Annals of the Institute of Statistical Mathematics (2021) 73:1089–1125
https://doi.org/10.1007/s10463-020-00781-0

A universal approach to estimate the conditional variance in semimartingale limit theorems

Mathias Vetter[1]

Received: 27 April 2020 / Revised: 30 October 2020 / Accepted: 5 November 2020 /
Published online: 1 January 2021

Abstract

The typical central limit theorems in high-frequency asymptotics for semimartingales are results on stable convergence to a mixed normal limit with an unknown conditional variance. Estimating this conditional variance usually is a hard task, in particular when the underlying process contains jumps. For this reason, several authors have recently discussed methods to automatically estimate the conditional variance, i.e. they build a consistent estimator from the original statistics, but computed at different time scales. Their methods work in several situations, but are essentially restricted to the case of continuous paths always. The aim of this work is to present a new method to consistently estimate the conditional variance which works regardless of whether the underlying process is continuous or has jumps. We will discuss the case of power variations in detail and give insight to the heuristics behind the approach.

Keywords Asymptotic conditional variance · High-frequency statistics · Itô semimartingale · Jumps · Stable convergence

1 Introduction

The asymptotic theory for functionals of semimartingales observed at high frequency is well understood now. Since the beginning of the century, a variety of laws of large numbers and accompanying central limit theorems has been stated in different situations, starting with power and bipower variation for continuous processes (Barndorff-Nielsen and Shephard (2003) or Barndorff-Nielsen et al. (2006)). Crucial generalizations involve the case of possible jumps in the process (Jacod 2008) or the discussion of observations with additional microstructure noise (Jacod et al. 2010). Later extensions regard truncated increments, multivariate processes or the

✉ Mathias Vetter
vetter@math.uni-kiel.de

[1] Mathematisches Seminar, Christian-Albrechts-Universität zu Kiel, Ludewig-Meyn-Str. 4, 24118 Kiel, Germany

treatment of irregularity and asynchronicity in the data. A general overview about these results and statistical applications can be found in the monographs Jacod and Protter (2012) and Aït-Sahalia and Jacod (2014).

Typically, the central limit theorems in these situations are stated as follows: one proves stable convergence in law of an appropriately rescaled statistic to a mixed normal limit, where the (asymptotic) conditional variance of the limiting variable is a random variable which depends in a complicated way on the underlying semimartingale. Once a consistent estimator for this conditional variance has been constructed, thanks to the properties of stable convergence in law, one can deduce the convergence in distribution of the standardized statistic to a standard Gaussian law. This opens the door for all kinds of statistical applications.

Constructing a consistent estimator for the conditional variance, however, is not always a simple task. Compared with the original object of interest for which the law of large numbers is shown, usually an integral of a power of volatility or a sum of a power of jumps, the variance is typically of a more complicated form and might depend on additional objects as well. In particular, apart from the case of power variations of continuous processes, it is not possible to estimate the variance by using similar statistics as for the corresponding law of large numbers. Hence, estimators are usually constructed based on the specific form of the conditional variance in the respective situations. This procedure has two major drawbacks: first, every newly proven central limit theorem requires new estimators for the conditional variances. Second, when the model is not correctly specified, it is likely that the proposed estimator does not work.

A different approach is to build an estimator which only requires knowledge of the original statistics and does not rely on the specific form of the conditional variance. For example, Jacod (2008) discusses statistics of the form

$$U_n = \sum_{i=1}^{n} f_n(\Delta_i^n X),$$

for simplicity over $[0, 1]$, where $\Delta_i^n X = X_{i\Delta_n} - X_{(i-1)\Delta_n}$ denotes the ith increment of the semimartingale X, $\Delta_n \to 0$, and where $f_n : \mathbb{R} \to \mathbb{R}$ is a function which may or may not depend on n. Several laws of large numbers and associated central limit theorems are shown in various cases. A universal estimator for the conditional variance in these very central limit theorems would then only depend on f_n, but not utilize the specific form of the conditional variance in the respective situations. Whether such estimators exist, and how they look like, is obviously an important question in the theoretical discussion of high-frequency statistics.

In recent years, two classes of such universal estimators have been proposed in the literature. Mykland and Zhang (2017) base their estimator on a comparison of local versions of U_n computed over neighbouring intervals of length $k_n\Delta_n$, $k_n \to \infty$ and $k_n\Delta_n \to 0$, whereas Christensen et al. (2017) use a subsampling approach which compares U_n with versions where only every k_nth increment is taken into account. Both estimators are shown to work in a variety of situations, but only when the semimartingale X does not jump (or when the jumps do not contribute to the limiting distribution), and it is rather simple to see that both procedures indeed do not work when the limiting distribution contains jumps.

Therefore, the question remains whether it is possible to construct a universal estimator for the conditional variance which works both in the continuous case and in the case involving jumps and, if yes, how it could be constructed. We will give positive answers to both questions, for simplicity in the case of power variations only, which means that X is a general Ito semimartingale including jumps and that f_n is essentially of the form $f_n(x) = |x|^p$, $p > 0$, up to a possible standardization. Already in this situation, we will see all different kinds of limiting behaviour, including conditional variances which only depend on the volatility or which depend jointly on jumps and volatility. It is to be expected that the same construction of a universal estimator works for most other statistics as well, as the main idea behind the proof of the respective central limit theorems usually is the same as for the corresponding power variations.

The paper is organized as follows: after introducing the setting in Sect. 2, we will discuss three novel universal estimators for the conditional variance in Sect. 3. While the first two estimators are rather simple to construct in practice, they have the deficiency that they do not work in all situations. In fact, the first one is consistent for continuous processes, but when jumps dominate it only converges stably in law to a random variable whose mean is the conditional variance. Similarly for the second estimator, but with different roles: the estimator is consistent in the jump case, but does not converge to the correct conditional variance for continuous processes. A remarkable exception is the case $p = 2$ in which it gives an alternative estimator for the conditional variance when the quadratic variation is to be estimated. Finally, the intuition behind both estimators is combined to construct the universal estimator which formally works in all situations. Its computation time is of order $n\binom{k_n}{\ell_n}$ for sequences k_n and ℓ_n converging to infinity, however, so it is of theoretical interest in the first place rather than being a serious alternative in all practical cases. The proofs are given in Sect. 5.

2 Setting

Suppose that we have a filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geq 0}, \mathbb{P})$ on which an Ito semimartingale of the form

$$\begin{aligned} X_t = X_0 + \int_0^t b_s \mathrm{d}s + \int_0^t \sigma_s \mathrm{d}W_s + \int_0^t \int \delta(s,z) \mathbb{1}_{\{|\delta(s,z)|\leq 1\}} (\mu - \nu)(\mathrm{d}s, \mathrm{d}z) \\ + \int_0^t \int \delta(s,z) \mathbb{1}_{\{|\delta(s,z)|>1\}} \mu(\mathrm{d}s, \mathrm{d}z) \end{aligned} \tag{1}$$

is defined, where W is a standard Brownian motion, μ is a Poisson random measure on $\mathbb{R}^+ \times \mathbb{R}$, and its predictable compensator satisfies $\nu(\mathrm{d}s, \mathrm{d}z) = \mathrm{d}s \otimes \lambda(\mathrm{d}z)$ for some σ-finite measure λ on $\mathbb{R}$ endowed with the Borelian σ-algebra. We further assume that b and σ are adapted processes and that δ is predictable on $\Omega \times \mathbb{R}^+ \times \mathbb{R}$. We write $\Delta X_s = X_s - X_{s-}$ with $X_{s-} = \lim_{t\nearrow s} X_t$ for a possible jump of X in s.

We will work in a high-frequency framework, so without loss of generality we assume to be on the fixed interval [0, 1]. Observations of X take place at the regular times $i\Delta_n$, $i = 0, \dots, n$, where we set $n = \Delta_n^{-1}$. Throughout the paper, $\Delta_n \to 0$ governs the asymptotics.

In order to prove asymptotic results for statistics based on increments of X, one typically needs additional assumptions on the semimartingale characteristics. Our aim in the following is not to be as general as possible, so we will state sufficient conditions in order to prove consistency of the statistics and associated central limit theorems, respectively. The first one is good enough for theorems on consistency, and it even is sufficient for some central limit theorems.

Condition 1 The process (b_s) is locally bounded and predictable, the process (σ_s) is càdlàg, and there exist a sequence (τ_n) of stopping times increasing to infinity and a sequence (γ_n) of deterministic real functions such that $1 \wedge |\delta(s,z)| \leq \gamma_n(z)$ for all $s \leq \tau_n$ and $\int \gamma_n(z)^2 \lambda(\mathrm{d}z) < \infty$ hold.

Stronger assumptions are typically needed when one is interested in a central limit theorem associated with a limit in probability which is governed by the continuous martingale part of X. What is always needed is that σ is positive and that it takes a form similar to (1).

Condition 2 We assume that the process (σ_s) is bounded below by a positive number and of the form

$$\sigma_t = \sigma_0 + \int_0^t \widetilde{b}_s \mathrm{d}s + \int_0^t \widetilde{\sigma}_s \mathrm{d}W_s + M_t + \sum_{0<s\leq t} \Delta\sigma_s \mathbb{1}_{\{|\Delta\sigma_s|>1\}}$$

with M being a local martingale with $|\Delta M_s| \leq 1$, orthogonal to W, and we assume that $\langle M, M\rangle_t = \int_0^t \alpha_s \mathrm{d}s$ as well as that the compensator of $\sum_{0<s\leq t} \Delta\sigma_s \mathbb{1}_{\{|\Delta\sigma_s|>1\}}$ takes the form $\int_0^t \alpha_s' \mathrm{d}s$. The processes (b_s) and $(\widetilde{\sigma}_s)$ are càdlàg, and the processes $(\widetilde{b}_s)$, (α_s) and (α_s') are locally bounded and predictable.

Even this condition is not general enough in the case where X has jumps as well; see Theorems 5.3.5 and 5.3.6 in Jacod and Protter (2012). We will therefore assume that X is continuous whenever we are concerned with central limit theorems associated with the continuous martingale part only. Condition 2 turns out to be sufficient then.

3 Results

3.1 Limit theorems for power variations

The typical object of interest in high-frequency statistics is a statistic of the form

$$U_n = \sum_{i=1}^{n} f_n(\Delta_i^n X),$$

where $\Delta_i^n X = X_{i\Delta_n} - X_{(i-1)\Delta_n}$ denotes the ith increment of X and $f_n : \mathbb{R} \to \mathbb{R}$ is a function which may or may not depend on n. Typical examples are power variations of the form

$$f_n(x) = |x|^p \quad \text{or} \quad f_n(x) = \Delta_n^{1-p/2}|x|^p$$

for some $p > 0$, where the latter scaling depends on the length of the interval over which the increment is computed. For those power variations and related statistics, a rule of thumb is: whenever a weak law of large numbers holds, the limit is of the form

$$U = \int_0^1 g(\sigma_s)\mathrm{d}s + \sum_{0<s\le 1} h(\Delta X_s),$$

where $g : \mathbb{R}^+ \to \mathbb{R}$ and $h : \mathbb{R} \to \mathbb{R}$ are suitable functions depending on f_n. Let us recall the results from Theorem 2.2 and Theorem 2.4 in Jacod (2008).

Theorem 1 *Let X be a semimartingale of the form* (1) *and assume that Condition* 1 *holds.*

(a) *Let $p < 2$ and $f_n(x) = \Delta_n^{1-p/2}|x|^p$. Then*

$$U_n \xrightarrow{\mathbb{P}} m_p \int_0^1 \sigma_s^p \mathrm{d}s$$

with $m_p = \mathbb{E}[|N|^p]$ for $N \sim \mathcal{N}(0,1)$.

(b) *Let $p > 2$ and $f_n(x) = |x|^p$ for any n. Then*

$$U_n \xrightarrow{\mathbb{P}} \sum_{0<s\le 1} |\Delta X_s|^p.$$

(c) *Let $f_n(x) = |x|^2$ for any n. Then*

$$U_n \xrightarrow{\mathbb{P}} [X,X]_1 = \int_0^1 \sigma_s^2 \mathrm{d}s + \sum_{0<s\le 1} |\Delta X_s|^2.$$

Remark 1 In the case where no jumps are present, the law of large numbers in part (a) also holds for $p \ge 2$. Similarly, if the continuous martingale part vanishes the claim in part (b) also holds for $p \in (1,2]$ and, under a further assumption on the drift, even for $p \le 1$. See again Jacod (2008). □

As noted above we have associated central limit theorems in all three cases, but for simplicity we will state the one connected to Theorem 1 (a) only in the case of a continuous X in which it holds irrespective of p. In general, such a result is expected to hold only with $p < 1$, but with additional assumptions regarding the jumps then. Similarly, the central limit theorem associated with Theorem 1 (b) only holds for $p > 3$. The mode of convergence is always ($\mathcal{F}$-)stable convergence in law, which means in particular that the limiting variables are typically defined on an appropriate extension of $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geq 0}, \mathbb{P})$. For details on stable convergence see Section 2.2.1 in Jacod and Protter (2012).

Theorem 2 *Let X be a semimartingale of the form* (1).

(a) *Suppose that X is continuous and assume that Condition 2 holds. With $f_n(x) = \Delta_n^{1-p/2}|x|^p$, we have the stable convergence*

$$\Delta_n^{-1/2}\Big(U_n - m_p \int_0^1 \sigma_s^p \mathrm{d}s\Big) \xrightarrow{\mathcal{L}-(s)} Y = \sqrt{m_{2p} - m_p^2} \int_0^1 \sigma_s^p \mathrm{d}W_s'$$

where W' denotes an independent Brownian motion on a suitable extension of the original probability space.

(b) *Let $p > 3$ and suppose that X allows for jumps and that X and σ never jump at the same time. Under Condition 1 and with $f_n(x) = |x|^p$ for all n, we have the stable convergence*

$$\Delta_n^{-1/2}\Big(U_n - \sum_{0<s\leq 1} |\Delta X_s|^p\Big) \xrightarrow{\mathcal{L}-(s)} Z = \sum_{r=1}^{\infty} p\,sign(\Delta X_{S_r})|\Delta X_{S_r}|^{p-1}\sigma_{S_r} N_r$$

where $(S_r)_{r\geq 1}$ denotes a sequence of stopping times exhausting the jumps of X over $[0, 1]$, and where $(N_r)_{r\geq 1}$ is a sequence of independent standard normal variables, also defined on a suitable extension of the original probability space.

