KERNEL ESTIMATION OF SPOT VOLATILITY WITH MICROSTRUCTURE NOISE USING PRE-AVERAGING

JOSÉ E. FIGUEROA-LÓPEZ
Washington University in St. Louis

BEI WU
Washington University in St. Louis

We revisit the problem of estimating the spot volatility of an Itô semimartingale using a kernel estimator. A central limit theorem (CLT) with an optimal convergence rate is established for a general two-sided kernel. A new pre-averaging/kernel estimator for spot volatility is also introduced to handle the microstructure noise of ultra high-frequency observations. A CLT for the estimation error of the new estimator is obtained, and the optimal selection of the bandwidth and kernel function is subsequently studied. It is shown that the pre-averaging/kernel estimator's asymptotic variance is minimal for two-sided exponential kernels, hence justifying the need of working with kernels of unbounded support. Feasible implementation of the proposed estimators with optimal bandwidth is developed as well. Monte Carlo experiments confirm the superior performance of the new method.

1. INTRODUCTION

Itô semimartingale models for the dynamics of asset returns have been widely used in financial econometrics. Such a process takes the form

$$dX_t = \mu_t dt + \sigma_t dW_t + dJ_t, \tag{1.1}$$

where $\{W_t\}_{t\geq 0}$ is a standard Brownian motion (BM) and $\{J_t\}_{t\geq 0}$ is the jump component. The spot volatility σ_t is a key feature of the model as it plays a crucial rule in option pricing, portfolio management, and financial risk management. Over the last decade, there has been growing interest in the estimation of volatility due to the wide availability of high-frequency data. In this work, we are concerned with spot volatility estimation in an Itô semimartingale model via kernel smoothing.

558

This is one of the most widely used nonparametric methods in statistics, dating back to the seminal work of Parzen (1962) and Rosenblatt (1956).

One of the earliest works on kernel-based estimation of spot volatility dates back to Foster and Nelson (1996), where they studied a weighted rolling window estimator, which is essentially a kernel estimator with compact support. Asymptotic normality was established under abstract conditions that were not directly stated in terms of the coefficients of the Itô semimartingale (1.1). Specifically, they worked with a discretized time series approximation of the model (1.1). Fan and Wang (2008) established the asymptotic normality for a general kernel estimator, this time working directly with the model (1.1) under relatively mild conditions on the coefficients, but without jumps. However, the result therein also required a certain condition on the convergence rate of the bandwidth to zero, which allowed them to neglect the "target error" coming from approximating the spot volatility by a kernel weighted volatility. As a result, the convergence rates of the estimators were suboptimal (see Section 6 of Figueroa-López and Li (2020a) for more details). Kristensen (2010) also proved a central limit theorem (CLT) for kernel-based estimators under the absence of jumps and a no-leverage condition (i.e., σ and W were assumed to be independent). Yu et al. (2014a) generalized Kristensen's result by allowing a jump component of finite activity (FA), but still assuming no-leverage effects. Mancini, Mattiussi, and Renò (2015) studied more general Itô semimartingales, but again FA jumps. All these works only considered CLTs with *suboptimal convergence rates*. Alvarez et al. (2012) proposed an estimator of σ_t^p by considering forward finite-difference approximations of the realized power variation process of order p, which is essentially a forward-looking kernel estimator with a uniform kernel. Jacod and Protter (2011) (see Section 13.3 therein) considered both backward and forward finite-difference approximations of the realized quadratic variation. Both works obtained the best possible convergence rates for their CLTs for a rather general Itô semimartingale model (in the case of Jacod and Protter (2011), also including jumps). We also refer to Aït-Sahalia and Jacod (2014, Chap. 8) for a more detailed review of the relevant literature.

More recently, Figueroa-López and Li (2020a) studied the leading order terms of the mean-squared error (MSE) of kernel-based estimators for continuous Itô semimartingales under a certain local condition on the covariance function of the spot variance σ_t^2, which covers not only Brownian driven volatilities, but also those driven by fractional BM and other Gaussian processes. Using the asymptotics for the MSE, the optimal convergence rate was established and formulas for the optimal bandwidth and kernel functions were derived under a no-leverage condition. CLTs for general right-sided kernel estimators were also obtained (see also Remark 8.10 in Aït-Sahalia and Jacod (2014), where a result for a general right-sided kernel with compact support was stated without proof). One of the objectives of the present work is then to extend the results of Figueroa-López and Li (2020a) and Yu et al. (2014a),[1] and prove a CLT for a general *two-sided*

---

[1] As explained above, Yu et al. (2014a) established a CLT for a general two-sided kernel but with a suboptimal converge rate, FA jumps, and no leverage.

kernel of unbounded support, with optimal convergence rate and in the presence of jumps and leverage effects. As proved in this paper in greater generality, such kernels can have better performance than either one-sided or compactly supported kernels. More specifically, we show that, at the optimal convergence rate regime (i.e., assuming the bandwidth b_n is such that $b_n \Delta_n^{-1/2} \overset{n \to \infty}{\longrightarrow} \beta \in [0, \infty]^2$ and choosing β to optimize the convergence rate of the estimation error), the double exponential kernel $K(x) = \exp(-|x|)/2$ minimizes the resulting asymptotic variance of the estimation error, hence extending the findings of Figueroa-López and Li (2020a) to a more general setting. This result was stated in Foster and Nelson (1996) partially based on a heuristic justification, since the formula of the asymptotic variance used to derive it required the kernel to be compactly supported, while the exponential kernel is clearly not of this type. In the context of kernel density estimation, van Eeden (1985) showed that double exponential kernels are optimal if the density is not continuous. As stated in Kristensen (2010), when the volatility has second derivatives that are Hölder continuous, the Epanechnikov kernel $\mathbf{1}_{[-1, 1]}(x)3(1 - x^2)/4$ is optimal (see also Section 4 of Figueroa-López and Li (2020b) for other related results).

While the results described in the previous paragraph are important for intermediate intraday frequencies (e.g., 1–5 min), it is widely accepted that financial returns at ultra high frequency are contaminated by market microstructure noise. Specifically, high-frequency asset prices exhibit several stylized features, which cannot be accounted for by Itô semimartingales, such as clustering noises, bid/ask bounce effects, and roundoff errors (cf. Aït-Sahalia and Jacod (2014, Chap. 2)). Such discrepancies between macro and micro movements are typically modeled by an additive noise. The literature of statistical estimation methods under microstructure noise has grown extensively since last decade and is still a highly researched subject (see Zhang, Mykland, and Aït-Sahalia (2005), Hansen and Lunde (2006), Bandi and Russell (2008), Barndorff-Nielsen et al. (2008), Podolskij and Vetter (2009), Jacod et al. (2009), and Mykland and Zhang (2012) for a few seminal works in the area as well as the monograph Aït-Sahalia and Jacod (2014)). Most of the existing literature on volatility estimation for high-frequency data with microstructure noise has mainly focused on the estimation of the integrated volatility or variance (IV), defined as $IV_T = \int_0^T \sigma_t^2 dt$. Zhang et al. (2005) showed that scaled by $(2n)^{-1}$, the realized variance estimator, the gold standard for IV estimation in the absence of microstructure noise, consistently estimates the variance of the microstructure noise, instead of the integrated volatility, as the sampling frequency n increases. There are several approaches to overcome this problem: the Two-Scale Realized Variance (TSRV) estimator by Zhang et al. (2005) and the efficient Multiscale Realized Variance by Zhang (2006); the Realized Kernel estimator by Barndorff-Nielsen et al. (2008); the pre-averaging method by Podolskij and Vetter (2009) and Jacod et al. (2009); and the Quasi-Maximun Likelihood Estimator by Xiu (2010).

---

[2]This condition always holds through a subsequence of b_n.

Spot volatility estimation is often viewed as a by-product of integrated volatility estimation since, in principle, we can recover the spot volatility σ_t^2 as a finite-difference approximation of an estimate of the integrated volatility. Following this idea, Zu and Boswijk (2014) constructed a Two-Scale Realized Spot Variance (TSRSV) estimator based on the TSRV integrated variance estimator of Zhang et al. (2005). They proved consistency and derived the asymptotic distribution of the estimation error with a convergence rate of $n^{-1/12}$, which is suboptimal.

The second objective of our work is to construct a kernel-based estimator of the spot volatility based on the pre-averaging integrated variance estimator of Jacod et al. (2009). The basic idea is simple and natural. If we denote $\widehat{IV}_t^{pre-av}$ the pre-averaging estimator of $IV_t = \int_0^t \sigma_s^2 ds$, our estimators combine this with a kernel localization technique as follows:

$$\hat{\sigma}_t^2 = \int_0^t \frac{1}{b_n} K\left(\frac{s-t}{b_n}\right) dIV_s^{pre-av},$$

where K is a suitable kernel function and $b_n > 0$ is the bandwidth, which should converge to 0 at an appropriate rate. We establish the asymptotic mix normality of our estimator and identify two asymptotic regimes for two different bandwidth convergence regimes. One of those regimes yields the optimal convergence rate of $n^{-1/8}$ for our estimator. It is important to point out that the asymptotic theory for the kernel/pre-averaging estimator cannot be derived from that for the pre-averaging integrated variance and also is substantially different and harder than that for kernel based estimators in the absence of microstructure noise.

Although combining pre-averaging and kernel smoothing is a natural idea, to the best of our knowledge, there are only two related results in the literature. Aït-Sahalia and Jacod (2014) (see Section 8.7 therein), stated, without proof, a stable convergence result of a pre-averaging estimator for the spot volatility of a continuous Itô semimartingale,[3] but only in the case of a one-sided uniform kernel $K(t) = \mathbf{1}_{[0,1]}(t)$ (see also Chen (2019) for a similar estimator). Here, we consider a truncated version to handle the jumps and a general two-side kernel (see below as to the need of considering such kernels). Yu et al. (2014b) also proposed a pre-averaging kernel estimator for the spot volatility, slightly different from our estimator. They established asymptotic normality with suboptimal convergence rate for their untruncated estimator in the case of a continuous Itô semimartingale, and for their truncated estimator in the presence of Lévy jumps of bounded variation. In both situations, a no-leverage condition was adopted. In our case, we consider not only leverage effects, but also more general jump processes, not necessarily of Lévy type and with no restriction in the index of jump activity, under both the suboptimal and optimal convergence rate regimes.

---

[3]The estimator therein is different from ours. Our estimator includes a debiasing term, which is omitted in Aït-Sahalia and Jacod (2014). Our Monte Carlo experiments show that such a correction is important in finite samples.

As an important application of our results, we study the problem of bandwidth and kernel function selection. More specifically, by assuming that the bandwidth b_n satisfies $b_n \Delta_n^{-1/4} \overset{n \to \infty}{\longrightarrow} \beta \in [0, \infty]$, we show that the choice $\beta \in (0, \infty)$ makes the convergence rate of the estimation error optimal (i.e., converging to 0 faster). We then determine the value of $\beta^\star \in (0, \infty)$ that minimizes the asymptotic variance of the estimation error. Finally, when choosing such a bandwidth $b_n^\star = \beta^\star \Delta_n^{1/4}$, we determine the kernel function that minimizes the limiting variance. As in the absence of microstructure noise, we deduce that the optimal kernel is a two-sided exponential or Laplace function $K(x) = \frac{1}{2} e^{-|x|}$. This fact justifies the necessity of developing the asymptotic theory for general kernels of unbounded support over the more widely used uniform kernels. If we were constrained to compactly supported kernels in the suboptimal asymptotic regime, a uniform kernel would be the best, but this is no longer the case if we allow kernels with unbounded support and/or consider an optimal convergence rate regime. Similarly, two-sided kernels will perform better than one-sided, even if compactly supported. Let us finally remark that the optimality criteria adopted in this work are not those of minimizing the MSE of the estimator, whose asymptotic behavior is much more challenging to derive in the model generality intended here. In the absence of microstructure noise, leverage effects, and jumps, Figueroa-López and Li (2020a) did obtain formulas for the leading terms of the MSE and the bandwidth that minimize them.

The implementation of the optimum bandwidth (at the optimum rate) is more challenging because it involves the volatility of volatility (i.e., the standard deviation of the volatility process) and the spot volatility itself. Hence, to implement it, we develop a new method, which iteratively estimates the spot volatility, the vol vol, and the optimal bandwidth. Using Monte Carlo simulation, we compare our estimator with the TSRSV estimator of Zu and Boswijk (2014) and show a significant improved accuracy. We also illustrate the improvement achieved by the optimal exponential kernel and the calibrated optimal bandwidth via our iterative method.

We finish the Introduction by giving one more reason for the importance of estimating the spot volatility. As mentioned above, while spot volatility estimation can, at least conceptually, be seen as a by-product of integrated variance estimation, interestingly enough, one can also use spot volatility estimation as an intermediate step toward the estimation of integrated volatility functionals of the form $I_T(g) := \int_0^T g(\sigma_s^2) ds$. Specifically, once an estimator $\hat{\sigma}_t^2$ of σ_t^2 has been developed, one can naturally devise an estimator for $I_T(g)$ of the form $\hat{I}_T(g) = \Delta_n \sum_{i=1}^n g(\hat{\sigma}_{t_i}^2)$, where $t_i = i\Delta_n$ and $\Delta_n = T/n$, followed by an appropriate bias correction adjustment. In the absence of noise, Jacod and Rosenbaum (2013), Mykland and Zhang (2009), and Li, Liu, and Xiu (2019) have developed methods for the estimation of these functionals (see also Li and Xiu (2016), Li, Todorov, and Tauchen (2017), and Aït-Sahalia and Xiu (2019) for related methods and other applications thereof). Recently, Chen (2019) developed an estimator for $\hat{I}_T(g)$ based on a forward finite-difference approximation of the standard pre-averaging estimator of the integrated variance.

The rest of the paper is organized as follows: Section 2 introduces the setting of the problem and the main result. Section 3 shows an application of our main theorem: the optimal parameter and kernel selection. The simulations are provided in Section 4. Some conclusions are given in Section 5. The proofs of our main results can be found in Appendixes A and B.

2. THE SETTING, ESTIMATOR, AND MAIN RESULTS

Throughout, we consider an Itô semimartingale of the form:

$$X_t = X_0 + \int_0^t \mu_s ds + \int_0^t \sigma_s dW_s$$
$$+ \int_0^t \int_E \delta(s,z) \mathbb{1}_{\{|\delta(s,z)| \le 1\}} (\mathfrak{p}-\mathfrak{q})(ds,dz) + \int_0^t \int_E \delta(s,z) \mathbb{1}_{\{|\delta(s,z)| > 1\}} \mathfrak{p}(ds,dz),$$
(2.1)

where all stochastic processes ($\mu := \{\mu_t\}_{t\ge 0}$, $\sigma := \{\sigma_t\}_{t\ge 0}$, $W := \{W_t\}_{t\ge 0}$, $\mathfrak{p} := \{\mathfrak{p}(B) : B \in \mathcal{B}(\mathbb{R}_+ \times E)\}$) are defined on a complete filtered probability space $(\Omega^{(0)}, \mathcal{F}^{(0)}, \mathbb{F}^{(0)}, \mathbb{P}^{(0)})$ with filtration $\mathbb{F}^{(0)} = (\mathcal{F}_t^{(0)})_{t\ge 0}$ and are assumed to satisfy standard conditions for X to be well defined. Here, W is a standard BM adapted to the filtration $\mathbb{F}^{(0)}$, and $\mathfrak{p}$ is a Poisson random measure on $\mathbb{R}_+ \times E$ for some arbitrary Polish space E with compensator $\mathfrak{q}(du, dx) = du \otimes \lambda(dx)$, where λ is a σ-finite measure on E having no atom. For further details regarding Itô semimartingales, see Section 2.1.4 of Jacod and Protter (2011).

We denote the spot variance process $c_t = \sigma_t^2$ and assume that it is also an Itô semimartingale with the following dynamics:

$$c_t = c_0 + \int_0^t \tilde{\mu}_s ds + \int_0^t \tilde{\sigma}_s dB_s + \int_0^t \int_E \tilde{\delta}(s,z)(\mathfrak{p}-\mathfrak{q})(ds,dz),$$
(2.2)

where $B := \{B_t\}_{t\ge 0}$ is a standard BM adapted to $\mathbb{F}^{(0)}$ so that $d\langle W,B\rangle_t = \rho_t dt$. Here, $\{\tilde{\mu}_t\}_{t\ge 0}$ is adapted, locally bounded; $\{\rho_t\}_{t\ge 0}$ is adapted, locally bounded, càdlàg; $\{\tilde{\sigma}_t\}_{t\ge 0}$ is adapted, càdlàg; and $\tilde{\delta}$ is a predictable function on $\mathbb{R}_+ \times E$ satisfying standard conditions for the process above to be well defined (see Jacod and Protter, 2011).

We now state the main assumption on the process X.

Assumption 1. The process X satisfies (2.1) with $c_t = \sigma_t^2$ satisfying (2.2) and, for some $r \in [0,2]$, measurable functions $\Gamma_m, \Lambda_m : E \to \mathbb{R}_+$, constants $C_m < \infty$, and a localizing sequence of stopping times $(\tau_m)_{m\ge 1}$ such that $\tau_m \to \infty$, we have

$$t \in [0, \tau_m] \Longrightarrow \begin{cases} |\mu_t| + |\sigma_t| + |\tilde{\mu}_t| + |\tilde{\sigma}_t| \le C_m, \\ |\delta(t,z)| \wedge 1 \le \Gamma_m(z), \quad \text{where } \int \Gamma_m(z)^r \lambda(dz) < \infty, \\ \left|\tilde{\delta}(t,z)\right| \wedge 1 \le \Lambda_m(z), \quad \text{where } \int \Lambda_m(z)^2 \lambda(dz) < \infty. \end{cases}$$

The parameter r plays a key role in our asymptotic results. In short, r determines the jump activity of the process: the larger r is, the more active or frequent are the small jumps of the process. When $r = 0$, the process exhibits finite many jumps in any bounded time interval (in that case, we say that the jumps are of FA). When $r < 1$ ($r > 1$), the jump component of the process is of bounded (unbounded) variation.

To establish the CLT for the kernel estimator $\hat{c}_t$, we need some assumptions on the kernel.

Assumption 2. The kernel function $K : \mathbb{R} \to \mathbb{R}$ is bounded, Lipschitz, and piecewise C^1 on $(-\infty, \infty)$ such that $\int K(x)dx = 1$, $\int |K(x)x|dx < \infty$, $K(x)x^2 \to 0$, as $|x| \to \infty$, and $\int |K'(x)|dx < \infty$.

For an arbitrary process $\{U_t\}_{t \geq 0}$ and a given time span $\Delta_n > 0$, we shall use the notation

$$U_i^n := U_{i\Delta_n}, \quad \Delta_i^n U := U_i^n - U_{i-1}^n.$$

Stable convergence in law is denoted by $\xrightarrow{st}$. See (2.2.4) in Jacod and Protter (2011) for the definition of this type of convergence. As usual, $a_n \sim b_n$ means that $a_n/b_n \to 1$ as $n \to \infty$.

Throughout the paper, we consider two settings: observations with and without market microstructure noise. In the absence of microstructure noise, we use standard kernel estimation, whereas to handle the noise, we propose a type of pre-averaging kernel estimator. These two settings together with the main results are presented in the following two subsections.

2.1. Observations Without Microstructure Noise

In this subsection, we assume that we can directly observe the process X in (2.1) at discrete times $t_i := t_{i,n} := i\Delta_n$, where $\Delta_n := T/n$ and $T \in (0, \infty)$ is a given fixed time horizon. We also consider a sequence of truncation levels v_n satisfying

$$v_n = \alpha \Delta_n^{\varpi} \quad \text{for some } \alpha > 0, \quad \varpi \in \left(0, \frac{1}{2}\right). \tag{2.3}$$

To estimate the spot volatility c_τ, at a given time $\tau \in (0, T)$, we adopt the kernel estimator, studied in Fan and Wang (2008) and Kristensen (2010) and its truncated version, studied in Yu et al. (2014a) and Mancini et al. (2015):

$$\hat{c}^n (m_n)_\tau := \sum_{i=1}^n K_{m_n \Delta_n} (t_{i-1} - \tau) \left(\Delta_i^n X\right)^2, \tag{2.4}$$

$$\hat{c}^n (m_n, v_n)_\tau := \sum_{i=1}^n K_{m_n \Delta_n} (t_{i-1} - \tau) \left(\Delta_i^n X\right)^2 \mathbb{1}_{\{|\Delta_i^n X| \leq v_n\}}, \tag{2.5}$$

where $K_b(x) := K(x/b)/b$, $m_n \in \mathbb{N}$, and $b_n := m_n \Delta_n$ is the bandwidth of the kernel function.[4] The asymptotic behavior of this estimator with one-sided uniform kernels (i.e., $K(x) = \mathbf{1}_{[0,1]}(x)$ or $K(x) = \mathbf{1}_{[-1,0]}(x)$) was studied in Jacod and Protter (2011). Yu et al. (2014a) showed a CLT for (2.5) at the suboptimal convergence rate ($\beta = 0$) under a nonleverage condition (i.e., $d\langle W, B\rangle_t = 0$) in (2.1) and (2.2) and a compound Poisson jump component. In this part, we extend these results for general two-sided kernels with possibly unbounded support, optimal convergence rate of the estimation error, and a more general type of Itô semimartingales. There is an important motivation for considering general unbounded kernels since, as proved in Figueroa-López and Li (2020a) in the no-leverage case and without jumps, exponential and some other nonuniform unbounded kernels can yield estimators with significantly better performance than those based on uniform kernels. In Section 3, we show that this is also true under the more general semimartingale models (2.1) and (2.2).

We now proceed to describe the limiting distribution of the estimation error of (2.4) and (2.5). Let V, V' be independent centered Gaussian variables, independent of $\mathcal{F}^{(0)}$, defined on a "very good" filtered extension $\left(\widetilde{\Omega}^{(0)}, \widetilde{\mathcal{F}}^{(0)}, \left(\widetilde{\mathcal{F}}_t^{(0)}\right)_{t \geq 0}, \widetilde{\mathbb{P}}^{(0)}\right)$ of $\left(\Omega^{(0)}, \mathcal{F}^{(0)}, \left(\mathcal{F}_t^{(0)}\right)_{t \geq 0}, \mathbb{P}^{(0)}\right)$ (see Jacod and Protter (2011) for definition) such that

$$\mathbb{E}\left(V^2\right) = 2\int K^2(u)du, \quad \mathbb{E}\left(V'^2\right) = \int L^2(t)dt, \tag{2.6}$$

where $L(t) = \int_t^\infty K(u)du\mathbf{1}_{\{t>0\}} - \int_{-\infty}^t K(u)du\mathbf{1}_{\{t \leq 0\}}$. Next, let $Z_\tau^{(0)}, Z_\tau'^{(0)}$ be defined as

$$Z_\tau^{(0)} = c_\tau V, \quad Z_\tau'^{(0)} = \tilde{\sigma}_\tau V'. \tag{2.7}$$

Now, we are ready to introduce our main theorem for a general kernel estimator in the absence of microstructure noise. The proof is given in Appendix A.

THEOREM 2.1. *Let the sequence $\{m_n\}_{n \geq 1}$ that controls the bandwidth of the kernel estimator be such that $m_n \to \infty$, $m_n \Delta_n \to 0$, and*

$$m_n\sqrt{\Delta_n} \to \beta, \quad with \quad \beta \in [0, \infty]. \tag{2.8}$$

Then, under Assumptions 1 and 2, at a given time $\tau \in [0, T]$, we have:

(a) *If X is continuous, both the truncated version (2.5) and the nontruncated version (2.4) satisfy the following stable convergence in law, as $n \to \infty$:*

$$\begin{align}
&\text{(i)} \quad \sqrt{m_n}\left(\hat{c}_\tau^n - c_\tau\right) \xrightarrow{st} Z_\tau^{(0)} + \beta Z_\tau'^{(0)}, \quad if \quad \beta < \infty, \\
&\text{(ii)} \quad \frac{1}{\sqrt{m_n\Delta_n}}\left(\hat{c}_\tau^n - c_\tau\right) \xrightarrow{st} Z_\tau'^{(0)}, \quad if \quad \beta = \infty,
\end{align} \tag{2.9}$$

where $Z_\tau^{(0)}, Z_\tau'^{(0)}$ are defined as in (2.7).

