Inference on the Lévy measure in case of noisy observations

Mathias Vetter[*]

April 16, 2013

Abstract

In this paper we are concerned with inference on the Lévy measure ν of a Lévy process in case of noisy high frequency observations. It is known that standard techniques for denoising, developed for diffusion settings, do not work in this case. For this reason, we provide an extension of the pre-averaging method which allows for a consistent estimation of the Lévy distribution function even under microstructure noise. Interestingly, the asymptotic behaviour of the novel estimator is the same as in the no-noise case. This is in sharp contrast to what is known for diffusions.

Keywords: Lévy process, microstructure noise, nonparametric statistics, weak convergence.

AMS 2000 subject classifications: primary, 60F05, 60G51; secondary, 62M09.

1 Introduction

Suppose one is interested in statistical inference on the Lévy measure ν of a univariate Lévy process X. Following the so-called Lévy-Itô decomposition, X can be written as

$$X_t = at + \sigma B_t + \int_0^t \int_{|u| \leq 1} u(\mu - \bar{\mu})(ds, du) + \int_0^t \int_{|u| > 1} u\mu(ds, du), \qquad (1.1)$$

where a is a scalar, the diffusion part consists of a standard Brownian motion B and some volatility $\sigma^2 > 0$, and $\mu(ds, du)$ denotes a Poisson random measure with compensator $\bar{\mu}(ds, du)$. It is well-known that the compensator takes the form $\mu(ds, du) = ds\,\nu(du)$ for some σ-finite measure ν, the latter being the Lévy measure of X which governs its jump behaviour. The triplet (a, σ^2, ν) completely determines the distribution of X.

Just as regular probability measures are uniquely determined by their distribution functions and vice versa, there is a one-to-one correspondence between the Lévy measure

ν and its Lévy distribution function $U : \overline{\mathbb{R}} \backslash \{0\} \to [0, \infty)$ defined by

$$U(t) = \begin{cases} \nu([t, \infty]), & t > 0, \\ \nu([-\infty, t]), & t < 0. \end{cases}$$

Bücher and Vetter (2013) used high-frequency data to infer on U, i.e. their estimator is built from observations $X_{i\Delta_n}$, $i = 1, \ldots, n$, where $\Delta_n \to 0$ and the covered time horizon satisfies $k_n = n\Delta_n \to \infty$. In their work, the Lévy measure was for brevity of notation assumed to be restricted to the positive half line only, in which case $U(t) = 0$ identically for $t < 0$. We will adopt their setting in this work.

Our aim is to provide similar results in the case of observations contaminated by microstructure noise. These effects are known to be a severe problem particularly in econometric practice, as financial data at a high frequency level suffers from observation errors due to discreteness of prices, bid-ask spreads or asymmetric information in the market. Suppose therefore that one does not observe the plain Lévy process, but

$$Z_{i\Delta_n} = X_{i\Delta_n} + V_{i\Delta_n},$$

where $V_{i\Delta_n}$, $i = 1, \ldots, n$, is an i.i.d. process, independent of X, which satisfies $\mathbb{E}[V_{i\Delta_n}] = 0$ and has moments of all order. The latter assumption is not too strict, as the noise is typically assumed to be responsible for a rather small deviation from the true process and might thus for example be modelled by a random variable with bounded support.

In this setting, the natural estimator proposed by Bücher and Vetter (2013) does not work anymore. Setting $\Delta_i^n Y = Y_{i\Delta_n} - Y_{(i-1)\Delta_n}$ for an arbitrary process Y, they use

$$U_n^X(y) = \frac{1}{k_n} \sum_{i=1}^{n} 1_{\{\Delta_i^n X \geq y\}}$$

as an estimator for $U(y)$, which counts large increments of X as they are likely to be caused by a jump over the corresponding small interval. Standardising with k_n instead of n is justified, since large jumps occur rarely, that is $P(\Delta_i^n X \geq y) \sim \Delta_n U(y)$ under mild assumptions on ν; see e.g. Corollary 8.9 in Sato (1999).

Suppose now that one observes Z instead. If there is a positive probability of $P(\Delta_i^n V \geq y)$, then $U_n^Z(y)$ (which is the same statistic as the previous one, but where increments of X are replaced by those involving Z) will not only count the few large jumps of X, but more often large increments of Z which are due to microstructure noise. Precisely, the statistic behaves like $\Delta_n^{-1} P(\Delta_i^n V \geq y)$ which diverges to infinity; recall $k_n = n\Delta_n$. Note also that if $P(\Delta_i^n V \geq y) = 0$, then $U_n^Z(y)$ is still bounded in probability, but we will rather estimate a convolution of jumps and noise than the plain jump measure.

