Kernel filtering of spot volatility in presence of Lévy jumps and market microstructure noise

Yu, Chao and Fang, Yue and Zhao, Xujie and Zhang, Bo

School of Statistics, University of International Business and Economics, Lunquist College of Business, University of Oregon, School of International Trade and Economics, University of International Business and Economics, School of Statistics, Renmin University of China

23 March 2013

Kernel Filtering of Spot Volatility in Presence of Lévy Jumps and Market Microstructure Noise*

Chao Yu[a], Yue Fang[b], Xujie Zhao[a] and Bo Zhang[c]

[a]*University of International Business and Economics, Beijing, P.R.China*

[b]*University of Oregon, Eugene, Oregon, U.S.A.*

[c]*Renmin University of China, Beijing, P.R.China*

Mar 10, 2014

Abstract

This paper considers the problem of estimating spot volatility in the simultaneous presence of Lévy jumps and market microstructure noise. We propose to use the pre-averaging approach and the threshold kernel-based method to construct a spot volatility estimator, which is robust to both microstructure noise and jumps of either finite or infinite activity. The estimator is consistent and asymptotically normal, with a fast convergence rate. Our estimator is general enough to include many existing kernel-based estimators as special cases. When the kernel bandwidth is fixed, our estimator leads to widely used estimators of integrated volatility. Monte Carlo simulations show that our estimator works very well.

Keywords: high-frequency data, spot volatility, Lévy jump, kernel estimation, microstructure noise, pre-averaging.

1 Introduction

How to estimate the volatility of a financial instrument has long been a central topic of great interest to e-conomists. The availability of high-frequency financial data has led to substantial improvements in modelling

*This research was supported in part by the Fundamental Research Funds for the Central Universities, the Research Funds of Renmin University of China (10XNL007), and the NSFC (71271210), the Humanities and Social Sciences Project of Chinese Ministry of Education (13YJC790217), and the Fundamental Research Funds for the Central Universities in UIBE (CXTD5-05). Address correspondences to Yue Fang, Lunquist College of Business, University of Oregon, Eugene, OR, 97403, USA; e-mail: yfang@uoregon.edu.

and estimating time-varying volatility (Aït-Sahalia and Jacod (2014)). Despite theoretical, computational and empirical advances, however, most studies have concentrated on Integrated Volatility (IV) over some arbitrarily fixed time period, typically one day in empirical applications, as a measure of volatility. The results developed in stochastic calculus show that the sum of squared returns is consistent for IV over a period of time if the process is observed continuously. Hence, within the setting of a continuous semimartingale, the IV as a model-free quantity is a natural choice as a volatility measurement and can be estimated consistently and nonparametrically. The nonparametric method is appealing because the asymptotic properties can be developed under fairly mild assumptions (Jacod (1994), Jacod and Protter (1998), Barndorff-Nielsen and Shephard (2002), and Mykland and Zhang (2006)). In this paper, we aim to use a kernel-weighted version of the realized volatility estimator to construct a spot volatility estimator by shrinking the bandwidth at an appropriate rate, resulting in desired asymptotic properties with a fast convergence rate.

In theory, estimating volatility using high-frequency data requires a large amount of data to be effective. However, when applied to data recorded at very high frequencies, volatility estimators including IV and spot volatility estimators are sensitive to market frictions (so-called market microstructure noise) and pronounced discontinuous patterns of the intraday returns (i.e., jumps).

A common practice for dealing with microstructure noise is to model the log price semimartingale as latent rather than as observed (see, for example, Fang (1996), Zhou (1996), Andersen, Bollerslev, Diebold and Labys (2000), Hansen and Lund (2006), Bandi and Russell (2008)). There are currently three main nonparametric approaches to estimating volatility in the presence of microstructure noise: the two-scale or multi-scale realized volatility approach based on subsampling (Zhang et al. (2005, 2006)); a realized kernel estimator based on a linear combination of autocovariances (Barndorff-Nielsen et al. (2008)); and the pre-averaging method, which uses local "pre-averaging" via a kernel function to produce a set of non-overlapping (asymptotically) noise-free observations (Podolskij and Vetter (2009) and Jacod, Li, Mykland (2009)). In fact, these three methods give rise to asymptotically equivalent IV estimators with the optimal convergence rate of $n^{-1/4}$, where n is the sample size of the time series. In this paper, we use the pre-averaging approach to construct the noise-robust spot volatility estimate; consult Jacod et al. (2009) for more about the advantages of the pre-averaging approach and the issues with implementing the method.

Another complication that usually arises in high-frequency financial data analysis is that the return series do not have continuous paths, but rather exhibit jumps. Recent empirical evidence points to the fact that jumps in returns may take on different forms, such as jumps with finite activity or infinite activity (Carr and Wu (2003, 2004, 2007), Li et al. (2008), Lee and Hannig (2010), Fan and Fan (2011), Jing, Kong and Liu (2011), Cont and Mancini (2011), Aït-Sahalia and Jacod (2009, 2011), and Lee and Mykland (2012)). In this paper, we consider Lévy jumps which are flexible in modeling various types of jumps, including infinite activity jumps that cannot be described by either diffusion processes or compound Poisson jumps. We adopt the threshold approach to construct spot volatility estimates robust to Lévy jumps. We show that the threshold approach works well for jumps with both finite and infinite activity.[1]

Econometric literature on spot volatility estimation with high frequency data was pioneered by Foster and Nelson (1996), who propose the use of rolling and block sampling filters to estimate the spot volatility in pure diffusion settings; see also Andreou and Ghysels (2002) for a study of the finite sample performance of Foster and Nelson's estimator. In a more recent study, Kristensen (2010) proposes a kernel-weighted version of the realized volatility estimator for spot volatility in the absence of jump and market microstructure noise; Yu et al. (2014) extends Kristensen's results to allow the presence of jumps with finite activity. Both of their estimators are asymptotically normal and have a convergence rate of $n^{-1/2}h^{-1/2}$, where h is the kernel bandwidth.

The closest results in other literature to the results given here for spot volatility estimation with the presence of microstructure noise are those of Mancini, Mattiussi and Renò (2012) and Zu and Boswijk (2014). Although Mancini, Mattiussi and Renò (2012) use the delta sequence approach and Zu and Boswijk (2014) adopt the kernel method in constructing spot volatility estimates, both estimators use the two-scale approach proposed by Zhang et al. (2005) to deal with microstructure noise. The asymptomatic normality of their estimators are established under similar assumptions to ours. Their estimators, however, have a convergence rate of $n^{-1/6}h^{-1/2}$, which is substantially slower than that of our proposed estimator. The slower convergence of their estimators is expected due to the suboptimal nature of the two-scale procedure (Zhang (2006)). In addition, as Zu and Boswijk (2014) point out, it is difficult to construct jump-robust estimators with the two-scale approach. We note that Mancini, Mattiussi and Renò (2012) also consider the case of jumps with finite activity and apply the threshold method to obtain the jump-robust estimator. They show the consistency of their estimator, but are unable to establish the asymptotic normality of the estimator in the presence of jumps.

This paper introduces a new type of spot volatility estimator based on high-frequency data, allowing for the presence of both Lévy jumps and market microstructure noise. The basic strategy is to combine the pre-averaging approach and the threshold kernel-based method: the averaging of observed prices over a local window allows us to asymptotically remove the market microstructure noise; while the kernel with an appropriate threshold allows us to filter out jumps and approximate the true volatility. We show that our estimator is asymptotically normal with a convergence rate of $n^{-1/4}h^{-1/2}$. This convergence rate is a natural blend of two causes, which makes it slower than the usual $n^{-1/2}$ rate: a $1/4$-exponent loss due to microstructure noise and the extra factor $h^{-1/2}$ due to kernel filtering of the spot volatility. However, for the problem discussed in this paper, the convergence rate is very fast. In the case of modelling σ_t^2 as a Brownian motion, the convergence rate of our estimator is nearly equal to $n^{-1/8}$, which is the best rate attainable by any spot volatility estimator based on data observed with noise (Hoffmann et al. (2010)).

It is well known that if microstructure noise is present but unaccounted for, then the optimal sampling fre-

[1] Alternative strategies based on bipower and multipower variation processes have been developed in Barndorff-Nielsen and Shephard (2004, 2006). The multipower variation estimator was first developed by Barndorff-Nielsen and Shephard (2006) under the assumption of finite activity jumps. Although results have been extended to the case of jumps with infinite jump activity, as Mancini (2009) pointed out, the extension may only work for very specific volatility cases.

quency in estimating the IV is finite. This is also true in the estimation of spot volatility here. The development of noise-robust estimators for spot volatility allows us not to discard a vast amount of data as a solution, but to diminish the impact of microstructure noise. Our finite sample simulations confirm this: sampling as often as possible will produce more efficient estimators for spot volatility. Our results also highlight the importance of choosing an estimator based on both the price dynamics and the sampling frequency. For example, our results indicate that if both jumps and microstructure noise are present, when the sampling frequency is low, say 5- or 10-min, the noise-robust estimator may be less efficient than the estimator which does not account for noises. Surprisingly, our results also show that for processes without jumps, the jump-robust estimator may perform better at certain frequencies than the estimator which does not account for jumps. Of course, data sampled at higher frequencies always allow the jump and noise-robust estimators to achieve better estimation results.

The remainder of this paper is organized as follows. Section 2 lays out the basic setup. In Section 3, we introduce our spot volatility estimators and establish their links with existing estimators. In Section 4, we provide central limit theorems for our estimators, allowing for market microstructure noise in both scenarios with no jumps and with jumps; in the presence of jumps, our estimator is applicable whether the jumps have finite or infinite activity. Section 5 provides a simulation study to demonstrate the proposed estimators' finite sample performances. Finally, Section 6 draws conclusions. All proofs are located in the Appendices.

2 SETTING AND ASSUMPTIONS

2.1 The Lévy Jump-diffusion Process

We consider the univariate logarithmic price process $(X_t)_{t\geq 0}$ of an asset defined on a filtered probability space $(\Omega^{(0)}, \mathcal{F}^{(0)}, (\mathcal{F}_t^{(0)})_{t\geq 0}, \mathbb{P}^{(0)})$, and assume that X_t evolves as

$$dX_t = b_t dt + \sigma_t dW_t + dJ_t, \tag{1}$$

where $W = (W_t)$ is a standard Brownian motion. The drift $b = (b_t)$ and the volatility $\sigma = (\sigma_t)$ are progressively measurable processes which guarantee that (1) has a unique, strong solution. $J = (J_t)$ is a Lévy jump process with a Lévy jump measure ν and is independent of W.

Assumption A1. (Properties of b and σ)

(a) Both b and σ are adapted, and càdlàg process, and jointly independent of W;

(b) The path of volatility $t \mapsto \sigma_t^2$ lies in $C^{m,\gamma}[0,T]$ for some $m \geq 0$ and $0 < \gamma < 1$, i.e. $t \mapsto \sigma_t^2$ are m times differentiable with the mth derivative $(\sigma_t^2)^{(m)}$ satisfying

$$|(\sigma_{t+\delta}^2)^{(m)} - (\sigma_t^2)^{(m)}| \leq L(t, |\delta|)|\delta|^\gamma + o(|\delta|^\gamma), \quad \delta \to 0 \quad (a.s.)$$

where $\delta \mapsto L(t, \delta)$ is a slowly varying (random) function at zero and $t \mapsto L(t, 0)$ is continuous.

Assumption A1(a) consists of regularity conditions of the local behavior of the spot drift and volatility processes. Clearly, this assumption is satisfied by a wide class of stochastic volatility models, including those in which b and σ have continuous trajectories (see, for example, Hull and White (1987) and Heston (1993)). It does, however, rule out leverage effects. Although the independence assumption does not appear to be strictly necessary, Kanaya and Kristensen (2010) demonstrate that in the case of the diffusion process without jumps, their spot volatility estimator remains consistent if one drops the independence assumption. To focus on developing a spot estimator robust to microstructure noises and jumps, we will restrict our analysis to the case without leverage effects.

Assumption A1(a) is typically required for integrated volatility estimations. It would suffice to derive the asymptotic properties of kernel volatility estimators in the fixed bandwidth setting. In the setting of estimating spot volatility, we require that $h \to 0$. In this case, we need to impose smoothness assumptions on the volatility process to control the bias. A standard approach to bias reduction is to assume the object of interest is differentiable up to a certain order. This assumption is, however, violated by standard stochastic volatility models. Following Kristensen (2010) and Yu et al. (2014), we introduce a more general smoothness condition in Assumption A1(b) that allows for process (σ_t^2) to have nondifferentiable trajectories as long as they are smooth of order $0 < \gamma < 1$ almost surely.

Assumption A1(b) is satisfied by diffusion processes commonly used in volatility literature. In the special case that σ_t is driven by Brownian motion, it holds with $m = 0$ and $0 < \gamma < 1/2$. A similar smoothness condition with $m \geq 2$ is imposed in Genon-Catalot et al. (1992). See also Genon-Catalot et al. (1992) for alternative definitions and assumptions regarding the smoothness conditions imposed on volatility processes.

All our requirements for the jump process are expressed in the next two assumptions. Note that J_t can be written as the sum of "large" jump and "small" jump components:

$$J_t = \int_0^t \int_{|x|>1} x\mu(ds, dx) + \int_0^t \int_{|x|\leq 1} x(\mu(ds, dx) - \nu(dx)ds) := J_{1t} + J_{2t}, \tag{2}$$

where μ is the Poisson random measure of J_t and $\tilde{\mu}(ds, dx) = \mu(ds, dx) - \nu(dx)ds$ is the compensated measure. J_{1t} is a compound Poisson process with finite activity of jump and can be further written as $J_{1t} = \sum_{i=1}^{N_t} Y_{\tau_i}$, where N_t is a Poisson process and Y_{τ_i} denotes the jump size at jump time τ_i. J_{2t} is a square integrable martingale with infinite activity of jump.

Assumption A2. (Finite activity jumps)

(a) N_t is independent of W_t;

(b) N_t has a constant intensity λ;

(c) Y_{τ_i} are i.i.d. and independent of N_t.

Assumption A3. (Infinite activity jumps)

$$\int_{|x|\leq\delta} x^2 \nu(dx) = O(\delta^{2-\alpha}), \text{ as } \delta \to 0,$$

$$\int_{\delta \le |x| \le 1} |x|^j \nu(dx) = O(jc + (-1)^j c\delta^{j-\alpha}),\ j = 0, 1,$$

where c is a constant, ν is the Lévy measure of J_t and α is the Blumenthal-Getoor index measuring the activity of small jumps of J_t, defined as

$$\alpha := \inf\{\delta \ge 0, \int_{|x|\le 1} |x|^\delta \nu(dx) < +\infty\}.$$

Note that J_t is a Lévy pure jump process. Thus, $\alpha \in [0, 2]$. An infinite activity jump process with Blumenthal-Getoor index $\alpha < 1$ has paths with finite variation. If $\alpha > 1$, the sample paths have infinite variation almost surely. For $\alpha = 1$, the sample paths have either finite or infinite variation.