---

[4]Here, m_n is equivalent to k_n in Theorem 13.3.7 of Jacod and Protter (2011), whereas $m_n\Delta_n$ is equivalent to the bandwidth h_n of Figueroa-López and Li (2020a).

(b) *Suppose*

$$m_n \Delta_n^a \to \beta' \in (0, \infty), \quad where \ a \in (0, 1), \tag{2.10}$$

so that (2.8) holds with $\beta = 0$ when $a < 1/2$, $\beta = \beta' \in (0, \infty)$ when $a = 1/2$, or $\beta = \infty$ when $a > 1/2$. Then, when X is discontinuous, we have (2.9) for the nontruncated version (2.4), as soon as

$$either \ r < \frac{4}{3}, \quad or \quad \frac{4}{3} \leq r < \frac{2}{1+a} \quad \left(and \ then \ a < \frac{1}{2} \right).$$

(c) *Under (2.10), when X is discontinuous, we have (2.9) for the truncated version, as soon as*

$$r < \frac{2}{1 + a \wedge (1-a)}, \quad \varpi > \frac{a \wedge (1-a)}{2(2-r)}. \tag{2.11}$$

Remark 2.1. The CLTs above generalize the results in Figueroa-López and Li (2020a), where only right-sided kernels were considered under the absence of jumps, in Jacod and Protter (2011) and Alvarez et al. (2012), where only one-sided uniform kernels (i.e., $K(x) = \mathbf{1}_{[0,1]}(x)$ or $K(x) = \mathbf{1}_{[-1,0]}(x)$) were studied, and in Aït-Sahalia and Jacod (2014), where a CLT for a general right-sided kernel with compact support was stated without proof. The proof of Theorem 2.1 is also different from that in Figueroa-López and Li (2020a) and is based on the approach of Jacod and Protter (2011). The case with $\beta = 0$ produces a CLT with convergence rate $m_n^{-1/2}$, which vanishes slower than $\Delta_n^{1/4}$, the optimal rate. In that case, our result generalizes Fan and Wang (2008), Kristensen (2010), Yu et al. (2014a), and Mancini et al. (2015) by allowing jumps of both finite and infinite activity and dependence between the volatility and the BM driving the log-return process X (leverage effects).

Remark 2.2. As stated by the points (b) and (c) above, in the presence of jumps, both estimators (2.4) and (2.5) can attain the optimal convergence rate of $\Delta_n^{1/4}$, but only if the index of jump activity is less than $\frac{4}{3}$. In the presence of higher jump activity, the estimators can only achieve the suboptimal convergence rate of $m_n^{-1/2} \gg \Delta_n^{1/4}$ (cases $a < 1/2$ and $\beta = 0$). It is worth noting the surprising fact that, even in the presence of jumps, the untruncated kernel estimator (2.4) can still consistently estimate the spot volatility. In the case of finitely many jumps, we may explain this fact by noting that in a small local window, there could be at most a finite number of jumps, while, in the limit, there are increasingly more increments that do not contain jumps.[5] Nevertheless, in practice and for better finite-sample performance, one typically would prefer the truncated version of the estimator.

Remark 2.3. As explained in the Introduction, it is critical to expand the results to general two-sided kernels of unbounded support since these kernels exhibit superior performance. For instance, in the suboptimal rate case ($\beta = 0$), the kernel

---

[5]We thank a referee for pointing out this interesting insight.

K with support $[0, 1]$ that minimizes the asymptotic variance $2 \int K^2(u)du$ is the uniform kernel $K_{unif}(x) = \mathbf{1}_{[0,1]}(x)$ since, by Jensen's inequality, $\int_0^1 K^2(u)du \geq (\int_0^1 K(x)dx)^2 = 1 = \int_0^1 K_{unif}^2(u)du$. However, there are many other kernels that are two-sided or of unbounded support and that attain smaller variance, even in the suboptimal rate case $\beta = 0$. For instance, both $K(x) = 2^{-1}\mathbf{1}_{[-1,1]}(x)$ and $K_{exp^+}(x) = e^{-x}\mathbf{1}_{(0,\infty)}(x)$ are such that $\int K^2(u)du = 1/2$. In the optimal rate case ($\beta \in (0, \infty)$) and when picking the optimal value of β, the optimal kernel is the two-sided exponential $K(x) = 2^{-1}e^{-|x|}$ as shown in Section 3.4.

2.2. Observations in the Presence of Microstructure Noise

In this part, we assume that our observations of X are contaminated by "microstructure" noise. That is, we assume that we observe

$$Y_{t_i} := X_{t_i} + \epsilon_{t_i}, \tag{2.12}$$

where $\epsilon = \{\epsilon_t\}$ is the noise process and, as before, $t_i := t_{i,n} := i\Delta_n$, $0 \leq i \leq n$, with $\Delta_n := T/n$ and a fixed time horizon $T \in (0, \infty)$. We allow the noise ϵ to depend on X, but in such a way that, conditionally on the whole process X, $\{\epsilon_t\}_{t \geq 0}$ is a family of independent, centered random variables. More formally, following the framework of Jacod and Protter (2011), for each time t, we consider a transition probability $Q_t\left(\omega^{(0)}, dz\right)$ from $\left(\Omega^{(0)}, \mathcal{F}_t^{(0)}\right)$ into $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$, and the canonical process $\{\epsilon_t\}_{t \geq 0}$ on $\mathbb{R}^{[0,\infty)}$ defined as $\epsilon_t(\tilde{\omega}) = \tilde{\omega}(t)$ for $t \geq 0$ and $\tilde{\omega} \in \mathbb{R}^{[0,\infty)}$. Next, we construct a new probability space $\left(\mathbb{R}^{[0,\infty)}, \mathbb{B}, \sigma\left(\epsilon_s : s \in [0,t)\right), \mathbb{Q}\right)$, where $\mathbb{B}$ is the product Borel σ-field and $\mathbb{Q} = \otimes_{t \geq 0}Q_t$. We then define an enlarged filtered probability space $\left(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t \geq 0}, \mathbb{P}\right)$ and a filtration $(\mathcal{H}_t)$ as follows:

$$\begin{cases} \Omega = \Omega^{(0)} \times \mathbb{R}^{[0,\infty)}, \\ \mathcal{F}_t = \mathcal{F}_t^{(0)} \otimes \sigma\left(\epsilon_s : s \in [0,t)\right), \quad \mathcal{H}_t = \mathcal{F}^{(0)} \otimes \sigma\left(\epsilon_s : s \in [0,t)\right), \\ \mathbb{P}(d\omega^{(0)}, d\tilde{\omega}) = \mathbb{P}^{(0)}(d\omega^{(0)})\mathbb{Q}(\omega^{(0)}, d\tilde{\omega}). \end{cases}$$

Any variable or process in either $\Omega^{(0)}$ or $\mathbb{R}^{[0,\infty)}$ can be extended in the usual way to a variable or a process on Ω. We now state the assumptions on the $\mathcal{F}^{(0)}$-conditional law of the noise process.

Assumption 3. All variables ($\epsilon_t : t \geq 0$) are independent conditionally on $\mathcal{F}^{(0)}$, and we have:

- $\mathbb{E}\left(\epsilon_t | \mathcal{F}^{(0)}\right) = 0$;
- for all $p > 0$, the process $\mathbb{E}\left(|\epsilon_t|^p | \mathcal{F}^{(0)}\right)$ is $\left(\mathcal{F}_t^{(0)}\right)$-adapted and locally bounded;
- the conditional variance process $\gamma_t = \mathbb{E}\left(|\epsilon_t|^2 \big| \mathcal{F}^{(0)}\right)$ is càdlàg.

Along the lines of Jacod and Protter (2011) (originally proposed in Jacod et al. (2009)), to construct the pre-averaging estimator, we need:

(i) a sequence of positive integers k_n, which represent the length of the pre-averaging window, satisfying

$$k_n = \frac{1}{\theta \sqrt{\Delta_n}} + o\left(\frac{1}{\Delta_n^{1/2}}\right), \quad \text{for some } \theta > 0; \tag{2.13}$$

(ii) a real-valued weight function g on $[0, 1]$, satisfying that g is continuous, piecewise C^1 with a piecewise Lipschitz derivative g' such that $g(0) = g(1) = 0$ and $\int_0^1 g(s)^2 ds = 1$;

(iii) a sequence v_n representing the truncation level, satisfying

$$v_n = \alpha \, (k_n \Delta_n)^\varpi \quad \text{for some } \alpha > 0, \, \varpi \in (0, \tfrac{1}{2}). \tag{2.14}$$

Next, for an arbitrary process U, we define the sequences:

$$\overline{U}_i^n = \sum_{j=1}^{k_n - 1} g\left(\frac{j}{k_n}\right) \Delta_{i+j-1}^n U = -\sum_{j=1}^{k_n} \left(g\left(\frac{j}{k_n}\right) - g\left(\frac{j-1}{k_n}\right)\right) U_{i+j-2}^n,$$

$$\widehat{U}_i^n = \sum_{j=1}^{k_n} \left(g\left(\frac{j}{k_n}\right) - g\left(\frac{j-1}{k_n}\right)\right)^2 \left(\Delta_{i+j-1}^n U\right)^2. \tag{2.15}$$

As seen from the definition, $\overline{U}_i^n$ is the weighted average of the increments $\Delta_{i+j-1} U, j = 1, \ldots, k_n - 1$, whereas $\widehat{U}_i^n$ is a debiasing term. For a weight function g as above, let

$$\phi_{k_n}(g) = \sum_{i=1}^{k_n} g(\frac{i}{k_n})^2, \quad \phi'_{k_n}(h) = \sum_{i=1}^{k_n} \left(g(\frac{i}{k_n}) - g(\frac{i-1}{k_n})\right)^2, \tag{2.16}$$

and note that

$$\phi_{k_n}(g) = k_n \int_0^1 g^2(s) ds + O(1) = k_n + O(1),$$

$$\phi'_{k_n}(g) = \frac{1}{k_n} \int_0^1 (g'(s))^2 ds + O\left(\frac{1}{k_n^2}\right). \tag{2.17}$$

Now, we can define the pre-averaging estimators of the spot variance c_τ at $\tau \in (0, T)$. We consider a nontruncated version, defined as

$$\hat{c}(k_n, m_n)_\tau = \frac{1}{\phi_{k_n}(g)} \sum_{j=1}^{n-k_n+1} K_{m_n \Delta_n}(t_{j-1} - \tau) \left(\left(\overline{Y}_j^n\right)^2 - \frac{1}{2} \widehat{Y}_j^n\right), \tag{2.18}$$

as well as, two truncated versions:

$$\hat{c}(k_n, m_n, v_n, 1)_\tau = \frac{1}{\phi_{k_n}(g)} \sum_{j=1}^{n-k_n+1} K_{m_n \Delta_n}(t_{j-1} - \tau) \left(\left(\overline{Y}_j^n\right)^2 \mathbb{1}_{\{|\bar{Y}_j^n| \leq v_n\}} - \frac{1}{2} \widehat{Y}_j^n\right),$$

$$\hat{c}(k_n, m_n, v_n, 2)_\tau = \frac{1}{\phi_{k_n}(g)} \sum_{j=1}^{n-k_n+1} K_{m_n \Delta_n}(t_{j-1} - \tau) \left(\left(\overline{Y}_j^n\right)^2 - \frac{1}{2} \widehat{Y}_j^n\right) \mathbb{1}_{\{|\bar{Y}_j^n| \leq v_n\}}. \tag{2.19}$$

The basic idea is the same as in the case where the efficient process X is observed without noise. We see $\overline{Y}_j^n$ as a noise-free proxy of the increment $\Delta_j^n X$. By properly choosing the truncation level $v_n \to 0$ (e.g., $v_n \gg \sqrt{u_n ln(1/u_n)}$ with $u_n := k_n \Delta_n$), the event $|\bar{Y}_j^n| > v_n$ will suggest the occurrence of a "big" jump happening during the time interval $[j\Delta_n, (j+k_n)\Delta_n]$ and, thus, we eliminate such a term from the summations in (2.19). The estimator $\hat{c}(k_n, m_n, v_n, 2)_\tau$ is closer to the one defined in Yu et al. (2014b), whereas $\hat{c}(k_n, m_n, v_n, 1)_\tau$ is similar to the one considered in Chen (2019), although therein only the one-sided kernel $K(x) = \mathbf{1}_{[0,1]}(x)$ is studied. It will be interesting to compare their statistical properties and finite-sample performance in our simulation study of Section 4.

Before giving the asymptotic behavior of the pre-averaging estimators (2.18) and (2.19), we introduce the limiting distributions. Below, Z_τ, Z_τ' are defined on a good extension $\left(\widetilde{\Omega}, \widetilde{\mathcal{F}}, \left(\widetilde{\mathcal{F}}_t\right)_{t>0}, \widetilde{\mathbb{P}}\right)$ of the space $\left(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geq0}, \mathbb{P}\right)$ so that, conditionally on $\mathcal{F}$, they are independent Gaussian random variables with conditional variance

$$\delta_1^2(\tau) := \widetilde{\mathbb{E}}\left(Z_\tau^2 | \mathcal{F}\right) = 4\left(\Phi_{22} c_\tau^2/\theta + 2\Phi_{12} c_\tau \gamma_\tau \theta + \Phi_{11}\gamma_\tau^2 \theta^3\right)\int K^2(u)du,$$

$$\delta_2^2(\tau) := \widetilde{\mathbb{E}}\left(Z_\tau'^2 | \mathcal{F}\right) = \tilde{\sigma}_\tau^2 \int L^2(t)dt, \tag{2.20}$$

with $\phi_1(s) = \int_s^1 g'(u)g'(u-s)\mathrm{d}u$, $\phi_2(s) = \int_s^1 g(u)g(u-s)\mathrm{d}u$, $\Phi_{ij} = \int_0^1 \phi_i(s)\phi_j(s)\mathrm{d}s$, and $L(t) = \int_t^\infty K(u)du\mathbf{1}_{\{t>0\}} - \int_{-\infty}^t K(u)du\mathbf{1}_{\{t\leq0\}}$. The following result establishes the asymptotic behavior of the estimation error for the proposed estimators. The proof is given in Appendix B.

THEOREM 2.2. *Let $\{m_n\}_{n\geq1}$ be a sequence of positive integers such that $m_n \to \infty$, $m_n\Delta_n \to 0$, $m_n\sqrt{\Delta_n} \to \infty$, and $m_n\Delta_n^{3/4} \to \beta$, for some $\beta \in [0,\infty]$, and let k_n, v_n, and g be as described in (i)–(iii) above. Then, under Assumptions 1–3, we have:*

1. *When X is continuous, the pre-averaging estimators (2.18) and (2.19) are all such that, as $n \to \infty$,*

   (i) $m_n^{1/2}\Delta_n^{1/4}\left(\hat{c}_\tau - c_\tau\right) \xrightarrow{st} Z_\tau + \beta Z_\tau'$, *if $\beta \in [0,\infty)$,*

   (ii) $\dfrac{1}{\sqrt{m_n\Delta_n}}\left(\hat{c}_\tau - c_\tau\right) \xrightarrow{st} Z_\tau'$, *if $\beta = \infty$.* $\qquad\qquad$ (2.21)

2. *When X is discontinuous and $r \in (0,2]$, with*

$$m_n\Delta_n^a \to \beta' \in (0,\infty), \quad \text{where } a \in \left(\frac{1}{2}, 1\right), \tag{2.22}$$

   and r, ϖ satisfying

$$r < \frac{5}{2} - 2\left[\left(a - \frac{1}{4}\right) \wedge \left(1 - a + \frac{1}{4}\right)\right], \quad \varpi \geq \frac{\left(a - \frac{1}{4}\right) \wedge \left(1 - \left(a - \frac{1}{4}\right)\right) - \frac{1}{4}}{2 - r}, \tag{2.23}$$

the truncated pre-averaging estimator $\hat{c}(k_n, m_n, v_n, 1)_\tau$ in (2.19) satisfies (2.21) with $\beta = 0$ when $a < 3/4$, with $\beta = \beta'$ when $a = 3/4$, or with $\beta = \infty$ when $3/4 < a < 1$.

3. When X is discontinuous and $r \in (0, 2]$, with (2.22) and r, ϖ satisfying

$$r \leq 4 - \frac{2}{a \vee (3/2 - a)}, \quad \frac{(a - \frac{1}{4}) \wedge (1 - (a - \frac{1}{4})) - \frac{1}{4}}{2 - r} \leq \varpi \leq \frac{(2 - 2a) \vee (2a - 1)}{r}, \tag{2.24}$$

the truncated pre-averaging estimator $\hat{c}(k_n, m_n, v_n, 2)_\tau$ in (2.19) satisfies (2.21) with $\beta = 0$ when $a < 3/4$, with $\beta = \beta' \in (0, \infty)$ when $a = 3/4$, and with $\beta = \infty$ when $3/4 < a < 1$.

Remark 2.4. The second and third points in the above theorem show that the truncated pre-averaging estimators (2.19) can achieve the optimal convergence rate of $\Delta_n^{1/8}$, but only if the index of jump activity is restricted to be $r < \frac{3}{2}$ for $\hat{c}(k_n, m_n, v_n, 1)$ and $r < \frac{4}{3}$ for $\hat{c}(k_n, m_n, v_n, 2)$. If the index of jump activity is larger than $\frac{3}{2}$ and $\frac{4}{3}$, respectively, the estimators can only achieve suboptimal convergence rates. When comparing their theoretical properties, $\hat{c}(k_n, m_n, v_n, 1)$ can, in principle, handle jumps with higher index r than $\hat{c}(k_n, m_n, v_n, 2)$ at the optimal bandwidth. However, as we will see in Section 4, $\hat{c}(k_n, m_n, v_n, 2)$ appears to be more effective at eliminating jumps when the jump size is large in the presence of FA jumps. The two estimators have similar performance when the jump's size is relatively small.

Remark 2.5. Let us give some intuition or heuristic explanation of the estimator (2.19) and its asymptotic behavior established above. For the estimation of the integrated variance (IV), $[X, X]_T = \int_0^T c_t dt$, Jacod et al. (2009) proposed the following pre-averaging estimator:

$$\widehat{[X, X]}_s := \frac{1}{\phi_{k_n}(g)} \frac{s}{s - k_n \Delta_n} \sum_{j=1}^{[s/\Delta_n] - k_n + 1} \left(\left(\overline{Y}_j^n \right)^2 - \frac{1}{2} \widehat{Y}_j^n \right), \quad s \in (0, T],$$

for a continuous Itô semimartingale X. It was shown that

$$\frac{1}{\Delta_n^{1/4}} \left(\widehat{[X, X]}_T - [X, X]_T \right) \xrightarrow{st} \mathcal{U}_T^{\text{noise}},$$

where $\mathcal{U}_T^{\text{noise}}$ is a centered Gaussian process with conditional variance

$$\delta_T := \mathbb{E} \left(\left(\mathcal{U}_T^{\text{noise}} \right)^2 | \mathcal{F} \right) = \int_0^T \zeta_t dt := \int_0^T 4 \left(\Phi_{22} c_t^2 / \theta + 2 \Phi_{12} c_t \gamma_t \theta + \Phi_{11} \gamma_t^2 \theta^3 \right) dt.$$

In the no-thresholding case ($v_n = \infty$), the spot volatility estimator (2.19) can be viewed as a localization of the IV process in that

$$\hat{c}_t \approx \int K_{m_n \Delta_n}(s - t) d\widehat{[X, X]}_s.$$

More specifically, the factor $\frac{s}{s-k_n\Delta_n}$ is omitted for the spot volatility estimator. If we use the representation $\mathcal{U}_t^{\text{noise}} = \int_0^t (\zeta_s)^{1/2} dB_s^U$, where B^U is a Wiener process, we can then heuristically argue that $\hat{c}_t - c_t \approx \int K_{m_n\Delta_n}(s-t) d(\widehat{[X,X]}_s - [X,X]_s) = \Delta_n^{1/4} \int K_{m_n\Delta_n}(s-t) dU_s^{\text{noise}} = \Delta_n^{1/4} \int K_{m_n\Delta_n}(s-t)(\zeta_s)^{1/2} dB_s^U$. Therefore, the variance of the estimation error at time t is expected to be close to

$$\sqrt{\Delta_n} \int K_{m_n\Delta_n}^2(s-t)\zeta_s ds \approx \frac{1}{m_n\sqrt{\Delta_n}} 4\left(\Phi_{22}c_t^2/\theta + 2\Phi_{12}c_t\gamma_t\theta + \Phi_{11}\gamma_t^2\theta^3\right) \int K^2(u)du,$$

which is indeed the case, but only when $m_n\Delta_n^{3/4} \to \beta = 0$ as formally shown in Theorem 2.2. It is important to remark that the proof of Theorem 2.2 does not rely on the heuristic arguments above.

3. AN APPLICATION: OPTIMAL PARAMETER TUNING

In this section, as an application of our main Theorems 2.1 and 2.2, we show how to tune the bandwidth parameter β and the pre-averaging parameter θ, as well as the kernel function K of the estimator, in order to minimize the asymptotic variance of the estimation error $\hat{c}_\tau - c_\tau$. Two possible approaches can be taken. Minimize the asymptotic variance of $\hat{c}_\tau$, say $\tilde{\delta}^2(\tau)$, at each time τ or minimize the *integrated asymptotic variance* $\int_0^T \tilde{\delta}^2(t)dt$ over the period $[0,T]$. In our simulations of Section 4, we implemented both methods and found out that the second method yields slightly better results. An explanation for this is given in Remark 4.1. Therefore, in this part, we focus on the second approach.

By necessity, the optimal choices of θ and β under the criterion of the previous paragraph will be expressed in terms of the integrated variance and quarticity, $IV_T := \int_0^T c_t dt$ and $QrT_T := \int_0^T c_t^2 dt$, respectively, the Integrated Volatility of Volatility (IVV), $\int_0^T \tilde{\sigma}_t^2 dt$, and the integrated variance of the noise ϵ_t, $\int_0^T \gamma_t dt$. We can estimate $\int_0^T \tilde{\sigma}_t^2 dt$ and $\int_0^T \gamma_t dt$ separately, whereas for IV_T and QrT_T, we propose an iterative procedure in which an initial rough estimate of c_t on a grid of $[0,T]$ is used to determine initial estimates of IV_T and QrT_T. These estimates are then used to find suitable estimates of the optimal values for θ and β. Finally, these estimated $\hat{\theta}$ and $\hat{\beta}$ are applied in the kernel pre-averaging estimator (2.19) to refine our estimates of c_t on the grid.

Remark 3.1. A related problem, that is not being considered here in detail, is that of tuning the truncation level v_n in the truncated estimators (2.5) and (2.19). Most of the literature about this problem has been in the context of estimating the integrated variance $IV_T = \int_0^T \sigma_s^2 ds$. It has been customary in econometric studies to adopt a power threshold of the form $v_n = c\Delta_n^\gamma$. The rule of thumb is to take a value of γ close to .5 and c that depends on an estimate of the volatility level. For instance, Jacod and Todorov (2014) took $\gamma = .49$ and $c = 4\sqrt{BPV}$, where $BPV := \frac{\pi}{2} \sum_{i=2}^n |\Delta_{i-1}^n X||\Delta_i^n X|$ is the bipower variation. More recently, this issue has also been studied in the literature using more objective and statistically valid

approaches, but only in the absence of microstructure noise. In the case of FA jumps and constant volatility σ, Figueroa-López and Mancini (2019) showed that the optimal threshold (in terms of minimizing the conditional MSE) is asymptotically equivalent to $\sqrt{2\sigma^2\Delta_n \ln(1/\Delta_n)}$ and proposed an iterative method to estimate σ. In the presence of small jumps that behave like those of an α-stable Lévy process, Figueroa-López, Gong, and Han (2022) showed that the optimal threshold is asymptotically equivalent to $\sqrt{(2-\alpha)\sigma^2\Delta_n \ln(1/\Delta_n)}$. Again, these results are in the absence of microstructure noise and for the problem of estimating the integrated variance. However, given the local nature of spot volatility estimation, one can imagine that similar results may hold for the estimators (2.5) and (2.19). We leave this problem for future research.