(c) *Suppose that X allows for jumps and that X and σ never jump at the same time. Under Condition 1 and with $f_n(x) = |x|^2$ for all n, we have the stable convergence*

$$\Delta_n^{-1/2}\big(U_n - [X, X]_1\big) \xrightarrow{\mathcal{L}-(s)} Y + Z,$$

with Y as in part (a) and Z as in part (b), where W' and $(N_r)_{r\geq 1}$ are defined on the same extended probability space and independent.

For a proof see Theorems 5.3.6, 5.1.2 and 5.4.2 of Jacod and Protter (2012).

Remark 2 The limiting variable in part (a) of Theorem 2 is mixed normal with conditional variance

$$V = (m_{2p} - m_p^2) \int_0^1 \sigma_s^{2p} \mathrm{d}s.$$

Given a consistent estimator V_n for V, Slutsky's lemma for stable convergence yields

$$\frac{\Delta_n^{-1/2}\left(U_n - m_p \int_0^1 \sigma_s^p \mathrm{d}s\right)}{\sqrt{V_n}} \xrightarrow{\mathcal{L}} \mathcal{N}(0,1). \tag{2}$$

In general, the central limit results connected with jumps do not allow for a mixed normal limit. An exception is the case where σ and X have no common jumps [compare, e.g. Proposition 5.1.1 in Jacod and Protter (2012)], which is why we work under this assumption. In this case, we obtain

$$V = \sum_{0<s\le 1} p^2 |\Delta X_s|^{2p-2} \sigma_s^2$$

for part (b) and

$$V = 2\int_0^1 \sigma_s^4 \mathrm{d}s + \sum_{0<s\le 1} 4|\Delta X_s|^2 \sigma_s^2$$

for part (c), respectively. The goal then again is to find a consistent estimator for V, from which central limit theorems similar to (2) can be concluded. □

Historically, estimators for the asymptotic conditional variances in Theorem 2 have been built using the exact representation of V and somewhat similar statistics as the original power variations. For example, in case (a) above it is obvious from Remark 1 that

$$\widehat{V}_n = \frac{m_{2p} - m_p^2}{m_{2p}} \sum_{i=1}^n g_n(\Delta_i^n X)$$

with $g_n(x) = \Delta_n^{1-p}|x|^{2p}$ consistently estimates V. In the other two cases, estimation of the conditional variances is possible, yet severely more complicated due to the mixture of jumps and volatility. Plain power variations cannot be used anymore, but a truncated version where only increments $\Delta_i^n X$ with $|\Delta_i^n X| > \alpha \Delta_n^{\varpi}$, $\varpi < 1/2$, $\alpha > 0$, are used, combined with a local estimator for the volatility, still does the trick. See, for example, Theorem 9.5.1 in Jacod and Protter (2012). This feature in fact is typical in high-frequency analysis: the conditional variance is often substantially more difficult to estimate than the original quantities of interest.

3.2 Universal estimators in the continuous case

Two competing procedures have recently been proposed in the literature which do not try to mimic the specific structure of the limiting conditional variance, but rather construct estimators directly from the form of the original statistics U_n. Let us remain in the framework of power variations, so

$$U_n = \sum_{i=1}^{n} f_n(\Delta_i^n X),$$

and let us write the limiting variables in Theorem 1 as

$$U = \sum_{i=1}^{n} \theta_{[(i-1)\Delta_n, i\Delta_n]},$$

so, for example,

$$\theta_{[(i-1)\Delta_n, i\Delta_n]} = [X,X]_{i\Delta_n} - [X,X]_{(i-1)\Delta_n} = \int_{(i-1)\Delta_n}^{i\Delta_n} \sigma_s^2 \mathrm{d}s + \sum_{(i-1)\Delta_n < s \le i\Delta_n} |\Delta X_s|^2$$

in case of part (c). The essential idea behind the estimator from Mykland and Zhang (2017) is the intuition that each summand $f_n(\Delta_i^n X)$ within U_n is in fact a local estimate for the corresponding $\theta_{[(i-1)\Delta_n, i\Delta_n]}$, and this intuition remains true if several increments are aggregated. Precisely,

$$\hat{\theta}_{[i\Delta_n, (i+k_n)\Delta_n]} = \sum_{j=1}^{k_n} f_n(\Delta_{i+j}^n X) \tag{3}$$

with an auxiliary sequence $k_n \to \infty$, $k_n \Delta_n \to 0$, serves as an estimator for

$$\theta_{[i\Delta_n, (i+k_n)\Delta_n]} = \sum_{j=1}^{k_n} \theta_{[(i+j-1)\Delta_n, (i+j)\Delta_n]}. \tag{4}$$

They therefore base their estimator on

$$QV_n(k_n) = \frac{1}{k_n} \sum_{i=k_n}^{n-k_n} \left(\hat{\theta}_{[(i-k_n)\Delta_n, i\Delta_n]} - \hat{\theta}_{[i\Delta_n, (i+k_n)\Delta_n]} \right)^2$$

which, using a simple decomposition, essentially mimics twice the asymptotic variance, plus an additional term due the difference of $\theta_{[i\Delta_n, (i+k_n)\Delta_n]}$ and $\theta_{[(i-k_n)\Delta_n, i\Delta_n]}$. When the latter approximation error is not too large compared with the other two terms, it is possible to get rid of it by working with a suitable linear combination of two different $QV_n(k_n)$. Among other possible linear combinations Mykland and Zhang (2017) choose

$$T_n = \frac{2}{3}\left(QV_n(k_n) - \frac{1}{4} QV_n(2k_n) \right).$$

An estimator for V is then given by nT_n.

The estimator from Christensen et al. (2017) is based on a subsampling procedure. They set

$$U_l^n = k_n \sum_{i=1}^{\lfloor \frac{n}{k_n} \rfloor} f_n(\Delta^n_{(i-1)k_n+l} X)$$

for each $l = 1, \dots, k_n$. Up to edge effects, this is the same estimator as the original one, but where only each k_nth increment is taken into account, thus the estimator is blown up by the factor k_n. Again, $f_n(\Delta^n_{(i-1)k_n+l} X)$ is a local estimator for $\theta_{[((i-1)k_n+l-1)\Delta_n,((i-1)k_n+l)\Delta_n]}$, and if neighbouring $\theta_{[((i-1)k_n+l-1)\Delta_n,((i-1)k_n+l)\Delta_n]}$ are close the each other, then U_l^n should behave in the same way as the original U_n. In particular, a central limit theorem should hold with the same asymptotic variance, but the rate of convergence should drop to $(k_n \Delta_n)^{1/2}$. Therefore, the subsampling estimator for the asymtotic variance is given by

$$\widehat{\Sigma}_n = \frac{1}{k_n} \sum_{l=1}^{k_n} (k_n \Delta_n)^{-1} (U_l^n - U_n)^2,$$

where U_n serves as an approximation for the unknown limit U. As the convergence of U_n to U happens at a faster rate than the convergence of U_l^n to U, this replacement does not cause any troubles in the limit.

Both estimators, nT_n and $\widehat{\Sigma}_n$, are known to work in a variety of situations if $k_n \to \infty$ and $k_n \Delta_n \to 0$ hold and are by no means restricted to power variations. Mykland and Zhang (2017) work with a structural assumption and show that their estimator works in most cases where the limiting variable takes the form

$$U = \int_0^1 \theta_s \mathrm{d}s$$

for some semimartingale θ, whereas Christensen et al. (2017) establish consistency of their subsampling estimators explicitly for power and bipower variations, including a truncated version when additional jumps are present in the process and a pre-averaged version when the process is only observed with noise. In particular, in both papers the case of a limit governed by jumps is excluded, intuitively because the implicit assumption fails that estimators close nearby will estimate the same quantity. In fact, they estimate very different quantities if a jump is present because it falls into just one interval and not into the next one.

Example 1 Suppose that $X_t = \sigma W_t + J_t$ for a constant $\sigma > 0$ and a Poisson process J with parameter $\lambda > 0$. Then, with

$$f_n(\Delta_i^n X) = |X_{i\Delta_n} - X_{(i-1)\Delta_n}|^2$$

and

$$U_n = \sum_{i=1}^{n} f_n(\Delta_i^n X),$$

we have

$$\Delta_n^{-1/2}\big(U_n - [X,X]_1\big) \xrightarrow{\mathcal{L}-(s)} Y + Z,$$

according to Theorem 2, where the limiting variance is given by

$$V = 2\sigma^4 + 4\sigma^2 J_1.$$

But, for any choice of $k_n \to \infty$ and $k_n\Delta_n \to 0$ we neither have $nT_n \xrightarrow{\mathbb{P}} V$ nor $\widehat{\Sigma}_n \xrightarrow{\mathbb{P}} V$. A proof of this result will be given in Sect. 5.1.

3.3 Three new universal estimators

In order to circumvent the problem that a jump falls into just one interval, we will present several novel estimators in the following, all of which are based on the following intuition: we fix a local interval $[i\Delta_n, (i+k_n)\Delta_n]$ first, and we will always compare two estimators constructed from increments within this interval only. These estimators are defined in such a way that a possible jump dominates both estimators in the same way, so that it is wiped out to first order. Afterwards, the local estimators based on $[i\Delta_n, (i+k_n)\Delta_n]$ are aggregated into a global estimator.

This procedure is explained easiest for a first estimator V_n which is not universal in the sense that $V_n \xrightarrow{\mathbb{P}} V$ holds in all three cases. Recall (3) and (4). We will use $\widehat{\theta}_{[i\Delta_n,(i+k_n)\Delta_n]}$ as a local estimator for $\theta_{[i\Delta_n,(i+k_n)\Delta_n]}$ again, but it will be compared with a local power variation based on the increment $X_{(i+k_n)\Delta_n} - X_{i\Delta_n}$ which, using the same $p > 0$, also is a local estimator for $\theta_{[i\Delta_n,(i+k_n)\Delta_n]}$. Recall that a possible scaling depends on the length of the interval over which the increment is computed, so the factor will be based on $k_n\Delta_n$ instead of Δ_n. This means, in the continuous case we set

$$U^n_{[i\Delta_n,(i+k_n)\Delta_n]} = (k_n\Delta_n)^{1-p/2}|X_{(i+k_n)\Delta_n} - X_{i\Delta_n}|^p$$

and

$$\widehat{\theta}_{[i\Delta_n,(i+k_n)\Delta_n]} = \sum_{j=1}^{k_n} \Delta_n^{1-p/2}|\Delta^n_{i+j}X|^p,$$

while otherwise

$$U^n_{[i\Delta_n,(i+k_n)\Delta_n]} = |X_{(i+k_n)\Delta_n} - X_{i\Delta_n}|^p, \quad \widehat{\theta}_{[i\Delta_n,(i+k_n)\Delta_n]} = \sum_{j=1}^{k_n} |\Delta^n_{i+j}X|^p.$$

The first estimator is then given by

$$\widehat{V}_n = \frac{n}{k_n(k_n-1)} \sum_{i=0}^{n-k_n} \Big(U^n_{[i\Delta_n,(i+k_n)\Delta_n]} - \widehat{\theta}_{[i\Delta_n,(i+k_n)\Delta_n]}\Big)^2.$$

Theorem 3 *Let X be of the form* (1) *and let* $k_n \to \infty$ *such that* $k_n = o(n)$.

(a) *Suppose that Xis continuous and assume that Condition 2 holds. We have*

$$\widehat{V}_n \xrightarrow{\mathbb{P}} V = (m_{2p} - m_p^2) \int_0^1 \sigma_s^{2p} \mathrm{d}s.$$

(b) *Let $p > 3$ and suppose that X allows for jumps and that X and σ never jump at the same time. Under Condition 1 , we have the stable convergence*

$$\widehat{V}_n \xrightarrow{\mathcal{L}-(s)} V^* = \sum_{r=1}^{\infty} p^2 |\Delta X_{S_r}|^{2p-2} \sigma_{S_r}^2 (1 + R_r)$$

where $(S_r)_{r\geq 1}$ denotes a sequence of stopping times exhausting the jumps of X over $[0, 1]$ and where $(R_r)_{r\geq 1}$ denotes a sequence of i.i.d. random variables distributed as a multiple Wiener-Itô integral

$$\int_{[0,2]^2} f(s,t) \mathrm{d}B_s \mathrm{d}B_t, \quad f(s,t) = (2 - s \vee t) \wedge 1 - (1 - s \wedge t) \vee 0,$$

independent of $\mathcal{F}$ and defined on a suitable extension of the original probability space.

(c) *Suppose that X allows for jumps and that X and σ never jump at the same time. For $p = 2$, under Condition 1 we have*

$$\widehat{V}_n \xrightarrow{\mathcal{L}-(s)} 2 \int_0^1 \sigma_s^4 \mathrm{d}s + \sum_{r=1}^{\infty} 4 |\Delta X_{S_r}|^2 \sigma_{S_r}^2 (1 + R_r),$$

with $(S_r)_{r\geq 1}$ and $(R_r)_{r\geq 1}$ as in (b).

Remark 3 Let us discuss the heuristics behind Theorem 3 by distinguishing the two cases of X being continuous and X having jumps. The mixed case typically just combines those arguments.

(i) In the continuous case, let us discuss the related, asymptotically equivalent, estimator

$$\widehat{V}_n^{(1)} = \frac{1}{k_n} \sum_{\ell=0}^{k_n-1} \frac{n}{k_n} \sum_{i=0}^{\lfloor \frac{n}{k_n} \rfloor - 1} \left(U^n_{[(ik_n+\ell)\Delta_n, ((i+1)k_n+\ell)\Delta_n]} - \widehat{\theta}_{[(ik_n+\ell)\Delta_n, ((i+1)k_n+\ell)\Delta_n]} \right)^2$$

which is the same as $\widehat{V}_n$ up to small order edge effects. Note that for each fixed ℓ, the estimator is based on observations from non-overlapping intervals. Later on these are aggregated in some type of sample mean. Then, if we set

$$\theta_{[u,v]} = m_p \int_u^v \sigma_s^p \mathrm{d}s$$

for $u < v$, following the same proof as Theorem 2 (a), it is easy to see that

$$\widehat{\theta}_{[(ik_n+\ell)\Delta_n,((i+1)k_n+\ell)\Delta_n]} - \theta_{[(ik_n+\ell)\Delta_n,((i+1)k_n+\ell)\Delta_n]} = o_{\mathbb{P}}(k_n\Delta_n),$$

uniformly in i and ℓ. Young's inequality allows us to replace one term by the other. As we work over disjoint intervals, we then use the intuition that the

$$\sqrt{\frac{n}{k_n}} \sum_{i=0}^{\lfloor \frac{n}{k_n} \rfloor - 1} \left(U^n_{[(ik_n+\ell)\Delta_n,((i+1)k_n+\ell)\Delta_n]} - \theta_{[(ik_n+\ell)\Delta_n,((i+1)k_n+\ell)\Delta_n]} \right)$$

obey the same central limit theorem as Theorem 2 (a). In particular, using conditional independence, it is no surprise that each

$$\frac{n}{k_n} \sum_{i=0}^{\lfloor \frac{n}{k_n} \rfloor - 1} \left(U^n_{[(ik_n+\ell)\Delta_n,((i+1)k_n+\ell)\Delta_n]} - \widehat{\theta}_{[(ik_n+\ell)\Delta_n,((i+1)k_n+\ell)\Delta_n]} \right)^2$$

estimates V. So does $\widehat{V}_n^{(1)}$.

