IDENTIFYING THE SUCCESSIVE BLUMENTHAL–GETOOR INDICES OF A DISCRETELY OBSERVED PROCESS

BY YACINE AÏT-SAHALIA[1] AND JEAN JACOD

Princeton University and Universitè Pierre and Marie Curie

This paper studies the identification of the Lévy jump measure of a discretely-sampled semimartingale. We define successive Blumenthal–Getoor indices of jump activity, and show that the leading index can always be identified, but that higher order indices are only identifiable if they are sufficiently close to the previous one, even if the path is fully observed. This result establishes a clear boundary on which aspects of the jump measure can be identified on the basis of discrete observations, and which cannot. We then propose an estimation procedure for the identifiable indices and compare the rates of convergence of these estimators with the optimal rates in a special parametric case, which we can compute explicitly.

1. Introduction. Let X be a one-dimensional semimartingale defined on a finite time interval $[0, T]$. Our objective is to make some progress toward the identification of the jump measure of X at high frequency. The motivation for what follows has its roots in a family of econometric problems, which can be stated as follows. We observe a single path of X, but not fully: although other observation schemes are possible, the most typical is one where we observe the variables $X_{i \Delta_n}$ for $i = 0, 1, \ldots, [T/\Delta_n]$, where $[x]$ denotes the integer part of the real x, over a fixed observation span T and where Δ_n is small. Asymptotic results are derived in the high-frequency limit where the sequence Δ_n going to 0. The overall objective is to find out what can be recovered, that is, identified, about the dynamics of X, in this setup where a single path, partially observed at a discrete time interval, is all that is available. For those parameters which can be identified, we also want asymptotically consistent estimators, with a rate whenever possible.

For the dynamics of X, we restrict our attention to Itô semimartingales, meaning that the characteristics (B, C, ν) of X can be written as follows:

$$
\begin{aligned}
B_t(\omega) &= \int_0^t b_s(\omega) \, ds, \\
C_t(\omega) &= \int_0^t c_s(\omega) \, ds, \qquad \nu(\omega, dt, dx) = dt \otimes F_t(\omega, dx)
\end{aligned}
$$

(1)

for some adapted processes b_t and c_t and measure $F_t(\omega, dx)$. Recall that B is the drift, C is the quadratic variation of the continuous martingale part and ν is the compensator of the jump measure μ of X [see Jacod and Shiryaev (2003) for more details on characteristics]. As is well known, these are the canonical models for arbitrage-free asset prices.

A sizeable part of the paper, however, is concerned with the much-restricted class of Lévy processes. A semimartingale X is a Lévy process if and only if (1) holds with $b_t(\omega) = b \in \mathbb{R}$ and $c_t(\omega) = c \geq 0$ and $F_t(\omega, dx) = F(dx)$ independent of ω and t. The measure F is the *Lévy measure*, and it integrates $x^2 \wedge 1$. The (deterministic) triple (b, c, F) is then the characteristic triple coming in the Lévy–Khintchine formula, providing the characteristic function of X_t,

$$(2) \qquad \mathbb{E}[e^{iuX_t}] = \exp t \left(iub - \frac{cu^2}{2} + \int \left(e^{iux} - 1 - iux 1_{\{|x| \leq 1\}} \right) F(dx) \right).$$

This completely characterizes the entire law of X.

Ultimately, we would like to identify as much as we can of the characteristics B, C and ν, and give consistent estimators for the identifiable parameters. The situation is well understood for the first two characteristics, B and C. When X is fully observed on $[0, T]$, one knows the jumps (size and location) occurring within the interval, and the quadratic variation of X on $[0, T]$, hence the function $t \mapsto C_t$ on $[0, T]$. On the other hand, and at least when C is strictly increasing (which is the case in almost all models used in practice), nothing can be said about the drift B. When the process is observed only at discrete times, C_t is no longer exactly known, but there are well established methods to estimate it in a consistent way as the observation mesh goes to 0, even in the presence of jumps.

We focus on the remaining open question, which concerns identifiability and estimation for the third characteristic, ν, or equivalently F_t, for a discretely sampled semimartingale. The measure F_t in a sense describes the law of a jump occurring at time t, conditionally on the past before t. There is a vast literature on identifying the Lévy measure when the time horizon T is asymptotically infinite, and when X is a Lévy process; see, for example, Basawa and Brockwell (1982), Figueroa-López and Houdré (2006), Nishiyama (2008), Neumann and Reiss (2009) and Comte and Genon-Catalot (2009). But over a finite time horizon T, we cannot reconstruct ν fully because there are only finitely many jumps on $[0, T]$ with size bigger than any $\varepsilon > 0$. The open question which we seek to address in this paper is: what can we and can we not identify about ν? High-frequency data analysis has proved a very fruitful area of research. As we will see, however, it is not able to achieve everything, and our objective in this paper is to pinpoint exactly the limitations, or frontier, involved in using high-frequency data over a fixed time span.

We can say something about the concentration of ν around 0. For example, we can decide for which $p \geq 0$ we have $\int_0^T ds \int F_t(\omega, dx)(|x|^p \wedge 1) < \infty$, because

outside a null set again these are exactly those p's for which $\sum_{s \leq T} |\Delta X_s(\omega)|^p < \infty$, where $\Delta X_s = X_s - X_{s-}$ is the size of the jump at time s, if any. The infimum of all such p's is a generalization of the *Blumenthal–Getoor* index (or BG index) of the process up to time T, and it is *known* when X is fully observed. Note that a priori it is random, and also increasing with T, and always with values in [0, 2]. However, in the Lévy process case, it reduces to $\inf\{p : \int F(dx)(|x|^p \wedge 1) < \infty\}$, and is nonrandom and independent of time. It was originally introduced by Blumenthal and Getoor (1961), and for a stable process the BG index is also the stability index of the process.