3.1. Optimal Selection of θ

Recall that we set $k_n = \frac{1}{\theta\sqrt{\Delta_n}} + o\left(\frac{1}{\Delta_n^{1/4}}\right)$ and, thus, the parameter θ determines the length of the pre-averaging window k_n. The following corollary, which follows easily from Theorem 2.2, gives us a method to tune θ up.

COROLLARY 3.1. *The optimal value θ^* of θ, which is set to minimize the integrated asymptotic variance of the pre-averaging kernel estimator (2.19), is such that*

$$(\theta^\star)^2 = \frac{\sqrt{\Phi_{12}^2\left(\int_0^T c_t\gamma_t dt\right)^2 + 3\Phi_{11}\Phi_{22}\int_0^T \gamma_t^2 dt \int_0^T c_t^2 dt} - \Phi_{12}\int_0^T c_t\gamma_t dt}{3\Phi_{11}\int_0^T \gamma_t^2 dt}. \tag{3.1}$$

Remark 3.2. Note that the local version of (3.1) (i.e., the value of θ that minimizes the spot asymptotic variance $\delta_1^2(t)$) is such that

$$(\theta_t^{\star,local})^2 = c_t\frac{\sqrt{\Phi_{12}^2 + 3\Phi_{11}\Phi_{22}} - \Phi_{12}}{3\Phi_{11}\gamma_t}. \tag{3.2}$$

In the context of integrated volatility estimation, Jacod and Mykland (2015) obtained the same formula (see equation (3.8) therein). It was also proposed a two-step procedure to implement it. However, in our simulation, we found out that the performance of the estimator is less sensitive to the choice of θ than to that of the bandwidth.

3.2. Optimal Bandwidth Selection

From Theorem 2.2, we can deduce that when m_n (the bandwidth in Δ_n units) is of the form $m_n = \beta\Delta_n^{-3/4}$ for some constant $\beta \in (0,\infty)$, the optimal convergence rate of $\Delta_n^{1/8}$ is attained and we further have

$$\Delta_n^{-1/8}\left(\hat{c}(k_n, m_n, v_n)_\tau - c_\tau\right) \xrightarrow{st} \beta^{-1/2}\left(Z_\tau + \beta Z_\tau'\right).$$

Therefore, the limiting distribution above has conditional variance $\bar{\delta}^2(\tau) := \frac{1}{\beta}\delta_1^2(\tau) + \beta\delta_2^2(\tau)$, where $\delta_1^2(\tau)$ and $\delta_2^2(\tau)$ are given as in (2.20). The following result gives the optimal value of β that minimizes $\int_0^T \bar{\delta}^2(\tau)d\tau$.

COROLLARY 3.2. *Let*

$$\Theta(\theta) := \Theta(\theta; g) := \frac{\Phi_{22}}{\theta}\int_0^T c_t^2 dt + 2\Phi_{12}\theta\int_0^T \gamma_t c_t dt + \Phi_{11}\theta^3\int_0^T \gamma_t^2 dt.$$

With the bandwidth $b_n = m_n\Delta_n = \beta\Delta_n^{1/4}$, the optimal value of b_n, which is set to minimize $\int_0^T \bar{\delta}_\tau^2 d\tau$, is given by

$$b_n^\star = \sqrt{\frac{\int_0^T \delta_1^2(t)dt}{\int_0^T \delta_2^2(t)dt}}\Delta_n^{1/4} = \Delta_n^{1/4}\sqrt{\frac{4\Theta(\theta)\int K^2(u)du}{\int_0^T \tilde{\sigma}_t^2 dt \int L^2(v)dv}}. \tag{3.3}$$

With this optimal bandwidth choice, the integrated variance $\int_0^T \bar{\delta}^2(\tau)d\tau$ of the limiting distribution for the scaled estimation error $\Delta_n^{-1/8}\left(\hat{c}(k_n, m_n, v_n)_\tau - c_\tau\right)$ is given by

$$2\sqrt{\int_0^T \delta_1^2(t)dt \int_0^T \delta_2^2(t)dt} = 4\sqrt{\Theta(\theta)\int_0^T \tilde{\sigma}_t^2 dt \int K^2(u)du \int L^2(v)dv}. \tag{3.4}$$

Note that $b_n^\star$ contains unknown theoretical quantities that need to be estimated in order to devise a plug in type estimator. Under the assumption of $\gamma_t \equiv \gamma$, the variance of the noise, γ, can be estimated using the estimator in Zhang et al. (2005):

$$\hat{\gamma} = \frac{1}{2n}\sum_{i=1}^n \left(Y_i^n - Y_{i-1}^n\right)^2.$$

For the estimation of the IVV, $\int_0^T \tilde{\sigma}_t^2 dt$, we start by obtaining a preliminary estimate of the spot variance c on the grid $\tau \in \{t_i\}_{i=0,\ldots,n}$, via the estimator (2.19), starting with some sensible initial estimates of the tuning parameter values. For example, we can set $b_n = m_n\Delta_n = \Delta_n^{1/4}$. Let us denote these initial estimates as $\hat{c}_{t_i,0}$. We then compute the sparse realized quadratic variation of the $\hat{c}_{t_i}$'s to estimate the IVV $= \int_0^T \tilde{\sigma}_t^2 dt$:

$$\widehat{IVV}_{T,0} := \sum_{i=0}^{[n/p]-1} (\hat{c}_{t_{(i+1)p},0} - \hat{c}_{t_{ip},0})^2,$$

for some positive integer $p \ll n$. We also implemented a pre-averaging integrated variance estimator for the IVV based on the spot variance estimates. However, the choice of tuning parameters here could be tricky and the performance is similar

to the simpler sparse Realized Variance estimator above. As for $\int_0^T c_t^2 dt$, we can simply compute the sum of squares of the preliminary estimates $\hat{c}_{t_i,0}^2$ and multiply by Δ_n.[6] Now, with these estimates, we can calculate an estimate of the optimal bandwidth $b_n^\star$ using the result of Corollary 3.2. Such an approximate optimal bandwidth can then be used to refine our estimates of the spot variance grid. Continuing this procedure iteratively, we hope to obtain good estimates of the optimal bandwidth.

Note that (3.3) sets the same bandwidth for the entire path of X. We can also consider a local or nonhomogeneous bandwidth: for $\tau \in [0, T]$, the local bandwidth is set to minimize the asymptotic variance of the estimation error at time τ. Concretely, by setting $m_n = \beta \Delta_n^{-3/4}$ and minimizing the asymptotic spot variance $\bar{\delta}^2(\tau) = \beta^{-1}\delta_1^2(\tau) + \beta\delta_2^2(\tau)$, the optimal bandwidth is given by

$$b_n^{\star,local}(\tau) = \frac{\delta_1(\tau)}{\delta_2(\tau)}\Delta_n^{1/4} = \Delta_n^{1/4}\sqrt{\frac{4\Theta_\tau(\theta)\int K^2(u)du}{\tilde{\sigma}_\tau^2 \int L^2(u)du}}, \qquad (3.5)$$

with $\delta_1(\tau)$ and $\delta_2(\tau)$ defined as in Theorem 2.2 and $\Theta_\tau(\theta)$ defined as

$$\Theta_\tau(\theta) := \frac{\Phi_{22}}{\theta}c_\tau^2 + 2\Phi_{12}\gamma_\tau\theta c_\tau + \Phi_{11}\gamma_\tau^2\theta^3.$$

With this optimal bandwidth, the variance of the limiting distribution for the estimation error is given by

$$2\delta_1(\tau)\delta_2(\tau) = 4\sqrt{\Theta_\tau(\theta)\tilde{\sigma}_\tau^2 \int K^2(u)du \int L^2(u)du}. \qquad (3.6)$$

Since the local bandwidth has the flexibility to adapt to the volatility level, we may expect that a data-driven estimate of the bandwidth $b_n^{\star,local}(\tau)$ in (3.5) should outperform a data-driven estimate of the homogeneous bandwidth $b_n^\star$ in (3.3). However, in our Monte Carlo simulations of Section 4, we found out that this is not always the case. A possible explanation for this is given below (see also Remark 4.1 for further analysis).

Remark 3.3. We can see the constant bandwidth (3.3) as an approximation of the optimal local bandwidth (3.5), where the mean values $\int_0^T \Theta_t(\theta)dt/T$ and $\int_0^T \tilde{\sigma}_t^2 dt/T$ are used as proxies of the spot values $\Theta_\tau(\theta)$ and $\tilde{\sigma}_\tau^2$, respectively. These global proxies have the advantages of being easier and more accurate to estimate. This may be one of the reasons why a data-driven estimate of the constant

---

[6]In the simulations, we also tried the preaveraged quarticity estimator of Jacod et al. (2009) (equation (3.14) therein), but the results were suboptimal.

bandwidth $b_n^\star$ may be able to outperform a data-driven estimate of the local version $b_n^{\star,local}(\tau)$ in some situations.

3.3. Optimal Kernel Function

With the optimal bandwidths of Section 3.2, we can now obtain a formula for the asymptotic variance, which enjoys an explicit dependence on the kernel function K. It is then natural to attempt to find the kernel that minimizes such a variance. As observed from (3.4) or (3.6), we only need to minimize

$$I(K) = \int K^2(u)du \int L^2(u)du = \int K^2(u)du \iint_{xy \geq 0} K(x)K(y)(|x| \wedge |y|)dxdy,$$

over all kernels K such that $\int K(u)du = 1$, where for the second equality above we used that $L(t) = \int_t^\infty K(u)du\mathbf{1}_{\{t>0\}} - \int_{-\infty}^t K(u)du\mathbf{1}_{\{t\leq 0\}}$. It has been proved in Figueroa-López and Li (2020a, Sect. 4.1) that, among all the kernel functions satisfying Assumption 2, the exponential kernel function $K^{\exp}(x) = \frac{1}{2}\exp(-|x|)$ is the one that minimizes the functional $I(K)$. Figueroa-López and Li (2020a) (see Remark 4.2 therein) showed that, compared with the two-sided uniform (resp. Epanechnikov) kernels, the integrated asymptotic variance can be reduced by approximately 14% (resp. 6%) when using exponential kernels. Figueroa-López and Li (2020a) also showed that exponential kernels have a computational advantage since they enable us to reduce the time complexity for estimating the volatility on all the grid points $t_1 < \cdots < t_n$, from $O(n^2)$ to $O(n)$. This property is particularly useful when working with high-frequency observations, where n is quite large.

3.4. Tuning Parameters Under the Absence of Microstructure Noise

By following the same arguments as above, we can determine the optimal bandwidth parameter and kernel function for the estimators (2.4) and (2.5) under the no-microstructure-noise models (2.1) and (2.2). Specifically, we first take a bandwidth of the form $b_n = \beta \Delta_n^{1/2}$ ($\beta \in (0, \infty)$), which, from Theorem 2.1, leads to the best possible rate of convergence $\Delta_n^{-1/4}$ of (2.4) and (2.5). In that case, the asymptotic variance will take the form $\bar{\delta}_\tau^2 = \beta^{-1}\delta_1^2(\tau) + \beta\delta_2^2(\tau)$, where

$$\delta_1^2(\tau) = 2c_\tau^2 \int K^2(u)du, \qquad \delta_2^2(\tau) = 2\tilde{\sigma}_\tau^2 \int L^2(t)dt.$$

Then, the optimal value of β that minimizes the asymptotic variance is $\beta^* = \delta_1(\tau)/\delta_2(\tau)$, leading to the optimal bandwidth

$$\tilde{b}_n^{\star,local} = \frac{\delta_1(\tau)}{\delta_2(\tau)}\Delta_n^{1/2} = \Delta_n^{1/2}\sqrt{\frac{2c_\tau^2 \int K^2(u)du}{\tilde{\sigma}_\tau^2 \int L^2(u)du}}. \tag{3.7}$$

Plugging β^* into $\bar{\delta}_\tau^2$ leads to the optimal asymptotic variance of

$$2\delta_1(\tau)\delta_2(\tau) = 4\sqrt{c_\tau^2\tilde{\sigma}_\tau^2 \int K^2(u)du \int L^2(u)du},$$

which, as before, is minimized by the two-sided exponential kernel $K(x) = 2^{-1}e^{-|x|}$. See the Introduction for other related results in the literature.

4. SIMULATION STUDY

In this section, we study the performance of the kernel pre-averaging estimators (2.18) and (2.19), together with the implementation procedure described in Section 3.2, and compare the results with the TSRSV estimator proposed in Zu and Boswijk (2014).

4.1. Simulation Design and Performance Metrics

We implemented two different data-generating models: a Heston model and a one-factor stochastic volatility (SV1F) model. More specifically, in Sections 4.2–4.5, we consider the Heston model:

$$\begin{aligned}
Y_{t_i} &= X_{t_i} + \varepsilon_{t_i}, \\
dX_t &= (\mu - c_t/2)\,dt + c_t^{1/2}dW_t + J_t^X dN_t^X, \\
dc_t &= \kappa\,(\alpha - c_t)\,dt + \gamma c_t^{1/2}dB_t + \sqrt{c_{t-}}J_t^c dN_t^c,
\end{aligned} \tag{4.1}$$

where we assume that $B_t = \rho W_t + \sqrt{1-\rho^2}\tilde{W}_t$, with $\tilde{W}$ being a BM independent with W. We adopt the same parameter values as in Zhang et al. (2005), but properly normalized so that the time unit is 1 day:

$$\mu = 0.05/252, \quad \kappa = 5/252, \quad \alpha = 0.04/252, \quad \gamma = 0.5/252, \quad \rho = -0.5. \tag{4.2}$$

We set the noise as $\epsilon_i^n := \epsilon_{t_i} \overset{i.i.d.}{\sim} \mathcal{N}\left(0, 0.0005^2\right)$, and the initial values to $X_0 = 1$ and $c_0 = 0.04/252$. The jump parameters are taken from Chen (2019) and set to be $J_t^X \overset{iid}{\sim} N(-0.01, 0.02^2)$, $N_{t+\Delta}^X - N_t^X \sim \text{Poisson}\,(36\Delta/252)$, $\log\left(J_t^c\right) \overset{iid}{\sim} N(-5, 0.8)$, and $N_{t+\Delta}^c - N_t^c \sim \frac{1}{\sqrt{252}}\text{Poisson}(12\Delta/252)$, with all these random processes being mutually independent.

We also consider the SV1F model (cf. Barndorff-Nielsen et al., 2008; Zu and Boswijk, 2014; Yu et al., 2014b):

$$\begin{aligned}
Y_i^n &= X_i^n + \epsilon_i^n \\
dX_t &= \mu\,dt + \exp\left(\beta_0 + \beta_1\gamma_t\right)\,dW_t + dJ_t, \\
d\gamma_t &= \alpha\gamma_t\,dt + dB_t.
\end{aligned} \tag{4.3}$$

The model above is adopted in Sections 4.6 and 4.7 with different parameter values that will be specified therein.

Throughout, we use the usual triangular weight function $g(x) = 2x \wedge (1-x)$. We simulate data for 1 day ($T = 1$), and assume that the data are observed once every second, with 6.5 trading hours per day. The number of observations is then $n = 23,400$. For the jth simulated path $\{X_{t_i}^{(j)} : 0 \leq i \leq n, t_i = iT/n\}$, we estimate the corresponding skeleton of the spot variance process, $\{c_{t_i,j}\}_{i=1,\ldots,n}$, for a given pre-averaging parameter θ and a bandwidth parameter $\tilde{\beta}$ (the bandwidth is then given by $\tilde{\beta}\Delta_n^{1/4}$). The estimated path is denoted as $\{\hat{c}_{t_i,j}\}_{i=1,\ldots,n}$. Next, we calculate the average of the squared errors (ASE),

$$ASE_j = \frac{1}{n - 2l + 1} \sum_{i=l}^{n-l} \left(\hat{c}_{t_i,j} - c_{t_i,j}\right)^2.$$

Here, $l = [0.1n]$ is used to further alleviate boundary effects. Then, we take the square root of the average of the *ASEs* over all the simulated paths:

$$\widehat{RMSE} = \sqrt{\frac{1}{m} \sum_{j=1}^{m} ASE_j},$$

where m is the number of simulations. This is an estimate of

$$RMSE = \sqrt{\mathbb{E}\left[\frac{1}{n - 2l + 1} \sum_{i=l}^{n-l} \left(\hat{c}_{t_i} - c_{t_i}\right)^2\right]}.$$

4.2. Elimination of Jumps and Truncation

In this subsection, we will show that the truncation in the estimator (2.19) does a good job in eliminating the jumps of the process (4.1). To this end, we compare the performance of the truncated estimator $\hat{c}(k_n, m_n, v_n, 1)$ in (2.19), with that of the nontruncated estimator

$$\hat{c}(k_n, m_n, Y^*)_\tau := \frac{1}{\phi_{k_n}(g)} \sum_{j=1}^{n-k_n+1} K_{m_n \Delta_n}(t_{j-1} - \tau) \left(\left(\overline{Y}_j^{*n}\right)^2 - \frac{1}{2}\widehat{Y}_j^{*n}\right), \tag{4.4}$$

applied to the continuous Heston model:

$$\begin{aligned}
Y_i^{*n} &= X_i^{*n} + \varepsilon_i^n, \\
dX_t^* &= \left(\mu - c_t^*/2\right) dt + \sqrt{c_t^*}\, dW_t, \\
dc_t^* &= \kappa\left(\alpha - c_t^*\right) dt + \gamma\sqrt{c_t^*}\, dB_t.
\end{aligned} \tag{4.5}$$

TABLE 1. Comparison between truncated and nontruncated estimators

	$\hat{c}(k_n, m_n, Y^*)$	$\hat{c}(k_n, m_n, Y)$	$\hat{c}(k_n, m_n, v_n)$
$\widehat{RMSE} \times 10^5$	5.483386	16.83420	5.419338

We set $\beta = 1$, $\theta = 5$, and $v_n = 1.8 \times \sqrt{BPV}(k_n \Delta_n)^{0.47}$, where $BPV = \frac{\pi}{2} \sum_{i=2}^{n} |\Delta_{i-1}^n X \| \Delta_i^n X|$.[7] The results are based on 2,000 simulated paths of both Y and Y^*. As proposed in Kristensen (2010), in order to alleviate the edge effects, we replace $K_{m_n \Delta_n}(t_{i-1} - \tau)$ in (2.19) and (4.4) with

$$K_{m_n \Delta_n}^{adj}(t_{i-1} - \tau) = \frac{K_{m_n \Delta_n}(t_{i-1} - \tau)}{\Delta_n \sum_{j=1}^{n-k_n+1} K_{m_n \Delta_n}(t_{j-1} - \tau)}.$$

The $\widehat{RMSE}$'s of the three estimators are reported in Table 1.

These results suggest that the truncation procedure can effectively eliminate the jumps under this Heston model, since the estimated RMSE of the truncated estimator for the model (4.1) is even less than that of the nontruncated estimator based on the continuous model (4.5).

4.3. Validity of the Asymptotic Theory and Necessity of Debiasing

We first show that the asymptotic behavior of the estimation error is consistent with our theoretical result. By Corollary 3.2, the optimal rate of convergence of the estimation error is attained when the bandwidth takes the form $m_n^\star \Delta_n = \beta \Delta_n^{1/4}$, for some $\beta \in (0, \infty)$, and, thus, we only analyze the case 1(i) ($\beta \in (0, \infty)$) of Theorem 2.2. We aim to estimate the spot variance $c_{0.5}$ in the Heston model (4.1) without jumps. Accordingly and for simplicity, we use the untruncated pre-averaging kernel estimator (2.18). We take $\beta = 1$ and exponential kernel. The histogram of the estimation errors, $\hat{c}_{0.5} - c_{0.5}$, based on 25,000 simulated paths, is shown in Figure 1. We also plot the theoretical density of the estimation error as prescribed by Theorem 2.2 but with the true parameter values for γ and θ, and replacing $c_{0.5}$ with the average value of $c_{0.5}$ over all 25,000 paths. As can be seen, the theoretical density is consistent with the empirical results.

To investigate the need of the bias correction term $\widehat{Y}_j^n$ in $\hat{c}(k_n, m_n, v_n, 1)_\tau$, let us consider a new estimator without the bias correction term, $\tilde{c}_\tau = \sum_{j=1}^{n-k_n+1} K_{m_n \Delta_n}(t_{j-1} - \tau) (\overline{Y}_j^n)^2 \mathbb{1}_{\{|\overline{Y}_j^n| \leq v_n\}}$. We show the histogram of the estimation errors $\tilde{c}_{0.5} - c_{0.5}$ for 25,000 simulated paths, and, for comparison, also plot the same theoretical asymptotic density function of Figure 1. As shown in the left panel of Figure 2, the estimator $\tilde{c}_{0.5}$ significantly overestimates the spot variance, which shows the necessity of the bias correction term $\widehat{Y}_j^n$ in (2.19).

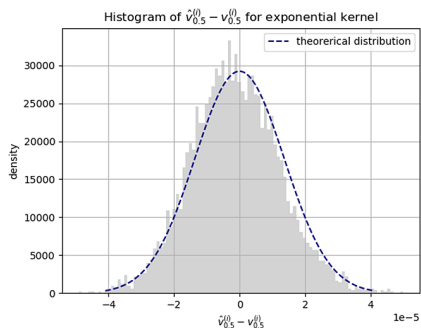[7] A similar threshold is applied in Jacod and Todorov (2014).

FIGURE 1. Histogram of $\hat{c}_t - c_t$ at $t = 0.5$ and the density of the theoretical limiting distribution.
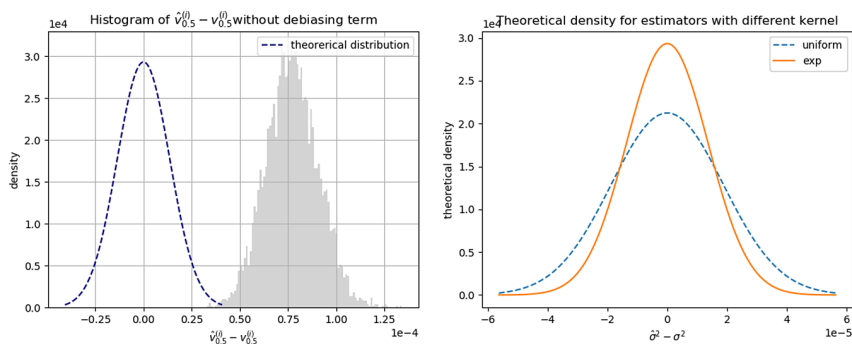

FIGURE 2. Left panel: The effect of bias correction term. Right panel: The comparison of the asymptotic distribution between uniform and exponential kernels.

4.4. Performance for Different Kernels

Before analyzing the empirical performance of the estimators for different kernels, we compare the theoretical asymptotic densities of the estimation error for the exponential and uniform kernels. This is shown in the right panel of Figure 2. We can see therein that, as predicted in Section 3.3, the exponential kernel estimator has smaller asymptotic variance.