(ii) Whenever jumps are present, the idea is to implicitly assume that there are only finitely many of them and that each interval $(i\Delta_n, (i+k_n)\Delta_n]$ contains either no jump or exactly one jump. The proof of Theorem 2 (b) shows, due to $p > 3$, that only those intervals with jumps play a role to first order in the asymptotics. For each jump time S_r and for each interval such that $S_r \in (i\Delta_n, (i+k_n)\Delta_n]$, a Taylor expansion gives

$$\begin{aligned}
&\widehat{\theta}_{[i\Delta_n,(i+k_n)\Delta_n]} - \theta_{[i\Delta_n,(i+k_n)\Delta_n]} \\
&\quad = p\, \mathrm{sign}(\Delta X_{S_r})|\Delta X_{S_r}|^{p-1}\sigma_{S_r}(W_{(i+k_n)\Delta_n} - W_{i\Delta_n}) + o_{\mathbb{P}}((k_n\Delta_n)^{1/2})
\end{aligned}$$

uniformly in i, where $\theta_{[i\Delta_n,(i+k_n)\Delta_n]} = |\Delta X_{S_r}|^p$ and by using that σ is continuous at S_r by assumption. Therefore,

$$\widehat{V}_n = \sum_r p^2 |\Delta X_{S_r}|^{2p-2}\sigma^2_{S_r} \frac{n}{k_n^2} \sum_{j=1}^{k_n} (W_{(i_r+k_n-j)\Delta_n} - W_{(i_r-j)\Delta_n})^2 \big(1 + o_{\mathbb{P}}(1)\big),$$

where $((i_r-1)\Delta_n, i_r\Delta_n]$ denotes the interval which includes S_r. Note that the second sum above consists of highly correlated Brownian increments, and it is easy to see that both its expectation and its variance are equal to one, at least to first order. So we do not have convergence to one in probability. □

The lesson told by Remark 3 is that we need less dependence between the Brownian increments over those intervals where jumps are detected. A natural second statistic therefore is given by

$$\widetilde{V}_n = \frac{n}{2} \sum_{i=0}^{n-k_n} \left(\frac{1}{\binom{k_n}{2}} \sum_{i<u<v\le i+k_n} \big(f_n(\Delta_u^n X + \Delta_v^n X) - (f_n(\Delta_u^n X) + f_n(\Delta_v^n X))\big)^2 \right),$$

where the scaling within f_n again depends on the length of the corresponding interval, so it becomes

$$f_n(\Delta_u^n X + \Delta_v^n X) = (2\Delta_n)^{1-p/2}|\Delta_u^n X + \Delta_v^n X|^p, \quad f_n(\Delta_u^n X) = \Delta_n^{1-p/2}|\Delta_u^n X|^p,$$

in the continuous case and

$$f_n(\Delta_u^n X + \Delta_v^n X) = |\Delta_u^n X + \Delta_v^n X|^p, \quad f_n(\Delta_u^n X) = |\Delta_u^n X|^p,$$

in the other cases.

Let us explain the main idea behind $\widetilde{V}_n$ by using the simplifying assumption again that there are only finitely many jumps which are separated in the sense that no interval $(i\Delta_n, (i+k_n)\Delta_n]$ contains more than one jump. Then in the jump case

$$\widetilde{V}_n = \sum_r \frac{n}{k_n(k_n-1)} \sum_{j=1}^{k_n} \Bigg(\sum_{\substack{i_r - j < v \le i_r + k_n - j \\ v \neq i_r}} \left(|\Delta_{i_r}^n X + \Delta_v^n X|^p - |\Delta_{i_r}^n X|^p\right)^2 \Bigg) \left(1 + o_{\mathbb{P}}(1)\right),$$

as only the cases with $u = i_r$ or $v = i_r$ give dominating terms to first order. If one now uses a Taylor expansion and keeps j fixed first, we obtain

$$\widetilde{V}_n = \sum_r \frac{1}{k_n} \sum_{j=1}^{k_n} \frac{n}{k_n - 1} \sum_{\substack{i_r - j < v \le i_r + k_n - j \\ v \neq i_r}} p^2 (\Delta X_{S_r})^{2p-2} \sigma_{S_r}^2 (\Delta_v^n W)^2 + o_{\mathbb{P}}(1),$$

and it is clear that we have indeed convergence in probability to the correct quantity.

The drawback, however, is that the statistic does not converge in probability to the correct variance if the continuous part dominates. The reason is simple: we now subtract $(f_n(\Delta_u^n X) + f_n(\Delta_v^n X))$ only which is just a sum of two terms. Previously, when discussing $\widehat{V}_n$, we subtracted a sum of k_n terms which asymptotically equals a functional of σ^p. This allowed us to mimic the arguments from the original central limit theorem. Now we estimate a quantity which is in general different from V. A remarkable exception is the case $p = 2$ where we exactly estimate the variance V.

Theorem 4 *Let X be of the form* (1) *and let $k_n \to \infty$ such that $k_n = o(n)$.*

(a) *Suppose that X is continuous and assume that Condition 2 holds. With*

$$c_p = 2\mathbb{E}\left[\left(\left|\frac{1}{\sqrt{2}}(N_1 + N_2)\right|^p - \frac{1}{2}(|N_1|^p + |N_2|^p)\right)^2\right]$$

for independent standard normal N_1, N_2 we have

$$\widetilde{V}_n \xrightarrow{\mathbb{P}} c_p \int_0^1 \sigma_s^{2p} \mathrm{d}s.$$

(b) *Let $p > 3$ and suppose that X allows for jumps and that X and σ never jump at the same time. Under Condition 1 , we have*

$$\widetilde{V}_n \xrightarrow{\mathbb{P}} V = \sum_{0<s\leq 1} p^2 |\Delta X_s|^{2p-2} \sigma_s^2.$$

(c) *Suppose that X allows for jumps and that X and σ never jump at the same time. For $p = 2$, under Condition 1 we have*

$$\widetilde{V}_n \xrightarrow{\mathbb{P}} V = 2 \int_0^1 \sigma_s^4 \mathrm{d}s + \sum_{0<s\leq 1} 4|\Delta X_s|^2 \sigma_s^2.$$

Remark 4 Note that Theorem 4 (c) proves that $\widetilde{V}_n$ is a consistent estimator for the asymptotic conditional variance when the quadratic variation is to be estimated. In this situation, various estimators are known in the literature which all mimic the specific form of the variance; see, for example, Chapter 9.5 in Jacod and Protter (2012) or Veraart (2010). □

The construction of a universal estimator which converges in probability to V in all three cases now combines the best from both worlds. Let $\ell_n \to \infty$ with $\ell_n = o(k_n)$ be another auxiliary sequence and set

$$V_n = \frac{n}{\ell_n(\ell_n - 1)} \sum_{i=0}^{n-k_n} \frac{1}{\binom{k_n}{\ell_n}} \sum_{i<j_1<\ldots<j_{\ell_n}\leq i+k_n} \left(f_n\Big(\sum_{m=1}^{\ell_n} \Delta_{j_m}^n X \Big) - \sum_{m=1}^{\ell_n} f_n(\Delta_{j_m}^n X) \right)^2,$$

so

$$f_n\Big(\sum_{m=1}^{\ell_n} \Delta_{j_m}^n X \Big) = |\ell_n \Delta_n|^{1-p/2} \Big| \sum_{m=1}^{\ell_n} \Delta_{j_m}^n X \Big|^p, \quad f_n(\Delta_{j_m}^n X) = \Delta_n^{1-p/2} |\Delta_{j_m}^n X|^p$$

for a continuous X and again

$$f_n\Big(\sum_{m=1}^{\ell_n} \Delta_{j_m}^n X \Big) = \Big| \sum_{m=1}^{\ell_n} \Delta_{j_m}^n X \Big|^p, \quad f_n(\Delta_{j_m}^n X) = |\Delta_{j_m}^n X|^p$$

otherwise. We see that a jump in $\Delta_{j_1}^n X$, say, comes together with a growing number of increments which are sufficiently independent from each other in order to ensure convergence in probability as for $\widetilde{V}_n$. Also, as we subtract $\sum_{m=1}^{\ell_n} f_n(\Delta_{j_m}^n X)$ we consistently estimate a local version of σ^p in the continuous case. Note that $\widehat{V}_n$ and $\widetilde{V}_n$ are special cases with $\ell_n = k_n$ and $\ell_n = 2$, respectively.

Theorem 5 *Let X be of the form (1) and let $\ell_n, k_n \to \infty$ with $\ell_n = o(k_n)$ and $k_n = o(n)$.*

(a) *Suppose that X is continuous and assume that Condition 2 holds. We have*

$$V_n \xrightarrow{\mathbb{P}} V = (m_{2p} - m_p^2) \int_0^1 \sigma_s^{2p} \mathrm{d}s.$$

(b) *Let $p > 3$ and suppose that X allows for jumps and that X and σ never jump at the same time. Under Condition 1 , we have*

$$V_n \xrightarrow{\mathbb{P}} V = \sum_{0<s\leq 1} p^2 |\Delta X_s|^{2p-2} \sigma_s^2.$$

(c) *Let $p = 2$ and suppose that X allows for jumps and that Xand σ never jump at the same time. Under Condition 1 , we have*

$$V_n \xrightarrow{\mathbb{P}} V = 2\int_0^1 \sigma_s^4 \mathrm{d}s + \sum_{0<s\leq 1} 4|\Delta X_s|^2 \sigma_s^2.$$

Remark 5

(i) An important question regards the choice of k_n and ℓ_n in Theorem 5. While the result itself is rather general and does not come with a rate of convergence per se, we can say more under some additional, mostly standard, assumptions. Let $p \geq 1$. If we assume finite activity jumps, and if for all stopping times $S \leq T$ and all $q > 0$ both

$$\mathbb{E}\Big[\sup_{s\in[S,T]} |\sigma_s - \sigma_S|^q\Big] \leq C_q \mathbb{E}\big[(T-S)^{q/2}\big]$$

and

$$\big|\mathbb{E}[\sigma_T^q - \sigma_S^q]\big| \leq C_q \mathbb{E}[T-S]$$

hold, where C_q is a constant which might depend on q, then a tedious but straightforward computation proves

$$V_n - V = O_{\mathbb{P}}\left(\sqrt{k_n\Delta_n} + \sqrt{\frac{1}{\ell_n}} + \sqrt{\frac{\ell_n}{k_n}}\right).$$

If k_n and ℓ_n are of the respective orders n^a and n^b then clearly $a = 2/3$, $b = 1/3$ gives $V_n - V = O_{\mathbb{P}}(n^{-1/6})$. Note further that in the purely continuous case (a) the latter error term disappears and the choice of $\ell_n = k_n$ becomes possible. In this case one chooses $a = 1/2$ and obtains $V_n - V = O_{\mathbb{P}}(n^{-1/4})$. The same order can be attained in the jump dominated case (b) where the second error term vanishes and $\ell_n = 2$ is a fine choice.

(ii) The approach presented in this work is tailored for a conditional variance associated with Theorem 2, so for sums of functionals computed over *non-*

overlapping intervals. In more general situations, the conditional variance in the central limit theorem contains terms which are due to the covariance between neighbouring intervals, and these terms need to be estimated as well. At least in the continuous case there is some remedy. Let us discuss bipower variation as a specific example where the statistics are of the form

$$U_n = \frac{1}{n} \sum_{i=1}^{n-1} F\Big(\frac{1}{\sqrt{n}} \Delta_i^n X, \frac{1}{\sqrt{n}} \Delta_{i+1}^n X) = \Delta_n^{1-(p+q)/2} \sum_{i=1}^{n-1} |\Delta_i^n X|^p |\Delta_{i+1}^n X|^q$$

with $F(x, y) = |x|^p |y|^q$, so (essentially) every $\Delta_i^n X$ appears twice within U_n. An associated central limit theorem can be found in Theorem 11.2.1 of Jacod and Protter (2012). In this case, the natural analogon for $\widehat{V}_n$ is built from

$$U^n_{[i\Delta_n,(i+2k_n)\Delta_n]} = (k_n \Delta_n)^{1-(p+q)/2} |X_{(i+k_n)\Delta_n - X_{i\Delta_n}}|^p |X_{(i+2k_n)\Delta_n} - X_{(i+k_n)\Delta_n}|^q$$

and

$$\widehat{\theta}_{[i\Delta_n,(i+2k_n)\Delta_n]} = \sum_{j=1}^{k_n} \Delta_n^{1-(p+q)/2} |\Delta^n_{i+2j-1} X|^p |\Delta^n_{i+2j} X|^q$$

and reads as

$$\widehat{V}_n = \frac{n}{k_n(k_n - 1)} \sum_{i=0}^{n-2k_n} \Big(U^n_{[i\Delta_n,(i+2k_n)\Delta_n]} - \widehat{\theta}_{[i\Delta_n,(i+2k_n)\Delta_n]} \Big)^2.$$

Clearly it is not consistent as it does not estimate the covariance due to $\Delta_i^n X$ appearing twice, but a version which also incorporates neighbouring intervals is. Precisely, the same methods as for Theorem 3 (a) prove that

$$\begin{aligned}
&\frac{n}{k_n(k_n-1)} \sum_{i=k_n}^{n-3k_n} \Big(U^n_{[i\Delta_n,(i+2k_n)\Delta_n]} - \widehat{\theta}_{[i\Delta_n,(i+2k_n)\Delta_n]} \Big) \\
&\times \Big(\big(U^n_{[(i-k_n)\Delta_n,(i+k_n)\Delta_n]} - \widehat{\theta}_{[(i-k_n)\Delta_n,(i+k_n)\Delta_n]} \big) + \big(U^n_{[i\Delta_n,(i+2k_n)\Delta_n]} - \widehat{\theta}_{[i\Delta_n,(i+2k_n)\Delta_n]} \big) \\
&+ \big(U^n_{[(i+k_n)\Delta_n,(i+3k_n)\Delta_n]} - \widehat{\theta}_{[(i+k_n)\Delta_n,(i+3k_n)\Delta_n]} \big) \Big)
\end{aligned}$$

consistently estimates the variance of bipower variation. Generalizations of this form are thus necessary when working with overlapping intervals. □

4 Conclusion

In this paper, we have presented a new class of estimators for the asymptotic (conditional) variance in limit theorems for semimartingales. These estimators are only based on the form of the original statistics

$$U_n = \sum_{i=1}^{n} f_n(\Delta_i^n X)$$

in the central limit theorem, and we have shown in Theorem 5 that they are consistent for power variations in all three possible regimes: for a dominating continuous martingale part, for dominating jumps, and for the quadratic variation.