The interest in identifying the BG index lies in the fact that the index allows for a classification of the processes from least active to most active: processes with BG index equal to 0 are either finitely active or infinitely active but with slow, sub-polynomial, divergence of ν near 0; processes with BG index strictly positive are all infinitely active; processes with BG index less than 1 have paths of finite variation; processes with BG index greater than 1 have paths of infinite variation; and in the limit, processes with continuous paths have an "activity index" (the analog of the BG index which no longer exists) equal to 2 when the volatility is not vanishing. In other words, jumps become more and more active as the BG index increases from 0 to 2, and we can think of this generalized BG index as an *index of jump activity*.

In the case of discrete observations at times $i\Delta_n$ with Δ_n going to 0, recovering the random BG index in full generality seems out of reach, but Aït-Sahalia and Jacod (2009a) constructed estimators of the nonrandom number β that are consistent as $\Delta_n \to 0$, under the main assumption that locally near 0, we have the behavior

$$(3) \qquad F_t(\omega, [-u, u]^c) \sim \frac{a_t(\omega)}{u^\beta} \qquad \text{as } u \downarrow 0$$

(plus a few technical hypotheses), where $a_t \geq 0$ is a process: in this case, β is the—deterministic—BG index at time t, on the set $\{\int_0^t a_s \, ds > 0\}$. We call this behavior "proto-stable," since it is similar to that of a stable process but only near 0. Away from a neighborhood of 0, the jump measure is completely unrestricted. We obtained the rate of convergence and a central limit theorem for the estimators, depending upon the rate in the approximations (3). Related estimators or tests for β include Belomestny (2010), Cont and Mancini (2011) and Todorov and Tauchen (2010).

We can think of (3) as providing the leading term, near 0, of the jump measure of X. Given that this term is identifiable, but that the full measure ν is not, our aim is to examine where the boundary between what can versus what cannot be identified lies. Toward this aim, one direction to go is to view (3) as giving the first term of the expansion of the "tail" $F_t(\omega, [-u, u]^c)$ near 0, and go further by assuming a series expansion such as

$$(4) \qquad F_t(\omega, [-u, u]^c) \sim \sum_{i \geq 1} \frac{a_t^i(\omega)}{u^{\beta_i}} \qquad \text{as } u \downarrow 0$$

(the precise assumption is given in Section 2), with successive powers $\beta_1 = \beta >$ $\beta_2 > \beta_3 > \cdots$. Those β_i's will be the "successive BG indices." This series expansion can, for example, result from the superposition of processes with different BG indices, in a model consisting of a sum of such processes.

The question then becomes one of identifying the successive terms in that expansion. The main theoretical result of the paper, which is somehow surprising, is as follows: the first index β_1 is always identifiable, as we already knew, but the subsequent indices β_i which are bigger than $\beta_1/2$ are identifiable, whereas those smaller are not. An intuition for this particular value of the "identifiability boundary" is as follows: in view of (4) the estimation of the β_i's can only be based on preliminary estimations of $F_t(\omega, [-u, u]^c)$, or of an integrated (in time) version of this, for a sequence $u_n \to 0$. It turns out that, even in idealized circumstances, an estimation of $F_t(\omega, [-u_n, u_n]^c)$ or of its integrated version has a rate of convergence $u_n^{-\beta_1/2}$ (there is a central limit theorem for this), so that any term contributing to $F_t(\omega, [-u_n, u_n]^c)$ by an amount less than $u_n^{-\beta_1/2}$ is fundamentally unreachable: we can only hope to estimate a further coefficient β_i if it leads to a number of increments greater than u_n (which is of order $u_n^{-\beta_i}$) that is larger than the sampling error in the number of terms generated by the first coefficient, implying that any $\beta_i < \beta_1/2$ cannot be identified. This shows that there are limits to our ability to identify these successive terms, even in the unrealistic situation where the process is fully observed, and the behavior of ν around 0 is only partly identifiable.

When the identifiability conditions are satisfied, and when the process is observed at discrete times with mesh Δ_n, we will construct estimators of the parameters which are consistent as $\Delta_n \to 0$, and determine their rate of convergence, which we will see are slow. In the case we have only two indices $\beta_1 > \beta_2$ with $\beta_2 > \beta_1/2$, we will further compare the rates of the estimators we exhibit, which are semiparametric, to the optimal rate achievable in a corresponding parametric sub-model (the sum of two stable processes, plus a drift and a Brownian motion).

The main results of the paper are summarized in Figure 1 for the two-component situation. We already noted that β_2 can be identified only if it is bigger than $\beta_1/2$; we will also see that the rate at which β_2 can be estimated increases as β_2 gets closer to β_1, and conversely decreases as β_2 gets closer to $\beta_1/2$, in the limit dropping to 0 as β_2 approaches $\beta_1/2$, consistently with the loss of identification that occurs at that point. Beyond the two-component model, we will provide general identifiability conditions and rates of convergence for the leading and higher order BG indices.