We now proceed to compare the finite-sample performance of the untruncated pre-averaging kernel estimator (2.18) for different kernels in the Heston model (4.1) without jumps. We assume both a no-leverage setting ($\rho = 0$) and a negative correlation setting ($\rho = -0.5$). We fix $\theta = 5$ and apply the iterative homogeneous bandwidth selection method introduced in Section 3.2 with different kernels. We report the estimated *RMSE* with the initial bandwidth $\beta = 1$ and the result of iterative bandwidth selection method after one iteration in Table 2 for the following

TABLE 2. Comparison of different kernel functions

	$\widehat{RMSE} \times 10^5$ ($\rho = 0$)	
Kernel	$\beta = 1$	Optimal bandwidth selection
$K_{\exp}$	1.400	1.068
K_{unif}	1.890	1.608
K_1	2.173	1.648
K_2	2.064	1.476

TABLE 3. Comparison between optimal bandwidth and suboptimal bandwidth

	$\widehat{RMSE} \times 10^5$ ($\rho = -0.5$)		
Bandwidth	h_1 (optimal)	h_2 (suboptimal)	h_3 (suboptimal)
$\beta = 1$	1.418	1.605	1.754
$\beta = 2$	1.133	1.225	1.308
$\beta = 3$	1.077	1.121	1.678
$\beta = 4$	1.050	1.073	1.104

four kernels:

$$K_{\exp}(x) = \frac{1}{2}e^{-|x|}, \quad K_{unif}(x) = \frac{1}{2}\mathbb{1}_{\{|x|<1\}},$$

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

This shows that, indeed, the exponential kernel provides the best performance.

4.5. Optimal Bandwidth

First, we show that the suboptimal bandwidth, which corresponds to $\beta = 0$ in Theorem 2.2, indeed performs worse than the optimal bandwidth, even though its asymptotic variance is easier to estimate without the $\beta Z'_\tau$ term. For simplicity, we again only consider the Heston model (4.1) without jumps and the untruncated pre-averaging kernel estimator (2.18). We will compare the truncated and untruncated versions in more detail below in Section 4.7.

In Table 3, we compare the optimal bandwidth $h_1 = \beta \Delta_n^{1/4}$ with the suboptimal bandwidths $h_2 = \beta \Delta_n^{0.28}$ and $h_3 = \beta \Delta_n^{0.3}$, using the exponential kernel with $\beta = 1, 2, 3, 4$, respectively, based on 1,000 simulated paths. The results show the advantage in using the optimal bandwidth for the same level of the coefficient β.

Next, we compare the results of the iterative homogeneous and local bandwidth selection methods, as discussed in Section 3.2. Based on some initial simulations, we observed that the parameter θ, which controls the length of the pre-averaging

TABLE 4. Comparison of different bandwidth selection methods based on 1,000 simulations. RMSE for initial bandwidth $\beta = 1$ is 1.4086×10^{-5}. Columns 2 and 3 show the results corresponding to the first and second iterations of bandwidth selection methods. Columns 4 and 5 show the results using oracle and semi-oracle bandwidths, respectively

	$\widehat{RMSE} \times 10^5 (\rho = -0.5)$			
	1st iter.	2nd iter	Oracle	Semi-oracle
Homogeneous	1.0530	1.0529	1.0540	1.0533
Local	1.0571	1.0551	1.0542	1.0547

window k_n as $k_n = \frac{1}{\theta\sqrt{\Delta_n}}$, has comparatively smaller effect on the performance of estimator than that of the bandwidth. Therefore, throughout this section, we fix $\theta = 5$, which is computed by (3.1) using true parameter values, and consider different bandwidth selection techniques.[8]

In Table 4, we report the estimated RMSE values for different bandwidth selection methods. For the homogeneous bandwidth selection method (3.3), we apply the realized variance of sparsely sampled (5 min) spot variance estimates $\{\hat{c}_{t_i}\}$ to estimate the vol vol $\int_0^T \tilde{\sigma}_t^2 dt$ as described in Section 3.2. We fix the estimated vol vol after the first iteration to prevent the increased variance brought by the iterative method. The first two iterations are shown in the first two columns of the table, and we can see that the second iteration does not improve the result significantly. Therefore, one iteration of the bandwidth selection method is sufficient in practice. For the local bandwidth method, we use $\int_0^T \tilde{\sigma}_t^2 dt/T$ as a proxy of $\tilde{\sigma}_\tau^2$ in the formula (3.5). As a reference, we also give the results of using an oracle optimal bandwidth, which is computed by the true parameter values and the simulated spot variance process with equations (3.3) and (3.5) for the optimal homogeneous and optimal local bandwidths, respectively. In the last column, we provide the result of a semi-oracle type of bandwidth, where we use the estimated spot variance "skeleton" $\{\hat{c}_{t_i}\}$ to estimate $\int_0^T c_t dt$ and $\int_0^T c_t^2 dt$, via Riemann sums[9] while using the true parameter of γ given in (4.2) to estimate $\int_0^T \tilde{\sigma}_t^2 dt = \gamma^2 \int_0^T c_t dt$. The last simplification is possible due to the special structure of the diffusion coefficient of variance process in the Heston model (4.1). A similar approach can be applied whenever the vol vol depends on the volatility. As we can see therein, the data-driven approaches (first two columns) are quite close to the oracle and semi-oracle estimates.

---

[8]We also considered other values of θ, and the results were similar.

[9]We also apply the pre-averaging estimate of quarticity given in Jacod, Podolskij, and Vetter (2010), but the results were less optimal.
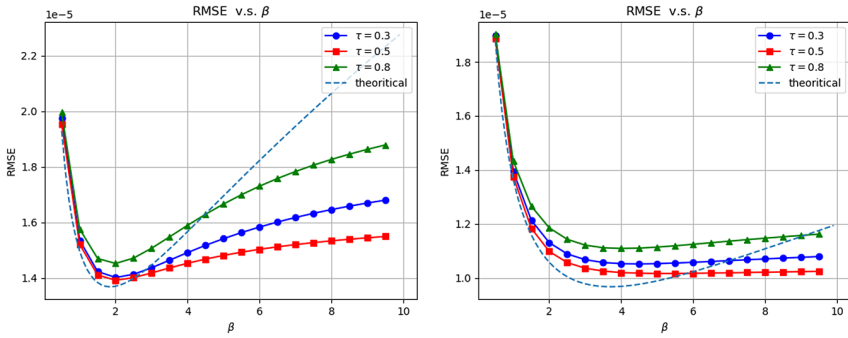
FIGURE 3. Left panel: MSE vs. bandwidth when $\gamma = \frac{1}{252}$. Right panel: MSE vs. bandwidth when $\gamma = \frac{0.5}{252}$.

Remark 4.1. The estimator with local bandwidth has the flexibility to adjust its bandwidth at different times based on the data. Therefore, theoretically, this estimator should be able to achieve a lower value of the integrated asymptotic variance $\int_0^T \bar{\delta}^2(\tau)d\tau$, which, as defined in Corollary 3.2, is given by $\Delta_n^{1/4} \int_0^T \left(\frac{1}{\beta_t} \delta_1^2(t) + \beta_t \delta_2^2(t) \right) dt$. However, our simulations show that the performance of the local bandwidth is almost the same as that of the homogeneous bandwidth. To further investigate this phenomenon, in the left panel of Figure 3, we show the estimated RMSE values for different times τ against the parameter β in the bandwidth formula $b_n = \beta \Delta_n^{1/4}$. As before, we simulate the Heston model (4.1) with the same parameters as in (4.2), but with the vol vol parameter $\gamma = \frac{1}{252}$. We can conclude from the figure that the optimal β-value is almost the same for different τ's, and this value is also close to the theoretical optimal homogeneous bandwidth based on the asymptotic variance of the estimator. Thus, an estimator with homogeneous bandwidth can achieve a similar result without extra computation cost. This trend is less obvious when the vol vol parameter γ is relatively small. In the right panel of Figure 3, we show the estimated RMSE vs. β when $\gamma = \frac{0.5}{252}$. In that case, the perceived almost flat trend as the bandwidth increases shows that the realized variance can serve as a good proxy of the spot volatility, at least for the purpose of tuning the parameters of the estimators, since the spot volatility estimator degenerates to the integrated volatility estimator when the bandwidth gets large.

4.6. Comparison with TSRSV

In this section, we adopt the model (4.3) with the same parameters as in Zu and Boswijk (2014):

$$\mu = 0.03, \ \beta_1 = 0.125, \ \alpha = -0.025, \ \rho = -0.3, \ \beta_0 = \beta_1^2/(2\alpha). \tag{4.6}$$

TABLE 5. Comparison between TSRSV and kernel pre-averaging estimator. We set the parameter θ in (2.13) to be 5, 5, and 1.5 for the noise levels $0.0001, 0.001$, and 0.01, respectively. For the TSRSV estimator, to reduce the computational cost, instead of choosing the initial bandwidth using the cross-validation method proposed in Kristensen (2010) as did in Section 4.2.1 of Zu and Boswijk (2014), we use the bandwidth already selected in Zu and Boswijk (2014, Tables 8–10), as the initial values to estimate the vol vol. Note that the obtained RMSEs of the smoothing TSRSV estimator under the different noise levels (0.0727, 0.1121, and 0.2345) match the results in Zu and Boswijk (2014), who reported the values 0.094, 0.118, and 0.223, respectively

Frequency	$\omega^2 = 0.0001$		$\omega = 0.001$		$\omega = 0.01$	
	$\hat{c}_{PA}$	$\hat{c}_{TSRSV}$	$\hat{c}_{PA}$	$\hat{c}_{TSRSV}$	$\hat{c}_{PA}$	$\hat{c}_{TSRSV}$
1 s	0.0411	0.0727	0.0546	0.1121	0.0634	0.2345
5 s	0.0505	0.1487	0.0649	0.1392	0.1066	0.3005

We also take $\gamma_0 \sim \mathcal{N}\left(0, -\frac{1}{2\alpha}\right)$ and $J = 0$. The microstructure noise ϵ_i^n is set to be $\epsilon_i^n := \epsilon_{t_i} \overset{i.i.d.}{\sim} \mathcal{N}\left(0, \omega^2\right)$, where, as in Zu and Boswijk (2014), ω^2 can take one of three possible levels: 0.0001, 0.001, and 0.01.

For the TSRSV estimator, we implement the smoothing version of the TSRSV (see Zu and Boswijk (2014, Sect. 3.1)), denoted by $\hat{c}_{TSRSV}$, and calculate the bandwidth and scale parameters according to Section 3.4 of Zu and Boswijk (2014). For our pre-averaging estimator, we implement the nontruncated version (denoted by $\hat{c}_{PA}$) with exponential kernel and use the iterative method described in Section 3.2 for bandwidth selection. We consider two sampling frequencies: 1 or 5 s. In Table 5, we report the RMSEs of the two estimators. As shown in the table, the pre-averaging estimator has a superior performance, especially when the noise level is large.

4.7. Comparison Between the Truncated and Untruncated Estimators

In this subsection, we study the two versions of the truncated estimators (2.19) and the nontruncated estimators (2.18) under various levels of jump size and sample frequency, using the simulation setting in Yu et al. (2014b). The parameters therein are chosen from Huang and Tauchen (2005):

$$\mu = 0.03, \ \beta_1 = 0.125, \ \alpha = -0.1, \ \rho = 0, \ \beta_0 = 0. \tag{4.7}$$

We also conduct our study in the same experiment design as in Yu et al. (2014b). More specifically, we consider three levels of jump activity (no jumps, compound Poisson jumps, and Variance Gamma jumps); three noise levels ($\omega = 0025, 0.035, 0.05$), and three different sample frequencies (one observation every 10 s, every 30 s, and every 60 s). In the case of FA jumps, $J_t = \sum_{j=1}^{N_t} Z_{\tau_j}$ with $Z_{\tau_j} \sim N(0, \sigma_Y^2)$ and $\{N_t\}_{t \geq 0} \sim Poisson(3)$, whereas in the case of infinite activity

TABLE 6. The RMSEs of the pre-averaging estimators. We set the parameter θ in (2.13) to be 5, 3, and 2 for 10-, 30-, and 60-s data, respectively. The truncation level is set to be $v_n = \alpha \sqrt{BPV \phi_{k_n}(g)}(\Delta_n)^{0.49}$, with $BPV = \frac{\pi}{2} \sum_{i=2}^{n} \left| \Delta_{i-1}^n X \right| \left| \Delta_i^n X \right|$ and calculated on sparsely sampled data (5-min frequency). When $\sigma_Y = 0$, $\alpha = 5, 4, 4$ for 10-, 30-, and 60-s data, respectively; when $\sigma_Y = 0.5$, $\alpha = 5.5, 3.5, 3$ for 10-, 30-, and 60-s data, respectively; when $\sigma_Y = 1.5$, $\alpha = 5.2, 3, 2.5$ for the respective frequencies; and, finally, in the case of jump with infinite activity, we set $\alpha = 6, 4, 4$ for 10-, 30-, and 60-s data, respectively

Frequency	$\omega = 0.025$			$\omega = 0.035$			$\omega = 0.05$		
	$\hat{c}_{T,2}$	$\hat{c}_{T,1}$	$\hat{c}_{Non}$	$\hat{c}_{T,2}$	$\hat{c}_{T,1}$	$\hat{c}_{Non}$	$\hat{c}_{T,2}$	$\hat{c}_{T,1}$	$\hat{c}_{Non}$
Scenario A: Diffusion with no jumps $\sigma_Y = 0$									
10 s	0.1421	0.1421	0.1420	0.1474	0.1474	0.1472	0.1689	0.1690	0.1677
30 s	0.1724	0.1724	0.1719	0.1763	0.1763	0.1758	0.1846	0.1846	0.1837
60 s	0.2057	0.2057	0.2054	0.2085	0.2085	0.2083	0.2161	0.2161	0.2160
Scenario B: Diffusion with small jumps $\sigma_Y = 0.5$									
10 s	0.1311	0.1630	1.0120	0.1347	0.1641	1.0004	0.1592	0.1805	0.9943
30 s	0.1660	0.1893	0.9758	0.1690	0.1913	0.9798	0.1797	0.1971	0.9783
60 s	0.2083	0.2209	0.9738	0.2113	0.2215	0.9732	0.2205	0.2255	0.9730
Scenario C: Diffusion with large jumps $\sigma_Y = 1.5$									
10 s	0.1952	0.9402	9.4648	0.2013	0.9397	9.3509	0.2090	0.9013	9.1702
30 s	0.2262	0.9183	9.0481	0.2256	0.9132	9.0438	0.2279	0.8990	8.9954
60 s	0.2640	0.8949	8.9794	0.2635	0.8892	8.9731	0.2590	0.8707	8.9514
Scenario D: Diffusion with jumps of infinite activity									
10 s	0.1246	0.1247	0.1271	0.1330	0.1331	0.1349	0.1529	0.1530	0.1547
30 s	0.1576	0.1577	0.1594	0.1587	0.1588	0.1598	0.1728	0.1729	0.1740
60 s	0.1940	0.1940	0.1952	0.1959	0.1959	0.1966	0.2063	0.2063	0.2076

jumps, $J_t = cG_t + \eta \tilde{W}_{G_t}$ with $G_t \sim Gamma(t/b, b)$, $b = 0.23$, $c = -0.2$, $\eta = 0.2$, and $\tilde{W}$ is an independent BM, as in Mancini (2009).

In Table 6, we report the RMSEs of our nontruncated estimators (2.18) and truncated estimators (2.19), denoted by $\hat{c}_{T,1}$, $\hat{c}_{T,2}$, and $\hat{c}_{Non}$, respectively. We observe that when there are no jumps, all three estimators have similar performance, with the nontruncated estimators giving slightly better results. However, when jumps are present, the truncated estimators have a much superior performance, especially when the jump size is large, and $\hat{c}_{T,2}$ appears to be more effective at eliminating jumps compared with $\hat{c}_{T,1}$.

Yu et al. (2014b) also proposed a pre-averaging kernel estimators for the spot volatility. As mentioned in the Introduction, their estimator has a different debiasing term, which could affect the finite-sample performance, and their asymptotic normality is only established with suboptimal convergence rate. Our results in

Table 6 are comparable with Yu et al. (2014b, Sect. 5). For example, the RMSE 0.1421 under $\omega = 0.025, \sigma_Y = 0$ with 10-s data is close to the RMSE $\sqrt{0.0201} = 0.1417$ in Yu et al. (2014b).

5. CONCLUSIONS

In this paper, we introduce high-frequency-based kernel estimators of the spot volatility under both the absence and presence of microstructure noise. One of the key differences of our results from those of earlier literature is to consider a general kernel in an asymptotic regime for the bandwidth that leads to optimal convergence rates for the resulting kernel estimators. Under this regime, kernels of unbounded support offer improved performance compared with uniform or other kernels with bounded support. General two-sided kernels of unbounded support were already advocated in the work of Figueroa-López and Li (2020a), where it was proved for the first time that exponential kernels are optimal, hence, formally validating an old conjecture of Foster and Nelson (1996). Unfortunately, Figueroa-López and Li (2020a) imposed strong assumptions for the validity of their results, the most important of which are the absence of leverage effects, microstructure noise, and jumps. These three effects are, of course, pervasive in real transaction data. In this work, we are able to relax all of those constraints and consider a rather general model. We further develop a feasible implementation of the proposed estimators. Via Monte Carlo experiments, we confirm the superior performance of the proposed estimators.

APPENDIX A. Proof of Theorem 2.1

We follow the steps in the proof of Theorem 13.3.3 in Jacod and Protter (2011) (which implies Theorem 13.3.7). By virtue of localization, without loss of generality, we assume throughout the proof that $|\delta(t,z)| \leq \Gamma(z), |\tilde{\delta}(\omega,t,z)| \wedge 1 \leq \Lambda(z)$ and

$$\Gamma(z) + \Lambda(z) + \int \Gamma(z)^r \lambda(dz) + \int \Lambda(z)^2 \lambda(dz) + |\mu_t| + |\sigma_t| + |X_t| + |\rho_t| + |\tilde{\sigma}_t| + |\tilde{\mu}_t| \leq A,$$

(see Section 4.4.1 and (6.2.1) of Jacod and Protter (2011) and Appendix A.5 of Aït-Sahalia and Jacod (2014) for details). We use C to represent a generic constant that may change from line to line.

A.1. Elimination of the Jumps and the Truncation

We denote the process $\int_0^t \int_{\mathbb{R}^d} \delta(s,x)\mu(ds,dx)$ by $\delta \star \mu_t$, and we set

$$X'' = \begin{cases} \delta \star \mathfrak{p}, & \text{if } r \leq 1, \\ \delta \star (\mathfrak{p} - \mathfrak{q}), & \text{if } r > 1, \end{cases} \qquad X' = X - X'', \qquad z_n = \begin{cases} \sqrt{m_n}, & \text{if } \beta < \infty, \\ \frac{1}{\sqrt{m_n \Delta_n}}, & \text{if } \beta = \infty. \end{cases}$$

To explicitly indicate the process Y for which the spot estimator is calculated, we use the notation $\hat{c}(m_n, v_n, Y)_\tau$. The proof of the following lemma is similar to the one of Theorem

13.3.3 in Jacod and Protter (2011) and is omitted for the sake of space (see details in Figueroa-López and Wu (2022)).

LEMMA A.1. *When $X = X'$ or when (2.10) and $\varpi \leq \frac{1-a}{r}$ hold, we have, as $n \to \infty$,*

$$\mathbb{P}\left(\hat{c}^n\left(m_n, v_n, X\right)_t \neq \hat{c}^n\left(m_n, X\right)_t\right) \to 0, \tag{A.1}$$

Furthermore, under (2.10) and (2.11), we have

$$z_n\left(\hat{c}^n\left(m_n, v_n, X\right)_t - \hat{c}^n\left(m_n, v_n, X'\right)_t\right) \xrightarrow{\mathbb{P}} 0. \tag{A.2}$$

A.2. Proof in the Continuous Case

With the previous lemma, it remains to prove the stable convergence (2.9) under the following assumption.

Assumption 4. We have (2.1) with X continuous, $c_t = \sigma_t^2$ satisfies (2.2), the processes $\mu, \tilde{\mu}, \sigma, \tilde{\sigma}$ are bounded, and $|\tilde{\delta}(\omega, t, z)| \wedge 1 \leq \Lambda(z)$ with a bounded function Λ on E satisfying $\int_E \Lambda(z)^2 \lambda(dz) < \infty$.

Now, we proceed our proof with the nontruncated estimator, which, for easiness of notation, is defined as $\hat{c}(m_n)_\tau := \sum_{i=1}^n K_{m_n \Delta_n}\left(t_{i-1} - \tau\right)\left(\Delta_i^n X\right)^2$. We first introduce some notation. Recall that $U_i^n := U_{i\Delta_n}$ and, for $t \in ((i-1)\Delta_n, i\Delta_n]$, let

$$
\begin{aligned}
V_t^n &:= \sum_{j=1}^n K_{m_n \Delta_n}\left(t_{j-1} - t\right)\left(\left(\Delta_j^n W\right)^2 - \Delta_n\right), \\
V_t'^n &:= \Delta_n \sum_{j=1}^n K_{m_n \Delta_n}\left(t_{j-1} - t\right)\left(B_j^n - B_i^n\right), \\
Z_t^n &:= c_i^n V_t^n, \quad Z_t'^n := \tilde{\sigma}_i^n V_t'^n, \quad Z_t''^n = \hat{c}(m_n)_t - c_t - Z_t^n - Z_t'^n.
\end{aligned}
\tag{A.3}
$$

All the three cases in Theorem 2.1 for the continuous case follow from the next two lemmas.

LEMMA A.2. *Under Assumptions 2 and 4, with $\Delta_n \to 0$, $m_n \Delta_n \to 0$, and $m_n \sqrt{\Delta_n} \to \infty$, we have the following stable convergence in law:*

$$\left(\sqrt{m_n} Z_t^n, \frac{1}{\sqrt{m_n \Delta_n}} Z_t'^n\right) \xrightarrow{st} \left(Z_t^{(0)}, Z_t'^{(0)}\right),$$

where Z_t^0 and $Z_t'^{(0)}$ are defined in (2.7).

LEMMA A.3. *Under Assumptions 2 and 4, we have, for all $t \in [0, T]$, $z_n^{(0)} Z_t''^n \xrightarrow{\mathbb{P}} 0$, where $z_n^{(0)} = m_n^{1/2}$ if $m_n \Delta_n^{1/2} \to \beta < \infty$ and $z_n^{(0)} = 1/\sqrt{m_n \Delta_n}$ if $m_n \Delta_n^{1/2} \to \beta = \infty$.*

We prove these two lemmas in the next two subsections.

