathbb{E}[(\hat{\sigma}^2_{\tau,n,h} - \sigma^2_\tau)^2]$. However, since the mapping $h \to MSE_n(h)$ is not continuous, it is possible that such a global minimum might not exist or be unique. Hence, in what follows, h_n^* is an extended nonnegative real number such that $h_n^* = \lim_{p\to\infty} h^*_{np}$ for a sequence $\{h^*_{np}\}_{p\geq 1}$ satisfying that $MSE_n(h^*_{np}) < \inf_{h\in\mathbb{R}_+} MSE_n(h) + \varepsilon_p$ and a sequence $\{\varepsilon_p\}_{p\geq 1}$ of positive reals converging to 0. Let us also recall that $h_n^{a,opt}$ denotes the approximated optimal bandwidth, which is defined as the minimizer of the MSE and is given by

$$h_n^{a,opt} = n^{-1/(\gamma+1)} \left[\frac{2T\mathbb{E}[\sigma^4_\tau] \int K^2(x)dx}{\gamma L(\tau) \iint K(x)K(y)C_\gamma(x,y)dxdy} \right]^{1/(\gamma+1)}, \tag{7.1}$$

Our goal is to find the relationship between h_n^* and $h_n^{a,opt}$, and between $MSE_n(h_n^*)$ and $MSE_n(h_n^{a,opt})$. We focus on the MSE, though it would be clear that the same arguments applies to the minimizer of the $IMSE$.

The problem is in general hard since the estimator is not continuous with respect to the bandwidth h, when the kernel function $K(\cdot)$ is not continuous in $\mathbb{R}$, which is an important case since kernel functions with finite supports are frequently used in practice (e.g., the uniform kernel function $K_{unif}(x) = 1_{[-1,1]}(x)$). Indeed, when $h \to (t_{i-1} - \tau)_-$, the summation appearing in (1.2) does not include $K_h(t_{i-1} - \tau)(\Delta_i X)^2$, while it does include this term when $h \to (t_{i-1} - \tau)_+$. Although it maybe hard to directly analyze the true MSE analytically, its the first order approximation is given by (**??**) and such an approximation is continuous with respect to h for given n, which makes the problem still tractable. However, the approximated MSE is expected to be close to the true MSE only when $\frac{\Delta}{h}, h \to 0$, but not in other situations, i.e., $h \nrightarrow 0$ or $\frac{\Delta}{h} \nrightarrow 0$. As we will show below, the latter situations are, however, irrelevant when the model under consideration is complex enough. It is worth to remark that typical non-parametric statistical problems consider parameter spaces that are at least as complex as $C^1([0,T])$. However, when the parameter space shrinks to a more trivial case, non-parametric methods may not perform as good as other simpler methods. Hence, in order to rule out some trivial cases, we do need an additional assumption on the complexity of the the model. The following assumption turns out to be enough for our purpose:

Assumption 4. *Assume that for any $t \in (0,T)$, the mapping $(r,s) \mapsto \mathbb{E}[(\sigma_r^2 - \sigma_t^2)(\sigma_s^2 - \sigma_t^2)], r,s \in [0,T]$ is positive definite, for any fixed $t \in (0,T)$.*

It is worth mentioning here that Assumption 4 is not necessary for the kernel estimator to be a consistent estimator. Such an assumption is solely for the purpose of ruling out trivial models so that we can compare the approximated optimal bandwidth with the true optimal bandwidth.

We also need the following simple lemma:

Lemma 7.1. *For the model (1.1) satisfying Assumptions 1, it is not possible to have $t \in (0,T)$, $n \in \mathbb{N}_+$ and $(\alpha_1, ..., \alpha_n) \in \mathbb{R}^n$, such that $\sum_{i=1}^n \alpha_i (\Delta_i X)^2 = \sigma_t^2$ a.s.*

Proof. If we define $\mathscr{G} = \sigma(\sigma_t : t \in [0,T])$, then it is enough to notice that the conditional distribution $\{\Delta_i X\}_{1 \le i \le n}$, given $\mathscr{G}$ is a collection of independent non-trivial Gaussian variables, while $\sigma_t^2 | \mathscr{G} = \sigma_t^2$ is a non-zero constant. □

We now give a simple example in which the Assumption 4 is not satisfied.