The paper is organized as follows. We first define the successive BG indices in Section 2. In Section 3, we study the identifiability of the parameters appearing in the expansion, from a theoretical viewpoint and in the special case of Lévy processes. Then we introduce consistent estimators for those parameters which we have found to be identifiable in the Lévy case, hence proving de facto their
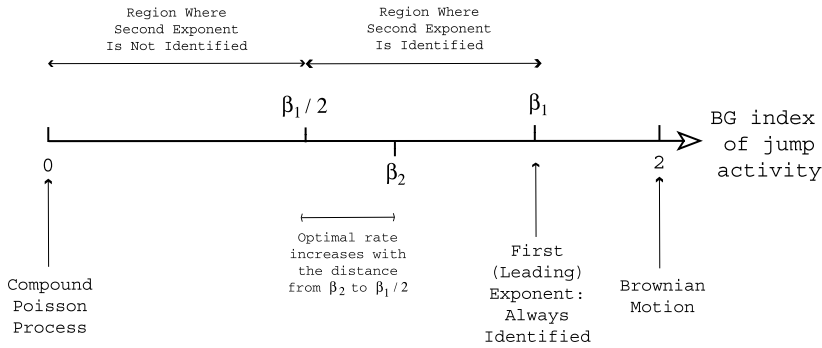
FIG. 1. *Two BG component model*: *regions where the components are identified versus not identified*, *and optimal rate of convergence*.

identifiability. This is done according to a two-step procedure, with preliminary estimators given in Section 4, and final estimators with much faster rates in Section 5. Unfortunately, although rates are given, we were not able to show a central limit theorem for these estimators, although such theorems ought to be available and would be crucial for obtaining confidence bounds.

In principle, those estimators could be used on real data, but the rates of convergence for the higher order indices are, by necessity, quite slow. We show in Section 6 that the slow nature of these rates of convergence is an inherent feature of the problem that cannot be improved upon. This is perhaps not too surprising since the range of values of the higher order indices that are identified is limited, and hence one would expect the rate of convergence to deteriorate all the way to zero as one approaches the region where identification disappears. We provide in Section 7 a simulation study for a model featuring a stochastic volatility plus two stable processes with different indices, the aim being to identify these two indices, especially the higher order one. A realistic application to high-frequency financial data, is out of the question for the typical sample sizes that are currently available, but may be useful in the future or in different fields of applications where semimartingales are used and where data are available in vast quantities, such as the study of Internet traffic or turbulence data in meteorology. The results do also present theoretical interest, especially as they set up bounds on what is asymptotically identifiable in the jump measure of a semimartingale, and consequently what is not.

2. The successive Blumenthal–Getoor indices. Throughout the paper, X is an Itô semimartingale with characteristics given by (1), on a filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t \geq 0}, \mathbb{P})$. The time horizon for the observations is $T > 0$, so the behavior of X after time T does not matter for us below.

Our first aim is to give a precise meaning to an hypothesis like (4). Instead of requiring an expansion like this for all times t, we rather use the "integrated

version" which uses the following family of (adapted, continuous and increasing) processes:

$$(5) \qquad u > 0 \quad \Rightarrow \quad \overline{A}(u)_t = \int_0^t F_s([-u, u]^c) \, ds.$$

The basic assumption is as follows:

ASSUMPTION 1. There are a *nonrandom* integer j, a strictly decreasing sequence $(\beta_i)_{1 \le i \le j+1}$ of numbers in $[0, 2)$ and a sequence $(A^i)_{1 \le i \le j+1}$ of processes such that

$$(6) \qquad t \in [0, T], \qquad 0 < u \le 1 \quad \Rightarrow \quad \left| \overline{A}(u)_t - \sum_{i=1}^j \frac{A_t^i}{u^{\beta_i}} \right| \le \frac{A_t^{j+1}}{u^{\beta_{j+1}}}.$$

Moreover, we have $A_T^i > 0$ for $i = 1, \ldots, j$.

If this assumption is satisfied with some $j \ge 2$, it is also satisfied with any smaller integer. The processes A^i and A' are nondecreasing nonnegative, and they can always be chosen to be predictable.

Clearly, $\beta = \beta_1$ is the BG index, as introduced before, and the following definition comes naturally in:

DEFINITION 1. Under Assumption 1, the numbers $\beta_1, \beta_2, \ldots, \beta_j$ are called the successive BG indices of the process X over the time interval $[0, T]$, and the variables A_T^i are called the associated integrated intensities.

EXAMPLE 1. Let $Y^1, \ldots, Y^j$ be independent stable processes with indices $\beta_1 > \cdots > \beta_j$. Then $X = Y^1 + \cdots + Y^j$ satisfies (6) with $A^{j+1} = 0$ and the successive indices and integrated intensities are β_i and $T a_i$, where $a_i = \lim_{u \to 0} u^{\beta_i} F^i([-u, u]^c)$, and F^i is the Lévy measure of Y^i.

If the Y^i's are tempered stable processes [see Rosiński (2007)] the same is true, provided $\beta_j > \beta_1 - 1$.

EXAMPLE 2. A semimartingale consisting of a continuous component and a jump part driven by a sum of such processes also satisfies (6). Let $X_t = X_0 + Z_t + \sum_{i=1}^j \int_0^t H_s^i \, dY_s^i$, with Z a continuous Itô semimartingale and Y^i as in the previous example and H^i locally bounded predictable processes with $\int_0^T |H_s^i|^{\beta_i} \, ds > 0$. The successive BG indices are again the β_i's, with the associated integrated intensities

$$A_T^i = a_i \int_0^T |H_s^i|^{\beta_i} \, ds.$$

REMARK 1. We have taken a finite family of possible indices β_i. Nothing prevents us from taking an infinite sequence: we simply have to assume that Assumption 1 holds for all j, with additionally $\lim_{i \to \infty} \beta_i = 0$. However, in view of the restriction imposed on the BG indices by our main theorems below about identifiability, this more general situation has no statistical interest.