Even though the estimator V_n discussed in Theorem 5 gives a positive answer to the question whether such universal estimators exist, its application in practice is difficult, as we need to compute statistics over each of the $\binom{k_n}{\ell_n}$ subintervals within $(i\Delta_n, (i+k_n)\Delta_n]$ in order to obtain V_n. From a computational point of view, this is certainly not a reasonable strategy, at least under the conditions $\ell_n \to \infty$ and $\ell_n = o(k_n)$. The other estimators $\widehat{V}_n$ and $\widetilde{V}_n$ are constructed with $\ell_n = k_n$ and $\ell_n = 2$, respectively, so they are computationally much less expensive, though not consistent in all situations.

Future research clearly needs to investigate the practical properties of this new class of estimators, for $\widehat{V}_n$ in comparison with Mykland and Zhang (2017) and Christensen et al. (2017) in the continuous case, but also with a focus towards the properties of $\widetilde{V}_n$ in the case of quadratic variation. This new estimator is consistent in all situations, with jumps or not, so one does not need to test in advance whether jumps are present in the path of X or not. Further, as hinted at in Remark 5, extensions of the estimator are necessary when working with overlapping intervals, such as for multipower variation or pre-averaging estimators when microstructure noise is present. While the continuous case can be treated with similar methods, additional research becomes particularly important when jumps dominate.

5 Proofs

Throughout the proofs we will assume that the processes (b_s), (σ_s) and (X_s) are bounded, and we will also assume that $|\delta(s,z)|$ is bounded by a deterministic function $\gamma(z)$ satisfying $\int \gamma^2(z)\lambda(dz) < \infty$. In fact, according to Condition 1 we know that (b_s) and $(\delta(s,z))$ safisfy such claims locally, and we also know that (σ_s) is càdlağ, and then a standard localization procedure as in Section 4.4.1 in Jacod and Protter (2012) shows that we may assume global bounds without loss of generality. Similarly, whenever we explicitly need Condition 2, we may further assume that $(\widetilde{\sigma}_s)$, $(\widetilde{b}_s)$, (α_s) and (α'_s) are bounded as well, and we may also assume that (σ_s) is bounded away from zero. Also, $C > 0$ denotes a universal constant which may change from line to line, and we write C_r whenever we want to emphasize dependence of the constant on an auxiliary parameter such as r.

We introduce the decomposition $X_t = X_0 + B(q)_t + X_t^c + M(q)_t + N(q)_t$ of the Itô semimartingale (1) with

$$\begin{aligned} B(q)_t &= \int_0^t \Big(b_s - \int (\delta(s,z)\mathbb{1}_{\{|\delta(s,z)|\le 1\}} - \delta(s,z)\mathbb{1}_{\{\gamma(z)\le 1/q\}})\lambda(\mathrm{d}z)\Big)\mathrm{d}s, \\ X_t^c &= \int_0^t \sigma_s \mathrm{d}W_s, \\ M(q)_t &= \int_0^t \int \delta(s,z)\mathbb{1}_{\{\gamma(z)\le 1/q\}}(\mu-\nu)(\mathrm{d}s,\mathrm{d}z), \\ N(q)_t &= \int_0^t \int \delta(s,z)\mathbb{1}_{\{\gamma(z)>1/q\}}\mu(\mathrm{d}s,\mathrm{d}z). \end{aligned}$$

Here q is a parameter which controls whether jumps are classified as small jumps or big jumps. We also set $X(q)_t = B(q)_t + X_t^c + M(q)_t$ and denote the derivative process of $B(q)$ with $b(q)$. From the integrability condition on γ one immediately obtains $|b(q)| \le Cq$.

5.1 Proof of Example 1

Let A be the subset of Ω such that J contains exactly one jump in $(0, 1)$ and that the jump time S is in $(0,1)\backslash\mathbb{Q}$. Obviously, $\mathbb{P}(A) > 0$, and it is sufficient to prove that both $nT_n\mathbb{1}_A$ and $\widehat{\Sigma}\mathbb{1}_A$ diverge to infinity in probability.

For T_n, on A, suppose that n is large enough such that $k_n\Delta_n < S < 1 - k_n\Delta_n$. Then, each

$$QV_n(k_n) = \frac{1}{k_n}\sum_{i=k_n}^{n-k_n}(\widehat{\theta}_{[(i-k_n)\Delta_n, i\Delta_n]} - \widehat{\theta}_{[i\Delta_n,(i+k_n)\Delta_n]})^2$$

consists of $2k_n$ summands which are affected by the one jump and of $n - 4k_n + 1$ summands which are not. Suppose, for example, that $i = \lceil nS \rceil$. Then,

$$\begin{aligned} &\widehat{\theta}_{[(i-k_n)\Delta_n, i\Delta_n]} - \widehat{\theta}_{[i\Delta_n,(i+k_n)\Delta_n]} \\ &\quad = \sum_{j=1}^{k_n-1} \sigma^2(|\Delta_{i-k_n+j}^n W|^2 - |\Delta_{i+j}^n W|^2) + |1 + \sigma\Delta_i^n W|^2 - \sigma^2|\Delta_{i+k_n}^n W|^2 \\ &\quad = 1 + 2\sigma\Delta_i^n W + \sum_{j=1}^{k_n} \sigma^2(|\Delta_{i-k_n+j}^n W|^2 - |\Delta_{i+j}^n W|^2) = 1 + O_{\mathbb{P}}(\sqrt{\Delta_n}), \end{aligned}$$

where we have used $k_n\Delta_n \to 0$. Consequently,

$$\frac{1}{k_n}\sum_{i=\lceil S\Delta_n^{-1}\rceil-k_n}^{\lceil S\Delta_n^{-1}\rceil+k_n-1}(\widehat{\theta}_{[(i-k_n)\Delta_n,i\Delta_n]}-\widehat{\theta}_{[i\Delta_n,(i+k_n)\Delta_n]})^2=2+O_{\mathbb{P}}(\sqrt{\Delta_n}).$$

The sum over the remaining $n-4k_n+1$ terms asymptotically behaves in the same way as the entire $QV_n(k_n)$ in the case without jumps and is of order Δ_n according to Theorem 4 of Mykland and Zhang (2017). Therefore,

$$T_n=\frac{2}{3}\big(QV_n(k_n)-\frac{1}{4}QV_n(2k_n)\big)=1+O_{\mathbb{P}}(\sqrt{\Delta_n}),$$

and nT_n diverges on A.

Similarly, on the set A we have that only one of the statistics U_l^n contains the increment with the one jump, whereas the remaining k_n-1 intervals are not affected by it. Therefore, each of the latter statistics satisfies $U_l^n-U_n=O_{\mathbb{P}}(1)$ as restricted to A both statistics converge in probability to σ^2 and σ^2+1, respectively. We conclude that

$$\widehat{\Sigma}_n=\frac{1}{k_n}\sum_{l=1}^{k_n}(k_n\Delta_n)^{-1}(U_l^n-U_n)^2=O_{\mathbb{P}}((k_n\Delta_n)^{-1})$$

on A, so it does not converge as well. □

5.2 Proof of Theorems 3, 4 and 5

We will proceed as follows: in all cases, we will only show parts (a) and (b), and we will discuss these in separate sections. The proof of part (c) mostly just combines the ideas from (a) and (b) after one separates intervals with and without jumps of $N(q)$. Within each section, we will start with the result from Theorem 5 which we will prove in essentially all details. Afterwards, we discuss the necessary changes for Theorems 3 and 4. Note that we can use analogous proofs for most parts because the estimators are essentially all the same, just with ℓ_n varying between 2 and k_n.

Before we begin with the proofs of the main theorems, we provide a key lemma which will be used extremely often throughout the remaining sections.

Lemma 1 *Let*

$$X_n=\sum_{i=1}^{n-k_n}(\chi_i^n)^2$$

and suppose that there exists

$$R_n=\sum_{i=1}^{n-k_n}(\rho_i^n)^2$$

such that $R_n\xrightarrow{\mathbb{P}}X$ *(or* $R_n\xrightarrow{\mathcal{L}-(s)}X$*) and*

$$\sum_{i=1}^{n-k_n} (\chi_i^n - \rho_i^n)^2 \xrightarrow{\mathbb{P}} 0. \tag{5}$$

Then $X_n \xrightarrow{\mathbb{P}} X$ (*or* $X_n \xrightarrow{\mathcal{L}} X$).

Proof For both claims we only have to show $X_n - R_n \xrightarrow{\mathbb{P}} 0$. Note that for each $\varepsilon > 0$ there exists some $C_\varepsilon > 0$ such that

$$|(x+y)^2 - x^2| \le \varepsilon x^2 + C_\varepsilon y^2, \tag{6}$$

which is a simple consequence of Young's inequality. Therefore,

$$|X_n - R_n| \le \varepsilon R_n + C_\varepsilon \sum_{i=1}^{n-k_n} (\chi_i^n - \rho_i^n)^2,$$

and we obtain

$$\limsup_{n\to\infty} \mathbb{P}(|X_n - R_n| > \delta) \le \limsup_{n\to\infty} \mathbb{P}\Big(R_n \ge \frac{\delta}{2\varepsilon}\Big) \le \mathbb{P}\Big(X \ge \frac{\delta}{2\varepsilon}\Big)$$

for each fixed ε, where we have first used (5) and the Portmanteau theorem plus $R_n \xrightarrow{\mathcal{L}} X$ afterwards. Letting $\varepsilon \to 0$ then finishes the proof. □

5.2.1 Proof of part (a)

We will start with Theorem 5 and discuss V_n. In the situation of a continuous X a simple computation using the respective standardization of f_n shows that the estimator reads as

$$V_n = \frac{\ell_n}{n(\ell_n - 1)} \sum_{i=0}^{n-k_n} \frac{1}{\binom{k_n}{\ell_n}} \sum_{1 \le j_1 < \ldots < j_{\ell_n} \le k_n} \Big(V^n_{i+j_1,\ldots,i+j_{\ell_n}} \Big)^2$$

with

$$V^n_{i+j_1,\ldots,i+j_{\ell_n}} = (\ell_n \Delta_n)^{-p/2} \Big| \sum_{m=1}^{\ell_n} \Delta^n_{i+j_m} X \Big|^p - \frac{1}{\ell_n} \sum_{m=1}^{\ell_n} \Delta_n^{-p/2} |\Delta^n_{i+j_m} X|^p.$$

The main strategy in the proof of $V_n \xrightarrow{\mathbb{P}} V$ is to apply Lemma 1 several times, which means that one successively replaces V_n by simpler terms until one ends up with

$$\overline{V}_n = \frac{1}{n} \sum_{i=0}^{n-k_n} \sigma^{2p}_{i\Delta_n} \frac{1}{\binom{k_n}{\ell_n}} \sum_{1 \le j_1 < \ldots < j_{\ell_n} \le k_n} \Big((\ell_n \Delta_n)^{-p/2} \Big| \sum_{m=1}^{\ell_n} \Delta^n_{i+j_m} W \Big|^p - m_p \Big)^2.$$

We first prove $\overline{V}_n \xrightarrow{\mathbb{P}} V$ for which we set $h_n(x_1, \ldots, x_{\ell_n}) = |\ell_n^{-1/2}(x_1 + \ldots + x_{\ell_n})|^p$ and

$$U_i^n = \frac{1}{\binom{k_n}{\ell_n}} \sum_{1\le j_1<\ldots<j_{\ell_n}\le k_n} \Big(h_n(\Delta_n^{-1/2}\Delta_{i+j_1}^n W,\ldots,\Delta_n^{-1/2}\Delta_{i+j_{\ell_n}}^n W) - m_p\Big)^2.$$

Clearly, $\mathbb{E}[U_i^n] = m_{2p} - m_p^2$, and because of conditional independence, boundedness of (σ_s) and the Cauchy–Schwarz inequality we also have

$$\mathbb{E}\Big[\Big(\frac{1}{n}\sum_{i=0}^{n-k_n} \sigma_{i\Delta_n}^{2p}(U_i^n - \mathbb{E}[U_i^n])\Big)^2\Big] \le \frac{C}{n^2}\sum_{i,r=0}^{n-k_n} \mathbb{1}_{\{|i-r|\le k_n\}}\sqrt{\mathrm{Var}\,(U_i^n)\,\mathrm{Var}\,(U_r^n)}. \quad (7)$$

Using Theorem 1.2.3 in Denker (1985) on an upper bound for the variance of a U statistic we obtain

$$\mathrm{Var}\,(U_i^n) \le \frac{\ell_n}{k_n}\,\mathrm{Var}\Big((h_n(\Delta_n^{-1/2}\Delta_{i+j_1}^n W,\ldots,\Delta_n^{-1/2}\Delta_{i+j_{\ell_n}}^n W) - m_p)^2\Big) \le C\frac{\ell_n}{k_n},$$

so as a consequence of $\ell_n\Delta_n \to 0$

$$\overline{V}_n = \frac{1}{n}\sum_{i=0}^{n-k_n}\sigma_{i\Delta_n}^{2p}U_i^n = \frac{1}{n}\sum_{i=0}^{n-k_n}\sigma_{i\Delta_n}^2(m_{2p}-m_p^2) + o_{\mathbb{P}}(1).$$

Convergence in probability of the latter quantity to V is standard.