Throughout the last decade an increasing amount of research has been concerned with microstructure noise, mostly in connection with diffusion processes. (An exception is the work by Comte and Genon-Catalot (2010), where inference for pure jump Lévy processes based on their characteristic exponent has been conducted.) Since it is typically the jump part of the underlying process which is responsible for heavy tails in standard

models in econometrics, there is a clear need for reliable tools in order to assess the Lévy measure. However, none of the standard methods for diffusion processes (e.g. the multiscale approach by Zhang et al. (2005) or the kernel-based from Barndorff-Nielsen et al. (2008)) can directly be applied in our context. Even the pre-averaging approach by Jacod et al. (2009), which provides a general concept for diminishing the influence of the noise by retaining information of the increments of X, fails when one is interested in the Lévy measure. See Section 5.4 in Bücher and Vetter (2013) for details.

Our aim in this work is therefore to provide an alternative concept which allows for a consistent estimation of the Lévy distribution function. It will be based on the intuition of pre-averaging, which is why we will shortly recall the approach from Jacod et al. (2009) in Section 2 and construct a modified version which turns out to work fine for noisy observations as well. The asymptotic theory for the novel estimator can be found in Section 3, alongside with a discussion of the result.

2   A modified pre-averaging estimator

We begin this section with some reminders on the pre-averaging approach: For an auxiliary sequence $l_n \to \infty$ with $n/l_n \to \infty$ and some piecewise differentiable function g on $[0, 1]$, which is positive on $(0, 1)$ and satisfies $g(0) = g(1) = 0$, one computes $\widetilde{Z_i^n} = \widetilde{X_i^n} + \widetilde{V_i^n}$, where

$$\widetilde{Y_i^n} = \sum_{j=1}^{l_n} g(j/l_n) \Delta_{i+j}^n Y$$

for an arbitrary process Y, $i = 1, \ldots, n - l_n$. The idea behind this approach is that $\widetilde{X_i^n}$ can be seen as some kind of generalised increment which still bears similar information as the plain $\Delta_i^n X$, while by assumption on g

$$\widetilde{V_i^n} = -g(1/l_n) V_{i\Delta_n} + \sum_{j=1}^{l_n-1} (g(j/l_n) - g((j+1)/l_n)) V_{(i+j)\Delta_n} = O_p\Big(l_n^{-1/2}\Big), \qquad (2.1)$$

which proves that the larger l_n becomes, the less important is the contamination by noise.

For this reason, estimation based on pre-averaging usually works in the way that one proceeds as usual, but replaces the standard estimators by ones based on $\widetilde{Z_i^n}$. If l_n is rather large compared with n, it is reasonable to replace $\widetilde{Z_i^n}$ with $\widetilde{X_i^n}$ in the asymptotics and thus to recover full information of X. In our setting this would lead to an estimator of the form

$$\widetilde{U_n^Z}(y) = \frac{1}{k_n} \sum_{i=0}^{\lfloor n/l_n \rfloor - 1} 1_{\{\widetilde{Z_{il_n}^n} \geq y\}},$$

where the $\widetilde{Z_{il_n}^n}$ are computed over non-overlapping intervals to retain i.i.d. terms in the sum. Intuitively, for large l_n it is equally well possible to discuss $\widetilde{U_n^X}(y)$ which, however, does not result in a consistent estimator for the Lévy distribution function $U(y)$. The reason is a rather technical one, connected with the fact that the main contribution to

$\widetilde{X_{il_n}^n}$ is coming from a single large jump within $[0, l_n\Delta_n]$ whose exact position influences standardisation with g. An argument similar to the proof of Lemma 3.2 below gives

$$P(\widetilde{X_{il_n}^n} \geq y) = \Delta_n \sum_{j=1}^{l_n} U(y/g(j/l_n)) + O(l_n^2\Delta_n^2) = l_n\Delta_n \int_0^1 U(y/g(x))dx + o(l_n\Delta_n),$$

which proves that $\widetilde{U_n^X}(y)$ estimates a functional of U depending on g and y rather than the plain $U(y)$. Even though one has freedom to choose the auxiliary function g above, it turns out that the choices which lead to an unbiased estimator for $U(y)$ contradict the necessary assumptions on g which secure that the noise terms become small in the asymptotics. For this reason, the standard pre-averaging approach cannot be applied. A modification, however, gives the desired result.