Assumption A3 is not as formidable as it appears. In fact, it is trivially satisfied for many commonly used models, such as NIG, Variance Gamma, tempered stable, α-stable, and GHL, among others. Similar requirements are given in Cont and Mancini (2011) and Mancini (2009).

2.2 Market Microstructure Noises

We assume that at any given time t_i, the observed log-price Z_{t_i} is

$$Z_{t_i} = X_{t_i} + \epsilon_{t_i}, \tag{3}$$

where ϵ_t is the market microstructure noise.

We further assume that for any $t \ge 0$, we have a transition probability $Q_t(\omega^{(0)}, dz)$ from $(\Omega^{(0)}, \mathcal{F}_t^{(0)})$ into $\mathbb{R}$, which satisfies

$$\int z Q_t(\omega^{(0)}, dz) = X_t(\omega^{(0)}). \tag{4}$$

We endow the space $\Omega^{(1)} = \mathbb{R}^{[0,\infty)}$ with the product Borel σ-field $\mathcal{F}^{(1)}$ and with the probability $\mathbb{Q}(\omega^{(0)}, d\omega^{(1)})$, which is the product $\otimes_{t\ge 0} Q_t(\omega^{(0)}, \cdot)$. Process $(Z_t)_{t\ge 0}$ is defined on $(\Omega^{(1)}, \mathcal{F}^{(1)})$ and the filtration $\mathcal{F}_t^{(1)} = \sigma(Z_s : s \le t)$. We work in the filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F})_{t\ge 0}, \mathbb{P})$ defined as follows:

$$\Omega = \Omega^{(0)} \times \Omega^{(1)},\ \mathcal{F} = \mathcal{F}^{(0)} \times \mathcal{F}^{(1)},\ \mathcal{F}_t = \cap_{s>t} \mathcal{F}_s^{(0)} \times \mathcal{F}_s^{(1)},$$
$$\mathbb{P}(d\omega^{(0)}, d\omega^{(1)}) = \mathbb{P}^{(0)}(d\omega^{(0)}) \mathbb{Q}(\omega^{(0)}, d\omega^{(1)}).$$

Assumption A4. (Market microstructure noise)
The ϵ_ts are i.i.d. and independent of W_t and J_t processes, with $E\epsilon_t = 0$ and $E|\epsilon_t|^8 < \infty$.[2]

Let $\alpha_t = E((Z_t)^2|\mathcal{F}^{(0)}) - (X_t)^2$. Assumption A4 implies that the process α_t is càdlàg, and $E((Z_t)^8|\mathcal{F}^{(0)})$ is a locally bounded process. Clearly, the noise process which meets the requirements given in Assumption A4 satisfies (4).

[2]Similar to most other literature, we start with the pure additive noise and make a few basic and mild assumptions on the noise process. As usual, we require some moment conditions. Note that the 8th moment condition requirement is primarily for reasons of tractability. Although our results may be modified to account for more general microstructure noise processes, these processes introduce intricate technical challenges without providing much more insight into the problem and are outside the scope of this paper.

2.3 Kernel and Threshold Functions

The requirements of the kernel function are presented in Assumption A5.[3]

Assumption A5. (The kernel function)

The kernel $K : \mathbb{R} \mapsto \mathbb{R}$ is continuously differentiable and bounded, with

(a) $\int_{\mathbb{R}} K(x)dx = 1$;

(b) $\int_{\mathbb{R}} x^i K(x)dx = 0$, where $i = 1, \cdots, r-1$, and $\int_{\mathbb{R}} |x|^r |K(x)| dx < \infty$ for some $r \geq 0$.

Assumption A5 is satisfied by most standard kernels where $r \leq 2$. In this paper, we use one-sided kernels, which require only information up to current time and generally lead to more precise estimates near boundaries; see, for example, Zhang and Karunamuni (1998) and Kristensen (2010).

The last assumption, Assumption A6, presents the requirements of the threshold function $r(x)$, which are essential for identifying the intervals where no jump occurred with noisy observation.

Assumption A6. (The threshold function)

The threshold function $r(x)$ is a deterministic function of the step length x, such that

(a) $\lim_{x \to 0} r(x) = 0$;

(b) $\lim_{x \to 0} \frac{x^{1/2}(\log \frac{1}{x})^2}{r(x)} = 0$.

Power functions $r(x) = \beta x^\alpha$ for any $\alpha \in (0, 1/2)$ and $\beta \in R$ are possible choices.

3 The Estimator

3.1 The Definition

Assume that observations of Z_t are sampled at discrete times $0 = t_0 < t_1 < ... < t_n = T$ over a fixed time interval $[0, T]$: $Z_0^n, Z_1^n, \cdots, Z_n^n$. For simplicity, we consider that observations are sampled at regularly spaced discrete times $t_i = i\Delta_n$ for $i = 0, 1, \cdots, n$. The goal is to estimate σ_τ^2 for $\tau \in [0, T]$ in (1). In the following, we use the shorthand notation $Z_i^n = Z_{i\Delta_n}$, $\Delta_i^n Z = Z_i^n - Z_{i-1}^n$.

Let $\bar{Z}_i^n$ denote the weighted average of k_n observations of $Z_i^n, Z_{i+1}^n, \cdots, Z_{i+k_n-1}^n$. More specifically, $\bar{Z}_i^n = \sum_{j=1}^{k_n-1} g_j^n \Delta_{i+j}^n Z_i^n$, with weights $g_j^n = g(j/k_n)$. The weighting function $g(x)$ is required to be continuous on $[0, 1]$, piecewise C^1 with a piecewise Lipschitz derivative g', and with $g(0) = g(1) = 0$, $\int_0^1 g(s)^2 ds > 0$. We further require that the integer sequence k_n satisfies $k_n \sqrt{\Delta_n} = \theta + o(\Delta_n^{1/4})$ for some constant $\theta > 0$. Our proposed spot volatility estimator takes on the general form

$$\hat{\sigma}^2(\tau) = \frac{\sqrt{\Delta_n}}{\theta\psi_2} \sum_{i=0}^{n-k_n+1} K_h(t_i - \tau)(\bar{Z}_i^n)^2 I_A(\bar{Z}_i^n) - \frac{\psi_1 \Delta_n}{2\theta^2 \psi_2} \sum_{i=1}^{n} K_h(t_i - \tau)(\Delta_i^n Z)^2 I_A(\bar{Z}_i^n), \tag{5}$$

[3] In our study, we use kernels as weights to construct estimators for the kernel-smoothed integrated volatility of both fixed and shrinking bandwidths. We note that the kernel technique is also used in Barndorff-Nielsen et al. (2008) to estimate integrated volatility. However, Barndorff-Nielsen et al. uses kernels to eliminate microstructure noise rather than to construct estimates for spot volatility.

where $K_h(x) = K(x/h)/h$ with bandwidth h, and I_A is an indicator function on set A, which takes different forms depending on whether or not jumps are present. The parameters ψ_1 and ψ_2 are constants associated with the weighting function $g(x)$ in the pre-averaging step and are defined in the Appendix A.

3.2 Special Cases

The proposed estimator (5) represents a very general class of spot volatility estimators. It defines new estimators and includes many existing kernel-based spot volatility estimators as special cases. It is also related to several popular integrated volatility estimators proposed in other literature. We will first introduce our two newly proposed estimators: $\hat{\sigma}^2_{PATKV}$ and $\hat{\sigma}^2_{PAKV}$.

If both market microstructure noise and jumps are present, we advocate the following jump- and noise-robust estimator for spot volatility: $\hat{\sigma}^2_{PATKV}$.

- $\hat{\sigma}^2_{PATKV}$: The pre-averaging threshold kernel estimator. Let $A = \{(\bar{Z}^n_i)^2 \le r(\Delta_n)\}$, where $r(\Delta_n)$ is a threshold function satisfying Assumption A6. Then, we have

$$\hat{\sigma}^2_{PATKV}(\tau) \equiv \frac{\sqrt{\Delta_n}}{\theta\psi_2} \sum_{i=0}^{n-k_n+1} K_h(t_i - \tau)(\bar{Z}^n_i)^2 I_{\{(\bar{Z}^n_i)^2 \le r(\Delta_n)\}} - \frac{\psi_1 \Delta_n}{2\theta^2\psi_2} \sum_{i=1}^{n} K_h(t_i - \tau)(\Delta^n_i Z)^2 I_{\{(\bar{Z}^n_i)^2 \le r(\Delta_n)\}}. \quad (6)$$

  This proposed $PATKV$ estimator is the main focus of the paper. It ia constructed by carefully combining the pre-averaging approach and the threshold kernel-based method. The asymptotic properties of $\hat{\sigma}^2_{PATKV}$ will be examined for jumps with finite activity in Section 4.2 and for jumps with infinite activity in Section 4.3.

If microstructure noise is present but jumps are absent, we advocate the following noise-robust estimator for spot volatility, $\hat{\sigma}^2_{PAKV}$, which is a special case of $\hat{\sigma}^2_{PATKV}$.

- $\hat{\sigma}^2_{PAKV}$: The pre-averaging kernel estimator. It is defined by selecting $A = \mathbb{R}$:

$$\hat{\sigma}^2_{PAKV}(\tau) \equiv \frac{\sqrt{\Delta_n}}{\theta\psi_2} \sum_{i=0}^{n-k_n+1} K_h(t_i - \tau)(\bar{Z}^n_i)^2 - \frac{\psi_1 \Delta_n}{2\theta^2\psi_2} \sum_{i=1}^{n} K_h(t_i - \tau)(\Delta^n_i Z)^2. \quad (7)$$

  The asymptotic properties of $PAKV$ will be studied in Section 4.1. The results will then be extended to $\hat{\sigma}^2_{PATKV}$, which allows for the presence of jumps, in Sections 4.2 and 4.3. Two closely related estimators are worth noticing. Working in the same setting as the $PAKV$ estimator, Mancini, Mattiussi and Renò (2012) and Zu and Boswijk (2014) propose the use of the two-scale approach to deal with microstructure noise. To the best of our knowledge, they are the first to show how to construct noise-robust estimators for spot volatility. Their estimators are consistent and asymptotically normal. However, their estimators have a slower convergence rate than that of $\hat{\sigma}^2_{PAKV}$.

Next, we will present several estimators that have been studied in other literature. If jumps are present but microstructure noise is absent, we can construct the kernel estimator directly with $(X_i)s$ and advocate the following jump-robust estimator for spot volatility.

- $\hat{\sigma}^2_{TKV}$: The threshold kernel estimator. It is defined as

$$\hat{\sigma}^2_{TKV}(\tau) \equiv \sum_{i=1}^{n} K_h(t_i - \tau)(\Delta_i X)^2 I_{\{(\Delta_i X)^2 \le r(\Delta_n)\}}, \tag{8}$$

  where the requirements of threshold function $r(\Delta_n)$ is different with Assumption A6. This TKV estimator has been studied by Yu et al. (2014). They show that TKV is jump-robust and is asymptotically normally distributed. See also Bandi and Renò (2010) for an alternative approach that localizes an integrated variance estimator to filter spot volatility in the presence of jumps.

The TKV estimator extends Kristensen's (2010) kernel estimator in a setting where both jumps and market microstructure noise are absent.

- $\hat{\sigma}^2_{KV}$: The kernel estimator. It is defined as

$$\hat{\sigma}^2_{KV}(\tau) \equiv \sum_{i=1}^{n} K_h(t_i - \tau)(\Delta_i X)^2. \tag{9}$$

  This KV estimator can be regarded as a Nadaraya-Watson-type kernel estimator and has been studied by Kristensen (2010). It also includes the rolling window estimator proposed by Foster and Nelson (1996) as a special case.

When we replace the kernel $K_h(t_i-\tau)$ with an arbitrary bounded weight function $w(t_i)$, our estimator $\hat{\sigma}^2(\tau)$ defined in (5) leads to widely used IV estimators. Consider, for example, the noise-robust estimator $\hat{C}^{n,w}_t$ and the jump- and noise-robust estimator $\hat{C}^{n,j}_t$.

- $\hat{C}^{n,w}_t$: The estimator for the weighted IV when market microstructure noise is present but jumps are absent. Let $A = \mathbb{R}$. Then, the estimator (5) takes the form

$$\hat{C}^{n,w}_T \equiv \frac{\sqrt{\Delta_n}}{\theta\psi_2} \sum_{i=0}^{n-k_n+1} w_i(\bar{Z}^n_i)^2 - \frac{\psi_1\Delta_n}{2\theta^2\psi_2} \sum_{i=1}^{n} w_i(\Delta^n_i Z)^2. \tag{10}$$

  As will be shown in Section 4.1, $\hat{C}^{n,w}_t$ is a consistent estimator for the weighted IV, $\int_0^T w(s)\sigma^2_s ds$. As a special case, if we let $w(x) = I_{\{0\le x\le T\}}(x)$, the estimator (5) takes the form

$$\hat{C}^n_T \equiv \frac{\sqrt{\Delta_n}}{\theta\psi_2} \sum_{i=0}^{n-k_n+1} (\bar{Z}^n_i)^2 - \frac{\psi_1\Delta_n}{2\theta^2\psi_2} \sum_{i=1}^{n} (\Delta^n_i Z)^2.$$

  $\hat{C}^n_T$ is the pre-averaging realized volatility estimator for $\int_0^T \sigma^2_s ds$ studied in Jacod et al. (2009).

- $\hat{C}_T^{n,j}$: The IV estimator when both market microstructure noise and jumps are present. If $A = \{(\bar{Z}_i^n)^2 \leq r(\Delta_n)\}$ and $w(x) = I_{\{0 \leq x \leq T\}}(x)$, the estimator (5) becomes the realized volatility estimator considered in Jing et al.(2014):

$$\hat{C}_T^{n,j} \equiv \frac{\sqrt{\Delta_n}}{\theta\psi_2} \sum_{i=0}^{n-k_n+1} (\bar{Z}_i^n)^2 I_{\{(\bar{Z}_i^n)^2 \leq r(\Delta_n)\}} - \frac{\psi_1 \Delta_n}{2\theta^2 \psi_2} \sum_{i=1}^{n} (\Delta_i^n Z)^2 I_{\{(\bar{Z}_i^n)^2 \leq r(\Delta_n)\}}.$$

4 ASYMPTOTIC PROPERTIES

4.1 The Case of Continuous Semimartingale

In this subsection, we consider the asymptotics of the pre-averaging kernel volatility estimator $\hat{\sigma}^2_{PAKV}$ for scenarios in which market microstructure noise is present but jumps are absent. We start by studying the asymptotic behavior of the general weighted version of the pre-averaging volatility estimator $\hat{C}_T^{n,w}$, defined in (10). We denote $C_T^w = \int_0^T w(s)\sigma_s^2 ds$.