Example 1. *For a complete filtered probability space* $(\Omega, \mathscr{F}, \mathbb{F} = \{\mathscr{F}_t\}_{t\geq 0}, \mathbb{P})$*, we define an* $\mathscr{F}$ *measurable random variable* $\xi \sim unif(-c, c)$ *and assume* $\mathscr{F}_t = \sigma(\xi, B_s : s \leq t)$*. Now we consider the following model for* $t \in [0, T]$*:*

$$dX_t = \sigma_t dB_t, \quad \sigma_t^2 = \sigma_0^2 + \xi\sigma_0^2 \sin(\frac{2\pi t}{T}),$$

where $\theta = (\sigma_0, c)$ *is the parameter in the parameter space* $\mathbb{R}_+ \times (0, 1)$*. Assumption 2 can be easily verified. Indeed, we have* $\gamma = 2$*,* $C_\gamma(r, s) = rs$ *and*

$$\mathbb{E}[(\sigma_{t+r}^2 - \sigma_t^2)(\sigma_{t+s}^2 - \sigma_t^2)] = \frac{4\pi^2\sigma_0^4\mathbb{E}[\xi^2]}{T^2} rs + o(r^2 + s^2).$$

We now consider the estimation of $\sigma_{T/2}^2$*. For this model, we actually have* $\sigma_{T/2}^2 = \sigma_0^2 = \int_0^T \sigma_t^2 \cdot \frac{1}{T} dt$*. We then consider the estimator*

$$\hat{\sigma}_{T/2}^2 = \frac{1}{T}\sum_{i=1}^n (\Delta_i X)^2.$$

The bias of such estimator is zero and the variance is given by

$$\begin{aligned} Var\left(\frac{1}{T}\sum_{i=1}^n (\Delta_i X)^2\right) &= \frac{1}{T^2}\left(\sum_{i=1}^n \mathbb{E}[(\Delta_i X)^4] + \sum_{i\neq j} \mathbb{E}[(\Delta_i X)^2(\Delta_j X)^2] - \sigma_{1/2}^4\right) \\ &= \frac{2}{T^2}\sum_{i=1}^n \mathbb{E}\left(\int_{t_{i-1}}^{t_i} \sigma_t^2 dt\right)^2 = O(n^{-1}). \end{aligned}$$

Note that we use the uniform kernel but we do not use a bandwidth that vanishes. The convergence rate here, $O(n^{-1})$*, is better than the one stated in Theorem 3.1 in Figueroa-López & Li (2018), when we consider the kernel estimation with a vanishing bandwidth. It is even better than the convergence rate if we use any "higher order" kernel. Therefore, for this model, a kernel estimator with vanishing bandwidth does not have good performance.*

With the additional assumption of model complexity, we are now able to show that the only possibility for the MSE of the kernel estimator to converge to zero is that both $\frac{\Delta}{h}$ and h converge to zero.

Proposition 7.1. *Define* $\{(n_k, h_k) : k \in \mathbb{N}\}$ *such that* $n_k \in \mathbb{N}$ *and* $h_k \in \mathbb{R}_+$ *and suppose that the model (1.1) satisfies Assumptions 1,* **??***, 2 and 4, and that the kernel function* K *satisfies Assumption 3. Then,* $\lim_{k\to\infty} MSE(n_k, h_{n_k}) = 0$ *if and only if* $\lim_{k\to\infty} \frac{T}{n_k h_{n_k}} = 0$ *and* $\lim_{k\to\infty} h_{n_k} = 0$*.*

We defer the proof of Proposition 7.1 to Appendix A. As we can see from Proposition 7.1, the kernel estimator only converges when the sample size $n \to \infty$. The following lemma enables us to consider the relationship between h_n^* and $h_n^{a,opt}$, whose proof is again given in Appendix A.