Define two auxiliary sequences m_n and l_n which satisfy $m_n = c\Delta_n^{-a} + o(\Delta_n^{-a})$ and $l_n = d\Delta_n^{-b} + o(\Delta_n^{-b})$ for some $c, d > 0$ and $0 < a < b < 1$. Furthermore, we introduce a family of weight functions depending on n as follows: Let

$$g_n(x) = \begin{cases} \frac{l_n}{m_n}x, & x \leq \frac{m_n}{l_n}, \\ 1, & \frac{m_n}{l_n} \leq x \leq 1 - \frac{m_n}{l_n}, \\ \frac{l_n}{m_n}(1-x), & 1 - \frac{m_n}{l_n} \leq x \leq 1, \end{cases}$$

which is well defined for $l_n \geq 2m_n$. The latter assumption is eventually satisfied due to $b > a$. For any process Y we then set

$$\overline{Y_i^n} = \sum_{j=1}^{l_n} g_n(j/l_n)\Delta_{i+j}^n Y.$$

Note first that the influence of the noise becomes small as m_n grows, as can be seen from an argument similar to the one leading to (2.1). Also, since l_n is much larger than m_n, a jump now typically occurs in the middle of the interval where no standardisation due to g as for the standard pre-averaging estimator takes place. Finally, an estimator over non-overlapping blocks is given by

$$\overline{U_n^Z}(y) = \frac{1}{k_n} \sum_{i=0}^{\lfloor n/l_n \rfloor - 1} 1_{\{\overline{Z_{il_n}^n} \geq y\}}. \tag{2.2}$$

3   The asymptotic theory

In order to derive asymptotics for $\overline{U_n^Z}(y)$, a smoothness condition on the Lévy measure is necessary. It secures that probabilities such as $P(\Delta_i^n X \geq y)$ are close to $\Delta_n U(y)$, and it also governs the order of the approximation error.

Assumption 3.1 *Let X be a Lévy process with the representation (1.1). We assume that ν has support on $(0, \infty)$, on which it is of the form $\nu(du) = s(u)du$ for a positive Lévy density s which satisfies $\sup_{u > \eta}(|s(u)| + |s'(u)|) < \infty$ for any $\eta \in (0, \infty)$.*

We are now interested in $P(\overline{X^n_{il_n}} \geq y)$, and due to stationarity and independence we discuss $i = 0$ only. The following claim contains an expansion of this probability to first order, where the approximation error comes from two sources: One is naturally due to the fact that a convolution of two or more (or no) large jumps may be responsible for a large value of $\overline{X^n_0}$ and appears in the no-noise case as well. The other one comes from effects on the boundary of the interval $[0, l_n\Delta_n]$ involving the auxiliary sequence m_n.

Lemma 3.2 *Suppose that Assumption 3.1 holds and let $\delta > 0$ be fixed. Then there exist a constant $K = K(\delta)$ such that for all $y \geq \delta$*

$$\left| P(\overline{X^n_0} \geq y) - l_n\Delta_n U(y) \right| \leq K_\delta l_n\Delta_n(m_n/l_n + l_n\Delta_n).$$

Proof. Before we begin with the proof, note that there are several different representations of $\overline{Y^n_0}$ which will all be used throughout the paper, namely

$$\overline{Y^n_0} = (Y_{(l_n-m_n)\Delta_n} - Y_{m_n\Delta_n}) + \sum_{j=1}^{m_n} \frac{j}{m_n}\Delta^n_j Y - \sum_{j=1}^{m_n-1} \frac{j}{m_n}\Delta^n_{l_n-j}Y \qquad (3.1)$$

$$= \frac{1}{m_n} \sum_{j=0}^{m_n-1} (Y_{(l_n-m_n+j)\Delta_n} - Y_{j\Delta_n}) = \frac{1}{m_n} \sum_{j=0}^{m_n-1} Y_{(l_n-m_n+j)\Delta_n} - \frac{1}{m_n} \sum_{j=0}^{m_n-1} Y_{j\Delta_n}.$$

We follow the ideas from Figueroa-López and Houdré (2009) who prove a similar result for plain increments of X. Let $\varepsilon < (\delta/r \wedge 1)$ for some $r > 0$ to be specified later and pick a smooth function $c_\varepsilon : \mathbb{R} \to \mathbb{R}$ satisfying

$$1_{[-\varepsilon/2,\varepsilon/2]}(u) \leq c_\varepsilon(u) \leq 1_{[-\varepsilon,\varepsilon]}(u).$$