It remains to prove that the simplification to $\overline{V}_n$ is adequate. We first show $\underline{V}_n \xrightarrow{\mathbb{P}} V$ for

$$\underline{V}_n = \frac{\ell_n}{n(\ell_n-1)}\sum_{i=0}^{n-k_n}\sigma_{i\Delta_n}^{2p}\frac{1}{\binom{k_n}{\ell_n}}\sum_{1\le j_1<\ldots<j_{\ell_n}\le k_n}\Big(\underline{V}_{i+j_1,\ldots,i+j_{\ell_n}}^n\Big)^2$$

with

$$\underline{V}_{i+j_1,\ldots,i+j_{\ell_n}}^n = (\ell_n\Delta_n)^{-p/2}\Big|\sum_{m=1}^{\ell_n}\Delta_{i+j_m}^n W\Big|^p - \frac{1}{\ell_n}\sum_{m=1}^{\ell_n}\Delta_n^{-p/2}|\Delta_{i+j_m}^n W|^p.$$

Using Lemma 1, boundedness of (σ_s) and $\overline{V}_n \xrightarrow{\mathbb{P}} V$ we just have to establish

$$\frac{1}{n}\sum_{i=0}^{n-k_n}\frac{1}{\binom{k_n}{\ell_n}}\sum_{1\le j_1<\ldots<j_{\ell_n}\le k_n}\Big(\frac{1}{\ell_n}\sum_{m=1}^{\ell_n}\Delta_n^{-p/2}|\Delta_{i+j_m}^n W|^p - m_p\Big)^2 \xrightarrow{\mathbb{P}} 0 \quad (8)$$

in order to show $\frac{\ell_n-1}{\ell_n}\underline{V}_n \xrightarrow{\mathbb{P}} V$, and the claim regarding $\underline{V}_n$ then follows from $\ell_n \to \infty$. Note that

$$\overline{T}^n_{i+j_1,\ldots,i+j_{\ell_n}} = \left(\frac{1}{\ell_n} \sum_{m=1}^{\ell_n} \Delta_n^{-p/2} |\Delta^n_{i+j_m} W|^p - m_p \right)^2$$

satisfies $\mathbb{E}[|\overline{T}^n_{i+j_1,\ldots,i+j_{\ell_n}}|^2] \leq C/\ell_n$ by independence of the Brownian increments, so that (8) follows from

$$\mathbb{E}\left[\frac{1}{n} \sum_{i=0}^{n-k_n} \frac{1}{\binom{k_n}{\ell_n}} \sum_{1\leq j_1<\ldots<j_{\ell_n}\leq k_n} \left(\frac{1}{\ell_n} \sum_{m=1}^{\ell_n} \Delta_n^{-p/2} |\Delta^n_{i+j_m} W|^p - m_p \right)^2 \right] \leq \frac{C}{\ell_n} \to 0.$$

Finally, another application of Lemma 1 together with $\underline{V}_n \xrightarrow{\mathbb{P}} V$, plus the obvious $(v+w)^2 \leq 2(v^2+w^2)$, shows that the proof of $V_n \xrightarrow{\mathbb{P}} V$ boils down to showing

$$\frac{\ell_n}{n(\ell_n-1)} \sum_{i=0}^{n-k_n} \frac{1}{\binom{k_n}{\ell_n}} \sum_{1\leq j_1<\ldots<j_{\ell_n}\leq k_n} \Delta_n^{-p} \left(\frac{1}{\ell_n} \sum_{m=1}^{\ell_n} \left(|\Delta^n_{i+j_m} X|^p - \sigma^p_{i\Delta_n} |\Delta^n_{i+j_m} W|^p \right) \right)^2 \xrightarrow{\mathbb{P}} 0 \tag{9}$$

as well as

$$\frac{\ell_n}{n(\ell_n-1)} \sum_{i=0}^{n-k_n} \frac{1}{\binom{k_n}{\ell_n}} \sum_{1\leq j_1<\ldots<j_{\ell_n}\leq k_n} \underline{T}^n_{i+j_1,\ldots,i+j_{\ell_n}} \xrightarrow{\mathbb{P}} 0 \tag{10}$$

with

$$\underline{T}^n_{i+j_1,\ldots,i+j_{\ell_n}} = (\ell_n \Delta_n)^{-p} \left(|\sum_{m=1}^{\ell_n} \Delta^n_{i+j_m} X|^p - \sigma^p_{i\Delta_n} |\sum_{m=1}^{\ell_n} \Delta^n_{i+j_m} W|^p \right)^2.$$

The proof is similar for both claims, and we will only prove (10) in detail.

To this end, let $\psi : \mathbb{R} \to \mathbb{R}$ be a smooth function such that

$$\mathbb{1}_{[1,\infty)}(x) \leq \psi(x) \leq \mathbb{1}_{[1/2,\infty)}(x),$$

and for any $A > 0$ and $p > 0$ we set

$$\psi_A(x) = \psi\left(\frac{|x|}{A}\right), \quad \psi'_A(x) = 1 - \psi_A(x), \quad \psi_{A,p}(x) = \psi_A(x)|x|^p, \quad \psi'_{A,p}(x) = \psi'_A(x)|x|^p.$$

Clearly,

$$\underline{T}^n_{i+j_1,\ldots,i+j_{\ell_n}} \leq 2(\underline{T}^n_{i+j_1,\ldots,i+j_{\ell_n},A} + \underline{T}'^n_{i+j_1,\ldots,i+j_{\ell_n},A})$$

with

$$\underline{T}^n_{i+j_1,\dots,i+j_{\ell_n},A} = \left(\psi_{A,p}\left(\frac{\sum_{m=1}^{\ell_n}\Delta^n_{i+j_m}X}{\sqrt{\ell_n\Delta_n}}\right) - \psi_{A,p}\left(\sigma_{i\Delta_n}\frac{\sum_{m=1}^{\ell_n}\Delta^n_{i+j_m}W}{\sqrt{\ell_n\Delta_n}}\right)\right)^2$$

and similarly for $\underline{T}'^n_{i+j_1,\dots,i+j_{\ell_n},A}$, but with $\psi_{A,p}$ replaced by $\psi'_{A,p}$. (10) then follows from

$$\lim_{A\to\infty}\limsup_{n\to\infty}\frac{\ell_n}{n(\ell_n-1)}\sum_{i=0}^{n-k_n}\frac{1}{\binom{k_n}{\ell_n}}\sum_{1\le j_1<\dots<j_{\ell_n}\le k_n}\mathbb{E}\left[\underline{T}^n_{i+j_1,\dots,i+j_{\ell_n},A}\right] = 0$$

and

$$\lim_{n\to\infty}\frac{\ell_n}{n(\ell_n-1)}\sum_{i=0}^{n-k_n}\frac{1}{\binom{k_n}{\ell_n}}\sum_{1\le j_1<\dots<j_{\ell_n}\le k_n}\mathbb{E}\left[\underline{T}'^n_{i+j_1,\dots,i+j_{\ell_n},A}\right] = 0$$

for every fixed $A>0$. The first claim can be quickly deduced from

$$\psi_{A,p}(x) = \psi\left(\frac{|x|}{A}\right)|x|^p \le \mathbb{1}_{\{2|x|\ge A\}}|x|^p \le \frac{2|x|^{p+1}}{A},$$

$(v+w)^2 \le 2(v^2+w^2)$ and, e.g.

$$\mathbb{E}\left[\left|\frac{\sum_{m=1}^{\ell_n}\Delta^n_{i+j_m}X}{\sqrt{\ell_n\Delta_n}}\right|^{2p+2}\right] \le C$$

which is a consequence of the Burkholder–Davis–Gundy inequality and the boundedness assumption for (b_s) and (σ_s).

So let finally be A fixed. It is easy to see that $\psi'_{A,p}$ is bounded and uniformly continuous, and it follows that

$$\theta(\varepsilon) = \sup_{x\in\mathbb{R},|y|\le\varepsilon}\left|\psi'_{A,p}(x+y)-\psi'_{A,p}(x)\right| \to 0$$

as $\varepsilon\to 0$. In particular,

$$\left|\psi'_{A,p}(x+y)-\psi'_{A,p}(x)\right| \le \theta(\varepsilon) + \left|\psi'_{A,p}(x+y)-\psi'_{A,p}(x)\right|\mathbb{1}_{\{|y|>\varepsilon\}} \le \theta(\varepsilon) + C_A\frac{y^2}{\varepsilon^2}.$$

By first letting $n\to\infty$ and then $\varepsilon\to 0$ it is thus sufficient to prove

$$\frac{\ell_n}{n(\ell_n-1)}\sum_{i=0}^{n-k_n}\frac{1}{\binom{k_n}{\ell_n}}\sum_{1\le j_1<\dots<j_{\ell_n}\le k_n}\mathbb{E}\left[\left(\frac{\sum_{m=1}^{\ell_n}\Delta^n_{i+j_m}X - \sigma_{i\Delta_n}\sum_{m=1}^{\ell_n}\Delta^n_{i+j_m}W}{\sqrt{\ell_n\Delta_n}}\right)^2\right] \to 0$$

as $n \to \infty$. Using $(v+w)^2 \le 2(v^2+w^2)$ once more, we can discuss the absolutely continuous part of the increments and the Brownian parts separately, and the proof for the first terms follows from

$$\frac{1}{\ell_n \Delta_n} \mathbb{E}\left[\left(\int_{(i+j_1-1)\Delta_n}^{(i+j_1)\Delta_n} b_s \mathrm{d}s + \ldots + \int_{(i+j_{\ell_n}-1)\Delta_n}^{(i+j_{\ell_n})\Delta_n} b_s \mathrm{d}s\right)^2\right] \le C\ell_n \Delta_n \to 0.$$

We can thus assume $dX_t = \sigma_t \mathrm{d}W_t$, and we will first prove the result in the case of a continuous σ. We have

$$\begin{aligned}
&\frac{1}{\ell_n \Delta_n} \mathbb{E}\left[\left(\int_{(i+j_1-1)\Delta_n}^{(i+j_1)\Delta_n} (\sigma_s - \sigma_{i\Delta_n}) \mathrm{d}W_s + \ldots + \int_{(i+j_{\ell_n}-1)\Delta_n}^{(i+j_{\ell_n})\Delta_n} (\sigma_s - \sigma_{i\Delta_n}) \mathrm{d}W_s\right)^2\right] \\
=&\frac{1}{\ell_n \Delta_n} \mathbb{E}\left[\int_{(i+j_1-1)\Delta_n}^{(i+j_1)\Delta_n} (\sigma_s - \sigma_{i\Delta_n})^2 \mathrm{d}s + \ldots + \int_{(i+j_{\ell_n}-1)\Delta_n}^{(i+j_{\ell_n})\Delta_n} (\sigma_s - \sigma_{i\Delta_n})^2 \mathrm{d}s\right],
\end{aligned}$$

so that

$$\begin{aligned}
&\frac{1}{\ell_n - 1} \sum_{i=0}^{n-k_n} \frac{1}{\binom{k_n}{\ell_n}} \sum_{1 \le j_1 < \ldots < j_{\ell_n} \le k_n} \mathbb{E}\left[\left(\sum_{m=1}^{\ell_n} \int_{(i+j_m-1)\Delta_n}^{(i+j_m)\Delta_n} (\sigma_s - \sigma_{i\Delta_n}) \mathrm{d}W_s\right)^2\right] \\
=&\frac{\ell_n}{\ell_n - 1} \frac{1}{k_n} \sum_{i=0}^{n-k_n} \int_{i\Delta_n}^{(i+k_n)\Delta_n} (\sigma_s - \sigma_{i\Delta_n})^2 \mathrm{d}s \le \int_0^1 \frac{2}{k_n} \sum_{m=0}^{k_n-1} \mathbb{E}[(\sigma_s - \sigma_{(\lfloor ns \rfloor - m)^+ \Delta_n})^2] \mathrm{d}s
\end{aligned}$$

where we have used that every interval $[(i+j_m-1)\Delta_n, (i+j_m)\Delta_n]$ appears $\binom{k_n - 1}{\ell_n - 1}$ times and

$$\ell_n \binom{k_n}{\ell_n} = k_n \binom{k_n - 1}{\ell_n - 1}, \tag{11}$$

plus $\ell_n \le 2(\ell_n - 1)$ for any $\ell_n \ge 2$. Convergence to zero in probability then follows from continuity of σ and dominated convergence.

In the general case we use the reasoning from Lemma 3.4.8 in Jacod and Protter (2012). A standard argument using $\int_0^1 \sigma_s^2 \mathrm{d}s \le C$ proves the existence of a sequence $\sigma(u)$ of adapted continuous processes $\sigma(u)$ such that

$$\mathbb{E}\left[\int_0^1 \left(\sigma_s - \sigma(u)\right)^2 \mathrm{d}s\right] \to 0 \tag{12}$$

as $u \to \infty$. Thus, setting $X(u)_t = X_0 + \int_0^t b_s \mathrm{d}s + \int_0^t \sigma(u)_s \mathrm{d}W_s$, we have already shown

$$V_n(u) \xrightarrow{\mathbb{P}} V(u) = (m_{2p} - m_p^2) \int_0^1 \sigma(u)_s^p \mathrm{d}s$$

as $n \to \infty$, where $V_n(u)$ denotes the statistic V_n, but based on $X(u)$. Clearly, $V(u) \xrightarrow{\mathbb{P}} V$ as $u \to \infty$ as well, so it remains to prove

$$\lim_{u\to\infty} \limsup_{n\to\infty} \mathbb{P}(|V_n - V_n(u)| > \eta) = 0$$

for every $\eta > 0$. Using (6) one has to deal with similar claims as (9) and (10), but with $\sigma = 1$ and where W becomes $X(u)$. Reproducing these lines the proof finally follows from (12).

