Power Variation from Second Order Differences for Pure Jump Semimartingales*

Viktor Todorov†

March 22, 2013

Abstract

We introduce power variation constructed from powers of the second-order differences of a discretely observed pure-jump semimartingale processes. We derive the asymptotic behavior of the statistic in the setting of high-frequency observations of the underlying process with a fixed time span. Unlike the standard power variation (formed from the first-order differences of the process), the limit of our proposed statistic is determined solely by the jump component of the process regardless of the activity of the latter. We further show that an associated Central Limit Theorem holds for a wider range of activity of the jump process than for the standard power variation. We apply these results to estimation of the jump activity as well as the integrated stochastic scale.

AMS 2000 subject classifications: 60F05, 60F17, 60G51, 60G07.

Keywords: Lévy process, Blumenthal-Getoor index, Jump Activity, Power variation, Stable Convergence.

*I would like to thank the editor (Per Mykland) and a referee for constructive comments and suggestions. Work partially supported by NSF Grant SES-0957330.

†Department of Finance, Kellogg School of Management, Northwestern University, Evanston, IL 60208; e-mail: v-todorov@northwestern.edu.

1 Introduction

Many stochastic processes of interest, e.g., asset prices in finance, are modeled using stochastic differential equations of the form

$$dX_t = \alpha_t dt + \sigma_{t-} dL_t + dY_t, \tag{1}$$

where L_t is a Lévy process, which is a martingale, if of infinite variation, and is a sum of jumps, in the finite variation case; α_t and σ_t are some processes with càdlàg paths, and Y_t is some "residual" jump component whose behavior over small time scales is dominated by the second term in (1). The leading example is the diffusion case in which L_t is a Brownian motion, but in some applications, particularly in finance, it is important to allow for more general driving Lévy processes that include jumps (in addition to the Brownian motion) or are even of pure-jump type, i.e., without Brownian motion. σ_t in (1) is typically referred to as stochastic volatility when L_t is a Brownian motion or more generally as stochastic scale when L_t is "locally" a stable process, i.e., when the small scale behavior of L_t is like that of a stable process (we will be precise about the "locally" stable processes in the next section).

When the process X is observed on a fine grid, $X_0, X_{\frac{1}{n}}, ..., X_{\frac{n}{n}}$, with mesh decreasing to zero, the realized power variation defined as

$$V_n(p, X) = \sum_{i=1}^{n} |\Delta_i^n X|^p, \quad \Delta_i^n X = X_{\frac{i}{n}} - X_{\frac{i-1}{n}}, \tag{2}$$

can be used to study the realized path of the latent stochastic scale, σ_t, as well as the small scale properties of the driving Lévy process L_t. In particular, if $\alpha_t = 0$ on the time interval $[0, 1]$, then with β denoting the stability parameter of the "locally" stable process L_t ($\beta \in (0, 2]$, and $\beta = 2$ means that the leading component of L_t over small scales is a Brownian motion), we have under certain regularity conditions, see e.g., Todorov and Tauchen (2011) and Woerner (2003, 2007),

$$\Delta_n^{1-p/\beta} V_n(p, X) \xrightarrow{\mathbb{P}} \mu_p(\beta) \int_0^1 |\sigma_s|^p ds, \quad p < \beta, \tag{3}$$

and their is an associated Central Limit theorem (CLT) (under some further restrictions on the "residual" components), see again Todorov and Tauchen (2011) and Woerner (2003, 2007),

$$\Delta_n^{-1/2} \left(\Delta_n^{1-p/\beta} V_n(p, X) - \mu_p(\beta) \int_0^1 |\sigma_s|^p ds \right) \xrightarrow{\mathcal{L}-s} \sqrt{\mu_{2p}(\beta) - \mu_p^2(\beta)} \sqrt{\int_0^1 |\sigma_s|^{2p} ds} Z, \quad p < \beta/2, \tag{4}$$

where Z is standard normal variable defined on an extension of the original probability space and $\mathcal{L} - s$ denotes stable convergence in law; $\mu_p(\beta)$ and $\mu_{2p}(\beta)$ are respectively the p-th and $2p$-th absolute moments of the leading stable component of L_t with further details provided later on. The result in (4) can be made also feasible by constructing estimators of the asymptotic variance and making use of the properties of the stable convergence.

Since the index β participates in the scaling of the realized power variation in (3), the latter computed over different scales can be used to construct (nearly) rate efficient nonparametric estimators of the activity index β of the driving Lévy process. Further efficiency gains can be made by picking adaptively the optimal power to estimate β from.

The limit results in (3) and (4) depend strongly on the assumption of zero drift, i.e., $\alpha_t = 0$. However, if α_t is not identically zero on the time interval $[0,1]$, the limit result in (3) continues to hold but only when $\beta > 1$. On the other hand, if $\beta < 1$ (and recall in this case L_t and Y_t are sums of jumps without a drift), we have, see e.g., Jacod (2008),

$$\Delta_n^{1-p} V_n(p,X) \xrightarrow{\mathbb{P}} \int_0^1 |\alpha_s|^p ds, \quad p < \beta, \tag{5}$$

that is the limit behavior of the realized power variation is governed by the drift term and not the driving pure-jump Lévy process L_t. Moreover, even when $\beta > 1$, the CLT result in (4) is unaffected by the presence of the drift term but only when $\beta > \sqrt{2}$. When $\beta \in (0, \sqrt{2})$, the presence of the drift term slows down the rate of convergence and invalidates feasible inference based on (4).

In this paper, we propose an extension of the original realized power variation that allows to make efficient nonparametric inference about the stochastic scale and the activity index β even in the case when a drift is present in X_t and the driving Lévy process L_t is of finite variation, i.e., when $\beta < 1$. The extension is based on replacing the first-order difference $\Delta_i^n X$ with the second-order difference $\Delta_i^n X - \Delta_{i-1}^n X$ in the construction of the realized power variation in (2). The effect of this is easiest to see in the simplest case when X_t is a Lévy process, i.e., when α_t and σ_t are constant. In this case $\Delta_i^n X - \Delta_{i-1}^n X$, unlike $\Delta_i^n X$, does not contain the drift. Moreover, the leading component of $\Delta_i^n L - \Delta_{i-1}^n L$ (under the "local" stability assumption for L_t) over small scales is a difference of (scaled) independent and identically distributed stable random variables which continues to be a stable random variable (but with a different scale). Intuitively, our modified realized power variation makes use of the difference in the pathwise behavior of the drift term and the driving Lévy process in (1) over small scales.

A law of large numbers (LLN) for the realized power variation based on second-order differences (with a limit that differs from the one in (3) only by a constant) continues to hold, but now the LLN holds without any restrictions regarding the presence of the drift term in (1), unlike the one for the original realized power variation in (2). Moreover, we have a CLT for our statistic for any value of the activity index β and regardless of the presence of a drift term when X_t is a Lévy process and for the relatively wide range $\beta > \frac{2}{3}$ in the general case when α_t and σ_t can be random.

The rate of the convergence of the realized power variation based on second-order differences of X is the square root of the high-frequency observations within the fixed time interval and is nearly the optimal one (the difference is $\ln(n)$) in the special case of estimating the scale and activity of a stable process from high-frequency observations, see e.g., Ait-Sahalia and Jacod (2008). The robustness of our realized power variation, based on $\Delta_i^n X - \Delta_{i-1}^n X$, to the presence of a drift term in X results in some loss of information when compared with the original power variation $V_n(p,X)$ in the case when X does not contain a drift term. This is reflected in a somewhat larger asymptotic variance for estimating the integrated power variation of the stochastic scale, as well as the loss of the information in the data regarding the potential asymmetry of the Lévy measure of L_t around the origin.

Finally, we note that in this paper we are only interested in the behavior of the power variation (and the modification proposed here) for powers below the activity β. It is well-known, see e.g., Lepingle (1976), that for $p > \beta \vee 1$, the power variation (without any scaling) converges in probability to the sum of the p-th absolute moments of the jumps on the time interval (see also Diop et al. (2013) for an associated CLT for the special case $p = 2$). It is easy to see that this limit result will continue to hold for our realized power variation based on the second-order differences of X.

The rest of the paper is organized as follows. Section 2 introduces the setup and the assumptions. In Section 3 we present the limit results for the realized power variation formed from second-order differences of the process both in the Lévy case and the more general case when the drift and the stochastic scale can be random. Section 4 applies the developed limit theory to propose (nearly) rate efficient and robust estimator of activity as well as the integrated power variation of the stochastic scale. Section 5 contains all the proofs.

2 Setting and Assumptions

We start with introducing the setting and stating the assumptions that we need for the results in the paper. We first recall that a Lévy process L_t with characteristic triplet (b, c, ν), with respect to truncation function κ (Definition II.2.3 in Jacod and Shiryaev (2003)), is a process with characteristic function given by

$$\mathbb{E}\left(e^{iuL_t}\right) = \exp\left[itub - tcu^2/2 + t\int_{\mathbb{R}}\left(e^{iux} - 1 - iu\kappa(x)\right)\nu(dx)\right]. \tag{6}$$

In what follows we will always assume for simplicity that $\kappa(-x) = -\kappa(x)$. Our assumption for the driving Lévy process in (1) as well as the "residual" jump component Y_t is given in assumption A.
Assumption A. *L_t in (1) is a Lévy process with characteristic triplet $(0, 0, \nu)$ for ν Lévy measure with density given by*

$$\nu(x) = \frac{A_-}{|x|^{1+\beta}}1_{\{x<0\}} + \frac{A_+}{|x|^{1+\beta}}1_{\{x>0\}} + \nu'(x), \quad A_\pm \geq 0 \text{ with } \max\{A_-, A_+\} > 0, \tag{7}$$

where $\nu'(x)$ is such that there exists $x_0 > 0$ with $|\nu'(x)| \leq C/|x|^{1+\beta'}$ for $|x| \leq x_0$ and some $\beta' < \beta$.

Y_t is Itô semimartingale with characterstic triplet (Jacod and Shiryaev (2003), Definition II.2.6) $\left(\int_0^t\int_{\mathbb{R}}\kappa(x)\nu_s^Y(dx)ds, 0, dt\otimes\nu_t^Y(dx)\right)$ when $\beta' < 1$ and $\left(0, 0, dt\otimes\nu_t^Y(dx)\right)$ otherwise, with $\int_{\mathbb{R}}(|x|^{\beta'+\iota}\wedge 1)\nu_t^Y(dx)$ being locally bounded predictable for some arbitrary small $\iota > 0$.

Assumption A implies that the Lévy measure of L_t around zero is dominated by that of a stable process. This, in particular means that the behavior of L_t over small scales is like that of a stable process. That is, for $\beta \in (1, 2)$ we have $h^{-1/\beta}L_{ht} \xrightarrow{\mathcal{L}} S_t$ as $h \to 0$, for S_t being a Lévy process with characteristic triplet $\left(-\int_{\mathbb{R}}\kappa'(x)\left(\frac{A_-}{|x|^{1+\beta}}1_{\{x<0\}} + \frac{A_+}{|x|^{1+\beta}}1_{\{x>0\}}\right)dx, 0, \left(\frac{A_-}{|x|^{1+\beta}}1_{\{x<0\}} + \frac{A_+}{|x|^{1+\beta}}1_{\{x>0\}}\right)dx\right)$ (and the convergence being in the space of càdlàg functions equipped with the Skorokhod topology and further $\kappa'(x) = x - \kappa(x)$). For $\beta \in (0, 1)$ we have the above convergence when $\kappa(x) = 0$ (no truncation) and S_t being a Lévy process with characteristic triplet $\left(0, 0, \left(\frac{A_-}{|x|^{1+\beta}}1_{\{x<0\}} + \frac{A_+}{|x|^{1+\beta}}1_{\{x>0\}}\right)dx\right)$. Finally for $\beta = 1$ we have $h^{-1/\beta}L_{ht} \xrightarrow{\mathcal{L}} S_t$ as $h \to 0$, for S_t being a Lévy process with characteristic triplet $\left(0, 0, \left(\frac{A_-}{|x|^{1+\beta}}1_{\{x<0\}} + \frac{A_+}{|x|^{1+\beta}}1_{\{x>0\}}\right)dx\right)$ provided $A_- = A_+$.