Theorem 1 *If Assumptions A1 and A4 hold, for any fixed $T > 0$, the sequence $\Delta_n^{-1/4}(\hat{C}_T^{n,w} - C_T^w)$ converges stably in law to a variable defined on an extension of the original space. This variable has the form*

$$Y_T = \int_0^T w_s \gamma_s dB_s,$$

where B is a standard Wiener process independent of $\mathcal{F}$ and γ_t is given by

$$\gamma_t^2 = \frac{4}{\psi_2^2}(\Phi_{22}\theta\sigma_t^4 + 2\Phi_{12}\frac{\sigma_t^2\alpha_t}{\theta} + \Phi_{11}\frac{\alpha_t^2}{\theta^3}).$$

Moreover, let

$$\begin{aligned}\Gamma_T^{n,w} &= \frac{4\Phi_{22}}{3\theta\psi_2^4} \sum_{i=0}^{n-k_n+1} w_i^2 (\bar{Z}_i^n)^4 \\ &+ \frac{4\Delta_n}{\theta^3}\left(\frac{\Phi_{12}}{\psi_2^3} - \frac{\Phi_{22}\psi_1}{\psi_2^4}\right) \sum_{i=0}^{n-2k_n+1} w_i^2 (\bar{Z}_i^n)^2 \sum_{j=i+k_n}^{i+2k_n-1} (\Delta_j^n Z)^2 \\ &+ \frac{\Delta_n}{\theta^3}\left(\frac{\Phi_{11}}{\psi_2^2} - \frac{2\Phi_{12}\psi_1}{\psi_2^3} + \frac{\Phi_{22}\psi_1^2}{\psi_2^4}\right) \sum_{i=1}^{n-2} w_i^2 (\Delta_i^n Z)^2 (\Delta_{i+2}^n Z)^2.\end{aligned}$$

Then,

$$\Gamma_T^{n,w} \xrightarrow{P} \int_0^T w_s^2 \gamma_s^2 ds.$$

Therefore, for any $T > 0$, the sequence $\frac{1}{\Delta_n^{1/4}\sqrt{\Gamma_T^{n,w}}}(\hat{C}_T^{n,w} - C_T^w)$ converges stably in law to a $N(0,1)$ variable independent of $\mathcal{F}$.

As noted in Section 3.2, if one lets $w(s) = I_{[0,T]}(s)$, $\hat{C}_t^{n,w}$ is equivalent to the IV estimator considered in Jacod et al. (2009). Theorem 1 covers Jacod's Theorem 3.1 as a special case.

Now, we are ready to study the asymptotic properties of $\hat{\sigma}^2_{PAKV}$. When we let $w(s) = K_h(s)$, we have the following theorem.

Theorem 2 *If Assumptions A1, A4, and A5 hold and kernel $K(x)$ satisfies A5 with $r \geq m + \gamma$, as $nh^2 \to \infty$ and $nh^{4(m+\gamma)+2} \to 0$, for any $\tau \in (0, T)$, we have*

$$\sqrt{\Delta_n^{-1/2} h}\left(\hat{\sigma}^2_{PAKV}(\tau) - \sigma^2_\tau\right) \xrightarrow{d} N(0, \gamma^2_\tau \int_{\mathbb{R}} K^2(s)ds).$$

Moreover, let

$$\begin{aligned}
\hat{\gamma}^2_{PAKV}(\tau) &= \frac{4\Phi_{22}}{3\theta\psi_2^4} \sum_{i=0}^{n-k_n+1} K_h(t_i - \tau)(\bar{Z}_i^n)^4 \\
&+ \frac{4\Delta_n}{\theta^3}\left(\frac{\Phi_{12}}{\psi_2^3} - \frac{\Phi_{22}\psi_1}{\psi_2^4}\right) \sum_{i=0}^{n-2k_n+1} K_h(t_i - \tau)(\bar{Z}_i^n)^2 \sum_{j=i+k_n}^{i+2k_n-1} (\Delta_j^n Z)^2 \\
&+ \frac{\Delta_n}{\theta^3}\left(\frac{\Phi_{11}}{\psi_2^2} - \frac{2\Phi_{12}\psi_1}{\psi_2^3} + \frac{\Phi_{22}\psi_1^2}{\psi_2^4}\right) \sum_{i=1}^{n-2} K_h(t_i - \tau)(\Delta_i^n Z)^2(\Delta_{i+2}^n Z)^2.
\end{aligned}$$

Then,

$$\hat{\gamma}^2_{PAKV}(\tau) \xrightarrow{P} \gamma^2_\tau \quad as \ \ h \to 0.$$

Kristensen (2010) considers the problem of spot volatility estimation in the absence of market microstructure noise. His estimator has a convergence rate of $n^{-1/2}h^{-1/2}$. The extra factor $h^{-1/2}$ (beyond the usual $n^{-1/2}$ convergence rate) is the result of spot volatility kernel filtering. Theorem 2 indicates that the convergence rate of $\hat{\sigma}^2_{PAKV}$ is $n^{-1/4}h^{-1/2}$, which is $1/4$-exponentially slower than that of Kristensen's estimator. The $n^{-1/4}$ efficiency loss due to microstructure noise coincides with $n^{-1/4}$ efficiency loss observed when estimating IV: the correction of microstructure noise tends to reduce the convergence rate of the estimators by $n^{-1/4}$ (see, for example, Zhang (2006), Barndorff-Nielsen et al. (2008), and Jacod et al. (2009)). This is in contrast to the convergence rate of $n^{-1/6}h^{-1/2}$ obtained from the two scale approach (Zu and Boswijk (2014)).

Theorem 2 indicates that the convergence rate of $\hat{\sigma}^2_{PAKV}$ depends on the smoothness of the volatility process. The convergence rate is, in general, very fast. Since we require that $nh^2 \to \infty$ and $nh^{4(m+\gamma)+2} \to 0$, $O_p(n^{-1/4}h^{-1/2})$ can be written as $O_p(n^{-1/4+\lambda})$ for any $\lambda > (8(m + \gamma) + 4)^{-1}$. Several authors studies model volatility as a smooth function that is m times differentiable (see, for example, Stanton (1997), Fan and Yao (1998), and Müller et al. (2011)). In this case, $\lambda = O_p(\frac{1}{8m})$. Thus, the convergence rate of $\hat{\sigma}^2_{PAKV}$ can be made arbitrarily close to $O_p(n^{-1/4})$ for large m. If the volatility is assumed to follow a diffusion process, it has continuous sample paths but is not differentiable (i.e., $m = 0$). In this case, our estimator can still have a fast convergence rate. In particularly, the convergence rate of $\hat{\sigma}^2_{PAKV}$ for Brownian motion ($m = 0$ and $\gamma < 1/2$) is nearly equal to $n^{-1/8}$, which is the best rate attainable for any spot volatility estimator in the presence of market microstructure noise (Hoffmann et al. (2010)).

4.2 The Case of Finite Activity Jumps

The following theorem presents the asymptotic properties of $\hat{\sigma}^2_{PATKV}$ in the presence of microstructure noise and jumps with finite activity.

Theorem 3 *If Assumptions A1-A2 and A4-A6 hold and kernel $K(x)$ satisfies A5 with $r \geq m+\gamma$, as $nh^2 \to \infty$ and $nh^{4(m+\gamma)+2} \to 0$, for any $\tau \in (0,T)$, we have*

$$\sqrt{\Delta_n^{-1/2}h}\left(\hat{\sigma}^2_{PATKV}(\tau) - \sigma^2_\tau\right) \xrightarrow{d} N(0, \gamma^2_\tau \int_{\mathbb{R}} K^2(s)ds).$$

Moreover, let

$$\begin{aligned}
\hat{\gamma}^2_{PATKV}(\tau) \quad = \quad & \frac{4\Phi_{22}}{3\theta\psi_2^4}\sum_{i=0}^{n-k_n+1} K_h(t_i-\tau)(\bar{Z}_i^n)^4 I_{\{(\bar{Z}_i^n)^2 \leq r(\Delta_n)\}} \\
& + \frac{4\Delta_n}{\theta^3}\left(\frac{\Phi_{12}}{\psi_2^3} - \frac{\Phi_{22}\psi_1}{\psi_2^4}\right)\sum_{i=0}^{n-2k_n+1} K_h(t_i-\tau)(\bar{Z}_i^n)^2 I_{\{(\bar{Z}_i^n)^2 \leq r(\Delta_n)\}} \sum_{j=i+k_n}^{i+2k_n-1} (\Delta_j^n Z)^2 I_{\{(\bar{Z}_{i+k_n}^n)^2 \leq r(\Delta_n)\}} \\
& + \frac{\Delta_n}{\theta^3}\left(\frac{\Phi_{11}}{\psi_2^2} - \frac{2\Phi_{12}\psi_1}{\psi_2^3} + \frac{\Phi_{22}\psi_1^2}{\psi_2^4}\right)\sum_{i=1}^{n-2} K_h(t_i-\tau)(\Delta_i^n Z)^2(\Delta_{i+2}^n Z)^2 I_{\{(\bar{Z}_i^n)^2 \leq r(\Delta_n)\}}.
\end{aligned}$$

Then,

$$\hat{\gamma}^2_{PATKV}(\tau) \xrightarrow{P} \gamma^2_\tau \quad as \quad h \to 0.$$

As Theorem 3 shows, $\hat{\sigma}^2_{PATKV}$ has the same asymptotic distribution as $\hat{\sigma}^2_{PAKV}$. In other words, the presence of jumps with finite activity does not affect the convergence rate and the asymptotic distribution of our spot volatility estimator, provided one uses an appropriate threshold to (asymptotically) identify and then exclude the intervals where jumps occur. A similar result is discussed in Yu et al. (2014) when the authors consider the problem of estimating spot volatility from observations without contamination of microstructure noise. Our result that the presence of finite-activity jumps does not affect the efficiency of spot volatility estimator is also consistent with the results in Mancini (2009), which studies the problem of IV estimation.

4.3 The Case of Infinite Activity Jumps

The following theorem directly extends the results of Theorem 3 to the case of infinite activity jumps.

Theorem 4 *Assume that Assumptions A1, A3 and A4-A6 hold, and kernel $K(x)$ satisfies A5 with $r \geq m+\gamma$. Let $r(\Delta_n) = \Delta_n^\beta$. As $nh^2 \to \infty$ and $nh^{4(m+\gamma)+2} \to 0$, for any $\tau \in (0,T)$,*

(a) if $\alpha < 1$ and $\beta > \frac{1}{4-2\alpha} \in (\frac{1}{4}, \frac{1}{2})$,

$$\sqrt{\Delta_n^{-1/2}h}\left(\hat{\sigma}^2_{PATKV}(\tau) - \sigma^2_\tau\right) \xrightarrow{d} N(0, \gamma^2_\tau \int_{\mathbb{R}} K^2(s)ds),$$

(b) if $\alpha \geq 1$, for any $\beta \in (0, 1/2)$,

$$\sqrt{\Delta_n^{-1/2}h}\left(\hat{\sigma}^2_{PATKV}(\tau) - \sigma^2_\tau\right) \xrightarrow{P} +\infty.$$

Note that $\hat{\sigma}^2_{PATKV}$ explodes if the sample path has infinite variation ($\alpha \geq 1$). When the sample path has finite variation, the degree of activity of jumps, α, does not impair the convergence rate. However, the choice of $r(\Delta_n)$ does depends on α.

To the best of our knowledge, Theorem 4 is the first contribution to solving the problem of estimating spot volatility with noisy observation in the presence of jumps with infinite activity. The result can be also useful in constructing tests for identifying the finer characteristics of jumps such as the degrees of jump activities.

5 SIMULATION FOR FINITE SAMPLE BEHAVIOR

To evaluate the finite sample performance of the proposed method, we conduct a simulation study for $\hat{\sigma}^2_{PAKV}$ in the absence of jumps in Section 5.1 and for $\hat{\sigma}^2_{PATKV}$ in the presence of jumps in Section 5.2. For comparison, we also report the results of the kernel-based filtering volatility estimator (Kristensen (2010)), $\hat{\sigma}^2_{KV}$, and the threshold kernel volatility estimator (Yu et al. (2014)), $\hat{\sigma}^2_{TKV}$. We note that both $\hat{\sigma}^2_{KV}$ and $\hat{\sigma}^2_{TKV}$ are designed to provide consistent estimates of spot volatility in cases where market microstructure noise is absent. Nevertheless, we include $\hat{\sigma}^2_{KV}$ and $\hat{\sigma}^2_{TKV}$ in our simulation comparisons to illustrate the risk of ignoring market microstructure noise when it is present in practice.[4]

The integrated mean squared error (IMSE) is used as the performance measure to evaluate the finite sample properties of the estimators in our simulation study:

$$IMSE = \int_{T_l}^{T_u} E[(\hat{\sigma}^2_t - \sigma^2_t)^2]dt,$$

where $0 \leq T_l < T_u \leq T$ and $\hat{\sigma}^2_t = PATKV$, $PAKV$, TKV, or KV.

5.1 The Case without Jumps

We consider the following stochastic volatility diffusion model, as studied by Banrdorff-Nielsen and Shephard (2004) and Huang and Tauchen (2005):

$$dX_t = udt + \exp[\beta_0 + \beta_1 v_t]dW_{1,t} \tag{11}$$

$$dv_t = \alpha_v v_t dt + dW_{2,t}, \tag{12}$$

where $W_{1,t}$ and $W_{2,t}$ are two independent, standard Brownian motions, and v_t is a stochastic volatility factor. We set $u = 0.03$, $\beta_0 = 0$, $\beta_1 = 0.125$, and $\alpha_v = -0.10$. These parameters were chosen to conform to other studies previously published in Andersen, Benzoni, and Lund (2002), Andersen, Bollerslev, and Diebold (2007), Huang and Tauchen (2005), and Chernov, Gallant and Ghysels et al.(2003).

[4]A one-sided kernel is adopted when constructing the various estimators in the simulation study. When jumps are present, a data-driven approach is used to select the optimal threshold; see Yu et al. (2014) for details.