Lemma 7.2. *Assume* $F : \mathbb{R}_+ \times \mathbb{R}_+ \to \mathbb{R}_+$ *and* $f(x, y) = Ax + By^\gamma$ *for* $A, B, \gamma > 0$*, such that* $F(x, y) - f(x, y) = o(x) + o(y^\gamma)$ *as* $x, y \to (0^+, 0^+)$*. Also assume that for all* $\delta > 0$*, there exists* $m > 0$*, such that for all* $x, y > \delta$*, we have* $F(x, y) > m$*. Suppose* $z_n \searrow 0$ *and, for each* $n \in \mathbb{N}_+$*,* y_n *and* y_n^* *minimize* $y \mapsto f(z_n/y, y)$ *and* $y \mapsto F(z_n/y, y)$*, respectively. Then we have*

$$\begin{aligned} &\inf_{y\in\mathbb{R}_+} F(z_n/y, y) \to 0, \quad y_n = y_n^* + o(y_n^*), \\ &F(z_n/y_n, y_n) = \inf_{y\in\mathbb{R}_+} F(z_n/y, y) + o(\inf_{y\in\mathbb{R}_+} F(z_n/y, y)), \end{aligned} \tag{7.2}$$

as $n \to \infty$*. Note that* F *might not be continuous, so we say that* y_n^* *minimize* $y \mapsto F(z_n/y, y)$ *in the sense that there exists* $\{y_{np}^* : p \in \mathbb{N}_+\}$*, such that* $\lim_{p\to\infty} y_{np}^* = y_n^*$ *and* $\lim_{p\to\infty} F(z_n/y_{np}^*, y_{np}^*) = \inf_{y\in\mathbb{R}_+} F(z_n/y, y)$*. Note that by assumptions on* F *and* f*,* y_n^* *is finite for* n *large enough.*

Remark 7.1. *The result* $F(z_n/y_n, y_n) = F(z_n/y_n^*, y_n^*) + o(F(z_n/y_n^*, y_n^*))$ *is quite important for our purpose. When connected to the kernel estimator, it means that the departure of approximated bandwidth from the true optimal bandwidth will not significantly affect the true MSE of the kernel estimator.*

With these in hand, we are ready for the result of the relationship between the approximated optimal bandwidth and the true optimal bandwidth.

Theorem 7.1. *For model (1.1) with μ and σ satisfying Assumptions 1, 2, and 4 and a kernel function $K(x)$ satisfying Assumption 3, we have*

$$\begin{aligned} h_n^{a,opt} &= h_n^* + o(h_n^*), \\ MSE_n(h_n^{a,opt}) &= \inf_h MSE_n(h) + o(\inf_h MSE_n(h)), \end{aligned} \tag{7.3}$$

where the superscript "$*$" denotes the true optimal bandwidth and MSE, while "a" denotes the approximated ones.

Proof. Now we write $MSE^*(n,h) = F(\frac{\Delta}{h}, h)$ and $MSE^a(n,h) = f(\frac{\Delta}{h}, h^\gamma) = A\frac{\Delta}{h} + Bh^\gamma$ where the value of A and B can be found in (**??**). From Theorem 3.1 in Figueroa-López & Li (2018)and Proposition 7.1, we know that $F(x,y)$ and $f(x,y)$ satisfy the requirements by Lemma 7.2, where $z_n = \Delta = \frac{T}{n}$. Then, it is immediate to obtain the desired result. □

Remark 7.2. *The theorem above also tells us a fact, that under our model setting, the kernel estimator generally perform better when we observe more data, i.e. the frequency of observation is higher. This seems to be an obvious fact, but is not always true. In the case of using realized variance to estimate the Integrated Volatility with market micro-structure noise, as proved in Zhang et al. (2005), there is an optimal frequency of the data. In such case, increasing the frequency does not yield better performance in general.*

A Other Technical Proofs

Proof of Proposition 2.1

To prove the first part of the result, we write (1.4) as $\mathbb{E}[(V_{t+r} - V_t)(V_{t+s} - V_t)] = C_\gamma(r,s;t) + D(r,s;t)$, where $D(r,s;t) = o((r^2+s^2)^{\gamma/2})$, as $r,s \to 0$. We first show that C_γ is non-negative definite. Indeed, for $n \in \mathbb{N}$, $(x_1, ..., x_n) \in \mathbb{R}^n$, $(c_1, ..., c_n) \in \mathbb{R}^n - \{0\}$ and $h \in \mathbb{R}_+$, we have