Assumption A is critical for what follows as it allows scaling of the increments of the process X_t in forming the power variation. It is satisfied by many parametric specifications of the Lévy process like the stable, the tempered stable and the generalized inverse Gaussian. Similar assumption has been made in related contexts in Ait-Sahalia and Jacod (2009), Jacod (2004), Todorov and Tauchen (2011, 2012), Woerner (2003, 2007).

Note further that $h^{-1/\beta}Y_{ht} \xrightarrow{\mathcal{L}} 0$ as $h \to 0$, and this is why Y_t is dominated over small time scales by L_t (and stochastic integrals with respect to it).

Remark. *The presence of Y_t in (1) allows to include in our setup also time-changed Lévy models when the time-change is absolute continuous and these types of models have been used in various financial applications, see e.g., Barndorff-Nielsen and Shiryaev (2010) and the many references therein. To see this, suppose that $\widehat{L}_v$ is pure-jump Lévy process with Lévy measure $\nu(x)dx$ (for $\nu(x)$ given in (7)) and $T_t = \int_0^t a_s ds$ is the time-change, for a_t being some predictable process. Then, the time-changed Lévy process $\widehat{L}_{T_t}$ is $\mathcal{F}_{T_t}$-adapted pure-jump process with jump compensator $a_t dt \otimes \nu(x)dx$. This follows from Theorem 10.27 in Jacod (1979). Then, assuming for simplicity that $\nu'(x) \geq 0$ (for the more general case one can use the jump decomposition in Section 5.1.2), we can write $\widehat{L}_{T_t} = X_0 + X_t^{(1)} + X_t^{(2)}$ where $X_t^{(1)}$ is pure-jump process with compensator $a_t dt \otimes \left(\frac{A_-}{|x|^{1+\beta}} 1_{\{x<0\}} + \frac{A_+}{|x|^{1+\beta}} 1_{\{x>0\}}\right) dx$ and $X_t^{(2)}$ is pure-jump process with compensator $a_t dt \otimes \nu'(x)dx$. The "residual" $X_t^{(2)}$ does not matter for our asymptotics (and corresponds to Y_t in (1)). On the other hand, $X_t^{(1)}$ can be represented in the form $\int_0^t \int_{\mathbb{R}_+} \kappa(a_s^{1/\beta} x)\underline{\widetilde{\mu}}(ds, dx) + \int_0^t \int_{\mathbb{R}_+} \kappa'(a_s^{1/\beta} x)\underline{\mu}(ds, dx)$, where $\underline{\mu}$ is homogenous Poisson measure with compensator $dt \otimes \left(\frac{A_-}{|x|^{1+\beta}} 1_{\{x<0\}} + \frac{A_+}{|x|^{1+\beta}} 1_{\{x>0\}}\right) dx$. This follows from a "representation theorem" for integer-valued random measures, see for example Theorem 2.1.2 and the discussion in Section 2.1.4 of the book of Jacod and Protter (2012). Thus the time-changed Lévy process $\widehat{L}_{T_t}$ has a representation of the form (1) with σ_t replaced by $a_t^{1/\beta}$.*

We next make an assumption regarding the variability in the processes α_t and σ_t.

Assumption B. *The processes α_t and σ_t are Itô semimartingales of the form*

$$\begin{aligned}
\alpha_t &= \alpha_0 + \int_0^t b_s^\alpha ds + \int_0^t \int_E \kappa(\delta^\alpha(s,x))\underline{\widetilde{\mu}}(ds,dx) + \int_E \kappa'(\delta^\alpha(s,x))\underline{\mu}(ds,dx), \\
\sigma_t &= \sigma_0 + \int_0^t b_s^\sigma ds + \int_0^t \int_E \kappa(\delta^\sigma(s,x))\underline{\widetilde{\mu}}(ds,dx) + \int_E \kappa'(\delta^\sigma(s,x))\underline{\mu}(ds,dx),
\end{aligned} \tag{8}$$

where $\kappa'(x) = x - \kappa(x)$, and

(a) *$|\sigma_t|^{-1}$ and $|\sigma_{t-}|^{-1}$ are strictly positive;*

(b) *$\underline{\mu}$ is Poisson measure on $\mathbb{R}_+ \times E$, having arbitrary dependence with the jump measure of L_t, with compensator $dt \otimes \lambda(dx)$;*

(c) *$\delta^\alpha(t,x)$ and $\delta^\sigma(t,x)$ are predictable, left-continuous with right limits in t with $|\delta^\alpha(t,x)| + |\delta^\sigma(t,x)| \leq \gamma_k(x)$ for all $t \leq T_k$, where $\gamma_k(x)$ is a deterministic function on $\mathbb{R}$ with $\int_{\mathbb{R}}(|\gamma_k(x)|^{\beta+\iota} \wedge 1)dx < \infty$ for arbitrary small $\iota > 0$ and β being the constant in (7), and T_k is a sequence of stopping times increasing to $+\infty$;*

(d) *b_t^α and b_t^σ are Itô semimartingales having dynamics as in (8) with coefficients satisfying the analogues of conditions (b) and (c) above.*

We note that the jump measure $\underline{\mu}$ does not need to coincide with the jump measure of L_t but it can have arbitrary dependence with it. Assumption B is satisfied in models where the triple $(X_t, \alpha_t, \sigma_t)$ is modeled via a Lévy-driven multivariate SDE with each of the elements of the driving Lévy process satisfying assumption A. Importantly assumption B allows for dependence between the innovations in α_t, σ_t and the driving Lévy process L_t which is of significant importance for financial applications. For example, assumption B is satisfied by the COGARCH model of Klüppelberg et al. (2004) in which the jumps in σ_t are proportional to the squared jumps in X_t or non-Gaussian

Ornstein-Uhlenbeck models for the stochastic scale in which the jumps in σ_t are proportional to the jumps in X_t as in Barndorff-Nielsen and Shephard (2001) or proportional to the squared price jumps as in Todorov (2011).

3 Limit Theory

We proceed with our limit results. Our asymptotics is for fixed time span and increasing sampling frequency of observations. In particular, we assume that the process X is observed on the equidistant grid $0, \frac{1}{n}, \frac{2}{n}, ..., 1$ with $n \to \infty$. We leave the extension to the case of random sampling times for future work.

Our power variation statistics constructed from the second-order differences of the process X is defined formally as

$$\widetilde{V}_n^1(p, X) = \sum_{i=2}^{n} |\Delta_i^n X - \Delta_{i-1}^n X|^p, \quad p > 0. \tag{9}$$

Since, we are going to apply the limit results for estimation of the activity index β, we also define here realized power variation formed from temporal aggregation of the second-order differences of the process X

$$\widetilde{V}_n^2(p, X) = \sum_{i=4}^{n} |\Delta_i^n X - \Delta_{i-1}^n X + \Delta_{i-2}^n X - \Delta_{i-3}^n X|^p, \quad p > 0. \tag{10}$$

The important thing about $\widetilde{V}_n^2(p, X)$ is that the temporal aggregation of the second-order differences results in its asymptotic limit being proportional to that of $\widetilde{V}_n^1(p, X)$ by a constant that depends solely on p and β. This allows for the inference of β from the ratio of $\widetilde{V}_n^1(p, X)$ and $\widetilde{V}_n^2(p, X)$. We should point out that there are other alternative ways of constructing $\widetilde{V}_n^2(p, X)$ that achieve the same goal. One such example is $\sum_{i=2}^n |\Delta_i^n X + \Delta_{i+1}^n X - 2\Delta_{i-1}^n X|^p$. The construction here is similar to the use of realized power variation in (2) on a coarser scale in Todorov and Tauchen (2011), but unlike that paper we make more efficient use of the data as intuitively we employ all temporally aggregated increments and not just the non-overlapping ones.

We proceed with introducing some more notation that we will use in stating the asymptotic results. We denote with S_i, for $i = 1, 2, ...$ a sequence of β-stable random variables that are independent of each other and whose distribution corresponds to the law at time 1 of a Lévy process with characteristic triplet $\left(0, 0, (\frac{A_-}{|x|^{1+\beta}} 1_{\{x<0\}} + \frac{A_+}{|x|^{1+\beta}} 1_{\{x>0\}})dx\right)$. We note that $S_1 - S_2$ has distribution that is equal to the law at time 1 of the Lévy process with characteristic triplet $\left(0, 0, \frac{A_- + A_+}{|x|^{1+\beta}} dx\right)$ which is a symmetric stable process.

We denote $\widetilde{\mu}_p(\beta) = \mathbb{E}|S_1 - S_2|^p$. Note here that $\widetilde{\mu}_p(\beta)$ is the p-th absolute moment of $S_1 - S_2$ and not of S_1. This is done for convenience of exposition as the asymptotic limits of $\widetilde{V}_n^1(p, X)$ and $\widetilde{V}_n^2(p, X)$ depend on the moments of the second-order differences $\Delta_i^n L - \Delta_{i-1}^n L$ and not on the moments of the first order differences $\Delta_i^n L$. We further set

$$\mathbf{Z}_i = \begin{pmatrix} |S_i - S_{i+1}|^p - \widetilde{\mu}_p(\beta) \\ |S_i - S_{i+1} + S_{i+2} - S_{i+3}|^p - 2^{p/\beta}\widetilde{\mu}_p(\beta) \end{pmatrix},$$

and then denote $\mathbf{\Sigma}_j(p, \beta) = \mathbb{E}(\mathbf{Z}_1 \mathbf{Z}_{1+j}')$ for $j = 1, 2, ...$ We note that $\mathbf{\Sigma}_j(p, \beta)$ depends on whether $A_- = A_+$ or not (but of course it does not depend on the particular choice of κ with respect to which characteristics are defined).

With this additional notation we are ready to state our results for the pair $(\widetilde{V}_n^1(p,X), \widetilde{V}_n^2(p,X))$. We first present the results for the basic case when X_t is a Lévy process and then extend them to the general case when α_t and σ_t can vary over time.

3.1 The Lévy case

Theorem 1 *Assume that X_t satisfies assumption A with α_t and σ_t constant.*

(a) For $p \in (0,\beta)$ and provided $\beta' < \beta$, we have

$$\widetilde{V}_n^1(p,X) \xrightarrow{\mathbb{P}} \widetilde{\mu}_p(\beta)|\sigma|^p, \qquad \widetilde{V}_n^2(p,X) \xrightarrow{\mathbb{P}} 2^{p/\beta}\widetilde{\mu}_p(\beta)|\sigma|^p. \tag{11}$$

(b) For $p \in \left(\frac{\beta\beta'}{2(\beta-\beta')}, \frac{\beta}{2}\right)$ and provided $\beta' < \beta/2$, we have

$$\sqrt{n}\begin{pmatrix} \widetilde{V}_n^1(p,X) - \widetilde{\mu}_p(\beta)|\sigma|^p \\ \widetilde{V}_n^2(p,X) - 2^{p/\beta}\widetilde{\mu}_p(\beta)|\sigma|^p \end{pmatrix} \xrightarrow{\mathcal{L}-s} |\sigma|^p \mathbf{\Xi}^{1/2}\mathbf{Z}, \tag{12}$$

where $\mathbf{Z}$ is two-dimensional standard normal random variable defined on an extension of the original probability space and independent from $\mathcal{F}$ and

$$\mathbf{\Xi} = \mathbf{\Sigma}_0(p,\beta) + \sum_{i=1}^{3}(\mathbf{\Sigma}_i(p,\beta) + \mathbf{\Sigma}_i'(p,\beta)). \tag{13}$$

Part (a) of Theorem 1 shows that our statistic $\widetilde{V}_n^1(p,X)$ estimates the integrated power variation of the scale (which in the Lévy case is a constant) regardless of the activity level β unlike the original realized power variation statistic $V_n(p,X)$ which does so only provided $\beta > 1$. We note that the first limit in (11) differs from that in (3) by a constant that equals the ratio of the p-th absolute moments of $S_1 - S_2$ and S_1. In the case when $A_- = A_+$, this ratio is simply $2^{p/\beta}$.