To evaluate the impact of microstructure noise on the performances of the estimators, we consider three scenarios of σ_ϵ: 0.025, 0.035, and 0.05. To get an idea of the magnitude of the microstructure noise, we compute the ratio of σ_ϵ to the variance of the observed return over the interval $[t_i - \Delta_n, t_i]$:

$$\pi(\Delta_n) = \frac{2\sigma_\epsilon^2}{\sigma_d^2 \Delta_n + 2\sigma_\epsilon^2},$$

where σ_d^2 is the unconditional daily variance of (X_t). Note that $\pi(\Delta_n)$ can be viewed as the percentage of the variance of the observed return attributed to microstructure noise (Aït-Sahalia et al. (2005)). In general, $\pi(\Delta_n)$ increases as σ_ϵ^2 and sampling frequency increase; see Table 1. When data are sampled every 10 minutes, $\pi(\Delta_n)$ is relatively small, ranging from 4.65% to 16.32%. When the sampling frequencies reach 30 and 10 seconds, the volatility of the observed return series is caused mainly by the variability of the microstructure noise. For example, $\pi(\Delta_n) = 92.13\%$ when $\sigma_\epsilon = 0.05$ and the sampling frequency is 10 seconds.

Table 1: $\pi(\Delta_n)$ under different values of σ_ϵ and sampling frequencies

Frequency	$\sigma_\epsilon = 0.025$	$\sigma_\epsilon = 0.035$	$\sigma_\epsilon = 0.05$
10 sec	0.7452	0.8515	0.9213
30 sec	0.4937	0.6565	0.7959
1 min	0.3277	0.4886	0.6610
5 min	0.0888	0.1604	0.2806
10 min	0.0465	0.0872	0.1632

In our simulation, we set $T = 1$ and $\Delta_n = 1/(6.5 \times 60 \times 60)$. Hence, each simulation is conducted over one trading day consisting of 6.5 trading hours. In each simulation, we simulate 23,400 second-by-second data by utilizing the first-order Euler discretization scheme of (11) and (12). We simulate one trajectory for (σ_t^2) and keep them fixed. Then, we run 1,000 Monte Carlo repetitions for (X_t), which evaluate the performance of various estimators based on the sampling frequencies, which range from 10 seconds to 10 minutes.

Table 2 reports the IMSEs of $\hat{\sigma}^2_{PAKV}$ and $\hat{\sigma}^2_{KV}$. As expected, $\hat{\sigma}^2_{PAKV}$ is noise-robust: the IMSEs of $\hat{\sigma}^2_{PAKV}$ decay as the sampling frequencies increase. We note that the most efficiency gains occur in the frequency range between 30 seconds and five minutes, in which the proportion of microstructure noise contributions increases sharply (see Table 1). For a given sampling frequency, the IMSEs of $\hat{\sigma}^2_{PAKV}$ increase slightly as the level of microstructure noise increases.

In contrast, the efficiency of suffers when market microstructure noise is present. For any given level of σ_ϵ, the curve of the IMSE of $\hat{\sigma}^2_{KV}$ exhibits a U-shaped pattern that highlights the trade-off between the use of more data and the microstructure noise effect. Our results indicate that the "optimal" sampling frequency for $\hat{\sigma}^2_{KV}$ is

Table 2: The IMSEs of $\hat{\sigma}^2_{PAKV}$ and $\hat{\sigma}^2_{KV}$: the case with no jumps

Frequency	$\sigma_\epsilon = 0.025$		$\sigma_\epsilon = 0.035$		$\sigma_\epsilon = 0.05$	
	$\hat{\sigma}^2_{PAKV}$	$\hat{\sigma}^2_{KV}$	$\hat{\sigma}^2_{PAKV}$	$\hat{\sigma}^2_{KV}$	$\hat{\sigma}^2_{PAKV}$	$\hat{\sigma}^2_{KV}$
10 sec	0.0260	4.598	0.0269	16.75	0.0304	71.01
30 sec	0.0272	0.5843	0.0309	2.139	0.0353	8.598
1 min	0.0588	0.2372	0.0725	0.7330	0.0842	3.050
5 min	0.2334	3.5524	0.2389	10.68	0.2405	58.40
10 min	0.2565	7.7822	0.2790	20.24	0.2882	114.4

about one minute, which is consistent with the results reported in other studies (for example, Bandi and Russell (2008) and Zhang et al. (2005)).

The results in Table 2 clearly highlight the importance of accounting for microstructure noise when estimating spot volatility. Even when the level of market microstructure noise is relatively low, the KV estimator at the "optimal" sampling frequency still performs much worse than $\hat{\sigma}^2_{PAKV}$. For example, when $\sigma_\epsilon = 0.025$, the optimal sampling frequency of $\hat{\sigma}^2_{KV}$ is 1 minute. In this case, the IMSE of $\hat{\sigma}^2_{KV} = 0.2372$, which is about 400% of that of $\hat{\sigma}^2_{PAKV}$.

5.2 The Case with Jumps

Consider the following finite activity jump diffusion model:

$$dX_t = udt + \exp[\beta_0 + \beta_1 v_t]dW_{1,t} + dJ_t \tag{13}$$

$$dv_t = \alpha_v v_t dt + dW_{2,t}, \tag{14}$$

where $J_t = \sum_{j=1}^{N_t} Y_{\tau_j}$ is a compound Poisson jump process. We further assume that N_t is a Poisson process with intensity $\lambda = 3$ and jump size $Y_{\tau_j} \sim N(0, \sigma_Y^2)$. To evaluate the impact of jump sizes on the performance of the estimators, we consider three scenarios: A. the case with no jumps ($\sigma_Y = 0.0$); B. the case with jumps of a relatively small size ($\sigma_Y = 0.5$); C. the case with jumps of a relatively large size ($\sigma_Y = 1.5$). All other parameters are kept the same as in Section 5.1.

Table 3 reports the estimation results of $\hat{\sigma}^2_{PATKV}$. For comparison, we also report the results of $\hat{\sigma}^2_{TKV}$, which is robust to jumps but does not account for the market microstructure noise (Yu et al. (2014)). As can be seen in Table 3, $\hat{\sigma}^2_{PATKV}$ is robust to jumps with both small and large jump sizes. As the sampling frequency increases, the IMSE of $\hat{\sigma}^2_{PATKV}$ improves. In contrast, the estimation error of $\hat{\sigma}^2_{TKV}$ increases sharply with the sampling frequency for a given level of microstructure noise, regardless of jump size.

When the sampling frequency is low and fixed, however, $\hat{\sigma}^2_{PATKV}$ may be less efficient than $\hat{\sigma}^2_{TKV}$. As can be seen in Table 3, if the sampling frequency is 5 minutes or lower, the IMSE of $\hat{\sigma}^2_{TKV}$ is smaller than that of $\hat{\sigma}^2_{PATKV}$ when $\sigma_\epsilon = 0.025$ or 0.035. For larger noise $\sigma_\epsilon = 0.05$, the sampling frequency needs to be 10 minutes or higher for $\hat{\sigma}^2_{PATKV}$ to outperform $\hat{\sigma}^2_{TKV}$.

Table 3: The IMSEs of $\hat{\sigma}^2_{PATKV}$ and $\hat{\sigma}^2_{TKV}$: the case with jumps of finite activity

Frequency	$\sigma_\epsilon = 0.025$		$\sigma_\epsilon = 0.035$		$\sigma_\epsilon = 0.05$	
	$\hat{\sigma}^2_{PATKV}$	$\hat{\sigma}^2_{TKV}$	$\hat{\sigma}^2_{PATKV}$	$\hat{\sigma}^2_{TKV}$	$\hat{\sigma}^2_{PATKV}$	$\hat{\sigma}^2_{TKV}$
Scenario A: Diffusion with no jumps $\sigma_Y = 0.0$						
10 sec	0.0201	4.5198	0.0153	15.089	0.0105	37.257
30 sec	0.0292	0.6206	0.0268	1.9994	0.0230	6.6617
1 min	0.0519	0.2236	0.0469	0.6587	0.0392	1.8445
5 min	0.2207	0.1257	0.2063	0.1924	0.1801	0.2864
10 min	0.2426	0.1541	0.2335	0.1690	0.2177	0.2078
Scenario B: Diffusion with small jumps $\sigma_Y = 0.5$						
10 sec	0.0184	4.476	0.0178	14.77	0.0108	37.48
30 sec	0.0307	0.5958	0.0291	2.095	0.0214	6.241
1 min	0.0489	0.2338	0.0479	0.6162	0.0397	3.041
5 min	0.2361	0.1519	0.2187	0.1589	0.1830	0.4135
10 min	0.2451	0.1740	0.2363	0.1808	0.2212	0.1977
Scenario C: Diffusion with large jumps $\sigma_Y = 1.5$						
10 sec	0.0191	4.507	0.0171	14.68	0.0112	36.37
30 sec	0.0319	0.6180	0.0271	1.938	0.0231	6.313
1 min	0.0490	0.2141	0.0489	0.6623	0.0402	3.025
5 min	0.2228	0.1205	0.2087	0.1613	0.1787	2.102
10 min	0.2460	0.1615	0.2316	0.1664	0.2205	0.2053

It is also interesting to observe that for a given sampling frequency, the higher the level of microstructure noise, the smaller the IMSEs of $\hat{\sigma}^2_{PATKV}$. For example, for data sampled every 10 seconds, the IMSEs of $\hat{\sigma}^2_{PATKV}$ in the small jump size scenario (Scenario B), where $\sigma_\epsilon = 0.025, 0.035$, and 0.05, are 0.0184, 0.0178, and 0.0108, respectively. The reason for this seemingly contradicting phenomenon is that in $\hat{\sigma}^2_{PATKV}$, "large" microstructure noises are identified as "jumps" and are removed when the threshold is applied. This is in complete contrast to the estimator $\hat{\sigma}^2_{PAKV}$, which is developed by assuming that jumps are absent: for any given sampling frequency, the performance of $\hat{\sigma}^2_{PAKV}$ gradually deteriorates as the the level of market microstructure noise increases (see Table 2).

Finally, we evaluate the quality of the finite sample performance of $\hat{\sigma}^2_{PATKV}$ in the presence of jumps with infinite activity. We consider the case that the jump component is modeled by a Variance Gamma (VG) process, which is a pure jump process with infinite activity and finite variation. More specifically, J_t is given by $cG_t+\eta W_{G_t}$, i.e., a composition of Brownian motion with drift and an independent Gamma process G. For each t, G_t at time t follows a gamma distribution: $G_t \sim Gamma(t/b, b)$, where c and η are constants. Now, the jump diffusion model (13) and (14) can be written as

$$dX_t = udt + \exp[\beta_0 + \beta_1 v_t]dW_{1,t} + cG_t + \eta W_{G_t} \tag{15}$$

$$dv_t = \alpha_v v_t dt + dW_{2,t}. \tag{16}$$

As did in Mancini (2009) and Madan (2001), we let $b = 0.23$, $c = -0.2$, and $\eta = 0.2$.

Table 4: The IMSEs of $\hat{\sigma}^2_{PATKV}$ and $\hat{\sigma}^2_{TKV}$: the case with jumps of infinite activity

Frequency	$\sigma_\epsilon = 0.025$		$\sigma_\epsilon = 0.035$		$\sigma_\epsilon = 0.05$	
	$\hat{\sigma}^2_{PATKV}$	$\hat{\sigma}^2_{TKV}$	$\hat{\sigma}^2_{PATKV}$	$\hat{\sigma}^2_{TKV}$	$\hat{\sigma}^2_{PATKV}$	$\hat{\sigma}^2_{TKV}$
10 sec	0.0187	4.4632	0.0138	15.03	0.0107	36.91
30 sec	0.0314	0.6230	0.0256	1.9860	0.0219	6.2279
1 min	0.0524	0.2312	0.0473	0.6482	0.0419	2.3760
5 min	0.2340	0.1223	0.2139	0.1567	0.1918	1.6820
10 min	0.2498	0.1468	0.2376	0.1614	0.2270	0.1872

Table 4 contains the results corresponding to the IMSEs of $\hat{\sigma}^2_{PATKV}$ and $\hat{\sigma}^2_{TKV}$ for the case with jumps of infinite activity. The finite sample performance of the jump-robust version, $\hat{\sigma}^2_{PATKV}$, is very similar to that of the case with jumps of finite activity. In particular, $\hat{\sigma}^2_{PATKV}$ provides consistent estimates of spot volatility: the higher the sampling frequency, the smaller the IMSE. Again, $\hat{\sigma}^2_{TKV}$ fails to provide consistent estimates: as the sampling frequency increases, the IMSE increases sharply. This result is hardly surprising as the variance of the observed series contains microstructure noise and is severely biased when data are sampled at frequencies higher than 5 minutes. However, if the sampling frequency is 5 minutes or lower, $\hat{\sigma}^2_{TKV}$ may outperform $\hat{\sigma}^2_{PATKV}$, as seen in the case of jumps with finite activity.

6 CONCLUSIONS

Since Foster and Nelson (1996) and Andreou and Ghysels (2002), among others, substantial attention has been devoted to the use of nonparametric estimation in spot volatility estimation. In particular, recent works by Mancini, Mattiussi and Renò (2012) and Zu and Boswijk (2014) provide theoretical justifications for using high-frequency data contaminated with microstructure noise to consistently estimate spot volatility of the efficient price process.

This paper introduces a general class of kernel-based estimators of spot volatility. The proposed estimators are robust to both microstructure noise and Lévy jumps with finite or infinite activity. The estimators are asymptotically normally distributed and have fast convergence rates. Monte Carlo simulations are conducted to study the finite sample properties of our estimators.

We demonstrate that in the context of spot volatility estimation, one should always use all available data and model jumps no matter whether jumps are present or not. If the sampling frequency is higher than 5 minutes, the noise- and jump-robust estimator, $\hat{\sigma}^2_{PATKV}$, performs well, particularly when the level of market microstructure noise is high. If the sampling frequency is 5 minutes or lower, one should use the simpler jump-robust estimator $\hat{\sigma}^2_{TKV}$, since $\hat{\sigma}^2_{TKV}$ provides a better bias and variance trade-off for data sampled at 5 minutes or lower frequencies.

References

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

Aït-Sahalia, Y., and J. Jacod (2009): "Testing for Jumps in a Discretely Observed Process," *Annals of Statistics*, 37 (1), 184-222.

Aït-Sahalia, Y., and J. Jacod (2011): "Testing Whether Jumps Have Finite or Infinite Activity," *Annals of Statistics*, 39 (3), 1689-1719.

Aït-Sahalia, Y., P. A. Mykland, and L. Zhang (2005): "How Often to Sample a Continuous-Time Process in the Presence of Market Microstructure Noise," *Review of Financial Studies*, 18 (2), 351-416.

Andersen, T. G., L. Benzoni, and J. Lund (2002): "An Empirical Investigation of Continuous-Time Equity Return Models," *Journal of Finance*, 57 (3), 1239-1284.