$$\sum_{i=1}^{n}\sum_{j=1}^{n} c_i c_j C_\gamma(x_i, x_j; t) = h^{-\gamma}\sum_{i=1}^{n}\sum_{j=1}^{n} c_i c_j C_\gamma(hx_i, hx_j; t)$$
$$= h^{-\gamma}\sum_{i=1}^{n}\sum_{j=1}^{n} c_i c_j \mathbb{E}[(V_{t+hx_i} - V_t)(V_{t+hx_j} - V_t)] - h^{-\gamma}\sum_{i=1}^{n}\sum_{j=1}^{n} c_i c_j D(hx_i, hx_j; t)$$
$$= h^{-\gamma}\mathbb{E}\left[\left(\sum_{i=1}^{n} c_i(V_{t+hx_i} - V_t)\right)^2\right] - h^{-\gamma}\sum_{i=1}^{n}\sum_{j=1}^{n} c_i c_j D(hx_i, hx_j; t).$$

On the right-hand side of the previous equation, we let $h \to 0_+$ and we have that the first term is always non-negative, while the second term converges to zero. This shows the non-negative definiteness of C_γ. The integral non-negative definiteness follows then, since the Riemann integration is defined to be the limit of finite sum, which is always non-negative.

We now show the second assertion of the proposition. First we prove the uniqueness of γ. Suppose there are γ, γ' such that $\gamma' > \gamma > 0$, and correspondingly, C_γ and $C'_{\gamma'}$, that satisfies (1.4). Since C_γ is non-zero, there exists $r, s \in \mathbb{R}$, such that $C_\gamma(r, s; t) \neq 0$. Then,

$$\begin{aligned}\mathbb{E}[(V_{t+rh} - V_t)(V_{t+sh} - V_t)] &= h^\gamma C_\gamma(r, s; t) + o(h^\gamma (r^2 + s^2)^{\gamma/2})\\ &= h^{\gamma'} C'_{\gamma'}(r, s; t) + o(h^{\gamma'}(r^2 + s^2)^{\gamma/2}), \quad h \to 0.\end{aligned}$$

Note now that in the right two parts, all the terms are $o(h^\gamma)$ except $h^\gamma C_\gamma(r, s; t)$. Since we have assumed that $C_\gamma(r, s; t) \neq 0$, this is impossible. Therefore, $\gamma = \gamma'$ and, thus, γ must be unique. Now with the same γ, suppose at some r, s, we have $C_\gamma(r, s) \neq C'_\gamma(r, s)$. Then, a similar argument shows that this leads to a contradiction. This proves the uniqueness of γ and C_γ.

Proof of Proposition 7.1

For notation simplicity, we shall write $\Delta = T/n_k$ and $h = h_{n_k}$. Since we already know $\lim_{k\to\infty} \frac{\Delta}{h} \left(= \lim_{k\to\infty} \frac{T}{n_k h_{n_k}}\right) = 0$ and $\lim_{k\to\infty} h \,(= \lim_{k\to\infty} h_{n_k}) = 0$ are sufficient for the convergence of the MSE$^*(n, h)$ from Theorem 3.1 in Figueroa-López & Li (2018), we only need to prove that the convergence fails in other situations. Note that it is enough to consider the case when both the limit of h and $\frac{\Delta}{h}$ exists (including convergence to infinity), since otherwise we can always choose a subsequence with the same limit of the true MSE. In what follows, we will prove that the true MSE cannot converge to zero in the following cases: (1) $h \to \infty$, (2) $\Delta \nrightarrow 0$, (3) $\Delta \to 0$ and $h \to h_0 > 0$, (4) $\Delta \to 0$, $h \to 0$ and $\frac{\Delta}{h} \to \alpha_0 > 0$.

Firstly, we prove that h cannot converge to infinity. To this end, we observe the following inequality:

$$\begin{aligned} &\mathbb{E}[(\hat{\sigma}_\tau^2 - \sigma_\tau^2)^2] \geq (\mathbb{E}[\hat{\sigma}_\tau^2] - \mathbb{E}[\sigma_\tau^2])^2 \\ &= \left[\sum_{i=1}^{n} K_h(t_{i-1} - \tau) \left(\mathbb{E}\left(\int_{t_{i-1}}^{t_i} \mu_t dt\right)^2 + \int_{t_{i-1}}^{t_i} \mathbb{E}[\sigma_t^2] dt\right) - \mathbb{E}[\sigma_\tau^2]\right]^2. \end{aligned} \tag{A.1}$$