A.2.1. *Proof of Theorem A.2.*   We first show

$$\left(\sqrt{m_n} V_t^n, \frac{1}{\sqrt{m_n \Delta_n}} V_t'^n \right) \xrightarrow{st} (V, V'), \tag{A.4}$$

where (V, V') are defined in (2.6). Denote the bandwidth of the kernel as $b_n := m_n \Delta_n$, recall that $t \in ((i-1)\Delta_n, i\Delta_n]$, and we can write the pair $\left(\sqrt{m_n} V_t^n, \frac{1}{\sqrt{m_n \Delta_n}} V_t'^n \right)$ as $\sum_{j=1}^n \left(\zeta_j^n(t), \zeta_j'^n(t) \right)$, where

$$\zeta_j^n(t) = \sqrt{m_n} K_{b_n} (t_{j-1} - t) \left(\left(\Delta_j^n W \right)^2 - \Delta_n \right),$$

$$\zeta_j'^n(t) = \frac{\Delta_n}{\sqrt{m_n \Delta_n}} \begin{cases} 0, & \text{if } j = 1, \\ -\left(\sum_{l=1}^{j-1} K_{b_n}(t_{l-1} - t) \right) \Delta_j^n B, & \text{if } 2 \le j \le i, \\ \left(\sum_{l=j}^n K_{b_n}(t_{l-1} - t) \right) \Delta_j^n B, & \text{if } i < j \le n. \end{cases}$$

Then we notice that $\left(\zeta_j^n(t), \zeta_j'^n(t) \right)$ is $\mathcal{F}^{(l)}{}_{t_j}$ measurable and with $\mathcal{F}_j := \mathcal{F}_{t_j}^{(0)}$, $\sum_{j=1}^n \mathbb{E} \left(\zeta_j^n(t) \middle| \mathcal{F}_{j-1}^{(0)} \right) = 0$, and $\sum_{j=1}^n \mathbb{E} \left(\zeta_j'^n(t) \middle| \mathcal{F}_{j-1}^{(0)} \right) = 0$. Recall that $\rho_s = d \langle W, B \rangle_s / ds$ is càdàg and bounded on the interval $[t_{j-1}, t_j]$. By Itô's lemma, the Cauchy–Schwarz inequality, and Doob's inequality, we have

$$\left| \mathbb{E} \left(\left(\Delta_j^n W \right)^2 \Delta_j^n B \middle| \mathcal{F}_{j-1}^{(0)} \right) \right| \le C \Delta_n^{3/2} \sqrt{ \mathbb{E} \left(\left(\rho_{t_j} - \rho_{t_{j-1}} \right)^2 \middle| \mathcal{F}_{j-1}^{(0)} \right) }.$$

Then, by a change of variable,

$$\left| \sum_{j=1}^n \mathbb{E} \left(\zeta_j^n(t) \zeta_j'^n(t) \middle| \mathcal{F}_{j-1} \right) \right|$$

$$\le C \Delta_n^2 \sum_{j=2}^i \left| K_{b_n} (t_{j-1} - t) \right| \left(\sum_{l=1}^{j-1} \left| K_{b_n}(t_{l-1} - t) \right| \right) \max_j \sqrt{ \mathbb{E} \left(\left(\rho_{t_j} - \rho_{t_{j-1}} \right)^2 \middle| \mathcal{F}_{j-1}^{(0)} \right) }$$

$$+ C \Delta_n^2 \sum_{j=i+1}^n \left| K_{b_n} (t_{j-1} - t) \right| \left(\sum_{l=j}^n \left| K_{b_n}(t_{l-1} - t) \right| \right) \max_j \sqrt{ \mathbb{E} \left(\left(\rho_{t_j} - \rho_{t_{j-1}} \right)^2 \middle| \mathcal{F}_{j-1}^{(0)} \right) }$$

$$\le C \int |K(u)| |L(u)| \, du \max_j \sqrt{ \mathbb{E} \left(\left(\rho_{t_j} - \rho_{t_{j-1}} \right)^2 \middle| \mathcal{F}_{j-1}^{(0)} \right) }.$$

We notice that ρ is right-continuous and uniformly bounded on [0,T] and, thus, $\sum_{j=1}^n \mathbb{E} \left(\zeta_j^n(t) \zeta_j'^n(t) \middle| \mathcal{F}_{j-1} \right) \to 0$, as $n \to \infty$. Next, we can deduce the following by the Riemann sum theorem and change of variables:

$$\sum_{j=1}^{n} \mathbb{E}\left(\zeta_j^n(t)^2 \,\middle|\, \mathcal{F}_{j-1}\right) = 2\sum_{j=1}^{n} m_n \Delta_n^2 K_{b_n}^2(t_{j-1} - t) \longrightarrow 2\int K^2(u)du,$$

$$m_n\Delta_n \sum_{j=1}^{n} \mathbb{E}\left(\zeta_j'^n(t)^2 \,\middle|\, \mathcal{F}_{j-1}\right) \sim \int_t^T \left(\int_v^T K_{b_n}(s-t)ds\right)^2 dv$$

$$+ \int_0^t \left(\int_0^v K_{b_n}(s-t)ds\right)^2 dv \longrightarrow \int L^2(u)du,$$

where $L(t) = \int_t^\infty K(u)du \mathbb{1}_{\{t>0\}} - \int_{-\infty}^t K(u)du \mathbb{1}_{\{t\leq 0\}}$. Note also that

$$\sum_{j=1}^{n} \{\mathbb{E}\left(\zeta_j^n(t)^4 \,\middle|\, \mathcal{F}_{i-1}\right) + \mathbb{E}\left(\zeta_j'^n(t)^4 \,\middle|\, \mathcal{F}_{i-1}\right)\} \leq \frac{C}{m_n}\int K^4(u)du + \frac{C}{m_n^2\Delta_n}\int L(u)^4 du \longrightarrow 0,$$

where U_j is a standard normal distribution and C is a generic constant. To apply Theorem 2.2.15 in Jacod and Protter (2011), we further need to show that

$$\text{(i)} \sum_{j=1}^{n} \mathbb{E}\left(\zeta_j^n(t)\left(M_{t_j} - M_{t_{j-1}}\right)\middle|\, \mathcal{F}_{j-1}\right) \to 0, \quad \text{(ii)} \sum_{j=1}^{n} \mathbb{E}\left(\zeta_j'^n(t)\left(M_{t_j} - M_{t_{j-1}}\right)\middle|\, \mathcal{F}_{j-1}\right) \to 0,$$

$$\text{(A.5)}$$

whenever M is either one of the components of (W, B) or is in the set $\mathcal{N}$ containing all bounded $\left(\mathcal{F}_t^{(0)}\right)$-martingales orthogonal (in the martingale sense) to (W, B). When $M = W$ or B, (A.5)(i)) holds true since it is the $\mathcal{F}_{(j-1)\Delta_n}$-conditional expectation of an odd function of the increments of the process W after time $(j-1)\Delta_n$. On the other hand, by the boundedness of the process ρ, we have $|\mathbb{E}\left(\Delta_j B \Delta_j W \middle|\, \mathcal{F}_{j-1}\right)| = \mathbb{E}\left(\left|\int_{t_{j-1}}^{t_j} \rho_s ds\right| \middle|\, \mathcal{F}_{|-\infty}\right) \leq C\Delta_n$, for some constant C and, thus, (A.5)(ii)) can be shown as follows:

$$\sum_{j=1}^{n} \mathbb{E}\left(\zeta_j'^n(t)\left(M_{t_j} - M_{t_{j-1}}\right)\middle|\, \mathcal{F}_{j-1}\right)$$

$$\leq \frac{\Delta_n^{3/2}}{\sqrt{m_n}}\left(\sum_{j=i+1}^{n}\left|\sum_{m=j}^{n} K_{b_n}\left(t_{m-1} - t\right)\right| + \sum_{j=2}^{i}\left|\sum_{m=1}^{j-1} K_{b_n}\left(t_{m-1} - t\right)\right|\right)$$

$$\leq C\frac{1}{m_n\Delta_n}\int |L(u)|\,du \to 0.$$

Suppose now that N is a bounded martingale, orthogonal to (W, B). By Itô's formula, we see that $\zeta_j^n(t)$ can be written as $\sqrt{m_n}K_{b_n}\left(t_{j-1} - t\right)\int_{t_{j-1}}^{t_j} 2\left(W_s - W_{t_{j-1}}\right)dW_s$, i.e., a stochastic integral with respect to W on the interval $[(j-1)\Delta_n, j\Delta_n]$. Similarly, $\zeta_j'^n(t)$ is a stochastic integral with respect to B on the same interval. Then the orthogonality of N and (W, B) implies (A.5). Now, we can apply Theorem 2.2.15 in Jacod and Protter (2011) and conclude that

$$\left(\sqrt{m_n}V_t^n, \frac{1}{\sqrt{m_n\Delta_n}}V_t'^n\right) \xrightarrow{st} \left(V, V'\right),$$

where V, V' are defined in (2.6). Finally, recall that $Z_t^n := c_i^n V_t^n$, and $Z_t'^n := \tilde{\sigma}_i^n V_t'^n$. From the càdlàg property of σ and $\tilde{\sigma}$, we see that $c_i^n \to c_t$ and $\tilde{\sigma}_i^n \to \tilde{\sigma}_t$, for $t \in ((i-1)\Delta_n, i\Delta_n]$. Then Lemma A.2 follows from (A.4).

A.2.2. *Proof of Lemma A.3.*   For $t \in ((i-1)\Delta_n, i\Delta_n]$, we can rewrite $Z_t''^n$ defined in (A.3) as $Z_t''^n = \sum_{j=1}^5 \zeta_j^n(t)$, where

$$\zeta_1^n(t) = c_i^n \Delta_n \sum_{j=1}^n K_{b_n}(t_{j-1} - t) - c_t,$$

$$\zeta_2^n(t) = \sum_{j=1}^n K_{b_n}(t_{j-1} - t)\left(\left(\Delta_j^n X\right)^2 - c_{j-1}^n \left(\Delta_j^n W\right)^2\right),$$

$$\zeta_3^n(t) = \sum_{j=1}^n K_{b_n}(t_{j-1} - t)\tilde{\sigma}_i^n \left(\left(\Delta_j^n W\right)^2 - \Delta_n\right)\left(B_j^n - B_i^n\right),$$

$$\zeta_4^n(t) = \sum_{j=1}^n K_{b_n}(t_{j-1} - t)\left(c_{j-1}^n - c_i^n - \tilde{\sigma}_i^n \left(B_j^n - B_i^n\right)\right)\left(\Delta_j^n W\right)^2.$$

Therefore, it is enough to prove that, for $l = 1, 2, 3, 4$ and all $t \in [0, T]$, we have

$$z_n^{(0)}\zeta_l^n(t) \xrightarrow{\mathbb{P}} 0. \tag{A.6}$$

Proof of (A.6) for $l = 1$. By Lemma 3.1 in Figueroa-López and Li (2020b) with $f = 1$ and Assumption 2,

$$\Delta_n \sum_{j=1}^n K_{b_n}(t_{j-1} - t) - \int_0^T K_{b_n}(s - t)ds = \frac{1}{2}\left(K(A^+) - K(B^-)\right)\frac{\Delta_n}{b} + \mathrm{o}\left(\frac{\Delta_n}{b}\right) = \mathrm{O}\left(\frac{\Delta_n}{b}\right),$$

where (A, B) is the support of K and $-\infty \le A < 0 < B \le \infty$. Therefore, the boundedness of c implies that

$$\zeta_1^n(t) = c_i^n \left(\int_0^T K_{b_n}(s - t)ds\right) - c_t + \mathrm{O}\left(\frac{\Delta_n}{b}\right)$$

$$= c_i^n - c_t + C\int_{(0,T)^c} K_{b_n}(t - \tau)dt + \mathrm{O}\left(\frac{\Delta_n}{b}\right).$$

Furthermore, we can deduce $\mathbb{E}(c_i - c_t)^2 \le C\Delta_n$, for $t \in ((i-1)\Delta_n, i\Delta_n]$ from (2.2). Assumption 2 implies that $x^{1/2}\int_x^\infty K(u)du \to 0$, as $x \to \infty$. We then have

$$b_n^{-1/2}\int_{(0,T)^c} K_{b_n}(t - \tau)dt = \frac{1}{\sqrt{b_n}}\left(\int_{-\infty}^{\frac{\tau}{b_n}} K(u)du + \int_{\frac{T-\tau}{b_n}}^\infty K(u)du\right) \to 0, \text{ as } n \to \infty. \quad \square$$

Proof of (A.6) for $l = 2$. Let $\rho_j^n(t) = \Delta_j^n X - \sigma_{j-1}^n \Delta_j^n W$. In view of (2.1.44) in Jacod and Protter (2011), for $q \ge 2$, we have

$$\mathbb{E}\left(\left|\rho_j^n(t)\right|^q\right) \le K_q \Delta_n^{1+q/2}, \quad \mathbb{E}\left(\left|\sigma_{j-1}\Delta_j^n W\right|^q\right) \le C\Delta_n^{q/2}.$$

Then, since $\left| \left(\Delta_j^n X \right)^2 - \sigma_{j-1}^2 \left(\Delta_j^n W \right)^2 \right| \leq 2 \left(\left| \rho_j^n(t) \right|^2 + \left| \rho_j^n(t) \right| \left| \sigma_{j-1}^2 \Delta_j^n W \right| \right)$, the inequalities above and the Cauchy–Schwarz inequality yield

$$\mathbb{E} \left| \zeta_2^n(t) \right| \leq 2 \sum_{j=1}^n \left| K_{b_n}(t_{j-1} - t) \right| \mathbb{E} \left(\left| \rho_j^n(t) \right|^2 + \sqrt{\mathbb{E} \left| \rho_j^n(t) \right|^2 \mathbb{E} \left| \sigma_{j-1}^2 \Delta_j^n W \right|^2} \right)$$

$$\leq C \sum_{j=1}^n \left| K_{b_n}(t_{j-1} - t) \right| \left(\Delta_n^2 + \Delta_n^{3/2} \right) \sim \int K(u) du \sqrt{\Delta_n}.$$

We then have the result since $z_n \sqrt{\Delta_n} \to 0$. $\qquad \square$

Proof of (A.6) for $l = 3$. $\zeta_3^n(t)$ can be written as $\tilde{\sigma}_i^n \Phi^n(t)$ where each $\tilde{\sigma}_i^n$ is bounded $\mathcal{F}_i^{(0)}$ measurable and $\Phi^n(t) = \sum_{j=1}^n K_{b_n}(t_{j-1} - t) \left(\left(\Delta_j^n W \right)^2 - \Delta_n \right) \left(B_{j-1}^n - B_i^n \right)$. We can compute that $\mathbb{E} \left(\Phi^n(t) \right) = 0$ and $\left| \mathbb{E} \left(\Delta_j^n W \Delta_j^n B \right) \right| = \left| \mathbb{E} \left(\int_{t_{j-1}}^{t_j} \rho_s ds \right) \right| \leq C \Delta_n$. Notice that $\left(\Delta_j^n W \right)^2 - \Delta_n, B_{j-1}^n - B_i^n$ are independent when $j \geq i+1$ and $\left(\Delta_j^n W \right)^2 - \Delta_n, B_i^n - B_j^n$ are independent when $j \leq i$. Then, by tower property, we have

$$\mathbb{E} \left(\Phi^n(t) \right)^2 \leq 2 \Delta_n^2 \sum_{j=i+1}^n K_{b_n}^2(t_{j-1} - t)(t_{j-1} - t_i) + \sum_{j=1}^i K_{b_n}^2(t_{j-1} - t) \left(2\Delta_n^2(t_i - t_j) + C_1 \Delta_n^3 \right)$$

$$\sim \Delta_n \left(\int_0^\infty K^2(u) u du - \int_{-\infty}^0 K^2(u) u du \right),$$

where $C_1^2 = \mathbb{E}(\chi_1^2 - 1)^4 \mathbb{E} \left(\chi_1^2 \right)^2$. Then $\frac{1}{\sqrt{\Delta_n}} \Phi^n(t)$ is bounded in probability, and the result follows, since $z_n \sqrt{\Delta_n} \to 0$. $\qquad \square$

Proof of (A.6) for $l = 4$. Let $\eta_j^n = \left(c_{j-1}^n - c_i^n - \tilde{\sigma}_i^n \left(B_{j-1}^n - B_i^n \right) \right) = \int_{t_i}^{t_{j-1}} \tilde{\mu}_s ds + \int_{t_i}^{t_{j-1}} \left(\tilde{\sigma}_s - \tilde{\sigma}_i^n \right) dB_s + M_{t_{j-1}} - M_i$, where $M = \tilde{\delta} \star (\mathfrak{p} - \mathfrak{q})$. Following the same argument for proof of (13.3.37) for $(j = 6)$ in Jacod and Protter (2011), on the set $\Omega(n, N, \varepsilon) = \{ |\Delta M_s| \leq \varepsilon, \forall s \in (i - N m_n \Delta_n, i + N m_n \Delta_n] \}$, with the notation $\gamma_j^n = \frac{1}{t_{j-1} - t_i} \mathbb{E} \left(\int_{t_i}^{t_{j-1}} \left| \tilde{\sigma}_s - \tilde{\sigma}_{t_i} \right|^2 ds \right)$, we deduce that, for $j \in (i - N m_n \Delta_n, i + N m_n \Delta_n]$,

$$\mathbb{E} \left(\left(\eta_j^n \right)^2 \mathbb{1}_{\Omega(n,i,\varepsilon)} \right) \leq C(t_{j-1} - t_i) \rho(n, j, \varepsilon), \quad \text{with } \rho(n, \varepsilon) = \frac{t_{j-1} - t_i}{\varepsilon} + \gamma_j^n + \phi(\varepsilon),$$

where $\phi(\epsilon) = \int_{\{\Lambda(z) < \varepsilon\}} \Lambda(z)^2 \lambda(dz)$ going to 0 as $\epsilon \to 0$. Since $\tilde{\sigma}$ is càdlàg and bounded, we see that $\gamma_j^n \to 0$ for all j and thus $\rho(n, j, \epsilon) \to 0$. From (2.1.44) in Jacod and Protter (2011), we also have, for all j, $\mathbb{E} \left(\left(\eta_j^n \right)^2 \right) \leq C(t_{j-1} - t_i)$ and, thus,

$$\mathbb{E} \left| \zeta_4^n(t) \right| = \mathbb{E} \left| \zeta_4^n(t) \right| \mathbb{1}_{\Omega(n,N,\varepsilon)} + \mathbb{E} \left| \zeta_4^n(t) \right| \mathbb{1}_{\Omega(n,N,\varepsilon)^c}$$

$$\leq CN \sqrt{N m_n \Delta_n} \rho(n, \epsilon) + C \sqrt{m_n \Delta_n} \left(\int_N^\infty K(u) \sqrt{u} du + \int_{-\infty}^{-N} K(u) \sqrt{u} du \right).$$

Additionally, we have $\lim_{\varepsilon \to 0} \limsup_n \rho(n, \varepsilon) = 0$ and $\int_N^\infty K(u)\sqrt{u}\,du + \int_{-\infty}^{-N} K(u)$ $\sqrt{u}\,du \overset{N \to \infty}{\longrightarrow} 0$. The result follows by $z_n \sqrt{b_n} < \infty$. $\qquad \square$

APPENDIX B. Proof of Theorem 2.2

Again, by virtue of localization, without loss of generality, we assume throughout the proof that $|\delta(t, z)| \leq \Gamma(z)$, $|\tilde{\delta}(\omega, t, z)| \wedge 1 \leq \Lambda(z)$ and

$$\Gamma(z) + \Lambda(z) + \int \Gamma(z)^r \lambda(dz) + \int \Lambda(z)^2 \lambda(dz) + |\mu_t| + |\sigma_t| + |X_t| + |\rho_t| + |\tilde{\sigma}_t| + |\tilde{\mu}_t| \leq A.$$

B.1. Elimination of the Jumps and the Truncation

We set

$$X'' = \begin{cases} \delta \star \mathfrak{p}, & \text{if } r \leq 1, \\ \delta \star (\mathfrak{p} - \mathfrak{q}), & \text{if } r > 1, \end{cases} \quad X' = X - X'', \quad \tilde{z}_n = \begin{cases} m_n^{1/2} \Delta_n^{1/4}, & \text{if } \beta < \infty, \\ \dfrac{1}{\sqrt{m_n \Delta_n}}, & \text{if } \beta = \infty. \end{cases}$$

Let $Y^* = Y - X + X'$ be the continuous process with microstructure noise and set

$$\hat{c}^*(k_n, m_n)_\tau = \frac{1}{\phi_{k_n}(g)} \sum_{j=1}^{n-k_n+1} K_{m_n \Delta_n}(t_{j-1} - \tau) \left(\left(\overline{Y}_j^{*n} \right)^2 - \frac{1}{2} \widehat{Y}_j^{*n} \right). \tag{B.1}$$

We need some preliminary estimates:

- By Corollary 2.1.9(a)–(c) in Jacod and Protter (2011), for $p > 0$ and $q \in [0, 1/r)$,

$$E\left[\sup_{u \in [0, s]} \left(\frac{|X''_{\tau+u} - X''_\tau|}{s^q} \wedge 1 \right)^p \Bigg| \mathcal{F}_\tau^{(0)} \right] \leq Cs^{(1-qr)(p/r \wedge 1)} a(s), \tag{B.2}$$

  where $a(s) \to 0$ as $s \to 0$. Let $g_n(t) = \sum_{j=1}^{k_n} g(j/k_n) \mathbb{1}_{((j-1)\Delta_n, j\Delta_n]}(t)$. With $u_n = k_n \Delta_n$ and $\bar{X}_i''^n = \int_{(i-1)\Delta_n}^{(i-1)\Delta_n + u_n} g_n(s - (i-1)\Delta_n) dX_s''$, the same as (9.2.13) in Jacod and Protter (2011), we deduce that

$$E\left[\left(\frac{|\bar{X}_i''^n|}{u_n^q} \wedge 1 \right)^p \Bigg| \mathcal{F}_{(i-1)\Delta_n}^{(0)} \right] \leq Cu_n^{(1-qr)(p/r \wedge 1)} a_n, \tag{B.3}$$

  where $a_n \to 0$ as $n \to \infty$.
- By the proof of Lemma 16.4.3 in Jacod and Protter (2011), for all $q > 0$,

$$\mathbb{E}\left(|\bar{X}_i'^n|^q \mid \mathcal{F}_{(i-1)\Delta_n} \right) \leq C_q \Delta_n^{q/4}, \quad \mathbb{E}\left(|\hat{x}_i'^n|^q \Big| \mathcal{F}_{(i-1)\Delta_n} \right) \leq C_q \Delta_n^{3q/2},$$

$$\mathbb{E}\left(|\hat{X}_i^n|^q \Big| \mathcal{F}_{(i-1)\Delta_n} \right) \leq C_q \Delta_n^{q/2} \left(\Delta_n^q + \Delta_n^{q \wedge 1} \right),$$

$$\mathbb{E}\left(|\bar{Y}_i^n|^q \Big| \mathcal{F}_{(i-1)\Delta_n} \right) \leq C_q \Delta_n^{q/4} \left(1 + \Delta_n^{-(q/4 - 1/2)+} \right), \tag{B.4}$$

$$\mathbb{E}\left(|\hat{Y}_i^n|^q \Big| \mathcal{F}_{(i-1)\Delta_n} \right) \leq C_q \Delta_n^{q/2} \left(\Delta_n^q + 1 + \Delta_n^{q \wedge 1} \right).$$

- From (16.2.3) in Jacod and Protter (2011), for all $p > 0$,

$$\mathbb{E}\left(\left|\bar{\epsilon}(g)_i^n\right|^p \mid \mathcal{H}_{(i-1)\Delta_n}\right) \le C_p \Delta_n^{p/4}, \quad \mathbb{E}\left(\left|\widehat{\epsilon}_i^n\right|^p \mid \mathcal{H}_{(i-1)\Delta_n}\right) \le C_p \Delta_n^{p/2}. \tag{B.5}$$

- Combining (B.4) and (B.5), and applying $(a+b)^q \le K_q(a^q + b^q)$, for all $q > 0$,

$$\mathbb{E}\left(\left|\bar{Y}_i^{*n}\right|^q \mid \mathcal{F}_{(i-1)\Delta_n}\right) \le C_q \mathbb{E}\left(\left|\bar{X}'(g)_i^n\right|^q + \left|\bar{\epsilon}(g)_i^n\right|^q \mid \mathcal{F}_{(i-1)\Delta_n}\right) \le C_q \Delta_n^{q/4},$$

$$\mathbb{E}\left(\left|\widehat{Y}_i^{*n}\right|^q \mid \mathcal{F}_{(i-1)\Delta_n}\right) \le C_q \mathbb{E}\left(\left|\widehat{X}_i^{\prime n}\right|^q + \left|\widehat{\epsilon}_i^n\right|^q \mid \mathcal{F}_{(i-1)\Delta_n}\right) \le C_q \Delta_n^{q/2}. \tag{B.6}$$

The following result will allow us to reduce the proof to the case of a continuous process and a kernel estimator without truncation.

LEMMA B.1. *Under Assumption 1, we have*

$$\tilde{z}_n \left|\hat{c}\left(k_n, m_n, v_n, l\right)_\tau - \hat{c}^*\left(k_n, m_n\right)_\tau\right| \overset{\mathbb{P}}{\longrightarrow} 0, \tag{B.7}$$

for both $l = 1, 2$ if $X = X'$, for $l = 1$ if (2.22) and (2.23) hold, and for $l = 2$ if (2.22) and (2.24) hold.