Andersen, T. G., T. Bollerslev, and F. X. Diebold (2007): "Roughing It Up: Including Jump Components in the Measurement, Modeling, and Forecasting of Return Volatility," *Review of Economics and Statistics*, 89 (4), 701-720.

Andersen, T. G., T. Bollerslev, F. X. Diebold, and P. Labys (2000): "Great Realizations," *Risk*, 13, 105-108.

Andreou, E., and E. Ghysels (2002): "Rolling-Sample Volatility Estimators: Some New Theoretical, Simulation, and Empirical Results," *Journal of Business and Economic Statistics*, 20 (3), 363-376.

Bandi, F. M., and J. R. Russell (2008): "Microstructure Noise, Realized Variance and Optimal Sampling," *Review of Economic Studies*, 75 (2), 339-369.

Bandi, F. M., and Renò, R. (2010): "Nonparametric Stochastic Volatility," Working paper.

Barndorff-Nielsen, O. E., and N. Shephard (2002): "Estimating Quadratic Variation Using Realized Variance," *Journal of Applied Econometrics*, 17 (5), 457-477.

Barndorff-Nielsen, O. E., and N. Shephard (2004): "Power and Bipower Variation with Stochastic Volatility and Jumps," *Journal of Financial Econometrics*, 2 (1), 1-37.

Barndorff-Nielsen, O. E., and N. Shephard (2006): "Econometrics of Testing for Jumps in Financial Economics Using Bipower Variation," *Journal of Financial Econometrics*, 4 (1), 1-30.

Barndorff-Nielsen, O. E., P. R. Hansen, A. Lunde, and N. Shephard (2008): "Designing Realised Kernels to Measure Ex-Post Variation of Equity Prices in the Presence of Noise," *Econometrica*, 76(6), 1481-1536.

Carr, P., and L. R. Wu (2003): "What Type of Process Underlies Options? A Simple Robust Test," *Journal of Finance*, 58 (6), 2581-2610.

Carr, P., and L. R. Wu (2004): "Time-Changed Lévy Processes and Option Pricing," *Journal of Financial Economics*, 71 (1), 113-141.

Carr, P., and L. R. Wu (2007): "Stochastic Skew in Currency Options," *Journal of Financial Economics*, 86 (1), 213-247.

Chernov, M., A. R. Gallant, E. Ghysels, and G. Tauchen (2003): "Alternative Models for Stock Price Dynamics," *Journal of Econometrics*, 116 (1-2), 225-257.

Cont, R., and C. Mancini (2011): "Nonparametric Tests for Pathwise Properties of Semimartingales," *Bernoulli*, 17 (2), 781-813.

Fan, Y. Y., and J. Q. Fan (2011): "Testing and Detecting Jumps Based on a Discretely Observed Process," *Journal of Econometrics*, 164 (2), 331-344.

Fan, J. Q., and Q. W. Yao (1998): "Efficient Estimation of Conditional Variance Functions in Stochastic Regression," *Biometrika*, 85 (3), 645-660.

Fang, Y. (1996): "Volatility Modeling and Estimation of High-Frequency Data with Gaussian Noise," Doctoral thesis, MIT, Sloan School of Management.

Foster, D. P., and D. B. Nelson (1996): "Continuous Record Asymptotics for Rolling Sample Variance Estimators," *Econometrica*, 64 (1), 139-174.

Genon-Catalot, V., C. Laredo, and D. Picard (1992): " Non-Parametric Estimation of the Diffusion Coefficient by Wavelets Methods," *Scandinavian Journal of Statistics*, 19 (4), 317-335.

Hansen, P. R., and A. Lunde (2006): "Realized Variance and Market Microstructure Noise," *Journal of Business and Economic Statistics*, 24 (2), 127-161.

Heston, S. L. (1993): "A Closed-Form Solution for Options with Stochastic Volatility with Applications to Bond and Currency Options," *Review of Financial Studies*, 6 (2), 327-343.

Hoffmann, M., A. Munk, and J. Schmidt-Hieber (2010): "Nonparametric Estimation of the Volatility under Microstructure Noise: Wavelet Adaptation," Working Paper.

Hull, J., and A. White (1987): "The Pricing of Options on Assets with Stochastic Volatilities" *Journal of Finance*, 42 (2), 281-300.

Huang, X., and G. Tauchen (2005): "The Relative Contribution of Jumps to Total Price Variance," *Journal of Financial Econometrics*, 3 (4), 456-499.

Jacod, J. (1994): "Limit of Random Measures Associated with the Increments of a Brownian Semimartingale," Technical Report, Université Paris VI.

Jacod, J., and P. Protter (1998): "Asymptotic Error Distributions for the Euler Method for Stochastic Differential Equations," *Annals of Probability*, 26 (1), 267-307.

Jacod, J., Y. Li, P. A. Mykland, M. Podolskij, and M. Vetter (2009): "Microstructure Noise in the Continuous Case: The Pre-Averaging Approach," *Stochastic Processes and their Applications*, 119 (7), 2249-2276.

Jing, B. Y., Z. Liu, and X. B. Kong (2014): "On the Estimation of Integrated Volatility with Jumps and Microstructure Noise," *Journal of Business and Economic Statistics*, 32 (3), 457-467.

Jing, B. Y., X. B. Kong, and Z. Liu (2011): "Estimating the Jump Activity Index under Noisy Observations Using High-Frequency Data," *Journal of the American Statistical Association*, 106 (494), 558-568.

Kanaya, S., Kristensen, D. (2010): "Estimation of Stochastic Volatility Models by Non-Parametric Filtering," Working paper.

Karatzas, I., and S. E. Shreve (1999). *Brownian Motion and Stochastic Calculus*. Springer, New York.

Kristensen, D. (2010): "Nonparametric Filtering of the Realized Spot Volatility: A Kernel-Based Approach," *Econometric Theory*, 26 (1), 60-93.

LEE, S.S., J. HANNIG,(2010):"Detecting Jumps from Lévy Jump Diffusion Processes," *Journal of Financial Economics*, 96, 271-290.

LEE, S.S., AND P. A. MYKLAND (2012): "Jumps in Equilibrium Prices and Market Microstructure Noise," *Journal of Econometrics*, 168 (2), 396-406.

LI, H. T., M. T. WELLS, AND C. L. YU (2008): "A Bayesian Analysis of Return Dynamics with Lévy Jumps," *Review of Financial Studies*, 21 (5), 2345-2378.

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, AND R. RENÒ (2012): "Spot Volatility Estimation Using Delta Sequences," Working paper.

MÜLLER, H-G., R. SEN, AND U. STADTMÜLLER (2011): "Functional Data Analysis for Volatility," *Journal of Econometrics*, 165 (2), 233-245.

MYKLAND, P. A., AND L. ZHANG (2006): "ANOVA for Diffusions and Itô Processes," *Annals of Statistics*, 34 (4), 1931-1963.

PODOLSKIJ, M. AND M. VETTER (2009): "Estimation of Volatility Functionals in the Simultaneous Presence of Microstructure Noise and Jumps," *Bernoulli*, 15 (3), 634-658.

REVUZ, D., AND M. YOR (2001). *Continuous Martingales and Brownian Motion*. Springer, New York.

STANTON, R. (1997): "A Nonparametric Model of Term Structure Dynamics and the Market Price of Interest Rate Risk," *Journal of Finance*, 52 (5), 1973-2002.

YU, C., Y. FANG, Z. LI, B. ZHANG, AND X. J. ZHAO (2014): "Threshold Kernel Estimation of Time-Dependent Spot Volatility in Jump-Diffusion Models," *Journal of Time Series Analysis*, 35 (6), 572-591.

ZHANG, L., P. A.MYKLAND, AND 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.

ZHANG, L. (2006): "Efficient Estimation of Stochastic Volatility Using Noisy Observations: A Multi-Scale Approach," *Bernoulli*, 12 (6), 1019-1043.

ZHANG, S. P., AND R.J. KARUNAMUNI (1998): "On Kernel Density Estimation Near Endpoints," *Journal of Statistical Planning and Inference*, 70 (2), 301-316.

ZHOU, B. (1996): "High-Frequency Data and Volatility in Foreign-Exchange Rates," *Journal of Business and Economic Statistics*, 14 (1), 45-52.

Zu, Y., and H. P. Boswijk (2014): "Estimating Spot Volatility with High-Frequency Financial Data," *Journal of Econometrics*, 181 (2), 117-135.

Appendix A: Preliminaries

We follow the notations of Jacod et al. (2009), and use the idea of that paper to obtain results of Lemma 1, which are the results of Lemmas 5.1 to 5.3 of Jacod et al. (2009) adapted to our setups.

For convenience, we use the shorthand notation g_i^n for $g(i/k_n)$, and we set $h_i^n = g_{i+1}^n - g_i^n$, $\bar{g} = \int_0^1 g(s)ds$. Define, on $\mathbb{R}_+$,

$$\begin{aligned}
&\phi_1(s) = \int_s^1 g'(u)g'(u-s)du, \;\; \phi_2(s) = \int_s^1 g(u)g(u-s)du, \text{ when } s \in [0,1], \\
&\phi_1(s) = 0, \;\; \phi_2(s) = 0, \text{ when } s > 1, \text{and} \\
&\Phi_{i,j} = \int_0^1 \phi_i(s)\phi_j(s)ds, \;\; \psi_i = \phi_i(0), \text{ for } i,j = 1,2.
\end{aligned}$$

For a given process $V = (V_t)_{t\geq 0}$, we write $V_i^n = V_{i\Delta_n}$, $\Delta_i^n V = V_i^n - V_{i-1}^n$, and $\bar{V}_i^n = \sum_{j=1}^{k_n-1} g_j^n \Delta_{i+j}^n V = -\sum_{j=0}^{k_n-1} h_j^n V_{i+j}^n$.

In the following, L denotes a constant, which may change from line to line and depend on $\sup_n k_n^2 \Delta_n$ and the bounds of various processes used in the proofs. We write it L_r if it depends on an additional parameter r. We also write $O_u(x)$ for a (possibly random) quantity smaller than Lx.

Unless otherwise stated, $p \geq 1$ denotes an integer and $q > 0$ a real number. For each n,

$$g_n(s) = \sum_{j=1}^{k_n-1} g_j^n I_{(j-1\Delta_n, j\Delta_n]}(s),$$

which is bounded uniformly in n and vanishes for $s > (k_n - 1)\Delta_n$ and $s \leq 0$.

Define the processes $X(n,s)_t$ and $C(n,s)_t$ as

$$\begin{aligned}
X(n,s)_t &= \int_0^t b_u g_n(u-s)du + \int_0^t \sigma_u g_n(u-s)dW_u \\
C(n,s)_t &= \int_0^t \sigma_u^2 (g_n(u-s))^2 du.
\end{aligned}$$

Both $X(n,s)_t$ and $C(n,s)_t$ are constant in time for $t \geq s + (k_n-1)\Delta_n$ but vanish for $t \leq s$. We also use the following shorthand notations:

$$\overline{X}_i^n = X(n, i\Delta_n)_{(i+k_n)\Delta_n},$$

and

$$c_i^n = C(n, i\Delta_n)_{(i+k_n)\Delta_n},$$

which equals to $\sum_{j=1}^{k_n-1} (g_j^n)^2 \Delta_{i+j}^n C$, where process $C = (C_t)_{t\geq 0}$, and $C_t = \int_0^t \sigma_s^2 ds$.

We further let

$$A_{i,j}^n = \sum_{m=i\vee j}^{i\wedge j+k_n-1} h_{m-i}^n h_{m-j}^n \alpha_m^n, \;\; A_i^n = A_{i,i}^n = \sum_{m=0}^{k_n-1} (h_m^n)^2 \alpha_{i+m}^n,$$

$$\begin{aligned}
&\widetilde{Z}_i'^n = (\overline{Z}_i^n)^2 - A_i^n - c_i^n, && \zeta(Z,p)_i^n = \sum_{j=i}^{i+pk_n-1} w_j \widetilde{Z}_j'^n, \\
&\zeta(X,p)_i^n = \sum_{j=i}^{i+pk_n-1} w_j \left((\overline{X}_j^n)^2 - c_j^n\right), && \zeta(W,p)_i^n = \sum_{j=i}^{i+pk_n-1} w_j \left((\sigma_i^n \overline{W}_j^n)^2 - c_j^n\right).
\end{aligned}$$

In addition, for any process V, let

$$\zeta'(V,p)_i^n = \sum_{(j,m):i\leq j\leq m\leq i+pk_n-1} w_j w_m \overline{V}_j^n \overline{V}_m^n \phi_1\left(\frac{m-j}{k_n}\right)$$

$$\zeta''(V)_i^n = (\overline{V}_i^n)^2 \sum_{j=i+k_n}^{i+2k_n-1} (\Delta_j^n V)^2.$$

We will consider the discrete time filtrations $\mathcal{F}_j^n = \mathcal{F}_{j\Delta_n}^{(0)} \otimes \mathcal{F}_{j\Delta_n-}^{(1)}$ and $\mathcal{F}_j'^n = \mathcal{F}^{(0)} \otimes \mathcal{F}_{j\Delta_n-}^{(1)}$ for $j \in \mathbb{N}$. We set $\mathcal{G}(p)_j^n = \mathcal{F}_{j(p+1)k_n}^n$ and $\mathcal{G}'(p)_j^n = \mathcal{F}_{j(p+1)k_n+pk_n}^n$, along with the following variables:

$$\begin{aligned}
\eta(p)_j^n &= \tfrac{\sqrt{\Delta_n}}{\theta\psi_2}\zeta(Z,p)_{j(p+1)k_n}^n, \quad \overline{\eta}(p)_j^n = E(\eta(p)_j^n|\mathcal{G}(p)_j^n),\\
\eta'(p)_j^n &= \tfrac{\sqrt{\Delta_n}}{\theta\psi_2}\zeta(Z,1)_{j(p+1)k_n+pk_n}^n, \text{ and } \overline{\eta}'(p)_j^n = E(\eta'(p)_j^n|\mathcal{G}'(p)_j^n).
\end{aligned}$$