by Assumption 1. If $h \to \infty$, then

$$\begin{aligned} &\left|\sum_{i=1}^{n} K_h(t_{i-1} - \tau) \left(\mathbb{E}\left(\int_{t_{i-1}}^{t_i} \mu_t dt\right)^2 + \int_{t_{i-1}}^{t_i} \mathbb{E}[\sigma_t^2] dt\right)\right| \\ &\quad \leq \sum_{i=1}^{n} |K_h(t_{i-1} - \tau)| \left(\Delta \int_{t_{i-1}}^{t_i} \mathbb{E}[\mu_t^2] dt + \int_{t_{i-1}}^{t_i} \mathbb{E}[\sigma_t^2] dt\right) \\ &\quad \leq \frac{1}{h} n M_K (\Delta^2 + \Delta) M_T \leq \frac{1}{h} M_K T(T+1) M_T \to 0, \quad h \to 0, \end{aligned}$$

where M_T is defined in Assumption 1 and M_K is such that $|K(x)| < M_K$, for all $x \in \mathbb{R}$, whose existence is guaranteed by Assumption 3. Therefore, the R.H.S of (A.1) converges to $(\mathbb{E}[\sigma_\tau^2])^2 > 0$ if $h \to \infty$. We are now able to conclude that we only need to consider that h converges to a finite limit.

Next, we prove that $\Delta \to 0$ must hold. First, assume that $h \to h_0 > 0$. If we do not have $\Delta \to 0$, since n_k can only take integer values, it is enough to consider the case that n_k and, thus, Δ, are fixed. In such a case, we have the following for k large enough:

$$\begin{aligned}\left(\sum_{i=1}^{n} K_h(t_{i-1}-\tau)(\Delta_i X)^2 - \sigma_\tau^2\right)^2 &\le 2\left(\sum_{i=1}^{n} K_h(t_{i-1}-\tau)(\Delta_i X)^2\right)^2 + 2\sigma_\tau^4 \\ &\le \frac{2M_K^2}{h_0^2} n \sum_{i=0}^{n} (\Delta_i X)^4 + 2\sigma_\tau^4,\end{aligned}$$

Note that h_0 and n are fixed and $(\Delta_i X)^4$ and σ_τ^4 have finite expectations by Assumption 1. Therefore, we can implement Dominate Convergence Theorem to conclude that

$$\liminf_{h\to h_0} \mathrm{MSE}_n^*(h) = \mathbb{E}\left[\left(\sum_{i=1}^{n} \liminf_{h\to h_0} K_h(t_{i-1}-\tau)(\Delta_i X)^2 - \sigma_\tau^2\right)^2\right].$$

This equals to zero if and only if

$$\sum_{i=1}^{n} \alpha_i (\Delta_i X)^2 - \sigma_\tau^2 = 0, \text{ a.s.},$$

for some $\alpha_i \in \mathbb{R}$, which is not possible by Lemma 7.1.

We now analyze the case of $h \to 0$ and $\Delta \nrightarrow 0$, where we may still assume a fixed Δ. Consider (A.1) again. From Assumption 3, we know that if $t_{i-1} \neq \tau$, we have

$$\lim_{h\to 0} K_h(t_{i-1}-\tau) = \frac{1}{t_{i-1}-\tau} \lim_{x\to\infty} xK(x) = 0.$$

Therefore, if there exists i_0 such that $t_{i_0} = \tau$, then $\mathbb{E}[\hat{\sigma}_\tau^2]$ converges to either infinity or zero, depending on if $K(0) \neq 0$ or $K(0) = 0$. Otherwise, $\mathbb{E}[\hat{\sigma}_\tau^2]$ always converges to zero. In both cases, we have that the true MSE of the kernel estimator does not converge to zero and, therefore, it must be true that $\Delta \to 0$.