Proof. Let E_j^n denote the conditional expectation with respect to $\mathcal{F}_{(j-1)\Delta_n}$. We first check the proof for the case of $l = 2$. The proof for $l = 1$ is shown below. We can write

$$\left|\left(\left(\bar{Y}_j^n\right)^2 - \frac{1}{2}\widehat{Y}_j^n\right) \mathbb{1}_{\left\{\left|\bar{Y}_j^n\right| \le v_n\right\}} - \left(\left(\bar{Y}_j^{*n}\right)^2 - \frac{1}{2}\widehat{Y}_j^{*n}\right)\right| \le \sum_{r=1}^4 \eta_j^{n,r},$$

where

$$\eta_j^{n,1} = \left|\left(\bar{Y}_j^n\right)^2 \mathbb{1}_{\left\{\left|\bar{Y}_j^n\right| \le v_n\right\}} - \left(\bar{Y}_j^{*n}\right)^2 \mathbb{1}_{\left\{\left|\bar{Y}_j^{*n}\right| \le v_n\right\}}\right|,$$

$$\eta_j^{n,2} = \frac{1}{2}\left|\widehat{Y}_j^n \mathbb{1}_{\left\{\left|\bar{Y}_j^n\right| \le v_n\right\}} - \widehat{Y}_j^{*n} \mathbb{1}_{\left\{\left|\bar{Y}_j^{*n}\right| \le v_n\right\}}\right|, \tag{B.8}$$

$$\eta_j^{n,3} = \left|\bar{Y}_j^{*n}\right|^2 \mathbb{1}_{\left\{\left|\bar{Y}_j^{*n}\right| > v_n\right\}}, \quad \eta_j^{n,4} = \frac{1}{2}\left|\widehat{Y}_j^{*n}\right| \mathbb{1}_{\left\{\left|\bar{Y}_j^{*n}\right| > v_n\right\}}.$$

When $X = X', \eta_j^{n,1} = 0$. When $r \in (0, 2]$, by the proof of Lemma 2 in Chen (2019),[10] under (B.3), there is a sequence $a_n \to 0$ such that[11] $E_j^n[|\eta_j^{n,1}|] \le C\Delta_n^{(\varpi+1/2)-r\varpi/2} a_n$. Next, $\eta_j^{n,2} = 0$ when $X = X'$. When $r \in (0, 2]$, we have

$$\left|\widehat{Y}_j^n - \widehat{Y}_j^{*n}\right| \le C\left(\widehat{X}_j^{\prime\prime n} + \sqrt{\widehat{Y}_j^{*n}\widehat{X}_j^{\prime\prime n}}\right).$$

If we set $\mu, \sigma = 0$, we have $\mathbb{E}_j^n \left|\widehat{X}_j^{\prime\prime n}\right|^q = \mathbb{E}_j^n \left|\widehat{X}_j^n\right|^q \le K_q \Delta_n^{q/2}\left(\Delta_n^q + \Delta_n^{q \wedge 1}\right)$ from (B.4) for $q > 0$. Combining with (B.6) and taking $q = 1$, we have $E_j^n|\widehat{Y}_j^n - \widehat{Y}_j^{*n}| \le C(\Delta_n^{3/2} +$

---

[10]Or follow the proof of Lemma 13.2.6 in Jacod and Protter (2011).

[11]The notation $\bar{Y}, \widehat{Y}$ is slightly different in Chen (2019), which results in the different form of the following inequality. The relation between ϖ in our setting and ρ in Chen (2019) is $\varpi/2 + 1/4 = \rho$.

$\Delta_n^{1/4}\Delta_n^{3/4}) \leq C\Delta_n$. By the notation (2.15) and the Cauchy–Schwarz inequality, we have, for arbitrary $\alpha > 0$,

$$\eta_j^{n,2} \leq \frac{1}{2}\left|\widehat{Y}_j^n - \widehat{Y}_j^{*n}\right| + \widehat{Y}_j^n \frac{\left|\bar{Y}_j^{*n}\right|^\alpha}{v_n^\alpha} + \widehat{Y}_j^{*n}\frac{\left|\bar{Y}_j^{*n}\right|^\alpha}{(v_n/2)^\alpha} + \widehat{Y}_j^{*n}\left(\frac{\left|\bar{X}_j^{\prime\prime n}\right|}{v_n/2} \wedge 1\right),$$ (B.9)

where the last term is because $2\left|\bar{Y}_j^{*n}\right| < v_n < \left|\bar{Y}_j^n\right|$ implies $\left|\bar{X}_j^{\prime\prime n}\right| > v_n/2$, and therefore, $\frac{\left|\bar{X}_j^{\prime\prime n}\right|}{v_n/2} \wedge 1 > \mathbb{1}_{\left\{2\left|\bar{Y}_j^{*n}\right| < v_n < \left|\bar{Y}_j^n\right|\right\}}$. Applying the Cauchy–Schwarz inequality and (B.3) and (B.4), we have

$$\eta_j^{n,2} \leq C\left(\Delta_n + \sqrt{\Delta_n \Delta_n^{\alpha(1/2-\varpi)}} + \sqrt{\Delta_n 2^\alpha \Delta_n^{\alpha(1/2-\varpi)}} + \sqrt{\Delta_n \Delta_n^{1/2-\varpi r/2} a_n}\right).$$ (B.10)

Since $\varpi < 1/2$, there exists α such that $\alpha(1/2-\varpi) \geq 1/2$. Then, $E_j^n|\eta_j^{n,2}| \leq C(\Delta_n + \Delta_n^{3/4-\varpi r/4} a_n^{1/2})$. By Cauchy–Schwarz's and Markov's inequalities and (B.6), for an arbitrary positive number m,

$$E_j^n\left|\eta_j^{n,3}\right| \leq \sqrt{E_j^n\left|\bar{Y}_j^{*n}\right|^4 \mathbb{P}_j^n\left(\left|\bar{Y}_j^{*n}\right| > v_n\right)} \leq C\sqrt{\Delta_n E_j^n\left|\bar{Y}_j^{*n}\right|^m /v_n^m} \leq C\sqrt{\Delta_n^{1+m/4-m\varpi/2}}.$$

Similarly, there exists m such that $E_j^n\left|\eta_j^{n,3}\right| \leq C\Delta_n$. By the same argument, $E_j^n\left|\eta_j^{n,4}\right| \leq C\Delta_n$. Now, we have the following when $r \in (0,2]$ by combining the above inequalities:

$$E\left(|\hat{c}(k_n,m_n,v_n,2)_\tau - \hat{c}^*(k_n,m_n)_\tau\right)$$
$$\leq C\Delta_n^{1/2} \sum_{j=1}^{n-k_n+1} K_{m_n\Delta_n}(t_{j-1}-\tau)\left[\Delta_n^{\varpi+1/2-r\varpi/2} a_n + \Delta_n + \sqrt{\Delta_n^{3/2-\varpi r/2} a_n}\right]$$
$$\leq C\left[\Delta_n^{\varpi-r\varpi/2} a_n + \sqrt{\Delta_n} + \Delta_n^{1/4-\varpi r/4}\sqrt{a_n}\right],$$

by the property of the kernel. Therefore, recalling (2.22), to show that $\tilde{z}_n\left|\hat{c}(k_n,m_n,v_n,2)_\tau - \hat{c}^*(k_n,m_n)_\tau\right|$ is $o_P(1)$, we need the condition (2.24). This concludes (B.7) for $l=2$.

When $l=1$, $\hat{c}(k_n,m_n,v_n,1) - \hat{c}^*(k_n,m_n)$ has a similar decomposition like (B.8) with $\eta_j^{n,2} = \frac{1}{2}\left|\widehat{Y}_j^n - \widehat{Y}_j^{*n}\right|$ and $\eta_j^{n,4} = 0$. Therefore, $E\left(|\hat{c}(k_n,m_n,v_n,1)_\tau - \hat{c}^*(k_n,m_n)_\tau\right) \leq C\left[\Delta_n^{\varpi-r\varpi/2} a_n + \sqrt{\Delta_n}\right]$. In that case, the second inequality in (2.23) gives (B.7) for $l=1$. Since $\varpi < 1/2$, we necessarily need that $r < \frac{5}{2} - 2\left[\left(a-\frac{1}{4}\right) \wedge \left(1-a+\frac{1}{4}\right)\right]$.

Finally, when $X=X'$, note that $E\left(|\hat{c}(k_n,m_n,v_n,l)_\tau - \hat{c}^*(k_n,m_n)_\tau\right) \leq C\sqrt{\Delta_n}$, for $l=1,2$, since $\eta_j^{n,1} = \eta_j^{n,2} = 0$. Therefore, we can conclude (B.7) in all the cases stated in the statement of the lemma.   □

B.2. Proof of Stable Convergence in Law for Continuous Process

With the lemma above, it suffices to prove Theorem 2.2 for the nontruncated estimator (B.1) under Assumption 4. We first introduce some notations needed for the proofs.

Define

$$
\begin{aligned}
\phi(Y)_i^n &= (\overline{Y}_i^n)^2 - \frac{1}{2}\hat{Y}_i^n = (\overline{X}_i^n + \overline{\epsilon}_i^n)^2 - \frac{1}{2}\hat{Y}_i^n, \\
\phi_{i,j}^n &= (\sigma_{(i-j-1)\Delta_n}\overline{W}_i^n + \overline{\epsilon}_i^n)^2 - \frac{1}{2}\hat{\epsilon}_i^n, \\
\Psi_{i,j}^n &= \mathbb{E}(\phi_{i,j}^n|\mathcal{H}_{(i-1)\Delta_n}) - (\sigma_{(i-j-1)\Delta_n}\overline{W}_i^n)^2.
\end{aligned}
\tag{B.11}
$$

With any process U, we associate the variables

$$
\Gamma(U)_i^n = \sup_{t\in[(i-1)\Delta_n, i\Delta_n+k_n\Delta_n]} |U_t - U_{(i-1)\Delta_n}|, \quad \Gamma'(U)_i^n = \left(\mathbb{E}\left((\Gamma(U)_i^n)^4\Big|\mathcal{F}_{(i-1)\Delta_n}\right)\right)^{1/4}.
$$

The following decomposition will be instrumental to deduce the behavior of the estimation error:

$$
\hat{c}(k_n, m_n)_\tau - c_\tau = \sum_{l=1}^{5} \overline{H}(l)^n,
$$

where, using the notation $\mathbb{E}_{i-1}[\cdot] = \mathbb{E}[\cdot|\mathcal{F}_{(i-1)\Delta_n}]$, $b_n := m_n\Delta_n$, and $\phi_n := \phi_{k_n}(g)$,

$$
\overline{H}(1)^n = \frac{1}{\phi_n}\sum_{i=1}^{n-k_n+1} K_{b_n}(t_{i-1}-\tau)\left(\phi_{i,0}^n - \mathbb{E}_{i-1}(\phi_{i,0}^n)\right), \quad \overline{H}(2)^n = \int_0^T K_{b_n}(t-\tau)c_t dt - c_\tau,
$$

$$
\overline{H}(3)^n = \frac{1}{\phi_n}\sum_{i=1}^{n-k_n+1} K_{b_n}(t_{i-1}-\tau)\left(\phi(Y)_i^n - \phi_{i,0}^n\right), \quad \overline{H}(4)^n = \frac{1}{\phi_n}\sum_{i=1}^{n-k_n+1} K_{b_n}(t_{i-1}-\tau)\mathbb{E}_{i-1}\left(\Psi_{i,0}^n\right),
$$

$$
\overline{H}(5)^n = \Delta_n\sum_{i=1}^{n-k_n+1} K_{b_n}(t_{i-1}-\tau)c_{t_{i-1}} - \int_0^T K_{b_n}(t-\tau)c_t dt.
$$

The first term is the statistical error, whereas the second term is the local approximation error. Each of these will contribute one term to the asymptotic variance in (2.21)(i).

We recall some needed estimates and preliminary results.

1. By Lemma 16.5.14 in Jacod and Protter (2011), for some constant C,

$$
\left|\mathbb{E}_{i-1}\left(\phi(Y)_i^n - \phi_{i,0}^n\right)\right| \leq C\Delta_n^{3/4}\left(\Delta_n^{1/4} + \Gamma'(\mu)_i^n + \Gamma'(\widetilde{\sigma})_i^n + \Gamma'(\gamma)_i^n\right).
\tag{B.12}
$$

2. As in Lemma 16.5.15 in Jacod and Protter (2011), if an array (δ_i^n) satisfies

$$
0 \leq \delta_i^n \leq K, \quad \Delta_n\mathbb{E}\sum_{i=1}^{n}\delta_i^n \to 0,
\tag{B.13}
$$

then, for any $q > 0$, the array $\left(|\delta_i^n|^q\right)$ also satisfies (B.13). Furthermore, if U is a càdlàg bounded process, the two arrays $\left(\Gamma(U)_i^n\right)$ and $\left(\Gamma'(U)_i^n\right)$ also satisfy (B.13).

3. Under Assumptions 3 and (2.2), by Lemma 16.5.13 in Jacod and Protter (2011), for all $q > 0$,

$$\mathbb{E}_{i-1}\left(\left|\phi\left(Y\right)_i^n\right|^q + \left|\phi_{i,0}^n\right|^q\right) \leq C_q \Delta_n^{q/2}, \tag{B.14}$$

$$\mathbb{E}_{i-1}\left(\left|\phi\left(Y\right)_i^n - \phi_{i,0}^n\right|^2\right) \leq C\Delta_n\left(\Delta_n^{1/2} + \left(\Gamma'(\sigma)_i^n\right)^2\right), \quad \mathbb{E}_{i-j-1}\left(\left|\phi_{i,j}^n\right|^q\right) \leq C_q \Delta_n^{q/2}. \tag{B.15}$$

4. Let $\gamma_t' = \mathbb{E}(|\epsilon_t|^3 |\mathcal{F}^{(0)})$. Under Assumptions 3 and (2.2), by Lemma 16.5.12 in Jacod and Protter (2011), $\Psi_{i,j}^n$ defined in (B.11) is such that

$$\mathbb{E}_{i-1}\left|\Psi_{i,j}^n\right| \leq C\Delta_n + C\Delta_n^{3/4}\left(\Gamma'(\gamma)_i^n + \Gamma'\left(\gamma'\right)_i^n\right), \quad \mathbb{E}_{i-1}\left|\Psi_{i,j}^n\right|^2 \leq C\Delta_n^{3/2}. \tag{B.16}$$

5. By Itô's Lemma and Burkholder–Davis–Gundy inequalities (see Section 2.1.5 of Jacod and Protter (2011)), we have, for all $s, t \geq 0$ and $p \geq 2$,

$$\mathbb{E}\left(\sup_{r \in [0,s]}\left|\sigma_{t+r}^2 - \sigma_t^2\right|^p \Big| \mathcal{F}_t\right) \leq C_p s, \quad \mathbb{E}\left(\sup_{r \in [0,s]}|\sigma_{t+r} - \sigma_t|^p \Big| \mathcal{F}_t\right) \leq C_p s, \tag{B.17}$$

$$\Gamma'(\sigma)_i^n = \mathbb{E}\left(\sup_{t \in [(i-1)\Delta_n, i\Delta_n + k_n\Delta_n]}\left|\sigma_t - \sigma_{(i-1)\Delta_n}\right|^4 \Big| \mathcal{F}_{(i-1)\Delta_n}\right)^{1/4} \leq C\left(k_n\Delta_n\right)^{1/4}. \tag{B.18}$$

Theorem 2.2 will then follow from the following lemmas.

LEMMA B.2. *Under Assumptions 2–4, with $m_n \to \infty$, $m_n\Delta_n \to 0$, and $m_n\sqrt{\Delta_n} \to \infty$, we have $m_n^{1/2}\Delta_n^{1/4}\overline{H}(1)^n \xrightarrow{st} Z_\tau$, where Z_τ is described in Theorem 2.2.*

LEMMA B.3. *Under Assumptions 2–4, with $m_n \to \infty$, $m_n\Delta_n \to 0$, and $m_n\Delta_n^{3/4} \to \beta \in [0,\infty)$, $m_n^{1/2}\Delta_n^{1/4}\left(\overline{H}(1)^n + \overline{H}(2)^n\right) \xrightarrow{st} Z_\tau + \beta Z_\tau'$, where Z_τ' is described in Theorem 2.2.*

LEMMA B.4. *Under Assumptions 2–4, assuming $m_n\Delta_n^{3/4} \to \beta \in [0,\infty]$, we have*

$$z_n\overline{H}(l)^n \xrightarrow{P} 0 \text{ for } l = 3, 4, 5, \tag{B.19}$$

where $z_n = m_n^{1/2}\Delta_n^{1/4}$ if $m_n\Delta_n^{3/4} \to \beta < \infty$, and $z_n = \frac{1}{\sqrt{m_n\Delta_n}}$ if $m_n\Delta_n^{3/4} \to \beta = \infty$.

We prove the lemmas above in three steps. In Step 1, we start to prove the last lemma which is more straightforward than the other two. In Step 2, we prove Lemma B.2. In Step 3, we show Lemma B.3.

Step 1. For $l = 3$, set $\zeta_i^n = \frac{1}{\phi_{k_n}(g)}K_{m_n\Delta_n}(t_{i-1} - \tau)\left(\phi\left(Y\right)_i^n - \phi_{i,0}^n\right)$ and $b_n := m_n\Delta_n$. By Lemma 2.2.10 in Jacod and Protter (2011), the result follows if the array $z_n\mathbb{E}\left(|\zeta_i^n| \,|\, \mathcal{F}_{(i-1)\Delta_n}\right)$ is asymptotically negligible. To this end, note that (B.12) yields

$$\mathbb{E}\left(|\zeta_i^n| \,|\, \mathcal{F}_{(i-1)\Delta_n}\right) \leq C\Delta_n^{1/4}\Delta_n|K_{m_n\Delta_n}(t_{i-1} - \tau)|\mathbb{E}\left(\left(\Delta_n^{1/4} + \Gamma'(\mu)_i^n + \Gamma'(\widetilde{\sigma})_i^n + \Gamma'(\gamma)_i^n\right)\right),$$

where recall that we are assuming that $\tilde{\sigma}$, μ, and γ are càdlàg bounded processes by localization. Thus, from Lemma 16.5.15 in Jacod and Protter (2011), $\left((\Gamma'(\tilde{\sigma})_i^n)^2\right)$, $\left((\Gamma'(\mu)_i^n)^2\right)$, and $\left((\Gamma'(\gamma)_i^n)^2\right)$ satisfy (B.13). By the Cauchy–Schwarz inequality,

$$\Delta_n \sum_{j=1}^{n-k_n+1} |K_{b_n}(t_{i-1}-\tau)| \,\mathbb{E}\left(\Gamma'(\mu)_i^n\right) = \mathrm{o}\left((m_n\Delta_n)^{-1/2}\right). \tag{B.20}$$

We can obtain similar results on $(\Gamma'(\tilde{\sigma})_i^n)$ and $\Gamma'(\Upsilon)_i^n$. Thus, $z_n \sum_{j=1}^{n-k_n+1} \mathbb{E}\left(|\zeta_i^n| \mid \mathcal{F}_{(i-1)\Delta_n}\right)$ is $\mathrm{O}(z_n\Delta_n^{1/2}) + \mathrm{o}(z_n/(m_n^{1/2}\Delta_n^{1/4}))$ and, hence, converges to 0. This finishes the proof for $l=3$. For $l=4$, by (2.13), (2.17), and (B.16), we have

$$|\overline{H}(4)^n| \le C\frac{1}{k_n} \sum_{j=1}^{n-k_n+1} |K_{b_n}(t_{j-1}-\tau)|\left(\Delta_n + \Delta_n^{3/4}\left(\Gamma'(\gamma)_i^n + \Gamma'(\gamma')_i^n\right)\right),$$

which is $\mathrm{O}\left(\Delta_n^{1/2}\right) + \mathrm{o}\left(\frac{\Delta_n^{1/4}}{\sqrt{m_n\Delta_n}}\right)$ by similar argument as in (B.20). For $l=5$, we have

$$\left|\overline{H}(5)^n\right| \le \int_{t_{n-k_n+1}}^T \left|K_{b_n}(t-\tau)\sigma_t^2\right| dt + \sum_{j=1}^{n-k_n+1} \int_{t_{j-1}}^{t_j} \left|K_{b_n}(s-\tau)\sigma_s^2 - K_{b_n}(t_{j-1}-\tau)\sigma_{(j-1)\Delta_n}^2\right| ds$$

$$\le C\frac{1}{m_n\sqrt{\Delta_n}} + (n-k_n-1)\Delta_n\left(\frac{1}{m_n\Delta_n^{1/2}} + \frac{1}{m_n^2\Delta_n}\right) = \mathrm{O}_P\left(\frac{1}{m_n\sqrt{\Delta_n}}\right), \tag{B.21}$$

where the first term in (B.21) follows from the boundedness of K and σ, whereas the second term in (B.21) can be deduced by (B.17) and Lipschitz property of K. Indeed, for $s \in [t_{j-1}, t_j]$,

$$\left|K_{b_n}(s-\tau)\sigma_s^2 - K_{b_n}(t_{j-1}-\tau)\sigma_{(j-1)\Delta_n}^2\right|$$
$$\le \left|K_{b_n}(s-\tau)\sigma_s^2 - K_{b_n}(s-\tau)\sigma_{(j-1)\Delta_n}^2\right| + \left|K_{b_n}(s-\tau)\sigma_{(j-1)\Delta_n}^2 - K_{b_n}(t_{j-1}-\tau)\sigma_{(j-1)\Delta_n}^2\right|,$$

which is $\mathrm{O}_P\left(1/m_n\Delta_n^{1/2}\right) + \mathrm{O}_P\left(1/m_n^2\Delta_n\right)$. So, we deduce (B.19) for $l=5$.

Step 2. To show Lemma B.2, we need several preliminary lemmas. We employ the "block splitting" method proposed in Jacod and Protter (2011) (see Section 16.5.4 and page 548 therein). Recall that

$$\overline{H}(1)^n = \sum_{i=1}^{n-k_n+1} \zeta_i^n,$$

where $\zeta_i^n = \frac{1}{\phi_{k_n}(g)} K_{m_n\Delta_n}(t_{i-1}-\tau)\left(\phi_{i,0}^n - \mathbb{E}\left(\phi_{i,0}^n \mid \mathcal{F}_{(i-1)\Delta_n}\right)\right)$. The variables ζ_i^n are not martingale differences. To use martingale methods, we fix an integer $m \ge 1$, and divide the summands in the definition of $\overline{H}(1)^n$ into blocks of size mk_n and k_n. Concretely, the ℓth big block, of size mk_n, contains the indices between $I(m,n,\ell) = (\ell-1)(m+1)k_n+1$ and $I(m,n,\ell)+mk_n-1$. The number of such blocks before time t is $l_n(m) = \left[\frac{n-k_n+1}{(m+1)k_n}\right]$. These big blocks are separated by small blocks of size k_n, and the "real" time corresponding to

the beginning of the ℓth big block is $t(m, n, \ell) = (I(m, n, \ell) - 1)\Delta_n$. Then we introduce the summand over all the big blocks,

$$Z^n(m) := \sum_{\ell=1}^{l_n(m)} \delta(m)_\ell^n := \sum_{\ell=1}^{l_n(m)} \sum_{r=0}^{mk_n-1} \zeta_{I(m,n,\ell)+r}^n. \tag{B.22}$$

Note that the sequence $(\delta(m)_\ell^n)$ are now martingale differences w.r.t. the discrete filtration $\mathcal{G}_\ell = \mathcal{F}_{(I(m,n,\ell+1)-1)\Delta_n}$, for $\ell = 1, \dots, l_n(m)$.

We now show that the contribution of the small blocks, i.e., $\overline{H}(1)^n - Z^n(m)$, is asymptotically "negligible" compared to $m_n^{-1/2}\Delta_n^{-1/4}$.