We also define the function $\bar{c}_\varepsilon$ via $\bar{c}_\varepsilon(u) = 1 - c_\varepsilon(u)$. It is straightforward to see that there exist independent processes X^ε and $\widetilde{X}^\varepsilon$ such that $X \sim X^\varepsilon + \widetilde{X}^\varepsilon$ and where $\widetilde{X}^\varepsilon$ is a compound Poisson process with intensity $\lambda_\varepsilon = \int \bar{c}_\varepsilon(u)\nu(du)$ and jump distribution $f_\varepsilon(du) = \bar{c}_\varepsilon(u)\nu(du)/\lambda_\varepsilon$ and X^ε is a Lévy process with triplet $(a_\varepsilon, \sigma^2, c_\varepsilon(u)\nu(du))$, where we set $a_\varepsilon = a - \int 1_{\{|u|\leq 1\}} u\bar{c}_\varepsilon(u)\nu(du)$. Let us work with this particular decomposition of X in the following.

Call $N^\varepsilon_{l_n\Delta_n}$ the number of jumps of $\widetilde{X}^\varepsilon$ up to time $l_n\Delta_n$ and set $f(u) = 1_{\{u\geq y\}}$. Using the law of total expectation we then have

$$\mathbb{E}[f(\overline{X^n_0})] = \sum_{k=0}^{\infty} e^{-\lambda_\varepsilon l_n\Delta_n} \frac{(\lambda_\varepsilon l_n\Delta_n)^k}{k!} \mathbb{E}[f(\overline{X^n_0})|N^\varepsilon_{l_n\Delta_n} = k]. \qquad (3.2)$$

Let us start with the case for $k = 0$. In this situation we are interested in $P(\overline{X^{\varepsilon,n}_0} \geq y)$, which can be bounded by

$$P\Big(\frac{1}{m_n} \sum_{j=0}^{m_n-1} (X^\varepsilon_{(l_n-m_n+j)\Delta_n} - X^\varepsilon_{j\Delta_n}) \geq y\Big) \leq \sum_{j=0}^{m_n-1} P\Big((X^\varepsilon_{(l_n-m_n+j)\Delta_n} - X^\varepsilon_{j\Delta_n}) \geq y\Big)$$

using the second identity in (3.1). Equation (3.3) in Figueroa-López and Houdré (2009) proves that the probability on the right hand side above can be bounded by $K_\delta(l_n\Delta_n)^r$, such that

$$P(\overline{X_0^{\varepsilon,n}} \geq y) \leq K_\delta m_n(l_n\Delta_n)^r \leq K_\delta l_n\Delta_n(m_n/l_n + l_n\Delta_n)$$

by choosing r large enough due to $0 < a < b < 1$. The same upper bound holds for $k > 2$,

$$\sum_{k=2}^{\infty} e^{-\lambda_\varepsilon l_n\Delta_n} \frac{(\lambda_\varepsilon l_n\Delta_n)^k}{k!} \mathbb{E}[f(\overline{X_0^n})|N_{l_n\Delta_n}^\varepsilon = k] \leq \sum_{k=2}^{\infty} e^{-\lambda_\varepsilon l_n\Delta_n} \frac{(\lambda_\varepsilon l_n\Delta_n)^k}{k!} \leq K_\delta(l_n\Delta_n)^2.$$

Thus, let us finally discuss $k = 1$ which corresponds to one single large jump within $[0, l_n\Delta_n]$. Call s the time and $\xi \sim f_\varepsilon$ the size of the jump. Then $\mathbb{E}[f(\overline{X_0^n})|N_{l_n\Delta_n}^\varepsilon = 1]$ can be written as

$$\mathbb{E}[f(\overline{X_0^n})|N_{l_n\Delta_n}^\varepsilon = 1, s \in [0, m_n\Delta_n)]P(s \in [0, m_n\Delta_n)|N_{l_n\Delta_n}^\varepsilon = 1)$$
$$+\mathbb{E}[f(\overline{X_0^n})|N_{l_n\Delta_n}^\varepsilon = 1, s \in ((l_n - m_n)\Delta_n, l_n\Delta_n]]P(s \in ((l_n - m_n)\Delta_n, l_n\Delta_n]|N_{l_n\Delta_n}^\varepsilon = 1)$$
$$+\mathbb{E}[f(\overline{X_0^n})|N_{l_n\Delta_n}^\varepsilon = 1, s \in [m_n\Delta_n, (l_n - m_n)\Delta_n]]P(s \in [m_n\Delta_n, (l_n - m_n)\Delta_n]|N_{l_n\Delta_n}^\varepsilon = 1).$$