For Theorem 3 the proof holds without any changes, because we have only used $\ell_n \Delta_n \to 0$ and $\ell_n \to \infty$ which holds for $\ell_n = k_n$ as well. The situation is different for Theorem 4 in which case

$$\underline{V}_n = \frac{1}{n} \sum_{i=0}^{n-k_n} \sigma_{i\Delta_n}^{2p} \underline{U}_i^n$$

for a U statistic of the form

$$\underline{U}_i^n = \frac{1}{\binom{k_n}{2}} \sum_{1 \le j_1 < j_2 \le k_n} 2\left(\underline{V}_{i+j_1,i+j_2}^n\right)^2$$

where

$$\begin{aligned}
\underline{V}_{i+j_1,i+j_2}^n &= (2\Delta_n)^{-p/2} |\Delta_{i+j_1}^n W + \Delta_{i+j_2}^n W|^p - \frac{1}{2}\Delta_n^{-p/2}(|\Delta_{i+j_1}^n W|^p + |\Delta_{i+j_2}^n W|^p) \\
&= \left|\frac{1}{\sqrt{2}}(N_{i+j_1} + N_{i+j_2})\right|^p - \frac{1}{2}(|N_{i+j_1}|^p + |N_{i+j_2}|^p)
\end{aligned}$$

and the latter equality is to be understood in distribution, with the N_{i+j} all independent standard normal. Setting

$$c_p = 2\mathbb{E}\left[\left(\left|\frac{1}{\sqrt{2}}(N_{i+j_1} + N_{i+j_2})\right|^p - \frac{1}{2}(|N_{i+j_1}|^p + |N_{i+j_2}|^p)\right)^2\right]$$

the same reasoning as for (7) gives

$$\underline{V}_n = c_p \frac{1}{n} \sum_{i=0}^{n-k_n} \sigma_{i\Delta_n}^{2p} + o_{\mathbb{P}}(1) = c_p \int_0^1 \sigma_s^{2p} \mathrm{d}s + o_{\mathbb{P}}(1).$$

The remainder of the proof remains unchanged. Note finally that

$$c_2 = \frac{1}{2}\mathbb{E}\Big[\big(|N_1 + N_2|^2 - (|N_1|^2 + |N_2|^2)\big)^2\Big] = 2\mathbb{E}[N_1^2 N_2^2] = 2.$$

□

5.2.2 Proof of part (b)

We define $L_m = \{z \mid \gamma(z) > 1/m\}$ for any $m \geq 1$, and let $\{S(m,j) \mid j \geq 1\}$ denote the jump times of the Poisson process $\mathbb{1}_{L_m \setminus L_{m-1}} \star \mu$ over [0, 1], where we use the notation from, e.g. Section 2.1.2 in Jacod and Protter (2012) to denote the integral process with respect to a jump measure. Then, if $(S_r)_{r\geq 1}$ is a reordering of the double sequence $(S(m,j))_{m,j\geq 1}$, we denote with P_q the set of all indices r such that $S_r = S(m,j)$ for some $m \leq q$. By definition, these are the jump times of $N(q)$ over [0, 1]. Further, let $\Omega(n,q)$ be the set of all ω on which $N(q)$ has at most one jump in each interval $[i\Delta_n, (i+k_n)\Delta_n]$, $i = 0, \ldots, n-k_n$, all jumps of $N(q)$ over [0, 1] occur within $[k_n\Delta_n, 1-k_n\Delta_n]$ and where

$$|X(q)(\omega)_{t+s} - X(q)(\omega)_t| \leq 2/q \text{ for all } t \in [0,1] \text{ and } s \in [0, k_n\Delta_n].$$

Since $X(q)$ is càdlàg with jumps bounded by $1/q$ (in absolute value) and $N(q)$ only possesses finitely many jumps on [0, 1], it is clear that $\mathbb{P}(\Omega(n,q)) \to 1$ as $n \to \infty$ for any $q > 0$. As we will typically let first $n \to \infty$ and then $q \to \infty$, we will sometimes assume $\omega \in \Omega(n,q)$.

We introduce the notation i_r to denote the interval $((i_r-1)\Delta_n, i_r\Delta_n]$ containing the rth jump ΔX_{S_r} of $N(q)$. In this case we have

$$V_n = \frac{n}{\ell_n(\ell_n - 1)} \sum_{i=0}^{n-k_n} \frac{1}{\binom{k_n}{\ell_n}} \sum_{1\leq j_1<\ldots<j_{\ell_n}\leq k_n} \left(\Big|\sum_{m=1}^{\ell_n} \Delta_{i+j_m}^n X\Big|^p - \sum_{m=1}^{\ell_n} |\Delta_{i+j_m}^n X|^p\right)^2$$

and the key to the proof will be the decomposition $V_n = V_n(q) + V_n'(q)$ with

$$V_n(q) = \frac{n}{\ell_n(\ell_n - 1)} \sum_{r\in P_q} \sum_{\alpha=1}^{k_n} Y_{r,\alpha}^{(n)}, \qquad V_n'(q) = V_n - V_n(q),$$

and where

$$Y_{r,\alpha}^{(n)} = \frac{1}{\binom{k_n}{\ell_n}} \sum_{\substack{1 \leq j_1 < \ldots < j_{\ell_n} \leq k_n \\ \{j_1,\ldots,j_{\ell_n}\} \cap \{\alpha\} \neq \emptyset}} \left(\Big|\sum_{m=1}^{\ell_n} \Delta_{i_r-\alpha+j_m}^n X\Big|^p - |\Delta_{i_r}^n X|^p\right)^2.$$

Clearly, the proof is finished once we have shown

$$V_n(q) \xrightarrow{\mathbb{P}} V(q) = \sum_{r \in P_q} p^2 |\Delta X_{S_r}|^{2p-2} \sigma_{S_r}^2 \tag{13}$$

as $n \to \infty$ for any fixed q,

$$V(q) \xrightarrow{\mathbb{P}} V = \sum_{0<s\le 1} p^2 |\Delta X_s|^{2p-2} \sigma_s^2 \tag{14}$$

as $q \to \infty$, as well as

$$\lim_{q\to\infty} \limsup_{n\to\infty} \mathbb{P}\big(|V_n'(q)| > \eta\big) = 0 \tag{15}$$

for all $\eta > 0$. Note that (14) is a direct consequence of monotone convergence. Regarding (15) we observe that increments of X and $X(q)$ coincide when no jump of $N(q)$ is present. Therefore, and using $\ell_n \le 2(\ell_n - 1)$ for $\ell_n \ge 2$, we have the inequality

$$|V_n'(q)| \le A_n(q) + B_n(q)$$

with

$$A_n(q) = \frac{2n}{\ell_n^2} \sum_{i=0}^{n-k_n} \frac{1}{\binom{k_n}{\ell_n}} \sum_{1\le j_1<\ldots<j_{\ell_n}\le k_n} \left(\Big| \sum_{m=1}^{\ell_n} \Delta_{i+j_m}^n X(q) \Big|^p - \sum_{m=1}^{\ell_n} |\Delta_{i+j_m}^n X(q)|^p \right)^2$$

and

$$B_n(q) = \frac{n}{\ell_n(\ell_n - 1)} \sum_{r\in P_q} \sum_{\alpha=1}^{k_n} |Z_{r,\alpha}^{(n)} - Y_{r,\alpha}^{(n)}|$$

for

$$Z_{r,\alpha}^{(n)} = \frac{1}{\binom{k_n}{\ell_n}} \sum_{\substack{1 \le j_1 < \ldots < j_{\ell_n} \le k_n \\ \{j_1,\ldots,j_{\ell_n}\} \cap \{\alpha\} \ne \emptyset}} \left(\Big| \sum_{m=1}^{\ell_n} \Delta_{i_r-\alpha+j_m}^n X \Big|^p - \sum_{m=1}^{\ell_n} |\Delta_{i_r-\alpha+j_m}^n X|^p \right)^2.$$

We will start with the first part of (15) and prove

$$\lim_{q\to\infty} \limsup_{n\to\infty} \mathbb{P}\big(A_n(q) > \eta\big) = 0, \tag{16}$$

for which we set

$$Y(q)_t = \int_{i\Delta_n}^t \mathbb{1}_{B^n_{i,j_1,\ldots,j_{\ell_n}}}(s) dX(q)_s, \quad t \ge i\Delta_n, \tag{17}$$

where we use the shorthand notation

$$B = B^n_{i,j_1,\ldots,j_{\ell_n}} = ((i+j_1-1)\Delta_n,(i+j_1)\Delta_n] \cup \ldots \cup ((i+j_{\ell_n}-1)\Delta_n,(i+j_{\ell_n})\Delta_n].$$

We will basically apply (5.1.19) in Jacod and Protter (2012) which is stated for increments of $X(q)$ rather than for $Y(q)$, but the proof works similarly in our situation. Let us introduce some notation. We set $f(x) = |x|^p$ as well as

$$k(x,y) = f(x+y) - f(x) - f(y), \quad g(x,y) = k(x,y) - f'(x)y.$$

Then, we obtain

$$\begin{aligned}
&|\Delta^n_{i+j_1} X(q) + \ldots + \Delta^n_{i+j_{\ell_n}} X(q)|^p = f(Y(q)_{(i+k_n)\Delta_n}) \\
&= \sum_{i\Delta_n < s \le (i+k_n)\Delta_n} f(\Delta X(q)_s)\mathbb{1}_B(s) + A(n,q,B)_{(i+k_n)\Delta_n} + M(n,q,B)_{(i+k_n)\Delta_n},
\end{aligned}$$

where $M(n, q, B)$ is a square-integrable martingale with predictable bracket $A'(n,q,B)$, and where

$$A(n,q,B)_t = \int_{i\Delta_n}^t a(n,q,B)_u \mathrm{d}u, \quad A'(n,q,B) = \int_{i\Delta_n}^t a'(n,q,B)_u \mathrm{d}u,$$

with

$$\begin{aligned}
a(n,q,B)_u =& f'(Y(q)_{u-})b(q)_u \mathbb{1}_B(u) + \frac{1}{2} f''(Y(q)_{u-})\sigma_u^2 \mathbb{1}_B(u) \\
&+ \int g(Y(q)_{u-}, \delta(u,z))\mathbb{1}_{\{\gamma(z)\le 1/q\}} \mathbb{1}_B(u)\lambda(\mathrm{d}z)
\end{aligned}$$

and

$$a'(n,q,B)_u = (f'(Y(q)_{u-}))^2 \sigma_u^2 \mathbb{1}_B(u) + \int k(Y(q)_{u-}, \delta(u,z))^2 \mathbb{1}_{\{\gamma(z)\le 1/q\}} \mathbb{1}_B(u)\lambda(\mathrm{d}z).$$

Similarly,

$$\begin{aligned}
\sum_{m=1}^{\ell_n} |\Delta^n_{i+j_m} X|^p =& \sum_{i\Delta_n < s \le (i+k_n)\Delta_n} f(\Delta X(q)_s)\mathbb{1}_B(s) \\
&+ \sum_{m=1}^{\ell_n} (A(n,q,i+j_m-1)_{(i+j_m)\Delta_n} + M(n,q,i+j_m-1)_{(i+j_m)\Delta_n})
\end{aligned}$$

with $A(n,q,i+j_m-1)$ and $M(n,q,i+j_m-1)$ defined as above, but with B being replaced by $((i+j_m-1)\Delta_n,(i+j_m)\Delta_n]$, also in the definition of $Y(q)$. Thus, as the respective sums over the jumps $f(\Delta X(q)_s)$ cancel, $A_n(q)$ becomes

$$\frac{2n}{\ell_n^2}\sum_{i=0}^{n-k_n}\frac{1}{\binom{k_n}{\ell_n}}\sum_{1\le j_1<\ldots<j_{\ell_n}\le k_n}\Big(A(n,q,B^n_{i,j_1,\ldots,j_{\ell_n}})_{(i+k_n)\Delta_n}+M(n,q,B^n_{i,j_1,\ldots,j_{\ell_n}})_{(i+k_n)\Delta_n}$$
$$-\sum_{m=1}^{\ell_n}(A(n,q,i+j_m-1)_{(i+j_m)\Delta_n}+M(n,q,i+j_m-1)_{(i+j_m)\Delta_n})\Big)^2 \tag{18}$$

and in order to show (16) it becomes important to bound quantities like

$$\mathbb{E}[(A(n,q,B)_{(i+k_n)\Delta_n})^2] \quad \text{and} \quad \mathbb{E}[(M(n,q,B)_{(i+k_n)\Delta_n})^2]=\mathbb{E}[A'(n,q,B)_{(i+k_n)\Delta_n}].$$

A Taylor expansion gives $|k(x,y)|\le C\big(|x||y|^{p-1}+|y||x|^{p-1}\big)$ as well as $|g(x,y)|\le C\big(|x||y|^{p-1}+y^2|x|^{p-2}\big)$. From the boundedness conditions and integrability of $\gamma(z)$, we obtain

$$|a(n,q,B)_u|\le C\mathbb{1}_B(u)\big(q|Y(q)_{u-}|^{p-1}+|Y(q)_{u-}|^{p-2}+\alpha_q|Y(q)_{u-}|\big)$$
$$a'(n,q,B)_u\le C\mathbb{1}_B(u)\big(|Y(q)_{u-}|^{2p-2}+\alpha_q|Y(q)_{u-}|^2\big)$$

for some sequence α_q with $\alpha_q\to 0$ as $q\to\infty$, and an argument similar to (15.2.22) in Jacod and Protter (2012) gives

$$\mathbb{E}\Big[\sup_{u\le(i+k_n)\Delta_n}|Y(q)_{u-}|^r\Big]\le C\big(q^r(\ell_n\Delta_n)^r+(\ell_n\Delta_n)^{r/2}+\alpha_q(\ell_n\Delta_n)^{1\wedge(r/2)}\big)$$

where we have used $|B|=\ell_n\Delta_n$. Therefore,

$$\mathbb{E}\Big[\sup_{u\le(i+k_n)\Delta_n}a(n,q,B)_u^2\Big]\le C_q\ell_n\Delta_n$$

and

$$\mathbb{E}\Big[\sup_{u\le(i+k_n)\Delta_n}a'(n,q,B)_u\Big]\le C_q(\ell_n\Delta_n)^2+\alpha_q\ell_n\Delta_n.$$

To summarize,

$$\mathbb{E}[(A(n,q,B)_{(i+k_n)\Delta_n})^2]\le(\ell_n\Delta_n)^2\mathbb{E}\Big[\sup_{u\le(i+k_n)\Delta_n}a(n,q,B)_u^2\Big]\le C_q(\ell_n\Delta_n)^3 \tag{19}$$

and

$$\mathbb{E}[(M(n,q,B)_{(i+k_n)\Delta_n})^2]=\mathbb{E}[A'(n,q,B)_{(i+k_n)\Delta_n}]\le\ell_n\Delta_n\mathbb{E}\Big[\sup_{u\le(i+k_n)\Delta_n}a'(n,q,B)_u\Big]$$
$$\le C_q(\ell_n\Delta_n)^3+\alpha_q(\ell_n\Delta_n)^2. \tag{20}$$

Similar inequalities hold for $A(n,q,i+j_m-1)$ and $M(n,q,i+j_m-1)$, but with $\ell_n=1$. Then,

$$\mathbb{E}\Big[\Big(\sum_{m=1}^{\ell_n} A(n,q,i+j_m-1)_{(i+j_m)\Delta_n}\Big)^2\Big] \leq C_q \ell_n^2 \Delta_n^3 \tag{21}$$

and

$$\begin{aligned}&\mathbb{E}\Big[\Big(\sum_{m=1}^{\ell_n} M(n,q,i+j_m-1)_{(i+j_m)\Delta_n}\Big)^2\Big] \\ &\quad = \sum_{m=1}^{\ell_n} \mathbb{E}\big[A'(n,q,i+j_m-1)_{(i+j_m)\Delta_n}\big] \leq C_q \ell_n \Delta_n^3 + \alpha_q \ell_n \Delta_n^2.\end{aligned} \tag{22}$$

From (18) and the bounds in (19)–(22), we obtain

$$\mathbb{E}\big[A_n(q)\big] \leq C\big(C_q \ell_n \Delta_n + \alpha_q\big),$$

and the right-hand side goes to zero as first $n \to \infty$ and then $q \to \infty$. This finishes the proof of (16).