Part (b) presents the CLT for our statistics. As for the law of large numbers, the CLT result holds for any value of β and this is significant improvement over the corresponding result for the power variation derived in Todorov and Tauchen (2011), and stated here in equation (4) in the introduction, that holds only for $\beta > \sqrt{2}$. The presence of the $\mathbf{\Sigma}_i(p,\beta)$ terms, for $i \geq 1$, in $\mathbf{\Xi}$ is due to the third-order autocorrelation in the summands of $(\widetilde{V}_n^1(p,X), \widetilde{V}_n^2(p,X))$.

Unlike the original power variation, the asymptotic limit of the power variation based on second-order differences depends only on the moments of a symmetric β-stable process regardless of whether $A_- = A_+$ or not. The reason is pretty straightforward: $\widetilde{V}_n^1(p,X)$ and $\widetilde{V}_n^2(p,X)$ depend only on $\Delta_i^n L - \Delta_{i-1}^n L$ which "symmetrizes" the potentially asymmetric Lévy process L. On the other hand, the asymptotic variance of $\widetilde{V}_n^1(p,X)$ and $\widetilde{V}_n^2(p,X)$ will depend on the potential asymmetry of the Lévy density of L around zero because the dependence between consecutive summands in $\widetilde{V}_n^1(p,X)$ and $\widetilde{V}_n^2(p,X)$ depends on the latter. Consistent estimators of the limiting variance, however, should be easy to form. This of course carries over to the activity estimation based on the limit theory developed here that we conduct in the next section. To make the limiting variance independent of the potential Lévy density asymmetry of L at the origin, we can keep every second summand in $\widetilde{V}_n^1(p,X)$ and every forth in $\widetilde{V}_n^2(p,X)$. We note also that, similar to the case of the realized power variation $V_n(p,X)$, there is a CLT for $\widetilde{V}_n^1(p,X)$ in the case when $p \in (\beta/2,\beta)$ with a limit that is stable and with a rate of convergence that is slower than the $\sqrt{n}$ rate in the case $p \in (0,\beta/2)$.

Finally, the differencing of $\Delta_i^n X$ leads naturally to some loss of efficiency in estimating $|\sigma|^p$ (or more generally the integrated power variation of the scale in the general case when σ_t varies) when β is known. On Figure 1 we plot the ratio of the asymptotic errors in estimating $|\sigma|^p$ using $\widetilde{V}_n^1(p,X)$ and $V_n(p,X)$ for different values of β in the case when $A_- = A_+$. As we can see from the figure, there is efficiency loss of up to around 40% in estimating the scale parameter from using second-order differences of X. The differences decrease for higher values of β. This loss of efficiency is the price to pay for robustifying the inference with respect to the presence of a drift term in the evolution of X.

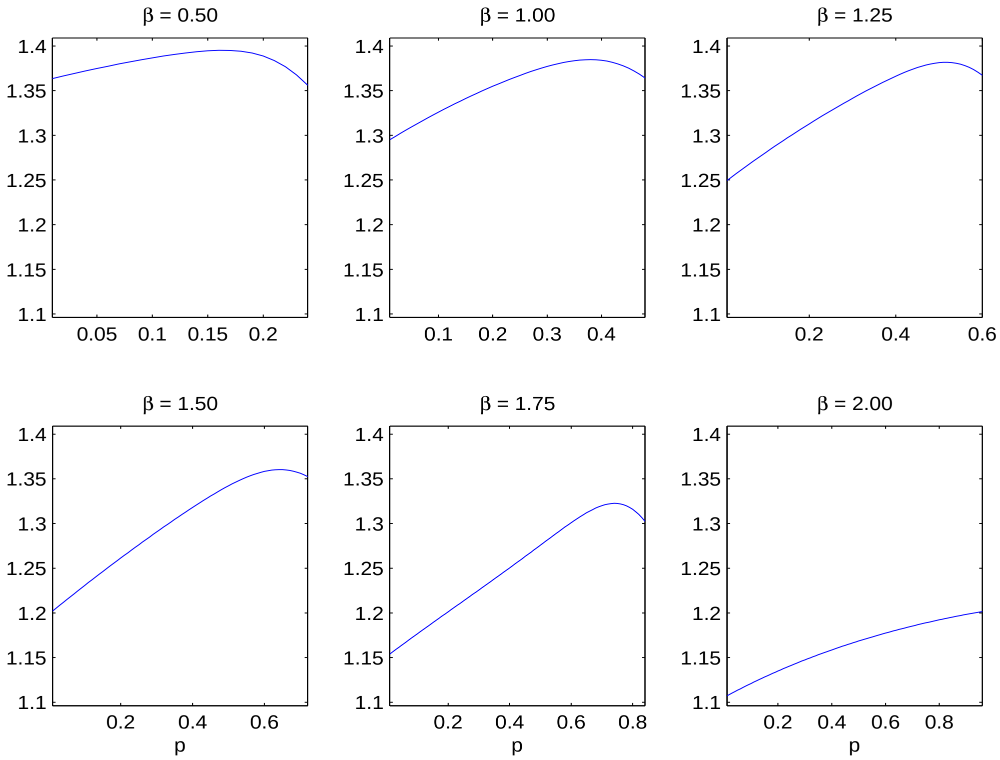

Figure 1: Ratio of the asymptotic standard deviations in estimating $|\sigma|^p$ using $\widetilde{V}_n^1(p,X)$ and $V_n(p,X)$ as a function of p, when X_t is a Lévy process (with no drift) and $A_- = A_+$. The ratio is given by $\sqrt{1 + \frac{2}{\mu_{2p}(\beta)-\mu_p^2(\beta)}\mathbb{E}\left\{\left(\frac{|S_1-S_2|^p}{2^{p/\beta}} - \mu_p(\beta)\right)\left(\frac{|S_2-S_3|^p}{2^{p/\beta}} - \mu_p(\beta)\right)\right\}}$. The limiting case $\beta = 2$ corresponds to the case when L_t is a Brownian motion and the asymptotic distribution of $|\sigma|^2$ in this case continues to be given by the first limit in (12).

3.2 The general case

We present next the analogue of Theorem 1 for the case when α_t and σ_t can be time varying.

Theorem 2 *Assume that X_t satisfies assumptions A and B.*

(a) For $p \in (0, \beta)$ and provided $\beta' < \beta$, we have

$$\Delta_n^{1-p/\beta} \widetilde{V}_n^1(p, X) \xrightarrow{\mathbb{P}} \widetilde{\mu}_p(\beta) \int_0^1 |\sigma_s|^p ds, \quad \Delta_n^{1-p/\beta} \widetilde{V}_n^2(p, X) \xrightarrow{\mathbb{P}} 2^{p/\beta} \widetilde{\mu}_p(\beta) \int_0^1 |\sigma_s|^p ds. \tag{14}$$

(b) For $p \in \left(\frac{|\beta - 1|}{2(\beta \wedge 1)} \bigvee \frac{\beta\beta'}{2(\beta - \beta')}, \frac{\beta}{2} \right)$ and provided $\beta' < \beta/2$ as well as $\beta > \frac{2}{3}$, we have

$$\sqrt{n} \begin{pmatrix} \Delta_n^{1-p/\beta} \widetilde{V}_n^1(p, X) - \widetilde{\mu}_p(\beta) \int_0^1 |\sigma_s|^p ds \\ \Delta_n^{1-p/\beta} \widetilde{V}_n^2(p, X) - 2^{p/\beta} \widetilde{\mu}_p(\beta) \int_0^1 |\sigma_s|^p ds \end{pmatrix} \xrightarrow{\mathcal{L}-s} \sqrt{\int_0^1 |\sigma_s|^{2p} ds} \, \mathbf{\Xi}^{1/2} \mathbf{Z}, \tag{15}$$

where $\mathbf{Z}$ is two-dimensional standard normal random variable defined on an extension of the original probability space and independent from $\mathcal{F}$ and $\mathbf{\Xi}$ is defined in (13).

There are few observations to be made. First, the law of large numbers result for $\widetilde{V}_p^1(p, X)$ and $\widetilde{V}_p^2(p, X)$ continues to hold without any further restrictions on the activity β or on the power p. The CLT result in (15), however, holds under the additional restriction of $\beta > \frac{2}{3}$ (and some additional restrictions on the power p). This additional restriction arises from bounding the effect due to the time variation in α_t and σ_t on the asymptotic behavior of $\widetilde{V}_p^1(p, X)$ and $\widetilde{V}_p^2(p, X)$. Its absence from Theorem 1 suggests that some additional structure on α_t and σ_t can allow to weaken it further. Nevertheless, Theorem 2 is still very general, generalizes significantly the corresponding result for $V_n(p, X)$ in Todorov and Tauchen (2011), and in particular allows for efficient inference for all infinite variation cases, i.e., when X satisfies assumption A with $\beta \in (1, 2)$.

4 Application to Activity and Stochastic Scale Estimation

We proceed next with applying the limit theory of the previous section to estimation of the activity index β as well as the integrated power variation of the stochastic scale when the activity β is unknown. We start with the estimation of the activity. Similar to Todorov and Tauchen (2011), we define the activity estimator from the ratio of $\widetilde{V}_n^2(p, X)$ and $\widetilde{V}_n^1(p, X)$ as

$$\widehat{\beta} = \frac{p \log(2)}{\log \left[\widetilde{V}_n^2(p, X) / \widetilde{V}_n^1(p, X) \right]} 1_{\{\widetilde{V}_n^1(p,X) \neq \widetilde{V}_n^2(p,X)\}}. \tag{16}$$

In the next Corollary we state the asymptotic behavior of $\widehat{\beta}$ in the Lévy case.

Corollary 1 *Assume that X_t satisfies assumption A with α_t and σ_t constant. For $p \in (0, \beta)$ we have $\widehat{\beta} \xrightarrow{\mathbb{P}} \beta$. If in addition $p \in \left(\frac{\beta\beta'}{2(\beta - \beta')}, \frac{\beta}{2} \right)$, and provided $\beta' < \beta/2$, we have*

$$\sqrt{n} \left(\widehat{\beta} - \beta \right) \xrightarrow{\mathcal{L}-s} \frac{\beta^2}{\widetilde{\mu}_p(\beta) p \log(2)} \sqrt{\widetilde{\Xi}} \, Z, \tag{17}$$

where $\widetilde{\Xi} = \mathbf{\Xi}^{(1,1)} - 2^{1-p/\beta} \mathbf{\Xi}^{(1,2)} + 2^{-2p/\beta} \mathbf{\Xi}^{(2,2)}$ and Z is univariate standard normal defined on an extension of the original probability space and independent of the σ-algebra $\mathcal{F}$.

On Figure 2 we plot the asymptotic standard deviation of the activity estimator $\widehat{\beta}$ defined in (16) as a function of the power p for different values of the activity index β. Comparing with the corresponding asymptotic standard errors for the activity estimator based on $V_n(p, X)$ computed over different scales proposed in Todorov and Tauchen (2011), we see that the standard errors are quite comparable. Of course, the proposed estimator here has the additional advantage of robustness against the presence of a drift term in the dynamics of X. We also note that the efficiency of the estimator depends critically on the power p and the true value of β. Therefore, an adaptive estimation strategy similar to Todorov and Tauchen (2011), based on Corollary 1, can be further developed.