Since the jump time is uniformly distributed over $[0, l_n\Delta_n]$, the first two terms can be bounded by m_n/l_n each. In the same way, the final term becomes

$$\mathbb{E}[f(\overline{X_0^n})|N_{l_n\Delta_n}^\varepsilon = 1, s \in [m_n\Delta_n, (l_n - m_n)\Delta_n]]\left(1 - 2\frac{m_n}{l_n}\right) = P(\overline{X_0^{\varepsilon,n}} + \xi \geq y)\left(1 - 2\frac{m_n}{l_n}\right),$$

where ξ is not affected by the weight function g_n, since the jump is assumed to occur away from the boundary of $[0, l_n\Delta_n]$. See the first identity in (3.1). By independence of X^ε and ξ one may set $h(u) = P(\xi \geq y - u)$ and compute $E[h(\overline{X_0^{\varepsilon,n}})]$ then. The same arguments as in Figueroa-López and Houdré (2009) show that

$$|E[h(\overline{X_0^{\varepsilon,n}})] - h(0)| \leq K_\delta l_n\Delta_n$$

with $h(0) = P(\xi \geq y) = U(y)/\lambda_\varepsilon$ by our choice of ε. Putting all bounds together and using (3.2) plus an expansion of $\exp(-\lambda_\varepsilon l_n\Delta_n) - 1$ the claim follows. □

We have now obtained an expression for the probability of $\overline{X_0^n}$ exceeding a deterministic bound, but in reality we observe $\overline{Z_0^n}$ instead. Therefore we need a result which proves that the error due to replacing the observable statistic based on Z by the same one based on the latent X is small compared to the rate of convergence. This is what we aim at in the the following lemma.

Lemma 3.3 *Suppose that Assumption 3.1 holds and that m_n and l_n are chosen in such a way that*

$$k_n = O(m_n^\gamma) \text{ for some } 0 < \gamma < 1 \quad and \quad \sqrt{k_n}(l_n\Delta_n + m_n/l_n) \to 0 \qquad (3.3)$$

hold. Then we have

$$\frac{1}{\sqrt{k_n}} \sum_{i=0}^{\lfloor n/l_n \rfloor - 1} \left(1_{\{\overline{Z_{il_n}^n} \geq y\}} - 1_{\{\overline{X_{il_n}^n} \geq y\}}\right) = o_p(1),$$

uniformly on $[\delta, \infty)$.

Proof. The summands above are identically distributed, which is why we begin with a discussion of the first one. We use the general relation

$$1_{\{p+q\geq y\}} - 1_{\{p\geq y\}} = 1_{\{q\geq y\}} + 1_{\{p<y,q<y,p+q\geq y\}} - 1_{\{p\geq y,p+q<y\}} - 1_{\{q\geq y,p+q<y\}} - 1_{\{p\geq y,q\geq y\}},$$

where the role of p and q is played by $\overline{X_0^n}$ and $\overline{V_0^n}$, respectively. Observe that

$$|1_{\{p+q\geq y\}} - 1_{\{p\geq y\}}| \leq 3 \times 1_{\{q\geq y\}} + 1_{\{p<y,q<y,p+q\geq y\}} + 1_{\{p\geq y,p+q<y\}}$$

$$\leq 5 \times 1_{\{|q|\geq d_n\}} + 2 \times 1_{\{y-d_n\leq p\leq y+d_n\}}$$

for any sequence $d_n < \delta/2$. For later reasons we set $d_n = m_n^{-\varpi}$ with some $\varpi < 1/2$ such that $2\varpi > \gamma$ holds. Therefore we have to show that $B_{n1} + \sup_{y\geq\delta} B_{n2}(y) = o_p(1)$, where

$$B_{n1} = \frac{1}{\sqrt{k_n}} \sum_{i=0}^{\lfloor n/l_n\rfloor-1} 1_{\{|\overline{V_0^n}|\geq d_n\}} \quad \text{and} \quad B_{n2}(y) = \frac{1}{\sqrt{k_n}} \sum_{i=0}^{\lfloor n/l_n\rfloor-1} 1_{\{y-d_n\leq \overline{X_0^n}\leq y+d_n\}}.$$