LEMMA B.5. *Under Assumptions 2–4,* $\lim_{m\to\infty} \limsup_{n\to\infty} \mathbb{E}\left(m_n^{1/2}\Delta_n^{1/4}|\overline{H}(1)^n - Z^n(m)|\right) = 0.$

Proof. Denote by $J(n, m)$ the set of all integers j between 1 and $n - k_n + 1$, which are *not* in the big blocks (i.e., those corresponding to the small blocks). We further divide $J(n, m)$ into k_n disjoint subsets $J(n, m, r)$ for $r = 1, \dots, k_n$, where $J(n, m, r)$ is the set of all $j \in J(n, m)$ equal to r modulo k_n. Then, $\overline{H}(1)^n - Z^n(m) = \sum_{r=1}^{k_n} \sum_{j \in J(n,m,r)} \zeta_j^n$. Observe that $\mathbb{E}\left(\zeta_j^n \,\middle|\, \mathcal{F}_{(j-1)\Delta_n}\right) = 0$ and ζ_j^n is $\mathcal{F}_{(j+k_n)\Delta_n}$ measurable. Then $\sum_{j \in J(n,m,r)} \zeta_j^n$ is the sum of martingale increments, because any two distinct indices in $J(n, m, r)$ are more than k_n apart. Therefore, by (B.14) and the fact that $\mathbb{E}\left(\zeta_j^n \,\middle|\, \mathcal{F}_{(j-1)\Delta_n}\right) = 0$, for some constant C (changing from line to line) and large enough n,

$$\mathbb{E}\left|\sum_{j \in J(n,m,r)} \zeta_j^n\right|^2 \leq C \sum_{j \in J(n,m,r)} \frac{\Delta_n}{\phi_{k_n}^2(g)} K_{b_n}^2(t_{j-1}-\tau) \leq C\frac{1}{(m+1)k_n^3 m_n \Delta_n} \int K^2(u)du \leq C\frac{\Delta_n^{1/2}}{mm_n},$$

where the last inequality holds because of (2.13) and the second inequality holds because, recalling that two consecutive j's in $J(n, m, r)$ are separated by $(m + 1)k_n$, $(m + 1)k_n m_n \Delta_n^2 \sum_{j \in J(n,m,r)} K_{b_n}^2(t_{j-1} - \tau)$ converges to $\int K^2(u)du$. Then, $\mathbb{E}\left(m_n^{1/2}\Delta_n^{1/4}|\overline{H}(1)^n - Z^n(m)|\right)$ is $O\left(1/\sqrt{m}\right)$, for large enough n. As $m \to \infty$, the above quantity goes to 0 and we get the result. $\square$

Next, we modify the "big-blocks" process $Z^n(m)$ defined in (B.22) in such a way that each summand involves the volatility at the beginning of the corresponding large block. Recalling the notation in (B.11), we set

$$\eta_{i,r}^n = \frac{1}{\phi_{k_n}(g)} K_{m_n \Delta_n}\left(t_{i-1} - \tau\right)\left(\phi_{i,r}^n - \mathbb{E}\left(\phi_{i,r}^n \,\middle|\, \mathcal{F}_{(i-r-1)\Delta_n}\right)\right), \tag{B.23}$$

$$\eta_{i,r}'^n = \frac{1}{\phi_{k_n}(g)} K_{m_n \Delta_n}\left(t_{i-1} - \tau\right)\left(\mathbb{E}\left(\phi_{i,r}^n \,\middle|\, \mathcal{F}_{(i-r-1)\Delta_n}\right) - \mathbb{E}\left(\phi_{i,r}^n \,\middle|\, \mathcal{F}_{(i-1)\Delta_n}\right)\right), \tag{B.24}$$

$$M^n(m) = \sum_{i=1}^{l_n(m)} \sum_{r=0}^{mk_n-1} \eta_{I(m,n,i)+r,r}^n, \quad M'^n(m) = \sum_{i=1}^{l_n(m)} \sum_{r=0}^{mk_n-1} \eta_{I(m,n,i)+r,r}'^n. \tag{B.25}$$

Lemma B.6. *Under Assumptions 2–4, for a fixed m,*

$$\lim_{n\to\infty} \mathbb{E}\left(m_n^{1/2} \Delta_n^{1/4} \left|Z^n(m) - M^n(m) - M'^n(m)\right|\right) = 0.$$

Proof. We use a similar method as in the previous lemma: Let $J'(n,m)$ the set of all integers j between 1 and $n - k_n + 1$, which are *inside* the big blocks, that is of the form $j = I(m,n,i) + l$ for some $i \geq 1$ and $l \in \{0, \ldots, mk_n - 1\}$. Let $J'(n,m,r)$ be the set of all $j \in J'(n,m)$ equal to r modulo k_n. We can then write $Z^n(m) - M^n(m) - M'^n(m) = \sum_{r=1}^{k_n} \sum_{j \in J'(n,m,r)} \theta_j^n$, where $\theta_j^n = \frac{1}{\phi_{k_n}(g)} K_{m_n \Delta_n}(t_{j-1} - \tau)\left(\phi_{j,0}^n - \phi_{j,l}^n - \mathbb{E}\left(\phi_{j,0}^n - \phi_{j,l}^n \big| \mathcal{F}_{(j-1)\Delta_n}\right)\right)$, when $j = I(m,n,i) + l$. Note $\phi_{j,0}$ and $\phi_{j,l}$ have the same noise part, $-\frac{1}{2}\hat{\epsilon}_j^n$, and the cross term $\overline{W}_j^n \overline{\epsilon}_j^n$ has expectation 0. Then, for some constant C and large enough n,

$$\mathbb{E}\left|\theta_j^n\right|^2 \leq \frac{1}{\phi_{k_n}^2(g)} K_{m_n \Delta_n}^2(t_{j-1} - \tau) \mathbb{E}\left|\phi_{j,0}^n - \phi_{j,l}^n\right|^2 \leq C K_{m_n \Delta_n}^2(t_{j-1} - \tau) m k_n \Delta_n^3,$$

for $j \in J'(n,m,r)$, where the last inequality follows by conditioning on $\mathcal{F}_{(j-1)\Delta_n}$, using that $\mathbb{E}\left[(\overline{W}_j^n)^4 | \mathcal{F}_{(j-1)\Delta_n}\right] = 3\phi_{k_n}(g)^2 \Delta_n^2$, and applying (B.17). As in the proof of the previous lemma,

$$\mathbb{E}\left|\sum_{j \in J'(n,m,r)} \theta_j^n\right|^2 \leq C\Delta_n^3 k_n \sum_{j \in J'(n,m,r)} K_{m_n \Delta_n}^2(t_{j-1} - \tau) \leq C\frac{\Delta_n}{m_n} \int K^2(u) du.$$

So we have $\mathbb{E}\left(m_n^{1/2} \Delta_n^{1/4} \left|Z^n(m) - M^n(m) - M'^n(m)\right|\right)$ is less than $C m_n^{1/2} \Delta_n^{1/4} k_n \sqrt{\frac{\Delta_n}{m_n}}$ and, hence, it converges to 0. $\square$

Now, we prove $M'^n(m)$, defined in (B.25), is asymptotically negligible.

Lemma B.7. *Under Assumptions 2–4,* $\lim_{n\to\infty} \mathbb{E}\left(m_n^{1/2} \Delta_n^{1/4} \left|M'^n(m)\right|\right) = 0.$

Proof. Recall that $b_n = m_n \Delta_n$ and $\Psi_{i,j}^n = \mathbb{E}(\phi_{i,j}^n | \mathcal{H}_{(i-1)\Delta_n}) - (\sigma_{(i-j-1)\Delta_n} \overline{W}_i^n)^2$ and, since $\mathcal{H}_t = \mathcal{F}^{(0)} \otimes \sigma\left(\epsilon_s : s \in [0,t)\right)$,

$$\mathbb{E}\left(\Psi_{i+r,r}^n \big| \mathcal{F}_{(i-1)\Delta_n}\right) = \mathbb{E}\left(\phi_{i+r,r}^n \big| \mathcal{F}_{(i-1)\Delta_n}\right) - \mathbb{E}\left((\sigma_{(i-1)\Delta_n} \overline{W}_{i+r}^n)^2 \big| \mathcal{F}_{(i-1)\Delta_n}\right),$$

$$\mathbb{E}\left(\Psi_{i+r,r}^n \big| \mathcal{F}_{(i+r-1)\Delta_n}\right) = \mathbb{E}\left(\phi_{i+r,r}^n \big| \mathcal{F}_{(i+r-1)\Delta_n}\right) - \mathbb{E}\left((\sigma_{(i-1)\Delta_n} \overline{W}_{i+r}^n)^2 \big| \mathcal{F}_{(i+r-1)\Delta_n}\right).$$

Since $\overline{W}_{i+r}^n$ is a linear combination of $W_{(i+r)\Delta_n}, \ldots, W_{(i+r+k_n-1)\Delta_n}$, we have

$$\eta_{i+r,r}'^n = \frac{1}{\phi_{k_n}(g)} K_{b_n}(t_{i+r-1} - \tau)\left(\mathbb{E}\left(\phi_{i+r,r}^n \big| \mathcal{F}_{(i-1)\Delta_n}\right) - \mathbb{E}\left(\phi_{i+r,r}^n \big| \mathcal{F}_{(i+r-1)\Delta_n}\right)\right)$$

$$= \frac{1}{\phi_{k_n}(g)} K_{b_n}(t_{i+r-1} - \tau)\left(\mathbb{E}\left(\Psi_{i+r,r}^n \big| \mathcal{F}_{(i-1)\Delta_n}\right) - \mathbb{E}\left(\Psi_{i+r,r}^n \big| \mathcal{F}_{(i+r-1)\Delta_n}\right)\right).$$

Next, note that, by (B.16), we have

$$\mathbb{E}\left|\mathbb{E}\left(\Psi_{i+r,r}^n\middle|\mathcal{F}_{(i-1)\Delta_n}\right) - \mathbb{E}\left(\Psi_{i+r,r}^n\middle|\mathcal{F}_{(i+r-1)\Delta_n}\right)\right|^2$$
$$\leq \mathbb{E}\left(\mathbb{E}\left(\Psi_{i+r,r}^n\middle|\mathcal{F}_{(i+r-1)\Delta_n}\right)^2\right) \leq \mathbb{E}\left(\mathbb{E}\left(\left(\Psi_{i+r,r}^n\right)^2\middle|\mathcal{F}_{(i+r-1)\Delta_n}\right)\right) \leq C\Delta_n^{3/2}.$$

We can then deduce that, for $r \neq l$,

$$\mathbb{E}\left(\eta_{i+r,r}''^n \eta_{i+l,l}''^n\right) \leq C\frac{1}{\phi_{k_n}(g)^2}|K_{b_n}(t_{i+r-1}-\tau)||K_{b_n}(t_{i+l-1}-\tau)|\Delta_n^{3/2}.$$

Therefore, denoting for simplicity $I_i = I(m,n,i) = (i-1)(m+1)k_n + 1$,

$$\mathbb{E}\left|\sum_{r=0}^{mk_n-1}\eta_{I(m,n,i)+r,r}''^n\right|^2 \leq C\Delta_n^{1/2}\left(\int_{t_{I_i-1}}^{t_{I_i+(mk_n-1)\Delta_n}}|K_{b_n}(s-\tau)|ds\right)^2.$$

The result is then proved since $m_n^{1/2}\Delta_n^{1/4}\mathbb{E}\left|M'^n(m)\right| \leq Cm_n^{1/2}\Delta_n^{1/2}\int|K(u)|du \to 0$. $\quad\square$

At this stage, we are ready to prove a CLT for the processes $M_n(m)$, for each fixed m. We follow the arguments of Jacod and Protter (2011, p. 550). For completeness, we outline them here. Let

$$L(g)_t = \int_t^{t+1}g(u-t)dW_u^1, \quad L'(g)_t = \int_t^{t+1}g'(u-t)dW_u^2, \tag{B.26}$$

where W^1 and W^2 are two independent one-dimensional BMs defined on an auxiliary space $\left(\tilde{\Omega},\tilde{\mathcal{F}},\left(\tilde{\mathcal{F}}_t\right)_{t\geq 0},\tilde{\mathbb{P}}\right)$. The processes $L(g)$ and $L'(g)$ are independent, stationary, centered, and Gaussian with covariances $\mathbb{E}\left(L(g)_t L(g)_s\right) = \int_{t\vee s}^{(t+1)\wedge(s+1)}g(u-t)g(u-s)du$, and $\mathbb{E}\left(L'(g)_t L'(g)_s\right) = \int_{t\vee s}^{(t+1)\wedge(s+1)}g'(u-t)g'(u-s)du$. Next, denoting $\tilde{\mathbb{E}}$ the expectation with respect to $\tilde{\mathbb{P}}$, let

$$\mu\left(v,v'\right) = \tilde{\mathbb{E}}\left(\left(vL(g)_s + v'L'(g)_s\right)^2 - v'^2\phi(g')\right),$$
$$\mu'\left(v,v';s,s'\right) = \tilde{\mathbb{E}}\left(\left(\left(vL(g)_s + v'L'(g)_s\right)^2 - v'^2\phi(g')\right)\left(\left(vL(g)_{s'} + v'L'(g)_{s'}\right)^2 - v'^2\phi(g')\right)\right),$$
$$R\left(v,v'\right) = \int_0^2\left(\mu'\left(v,v';1,s\right) - \mu\left(v,v'\right)\mu\left(v,v'\right)\right)ds.$$

As argued in the proof of Theorem 7.20 in Aït-Sahalia and Jacod (2014), one can show that $\frac{1}{\theta}R(\sigma_t, \theta v_t)$ equals $4\left(\Phi_{22}\sigma_t^4/\theta + 2\Phi_{12}\sigma_t^2\gamma_t\theta + \Phi_{11}\gamma_t^2\theta^3\right)$, where $v_t = \sqrt{\gamma_t}$ is the conditional standard deviation for ϵ_t. For a fixed m and $t \in [0,T]$, let $\gamma(m)_t = m\mu(\sigma_t, \theta v_t)$, and $\gamma'(m)_t = \int_0^m ds\int_0^m ds'\mu'(\sigma_t, \theta v_t; s, s')$.

LEMMA B.8. *Under Assumptions 2–4, for each $m \geq 1$, as $n \to \infty$, the process $m_n^{1/2}\Delta_n^{1/4}M^n(m)$ converges in law to an random variable (r.v.) $\overline{Y}(m)$, which conditionally on $\mathcal{F}$ is a centered Gaussian r.v. with variance*

$$\mathbb{E}\left(\left(\overline{Y}(m)\right)^2\middle|\mathcal{F}\right) = \frac{1}{m+1}\frac{1}{\theta}\left(\gamma'(m)_\tau - \gamma(m)_\tau^2\right)\int K^2(u)du.$$

Proof. For $i = 1, \ldots, l_n(m)$, let $\eta(m)_i^n := m_n^{1/2} \Delta_n^{1/4} \sum_{r=0}^{mk_n-1} \eta_{I(m,n,i)+r,r}^n$, $\mathcal{G}_i^n = \mathcal{F}_{(I(m,n,i+1)-1)\Delta_n}$, $\mathbb{E}_{i-1}[\cdot] = \mathbb{E}[\cdot|\mathcal{G}_{i-1}^n]$, $b_n = m_n \Delta_n$, and $\phi_n = \phi_{k_n}(g)$. For simplicity, we write $I_i = I(m,n,i)$. Note that $\eta(m)_i^n$ is $\mathcal{G}_i^n$-measurable and, furthermore, $\mathbb{E}[\eta(m)_i^n|\mathcal{G}_{i-1}^n] = 0$. We will apply Theorem 2.2.15 in Jacod and Protter (2011) to the martingale increments $\eta(m)_i^n, i = 1, \ldots, l_n(m)$. By the Jensen-type inequality and (B.15), we have, for each fixed m,

$$
\begin{aligned}
\sum_{i=1}^{l_n(m)} \mathbb{E}_{i-1} \left| \eta(m)_i^n \right|^4 &\leq C \sum_{i=1}^{l_n(m)} m_n^2 \Delta_n^3 \left(\sum_{r=0}^{mk_n-1} \frac{1}{|\phi_n|} |K_{b_n}(t_{I_i+r-1} - \tau)| \right)^4 \\
&\leq C \sum_{i=1}^{l_n(m)} m^4 m_n^2 \Delta_n^3 \frac{1}{mk_n \Delta_n} \int_{t_{I_i-1}}^{t_{I_i-1}+mk_n\Delta_n} K_{b_n}^4(s-\tau) ds \to 0.
\end{aligned}
\tag{B.27}
$$

Therefore, for every $\varepsilon > 0$, $\sum_{i=1}^{l_n(m)} \mathbb{E}_{i-1}\left(|\eta(m)_i^n|^2 \mathbb{1}_{|\eta(m)_i^n|^2 \geq \varepsilon} \right) \leq \frac{1}{\epsilon} \sum_{i=1}^{l_n(m)} \mathbb{E}_{i-1} \left(|\eta(m)_i^n|^4 \right) \stackrel{n \to \infty}{\longrightarrow} 0$. It remains to prove that, for a fixed m,

$$
S_n := \sum_{i=1}^{l_n(m)} \mathbb{E}_{i-1} \left((\eta(m)_i^n)^2 \right) \stackrel{\mathbb{P}}{\longrightarrow} \frac{1}{(m+1)\phi^2(g)} \frac{1}{\theta} \int K^2(u) du \left(\gamma'(m)_\tau - \gamma(m)_\tau^2 \right),
\tag{B.28}
$$

and, for any bounded $\mathcal{F}_t$-martingale N that is orthogonal to W, or for $N = W$,

$$
\sum_{i=1}^{l_n(m)} \mathbb{E} \left(\eta(m)_i^n \left(N_{(I_{i+1}-1)\Delta_n} - N_{(I_i-1)\Delta_n} \right) \Big| G_{i-1}^n \right) \stackrel{\mathbb{P}}{\to} 0.
\tag{B.29}
$$

We start by proving (B.28). Let

$$
\begin{aligned}
\alpha_i^n &:= \frac{1}{k_n^2 \Delta_n} \sum_{r=0}^{mk_n-1} K_{b_n}(t_{I_i+r-1} - \tau) \phi_{I_i+r,r}^n \\
&= \frac{1}{k_n^2 \Delta_n} K_{b_n}(t_{I_i-1} - \tau) \sum_{r=0}^{mk_n-1} \phi_{I_i+r,r}^n + O_P\left(\frac{1}{m_n^2 \Delta_n^{3/2}} \right).
\end{aligned}
\tag{B.30}
$$

For $(I(m,n,i) - 1)\Delta_n \leq s < (I(n,m,i+1) - 1)\Delta_n$, set

$$
\gamma_s^n = \mathbb{E}_{i-1} \left(\frac{1}{k_n^2 \Delta_n} \sum_{r=0}^{mk_n-1} \phi_{I_i+r,r}^n \right), \quad \gamma_s'^n = \mathbb{E}_{i-1} \left(\left(\frac{1}{k_n^2 \Delta_n} \sum_{r=0}^{mk_n-1} \phi_{I_i+r,r}^n \right)^2 \right).
$$

Then, we have

$$
S_n = m_n \Delta_n^{1/2} \frac{k_n^4 \Delta_n^2}{\phi_{k_n}^2(g)} \sum_{i=1}^{l_n(m)} K_{m_n\Delta_n}^2(t_{I_i-1} - \tau) \left(\gamma_{t_{I_i-1}}'^n - \left(\gamma_{t_{I_i-1}}^n \right)^2 \right) + O_P\left(\frac{1}{m_n\sqrt{\Delta_n}} \right).
$$

If we can show that for any $s \in [0, T]$,

$$
\gamma_s^n \stackrel{\mathbb{P}}{\longrightarrow} \gamma(m)_s, \quad \gamma_s'^n \stackrel{\mathbb{P}}{\longrightarrow} \gamma'(m)_s,
\tag{B.31}
$$

we can obtain (B.28); indeed,

$$S_n = \frac{1}{(m+1)k_n\Delta_n} m_n \Delta_n^{1/2} k_n^2 \Delta_n^2 \int_0^T K_{m_n\Delta_n}^2 (s-\tau) \left(\gamma'(m)_s - \gamma(m)_s^2 \right) ds + o_P(1)$$

$$= \frac{1}{\theta(m+1)} \int_{\frac{-\tau}{m_n\Delta_n}}^{\frac{T-\tau}{m_n\Delta_n}} K^2(u) \left(\gamma'(m)_{\tau+um_n\Delta_n} - \gamma(m)_{\tau+um_n\Delta_n}^2 \right) du + o_P(1)$$

$$\xrightarrow{\mathbb{P}} \frac{1}{\theta(m+1)} \int K^2(u) du \left(\gamma'(m)_\tau - \gamma(m)_\tau^2 \right).$$

To show (B.31), we fix $s \in [0, T]$ and apply Lemma 16.3.9 in Jacod and Protter (2011) with the sequence $i_n = I(m, n, i)$, $T_n = (I(m, n, i) - 1)\Delta_n$ if $I(m, n, i-1)\Delta_n \le s < I(m, n, i)\Delta_n$. Concretely, with the notation $L_u^n = \frac{1}{\sqrt{k_n\Delta_n}}\overline{W}_{i_n+[k_nu]}^n$, $L_u'^n = \sqrt{k_n}\overline{\epsilon}_{i_n+[k_nu]}^n$, $\widehat{L}_u^n = k_n\widehat{\epsilon}_{i_n+[k_nu]}^n$, for $u \in [0, m]$, we have

$$\frac{1}{k_n^2\Delta_n} \sum_{r=0}^{mk_n-1} \phi_{i+r,r}^n = F_n \left(\sigma_{T_n} L^n, L'^n, \widehat{L}^n \right),$$

where F_n is the function on $\mathbb{D} \times \mathbb{D} \times \mathbb{D}$ (here $\mathbb{D} = \mathbb{D}\left([0, m] : \mathbb{R}^1\right)$ is the Skorokhod space), defined as

$$F_n(x, y, z) = \frac{1}{k_n} \sum_{r=0}^{mk_n-1} \left(\left(x\left(\frac{r}{k_n}\right) + \frac{1}{\sqrt{k_n^2\Delta_n}} y\left(\frac{r}{k_n}\right) \right)^2 - \frac{1}{2k_n^2\Delta_n} z\left(\frac{r}{k_n}\right) \right). \tag{B.32}$$

Note that the functions F_n, F_n^2 converge pointwise to F, F^2, respectively, where

$$F(x, y, z) = \int_0^m \left\{ (x(s) + \theta y(s))^2 - \frac{1}{2}\theta^2 z(s) \right\} ds.$$

Now, we deduce from Lemma 16.3.9 in Jacod and Protter (2011) that with $Z = 1$, $\phi(f) = \int_0^1 f^2(u) du$ and the notation from (B.26):[12]

$$\mathbb{E}\left(F_n \left(\sigma_{T_n} L^n, L'^n, \widehat{L}^n \right) \big| \mathcal{G}_{(i-1)} \right) \xrightarrow{\mathbb{P}} \mathbb{E}\left(F\left(\sigma_s L, v_s L', 2\phi(g')\gamma_s \right) \right) = \gamma(m)_s.$$

Similarly,

$$\mathbb{E}\left(F_n^2 \left(\sigma_{T_n} L^n, L'^n, \widehat{L}^n \right) \big| \mathcal{G}_{(i-1)} \right) \xrightarrow{\mathbb{P}} \gamma(m)_s',$$

and we conclude (B.31). This finishes the proof for (B.28). Now, we show (B.29). Let

$$\zeta_i^n = \frac{m_n^{1/2}\Delta_n^{1/4}}{\phi_{k_n}(g)} \sum_{r=0}^{mk_n-1} K_{m_n\Delta_n} \left(t_{I_i+r,r} - \tau \right) \phi_{I_i+r,r}^n,$$

---

[12] Below, we assume that the space $\left(\tilde{\Omega}, \tilde{\mathcal{F}}, \tilde{\mathbb{P}} \right)$, where W^1 and W^2 (hence, L and L') are defined, is an extension of the space $(\Omega, \mathcal{F}, \mathbb{P})$ and that W^1 and W^2 are independent of X and ϵ.