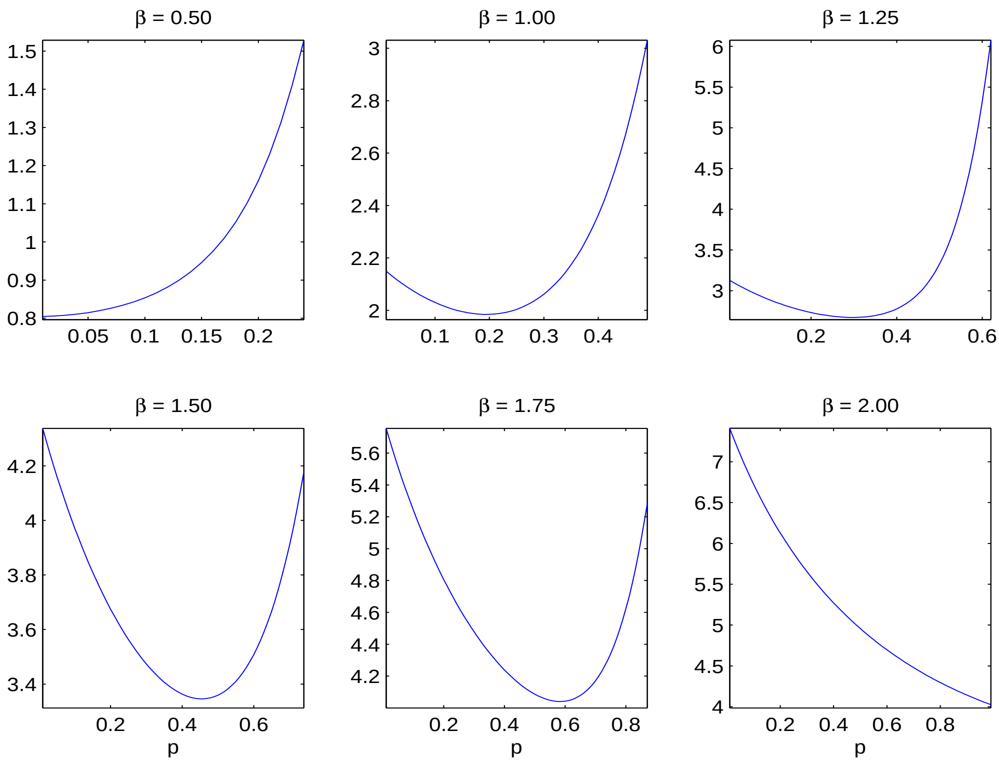

Figure 2: Asymptotic standard deviation of $\widehat{\beta}$, given in (16), as a function of p when X_t is Lévy process and $A_- = A_+$. The limiting case $\beta = 2$ corresponds to the case when L_t is a Brownian motion and the asymptotic distribution of $\widehat{\beta}$ in this case continues to be given by the right-hand side of (17).

The estimator of the activity β proposed here, as well as the original one based on the realized power variation $V_n(p, X)$, make use of the self-similarity of the (strictly) stable process and is based on the p-th absolute power of the stable distribution. One can of course consider other activity estimators based on different moments of the stable distribution. Examples include the logarithmic moments or the tail moments of the stable distribution (which are actually known in closed form). In analogy with the very different asymptotic standard deviation of our estimator $\widehat{\beta}$ based on different powers, evident from Figure 2, we expect these alternative estimators to differ in terms of efficiency and robustness towards presence of additional terms in the price process X_t (in addition to the leading stable component of L_t) like the drift, residual jump components (controlled here by the parameter β') or even a presence of a diffusion component. In any case, however, the approach proposed here of using second-order differences of X_t can be readily adopted to those alternative estimators as well and should help of removing a bias due to the presence of the drift term, which as we saw earlier can in certain cases slows down the rate of convergence of the estimator (or even lead to a limit that is determined by the drift as is the case for the estimator based on the power variation $V_n(p, X)$ when $\beta < 1$).

We next state the asymptotic behavior of the activity estimator $\widehat{\beta}$ in the more general case when α_t and σ_t can vary.

Corollary 2 *Assume that X_t satisfies assumptions A and B. For $p \in (0, \beta)$ we have $\widehat{\beta} \xrightarrow{\mathbb{P}} \beta$. If in addition $p \in \left(\frac{|\beta-1|}{2(\beta\wedge 1)} \bigvee \frac{\beta\beta'}{2(\beta-\beta')}, \frac{\beta}{2}\right)$, and provided $\beta' < \beta/2$ as well as $\beta > \frac{2}{3}$, we have*

$$\sqrt{n}\left(\widehat{\beta} - \beta\right) \xrightarrow{\mathcal{L}-s} \frac{\sqrt{\int_0^1 |\sigma_s|^{2p} ds}}{\int_0^1 |\sigma_s|^p ds} \frac{\beta^2}{\widetilde{\mu}_p(\beta) p \log(2)} \sqrt{\widetilde{\Xi}}\, Z, \tag{18}$$

where Z is univariate standard normal defined on an extension of the original probability space and independent of the σ-algebra $\mathcal{F}$ and $\widetilde{\Xi}$ is defined in Corollary 1.

Corollary 2 can be made feasible by using consistent estimators of $\int_0^1 |\sigma_s|^p ds$ and $\int_0^1 |\sigma_s|^{2p} ds$. We conclude this section by proposing a feasible estimator of the integrated power variation of the scale in the realistic case when β is not known and has to be inferred from the data first. The result is given in the following theorem.

Theorem 3 *Assume that X_t satisfies assumptions A and B.*

(a) For $p \in (0, \beta)$ we have

$$\Delta_n^{1-p/\widehat{\beta}} \widetilde{V}_n^1(p, X) \xrightarrow{\mathbb{P}} \widetilde{\mu}_p(\beta) \int_0^1 |\sigma_s|^p ds. \tag{19}$$

(b) For $p \in \left(\frac{|\beta-1|}{2(\beta\wedge 1)} \bigvee \frac{\beta\beta'}{2(\beta-\beta')}, \frac{\beta}{2}\right)$, and provided $\beta' < \beta/2$ as well as $\beta > \frac{2}{3}$, we have

$$\begin{pmatrix} \sqrt{n}\left(\Delta_n^{1-p/\widehat{\beta}} \widetilde{V}_n^1(p, X) - \widetilde{\mu}_p(\beta) \int_0^1 |\sigma_s|^p ds\right) - \frac{p}{\beta^2} \widetilde{\mu}_p(\beta) \int_0^1 |\sigma_s|^p ds \log(\Delta_n) \sqrt{n}(\widehat{\beta} - \beta) \\ \sqrt{n}(\widehat{\beta} - \beta) \end{pmatrix}$$

$$\xrightarrow{\mathcal{L}-s} \begin{pmatrix} \int_0^1 |\sigma_s|^{2p} ds \Xi^{(1,1)} & \frac{\int_0^1 |\sigma_s|^{2p} ds}{\int_0^1 |\sigma_s|^p ds} \frac{\beta^2}{\widetilde{\mu}_p(\beta) p \log(2)} (\Xi^{(1,1)} - 2^{-p/\beta} \Xi^{(1,2)}) \\ \frac{\int_0^1 |\sigma_s|^{2p} ds}{\int_0^1 |\sigma_s|^p ds} \frac{\beta^2}{\widetilde{\mu}_p(\beta) p \log(2)} (\Xi^{(1,1)} - 2^{-p/\beta} \Xi^{(1,2)}) & \frac{\int_0^1 |\sigma_s|^{2p} ds}{(\int_0^1 |\sigma_s|^p ds)^2} \left(\frac{\beta^2}{\widetilde{\mu}_p(\beta) p \log(2)}\right)^2 \widetilde{\Xi} \end{pmatrix}^{1/2} \mathbf{Z}, \tag{20}$$

where $\mathbf{Z}$ is two-dimensional standard normal random variable defined on an extension of the original probability space and independent from $\mathcal{F}$ and $\widetilde{\Xi}$ is defined in Corollary 1.

Comparing Theorem 3 and Theorem 2, we can see that the estimation of β has a first-order effect on the inference for the integrated power variation of the stochastic scale. That is the rate of convergence of $\Delta_n^{1-p/\widehat{\beta}}\widetilde{V}_n^1(p,X)$ is driven by that of $\widehat{\beta}$. The above theorem gives higher-order asymptotic expansion which contains also the faster converging component due to $\Delta_n^{1-p/\beta}\widetilde{V}_n^1(p,X)$. The difference between that component and the slower converging piece due to the estimation of β is only $\log(\Delta_n)$, so in a practical application it is advisable to incorporate the asymptotic effect of the former as well.

To this end, from part (b) of the above theorem, we easily get

$$\frac{\sqrt{n}\left(\Delta_n^{1-p/\widehat{\beta}}\widetilde{V}_n^1(p,X) - \widetilde{\mu}_p(\beta)\int_0^1|\sigma_s|^p ds\right)}{\sqrt{\left(1+\frac{\log(\Delta_n)}{\log(2)}\right)^2\Xi^{(1,1)} + \frac{\log^2(\Delta_n)}{\log^2(2)4^{p/\beta}}\Xi^{(2,2)} - \left(1+\frac{\log(\Delta_n)}{\log(2)}\right)\frac{2^{1-p/\beta}\log(\Delta_n)}{\log(2)}\Xi^{(1,2)}}\sqrt{\int_0^1|\sigma_s|^{2p}ds}} \xrightarrow{\mathcal{L}-s} Z, \tag{21}$$

where Z is standard normal variable defined on an extension of the original probability space. Also, it is straightforward to replace $\int_0^1|\sigma_s|^{2p}ds$ in the above with a consistent estimate using part(a) of Theorem 3, which allows for feasible inference regarding the integrated power variation of the stochastic scale.

5 Proofs

In the proofs we use the shorthand notation $\mathbb{E}_i^n(\cdot) \equiv \mathbb{E}(\cdot|\mathcal{F}_{i\Delta_n})$ and $\mathbb{P}_i^n(\cdot) \equiv \mathbb{P}(\cdot|\mathcal{F}_{i\Delta_n})$. We also denote with K a positive constant that does not depend on n and might change from line to line in the inequalities that follow. We start with establishing some preliminary results and then proceed with the proofs of the theorems and corollaries in the paper.

5.1 Preliminary results

5.1.1 Localization

Throughout we prove results under the stronger assumption:

Assumption SB. *We have assumption B and in addition*

(a) *the processes $|\sigma_t|$ and $|\sigma_t|^{-1}$ are uniformly bounded;*

(b) *the processes b_t^α and b_t^σ are uniformly bounded;*

(c) $|\delta^\alpha(t,x)| + |\delta^\sigma(t,x)| \leq \gamma(x)$ *for all t, where $\gamma(x)$ is a deterministic bounded function on $\mathbb{R}$ with $\int_{\mathbb{R}}|\gamma(x)|^{\beta+\iota}dx < \infty$ for $\iota \in (\beta,2)$ and β being the constant in (7);*

(d) *the coefficients in the Itô semimartingale representation of b_t^α and b_t^σ satisfy the analogues of conditions (b) and (c) above;*

(e) *the process $\int_{\mathbb{R}}(|x|^{\beta'+\iota}\wedge 1)\nu_t^Y(dx)$ is bounded and the jumps of L and Y are bounded.*

Once we establish the results of the paper under the stronger assumption SB, extending them to the case of the weaker assumption B follows by a standard localization argument, see e.g., Section 4.4.1 of Jacod and Protter (2012).

5.1.2 Jump Representation

In what follows it is convenient to extend appropriately the probability space and then decompose the driving Lévy process L_t as follows

$$L_t + \widehat{S}_t = S_t + \widetilde{S}_t, \tag{22}$$

where S_t, $\widehat{S}_t$ and $\widetilde{S}_t$ are pure-jump Lévy processes with first two characteristics zero (with respect to the truncation function $\kappa(\cdot)$) and Lévy densities $\frac{A_-}{|x|^{1+\beta}}1_{\{x<0\}} + \frac{A_+}{|x|^{1+\beta}}1_{\{x>0\}}$, $2|\nu'(x)|1_{\{\nu'(x)<0\}}$, and $|\nu'(x)|$ respectively. We denote the associated counting jump measures with μ, μ_1 and μ_2 (note that there can be dependence between μ, μ_1 and μ_2).

S_t is β-stable process and $\widehat{S}_t$ and $\widetilde{S}_t$ are "residual" components whose effect on our statistic, as will be shown, is negligible (under suitable conditions). The proof of the decomposition in (22) as well as the explicit construction of S_t, $\widehat{S}_t$ and $\widetilde{S}_t$ can be found in Section 1 of the supplementary appendix of Todorov and Tauchen (2012).