Let $j_n(p,T)$ and $i_n(p,T)$ denote $\left[\frac{T+\Delta_n}{(p+1)k_n\Delta_n}\right]-1$ and $(j_n(p,T)+1)(p+1)k_n$, respectively. Then, for all $p \geq 1$ we have the following identity as in Jacod et al. (2009):

$$\hat{C}_T^{n,w} - C_T^w = M(p)_t^n + M'(p)_T^n + F(p)_T^n + F'(p)_T^n + \widehat{C}(p)_T^n + \widehat{C}'(p)_T^n + \widehat{C}_T''^n, \tag{A.1}$$

where

$$\begin{aligned}
F(p)_t^n &= \textstyle\sum_{j=0}^{j_n(p,T)} \overline{\eta}(p)_j^n, \quad M(p)_t^n = \sum_{j=0}^{j_n(p,T)} (\eta(p)_j^n - \overline{\eta}(p)_j^n),\\
F'(p)_t^n &= \textstyle\sum_{j=0}^{j_n(p,T)} \overline{\eta}'(p)_j^n, \quad M'(p)_t^n = \sum_{j=0}^{j_n(p,T)} (\eta'(p)_j^n - \overline{\eta}'(p)_j^n),
\end{aligned}$$

$$\widehat{C}(p)_T^n = \frac{\sqrt{\Delta_n}}{\theta\psi_2} \sum_{i=i_n(p,T)}^{n-k_n+1} w_i \widetilde{Z}_i'^n,$$

$$\widehat{C}'(p)_T^n = \frac{\sqrt{\Delta_n}}{\theta\psi_2} \sum_{i=0}^{n-k_n+1} w_i A_i^n - \frac{\psi_1\Delta_n}{2\theta^2\psi_2} \sum_{i=1}^{n} w_i(\Delta_i^n Z)^2,$$

and

$$\widehat{C}_T''^n = \frac{\sqrt{\Delta_n}}{\theta\psi_2} \sum_{i=0}^{n-k_n+1} w_i c_i^n - C_T^w.$$

Finally, let

$$\begin{aligned}
\beta(p)_i^n &= \sup\nolimits_{s,t\in[i\Delta_n,(i+(p+2)k_n)\Delta_n]} \left(|b_s - b_t| + |\sigma_s - \sigma_t| + |\alpha_s - \alpha_t|\right),\\
\chi(p)_i^n &= \Delta_n^{1/4} + \sqrt{E((\beta(p)_i^n)^2|\mathcal{F}_i^n)},\\
\Xi_{ij} &= -\textstyle\int_0^1 s\phi_i(s)\phi_j(s)ds.
\end{aligned}$$

Lemma 1 We have

$$E\left((\zeta(W,p)_i^n)^2|\mathcal{F}_i^n\right) = 4(pw_{i*}^2\Phi_{22} + w_i^2\Xi_{22})k_n^4\Delta_n^2(\sigma_i^n)^4 + O_u(p^2\chi(p)_i^n),$$

$$E\left((\zeta'(W,p)_i^n)|\mathcal{F}_i^n\right) = (pw_{i*}^2\Phi_{12} + w_i^2\Xi_{12})k_n^3\Delta_n + O_u(p\Delta_n^{-1/4}),$$

$$|E(\zeta(X,p)_i^n|\mathcal{F}_i^n)| \leq L_{p,w}\Delta_n^{-1/4}\chi(p)_i^n,$$

$$\left|E\left((\zeta(X,p)_i^n)^2|\mathcal{F}_i^n\right) - 4(pw_{i*}^2\Phi_{22} + w_i^2\Xi_{22})k_n^4\Delta_n^2(\sigma_i^n)^4\right| \leq L_{p,w}\chi(p)_i^n,$$

$$\left|E\left((\zeta'(X,p)_i^n)|\mathcal{F}_i^n\right) - (pw_{i*}^2\Phi_{12} + w_i^2\Xi_{12})k_n^3\Delta_n(\sigma_i^n)^2\right| \leq L_{p,w}\Delta_n^{-1/2}\chi(p)_i^n,$$

where $i* \in (i, i+pk_n)$ satisfies the mean value theorem $\int_{t_i}^{t_{i+pk_n}} w^2(s)ds = w^2(t_{i*})pk_n\Delta_n$.

Proof: Since the weighting function $w(s)$ is bounded, the arguments used in the proofs of Lemmas 5.1 - 5.3 of Jacod et al. (2009) work in the present context, so we omit the proof.

Appendix B: Proofs

Proof of Theorem ??.

Given the identity of (A.1), the proof of Theorem **??** hinges on the asymptotic properties of the six terms on the right-hand side of (A.1), which are evaluated one by one.

First, for any fixed $p \geq 1$, we have $\Delta_n^{-1/4}F(p)_T^n \xrightarrow{P} 0$ and $\Delta_n^{-1/4}F'(p)_T^n \xrightarrow{P} 0$. Since the weighting function $w(s)$ is bounded, in a similar way as proving (5.40) in Jacod et al. (2009), we have $E(\zeta(Z,p)_i^n|\mathcal{F}_i'^n) = \zeta(X,p)_i^n$, $E((\zeta(Z,p)_i^n)^4|\mathcal{F}_i^n) \leq L_{p,w}$, and $|E(\zeta(Z,p)_i^n|\mathcal{F}_i'^n)| \leq L_{p,w}\Delta_n^{1/4}\chi(p)_i^n$. Combining those results with Lemma 1 in Appendix A and Lemma 5.4 in Jacod et al. (2009) yields the desired results.

Second, since $E(\widetilde{Z}_i'^n|\mathcal{F}_i'^n) = (\overline{X}_i^n)^2 - c_i^n$, $|c_i^n| \leq L\sqrt{\Delta_n}$, and $E(|\overline{X}_i^n|^q|\mathcal{F}_i^n) \leq L_q\Delta_n^{q/4}$, we have

$$\begin{aligned}
E(\Delta_n^{-1/4}\widehat{C}(p)_T^n|\mathcal{F}_i^n) &= E(E(\Delta_n^{-1/4}\frac{\sqrt{\Delta_n}}{\theta\psi_2}\sum_{i=i_n(p,T)}^{n-k_n+1} w_i\widetilde{Z}_i'^n|\mathcal{F}_i'^n)|\mathcal{F}_i^n) \\
&= \Delta_n^{-1/4}\frac{\sqrt{\Delta_n}}{\theta\psi_2}\sum_{i=i_n(p,T)}^{n-k_n+1} w_iE(((\overline{X}_i)^2 - c_i^n)|\mathcal{F}_i^n) \\
&\leq \Delta_n^{-1/4}\frac{\sqrt{\Delta_n}}{\theta\psi_2}L_w\frac{L_p}{\sqrt{\Delta_n}}L_q\sqrt{\Delta_n} \to 0.
\end{aligned}$$

Then, we readily deduce the result that $\Delta_n^{-1/4}\widehat{C}(p)_T^n \xrightarrow{P} 0$.

Now we proceed with term $\Delta_n^{-1/4}\hat{C}_T''^n$. By mean value theorem, for $s_1 \in [(j+i-1)\Delta_n, (j+i)\Delta_n]$ and $s_2 \in [j\Delta_n, (j+1)\Delta_n]$, we have

$$\begin{aligned}
\sum_{i=0}^{n-k_n+1} w_ic_i^n &= \sum_{i=1}^{k_n-1}(g_i^n)^2\sum_{j=i}^{i+n-k_n+1} w_{j-i}\Delta_j^nC \\
&= \sum_{i=1}^{k_n-1}(g_i^n)^2\sum_{j=i}^{i+n-k_n+1} w_{j-i}(\Delta_{j-i+1}^nC + \Delta_j^nC - \Delta_{j-i+1}^nC) \\
&= \sum_{i=1}^{k_n-1}(g_i^n)^2\sum_{j=0}^{n-k_n+1} w_j\Delta_{j+1}^nC + \sum_{i=1}^{k_n-1}(g_i^n)^2\sum_{j=0}^{n-k_n+1} w_{j-i}(\Delta_j^nC - \Delta_{j-i+1}^nC)
\end{aligned}$$

$$= \sum_{i=1}^{k_n-1}(g_i^n)^2\sum_{j=0}^{n-k_n+1} w_j\int_{j\Delta_n}^{(j+1)\Delta_n}\sigma_s^2ds + \sum_{i=1}^{k_n-1}(g_i^n)^2\sum_{j=0}^{n-k_n+1} w_j(\Delta_{j+i}^nC - \Delta_{j+1}^nC)$$

$$= \sum_{i=1}^{k_n-1}(g_i^n)^2\left(\int_0^t w(s)\sigma_s^2 ds + O(\Delta_n + k_n\Delta_n)\right) + \sum_{i=1}^{k_n-1}(g_i^n)^2\sum_{j=0}^{n-k_n+1} w_j\left(\sigma_{s_1}^2 - \sigma_{s_2}^2\right)\Delta_n.$$

Denote $a_n = \sum_{j=1}^{k_n-1}(g_j^n)^2$. Then we have $a_n = k_n\psi_2 + O_u(1)$. In addition, since $|\sigma_{s1}^2 - \sigma_{s2}^2| \le k_n\Delta_n$, $\Delta_n^{-1/4}\hat{C}_T''^{n} \xrightarrow{P} 0$ holds.

Next, we show that $\Delta_n^{-1/4}\hat{C}'(p)_T^n \xrightarrow{P} 0$. Let $\zeta_i^n = w_i((\Delta_i^n Z)^2 - (\alpha_{i+1}^n + \alpha_i^n))$. For $1 \le i \le j-2$, we have $E(\zeta_i^n) = E(w_i(\Delta_i^n X)^2) = O_u(\Delta_n)$, $E(\zeta_i^n\zeta_j^n) = E(w_i w_j(\Delta_i^n X)^2(\Delta_j^n X)^2) = O_u(\Delta_n^2)$, and $E(|\zeta_i^n|^2) \le L$, as in Lemma 5.6 of Jacod et al. (2009). Obviously, $E((\sum_{i=1}^n \zeta_i^n)^2) \le L/\Delta_n$. Thus, we have $G_n := \frac{\psi\Delta_n^{3/4}}{2\theta^2\psi_2}\sum_{i=1}^n \zeta_i^n \xrightarrow{P} 0$. Now, it suffices to show that $\frac{1}{\Delta_n^{1/4}}\hat{C}'(p)_T^n + G_n \xrightarrow{P} 0$. We write $\frac{1}{\Delta_n^{1/4}}\hat{C}'(p)_T^n + G_n = U_n + V_n$, where

$$U_n = \left(\frac{\Delta_n^{1/4}}{\theta\psi_2}\left(\sum_{l=0}^{k_n-1}(h_l^n)^2\right) - \frac{\psi_1\Delta_n^{3/4}}{\theta^2\psi_2}\right)\sum_{i=k_n}^{i_n(p,T)-1} w_i\alpha_i^n + \frac{\Delta_n^{1/4}}{\theta\psi_2}\sum_{l=0}^{k_n-1}(h_l^n)^2\sum_{i=k_n}^{i_n(p,T)-1}(w_{i-l} - w_i)\alpha_i^n$$

and

$$\begin{aligned} V_n &= \frac{\Delta_n^{1/4}}{\theta\psi_2}\left(\sum_{i=0}^{k_n-1}\alpha_i^n\sum_{l=0}^{i} w_{i-l}(h_l^n)^2 + \sum_{i=i_n(p,T)}^{i_n(p,T)+k_n-2}\alpha_i^n\sum_{l=i+1-i_n(p,T)}^{k_n-1} w_{i-l}(h_l^n)^2\right) \\ &\quad -\frac{\psi_1\Delta_n^{3/4}}{2\theta^2\psi_2}\left(w_0\alpha_0^n + 2\sum_{i=1}^{k_n-1}(w_i + w_{i+1})\alpha_i^n + 2\sum_{i=i_n(p,T)}^{n-1}(w_i + w_{i+1})\alpha_i^n + w_n\alpha_n^n\right). \end{aligned}$$

Note that $|V_n| \to 0$, because α_t and w_i are bounded, and $|h_l^n| \le L\sqrt{\Delta_n}$. The result that $U_n \to 0$ pointwise follows from the fact that $\sum_{l=0}^{k_n-1}(h_l^n)^2 = \frac{\psi_1}{k_n} + O(\Delta_n)$, whereas $\sum_{i=k_n}^{i_n(p,T)-1} w_i\alpha_i^n \le L/\Delta_n$ and $\sum_{i=k_n}^{i_n(p,T)-1}(w_{i-l} - w_i)\alpha_i^n \le k_n$.

Finally, we evaluate the two remaining terms $M(p)_T^n$ and $M'(p)_T^n$, which are sums of martingale differences. In a similar way as in Jacod et al. (2009), we have

$$\begin{aligned} &E\left(\left(\zeta(Z,p)_i^n\right)^2|\mathcal{F}_i'^n\right) = \sum_{i\le j,m\le i+pk_n-1} w_j w_m E\left(\widetilde{Z}_j'^n\widetilde{Z}_m'^n|\mathcal{F}_i'^n\right) \\ = &\sum_{i\le j,m\le i+pk_n-1} w_j w_m\left(\left(\overline{X}_j^n\right)^2 - c_j^n\right)\left(\left(\overline{X}_m^n\right)^2 - c_m^n\right) + \sum_{i\le j,m\le i+pk_n-1} w_j w_m\overline{X}_j^n\overline{X}_m^n A_{j,m}^n \\ &+\sum_{i\le j,m\le i+pk_n-1} 2w_j w_m\left(A_{j,m}^n\right)^2 + \sum_{i\le j,m\le i+pk_n-1} \mathrm{O}_u\left(\Delta_n^{3/2} + \Delta_n|\overline{X}_j^n|\Delta_n|\overline{X}_m^n|\right) \\ = &\left(\sum_{j=i}^{i+pk_n-1} w_j\left(\left(\overline{X}_j^n\right)^2 - c_j^n\right)\right)^2 + \frac{4}{k_n}\sum_{i\le j,m\le i+pk_n-1} w_j w_m\overline{X}_j^n\overline{X}_m^n\left(\alpha_i^n\phi_1\left(\frac{m-j}{k_n}\right)\right) \\ &+\sum_{i\le j,m\le i+pk_n-1} 2w_j w_m\left(A_{j,m}^n\right)^2 + \sum_{i\le j,m\le i+pk_n-1} w_j w_m\overline{X}_j^n\overline{X}_m^n\mathrm{O}_u(p\Delta_n + \sqrt{\Delta_n}\beta(p)_i^n) \\ &+\sum_{i\le j,m\le i+pk_n-1} \mathrm{O}_u\left(\Delta_n^{3/2} + \Delta_n|\overline{X}_j^n|\Delta_n|\overline{X}_m^n|\right) \\ = &\left(\zeta(X,p)_i^n\right)^2 + \frac{8}{k_n}\alpha_i^n\zeta'(X,p)_i^n + 4(\alpha_i^n)^2(pw_{i*}^2\Phi_{11} + w_i^2\Xi_{11}) \\ &+p^3\mathrm{O}_u\left(\left(\sqrt{\Delta_n} + \beta(p)_i^n\right)\left(1 + \sum_{j=i}^{i+pk_n-1}|\overline{X}_j^n|^2\right)\right). \end{aligned}$$