REMARK 2. Assumption 1 imposes a certain structure on the behavior of the jump measure of the process near 0. It is important to note that it does not restrict in any way the behavior of the jump measure away from 0. Although most models used in practice and with infinite activity jumps satisfy this assumption, the Gamma process does not: although it (barely) exhibits infinite activity, its BG index is 0, and $\overline{A}(u)_t$ is of order $\log(1/u)$.

In Assumption 1, expansion (6) is central, but one may wonder about the additional requirement $A_T^i > 0$. So, we end this section with some comments and extensions, which may look complicated and are not necessary for the rest of the paper, but which we think are useful and somewhat enlightening.

EXTENSION 1. In Assumption 1 positive and negative jumps are treated in the same way. In practice, it might be useful for modeling purposes to establish the behavior of positive and negative jumps separately. Toward this end, one can replace (5) by

$$\overline{A}(u)_t^{(+)} = \int_0^t F_s((u, \infty)) \, ds, \qquad \overline{A}(u)_t^{(+)} = \int_0^t F_s((-\infty, -u)) \, ds.$$

Then, if one is interested in positive jumps only, say, one replaces (6) by a similar expansion for $\overline{A}(u)_t^{(+)}$: all the content of the paper still holds, mutatis mutandis, under this modified assumption, for positive jumps. The same is true of negative jumps, and the "positive" and "negative" successive BG indices can of course be different.

EXTENSION 2. Now we come to the requirement $A_T^i > 0$, which in Assumption 1 is supposed to hold for all (or, almost all) ω. This is of course unlikely to hold for the terminal time T, unless it holds for all $t > 0$, and even unless the processes A^i are strictly increasing. In Example 2, this amounts to suppose that none of processes H^i vanishes. However, it might be relevant in practice to allow for each H^i to vanish on some (possibly random) time intervals: we then can have different components of the model turned on and off at different times.

Thus, let us examine what happens if we relax the requirements $A_T^i > 0$. For any particular outcome ω, the (first) BG index of the process X is β_i, where i is the smallest integer such that $A_T^i > 0$, and if all of them vanish one only knows that the BG index is not bigger than β_{j+1}. The same applies to further indices. In

other words, one can define a partition of Ω indexed by all subsets D of $\{1, \ldots, j\}$ as follows:

$$(7) \qquad \Omega_T(D) = \left(\bigcap_{i \in D} \{A_T^i > 0\}\right) \cap \left(\bigcap_{i \in \{1, \ldots, j\} \setminus D} \{A_T^i = 0\}\right).$$

Then, for any ω, the successive BG indices of X over $[0, T]$ and the associated intensities are the numbers $\beta_1'(\omega), \ldots, \beta_J'(\omega)$ and $\Gamma_i(\omega)$, defined as

$$(8) \qquad \begin{aligned} J(\omega) &= m, \qquad \beta_i'(\omega) = \beta_{l_i}, \\ \Gamma_i(\omega) &= A_t^{l_i}(\omega) \qquad \text{if } \omega \in \Omega_t(\{l_1, \ldots, l_m\}). \end{aligned}$$

On the set $\Omega_T(\varnothing)$, which is not necessarily empty, we have $J = 0$ and no β_i''s.

All results of this paper are true if we relax $A_T^i > 0$ in Assumption 1, provided we replace j by J and the β_i's by the β_i''s, *in restriction to the set* $\Omega_T(D)$: this is indeed very easy, because on this set the process X coincides at all times $t \in [0, T]$ with a process X' with satisfies Assumption 1 as stated above, with $(j, \beta_1, \ldots, \beta_j, \beta_{j+1})$ substituted with $(m, \beta_{l_1}, \ldots, \beta_{j_m}, \beta_{j+1})$, when $D = \{l_1, \ldots, l_m\}$.

3. Identifiability in the Lévy case.

Loosely speaking, in an asymptotic statistical framework, identifiability of a parameter means the existence of a sequence of estimators which is (weakly) consistent. Identifiability can be "proved" by exhibiting such a sequence. It can be "disproved" by theoretical arguments, such as the fact that if the parameter is identifiable in our high-frequency observations setting, then, were the path $t \mapsto X_t$ fully observed on $[0, T]$, it would enjoy "nonasymptotic" identifiability in the sense that its value is almost surely known. For example, in the simple model $X_t = bt + W_t$ the parameter b does not enjoy this nonasymptotic property because the laws of the process X (restricted to $[0, T]$) are all equivalent when b varies, and thus b is even less identifiable in the asymptotic setting.

Disproving identifiability is usually a hard task, especially in a nonparametric setting. However, if a parameter is not identifiable for a certain class of models, it is of course not identifiable for any wider class.

These arguments lead us to consider the very special situation of a Lévy processes X, with Lévy–Khintchine characteristics (b, c, F) [see (2)] when the path $t \mapsto X_t$ is fully observed on $[0, T]$. In this section we are interested in nonasymptotic identifiability of those characteristics, or functions of them. Note that, were T infinite, the triple (b, c, F) would be identifiable because, for example, one would know the values of all the i.i.d. increments $X_{n+1} - X_n$, giving us almost surely the law of X_1, which in turn determines the triple (b, c, F).