Next, we prove that it is not possible that $\Delta \to 0$ but $h \to h_0 > 0$. Using similar arguments as the proof of Theorem 3.1 in Figueroa-López & Li (2018), we have

$$\mathbb{E}\left(\sum_{i=1}^{n} K_h(t_{i-1}-\tau)(\Delta_i X)^2 - \int_0^T K_h(t-\tau)\sigma_t^2 dt\right)^2 = o(1),$$
$$\mathbb{E}\left(\int_0^T K_h(t-\tau)\sigma_t^2 dt - \int_0^T K_{h_0}(t-\tau)\sigma_t^2 dt\right)^2 = o(1).$$

In the first equality, we use Lemma 3.1 and in the second equality, we notice that $K_{h_0}(t-\tau) \neq \lim K_h(t-\tau)$ for only finite many t. By Assumption 4, $\mathbb{E}\left(\int_0^T K_{h_0}(t-\tau)\sigma_t^2 dt - \sigma_\tau^2\right)^2 \neq 0$. As a result, we have proved that the third case is not possible.

Finally, we need to consider the case that $\Delta \to 0$, $h \to 0$ and $\frac{\Delta}{h} \to \alpha_0 > 0$. We notice that

$$Cov((\Delta_i X)^2, (\Delta_j X)^2 | \sigma(\mu, \sigma)) = 0,$$
$$Cov((\Delta_i X)^2, (\Delta_i X)^2 | \sigma(\mu, \sigma)) = 2\left[\int_{t_{i-1}}^{t_i} \sigma_t^2 dt\right]^2,$$

Thus, we have

$$\begin{aligned} &Var\left[\sum_{i=1}^{n} K_h(t_{i-1}-\tau)(\Delta_i X)^2 | \sigma(\mu, \sigma)\right] \\ &= \sum_{i=1}^{n}\sum_{j=1}^{n} K_h(t_{i-1}-\tau)K_h(t_{j-1}-\tau) Cov((\Delta_i X)^2, (\Delta_j X)^2 | \sigma(\mu, \sigma)) \\ &= 2\sum_{i=1}^{n} K_h^2(t_{i-1}-\tau)\left[\int_{t_{i-1}}^{t_i} \sigma_t^2 dt\right]^2. \end{aligned}$$

Above, there are two possibilities. The first one is that $\lim_{n\to\infty}\sum_{i=1}^n K^2\left(\frac{t_{i-1}-\tau}{h}\right) = 0$, which implies $K\frac{t_{i-1}-\tau}{h} \to 0$ and thus, by Dominate Convergence Theorem, the estimator $\hat{\sigma}_\tau^2$ converges to zero in probability. The second case is that $\lim_{n\to\infty}\sum_{i=1}^n K^2\left(\frac{t_{i-1}-\tau}{h}\right) > 0$, in which case $\sum_{i=1}^n K_h^2(t_{i-1}-\tau)\left[\int_{t_{i-1}}^{t_i} \sigma_t^2 dt\right]^2$ is bounded away from zero and thus the conditional variance is not zero. In both cases, the estimator does not converge to the true spot volatility.

Proof of Lemma 7.2

Fix an arbitrary $\epsilon > 0$. Then, there exists $\delta > 0$, such that

$$(1-\epsilon)f(x,y) < F(x,y) < (1+\epsilon)f(x,y), \text{ for all } (x,y) \in (0,\delta)\times(0,\delta).$$

Let $m > 0$ be such that

$$F(x,y) > m, \text{ for all } (x,y) \in \mathbb{R}_+ \times \mathbb{R}_+ - (0,\delta)\times(0,\delta).$$

and let $I_n(\delta) := \left\{ y \in \mathbb{R}_+ : \left(\frac{z_n}{y}, y\right) \in (0,\delta)\times(0,\delta) \right\}$. Notice that $z_n \searrow 0$, so there exists $N \in \mathbb{N}_+$, such that for all $n > N$,

$$\max\left(\left(\frac{B\gamma z_n^\gamma}{A}\right)^{1/(\gamma+1)}, \left(\frac{Az_n}{B\gamma}\right)^{1/(\gamma+1)} \right) < \delta,$$
$$(1+\epsilon)\left[(\gamma A^\gamma B z_n^\gamma)^{1/(\gamma+1)} + (A^\gamma B z_n^\gamma/\gamma^\gamma)^{1/(\gamma+1)} \right] < m.$$