From the third identity in (3.1) we see that

$$\mathbb{E}[|\overline{V_0^n}|^{2r}] = \mathbb{E}\left[\left|\frac{1}{m_n}\sum_{j=0}^{m_n-1} V_{l_n-m_n+j} - \frac{1}{m_n}\sum_{j=0}^{m_n-1} V_j\right|^{2r}\right] \leq C_r m_n^{-r} \tag{3.4}$$

for any integer $r > 0$ which proves using Markov inequality that $P(|\overline{V_0^n}| \geq d_n)$ can be made arbitrarily small, due to $d_n = m_n^{-\varpi}$ with $\varpi < 1/2$. $B_{n1} = o_p(1)$ then follows. Thus, let us discuss the claim involving $B_{n2}(y)$ and fix $y \geq \delta$ first. From Lemma 3.2 we have

$$\left|P(y - d_n \leq \overline{X_0^n} \leq y + d_n) - l_n\Delta_n\nu([y-d_n, y+d_n])\right| \leq K_\delta l_n\Delta_n(m_n/l_n + l_n\Delta_n).$$

Since ν has a bounded Lévy density on $[\delta/2, \infty)$, we obtain

$$\frac{1}{\sqrt{k_n}} \sum_{i=0}^{\lfloor n/l_n\rfloor-1} P(y - d_n \leq \overline{X_{il_n}^n} \leq y + d_n) \leq K_\delta \frac{n}{l_n\sqrt{k_n}} l_n\Delta_n(d_n + m_n/l_n + l_n\Delta_n).$$

The latter two terms converge to zero due to (3.3). For the remainder, note that we have chosen ϖ such that $2\varpi > \gamma$. We obtain

$$K_\delta \frac{n}{l_n\sqrt{k_n}} l_n\Delta_n d_n = K_\delta\sqrt{k_n}d_n \leq K_\delta m_n^{\gamma/2-\varpi} \to 0,$$

which proves the claim for each fixed y, i.e. convergence of the finite dimensional distributions of $B_{n2}(y)$ to those of the zero process. Therefore we are left to establish asymptotic tightness which can be shown using the conditions in Theorem 11.16 of Kosorok (2008). Here, we use the fact that indicator functions over sets $(a, b), -\infty \leq a < b \leq \infty$ shatter no set of size 3 to obtain manageability due to a finite VC index. $\square$

We now come to the result on weak convergence of the process

$$\overline{\alpha}_n(y) = \sqrt{k_n}\left(\overline{U_n^Z}(y) - U(y)\right)$$

in the space $\mathcal{B}_\infty(0, \infty]$ of all real functions on $(0, \infty]$ that are bounded on sets which are bounded away from the origin. We equip this space with the metric $d(f, g) = \sum_{k=1}^\infty 2^{-k}(\|f - g\|_{T_k} \wedge 1)$, where $T_k = [1/k, \infty]$ and $\|\cdot\|_{T_k}$ denotes the uniform norm.

Theorem 3.4 *Assume that X is a Lévy process satisfying Assumption 3.1. If (3.3) holds, then we have $\overline{\alpha}_n(y) \xrightarrow{w} \mathbb{B}(y)$ in $(\mathcal{B}_\infty(0,\infty], d)$. Here, $\mathbb{B}$ is a tight, centered Gaussian process with covariance $\mathbb{E}\left[\mathbb{B}(x)\mathbb{B}(y)\right] = U(x \vee y)$. The sample paths of $\mathbb{B}$ are uniformly continuous on each T_k with respect to the pseudo distance*

$$\rho(x,y) = \mathbb{E}\left[(\mathbb{B}(x) - \mathbb{B}(y))^2\right]^{1/2} = |U(x) - U(y)|^{1/2}.$$

Proof. Note that, due to Theorem 1.6.1 in van der Vaart and Wellner (1996), weak convergence in $(\mathcal{B}_\infty(0,\infty], d)$ is equivalent to weak convergence on each $\ell^\infty(T_k)$. Therefore, it is possible to fix one such T_k throughout the rest of the proof. We set $\widetilde{\alpha}_n(y) = \sqrt{k_n}\left(U_n^X(y) - \mathbb{E}[U_n^X(y)]\right)$. Based on Lemma 3.2 and Lemma 3.3 it is possible to show that weak convergence of $\overline{\alpha}_n(y)$ and of $\widetilde{\alpha}_n(y)$ are equivalent. Indeed, we have

$$|\overline{\alpha}_n(y) - \widetilde{\alpha}_n(y)| \leq \sqrt{k_n}\left|U_n^Z(y) - U_n^X(y)\right| + \sqrt{k_n}\left|\mathbb{E}[U_n^X(y)] - U(y)\right|.$$