5.2 Proof of Theorems 1 and 2

We first decompose

$$\begin{pmatrix} \Delta_n^{1-p/\beta}\widetilde{V}_n^1(p,X) - \widetilde{\mu}_p(\beta)\int_0^1|\sigma_s|^p ds \\ \Delta_n^{1-p/\beta}\widetilde{V}_n^2(p,X) - 2^{p/\beta}\widetilde{\mu}_p(\beta)\int_0^1|\sigma_s|^p ds \end{pmatrix} = A_1 + A_2 + A_3 + \tilde{A}, \tag{23}$$

$$A_1 = \Delta_n \sum_{i=4}^n |\sigma_{(i-4)\Delta_n-}|^p \begin{pmatrix} \Delta_n^{-p/\beta}|\Delta_i^n S - \Delta_{i-1}^n S|^p - \widetilde{\mu}_p(\beta) \\ \Delta_n^{-p/\beta}|\Delta_i^n S - \Delta_{i-1}^n S + \Delta_{i-2}^n S - \Delta_{i-3}^n S|^p - 2^{p/\beta}\widetilde{\mu}_p(\beta) \end{pmatrix},$$

$$A_2 = \sum_{i=4}^n \left(\Delta_n|\sigma_{(i-4)\Delta_n-}|^p - \int_{(i-1)\Delta_n}^{i\Delta_n}|\sigma_s|^p ds\right) \otimes \begin{pmatrix} \widetilde{\mu}_p(\beta) \\ 2^{p/\beta}\widetilde{\mu}_p(\beta) \end{pmatrix},$$

$$A_3 = \Delta_n^{1-p/\beta}\sum_{i=4}^n \begin{pmatrix} |\Delta_i^n X - \Delta_{i-1}^n X|^p - |\sigma_{(i-4)\Delta_n-}|^p|\Delta_i^n S - \Delta_{i-1}^n S|^p \\ |\Delta_i^n X - \Delta_{i-1}^n X + \Delta_{i-2}^n X - \Delta_{i-3}^n X|^p - |\sigma_{(i-4)\Delta_n-}|^p|\Delta_i^n S - \Delta_{i-1}^n S + \Delta_{i-2}^n S - \Delta_{i-3}^n S|^p \end{pmatrix}.$$

$$\tilde{A} = \Delta_n^{1-p/\beta}\sum_{i=2}^3 \begin{pmatrix} |\Delta_i^n X - \Delta_{i-1}^n X|^p \\ 0 \end{pmatrix}.$$

We analyze each of the terms on the right-hand side of (23) separately. We note that the term A_2 is different from zero only under the general case of Theorem 2. The case $\tilde{A}$ is easiest. We have $\mathbb{E}|\tilde{A}| \le K\Delta_n^{1-\iota}$ for $\iota > 0$ sufficiently small by an application of Hölder and Burkholder-Davis-Gundy inequalities.

5.2.1 The term A_1

We will prove

$$\sqrt{n}A_1 \stackrel{\mathcal{L}-s}{\longrightarrow} \sqrt{\int_0^1 |\sigma_s|^{2p} ds}\,\mathbf{\Xi}^{1/2}\mathbf{Z}. \tag{24}$$

First, we set

$$\chi_i^n = \Delta_n^{-p/\beta}\begin{pmatrix} |\Delta_i^n S - \Delta_{i-1}^n S|^p \\ |\Delta_i^n S - \Delta_{i-1}^n S + \Delta_{i-2}^n S - \Delta_{i-3}^n S|^p \end{pmatrix},$$

and with this notation we then denote

$$\eta_i^n = |\sigma_{(i-4)\Delta_n-}|^p \left(\chi_i^n - \mathbb{E}_{i-1}^n\chi_i^n\right) + |\sigma_{(i-3)\Delta_n-}|^p \left(\mathbb{E}_i^n\chi_{i+1}^n - \mathbb{E}_{i-1}^n\chi_{i+1}^n\right)$$
$$+ |\sigma_{(i-2)\Delta_n-}|^p \left(\mathbb{E}_i^n\chi_{i+2}^n - \mathbb{E}_{i-1}^n\chi_{i+2}^n\right) + |\sigma_{(i-1)\Delta_n-}|^p \left(\mathbb{E}_i^n\chi_{i+3}^n - \mathbb{E}_{i-1}^n\chi_{i+3}^n\right).$$

We easily have

$$A_1 - \Delta_n\sum_{i=4}^n \eta_i^n = -\Delta_n\Big\{|\sigma_{0-}|^p(\mathbb{E}_0^n\chi_4^n - \mathbb{E}_3^n\chi_4^n) + |\sigma_{\Delta_n-}|^p(\mathbb{E}_1^n\chi_5^n - \mathbb{E}_3^n\chi_5^n) + |\sigma_{2\Delta_n-}|^p(\mathbb{E}_2^n\chi_6^n - \mathbb{E}_3^n\chi_6^n)$$
$$- |\sigma_{(n-3)\Delta_n-}|^p(\mathbb{E}_{n-3}^n\chi_{n+1}^n - \mathbb{E}_n^n\chi_{n+1}^n) - |\sigma_{(n-2)\Delta_n-}|^p(\mathbb{E}_{n-2}^n\chi_{n+2}^n - \mathbb{E}_n^n\chi_{n+2}^n)$$
$$- |\sigma_{(n-1)\Delta_n-}|^p(\mathbb{E}_{n-1}^n\chi_{n+3}^n - \mathbb{E}_n^n\chi_{n+3}^n)\Big\}.$$

Using the boundedness of σ_t, as well as the self-similarity of the strictly stable process (and the fact that the p-th absolute moment of a stable random variable exists when $p < \beta$), we easily get

$$\mathbb{E}\left|\left|A_1 - \Delta_n\sum_{i=4}^n \eta_i^n\right|\right| \leq K\Delta_n \quad \Longrightarrow \quad A_1 - \Delta_n\sum_{i=4}^n \eta_i^n = o_p\left(1/\sqrt{n}\right). \tag{25}$$

Therefore, to prove (24), it suffices to look at $\frac{1}{\sqrt{n}}\sum_{i=4}^n \eta_i^n$. Using Theorem IX.7.28 of Jacod and Shiryaev (2003), we will be done if we can show

$$\begin{cases} \frac{1}{\sqrt{n}}\sum_{i=4}^n \mathbb{E}_{i-1}^n\eta_i^n \stackrel{\mathbb{P}}{\longrightarrow} \mathbf{0}, \\ \frac{1}{n}\sum_{i=4}^n \mathbb{E}_{i-1}^n\left[\left(\eta_i^n - \mathbb{E}_{i-1}^n\eta_i^n\right)\left(\eta_i^n - \mathbb{E}_{i-1}^n\eta_i^n\right)'\right] \stackrel{\mathbb{P}}{\longrightarrow} \mathbf{\Xi}\int_0^1 |\sigma_s|^{2p}ds, \\ \frac{1}{n^{1+\iota/2}}\sum_{i=4}^n \mathbb{E}_{i-1}^n||\eta_i^n||^{2+\iota} \stackrel{\mathbb{P}}{\longrightarrow} \mathbf{0}, \quad \iota \in (0, \beta/p - 2), \\ \frac{1}{n}\sum_{i=4}^n \mathbb{E}_{i-1}^n\left(\eta_i^n\Delta_i^n N\right) \stackrel{\mathbb{P}}{\longrightarrow} \mathbf{0}, \end{cases} \tag{26}$$

where N is any bounded martingale defined on the original probability space.

The first convergence result in (26) follows, since using the law of iterated expectations, we trivially have

$$\mathbb{E}_{i-1}^n\left(\eta_i^n\right) = 0.$$

We next show the third convergence result in (26). Using Jensen's inequality, the self-similarity property of the strictly stable process, and the fact that $p < \beta/2$ under parts (b) of Theorems 1 and 2, we have

$$\mathbb{E}||\eta_i^n||^{2+\iota} \leq K, \quad 0 < \iota < \beta/p - 2.$$

From here the third limit result in (26) trivially follows.

We continue with the second convergence result in (26). We denote

$$\widetilde{\eta}_i^n = |\sigma_{(i-4)\Delta_n-}|^p \times \Bigg[\left(\chi_i^n - \mathbb{E}_{i-1}^n\chi_i^n\right) + \left(\mathbb{E}_i^n\chi_{i+1}^n - \mathbb{E}_{i-1}^n\chi_{i+1}^n\right)$$
$$+ \left(\mathbb{E}_i^n\chi_{i+2}^n - \mathbb{E}_{i-1}^n\chi_{i+2}^n\right) + \left(\mathbb{E}_i^n\chi_{i+3}^n - \mathbb{E}_{i-1}^n\chi_{i+3}^n\right)\Bigg].$$

With this notation, it is clear that the second convergence result in (26) will be proved if we can establish the following three results

$$\begin{cases} \frac{1}{n}\sum_{i=4}^{n}\mathbb{E}_{i-1}^{n}\left(\eta_i^n(\eta_i^n)' - \widetilde{\eta}_i^n(\widetilde{\eta}_i^n)'\right) \stackrel{\mathbb{P}}{\longrightarrow} \mathbf{0}, \\ \frac{1}{n}\sum_{i=4}^{n}\left(\mathbb{E}_{i-1}^{n}(\widetilde{\eta}_i^n(\widetilde{\eta}_i^n)') - |\sigma_{(i-4)\Delta_n-}|^{2p}\mathbf{\Xi}\right) \stackrel{\mathbb{P}}{\longrightarrow} \mathbf{0}, \\ \frac{1}{n}\sum_{i=4}^{n}|\sigma_{(i-4)\Delta_n-}|^{2p} \stackrel{\mathbb{P}}{\longrightarrow} \int_0^1|\sigma_s|^{2p}ds. \end{cases} \tag{27}$$

The first convergence result in (27) follows by application of the algebraic inequality $||a+b|^p - |a|^p| \leq |b|^p$ for $a,\ b\in\mathbb{R}$ and $p\in(0,1)$, as well as the following result

$$\mathbb{E}|\sigma_t - \sigma_s|^p \leq K|t-s|^{\frac{p}{\beta}\wedge 1 - \iota 1_{\{p\leq\beta\}}},\ s,\ t>0,\quad \forall\iota>0. \tag{28}$$

The above follows by an application of Hölder inequality, an application of the Burkholder-Davis-Gundy inequality, the restriction on the jumps in σ_t in assumption SB, as well as the algebraic inequality $|\sum_i a_i|^p \leq \sum_i |a_i|^p$ for $p\in(0,1)$.

The second result in (27) follows trivially from the definition of $\mathbf{\Xi}$ and the boundedness of σ_t. The third result in (27) follows by Riemann integrability. This proves (27) and hence the second result in (26).

Finally, the last result in (26) can be proved exactly as Lemma 6.1 of Todorov and Tauchen (2011). Combining (25) and (26), we have (24).

5.2.2 The term A_2

We first use the decomposition

$$\begin{aligned}|\sigma_s|^p - |\sigma_{(i-4)\Delta_n}|^p = |\sigma_s|^p - &\left|\sigma_{(i-4)\Delta_n} + \int_{(i-4)\Delta_n}^s b_u^\sigma du + \int_{(i-4)\Delta_n}^s\int_{\mathbb{R}}\kappa(\delta^\sigma(u,x))\underline{\widetilde{\mu}}(du,dx)\right|^p \\ &+ \left|\sigma_{(i-4)\Delta_n} + \int_{(i-4)\Delta_n}^s b_u^\sigma du + \int_{(i-4)\Delta_n}^s\int_{\mathbb{R}}\kappa(\delta^\sigma(u,x))\underline{\widetilde{\mu}}(du,dx)\right|^p - |\sigma_{(i-4)\Delta_n}|^p,\end{aligned}$$

for $s\in[(i-1)\Delta_n, i\Delta_n]$.