These implies that for all $n > N$, we have

$$\begin{aligned}\inf_{y\in I_n(\delta)} F(z_n/y, y) &\le \min_{y\in I_n(\delta)} f(z_n/y,y)(1+\epsilon) \\ &= f\left(\left(\frac{B\gamma z_n^\gamma}{A}\right)^{1/(\gamma+1)}, \left(\frac{Az_n}{B\gamma}\right)^{1/(\gamma+1)} \right)(1+\epsilon) \\ &= (1+\epsilon)\left[(\gamma A^\gamma B z_n^\gamma)^{1/(\gamma+1)} + (A^\gamma B z_n^\gamma/\gamma^\gamma)^{1/(\gamma+1)} \right] < m.\end{aligned}$$

Combining these three inequalities, we have that for $n > N$,

$$\begin{aligned}\inf_{y\in\mathbb{R}_+} F(z_n/y,y) &= \min\left(\inf_{y\in I_n(\delta)} F(z_n/y,y), \inf_{y\in I_n(\delta)^C} F(z_n/y,y) \right) \\ &= \inf_{y\in I_n(\delta)} F(z_n/y,y) < m.\end{aligned}$$

Without loss of generality, we may assume that the above holds for all n. Now we define $y_n = \left(\frac{Az_n}{B\gamma}\right)^{1/(\gamma+1)}$ and we have

$$(1-2\epsilon)f(z_n/y_n, y_n) \le \inf_{y\in\mathbb{R}_+} F(z_n/y, y) \le F(z_n/y_n, y_n) < (1+2\epsilon)f(z_n/y_n, y_n).$$

Therefore, we have

$$\inf_{y\in\mathbb{R}_+} F(z_n/y, y) \to 0,$$
$$F(z_n/y_n, y_n) = \inf_{y\in\mathbb{R}_+} F(z_n/y, y) + o(\inf_{y\in\mathbb{R}_+} F(z_n/y, y)), \quad n \to \infty.$$

Then, by definition of y_n^*, there exists y_n^{**} such that $y_n^*/y_n^{**} \to 1$ and the following holds:

$$(1-2\epsilon)f(z_n/y_n, y_n) \le \inf_{y\in\mathbb{R}_+} F(z_n/y, y) \le F(z_n/y_n^{**}, y_n^{**}) < (1+2\epsilon)f(z_n/y_n, y_n).$$

The existence of such y_n^{**} is guaranteed by $\inf_{y\in\mathbb{R}_+} F(z_n/y, y) \le (1+\epsilon)f(z_n/y_n, y_n) < (1+2\epsilon)f(z_n/y_n, y_n)$ and the fact that $\{y_{np}^* : p \in \mathbb{N}\} \cap \{y : F(z_n/y, y) < (1+2\epsilon)f(z_n/y_n, y_n)\} \cap \left(y_n^*\left(1-\frac{1}{n}\right), y_n^*\left(1+\frac{1}{n}\right)\right)$ is not empty.

We claim that the inequalities above imply

$$f(z_n/y_n^{**}, y_n^{**}) < \alpha f(z_n/y_n, y_n),$$

where $\alpha = \frac{1+2\epsilon}{1-2\epsilon}$. Otherwise, we will have

$$F(z_n/y_n^{**}, y_n^{**}) > (1-2\epsilon)f(z_n/y_n^{**}, y_n^{**}) > (1+2\epsilon)f(z_n/y_n, y_n) > F(z_n/y_n^{**}, y_n^{**}),$$

which is a contradiction. Since this is true for all $\epsilon > 0$, we then have $\lim_{n\to\infty} y_n/y_n^{**} = 1$, which implies $\lim_{n\to\infty} y_n/y_n^* = 1$. This completes the proof.

References

Figueroa-López, J. & Li, C. (2018), 'Supplement to "optimal kernel estimation of spot volatility of stochastic differential equations"'.

Jacod, J. & Shiryaev, A. N. (2003), *Limit Theorems for stochastic processes*, Springer-Verlag, Berlin.

Mancini, C., Mattiussi, V. & Renò, R. (2004), 'Spot volatility estimation using delta sequences', *Scandinavian Actuarial Journal* **2004**(1), 42–52.

Zhang, L., Mykland, P. & Aït-Sahalia, Y. (2005), 'A tale of two time scales', *Journal of the American Statistical Association* **100**(472).