This is no longer the case when, as in this paper, the time interval $[0, T]$ is finite. In this case, we give a formal definition of identifiability. We use $Q_{b,c,F}$ to denote the law of the process X, restricted to the interval $[0, T]$ (T is kept fixed all throughout). So $Q_{b,c,F}$ is a probability measure on the Skorokhod space $\mathbb{D} = \mathbb{D}([0, T], \mathbb{R})$. We also let $\mathcal{T}$ be some given subset of all possible triples (b, c, F).

DEFINITION 2.   A function H is *identifiable on the class* $\mathcal{T}$ if, for any two (b, c, F) and (b', c', F') in $\mathcal{T}$ such that $H(b', c', F') \neq H(b, c, F)$, we have $Q_{b,c,F} \perp Q_{b',c',F'}$ (*i.e., the two measures $Q_{b,c,F}$ and $Q_{b',c',F'}$ are mutually singular*).

The rationale behind this definition is as follows: if H is identifiable and $(b, c, F) \in \mathcal{T}$, and X is drawn according to the law $Q_{b,c,F}$, then we can discard with probability 1 any fixed $(b', c', F') \in \mathcal{T}$ such that $H(b', c', F') \neq H(b, c, F)$. Unfortunately, this does not mean that we can (almost surely) reject all (b', c', F') with $H(b', c', F') \neq H(b, c, F)$ simultaneously: this stronger property is (almost) never satisfied.

There exists a criterion for mutual singularity of $Q_{b,c,F}$ and $Q_{b',c',F'}$; see Remark IV.4.40 of Jacod and Shiryaev (2003). We have a Lebesgue decomposition $F' = f \bullet F + F'^{\perp}$ of F' with respect to F, with f a nonnegative Borel function and $F'^{\perp}$ a measure supported by an F-null set. Then $Q_{b',c',F'} \perp Q_{b,c,F}$ if and only if at least one of the following five properties is violated:

(9)
$$
\begin{cases}
F'^{\perp}(\mathbb{R}) < \infty, \\
\alpha(F, F') = \int \left(|f(x) - 1|^2 \wedge |f(x) - 1| \right) F(dx) < \infty, \\
\alpha'(F, F') = \int_{\{|x| \leq 1\}} |x| |f(x) - 1| F(dx) < \infty, \\
c = 0 \quad \Rightarrow \quad b' = b - \int_{\{|x| \leq 1\}} x(f(x) - 1) F(dx), \\
c' = c.
\end{cases}
$$

It clearly follows that the function $H(b, c, F) = c$ is identifiable on any class $\mathcal{T}$ (a well-known fact). The function $H(b, c, F) = b$ is not identifiable in general; however, on the class of all (b, c, F) having $c = 0$ and $\int_{\{|x| \leq 1\}} |x| F(dx) < \infty$ the function $H(b, c, F) = \widehat{b} = b - \int_{\{|x| \leq 1\}} x F(dx)$ (which is the "real" drift, in the sense that $X_t = \widehat{b} t + \sum_{s \leq t} \Delta X_s$) is identifiable.

In the sequel we are not interested in b or c, but in F only. That is, we are looking at functions $H = H(F)$. This leads us to consider classes of the form

(10) $\mathcal{T} = \mathbb{R} \times \mathbb{R}_+ \times \mathcal{T}_3$ where $\mathcal{T}_3$ is a set of Lévy measures.

In words, we want no restriction on the parameters b and c. Of course $\mathcal{T}_3$ should not be a singleton, and $H(F)$ should not be constant on $\mathcal{T}_3$, otherwise the identifiability problem is empty.

The following example is clear:

EXAMPLE 3.   If $\mathcal{T}_3$ is a set of measures which coincide with some given F on a neighborhood of 0, then by (9) no nontrivial $H(F)$ is identifiable on $\mathcal{T}$.

This implies that, in the best-case scenario, a function $H(F)$ can be identifiable only if it depends on the "behavior of the measure F around 0." Giving a necessary and sufficient condition for identifiability of such a function, other than saying that one of the properties in (9) fails when $H(F) \neq H(F')$, seems out of reach. However, this is possible for some specific, but relatively large, classes of sets $\mathcal{T}_3$, with a priori relatively surprising results. Below we introduce such a class, in order to illustrate the nature of the available results.

DEFINITION 3 (The class $\mathcal{T}_3^{(1)}$ of Lévy measures).   We say that a Lévy measure F belongs to this class if we have

(11)
$$
\begin{cases}
F(dx) = \widetilde{F}(dx) + \displaystyle\sum_{i=1}^{\infty} \frac{a_i \beta_i}{|x|^{1+\beta_i}} 1_{[-\eta,\eta]}(x)\, dx, & \text{where } \eta > 0 \text{ and} \\
\quad \text{(i)} \quad 0 \le \beta_{i+1} \le \beta_i < 2, \qquad \beta_i > 0 \quad \Rightarrow \quad \beta_i > \beta_{i+1}, \\
\qquad\quad \lim_{i\to\infty} \beta_i = 0, \\
\quad \text{(ii)} \quad a_i > 0 \quad \Leftrightarrow \quad \beta_i > 0, \\
\quad \text{(iii)} \quad 0 < \displaystyle\sum_{i=1}^{\infty} a_i < \infty, \\
\quad \text{(iv)} \quad \widetilde{F} \text{ is a finite measure supported by } [-\eta,,\eta]^c.
\end{cases}
$$

Parts (i) and (ii) together ensure the uniqueness of the numbers (a_i, β_i) in the representation of F, whereas if this representation holds for some $\eta > 0$, it also holds for all $\eta' \in (0, \eta)$, with the same (a_i, β_i). Part (iii) ensures that the infinite sum in the representation converges, without being zero (so equivalently, $a_1 > 0$, or $\beta_1 > 0$).