Then, an application of the algebraic inequality $||a+b|^p - |a|^p| \leq |b|^p$ for $a,\ b\in\mathbb{R}$ and $p\in(0,1)$, the algebraic inequality $|\sum_i a_i|^p \leq \sum_i |a_i|^p$ for $p\in(0,1)$, together with the fact that the function κ' is zero around zero, we get

$$\begin{aligned}&\mathbb{E}\left||\sigma_s|^p - \left|\sigma_{(i-4)\Delta_n} + \int_{(i-4)\Delta_n}^s b_u^\sigma du + \int_{(i-4)\Delta_n}^s\int_{\mathbb{R}}\kappa(\delta^\sigma(u,x))\underline{\widetilde{\mu}}(du,dx)\right|^p\right| \\ &\quad\leq \mathbb{E}\left|\int_{(i-4)\Delta_n}^s\int_{\mathbb{R}}\kappa'(\delta^\sigma(u,x))\underline{\mu}(du,dx)\right|^p \leq K\Delta_n,\quad s\in[(i-1)\Delta_n, i\Delta_n].\end{aligned} \tag{29}$$

To proceed further, we make use of the following algebraic inequality

$$||a+b|^p - |a|^p| \leq |b|^p 1_{\{|b|>0.5|a|\}} + K|a|^{p-1}|b|,\quad a,b\in\mathbb{R},\quad p\in(0,1).$$

Applying the above inequality with $a = \sigma_{(i-4)\Delta_n}$ and $b = \int_{(i-4)\Delta_n}^s b_u^\sigma du + \int_{(i-4)\Delta_n}^s \int_{\mathbb{R}} \kappa(\delta^\sigma(u,x))\underline{\widetilde{\mu}}(du,dx)$, together with the fact that $|\sigma_{(i-4)\Delta_n}|$ is bounded both from below and above by a positive constant, we get

$$\begin{aligned}
&\mathbb{E}\left|\left|\sigma_{(i-4)\Delta_n} + \int_{(i-4)\Delta_n}^s b_u^\sigma du + \int_{(i-4)\Delta_n}^s \int_{\mathbb{R}} \kappa(\delta^\sigma(u,x))\underline{\widetilde{\mu}}(du,dx)\right|^p - |\sigma_{(i-4)\Delta_n}|^p\right| \\
&\quad \le K\mathbb{E}\left|\int_{(i-4)\Delta_n}^s b_u^\sigma du + \int_{(i-4)\Delta_n}^s \int_{\mathbb{R}} \kappa(\delta^\sigma(u,x))\underline{\widetilde{\mu}}(du,dx)\right| \le K\Delta_n^{1/(\beta+\iota)\wedge 1}, \quad \iota \in (0, 2-\beta),
\end{aligned} \tag{30}$$

for $s \in [(i-1)\Delta_n, i\Delta_n]$ and where for the last inequality we made use of the Burkholder-Davis-Gundy inequality (when $\beta \ge 1$).

Combining (29) and (30), together with the fact that the probability of σ_t jumping at fixed times is zero, we get altogether,

$$\mathbb{E}||A_2|| \le K\Delta_n^{1/(\beta+\iota)\wedge 1}, \quad \iota \in (0, 2-\beta). \tag{31}$$

5.2.3 The term A_3

Here we derive a bound for the first absolute moment of the first element of the vector A_3, the analogous moment for the second element being done in exactly the same way. First, we split

$$S_t = S_t^{(1)} + S_t^{(2)}, \quad S_t^{(1)} = \int_0^t \int_{\mathbb{R}} \kappa(x)\widetilde{\mu}(ds,dx) \text{ and } S_t^{(2)} = \int_0^t \int_{\mathbb{R}} \kappa'(x)\mu(ds,dx),$$

$$S_t^{(1,a)} = \begin{cases} \int_0^t \int_{\mathbb{R}} \kappa(x)\widetilde{\mu}(ds,dx), & \text{if } \beta \ge 1 \\ \int_0^t \int_{\mathbb{R}} \kappa(x)\mu(ds,dx), & \text{if } \beta < 1 \end{cases}, \quad S_t^{(1,b)} = S_t^{(1)} - S_t^{(1,a)}.$$

Similarly

$$\widehat{S}_t^{(a)} = \begin{cases} \int_0^t \int_{\mathbb{R}} x\widetilde{\mu}_1(ds,dx), & \text{if } \beta' \ge 1 \\ \int_0^t \int_{\mathbb{R}} x\mu_1(ds,dx), & \text{if } \beta' < 1 \end{cases}, \quad \widehat{S}_t^{(b)} = \widehat{S}_t - \widehat{S}_t^{(a)},$$

and we similarly decompose $\widetilde{S}_t = \widetilde{S}_t^{(a)} + \widetilde{S}_t^{(b)}$. Note that $\int_{\mathbb{R}} \frac{|\kappa(x)|}{|x|^{1+\beta}}dx < \infty$ when $\beta < 1$ and therefore $S_t^{(1,a)}$ and $S_t^{(1,b)}$ are well defined when $\beta < 1$. Similar argument applies for $\widehat{S}_t^{(a)}$ and $\widetilde{S}_t^{(a)}$. Note also that $\widehat{S}_t^{(a)}$ and $\widetilde{S}_t^{(a)}$ are well defined because the jumps of L are bounded (assumption SB(e)). We further set

$$\widetilde{b}_t^\iota = b_t^\iota - 1_{\{\beta<1\}} \int_0^t \int_E \kappa(\delta^\iota(s,x))ds\lambda(dx), \quad \iota = \alpha, \sigma,$$

$$\alpha_u^n = \alpha_u - \alpha_{(i-4)\Delta_n} - \widetilde{b}_{(i-4)\Delta_n}^\alpha(u-(i-4)\Delta_n), \;\; \sigma_u^n = \sigma_u - \sigma_{(i-4)\Delta_n} - \widetilde{b}_{(i-4)\Delta_n}^\sigma(u-(i-4)\Delta_n), \;\; u \ge (i-4)\Delta_n,$$

$$\psi = -1_{\{\beta<1\}}\left(\int_{x<0} \frac{A_-\kappa(x)}{|x|^{1+\beta}}dx + \int_{x>0} \frac{A_+\kappa(x)}{|x|^{1+\beta}}dx\right) - 1_{\{\beta'<1\}} \int_{\mathbb{R}} \kappa(x)\nu'(x)dx + 1_{\{\beta'\ge 1\}} \int_{\mathbb{R}} \kappa'(x)\nu'(x)dx.$$

$$\mathcal{S}_i = \left\{\omega : |\sigma_{(i-4)\Delta_n-}(\Delta_i^n S - \Delta_{i-1}^n S)| \le 2\Delta_n^2 \left|\widetilde{b}_{(i-4)\Delta_n-}^\alpha + \widetilde{b}_{(i-4)\Delta_n-}^\sigma \psi\right|\right\}.$$

With this additional notation, we decompose $\Delta_i^n X - \Delta_{i-1}^n X = \sum_{j=1}^4 \xi_i^{(j)}$ where

$$\xi_i^{(1)} = \sigma_{(i-4)\Delta_n -}(\Delta_i^n S - \Delta_{i-1}^n S), \quad \xi_i^{(2)} = \Delta_n^2 \left(\widetilde{b}^{\alpha}_{(i-4)\Delta_n -} + \widetilde{b}^{\sigma}_{(i-4)\Delta_n -}\psi\right),$$

$$\begin{aligned}
\xi_i^{(3)} =& \int_{(i-1)\Delta_n}^{i\Delta_n} (\sigma_{u-} - \sigma_{(i-4)\Delta_n -}) dS_u^{(1,a)} - \int_{(i-2)\Delta_n}^{(i-1)\Delta_n} (\sigma_{u-} - \sigma_{(i-4)\Delta_n -}) dS_u^{(1,a)} \\
&+ \int_{(i-1)\Delta_n}^{i\Delta_n} \sigma_u^n dS_u^{(1,b)} - \int_{(i-2)\Delta_n}^{(i-1)\Delta_n} \sigma_u^n dS_u^{(1,b)} + \int_{(i-1)\Delta_n}^{i\Delta_n} (\alpha_u^n - \alpha_{u-\Delta_n}^n) du \\
&+ \int_{(i-1)\Delta_n}^{i\Delta_n} \sigma_u^n d\widetilde{S}_u^{(b)} - \int_{(i-2)\Delta_n}^{(i-1)\Delta_n} \sigma_u^n d\widetilde{S}_u^{(b)} - \int_{(i-1)\Delta_n}^{i\Delta_n} \sigma_u^n d\widehat{S}_u^{(b)} + \int_{(i-2)\Delta_n}^{(i-1)\Delta_n} \sigma_u^n d\widehat{S}_u^{(b)}, \\
\xi_i^{(4)} =& \int_{(i-1)\Delta_n}^{i\Delta_n} \sigma_{u-} d\widetilde{S}_u^{(a)} - \int_{(i-2)\Delta_n}^{(i-1)\Delta_n} \sigma_{u-} d\widetilde{S}_u^{(a)} - \int_{(i-1)\Delta_n}^{i\Delta_n} \sigma_{u-} d\widehat{S}_u^{(a)} + \int_{(i-2)\Delta_n}^{(i-1)\Delta_n} \sigma_{u-} d\widehat{S}_u^{(a)} \\
&+ \int_{(i-1)\Delta_n}^{i\Delta_n} (\sigma_{u-} - \sigma_{(i-4)\Delta_n -}) dS_u^{(2)} - \int_{(i-2)\Delta_n}^{(i-1)\Delta_n} (\sigma_{u-} - \sigma_{(i-4)\Delta_n -}) dS_u^{(2)} + \Delta_i^n Y.
\end{aligned}$$

We finally set $\widetilde{\xi}_i^{(j)} = \Delta_n^{-1/\beta} \xi_i^{(j)}$ for $j = 1, ..., 4$. We then have for the first element of the vector A_3

$$A_3(1) = \Delta_n \sum_{i=4}^n \left\{ \left| \sum_{j=1}^4 \widetilde{\xi}_i^{(j)} \right|^p - \left| \widetilde{\xi}_i^{(1)} \right|^p \right\}.$$

We first can decompose the difference

$$\sum_{i=4}^n \left(|\widetilde{\xi}_i^{(1)} + \widetilde{\xi}_i^{(2)}|^p - |\widetilde{\xi}_i^{(1)}|^p \right) = B_n + R_n, \quad B_n = \sum_{i=4}^n \zeta_i, \quad \zeta_i = p|\widetilde{\xi}_i^{(1)}|^{p-1} \text{sign}(\widetilde{\xi}_i^{(1)}) \widetilde{\xi}_i^{(2)} 1_{\{|\widetilde{\xi}_i^{(1)}| \geq 2|\widetilde{\xi}_i^{(2)}|\}},$$

$$\text{with} \qquad |R_n| \leq K \sum_{i=4}^n \left(|\widetilde{\xi}_i^{(2)}|^p 1_{\{|\widetilde{\xi}_i^{(1)}| < 2|\widetilde{\xi}_i^{(2)}|\}} + |\widetilde{\xi}_i^{(1)}|^{-1+\iota} |\widetilde{\xi}_i^{(2)}|^{p+1-\iota} \right), \quad \forall \iota > 0.$$

Making use of the fact that by successive conditioning $\mathbb{E}_i^n(\zeta_{i+j}) = 0$ for $j \geq 2$, we can further write

$$\mathbb{E}\left(\frac{1}{\sqrt{n}} B_n\right)^2 = \frac{1}{n} \mathbb{E}\left(\sum_{i=4}^n \zeta_i^2 + 2 \sum_{i=5}^n \zeta_{i-1}\zeta_i \right) \leq \frac{3}{n} \mathbb{E}\left(\sum_{i=4}^n \zeta_i^2 \right),$$

$$\zeta_i^2 \leq K |\widetilde{\xi}_i^{(1)}|^{-1+\iota+(2p-1)\vee 0} |\widetilde{\xi}_i^{(2)}|^{2+(2p-1)\wedge 0 - \iota} \implies \frac{1}{\sqrt{n}} B_n = O_p\left(\Delta_n^{(2-1/\beta)(1+(p-1/2)\wedge 0 - \iota)}\right), \quad \forall \iota > 0.$$

$$\frac{1}{n}\mathbb{E}|R_n| \leq \frac{K}{n} \sum_{i=4}^n \mathbb{E}\left(|\widetilde{\xi}_i^{(1)}|^{-1+\iota} |\widetilde{\xi}_i^{(2)}|^{p+1-\iota} \right) \implies \frac{1}{n} R_n = O_p\left(\Delta_n^{(2-1/\beta)(p+1-\iota)}\right), \quad \forall \iota > 0.$$

Therefore altogether we have

$$\frac{1}{n} \sum_{i=4}^n \left(|\widetilde{\xi}_i^{(1)} + \widetilde{\xi}_i^{(2)}|^p - |\widetilde{\xi}_i^{(1)}|^p \right) = O_p\left(\Delta_n^{(2-1/\beta)(p+1-\iota)} \vee \Delta_n^{1/2+\iota'}\right), \quad \forall \iota > 0 \text{ and some } \iota' > 0. \quad (32)$$