Thus

$$\Big|E\left(\left(\zeta(Z,p)_i^n\right)^2|\mathcal{F}_i^n\right) - 4(pw_{i*}^2\Phi_{22} + w_i^2\Xi_{22})k_n^4\Delta_n^2(\sigma_i^n)^4$$
$$-8\alpha_i^n(\sigma_i^n)^2(pw_{i*}^2\Phi_{12} + w_i^2\Xi_{12})k_n^2\Delta_n - 4(\alpha_i^n)^2(pw_{i*}^2\Phi_{11} + w_i^2\Xi_{11})\Big| \le L_{p,w}\chi(p)_i^n, \tag{B.1}$$

and hence

$$E\left(\sup_{s\le T}\left|M'(p)_s^n\right|^2\right) \le \frac{LT}{p}\sqrt{\Delta_n}.$$

Now we turn to the last term in (A.1), $M(p)_T^n$. Again, since w_i is bounded, (5.57) and (5.58) in Jacod et al. (2009) also hold here, and we have

$$\frac{1}{\sqrt{\Delta_n}}\sum_{j=0}^{j_n(p,T)}\left(E\left(\left(\eta(p)_j^n\right)^2|\mathcal{G}(p)_j^n\right) - \left(\overline{\eta}(p)_j^n\right)^2\right)$$
$$= \frac{\sqrt{\Delta_n}}{\theta^2\psi_2^2}\sum_{j=0}^{j_n(p,T)}E\left(\left(\zeta(Z,p)_{j(p+1)k_n}^n\right)^2|\mathcal{G}(p)_j^n\right) - \frac{\sqrt{\Delta_n}}{\theta^2\psi_2^2}\sum_{j=0}^{j_n(p,T)}\left(E\left(\zeta(Z,p)_{j(p+1)k_n}^n|\mathcal{G}(p)_j^n\right)\right)^2.$$

By a Riemann sums argument and (B.1), we readily deduce the following convergence results:

$$\frac{1}{\sqrt{\Delta_n}}\sum_{j=0}^{j_n(p,T)}\left(E\left(\left(\eta(p)_j^n\right)^2|\mathcal{G}(p)_j^n\right) - \left(\overline{\eta}(p)_j^n\right)^2\right) \xrightarrow{P} \int_0^T w_s^2\gamma(p)_s^2 ds.$$

Then we obtain the result that for any fixed $p \ge 2$, the sequence $\frac{1}{\Delta_n^{1/4}}M(p)^n$ of processes converges stably in law to

$$Y(p)_T = \int_0^T w_s\gamma(p)_s dB_s,$$

where B is a standard Wiener process independent of $\mathcal{F}$, and $\gamma(p)_t$ is the square root of

$$\gamma(p)_t^2 = \frac{4}{\psi_2^2}\left(\left(\frac{p}{p+1}\Phi_{22} + \frac{1}{p+1}\Psi_{22}\right)\theta\sigma_t^4 + 2\left(\frac{p}{p+1}\Phi_{12} + \frac{1}{p+1}\Psi_{12}\right)\frac{\sigma_t^2\alpha_t}{\theta}\right.$$
$$\left.\left(\frac{p}{p+1}\Phi_{11} + \frac{1}{p+1}\Psi_{11}\right)\frac{\alpha_t^2}{\theta^3}\right).$$

This establishes the first claim of Theorem **??**. In view of the proof of (3.10) in Jacod et al. (2009), it's easy to obtain $\Gamma_T^n \xrightarrow{P} \int_0^T w_s^2\gamma_s^2 ds$, and hence the proof of Theorem **??** is complete.

Proof of Theorem 4.1.

First, we write

$$\frac{\sqrt{h}(\hat{\sigma}_{PAKV}^2(\tau) - \sigma_\tau^2)}{\Delta_n^{1/4}\sqrt{\gamma_\tau^2\int_{\mathbb{R}}K^2(s)ds}} = \frac{\hat{\sigma}_{PAKV}^2(\tau) - \int_0^T K_h(s-\tau)\sigma_s^2 ds}{\Delta_n^{1/4}\sqrt{\int_0^T K_h^2(s-\tau)\gamma_s^2 ds}} \times \frac{\sqrt{h\int_0^T K_h^2(s-\tau)\gamma_s^2 ds}}{\sqrt{\gamma_\tau^2\int_{\mathbb{R}}K^2(s)ds}}$$
$$+\frac{\sqrt{h}\left(\int_0^T K_h(s-\tau)\sigma_s^2 ds - \sigma_\tau^2\right)}{\Delta_n^{1/4}\sqrt{\gamma_\tau^2\int_{\mathbb{R}}K^2(s)ds}}. \tag{B.2}$$

For $f \in C^{m,\gamma}[0,T]$, we have by Taylor expansion that there exists $\bar{s} \in [s,\tau]$ such that

$$\begin{aligned}\int_0^T K_h(s-\tau)f(s)ds &= f(\tau)\int_0^T K_h(s-\tau)ds + \sum_{k=1}^{m-1}\frac{f^{(k)}(\tau)}{k!}\int_0^T K_h(s-\tau)(s-\tau)^k ds \\ &\quad + \int_0^T \frac{f^{(m)}(\bar{s})}{m!}K_h(s-\tau)(s-\tau)^m ds.\end{aligned}$$

Since $\int_0^T K_h(s-\tau)(s-\tau)^k ds = h^k \int_{-\tau/h}^{(T-\tau)/h} K(z)z^k dz, k = 0, \cdots, m$, and

$$\begin{aligned}&\int_0^T K_h(s-\tau)(s-\tau)^m \left(f^{(m)}(\bar{s}) - f^{(m)}(\tau)\right) ds \\ &= \int_0^T K_h(s-\tau)(s-\tau)^m \left(L_f(\tau, |\bar{s}-\tau|)|\bar{s}-\tau|^\gamma + o(|\bar{s}-\tau|)\right) ds \\ &= h^{m+\gamma}L_f(\tau,0)\int_{\mathbb{R}} K(z)z^{m+\gamma}dz + o(h^{m+\gamma}),\end{aligned}$$

we obtain

$$\int_0^T K_h(s-\tau)f(s)ds = f(\tau) + h^{m+\gamma}L_f(\tau,0)\int_{\mathbb{R}} K(z)z^{m+\gamma}dz + o(h^{m+\gamma}).$$

In a similar way, we get

$$\int_0^T K_h^2(s-\tau)f^2(s)ds = \frac{f^2(\tau)}{h}\int_{\mathbb{R}} K^2(z)dz + h^{m+\gamma-1}L_{f^2}(\tau,0)\int_{\mathbb{R}} K^2(z)z^{m+\gamma}dz + o(h^{m+\gamma-1}).$$

Combining (B.2) and the results in Theorem **??** by substituting the weighting function $K_h(s)$ for $w(s)$, the first part of Theorem 4.1 follows. In addition, with the same arguments above, we can easily show that $\hat{\gamma}^2_{PAKV}(\tau) \xrightarrow{P} \gamma^2_\tau$. Then the proof of Theorem 4.1 is complete.

Proof of Theorem 4.2.

Denote X_{0t} the continuous diffusion part of price process. Then $Z_t = X_{0t} + \epsilon_t + J_{1t} := Z_{0t} + J_{1t}$, $\bar{X}^n_{0i} = \sum_{j=1}^{k_n-1} g^n_j \Delta^n_{i+j} X_0$, and $\bar{Z}^n_{0i} = \sum_{j=1}^{k_n-1} g^n_j \Delta^n_{i+j} Z_0$. By Lévy law for modulus of continuity of Brownian motion paths (see Theorem 9.25, Karatzas and Shreve, 1999) and time-changed Brownian motion (Theorems 1.9-1.10, Revuz and Yor, 2001), we have that for small Δ_n,

$$\sup_{i\in\{i=1,\cdots,n-k_n+1\}} \frac{|\bar{X}^n_{0i}|}{\Delta_n^{1/4}\sqrt{\log\frac{1}{\Delta_n}}} \le \Lambda(\omega), \text{ a.s.}$$

In addition, the central limit theorem implies that

$$P(|\bar{\epsilon}^n_i| \le \sqrt{1/k_n}\log\frac{1}{\Delta_n}) = P(|\bar{\epsilon}^n_i| \le \Delta_n^{1/4}\log\frac{1}{\Delta_n}) = 1 - o(\Delta_n^{1/4}).$$

Then we can easily show that a.s., $\exists\Delta > 0$ such that $\forall\Delta_n \le \Delta$, $I_{\{(\bar{Z}^n_i)^2 \le r(\Delta_n)\}} = I_{\{\cap_{j=i+1}^{i+k_n-1}(\Delta_j N=0)\}}$. Hence,

$$\hat{\sigma}^2_{PATKV}(\tau) = \frac{\sqrt{\Delta_n}}{\theta\psi_2}\sum_{i=0}^{n-k_n+1} K_h(t_i-\tau)(\bar{Z}^n_{0i})^2 I_{\{(\bar{Z}^n_i)^2 \le r(\Delta_n)\}}$$

$$
\begin{aligned}
&-\frac{\psi_1\Delta_n}{2\theta^2\psi_2}\sum_{i=1}^{n}K_h(t_i-\tau)(\Delta_i^n Z_0)^2 I_{\{(\bar{Z}_i^n)^2\le r(\Delta_n)\}}\\
=\ &\hat{\sigma}^2_{PAKV}(\tau)-\frac{\sqrt{\Delta_n}}{\theta\psi_2}\sum_{i=0}^{n-k_n+1}K_h(t_i-\tau)(\bar{Z}_{0i}^n)^2 I_{\{(\bar{Z}_i^n)^2\ge r(\Delta_n)\}}\\
&+\frac{\psi_1\Delta_n}{2\theta^2\psi_2}\sum_{i=1}^{n}K_h(t_i-\tau)(\Delta_i^n Z_0)^2 I_{\{(\bar{Z}_i^n)^2\ge r(\Delta_n)\}}\\
:=\ &\hat{\sigma}^2_{PAKV}(\tau)+R_T^n(\tau).
\end{aligned}
$$

Since

$$
\Delta_n^{-1/4}\sqrt{h}R_T^n(\tau)\le \Delta_n^{1/4}\sqrt{h}\times Lk_nN_T\times\left(\sup_i(\bar{Z}_{0i}^n)^2+\Delta_n^{1/2}\sup_i(\Delta_i^n Z_0)^2\right)\xrightarrow{P}0,
$$

the first part of Theorem 4.2 follows by Theorem 4.1.

Next, we proceed to evaluate $\hat{\gamma}^2_{PATKV}(\tau)$. In fact, we can have the following decomposition

$$
\hat{\gamma}^2_{PATKV}(\tau)=\hat{\gamma}^2_{PAKV}(\tau)+R_1+R_2+R_3,
$$

where

$$
R_1=-\frac{4\Phi_{22}}{3\theta\psi_2^4}\sum_{i=0}^{n-k_n+1}K_h(t_i-\tau)(\bar{Z}_{0i}^n)^4 I_{\{(\bar{Z}_i^n)^2>r(\Delta_n)\}},
$$

$$
\begin{aligned}
R_2\ =\ &\frac{4\Delta_n}{\theta^3}\left(\frac{\Phi_{12}}{\psi_2^3}-\frac{\Phi_{22}\psi_1}{\psi_2^4}\right)\Bigg(-\sum_{i=0}^{n-2k_n+1}K_h(t_i-\tau)(\bar{Z}_{0i}^n)^2 I_{\{(\bar{Z}_i^n)^2>r(\Delta_n)\}}\sum_{j=i+k_n}^{i+2k_n-1}(\Delta_j^n Z_0)^2\\
&-\sum_{i=0}^{n-2k_n+1}K_h(t_i-\tau)(\bar{Z}_{0i}^n)^2\sum_{j=i+k_n}^{i+2k_n-1}(\Delta_j^n Z_0)^2 I_{\{(\bar{Z}_{i+k_n}^n)^2>r(\Delta_n)\}}\\
&+\sum_{i=0}^{n-2k_n+1}K_h(t_i-\tau)(\bar{Z}_{0i}^n)^2 I_{\{(\bar{Z}_i^n)^2>r(\Delta_n)\}}\sum_{j=i+k_n}^{i+2k_n-1}(\Delta_j^n Z_0)^2 I_{\{(\bar{Z}_{i+k_n}^n)^2>r(\Delta_n)\}}\Bigg),
\end{aligned}
$$

and

$$
R_3=-\frac{\Delta_n}{\theta^3}\left(\frac{\Phi_{11}}{\psi_2^2}-\frac{2\Phi_{12}\psi_1}{\psi_2^3}+\frac{\Phi_{22}\psi_1^2}{\psi_2^4}\right)\sum_{i=1}^{n-2}K_h(t_i-\tau)(\Delta_i^n Z_0)^2(\Delta_{i+2}^n Z_0)^2 I_{\{(\bar{Z}_i^n)^2>r(\Delta_n)\}}.
$$

It is easy to show that $\Delta_n^{-1/4}\sqrt{h}R_i\xrightarrow{P}0$ for $i=1,2,3$. Therefore, we have $\hat{\gamma}^2_{PATKV}(\tau)\xrightarrow{P}\gamma_\tau^2$. This concludes the proof of Theorem 4.2.

Proof of Theorem 4.3.

Denote $Z_{1t}=X_{0t}+J_{1t}+\epsilon_t$. Then we have $Z_t=Z_{1t}+J_{2t}$. Consider the following decomposition:

$$
\frac{\sqrt{\Delta_n}}{\theta\psi_2}\sum_{i=0}^{n-k_n+1}K_h(t_i-\tau)(\bar{Z}_i^n)^2 I_{\{(\bar{Z}_i^n)^2\le r(\Delta_n)\}}=\frac{\sqrt{\Delta_n}}{\theta\psi_2}\sum_{i=0}^{n-k_n+1}K_h(t_i-\tau)(\bar{Z}_{1i}^n)^2 I_{\{(\bar{Z}_{1i}^n)^2\le 4r(\Delta_n)\}}\tag{B.3}
$$

$$
\begin{aligned}
&+\frac{\sqrt{\Delta_n}}{\theta\psi_2}\sum_{i=0}^{n-k_n+1} K_h(t_i-\tau)(\bar{Z}_{1i}^n)^2\left(I_{\{(\bar{Z}_i^n)^2\le r(\Delta_n)\}}-I_{\{(\bar{Z}_{1i}^n)^2\le 4r(\Delta_n)\}}\right)\\
&+\frac{2\sqrt{\Delta_n}}{\theta\psi_2}\sum_{i=0}^{n-k_n+1} K_h(t_i-\tau)\bar{Z}_{1i}^n\bar{J}_{2i}^n I_{\{(\bar{Z}_i^n)^2\le r(\Delta_n)\}}+\frac{\sqrt{\Delta_n}}{\theta\psi_2}\sum_{i=0}^{n-k_n+1} K_h(t_i-\tau)(\bar{J}_{2i}^n)^2 I_{\{(\bar{Z}_i^n)^2\le r(\Delta_n)\}}\\
&:= I_1+I_2+I_3+I_4.
\end{aligned}
$$

Since the results of Theorem 4.2 can be applied to the term I_1, we only need to exhibit the limits of the remaining terms in (B.3) and show that they either tend to zero or to infinity.