The class $\mathcal{T}_3^{(1)}$ contains all sums of symmetric stable Lévy measures. On the other hand, it is contained in the class of all Lévy measures F of a Lévy process satisfying Assumption 1: the latter is the class $\mathcal{T}_3^{(2)}$ of all F such that

(12)
$$
u \le 1 \quad \Rightarrow \quad \left| F([-u,u]^c) - \sum_{i=1}^{j} \frac{a_i}{u^{\beta_i}} \right| \le \frac{a'}{u^{\beta_{j+1}}}
$$

for $2 > \beta_1 > \cdots > \beta_{j+1} \ge 0$ and $a_i > 0$ for $i = 1, \ldots, j$ and $a' \ge 0$, and those conditions are implied by (11), for any $j \le \sup(i : \beta_i > 0)$, with the same β_i and a_i.

Considering a_i and β_i as functions on $\mathcal{T}_3^{(1)}$, the identifiability result goes as follows:

THEOREM 1.   *In the previous setting, the following holds*:

(i) *The functions β_1 and a_1 are identifiable on the set $\mathcal{T}_3^{(1)}$.*

(ii) *For any given $i \geq 2$, the functions β_i and a_i are identifiable on the subset $\mathcal{T}_3^{(1)}(i) = \{F \in \mathcal{T}_3^{(1)} : \beta_i(F) \geq \beta_1(F)/2\}$ of $\mathcal{T}_3^{(1)}$, and they are not on the complement $\mathcal{T}_3^{(1)} \setminus \mathcal{T}_3^{(1)}(i)$.*

REMARK 3.   As mentioned in the "first extension" described in the previous section, a similar statement is true if we replace the first line of (11) by

$$F(dx) = \widetilde{F}(dx) + \sum_{j=1}^{\infty} \left(\frac{a_i^{(+)} \beta_i^{(+)}}{|x|^{1+\beta_i}} 1_{(0,\eta]}(x) + \frac{a_i^{(-)} \beta_i^{(-)}}{|x|^{1+\beta_i}} 1_{(-\eta,0)}(x) \right) dx$$

with both families $(\beta_i^{(\pm)}, a_i^{(\pm)})$ satisfying (i)–(iii). Then the theorem above holds for both these families, with exactly the same proof.

REMARK 4.   As said before, any Lévy process X whose Lévy measure F is in $\mathcal{T}_3^{(1)}$ satisfies Assumption 1, but the converse is far from being true, so, even for Lévy processes, the identifiability question is not completely solved under Assumption 1. More precisely, as the estimation results will show below, (12) implies the "positive" identifiability results [(i) and the first part of (ii) of Theorem 1] for Lévy processes, but *not* the "negative" results [second part of (ii)].

For example, consider the class $\mathcal{T}_3^{(3)}$ of all measure of the form

$$F(dx) = \frac{a_1 \beta_1}{x^{1+\beta_1}} 1_{(0,1]}(x) \, dx + G(dx) \qquad \text{with } G = a_2 \sum_{n \geq 1} \varepsilon_{1/n^{1/\beta_2}}(dx)$$

and $0 < \beta_2 < \beta_1 < 2$ and $a_1, a_2 > 0$. Any such F satisfies (33), but not (11). On $\mathcal{T}_3^{(3)}$, all four parameters $\beta_1, \beta_2, a_1, a_2$ are identifiable without the restriction $\beta_2 \geq \beta_1/2$. This is of course due to the fact that the measure G is singular, and any two measures G and G' of the same type with $(\beta_2, a_2) \neq (\beta_2', a_2')$ have a Lebesgue decomposition $G' = g \bullet G + G'^{\perp}$ with $G'^{\perp}(\mathbb{R}) = \infty$ when $\beta_2 \neq \beta_2'$ and $\alpha(G, G') = \infty$ when $\beta_2 = \beta_2'$ and $a_2 \neq a_2'$.

We emphasize again that this example is quite singular, and verify here the fairly general principle that the less regular a statistical problem is, the easier it is to solve in the sense that more parameters can be estimated, and often with faster rates.

REMARK 5.   The class $\mathcal{T}_3^{(2)}$ may be bigger than $\mathcal{T}_3^{(1)}$, but it is *very far* from containing all possible Lévy measures. Indeed, any decreasing right-continuous function f on $(0, \infty)$ with $f(x) \to 0$ as $x \to \infty$ and $f(x) \leq K/x^{\alpha}$ for $x \in (0, 1]$, for some constants $K > 0$ and $\alpha \in (0, 2)$, is the symmetrical tail $f(x) = F([-x, x]^c)$ of a Lévy measure, although of course it does not need to be equivalent to a/x^{β} as $x \to 0$ for some $\beta \in (0, 2)$ and $a > 0$: so (6) may fail even with $j = 1$.