The proof of (15) is completed by showing

$$\lim_{q\to\infty} \limsup_{n\to\infty} \mathbb{P}\big(B_n(q) > \eta\big) = 0 \tag{23}$$

which we will do under the assumption that (13) holds. The proof of the latter claim will finish the entire section. Thus, let $\kappa > 0$ be arbitrary. Then there exists $K > 0$ such that $\mathbb{P}(V \geq K) \leq \kappa$, and from the Portmanteau theorem we deduce

$$\limsup_{q\to\infty} \limsup_{n\to\infty} \mathbb{P}(V_n(q) \geq K) \leq \limsup_{q\to\infty} \mathbb{P}(V(q) \geq K) \leq \mathbb{P}(V \geq K) \leq \kappa.$$

Let $\varepsilon \leq \frac{\eta}{3K}$. Then, using (6), we obtain

$$\begin{aligned}\mathbb{P}\big(B_n(q) > \eta\big) &\leq \mathbb{P}(\varepsilon V_n(q) \mathbb{1}_{\{V_n(q)\geq K\}} > \eta/3) + \mathbb{P}(\varepsilon V_n(q) \mathbb{1}_{\{V_n(q)<K\}} > \eta/3) \\ &+ \mathbb{P}\Bigg(C_\varepsilon \frac{n}{\ell_n(\ell_n-1)} \sum_{r\in P_q} \sum_{\alpha=1}^{k_n} \frac{1}{\binom{k_n}{\ell_n}} \sum_{\substack{1\leq j_1<\ldots<j_{\ell_n}\leq k_n \\ \{j_1,\ldots,j_{\ell_n}\}\cap\{\alpha\}\neq\emptyset}} \Big(|\Delta_{i_r}^n X|^p - \sum_{m=1}^{\ell_n} |\Delta_{i_r-\alpha+j_m}^n X|^p\Big)^2 > \eta/3\Bigg).\end{aligned}$$

For the first summand, we have

$$\limsup_{q\to\infty} \limsup_{n\to\infty} \mathbb{P}(\varepsilon V_n(q) \mathbb{1}_{\{V_n(q)\geq K\}} > \eta/3) \leq \mathbb{P}(V \geq K) \leq \kappa,$$

while for the second term

$$\mathbb{P}(\varepsilon V_n(q) \mathbb{1}_{\{V_n(q)<K\}} > \eta/3) \leq \mathbb{P}(\varepsilon K > \eta/3) = 0$$

by construction. As κ was arbitrary (23) follows, using $\ell_n \leq 2(\ell_n - 1)$ again, once we have shown

$$\lim_{q\to\infty}\limsup_{n\to\infty}\mathbb{P}\Bigg(\frac{2n}{\ell_n^2}\sum_{r\in P_q}\sum_{\alpha=1}^{k_n}\frac{1}{\binom{k_n}{\ell_n}}\sum_{\substack{1\le j_1<\ldots<j_{\ell_n}\le k_n\\ \{j_1,\ldots,j_{\ell_n}\}\cap\{\alpha\}\neq\emptyset}}\Big(|\Delta_{i_r}^nX|^p-\sum_{m=1}^{\ell_n}|\Delta_{i_r-\alpha+j_m}^nX|^p\Big)^2>\delta\Bigg)=0$$

for any $\delta>0$, and we may assume to live on $\Omega(n,q)$ without loss of generality. On this set the decomposition

$$\begin{aligned}|\Delta_{i_r}^nX|^p-\sum_{m=1}^{\ell_n}|\Delta_{i_r-\alpha+j_m}^nX|^p&=\bigg(\Big|\sum_{m=1}^{\ell_n}\Delta_{i_r-\alpha+j_m}^nX(q)\Big|^p-\sum_{m=1}^{\ell_n}|\Delta_{i_r-\alpha+j_m}^nX(q)|^p\bigg)\\&\quad-\Big|\sum_{m=1}^{\ell_n}\Delta_{i_r-\alpha+j_m}^nX(q)\Big|^p+|\Delta_{i_r}^nX(q)|^p\\&=I(n,q,i_r,\alpha,j_1,\ldots,j_{\ell_n})-II(n,q,i_r,\alpha,j_1,\ldots,j_{\ell_n})+III(n,q,i_r)\end{aligned}$$

holds, because each interval $[i\Delta_n,(i+k_n)\Delta_n]$, $i=0,\ldots,n-k_n$, contains at most one jump of $N(q)$. We will now prove

$$\lim_{q\to\infty}\limsup_{n\to\infty}\mathbb{P}\Bigg(\frac{2n}{\ell_n^2}\sum_{r\in P_q}\sum_{\alpha=1}^{k_n}\frac{1}{\binom{k_n}{\ell_n}}\sum_{\substack{1\le j_1<\ldots<j_{\ell_n}\le k_n\\ \{j_1,\ldots,j_{\ell_n}\}\cap\{\alpha\}\neq\emptyset}}I(n,q,i_r,\alpha,j_1,\ldots,j_{\ell_n})^2>\delta\Bigg)=0, \tag{24}$$

$$\lim_{q\to\infty}\limsup_{n\to\infty}\mathbb{P}\Bigg(\frac{2n}{\ell_n^2}\sum_{r\in P_q}\sum_{\alpha=1}^{k_n}\frac{1}{\binom{k_n}{\ell_n}}\sum_{\substack{1\le j_1<\ldots<j_{\ell_n}\le k_n\\ \{j_1,\ldots,j_{\ell_n}\}\cap\{\alpha\}\neq\emptyset}}II(n,q,i_r,\alpha,j_1,\ldots,j_{\ell_n})^2>\delta\Bigg)=0, \tag{25}$$

$$\lim_{q\to\infty}\limsup_{n\to\infty}\mathbb{P}\Bigg(\frac{2n}{\ell_n}\sum_{r\in P_q}III(n,q,i_r)^2>\delta\Bigg)=0, \tag{26}$$

and again restricted to $\Omega(n,q)$ if necessary. Note that the simplification in (26) is due to

$$\frac{n}{\ell_n^2}k_n\frac{1}{\binom{k_n}{\ell_n}}\binom{k_n-1}{\ell_n-1}=\frac{n}{\ell_n}. \tag{27}$$

Clearly, (24) is a simple consequence of (16), and the proof of (26) is essentially the same as for (25), but with $\ell_n=1$.

Thus, we will only prove (25), and we further introduce an auxiliary parameter $L\in\mathbb{N}$ and formally prove the equivalent convergence of (25) as first $n\to\infty$, then $L\to\infty$ and finally $q\to\infty$. Introducing the events $\mathbb{1}_{\{|P_q|\le L\}}$ and $\mathbb{1}_{\{|P_q|>L\}}$, where $|A|$ denotes the cardinality of a discrete set A, and from the fact that

$$\lim_{L\to\infty} \mathbb{P}(|P_q| > L) = 0$$

for any fixed q, it is clear that (25) follows from

$$\lim_{n\to\infty} \mathbb{P}\Big(\frac{2n}{\ell_n^2} \sum_{\substack{r\in P_q \\ |P_q|\le L}} \sum_{\alpha=1}^{k_n} \frac{1}{\binom{k_n}{\ell_n}} \sum_{\substack{1\le j_1<\ldots<j_{\ell_n}\le k_n \\ \{j_1,\ldots,j_{\ell_n}\}\cap\{\alpha\}\neq\emptyset}} II(n,q,i_r,\alpha,j_1,\ldots,j_{\ell_n})^2 > \delta\Big) = 0 \tag{28}$$

for any fixed q and L. As the sum over r is then finite, we may focus on a single arbitrary index i_r, and by properties of a Poisson measure we can also drop the dependence on the jumps of $\mathbb{1}_{L_q} \star \mu$ and simply write i. With the notation (17) we have

$$II(n,q,i,j_1,\ldots,j_{\ell_n}) \le \left| \int_{(i-k_n)\Delta_n}^{(i+k_n)\Delta_n} \mathbb{1}_{B^n_{i-\alpha,j_1,\ldots,j_{\ell_n}}}(s) \mathrm{d}X(q)_s \right|^p.$$

By definition $X(q)$ consists of three terms, and we will discuss each of them separately. The first two are easier to deal with, and we have

$$\mathbb{E}\Big[\Big| \int_{(i-k_n)\Delta_n}^{(i+k_n)\Delta_n} \mathbb{1}_{B^n_{i-\alpha,j_1,\ldots,j_{\ell_n}}}(s) b(q)_s \mathrm{d}s \Big|^{2p}\Big] \le C_q(\ell_n\Delta_n)^{2p}$$

and

$$\mathbb{E}\Big[\Big| \int_{(i-k_n)\Delta_n}^{(i+k_n)\Delta_n} \mathbb{1}_{B^n_{i-\alpha,j_1,\ldots,j_{\ell_n}}}(s) \sigma_s \mathrm{d}W_s \Big|^{2p}\Big] \le C(\ell_n\Delta_n)^p. \tag{29}$$

Together with (27) it is clear that (28) follows from

$$\mathbb{E}\Big[\Big(\frac{2n}{\ell_n^2} \sum_{\alpha=1}^{k_n} \frac{1}{\binom{k_n}{\ell_n}} \sum_{\substack{1\le j_1<\ldots<j_{\ell_n}\le k_n \\ \{j_1,\ldots,j_{\ell_n}\}\cap\{\alpha\}\neq\emptyset}} \Big(\int_{(i-k_n)\Delta_n}^{(i+k_n)\Delta_n} \int \delta(s,z)\mathbb{1}_B(s)\mathbb{1}_{\{\gamma(z)\le 1/q\}}(\mu-\nu)(\mathrm{d}s,\mathrm{d}z)\Big)^{2p}\Big) \wedge 1\Big] \to 0$$

where we again use the notation $B = B^n_{i-\alpha,j_1,\ldots,j_{\ell_n}}$. For any $0<\varepsilon<1$ and any $t \ge (i-k_n)\Delta_n$, we decompose the above integral into three terms and set

$$\begin{aligned}
\widetilde{N}(\varepsilon)_t &= \int_{(i-k_n)\Delta_n}^{t} \int \mathbb{1}_{\{\gamma(z)>\varepsilon\}} \mu(\mathrm{d}s,\mathrm{d}z), \\
\widetilde{M}(\varepsilon)_t &= \int_{(i-k_n)\Delta_n}^{t} \int \mathbb{1}_{\{\gamma(z)\le\varepsilon\}} \delta(s,z)\mathbb{1}_B(s)\mathbb{1}_{\{\gamma(z)\le 1/q\}}(\mu-\nu)(\mathrm{d}s,\mathrm{d}z), \\
\widetilde{B}(\varepsilon)_t &= -\int_{(i-k_n)\Delta_n}^{t} \int \mathbb{1}_{\{\gamma(z)>\varepsilon}\}\delta(s,z)\mathbb{1}_B(s)\mathbb{1}_{\{\gamma(z)\le 1/q\}}\lambda(\mathrm{d}z)\mathrm{d}s.
\end{aligned}$$

By integrability of γ^2 we have

$$\mathbb{P}(\widetilde{N}(\varepsilon)_{(i+k_n)\Delta_n} \geq 1) \leq \mathbb{E}[\widetilde{N}(\varepsilon)_{(i+k_n)\Delta_n}] = \mathbb{E}\Big[\int_{(i-k_n)\Delta_n}^{(i+k_n)\Delta_n} \int \mathbb{1}_{\{\gamma(z)>\varepsilon\}} \lambda(\mathrm{d}z)\mathrm{d}s\Big] \leq C\frac{k_n\Delta_n}{\varepsilon^2},$$

and similarly we can deduce $|\widetilde{B}(\varepsilon)_{(i+k_n)\Delta_n}| \leq C\frac{\ell_n\Delta_n}{\varepsilon}$. Finally, from Lemma 2.1.5 in Jacod and Protter (2012) we obtain

$$\begin{aligned}
&\mathbb{E}[|\widetilde{M}(\varepsilon)_{(i+k_n)\Delta_n}|^{2p}] \\
&\leq C\Big(\mathbb{E}\Big[\int_{(i-k_n)\Delta_n}^{(i+k_n)\Delta_n} \int \mathbb{1}_{\{\gamma(z)\leq\varepsilon\}} |\delta(s,z)|^{2p} \mathbb{1}_B(s) \mathbb{1}_{\{\gamma(z)\leq 1/q\}} \lambda(\mathrm{d}z)\mathrm{d}s\Big] \\
&\quad + \mathbb{E}\Big[\Big(\int_{(i-k_n)\Delta_n}^{(i+k_n)\Delta_n} \int \mathbb{1}_{\{\gamma(z)\leq\varepsilon\}} |\delta(s,z)|^{2} \mathbb{1}_B(s) \mathbb{1}_{\{\gamma(z)\leq 1/q\}} \lambda(\mathrm{d}z)\mathrm{d}s\Big)^p\Big]\Big) \\
&\leq C\Big(\varepsilon^{2p-2}\ell_n\Delta_n \int \gamma(z)^2\lambda(\mathrm{d}z) + (\ell_n\Delta_n)^p\Big(\int \gamma(z)^2\lambda(\mathrm{d}z)\Big)^p\Big) \leq C(\varepsilon^{2p-2}\ell_n\Delta_n + (\ell_n\Delta_n)^p).
\end{aligned} \tag{30}$$

Then

$$\begin{aligned}
&\mathbb{E}\Big[\Big(\frac{2n}{\ell_n^2}\sum_{\alpha=1}^{k_n} \frac{1}{\binom{k_n}{\ell_n}} \sum_{\substack{1\leq j_1<\ldots<j_{\ell_n}\leq k_n \\ \{j_1,\ldots,j_{\ell_n}\}\cap\{\alpha\}\neq\emptyset}} \Big(\int_{(i-k_n)\Delta_n}^{(i+k_n)\Delta_n} \int \delta(s,z)\mathbb{1}_B(s)\mathbb{1}_{\{\gamma(z)\leq 1/q\}}(\mu-\nu)(\mathrm{d}s,\mathrm{d}z)\Big)^{2p}\Big) \wedge 1\Big] \\
&\leq \mathbb{P}(\widetilde{N}(\varepsilon)_{(i+k_n)\Delta_n} \geq 1) + C_p\frac{n}{\ell_n^2}\sum_{\alpha=1}^{k_n} \frac{1}{\binom{k_n}{\ell_n}} \sum_{\substack{1\leq j_1<\ldots<j_{\ell_n}\leq k_n \\ \{j_1,\ldots,j_{\ell_n}\}\cap\{\alpha\}\neq\emptyset}} \mathbb{E}[|\widetilde{B}(\varepsilon)_{(i+k_n)\Delta_n}|^{2p} + |\widetilde{M}(\varepsilon)_{(i+k_n)\Delta_n}|^{2p}] \\
&\leq C_p\Big(\frac{k_n\Delta_n}{\varepsilon^2} + \big(\ell_n\Delta_n\big)^{2p-1}\varepsilon^{-2p} + \varepsilon^{2p-2} + (\ell_n\Delta_n)^{p-1}\Big),
\end{aligned}$$

where we have used (27). Choosing $\varepsilon_n \to 0$ small enough then ends the proof of (28).