To get a bound on $A_3(1)$, we are left with $\frac{1}{n}\sum_{i=4}^{n}\left(\left|\sum_{j=1}^{4}\widetilde{\xi}_i^{(j)}\right|^p - |\widetilde{\xi}_i^{(1)}+\widetilde{\xi}_i^{(2)}|^p\right)$. To proceed further we use the inequality

$$\left|\left|\sum_{j=1}^{4}\widetilde{\xi}_i^{(j)}\right|^p - |\widetilde{\xi}_i^{(1)}+\widetilde{\xi}_i^{(2)}|^p\right| \le |\widetilde{\xi}_i^{(4)}|^p + \left||\widetilde{\xi}_i^{(1)}+\widetilde{\xi}_i^{(2)}+\widetilde{\xi}_i^{(3)}|^p - |\widetilde{\xi}_i^{(1)}+\widetilde{\xi}_i^{(2)}|^p\right|. \qquad (33)$$

Before analyzing each of the terms on the right-hand side of the above inequality, we derive several bounds that we make use of. Using Burkholder-Davis-Gundy inequality, the algebraic one $|\sum_i a_i|^p \le \sum_i |a_i|^p$ for $p\in(0,1)$, the boundedness of $\{\sigma_t\}_{t\in[0,1]}$, as well as the bound in (28), we have

$$\begin{aligned}\mathbb{E}\left|\int_{(i-1)\Delta_n}^{i\Delta_n}(\sigma_{u-}-\sigma_{(i-4)\Delta_n-})dS_u^{(1,a)}\right|^q &\le K\mathbb{E}\left(\int_{(i-1)\Delta_n}^{i\Delta_n}\int_{\mathbb{R}}|\sigma_{u-}-\sigma_{(i-4)\Delta_n-}|^q|\kappa(x)|^q\mu(du,dx)\right)\\ &\le K\Delta_n^{1+q\wedge 1},\quad q\in(\beta,1+1_{\{\beta\ge 1\}}).\end{aligned} \qquad (34)$$

$$\begin{aligned}\mathbb{E}\left|\int_{(i-1)\Delta_n}^{i\Delta_n}\sigma_u^n dS_u^{(1,b)}\right|^q &\le K\mathbb{E}\left|\int_{(i-1)\Delta_n}^{i\Delta_n}\int_{(i-4)\Delta_n}^{u}(\widetilde{b}_s^\sigma-\widetilde{b}_{(i-4)\Delta_n}^\sigma)dsdu\right|^q\\ &\quad + K\mathbb{E}\left|\int_{(i-1)\Delta_n}^{i\Delta_n}\int_{(i-4)\Delta_n}^{u}\int_E\delta^\sigma(s,x)\underline{\mu}(ds,dx)du\right|^q,\quad q\in(\beta,1).\end{aligned} \qquad (35)$$

By changing the order of integration (note that the integral with respect to the random measure $\underline{\mu}$ is a standard Lebesgue-Stieltjes integral and one can apply Fubini's theorem), we can further write

$$\begin{aligned}\mathbb{E}\left|\int_{(i-1)\Delta_n}^{i\Delta_n}\int_{(i-4)\Delta_n}^{u}\int_E\delta^\sigma(s,x)\underline{\mu}(ds,dx)du\right|^q &\le K\Delta_n^q\mathbb{E}\left(\int_{(i-4)\Delta_n}^{(i-1)\Delta_n}\int_E|\delta^\sigma(s,x)|^q\underline{\mu}(ds,dx)\right)\\ &\quad + K\mathbb{E}\left(\int_{(i-1)\Delta_n}^{i\Delta_n}\int_E(i\Delta_n-s)^q|\delta^\sigma(s,x)|^q\underline{\mu}(ds,dx)\right)\\ &\le K\Delta_n^{1+q},\quad q\in(\beta,1),\end{aligned} \qquad (36)$$

$$\begin{aligned}\mathbb{E}\left|\int_{(i-1)\Delta_n}^{i\Delta_n}\int_{(i-4)\Delta_n}^{u}(\widetilde{b}_s^\sigma-\widetilde{b}_{(i-4)\Delta_n}^\sigma)dsdu\right|^q &\le K\Delta_n^q\mathbb{E}\left|\int_{(i-4)\Delta_n}^{(i-1)\Delta_n}(\widetilde{b}_s^\sigma-\widetilde{b}_{(i-4)\Delta_n}^\sigma)ds\right|^q\\ &\quad + K\mathbb{E}\left|\int_{(i-1)\Delta_n}^{i\Delta_n}(i\Delta_n-s)(\widetilde{b}_s^\sigma-\widetilde{b}_{(i-4)\Delta_n}^\sigma)ds\right|^q\\ &\le K\Delta_n^{3q},\quad q\in(\beta,1),\end{aligned} \qquad (37)$$

where we made use of $\mathbb{E}|\widetilde{b}_s^\sigma-\widetilde{b}_{(i-4)\Delta_n}^\sigma| \le K|s-(i-4)\Delta_n|$ which follows from the Itô semimartingale assumption for b_t^σ in assumption SB(d) (note also that in this case $\beta<1$ and further $|\delta^\sigma(t,x)|\le\gamma(x)$ uniformly in $t\in[0,1]$ from assumption SB(c)). Similarly,

$$\mathbb{E}\left\{\left|\int_{(i-1)\Delta_n}^{i\Delta_n}\sigma_u^n d\tilde{S}_u^{(1,b)}\right|^q + \left|\int_{(i-1)\Delta_n}^{i\Delta_n}\sigma_u^n d\hat{S}_u^{(1,b)}\right|^q\right\} \le K\Delta_n^{1+q\wedge 1},\quad q\in(\beta,1+1_{\{\beta\ge 1\}}). \qquad (38)$$

Next, when $\beta < 1$, upon interchanging the order of integration, we have

$$
\begin{aligned}
\mathbb{E}\left|\int_{(i-1)\Delta_n}^{i\Delta_n}(\alpha_u^n-\alpha_{u-\Delta_n}^n)du\right|^q &\leq \left(\int_{(i-1)\Delta_n}^{i\Delta_n}\int_{u-\Delta_n}^{u}\mathbb{E}|\widetilde{b}_s^\alpha-\widetilde{b}_{(i-4)\Delta_n}^\alpha|dsdu\right)^q \\
&+K\int_{(i-2)\Delta_n}^{(i-1)\Delta_n}\int_E|s-(i-2)\Delta_n|^q|\gamma(x)|^q ds\lambda(dx)+K\int_{(i-1)\Delta_n}^{i\Delta_n}\int_E|i\Delta_n-s|^q|\gamma(x)|^q ds\lambda(dx) \\
&\leq K\Delta_n^{1+q},\quad q\in(\beta,1),
\end{aligned}
\tag{39}
$$

where we made use of $\mathbb{E}|\widetilde{b}_s^\alpha-\widetilde{b}_{(i-4)\Delta_n}^\alpha|\leq K|s-(i-4)\Delta_n|$ which again follows from assumption SB(d) about the process b_t^α.

When $\beta\geq 1$, using Burkholder-Davis-Gundy inequality and assumption SB(c), we have

$$
\begin{aligned}
\mathbb{E}\left|\int_{(i-1)\Delta_n}^{i\Delta_n}(\alpha_u^n-\alpha_{u-\Delta_n}^n)du\right|^q &\leq K\mathbb{E}\left|\int_{(i-1)\Delta_n}^{i\Delta_n}\int_{u-\Delta_n}^{u}\int_E\delta^\alpha(s,x)\underline{\widetilde{\mu}}(ds,dx)du\right|^q \\
&+K\Delta_n^{2q-2}\int_{(i-1)\Delta_n}^{i\Delta_n}\int_{u-\Delta_n}^{u}\mathbb{E}\left(|\widetilde{b}_s^\alpha-\widetilde{b}_{(i-4)\Delta_n}^\alpha|^q+\left|\int_E\kappa'(\delta^\alpha(s,x))\lambda(dx)\right|^q\right)dsdu\leq K\Delta_n^{1+q},\quad q\in(\beta,2).
\end{aligned}
\tag{40}
$$

Combining the results in (34)-(40), we get altogether

$$
\mathbb{E}|\widetilde{\xi}_i^{(3)}|^q\leq K\Delta_n^{1+1\wedge q-q/\beta},\quad q\in(\beta,1+1_{\{\beta\geq 1\}}). \tag{41}
$$

We are now ready to bound the L^1 norms on the terms on the right-hand side of (33). We start with the first one. Making use of the algebraic inequality $|\sum_i a_i|^p\leq\sum_i|a_i|^p$ for $p\in(0,1)$ (note that Y, $\widehat{S}^{(a)}$ and $\widetilde{S}^{(a)}$ are all sums of jumps when $\beta'<1$), Hölder inequality and Burkholder-Davis-Gundy inequality if $\beta'\geq 1$, which is possible for parts (a) of Theorem 1 and Theorem 2, as well as assumptions A and SB(e), we get

$$
\mathbb{E}|\widetilde{\xi}_i^{(4)}|^p\leq K\Delta_n^{p/\beta'\wedge 1-p/\beta-\iota}\implies\frac{1}{n}\sum_{i=4}^n\widetilde{\xi}_i^{(4)}=O_p(\Delta_n^{p/\beta'\wedge 1-p/\beta-\iota}),\quad p<\beta\wedge 1,\quad\forall\iota>0. \tag{42}
$$

Proceeding with the second term on the right-hand side of (33), we make use of the following algebraic inequality

$$
||a+b|^p-|a|^p|\leq K\left(|a|^{p-1}|b|1_{\{|a|>\epsilon,\ |b|<0.5\epsilon\}}+|b|^p1_{\{|a|\leq\epsilon\}}+|b|^p1_{\{|b|\geq 0.5\epsilon\}}\right),\ a,b\in\mathbb{R},\ \epsilon>0,\ p\in(0,1).
$$

Applying the above inequality with $a=\widetilde{\xi}_i^{(1)}+\widetilde{\xi}_i^{(2)}$ and $b=\widetilde{\xi}_i^{(3)}$, we have

$$
\begin{aligned}
\left||\widetilde{\xi}_i^{(1)}+\widetilde{\xi}_i^{(2)}+\widetilde{\xi}_i^{(3)}|^p-|\widetilde{\xi}_i^{(1)}+\widetilde{\xi}_i^{(2)}|^p\right| &\leq K|\widetilde{\xi}_i^{(1)}+\widetilde{\xi}_i^{(2)}|^{p-1}|\widetilde{\xi}_i^{(3)}|1_{\{|\widetilde{\xi}_i^{(1)}+\widetilde{\xi}_i^{(2)}|>\epsilon,\ |\widetilde{\xi}_i^{(3)}|<0.5\epsilon\}}1_{\{\mathcal{S}_i^c\}} \\
&+K\left(|\widetilde{\xi}_i^{(3)}|^p1_{\{|\widetilde{\xi}_i^{(1)}+\widetilde{\xi}_i^{(2)}|\leq\epsilon\}}+|\widetilde{\xi}_i^{(3)}|^p1_{\{|\widetilde{\xi}_i^{(3)}|\geq 0.5\epsilon\}}\right)1_{\{\mathcal{S}_i^c\}}+K|\widetilde{\xi}_i^{(3)}|^p1_{\{\mathcal{S}_i\}}.
\end{aligned}
\tag{43}
$$

We bound the expected value of each of the terms on the right-hand side of (43). Starting with the first term, in the case of $\beta>1$, we can use the fact that on the set $\mathcal{S}_i^c$, $\frac{2}{3}|\widetilde{\xi}_i^{(1)}+\widetilde{\xi}_i^{(2)}|\leq|\widetilde{\xi}_i^{(1)}|\leq$