4. Discretely observed semimartingales: Preliminary estimators. Now we turn to the more general case of semimartingales. The process X is observed at the times $i\Delta_n$ for $i = 0, 1, \ldots, [T/\Delta_n]$ (where $[x]$ denotes the integer part of the real x). We thus observe the increments

$$(13) \qquad \Delta_i^n X = X_{i\Delta_n} - X_{(i-1)\Delta_n}.$$

The BG indices describes some properties of jumps, which are not observed. However, when an increment $\Delta_i^n X$ is relatively large, say bigger than u_n with $u_n \gg \sqrt{\Delta_n}$, it is likely to be due to jumps because the drift plus the continuous martingale part have increments of order of magnitude $\sqrt{\Delta_n}$. Moreover it turns out that it is usually due to a single "large" jump of size bigger than u_n, although of course the observed value $\Delta_i^n X$ is not exactly the jump size. So one may expect the number of jumps with size bigger than $u_n = u$, over the time interval $[0, t]$, to be the following number, or be relatively close to it:

$$(14) \qquad U(u, \Delta_n)_t = \sum_{i=1}^{[t/\Delta_n]} 1_{\{\Delta_i^n X > u\}}.$$

In order for the previous statement to actually be true, we need some additional assumptions, though. Those are given in the following:

ASSUMPTION 2. The process X is an Itô semimartingale, and:

(a) The processes b_t, c_t are locally bounded.
(b) We have Assumption 1 with $A_t^i = \int_0^t a_s^i \, ds$ for $i = 1, \ldots, j + 1$, where the processes a^i are locally bounded.
(c) We have $\beta_j > \beta_1/2$.

Assumption 2(c) above may look strange, or too strong. However, in view of the identifiability results of the previous section, we cannot estimate consistently β_i if it is strictly smaller than $\beta_1/2$, and as a matter of fact, the estimators described below are consistent only if $\beta_i > \beta_1/2$. Hence, since Assumption 1 for j implies the same for all $j' < j$, (c) above is really *not* a restriction, but amounts to replacing j in this assumption by $j \wedge \sup\{i : \beta_i > \beta_1/2\}$.

Apart from (c), this assumption is satisfied in Examples 1 and 2, and also by any Lévy process satisfying Assumption 1.

The estimation procedure is a two-step procedure, and in this section we describe the first—preliminary—estimators. These estimators will be consistent, but with very slow rates of convergence. This is why, in the next subsection, we will derive final estimators which exhibit much faster (although still slow) rates.

Those preliminary estimators require the knowledge of a number $\varepsilon > 0$ which satisfies

$$(15) \qquad i = 1, \ldots, j-1 \quad \Rightarrow \quad \beta_i - \beta_{i+1} \geq \varepsilon.$$

Such an ε always exists, but here we suppose that it is known, somewhat in contradiction with the fact that the β_i are unknown. It it is obviously quite difficult to estimate properly two contiguous indices β_i and β_{i+1} when they are very close to to one another. So from a statistical viewpoint, the assumption $\beta_i - \beta_{i+1} > \varepsilon$ for some fixed $\varepsilon > 0$ is natural. Moreover, since we do not know a priori which ω is observed, this amounts to supposing that all possible values of the BG indices in the model satisfy this restriction. For models used in practice, this is not really a restriction since these models rely on at most a small number of indices that are separated from one another.

The key ingredient for constructing the estimators is the counting process defined in (14), evaluated at the terminal time T and for suitable values of u. In particular, we choose a sequence u_n satisfying

(16)
$$u_n \to 0, \qquad \Delta_n^\rho \le K u_n$$
$$\text{with } \rho < \frac{1}{2 + \beta_1} \wedge \frac{2}{\beta_1(3 + \beta_1)} \wedge \frac{4}{\beta_1(5 + 3\beta_1)}.$$

Of course $\rho > 0$ above (otherwise $u_n \to 0$ would fail). The infimum of the upper bound for ρ over all $\beta_1 < 2$ is $2/11$. Therefore, since we do not a priori know the values of β_1, whereas as we will see the rates improve when the sequence u_n becomes smaller (termwise), it is thus advisable to take $\rho = 2/11$ above.

The first-step estimation is done by induction on i. We choose $\gamma > 1$, and the estimators for β_1 and A_T^1 are

(17)
$$\widetilde{\beta}_1^n = \begin{cases} \dfrac{\log(U(u_n, \Delta_n)_T / U(\gamma u_n, \Delta_n)_T)}{\log \gamma}, & \text{if } U(\gamma u_n, \Delta_n)_T > 0, \\ -1, & \text{otherwise,} \end{cases}$$
$$\widetilde{\Gamma}_i^n = (u_n)^{\widetilde{\beta}_1^n} U(u_n, \Delta_n)_T.$$

For constructing the subsequent estimators, and with ε in (15), we set

(18)
$$u_{n,i} = u_n^{(\varepsilon/2)^{i-1}}$$

(so $u_{n,1} = u_n$). We denote by $I(k, l)$ the set of all subsets of $\{1, \dots, k\}$ having l elements. Assuming that we know $\widehat{\beta}_i^n$ and $\widehat{\Gamma}_i^n$ for $i = 1, \dots, k - 1$, for some $k \in \{2, \dots, j\}$, we set

$$x \ge 1 \quad \Rightarrow \quad U^n(k, x) = \sum_{l=0}^{k-1} (-1)^l U(x\gamma^l u_{n,k}, \Delta_n)_T \sum_{J \in I(k-1, l)} \gamma^{\sum_{i \in J} \widetilde{\beta}_i^n},$$

(19)
$$\widetilde{\beta}_k^n = \begin{cases} \dfrac{\log(U^n(k, 1)/U^n(k, \gamma))}{\log(\gamma)}, & \text{if } U^n(k, 1) > 0, U^n(k, \gamma) > 0, \\ -1, & \text{otherwise,} \end{cases}$$

$$\widetilde{\Gamma}_k^n = u_{n,k}^{\widetilde{\beta}_k^n} \left(U(u_{n,k}, \Delta_n)_T - \sum_{l=1}^{k-1} \widetilde{\Gamma}_l^n u_{n,k}^{-\widetilde{\beta}_l^n} \right).$$