The first term on the right hand side above becomes uniformly small on T_k due to Lemma 3.3, whereas the second one is smaller than

$$\frac{1}{\sqrt{k_n}}\sum_{i=0}^{\lfloor n/l_n \rfloor - 1}\left(\left|P(\overline{X}_{il_n}^n \geq y) - l_n\Delta_n U(y)\right| + \Delta_n U(y)\left|l_n - \frac{n}{\lfloor n/l_n \rfloor}\right|\right).$$

Using Lemma 3.2 the first summand can be bounded by

$$K\sqrt{k_n}(l_n\Delta_n + m_n/l_n),$$

whereas the second term is smaller than

$$K\frac{1}{\sqrt{k_n}}\Delta_n|l_n\lfloor n/l_n \rfloor - n| \leq K\frac{1}{\sqrt{k_n}}l_n\Delta_n,$$

and both claims hold uniformly on T_k. The latter bound is obviously smaller than the previous one, thus all terms converge to zero using (3.3).

In order to prove weak convergence of $\widetilde{\alpha}_n(y)$ we define a class of functions $\mathcal{F}_n = \{f_{n,y} : y \in T_k\}$ with corresponding envelope F_n via

$$f_{n,y}(u) = \sqrt{\frac{\lfloor n/l_n \rfloor}{k_n}}\mathbf{1}_{\{u \geq y\}} \quad \text{and} \quad F_n(u) = \sqrt{\frac{\lfloor n/l_n \rfloor}{k_n}}\mathbf{1}_{\{u \in T_k\}}.$$

With this notation we obtain

$$\widetilde{\alpha}_n(y) = (\lfloor n/l_n \rfloor)^{-1/2}\sum_{i=0}^{\lfloor n/l_n \rfloor - 1}\left(f_{n,y}(\overline{X}_{il_n}^n) - \mathbb{E}[f_{n,y}(\overline{X}_{il_n}^n)]\right).$$

Since the summands form a sequence of independent and identically distributed random variables, we are in a position to employ Theorem 11.20 in Kosorok (2008), for which six intermediate step have to the proven. Note first that the entropy condition on $\mathcal{F}_n$

follows from Lemma 11.21 in Kosorok (2008), since F_n is integrable with respect to any probability measure and the indicator functions $f_{n,y}(u)$ form a VC class with index 2. Furthermore, the sequence $\mathcal{F}_n$ is almost measurable Suslin due to the fact that the $f_{n,y}(u)$ are indicator functions; see the proof of Proposition 4.2 in Bücher and Vetter (2013) for an explicit reasoning. The remaining conditions are concerned with the computation of several (co)variances. We have

$$\lim_{n\to\infty} \mathbb{E}[\widetilde{\alpha}_n(x)\widetilde{\alpha}_n(y)] = \lim_{n\to\infty} (\mathbb{E}[f_{n,x}(\overline{X_0^n})f_{n,y}(\overline{X_0^n})] - \mathbb{E}[f_{n,x}(\overline{X_0^n})]\mathbb{E}[f_{n,y}(\overline{X_0^n})])$$

$$= \lim_{n\to\infty} \mathbb{E}[f_{n,x}(\overline{X_0^n})f_{n,y}(\overline{X_0^n})] = \lim_{n\to\infty} \lfloor n/l_n \rfloor \frac{l_n}{n} U(x \vee y) = U(x \vee y),$$

since $n/l_n \to \infty$ follows from $b < 1$ and $k_n = n\Delta_n \to \infty$. Also,

$$\limsup_{n\to\infty} \mathbb{E}[F_n^2(\overline{X_0^n})] \leq \limsup_{n\to\infty} \lfloor n/l_n \rfloor (l_n/n) U(1/k) = U(1/k) < \infty$$

and

$$\lim_{n\to\infty} \mathbb{E}[F_n^2(\overline{X_0^n})1_{\{F_n(\overline{X_0^n})>\varepsilon\sqrt{\lfloor n/l_n \rfloor}\}}] \leq \lim_{n\to\infty} \mathbb{E}[F_n^2(\overline{X_0^n})1_{\{k_n^{-1/2}>\varepsilon\rfloor}] = 0.$$