$2|\widetilde{\xi}_i^{(1)}+\widetilde{\xi}_i^{(2)}|$, the fact that $\widetilde{\xi}_i^{(1)}$ is symmetric stable random variable and the latter has moments of order $-1+\iota$ finite for $\iota\in(0,1+\beta)$, the Hölder inequality, and finally (41), to get

$$\begin{aligned}\mathbb{E}\left||\widetilde{\xi}_i^{(1)}+\widetilde{\xi}_i^{(2)}|^{p-1}|\widetilde{\xi}_i^{(3)}|1_{\{|\widetilde{\xi}_i^{(1)}+\widetilde{\xi}_i^{(2)}|>\epsilon,\ |\widetilde{\xi}_i^{(3)}|<0.5\epsilon\}}1_{\{\mathcal{S}_i^c\}}\right| \\ \leq K\left[\mathbb{E}\left(|\widetilde{\xi}_i^{(1)}|^{(p-1)\frac{\beta}{\beta-1}}1_{\{|\widetilde{\xi}_i^{(1)}|>\frac{2}{3}\epsilon\}}\right)\right]^{\frac{\beta-1}{\beta}}\left[\mathbb{E}\left(|\widetilde{\xi}_i^{(3)}|^{\beta}\right)\right]^{\frac{1}{\beta}} \\ \leq K\epsilon^{-(1/\beta-p)\vee 0-\iota}\Delta_n^{1/\beta-\iota},\quad \text{when } \beta>1,\end{aligned}\tag{44}$$

where the constant K above does not depend on ϵ and $\iota>0$ is any sufficiently small number. In the case when $\beta\leq 1$, we can first note that (using again the definition of the set $\mathcal{S}_i^c$)

$$\left||\widetilde{\xi}_i^{(1)}+\widetilde{\xi}_i^{(2)}|^{p-1}|\widetilde{\xi}_i^{(3)}|1_{\{|\widetilde{\xi}_i^{(1)}+\widetilde{\xi}_i^{(2)}|>\epsilon,\ |\widetilde{\xi}_i^{(3)}|<0.5\epsilon\}}1_{\{\mathcal{S}_i^c\}}\right|\leq K|\widetilde{\xi}_i^{(1)}|^{p-\beta}|\widetilde{\xi}_i^{(3)}|^{\beta}.$$

Using this inequality as well as (41), we get

$$\mathbb{E}\left||\widetilde{\xi}_i^{(1)}+\widetilde{\xi}_i^{(2)}|^{p-1}|\widetilde{\xi}_i^{(3)}|1_{\{|\widetilde{\xi}_i^{(1)}+\widetilde{\xi}_i^{(2)}|>\epsilon,\ |\widetilde{\xi}_i^{(3)}|<0.5\epsilon\}}1_{\{\mathcal{S}_i^c\}}\right|\leq K\epsilon^{p-\beta}\mathbb{E}|\widetilde{\xi}_i^{(3)}|^{\beta}\leq K\epsilon^{-(\beta-p)}\Delta_n^{\beta-\iota},\quad\text{when } \beta\leq 1,\tag{45}$$

where the constant K does not depend on ϵ and $\iota>0$ is any sufficiently small number.

Turning to the second term on the right-hand side of (43), applying Hölder inequality, using the definition of the set $\mathcal{S}_i^c$, and the bound in (41), we get

$$\mathbb{E}\left||\widetilde{\xi}_i^{(3)}|^p 1_{\{|\widetilde{\xi}_i^{(1)}+\widetilde{\xi}_i^{(2)}|\leq\epsilon\}}1_{\{\mathcal{S}_i^c\}}\right|\leq K\left[\mathbb{E}|\widetilde{\xi}_i^{(3)}|^{\beta}\right]^{p/\beta}\left[\mathbb{P}\left(|\widetilde{\xi}_i^{(1)}|\leq 2\epsilon\right)\right]^{1-p/\beta}\leq K\Delta_n^{(1\wedge\beta)\frac{p}{\beta}-\iota}\epsilon^{1-p/\beta-\iota},\tag{46}$$

where the constant K does not depend on ϵ and $\iota>0$ is any sufficiently small number.

Next, for the third term on the right-hand side of (43), making use again of (41), we have

$$\mathbb{E}\left||\widetilde{\xi}_i^{(3)}|^p 1_{\{|\widetilde{\xi}_i^{(3)}|\geq 0.5\epsilon\}}\right|\leq K\epsilon^{-(\beta-p)}\mathbb{E}|\widetilde{\xi}_i^{(3)}|^{\beta}\leq K\epsilon^{-(\beta-p)}\Delta_n^{1\wedge\beta-\iota},\tag{47}$$

where the constant K does not depend on ϵ.

Finally, for the last term on the right-hands side of (43), using the boundedness of $\{\sigma_t\}_{t\in[0,1]}$ from above and below as well as the boundedness from above of $\{\widetilde{b}_t^{\alpha}\}_{t\in[0,1]}$ and $\{\widetilde{b}_t^{\sigma}\}_{t\in[0,1]}$, we first have

$$\mathbb{P}(\mathcal{S}_i)\leq K\mathbb{E}|\Delta_n^{-1/\beta}(\Delta_i^n S-\Delta_{i-1}^n S)|^{-1+\iota}\Delta_n^{(2-1/\beta)(1-\iota)}\leq K\Delta_n^{(2-1/\beta)(1-\iota)},\quad \iota\in(0,1+\beta).\tag{48}$$

From here, using Hölder inequality and also (41), we have

$$\mathbb{E}\left||\widetilde{\xi}_i^{(3)}|^p 1_{\{\mathcal{S}_i\}}\right|\leq K\left[\mathbb{E}|\widetilde{\xi}_i^{(3)}|\right]^p[\mathbb{P}(\mathcal{S}_i)]^{1-p}\leq K\Delta_n^{(2-1/\beta)(1-\iota)},\tag{49}$$

for $\iota>0$ any sufficiently small number. Now, we can set $\epsilon=\Delta_n^x$ for $x=\frac{\beta\wedge 1}{\beta+1}$ in (44)-(47), and this together with (49), yields

$$\frac{1}{n}\sum_{i=4}^{n}\left(|\widetilde{\xi}_i^{(1)}+\widetilde{\xi}_i^{(2)}+\widetilde{\xi}_i^{(3)}|^p-|\widetilde{\xi}_i^{(1)}+\widetilde{\xi}_i^{(2)}|^p\right)=O_p(\Delta_n^{\frac{p+1}{\beta+1}(\beta\wedge 1)-\iota}+\Delta_n^{(2-1/\beta)(1-\iota)}+\Delta_n^{1/\beta-\iota}),\ \text{for }\forall\iota>0.\tag{50}$$

One can then easily show that for $p \in \left(\frac{|\beta-1|}{2(\beta\wedge 1)}, \frac{\beta}{2}\right)$ and provided $\beta > \frac{\sqrt{5}-1}{2}$, we have $\frac{p+1}{\beta+1}(\beta\wedge 1) > 1/2$.

Combining the result in (24) for A_1 as well as the bound in (42), we prove Theorem 1 (note that in the Lévy case we do not have the term A_2 as well as the components of A_3 involving $\widetilde{\xi}_i^{(1)}$, $\widetilde{\xi}_i^{(2)}$ and $\widetilde{\xi}_i^{(3)}$).

Similarly, combining the results for A_1 in (24), for A_2 in (31), and for A_3 in (32), (42) and (50), we get parts (a) and (b) of Theorem 2. □

5.3 Proof of Corollaries 1 and 2

The consistency of $\widehat{\beta}$ follows from part(a) of Theorems 1 and 2 and continuous mapping theorem. Using next Taylor series expansion and the limiting results in Theorems 1 and 2, we have

$$\begin{aligned}\sqrt{n}(\widehat{\beta}-\beta) &= \frac{\beta^2}{p\log(2)}\frac{1}{\widetilde{\mu}_p(\beta)\int_0^1|\sigma_s|^p ds}\sqrt{n}\left(\Delta_n^{1-p/\beta}\widetilde{V}_n^1(p,X) - \widetilde{\mu}_p(\beta)\int_0^1|\sigma_s|^p ds\right)\\ &- \frac{\beta^2}{p\log(2)}\frac{1}{2^{p/\beta}\widetilde{\mu}_p(\beta)\int_0^1|\sigma_s|^p ds}\sqrt{n}\left(\Delta_n^{1-p/\beta}\widetilde{V}_n^2(p,X) - 2^{p/\beta}\widetilde{\mu}_p(\beta)\int_0^1|\sigma_s|^p ds\right) + o_p(1).\end{aligned} \tag{51}$$

From here (17) and (18) readily follow from the limits in parts (b) of Theorems 1 and 2. □

5.4 Proof of Theorem 3

Using Taylor series expansion we have on a set with probability approaching one (the set on which $\widehat{\beta}$ is above zero)

$$\Delta_n^{1-p/\widehat{\beta}}\widetilde{V}_n^1(p,X) = \Delta_n^{1-p/\beta}\widetilde{V}_n^1(p,X) + \frac{p}{(\beta^*)^2}\Delta_n^{p/\beta-p/\beta^*}\Delta_n^{1-p/\beta}\widetilde{V}_n^1(p,X)\log(\Delta_n)(\widehat{\beta}-\beta), \tag{52}$$

where β^* is a value between $\widehat{\beta}$ and β. Next, since $\widehat{\beta}-\beta = o_p(\Delta_n^{\alpha})$ for some $\alpha > 0$, we have $\Delta_n^{p/\beta-p/\beta^*} \xrightarrow{\mathbb{P}} 1$. Thus, applying part(a) of Theorem 2, we get from (52) the result in (19).

Continuing with part(b) of Theorem 3, using (52) and (51) as well as the convergence result in (15), we easily get (20). □

References

Ait-Sahalia, Y. and J. Jacod (2008). Fisher's Information for Discretely Sampled Lévy Processes. *Econeomtrica 76*, 727–761.

Ait-Sahalia, Y. and J. Jacod (2009). Estimating the Degree of Activity of Jumps in High Frequency Financial Data. *Annals of Statistics 37*, 2202–2244.

Barndorff-Nielsen, O. and N. Shephard (2001). Non-Gaussian Ornstein–Uhlenbeck-Based Models and some of Their Uses in Financial Fconomics. *Journal of the Royal Statistical Society Series B, 63*, 167–241.

Barndorff-Nielsen, O. and A. Shiryaev (2010). *Change of Time and Change of Measure.* World Scientific.

Diop, A., J. Jacod, and V. Todorov (2013). Central Limit Theorems for Approximate Quadratic Variations of Pure Jump Ito Semimrtingales. *Stochastic Processes and their Applications 123*, 839–886.

Jacod, J. (1979). *Calcul Stochastique et Problèmes de Martingales.* Lecture notes in Mathemtatics 714. Berlin Heidelberg New York: Springer-Verlag.

Jacod, J. (2004). The Euler Scheme for Levy Driven Stochastic Differential Equations: Limit Theorems. *Annals of Probability 32*, 1830–1872.

Jacod, J. (2008). Asymptotic Properties of Power Variations and Associated Functionals of Semimartingales. *Stochastic Processes and their Applications 118*, 517–559.

Jacod, J. and P. Protter (2012). *Discretization of Processes.* Berlin: Springer-Verlag.

Jacod, J. and A. Shiryaev (2003). *Limit Theorems For Stochastic Processes* (2nd ed.). Berlin: Springer-Verlag.

Klüppelberg, C., A. Lindner, and R. Maller (2004). A Continuous Time GARCH Process Driven by a Lévy Process: Stationarity and Second Order Behavior. *Journal of Applied Probability 41*, 601–622.

Lepingle, D. (1976). La variation d'ordre p des semi-martingales. *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete 36*, 295–316.

Todorov, V. (2011). Econometric Analysis of Jump-Driven Stochastic Volatility Models. *Journal of Econometrics 160*, 12–21.

Todorov, V. and G. Tauchen (2011). Limit Theorems for Power Variations of Pure-Jump Processes with Application to Activity Estimation. *Annals of Applied Probability 21*, 546–588.

Todorov, V. and G. Tauchen (2012). Realized Laplace Transforms for Pure-Jump Semimartingales. *Annals of Statistics 40*, 1233–1262.

Woerner, J. (2003). Variational Sums and Power Variation: A Unifying Approach to Model Selection and Estimation in Semimartingale Models. *Statistics and Decisions 21*, 47–68.

Woerner, J. (2007). Inference in Lévy-type Stochastic Volatility Models. *Advances in Applied Probability 39*, 531–549.