We will finish the proof by showing (13), for which we use the following Taylor expansion for $f(x) = |x|^p$: on $\Omega(n,q)$ we have

$$\begin{aligned}
\Big|\sum_{m=1}^{\ell_n} \Delta^n_{i_r-\alpha+j_m} X\Big|^p - |\Delta^n_{i_r}X|^p &= f'(\Delta X_{S_r}) \sum_{\substack{m=1 \\ j_m\neq\alpha}}^{\ell_n} \Delta^n_{i_r-\alpha+j_m} X(q) \\
&\quad + f''(\kappa^n_{i_r})\Delta^n_{i_r}X(q) \sum_{\substack{m=1 \\ j_m\neq\alpha}}^{\ell_n} \Delta^n_{i_r-\alpha+j_m} X(q) + \frac{1}{2}f''(\xi^n_{i_r-\alpha,j_1,\ldots,j_{\ell_n}})\Big|\sum_{\substack{m=1 \\ j_m\neq\alpha}}^{\ell_n} \Delta^n_{i_r-\alpha+j_m} X(q)\Big|^2
\end{aligned}$$

for some intermediate $\kappa^n_{i_r}$ between $\Delta^n_{i_r}X$ and ΔX_{S_r} and $\xi^n_{i_r-\alpha,j_1,\ldots,j_{\ell_n}}$ between $\sum_{m=1}^{\ell_n} \Delta^n_{i_r-\alpha+j_m} X$ and $\Delta^n_{i_r}X$. On $\Omega(n,q)$ both are bounded by C_q. Obviously, one can show

$$\frac{2n}{\ell_n^2}\sum_{r\in P_q}\sum_{\alpha=1}^{k_n}\frac{1}{\binom{k_n}{\ell_n}}\sum_{\substack{1\le j_1<\ldots<j_{\ell_n}\le k_n\\ \{j_1,\ldots,j_{\ell_n}\}\cap\{\alpha\}\neq\emptyset}}\Big|\sum_{\substack{m=1\\ j_m\neq\alpha}}^{\ell_n}\Delta^n_{i_r-\alpha+j_m}X(q)\Big|^4\xrightarrow{\mathbb{P}}0$$

as $n\to\infty$ for any fixed q along the same lines as the ones from the proof of (25) with $p=2$, and similarly

$$\frac{2n}{\ell_n^2}\sum_{r\in P_q}\sum_{\alpha=1}^{k_n}\frac{1}{\binom{k_n}{\ell_n}}\sum_{\substack{1\le j_1<\ldots<j_{\ell_n}\le k_n\\ \{j_1,\ldots,j_{\ell_n}\}\cap\{\alpha\}\neq\emptyset}}|\Delta^n_{i_r}X(q)|^2\Big|\sum_{\substack{m=1\\ j_m\neq\alpha}}^{\ell_n}\Delta^n_{i_r-\alpha+j_m}X(q)\Big|^2\xrightarrow{\mathbb{P}}0.$$

Lemma 1 then suggests that we only need to prove

$$\frac{n}{\ell_n(\ell_n-1)}\sum_{r\in P_q}\sum_{\alpha=1}^{k_n}\widehat{Y}^{(n)}_{r,\alpha}\xrightarrow{\mathbb{P}}V(q)$$

where

$$\widehat{Y}^{(n)}_{r,\alpha}=\frac{1}{\binom{k_n}{\ell_n}}\sum_{\substack{1\le j_1<\ldots<j_{\ell_n}\le k_n\\ \{j_1,\ldots,j_{\ell_n}\}\cap\{\alpha\}\neq\emptyset}}f'(\Delta X_{S_r})^2\Big|\sum_{\substack{m=1\\ j_m\neq\alpha}}^{\ell_n}\Delta^n_{i_r-\alpha+j_m}X(q)\Big|^2.$$

The penultimate step is yet another application of Lemma 1, namely to first prove

$$\frac{2n}{\ell_n^2}\sum_{r\in P_q}\sum_{\alpha=1}^{k_n}\frac{1}{\binom{k_n}{\ell_n}}\sum_{\substack{1\le j_1<\ldots<j_{\ell_n}\le k_n\\ \{j_1,\ldots,j_{\ell_n}\}\cap\{\alpha\}\neq\emptyset}}\Big(\sum_{\substack{m=1\\ j_m\neq\alpha}}^{\ell_n}(\Delta^n_{i_r-\alpha+j_m}X(q)-\sigma_{S_r}\Delta^n_{i_r-\alpha+j_m}W)\Big)^2\xrightarrow{\mathbb{P}}0$$

as $n\to\infty$ for any fixed q and to use boundedness of the jumps of $N(q)$ by some C_q. This proof also works in the same way as (25) with $p=1$, but with two differences: first, instead of (29) we discuss

$$\mathbb{E}\Big[\Big|\int_{(i_r-k_n)\Delta_n}^{(i_r+k_n)\Delta_n}\mathbb{1}_{B^n_{i-\alpha,j_1,\ldots,j_{\ell_n}}}(s)(\sigma_s-\sigma_{S_r})\mathrm{d}W_s\Big|^2\Big]\le\ell_n\Delta_n\mathbb{E}[\sup_{|u|\le k_n\Delta_n}|\sigma_{S_r-u}-\sigma_{S_r}|^2],$$

and we apply additionally continuity of σ in S_r plus dominated convergence, and second the upper bound in (30) now becomes $\ell_n\Delta_n\int\mathbb{1}_{\{\gamma(z)\le\varepsilon\}}\gamma(z)^2\lambda(\mathrm{d}z)$. Therefore, from Lemma 1 it is sufficient to prove convergence in probability of

$$\frac{n}{\ell_n(\ell_n-1)}\sum_{r\in P_q}(f'(\Delta X_{S_r}))^2\sigma^2_{S_r}\sum_{\alpha=1}^{k_n}\frac{1}{\binom{k_n}{\ell_n}}\sum_{\substack{1\le j_1<\ldots<j_{\ell_n}\le k_n\\ \{j_1,\ldots,j_{\ell_n}\}\cap\{\alpha\}\neq\emptyset}}\Big|\sum_{\substack{m=1\\ j_m\neq\alpha}}^{\ell_n}\Delta^n_{i_r-\alpha+j_m}W\Big|^2 \tag{31}$$

to $V(q)$ as $n \to \infty$. Using $f'(x) = px^{p-1}$ and (11) we are left to show $\frac{1}{k_n}\sum_{\alpha=1}^{k_n} Z_{i,\alpha}^n \xrightarrow{\mathbb{P}} 0$ for any fixed i, where

$$Z_{i,\alpha}^n = \frac{1}{\binom{k_n-1}{\ell_n-1}} \sum_{\substack{1 \le j_1 < \ldots < j_{\ell_n-1} \le k_n \\ \{j_1,\ldots,j_{\ell_n-1}\} \cap \{\alpha\} = \emptyset}} \left(\frac{n}{\ell_n - 1} \Big| \sum_{m=1}^{\ell_n-1} \Delta_{i-\alpha+j_m}^n W \Big|^2 - 1 \right).$$

Note that we can again drop the dependence on r by properties of a Poisson random measure. Using $\mathbb{E}[Z_{i_r,\alpha}^n] = 0$ and

$$\operatorname{Var}\left(\frac{1}{k_n}\sum_{\alpha=1}^{k_n} Z_{i,\alpha}^n\right) = \frac{1}{k_n^2}\sum_{\alpha_1,\alpha_2=1}^{k_n} \operatorname{Cov}(Z_{i,\alpha_1}^n, Z_{i,\alpha_2}^n) \le \left(\frac{1}{k_n}\sum_{\alpha=1}^{k_n}\sqrt{\operatorname{Var}(Z_{i,\alpha}^n)}\right)^2$$

we are left to show $\operatorname{Var}(Z_{i,\alpha}^n) \le \eta_n \to 0$. In distribution, $Z_{i,\alpha}^n$ equals the U statistic

$$U_n = \frac{1}{\binom{k_n-1}{\ell_n-1}} \sum_{1 \le j_1 < \ldots < j_{\ell_n-1} \le k_n - 1} \left(\Big| \sum_{m=1}^{\ell_n-1} N_{j_m}^n \Big|^2 - 1 \right)$$

for i.i.d. standard normal N_i. Using Theorem 1.2.3 in Denker (1985) again we obtain

$$\operatorname{Var}(U_n) \le C\frac{\ell_n - 1}{k_n - 1} \to 0$$

which finishes the proof for V_n.

We will finally discuss the necessary changes for $\widehat{V}_n$ and $\widetilde{V}_n$, and this time the entire proof goes through in exactly the same way when $\ell_n = 2$. For $\ell_n = k_n$ the proof of (16) goes through without any changes, whereas for (23) we cannot apply (13) because we do not have convergence in probability in the end. Nevertheless, we only use (13) in an application of the Portmanteau theorem, and this goes through under weak convergence as well. So we only need to discuss the stable convergence of (13), as (14) finally follows from monotone convergence again.

The proof of (13) can always be reproduced until one arrives at (31) which, because of $k_n \to \infty$, becomes

$$\begin{aligned}
&\sum_{r \in P_q} (f'(\Delta X_{S_r}))^2 \sigma_{S_r}^2 \frac{1}{k_n} \sum_{\alpha=1}^{k_n} \frac{n}{k_n} \Big| \sum_{m=1}^{k_n} \Delta_{i_r-\alpha+m}^n W \Big|^2 \big(1 + o_{\mathbb{P}}(1)\big) \\
&\quad = \sum_{r \in P_q} (f'(\Delta X_{S_r}))^2 \sigma_{S_r}^2 w_{n,r} \big(1 + o_{\mathbb{P}}(1)\big)
\end{aligned}$$

with

$$w_{n,r} = \frac{n}{k_n^2} \sum_{j=0}^{k_n-1} (W_{(i_r+k_n-j)\Delta_n} - W_{(i_r-j)\Delta_n})^2.$$

The final step therefore is to prove the stable convergence

$$(w_{n,r})_{r\geq 1} \xrightarrow{\mathcal{L}-(s)} (1+R_r)_{r\geq 1},$$

which follows as in the proof of Theorem 4.3.1 in Jacod and Protter (2012) and can be traced back to convergence in distribution of each fixed $w_{n,r}$ to $1+R_r$. A simple computation gives

$$w_{n,r} = \frac{n}{k_n^2} \sum_{u,v=1}^{2k_n-1} \big((2k_n - u \vee v) \wedge k_n - (k_n - u \wedge v) \vee 0\big) \Delta^n_{i_r+u-(k_n-1)} W \Delta^n_{i_r+v-(k_n-1)} W$$

which equals

$$\widetilde{w}_{n,r} = \sum_{u,v=1}^{2k_n-1} f(u/k_n, v/k_n)(B_{u/k_n} - B_{(u-1)/k_n})(B_{v/k_n} - B_{(v-1)/k_n})$$

in distribution, where $f(s,t) = (2 - s \vee t) \wedge 1 - (1 - s \wedge t) \vee 0$ and B is a standard Brownian motion. Clearly,

$$\sum_{u=1}^{2k_n-1} f(u/k_n, u/k_n)(B_{u/k_n} - B_{(u-1)/k_n})^2 \xrightarrow{\mathbb{P}} \int_0^2 f(s,s)\mathrm{d}s = 1,$$

and the proof is finally finished because

$$\sum_{u,v=1}^{2k_n-1} \mathbb{1}_{\{u\neq v\}} f(u/k_n, v/k_n)(B_{u/k_n} - B_{(u-1)/k_n})(B_{v/k_n} - B_{(v-1)/k_n})$$

converges in probability to $\int_{[0,2]^2} f(s,t)\mathrm{d}B_s \mathrm{d}B_t$ by definition of a multiple Wiener-Itô integral. □

Acknowledgements The author wants to thank two anonymous referees for their very valuable comments which have led to a substantial improvement of this work. It was particularly helpful to be provided with a simple representation for the distribution of R_r in Theorem 3 (b) and (c).

References

Aït-Sahalia, Y., Jacod, J. (2014). *High-frequency finanicial econometrics*. Princeton: Princeton University Press.

Barndorff-Nielsen, O. E., Shephard, N. (2003). Realized power variation and stochastic volatility models. *Bernoulli*, *9*(2), 243–265.

Barndorff-Nielsen, O. E., Graversen, S. E., Jacod, J., Podolskij, M., Shephard, N. (2006). A central limit theorem for realised power and bipower variations of continuous semimartingales. *From stochastic calculus to mathematical finance*, pp. 33–68. Berlin: Springer.

Christensen, K., Podolskij, M., Thamrongrat, N., Veliyev, B. (2017). Inference from high-frequency data: A subsampling approach. *Journal of Econometrics*, *197*(2), 245–272.

Denker, M. (1985). *Asymptotic distribution theory in nonparametric statistics. Advanced lectures in mathematics*. Braunschweig: Friedr. Vieweg & Sohn.

Jacod, J. (2008). Asymptotic properties of realized power variations and related functionals of semimartingales. *Stochastic Processes and their Applications*, *118*(4), 517–559.

Jacod, J., Protter, P. (2012). *Discretization of processes*. Berlin: Springer.

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

Mykland, P. A., Zhang, L. (2017). Assessment of uncertainty in high frequency data: The observed asymptotic variance. *Econometrica*, *85*(1), 197–231.

Veraart, A. E. D. (2010). Inference for the jump part of quadratic variation of Itô semimartingales. *Econometric Theory*, *26*(2), 331–368.