Finally, the convergence

$$\rho_n^2(x,y) = \mathbb{E}[|f_{n,x}(\overline{X_0^n}) - f_{n,y}(\overline{X_0^n})|^2]$$

$$\to |U(x) - 2U(x \vee y) + U(y)|^2 = |U(x) - U(y)|^2 = \rho^2(x,y)$$

holds uniformly on T_k which proves the final condition for an application of Theorem 11.20 in Kosorok (2008). □

Remark 3.5 There are several valid choices for a and b such that (3.3) is satisfied. We propose one for which the finite sample bias due to all of these is roughly of the same order. To this end, note that (3.3) is satisfied whenever

$$\sqrt{k_n}\left(\Delta_n^{\eta a/2} + \Delta_n^{\eta(b-a)} + \Delta_n^{\eta(1-b)}\right) \to 0$$

for some $0 < \eta < 1$. All three exponents are the same for $a = 1/2$ and $b = 3/4$. In this case, Theorem 3.4 holds, if there exists some $0 < \eta < 1$ such that

$$k_n\Delta_n^{\eta/2} = n\Delta_n^{1+\eta/2} \to 0 \tag{3.5}$$

is satisfied. □

Remark 3.6 A key result in the work of Bücher and Vetter (2013) was

$$\alpha_n(y) = \sqrt{k_n}\left(U_n^X(y) - U(y)\right) \xrightarrow{w} \mathbb{B}(y)$$

in $(\mathcal{B}_\infty(0,\infty], d)$, where $\mathbb{B}$ is (in distribution) the same Gaussian process as the one in Theorem 3.4. Therefore, the asymptotic behaviour of $\overline{U_n^Z}(y)$ is the same as in the case of non-noisy observations. This is remarkable compared with the diffusion case where

already the optimal rate of convergence is of a larger order than the one without noise. See e.g. Gloter and Jacod (2001). What changes to the setting without noise, however, are the assumptions on the choice of n and Δ_n. Whereas in Bücher and Vetter (2013) the crucial condition besides $\Delta_n \to 0$ and $n\Delta_n \to \infty$ was $\sqrt{k_n}\Delta_n = \sqrt{n\Delta_n^3} \to 0$, the precedingly obtained assumption in (3.5) is much tighter. $\qquad\square$

Remark 3.7 A natural alternative to the estimator $\overline{U_n^Z}(y)$ is to use

$$\widehat{U_n^Z}(y) = \frac{1}{k_n l_n} \sum_{i=0}^{n-l_n} 1_{\{\widetilde{Z_{il_n}^n} \geq y\}},$$

for which all pre-averaging statistics are used and not only independent ones computed over non-overlapping intervals. This makes a proof of weak convergence much more involved, and, more importantly, even though we are convinced that such a claim holds, no gain in terms of a lower variance is to be expected, since the limiting process in Theorem 3.4 already takes the same form as the one in the absence of noise. This intuition is supported by some tedious computations involving blocking techniques which prove that the finite dimensional distributions of

$$\widehat{\alpha}_n(y) = \sqrt{k_n}\left(\widehat{U_n^Z}(y) - U(y)\right)$$

convergence to the same ones as the corresponding ones of $\overline{\alpha}_n(y)$. $\qquad\square$

References

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

Bücher, A. and Vetter, M. (2013). Nonparametric inference on Lévy measures and copulas. *Annals of Statistics*. to appear.

Comte, F. and Genon-Catalot, V. (2010). Non-parametric estimation for pure jump irregularly sampled or noisy Lévy processes. *Stat. Neerl.*, 64(3):290–313.

Figueroa-López, J. E. and Houdré, C. (2009). Small-time expansions for the transition distributions of Lévy processes. *Stochastic Process. Appl.*, 119(11):3862–3889.

Gloter, A. and Jacod, J. (2001). Diffusions with measurement errors. I. Local asymptotic normality. *ESAIM Probab. Statist.*, 5:225–242 (electronic).

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

Kosorok, M. R. (2008). *Introduction to Empirical Processes and Semiparametric Inference*. Springer, New York.

Sato, K.-i. (1999). *Lévy processes and infinitely divisible distributions*, volume 68 of *Cambridge Studies in Advanced Mathematics*. Cambridge University Press, Cambridge. Translated from the 1990 Japanese original, Revised by the author.

van der Vaart, A. W. and Wellner, J. A. (1996). *Weak Convergence and Empirical Processes - Springer Series in Statistics*. Springer, New York.

Zhang, L., Mykland, P. A., and Aït-Sahalia, Y. (2005). A tale of two time scales: determining integrated volatility with noisy high-frequency data. *J. Amer. Statist. Assoc.*, 100(472):1394–1411.