We start with I_2. On $\{(\bar{Z}_i^n)^2\le r(\Delta_n),(\bar{Z}_{1i}^n)^2>4r(\Delta_n)\}$, we have

$$
\sqrt{r(\Delta_n)}\ge|\bar{Z}_i^n|>|\bar{Z}_{1i}^n|-|\bar{J}_{2i}^n|>2\sqrt{r(\Delta_n)}-|\bar{J}_{2i}^n|.
$$

Thus, $|\bar{J}_{2i}^n|>\sqrt{r(\Delta_n)}$. Moreover, if $|\bar{Z}_{1i}^n|>2\sqrt{r(\Delta_n)}$, we necessarily have $\Delta_j N\neq 0$ for some $j\in\{i+1,i+2,\cdots,i+k_n-1\}$. It follows that

$$
\begin{aligned}
&P\left(\frac{\sqrt{\Delta_n}}{\Delta_n^{1/4}\theta\psi_2}\sum_{i=0}^{n-k_n+1}K_h(t_i-\tau)(\bar{Z}_{1i}^n)^2 I_{\{(\bar{Z}_i^n)^2\le r(\Delta_n),(\bar{Z}_{1i}^n)^2>4r(\Delta_n)\}}\neq 0\right)\\
&\le \Delta_n^{1/4}nP(|\bar{J}_{2i}^n|>\sqrt{r(\Delta_n)},\Delta_i N\neq 0)\le \Delta_n^{1/4}nO(\Delta_n)\frac{E[(\bar{J}_{2i}^n)^2]}{r(\Delta_n)}=O\left(\frac{\Delta_n^{3/4}}{r(\Delta_n)}\right)\to 0.
\end{aligned}
$$

In addition, we have that on $\{(\bar{Z}_{1i}^n)^2\le 4r(\Delta_n)\}$, $\Delta_i^{i+k_n}N=0$ for sufficiently small Δ_n. It follows that

$$
\begin{aligned}
&\{(\bar{Z}_i^n)^2>r(\Delta_n),(\bar{Z}_{1i}^n)^2\le 4r(\Delta_n)\}\subset\{(\bar{Z}_{0i}^n+\bar{J}_{2i}^n)^2>r(\Delta_n))\}\\
&\subset\{(\bar{Z}_{0i}^n)^2>r(\Delta_n)/4\}\cup\{(\bar{J}_{2i}^n)^2>r(\Delta_n)/4\}.
\end{aligned}
$$

Thus,

$$
\begin{aligned}
&\frac{\sqrt{\Delta_n}}{\Delta_n^{1/4}\theta\psi_2}\sum_{i=0}^{n-k_n+1}K_h(t_i-\tau)(\bar{Z}_{1i}^n)^2 I_{\{(\bar{Z}_i^n)^2>r(\Delta_n),(\bar{Z}_{1i}^n)^2\le 4r(\Delta_n)\}}\\
&\le \frac{\sqrt{\Delta_n}}{\Delta_n^{1/4}\theta\psi_2}\sum_{i=0}^{n-k_n+1}K_h(t_i-\tau)(\bar{Z}_{0i}^n)^2 I_{\{(\bar{J}_{2i}^n)^2>r(\Delta_n)/4\}}\\
&\le \Delta_n^{1/4}\Delta_n^{1/2}\left(\log\frac{1}{\Delta_n}\right)^2\sum_{i=0}^{n-k_n+1}I_{\{(\bar{J}_{2i}^n)^2>r(\Delta_n)/4\}}=O\left(\Delta_n^{\frac{1}{4}-\frac{\alpha\beta}{2}}\left(\log\frac{1}{\Delta_n}\right)^2\right).
\end{aligned}
\tag{B.4}
$$

The last equality in (B.4) will follow if we prove

$$
P\left(|\bar{J}_{2i}^n|>\sqrt{r(\Delta_n)}/2\right)=O(\Delta_n^{1/2-\alpha\beta/2}).
\tag{B.5}
$$

To prove (B.5), we define $\tilde{N}_t:=\sum_{s\le t}I_{\{|\Delta J_{2s}|>\sqrt{r(\Delta_n)}/2\}}$. Hence, we have

$$
\begin{aligned}
&P\left(\left|\bar{J}_{2i}^n\right|>\sqrt{r(\Delta_n)}/2\right)\le P\left(|\sum_{j=1}^{k_n-1}\Delta_{i+j}J_2|>\sqrt{r(\Delta_n)}/2\right)\\
=\;&P\left(\Delta_i^{i+k_n-1}\tilde{N}=0,|\Delta_i^{i+k_n-1}J_2|>\sqrt{r(\Delta_n)}/2\right)+P\left(\Delta_i^{i+k_n-1}\tilde{N}\ge 1,|\Delta_i^{i+k_n-1}J_2|>\sqrt{r(\Delta_n)}/2\right)\\
\le\;&P\left(\Delta_i^{i+k_n-1}\tilde{N}=0,|\Delta_i^{i+k_n-1}J_2|>\sqrt{r(\Delta_n)}/2\right)+P\left(\Delta_i^{i+k_n-1}\tilde{N}\ge 1\right).
\end{aligned}
$$

The two quantities in the last inequality can be easily evaluated:

$$\begin{aligned}
&P\left(\Delta_i^{i+k_n-1}\tilde{N}=0, |\Delta_i^{i+k_n-1}J_2| > \sqrt{r(\Delta_n)}/2\right) \\
&\leq P\left(|\Delta_i^{i+k_n-1}J_2| > \sqrt{r(\Delta_n)}/2, |\Delta J_{2s}| \leq \sqrt{r(\Delta_n)}/2, \text{for all } s \in (i\Delta_n, (i+k_n-1)\Delta_n]\right) \\
&\leq 2\frac{E\left[(\Delta_i^{i+k_n-1}J_2)^2 I_{\{|\Delta J_{2s}|\leq \sqrt{r(\Delta_n)}/2, \text{for all } s\in(i\Delta_n,(i+k_n-1)\Delta_n]\}}\right]}{r(\Delta_n)} = O\left(\Delta_n^{1/2-\alpha\beta/2}\right)
\end{aligned}$$

and

$$P\left(\Delta_i^{i+k_n-1}\tilde{N}\geq 1\right) = O(k_n\Delta_n\nu\{|x| > \sqrt{r(\Delta_n)}/2\}) = O(\Delta_n^{1/2-\alpha\beta/2}).$$

This establishes (B.5). Therefore, $I_2 = O\left(\Delta_n^{\frac{1}{4}-\frac{\alpha\beta}{2}}\left(\log\frac{1}{\Delta_n}\right)^2\right)$.

Now we proceed with I_3. We first write I_3 as the following decomposition

$$I_3 = \frac{2\sqrt{\Delta_n}}{\theta\psi_2}\sum_{i=0}^{n-k_n+1} K_h(t_i-\tau)\bar{Z}_{1i}^n\bar{J}_{2i}^n\left[I_{\{(\bar{Z}_i^n)^2\leq r(\Delta_n),(\bar{J}_i^n)^2\leq 4r(\Delta_n)\}} + I_{\{(\bar{Z}_i^n)^2\leq r(\Delta_n),(\bar{J}_i^n)^2> 4r(\Delta_n)\}}\right].$$

On one hand, we have

$$\begin{aligned}
&P\left(\frac{2\sqrt{\Delta_n}}{\theta\psi_2\Delta_n^{1/4}}\sum_{i=0}^{n-k_n+1} K_h(t_i-\tau)\bar{Z}_{1i}^n\bar{J}_{2i}^n I_{\{(\bar{Z}_i^n)^2\leq r(\Delta_n),(\bar{J}_i^n)^2> 4r(\Delta_n)\}} \neq 0\right) \\
&\leq \Delta^{1/4}nP(|\bar{J}_2^n| > \sqrt{r(\Delta_n)}, \Delta_i N\neq 0) = O\left(\frac{\Delta_n^{3/4}}{r(\Delta_n)}\right)\to 0.
\end{aligned}$$

On the other hand, if $I_{\{(\bar{Z}_i^n)^2\leq r(\Delta_n),(\bar{J}_i^n)^2\leq 4r(\Delta_n)\}} = 1$, then for sufficiently small Δ_n, we have $\Delta_i^{i+k_n-1}N = 0$. By the Cauchy-Schwartz inequality, we have

$$\begin{aligned}
&\frac{\sqrt{\Delta_n}}{\Delta_n^{1/4}}\sum_{i=0}^{n-k_n+1} K_h(t_i-\tau)\int_{i\Delta_n}^{(i+k_n)\Delta_n} g_n(u-i\Delta_n)b_u du \bar{J}_{2i}^n I_{\{(\bar{Z}_i^n)^2\leq r(\Delta_n),(\bar{J}_i^n)^2\leq 4r(\Delta_n)\}} \\
&\leq \frac{\Delta_n^{1/4}\sqrt{\sum_{i=0}^{n-k_n+1}(\int_{i\Delta_n}^{(i+k_n)\Delta_n} g_n(u-i\Delta_n)b_u du)^2}}{\Delta_n^{1/4}}\sqrt{\sqrt{\Delta_n}\sum_{i=0}^{n-k_n+1}(\bar{J}_{2i}^n)^2 I_{\{(\bar{J}_i^n)^2\leq 4r(\Delta_n)\}}} \\
&= O(\eta(\sqrt{r(\Delta_n)})),
\end{aligned}$$

where $\eta^2(\delta) := \int_{|x|\leq\delta} x^2\nu(dx)$. We also have that as $\Delta_n\to 0$,

$$\eta^2\left(2\sqrt{r(\Delta_n)}\right) = \int_{|x|\leq 2\sqrt{r(\Delta_n)}} x^2\nu(dx) = O(r(\Delta_n)^{1-\frac{\alpha}{2}})$$

by Assumptions A3 and A6.

Moreover, a direct computation yields

$$\frac{\sqrt{\Delta_n}}{\Delta_n^{1/4}}\sum_{i=0}^{n-k_n+1} K_h(t_i-\tau)\int_{i\Delta_n}^{(i+k_n)\Delta_n} g_n(u-i\Delta_n)\sigma_u dW_u \bar{J}_{2i}^n I_{\{(\bar{Z}_i^n)^2\leq r(\Delta_n),(\bar{J}_i^n)^2\leq 4r(\Delta_n)\}}$$

$$
\begin{aligned}
&= \frac{\sqrt{\Delta_n}}{\Delta_n^{1/4}} \sum_{i=0}^{n-k_n+1} K_h(t_i-\tau) \int_{i\Delta_n}^{(i+k_n)\Delta_n} g_n(u-i\Delta_n)\sigma_u dW_u \bar{J}_{2i}^{n,m} \qquad \text{(B.6)}\\
&- \frac{\sqrt{\Delta_n}}{\Delta_n^{1/4}} \sum_{i=0}^{n-k_n+1} K_h(t_i-\tau) \int_{i\Delta_n}^{(i+k_n)\Delta_n} g_n(u-i\Delta_n)\sigma_u dW_u \bar{J}_{2i}^{n,c}\\
&- \frac{\sqrt{\Delta_n}}{\Delta_n^{1/4}} \sum_{i=0}^{n-k_n+1} K_h(t_i-\tau) \int_{i\Delta_n}^{(i+k_n)\Delta_n} g_n(u-i\Delta_n)\sigma_u dW_u \bar{J}_{2i}^{n} I_{\{(\bar{Z}_i^n)^2 > r(\Delta_n), (\bar{J}_i^n)^2 \le 4r(\Delta_n)\}},
\end{aligned}
$$

where $\bar{J}_{2i}^{n,m} = \int_{i\Delta_n}^{(i+k_n)\Delta_n} \int_{|x|\le 2r(\Delta_n)/g_n(u-i\Delta_n)} g_n(u-i\Delta_n)x\tilde{\mu}(dx,dt)$, $\bar{J}_{2i}^{n,c} = \int_{i\Delta_n}^{(i+k_n)\Delta_n} \int_{2r(\Delta_n)/g_n(u-i\Delta_n)\le|x|\le 1} g_n(u-i\Delta_n)x\nu(dx)dt$. Hence, we have $\bar{J}_{2i}^n = \bar{J}_{2i}^{n,m} - \bar{J}_{2i}^{n,c}$. Since it's easy to show that each term in (B.6) tends to zero in probability, we have $\frac{1}{\Delta_n^{1/4}} I_3 \xrightarrow{P} 0$.

Finally, for I_4, we have

$$
\begin{aligned}
\operatorname{P}\lim_{n\to 0} \frac{\sqrt{\Delta_n}\sum_{i=0}^{n-k_n+1} (\bar{J}_{2i}^n)^2 I_{\{(\bar{J}_i^n)^2\le 4r(\Delta_n)\}}}{r(\Delta_n)^{1-\alpha/2}} &= \operatorname{P}\lim_{n\to 0} \frac{\sqrt{\Delta_n}\sum_{i=0}^{n-k_n+1} (\bar{J}_{2i}^n)^2 I_{\{(\bar{Z}_i^n)^2\le r(\Delta_n),(\bar{J}_i^n)^2\le 4r(\Delta_n)\}}}{r(\Delta_n)^{1-\alpha/2}}\\
&\le \operatorname{P}\lim_{n\to 0} \frac{\sqrt{\Delta_n}\sum_{i=0}^{n-k_n+1} (\bar{J}_{2i}^n)^2 I_{\{(\bar{J}_i^n)^2\le 9r(\Delta_n)/4\}}}{r(\Delta_n)^{1-\alpha/2}} = L.
\end{aligned}
$$

When $\alpha < 1$ and $\beta > \frac{1}{4-2\alpha} \in (1/4, 1/2)$, $\frac{r(\Delta_n)^{1-\alpha/2}}{\Delta_n^{1/4}} \to 0$. In view of the results in Theorem 4.2, the claim (a) of Theorem 4.3 follows. When $\alpha > 1$, we have $\frac{\sqrt{h}r(\Delta_n)^{1-\alpha/2}}{\Delta_n^{1/4}} \to +\infty$, hence Theorem 4.3 (b) results.