and set $D_i^n(N) = N_{(I_{i+1}-1)\Delta_n} - N_{(I_i-1)\Delta_n}$. Since $\mathbb{E}\left(D(N)_i^n | \mathcal{G}_{i-1}^n\right) = 0$, we only need to prove that, for any bounded martingale N,

$$\sum_{i=1}^{l_n(m)} \mathbb{E}\left(\zeta_i^n D_i^n(N) \Big| G_{i-1}^n\right) \xrightarrow{\mathbb{P}} 0. \tag{B.33}$$

Following the same argument of (B.27) and inequality (B.15), we have

$$\sum_{i=1}^{l_n(m)} \mathbb{E}\left(\zeta_i^n\right)^2 = \frac{m_n \Delta_n^{1/2}}{\phi_{k_n}^2(g)} \sum_{i=1}^{l_n(m)} \mathbb{E}_{i-1}\left(\left(\sum_{r=0}^{mk_n-1} K_{m_n\Delta_n}\left(t_{I_i+r-1} - \tau\right) |\phi_{I_i+r,r}^n|\right)^2\right)$$

$$\leq C \sum_{i=1}^{l_n(m)} m^2 m_n \Delta_n^{3/2} \frac{1}{mk_n\Delta_n} \int_{t_{I_i-1}}^{t_{I_i-1}+mk_n\Delta_n} K_{m_n\Delta_n}^2(s-\tau)\,ds = \mathrm{O}(1). \tag{B.34}$$

If N is a square-integrable martingale, the Cauchy–Schwarz inequality yields

$$\left(\sum_{i=1}^{l_n(m)} \mathbb{E}_{i-1}\left(\zeta_i^n D_i^n(N)\right)\right)^2 \leq \left(\sum_{i=1}^{l_n(m)} \mathbb{E}\left(\zeta_i^n\right)^2\right)\left(\sum_{i=1}^{l_n(m)} \mathbb{E}\left(D_i^n(N)\right)^2\right) \leq C\sqrt{\mathbb{E}N_T^2}.$$

Note with notation (B.11) and

$$\zeta_i'^n = \frac{m_n^{1/2}\Delta_n^{1/4}}{\phi_{k_n}(g)} \sum_{r=0}^{mk_n-1} K_{m_n\Delta_n}\left(t_{I_i+r,r} - \tau\right) \Psi_{I_i+r,r}^n,$$

the same argument also yields $\mathbb{E}_{i-1}\left(\zeta_i'^n D_i^n(N)\right) \leq C\Delta_n^{1/4}\sqrt{\mathbb{E}N_T^2}$. As shown on page 552 of Jacod and Protter (2011), we just need to prove this for $N \in \mathcal{N}^{(i)}, i = 0, 1$, where $\mathcal{N}^{(0)}$ is the set of all bounded $(\mathcal{F}_t^{(0)})$-martingales orthogonal to W and $\mathcal{N}^{(1)}$ is the set of all martingales having $N_\infty = h\left(\chi_{t_1}, \ldots, \chi_{t_w}\right)$, where h is a Borel bounded function on $\mathbb{R}^w$ and $t_1 < \cdots < t_w$ and $w \geq 1$. When N is either W or in $\mathcal{N}^{(0)}$, $D(N)_i^n$ is $\mathcal{H}_\infty$ measurable. Therefore, $\mathbb{E}_{i-1}\left(\zeta_{I_i}^n D(N)_i^n\right)$ is equal to

$$\mathbb{E}_{i-1}\left(\zeta_{I(m,n,i)}'^n D(N)_i^n\right) + \frac{m_n^{1/2}\Delta_n^{1/4}}{\phi_{k_n}(g)} \mathbb{E}_{i-1}\left(\sum_{r=0}^{mk_n-1} K_{b_n}\left(t_{I_i+r,r} - \tau\right)\left(\sigma_{(I_i-1)\Delta_n}\overline{W}_{I_i+r}^n\right)^2 D(N)_i^n\right).$$

The second term vanishes when $N = W$ since it is the $\mathcal{F}_{(I_i-1)\Delta_n}$-conditional expectation of an odd function of the increments of the process W after time $(I_i - 1)\Delta_n$. Suppose now that N is a bounded martingale, orthogonal to W. By Itô's formula, we see that $(\overline{W}_j^n)^2$ is the sum of a constant (depending on n) and of a martingale which is a stochastic integral with respect to W, B on the interval $[(j-1)\Delta_n, (j+k_n-1)\Delta_n]$. Then, the orthogonality of N and W implies this second term above vanishes as well. So, we have that $\mathbb{E}\left(\zeta_{I_i}^n D(N)_i^n | \mathcal{G}_{i-1}^n\right) \leq C\Delta_n^{1/4}\sqrt{\mathbb{E}N_T^2}$. When $N \in \mathcal{N}^{(1)}$ is associated with h and w and the t_i's, the same argument in Jacod and Protter (2011) and the inequality $\mathbb{E}\left(\zeta_i^n\right)^2 \leq C\frac{1}{m_n\sqrt{\Delta_n}}$

deduced from (B.34) yield

$$\mathbb{E}\left(\sum_{i=1}^{l_n(m)} \left| \mathbb{E}_{i-1}\left(\zeta_{I_i}^n D(N)_i^n\right)\right|\right) \leq Cw\left(\Delta_n^{1/4} + \frac{1}{m_n\sqrt{\Delta_n}}\right),$$

and (B.29) is shown. This finishes the proof of Lemma B.8.   □

The only thing left to prove Lemma B.2 is the stable convergence in law $\overline{Y}(m) \xrightarrow{st} Z_\tau$, as $m \to \infty$. For this, we only need to show that, as $m \to \infty$, $\frac{1}{m+1}\left(\gamma'(m)_\tau - \gamma(m)_\tau^2\right) \xrightarrow{st} R(\sigma_\tau, \theta v_\tau)$. Recall that the process (L, L') is stationary, and the variables (L_t, L_t') and (L_s, L_s') are independent if $|s - t| \geq 1$. So, $\mu'\left(v, v'; s, s'\right) = \left(\mu\left(v, v'\right)\right)^2$ when $|s - s'| \geq 1$ and $\mu'\left(v, v'; s, s'\right) = \mu'\left(v, v'; 1, s' + 1 - s\right)$, for all $s, s' \geq 0$ with $s' + 1 - s \geq 0$. Then, if $m \geq 2$ and letting $\mu = \mu\left(\sigma_\tau, \theta v_\tau\right)$ and $\mu'\left(s, s'\right) = \mu'\left(\sigma_\tau, \theta v_\tau; s, s'\right)$, we have

$$\frac{1}{m+1}\left(\gamma'(m)_\tau - \gamma(m)_\tau \gamma(m)_\tau\right) = \frac{1}{m+1}\int_0^m ds \int_0^m \mu'\left(s, s'\right) ds' - m^2 \mu^2$$

$$= \frac{m-1}{m+1}\int_0^2 \left(\mu'\left(1, s'\right) - \mu^2\right) ds' + \frac{1}{m+1}\int_0^1 ds \int_{1-s}^2 \left(\mu'\left(1, s' + 1 - s\right) - \mu^2\right) ds',$$

which converges to $R(\sigma_\tau, \theta v_\tau)$, since μ, μ' are bounded. This finishes the proof of Lemma B.2.

Step 3. We now show Lemma B.3.

Proof of Lemma B.3. Let $b_n = m_n \Delta_n$ and $t(i) = (I(m, n, i) - 1)\Delta_n$, where the notation for $I(m, n, i)$ can be found after Step 2 above. From the proof of Theorem 6.2 in Figueroa-López and Li (2020a) and recalling we have bounded jumps, we have

$$b_n^{-1/2}\int_0^T K_{b_n}(t-\tau)\left(\sigma_t^2 - \sigma_\tau^2\right) dt = b_n^{-1/2}\Lambda_{\tau-\sqrt{b_n}}\int_{\tau-\sqrt{b_n}}^T L\left(\frac{t-\tau}{b_n}\right) dB_t + o_P(1),$$

where $L(t) = \int_t^\infty K(u) du \mathbb{1}_{\{t>0\}} - \int_{-\infty}^t K(u) du \mathbb{1}_{\{t\leq 0\}}$. Furthermore, we have

$$b_n^{-1/2}\int_{(0,T)^c} K_{b_n}(t-\tau) dt = \frac{1}{\sqrt{b_n}}\left(\int_{-\infty}^{\frac{-\tau}{b_n}} K(u) du + \int_{\frac{T-\tau}{b_n}}^\infty K(u) du\right) \to 0, \text{ as } n \to \infty,$$

since Assumption 2 implies that $x^{1/2}\int_x^\infty K(u) du \to 0$, as $x \to \infty$. So, for a fixed m, we can rewrite $m_n^{1/2}\Delta_n^{1/4}\overline{H}(2)^n$ as

$$\beta b_n^{-1/2}\Lambda_{\tau-\sqrt{b_n}}\sum_{i:t(i)>\tau-\sqrt{b_n}}^{l_n(m)}\int_{t(i)}^{t(i)+(m+1)k_n\Delta_n} L\left(\frac{t-\tau}{b_n}\right) dB_t + o_P(1) =: \sum_{i=1}^{l_n(m)} \alpha(m)_i^n + o_P(1),$$

$$\text{(B.35)}$$

with $\alpha(m)_i^n = 0$ if i is such that $t(i) \leq \tau - \sqrt{b_n}$. Combining with the proof of Lemma B.2, we can deduce the following lemma.

LEMMA B.9. *Under Assumptions 2 and 3 and (2.2), with $m_n \to \infty$ and $m_n \Delta_n^{3/4} \to \beta \in (0, \infty)$,*

$$\lim_{m \to \infty} \limsup_{n \to \infty} m_n^{1/2} \Delta_n^{1/4} \mathbb{E} \left| \overline{H}(1)^n + \overline{H}(2)^n - \sum_{i=1}^{l_n(m)} \zeta(m)_i^n - \sum_{i=1}^{l_n(m)} \alpha(m)_i^n \right| = 0,$$

where $\zeta(m)_i^n := m_n^{1/2} \Delta_n^{1/4} \frac{1}{\phi_{k_n}(g)} K_{b_n} (t(i) - \tau) \sum_{r=1}^{mk_n - 1} \phi_{I_i+r,r}^n$ with notation (B.11).   □

Now Lemma B.3 follows if we apply Theorem 2.2.15 in Jacod and Protter (2011) to the sum of martingale differences $(\zeta(m)_i^n + \alpha(m)_i^n)$ and the filtration $\mathcal{G}_i = \mathcal{F}_{(I_{i+1}-1)\Delta_n}$, and show that $\sum_{i=1}^{l_n(m)} (\zeta(m)_i^n + \alpha(m)_i^n) \xrightarrow{st} Z_\tau + \beta Z_\tau'$. To this end, we first need to show, for a fixed m,

$$\sum_{i=1}^{l_n(m)} \mathbb{E}_{i-1} \left((\zeta(m)_i^n)^2 \right) \to \frac{1}{m+1} \frac{1}{\theta} \left(\gamma'(m)_\tau - \gamma(m)_\tau^2 \right) \int K^2(u) du, \tag{B.36}$$

$$\sum_{i=1}^{l_n(m)} \mathbb{E}_{i-1} \left((\alpha(m)_i^n)^2 \right) \to \beta^2 \Lambda_\tau^2 \int L^2(u) du, \tag{B.37}$$

$$\sum_{i=1}^{l_n(m)} \mathbb{E}_{i-1} \left((\zeta(m)_i^n \alpha(m)_i^n) \right) \to 0. \tag{B.38}$$

The proof of (B.36) can be found in the proof of Lemma B.2. (B.37) can be directly derived from the definition (B.35). So, we only need to show (B.38). With the notation (B.11), we have

$$\mathbb{E}_{i-1} \left(\left(\sum_{r=0}^{mk_n-1} \phi_{I_i+r,r}^n \right) \int_{t(i)}^{t(i)+(m+1)k_n\Delta_n} L \left(\frac{t-\tau}{b_n} \right) dB_t \right)$$

$$= \sigma_{t(i)}^2 \mathbb{E}_{i-1} \left(\left(\sum_{r=0}^{mk_n-1} \left(\overline{W}_{t(i)+r}^n \right)^2 \right) \int_{t(i)}^{t(i)+(m+1)k_n\Delta_n} L \left(\frac{t-\tau}{b_n} \right) dB_t \right)$$

$$+ \mathbb{E}_{i-1} \left(\left(\sum_{r=0}^{mk_n-1} \Psi_{t(i)+r,r} \right) \int_{t(i)}^{t(i)+(m+1)k_n\Delta_n} L \left(\frac{t-\tau}{b_n} \right) dB_t \right) := A_i + B_i.$$

Let $U_{i,r}^s = \int_{t(i)+r\Delta_n}^s g_n \left(\frac{u-(t(i)+r\Delta_n)}{k_n\Delta_n} \right) dW_u$, $g_n(t) = \sum_{r=1}^{k_n} g\left(\frac{r}{k_n} \right) \mathbb{1}_{\left[\frac{(r-1)\Delta_n}{k_n\Delta_n}, \frac{r\Delta_n}{k_n\Delta_n}, \right]}(t)$. By the Itô lemma, we have, when $t(i) > \tau - \sqrt{b_n}$,

$$A_i = \frac{1}{k_n^2 \Delta_n} \sigma_{t(i)}^2 \mathbb{E}_{i-1} \left(\sum_{r=0}^{mk_n-1} \int_{t(i)+r\Delta_n}^{t(i)+(r+k_n)\Delta_n} U_{i,r}^s g_n \left(\frac{s-(t(i)+r\Delta_n)}{k_n\Delta_n} \right) L \left(\frac{s-\tau}{b_n} \right) \rho_s ds \right) = 0,$$

since $\mathbb{E}\left(U_{i,r}^s\big|\mathcal{G}_{i-1}\right) = 0$. As for B_i, we can apply the Cauchy–Schwarz inequality. By (B.16) and the boundedness of L,

$$B_i^2 \leq \mathbb{E}_{i-1}\left(\sum_{r=0}^{mk_n-1}\Psi_{t(i)+r,r}\right)^2\int_{t(i)}^{t(i)+(m+1)k_n\Delta_n}L^2\left(\frac{s-\tau}{b_n}\right)ds \leq C\Delta_n.$$

Finally, we can show

$$\sum_{i=1}^{l_n(m)}\mathbb{E}\left(\left(\zeta(m)_i^n\alpha(m)_i^n\right)|\mathcal{G}_{i-1}\right) = \sum_{i:t(i)>\tau-\sqrt{b_n}}^{l_n(m)}Cb_nK_{b_n}(t(i)-\tau)\Lambda_{\tau-\sqrt{b_n}}(A_i+B_i)$$

$$\leq Cb_n\sum_{i=1}^{l_n(m)}\left|K_{b_n}(t(i)-\tau)\right|\left(\Delta_n^{1/2}\right) = O(m_n\Delta_n) \to 0.$$

Now, we single out a two-dimensional BM $\tilde{W} = (W,B)$, and a subset $\mathcal{N}$ of bounded martingales, all orthogonal to $\tilde{W}$. Let $D_i^n(N) = N_{(I_{i+1}-1)\Delta_n} - N_{(I_i-1)\Delta_n}$. We need to prove $\sum_{i=1}^{l_n(m)}\mathbb{E}\left(\left(\zeta(m)_i^n + \alpha(m)_i^n\right)D_i^n(N)|\mathcal{G}_{i-1}\right) \overset{\mathbb{P}}{\to} 0$, whenever N is one of the components of $\tilde{W}$ or is in the set $\mathcal{N}$. Since $[W_t, B_t] \leq [W_t, W_t] = t$, we can deduce $\sum_{i=1}^{l_n(m)}\mathbb{E}\left(\left(\zeta(m)_i^n + \alpha(m)_i^n\right)D_i^n(N)|\mathcal{G}_{i-1}^n\right) \overset{\mathbb{P}}{\to} 0,$ for the same reason as in proving (B.29). Next, $\sum_{i=1}^{l_n(m)}\mathbb{E}\left(\left(\zeta(m)_i^n + \alpha(m)_i^n\right)^4|\mathcal{G}_{i-1}\right) \overset{\mathbb{P}}{\to} 0$ can be easily deduced by straightforward computation and (B.27). Thus, letting $m \to \infty$, we can conclude that $m_n^{1/2}\Delta_n^{1/4}\left(\overline{H}(1)^n + \overline{H}(2)^n\right)$ converges stably in law to a random variable defined on a good extension $\left(\tilde{\Omega}, \tilde{\mathcal{F}}, \left(\tilde{\mathcal{F}}_t\right)_{t>0}, \tilde{\mathbb{P}}\right)$ of the space $\left(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geq 0}, \mathbb{P}\right)$, which, conditionally on $\mathcal{F}$, is Gaussian with a conditional variance $\delta_1^2 + \delta_2^2$. Combining with Lemma B.2, we can finally deduce that $m_n^{1/2}\Delta_n^{1/4}\left(\overline{H}(1)^n + \overline{H}(2)^n\right) \overset{st}{\longrightarrow} Z_\tau + \beta Z_\tau'$, where Z_τ, Z_τ' are defined on $\left(\tilde{\Omega}, \tilde{\mathcal{F}}, \left(\tilde{\mathcal{F}}_t\right)_{t>0}, \tilde{\mathbb{P}}\right)$ and conditionally independent with $\widetilde{\mathbb{E}}_{i-1}\left(Z_\tau^2\right) = \delta_1^2$ and $\widetilde{\mathbb{E}}_{i-1}\left(Z_\tau'^2\right) = \delta_2^2$.

REFERENCES

Aït-Sahalia, Y. & J. Jacod (2014) *High-Frequency Financial Econometrics*. Princeton University Press.

Aït-Sahalia, Y. & D. Xiu (2019) Principal component analysis of high-frequency data. *Journal of the American Statistical Association* 114(525), 287–303.

Alvarez, A., F. Panloup, M. Pontier, & N. Savy (2012) Estimation of the instantaneous volatility. *Statistical Inference for Stochastic Processes* 15(1), 27–59.

Bandi, F. & J. Russell (2008) Microstructure noise, realized volatility and optimal sampling. *Review of Economic Studies* 75, 339–369.

Barndorff-Nielsen, O.E., P.R. Hansen, A. Lunde, & N. Shephard (2008) Designing realized kernels to measure the ex post variation of equity prices in the presence of noise. *Econometrica* 76(6), 1481–1536.

Chen, R.Y. (2019) Inference for Volatility Functionals of Multivariate Itô Semimartingales Observed with Jump and Noise. Working paper. Available at arXiv:1810.04725v2.

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

Figueroa-López, J.E., R. Gong, & Y. Han (2022) Estimation of a tempered stable Lévy model of infinite variation. *Methodology and Computing in Applied Probability* 24(2), 713–747.

Figueroa-López, J.E. & C. Li (2020a) Optimal kernel estimation of spot volatility of stochastic differential equations. *Stochastic Processes and their Applications* 130(8), 4693–4720.

Figueroa-López, J.E. & C. Li (2020b) Supplement to "Optimal kernel estimation of spot volatility of stochastic differential equations." Available at https://sites.wustl.edu/figueroa/.

Figueroa-López, J.E. & C. Mancini (2019) Optimum thresholding using mean and conditional mean square error. *Journal of Econometrics* 208(1), 179–210.

Figueroa-López, J.E. & B. Wu (2022) Kernel estimation of spot volatility with microstructure noise using pre-averaging. *Preprint*, arXiv:2004.01865v3.

Foster, D. & D. Nelson (1996) Continuous record asymptotics for rolling sample variance estimators. *Econometrica* 64(1), 139–174.

Hansen, P. & A. Lunde (2006) Realized variance and market microstructure noise. *Journal of Business and Economic Statistics* 24, 127–218.

Huang, X. & G. Tauchen (2005) The relative contribution of jumps to total price variance. *Journal of Financial Econometrics* 3(4), 456–499.

Jacod, J., Y. Li, P.A. Mykland, M. Podolskij, & M. Vetter (2009) Microstructure noise in the continuous case: The pre-averaging approach. *Stochastic Processes and their Applications* 119(7), 2249–2276.

Jacod, J. & P.A. Mykland (2015) Microstructure noise in the continuous case: Approximate efficiency of the adaptive pre-averaging method. *Stochastic Processes and their Applications* 125(8), 2910–2936.

Jacod, J., M. Podolskij, & M. Vetter (2010) Limit theorems for moving averages of discretized processes plus noise. *Annals of Statistics* 38(3), 1478–1545.

Jacod, J. & P. Protter (2011) *Discretization of Processes*. Springer Science & Business Media.

Jacod, J. & M. Rosenbaum (2013) Quarticity and other functionals of volatility: Efficient estimation. *Annals of Statistics* 41(3), 1462–1484.

Jacod, J. & V. Todorov (2014) Efficient estimation of integrated volatility in presence of infinite variation jumps. *Annals of Statistics* 42(3), 1029–1069.

Kristensen, D. (2010) Nonparametric filtering of the realized spot volatility: A kernel-based approach. *Econometric Theory* 26(1), 60–93.

Li, J., Y. Liu, & D. Xiu (2019) Efficient estimation of integrated volatility functionals via multiscale jackknife. *Annals of Statistics* 47(1), 156–176.

Li, J., V. Todorov, & G. Tauchen (2017) Adaptive estimation of continuous-time regression models using high-frequency data. *Journal of Econometrics* 200(1), 36–47.

Li, J. & D. Xiu (2016) Generalized method of integrated moments for high-frequency data. *Econometrica* 84(4), 1613–1633.

Mancini, C. (2009) Non-parametric threshold estimation for models with stochastic diffusion coefficient and jumps. *Scandinavian Journal of Statistics* 36(2), 270–296.

Mancini, C., V. Mattiussi, & R. Renò (2015) Spot volatility estimation using delta sequences. *Finance and Stochastics* 19(2), 261–293.

Mykland, P. & L. Zhang (2012) The econometrics of high-frequency data. In M. Kessler, A. Lindner, & M. Sorensen (eds.), *Statistical Methods for Stochastic Differential Equations*. Chapman and Hall/CRC, pp. 109–190.

Mykland, P.A. & L. Zhang (2009) Inference for continuous semimartingales observed at high frequency. *Econometrica* 77(5), 1403–1445.

Parzen, E. (1962) On estimation of a probability density function and mode. *The Annals of Mathematical Statistics* 33(3), 1065–1076.

Podolskij, M. & M. Vetter (2009) Estimation of volatility functionals in the simultaneous presence of microstructure noise and jumps. *Bernoulli* 15(3), 634–658.

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

van Eeden, C. (1985) Mean integrated squared error of kernel estimators when the density and its derivative are not necessarily continuous. *Annals of the Institute of Statistical Mathematics* 37, 461–472.

Xiu, D. (2010) Quasi-maximum likelihood estimation of volatility with high frequency data. *Journal of Econometrics* 159(1), 235–250.

Yu, C., Y. Fang, Z. Li, B. Zhang, & X. Zhao (2014a) Non-parametric estimation of high-frequency spot volatility for Brownian semimartingale with jumps. *Journal of Time Series Analysis* 35, 572–591.

Yu, C., Y. Fang, Z. Li, B. Zhang, & X. Zhao (2014b) Kernel filtering of spot volatility in presence of Lévy jumps and market microstructure noise. *Preprint*. Available at https://mpra.ub.uni-muenchen.de/63293/1/MPRA_paper_63293.pdf.

Zhang, L. (2006) Efficient estimation of stochastic volatility using noisy observations: A multi-scale approach. *Bernoulli* 12(6), 1019–1043.

Zhang, L., P.A. Mykland, & Y. Aït-Sahalia (2005) A tale of two time scales: Determining integrated volatility with noisy high-frequency data. *Journal of the American Statistical Association* 100(472), 1394–1411.

Zu, Y. & H.P. Boswijk (2014) Estimating spot volatility with high-frequency financial data. *Journal of Econometrics* 181(2), 117–135.