Finally, in order to state the result, we need a further notation, for $i = 1, \ldots, j - 1$ (so when $j = 1$ the following is empty):

$$(20) \qquad H_i = \frac{A_T^{i+1}}{A_T^i \log \gamma} \frac{\prod_{l=1}^{i}(\gamma^{\beta_l - \beta_{i+1}} - 1)}{\prod_{l=1}^{i-1}(\gamma^{\beta_l - \beta_i} - 1)}.$$

THEOREM 2. *Under Assumption* 2 *and* (15), *for all* $i = 1, \ldots, j - 1$ *such that* $\beta_{i+1} > \beta_1/2$, *we have*

$$(21) \qquad \frac{\widetilde{\beta}_i^n - \beta_i}{u_{n,i}^{\beta_i - \beta_{i+1}}} \xrightarrow{\mathbb{P}} -H_i, \qquad \frac{\widetilde{\Gamma}_i^n - A_T^i}{u_{n,i}^{\beta_i - \beta_{i+1}} \log(1/u_{n,i})} \xrightarrow{\mathbb{P}} \Gamma_i H_i.$$

Moreover if $\eta = \beta_j - \beta_{j+1} \vee \frac{\beta_1}{2} > 0$, *the following variables are bounded in probability*:

$$(22) \qquad \frac{\widetilde{\beta}_j^n - \beta_i}{u_{n,j}^{\eta}}, \qquad \frac{\widetilde{\Gamma}_j^n - A_T^j}{u_{n,j}^{\eta} \log(1/u_{n,j})}.$$

The estimator $\widetilde{\beta}_1^n$ is exactly the estimator proposed in Aït-Sahalia and Jacod (2009a) for the leading BG index β_1. So, not only does it satisfy (21) when $j \geq 2$ or the tightness of (22) when $j = 1$, but it also enjoys a central limit theorem centered at β_1 and with rate $u_n^{\beta_1/2}$ as soon as $\beta_2 < \beta_1/2$ (this property implies $j = 1$ here). Moreover, in this case one could prove that $\widetilde{\Gamma}_1^n$ also satisfies a CLT with the rate $u_n^{\beta_1/2} \log(1/u_n)$, although we will not prove it, since the emphasis here is on the case of several BG indices.

Some remarks are in order here:

REMARK 6. It is possible for the estimator $\widetilde{\Gamma}_i^n$ to be negative, in which case we may replace it by 0, or by any other positive number. It may also happen that the sequence $\widetilde{\beta}_i^n$ is not decreasing, and we can then reorder the whole family as to obtain a decreasing family (we relabel the estimators of A_T^i accordingly, of course). All these modifications are asymptotically immaterial.

REMARK 7. As mentioned in the Extension 2 at the end of Section 2, we can relax $A_T^i > 0$ in Assumption 1. Then the above theorem is still valid, in restriction to the set $\Omega_T(\{l_1, \ldots, l_m\})$ of (7), as soon as $\beta_{l_m} > \beta_{l_1}/2$.

REMARK 8. Suppose that $j \geq 2$. The limits in (21) are pure bias, hence precluding the existence of a proper central limit theorem. Note that $H_i > 0$ if $i < j$, so the bias for $\widehat{\beta}_i^n$ and for $\widehat{\Gamma}_i^n$ are always negative and positive, respectively.

Note also that the rate of convergence for estimating β_i when $i \leq j - 1$, say, is $u_{n,i}^{\beta_i - \beta_{i+1}}$, that is $u_n^{(\beta_i - \beta_{i+1})(\varepsilon/2)^{i-1}}$. This is exceedingly small, indeed. For example,

suppose that we have three indices $\beta_1 > \beta_2 > \beta_3 > \frac{\beta_1}{2}$. Then (15) implies necessarily $\varepsilon < \frac{\beta_1}{2}$, so the best possible rate for $i = 2$ would be less than, but close to, $u_n^{(\beta_2 - \beta_3)\beta_1/4}$, upon taking ε close to $\frac{\beta_1}{2}$, which is of course impossible because we do not know β_1 to start with.

In the previous example, if we suspect that β_1 is bigger than 1, say, it becomes (perhaps) not totally unreasonable to choose $\varepsilon = 0.1$; the rates for $i = 2$ and $i = 3$ thus become $u_n^{(\beta_2 - \beta_3)/10}$ and $u_n^{(\beta_3 - \beta_1/2)/100}$. This is of course on top of the fact that, because of (16), u_n is of order of magnitude $\Delta_n^{2/11}$, by a conservative choice of ρ.

Practical considerations. Letting aside the slow convergence rates, the previous result suffers from two main drawbacks:

(1) It requires to know the number of indices to be estimated (this is implicit in Assumption 2).

(2) It requires to know a number $\varepsilon > 0$ satisfying (15).

About the first problem above, in real world one does not know the number of indices. On the other hand, if Assumption 1 holds, it seems reasonable to suppose that it holds for all j, whereas the estimation is made for those β_i which are bigger than $\beta_1/2$ only. In connection with this, we assume $\beta_i - \beta_{i+1} \geq \varepsilon$ for all $i \leq j :=$ $\sup(k : \beta_k > \beta_1/2)$, plus the property $\beta_j > \beta_1/2 + \varepsilon$. Then the aim becomes to estimate β_i and A_T^i for all $i \leq j$, with j unknown.

Since the estimation procedure is done by induction on the successive indices, one can start the induction as described above, and stop it at the first i such that $\widetilde{\beta}_i \leq \varepsilon + \widetilde{\beta}_1/2$. Asymptotically, this procedure will deliver the "correct" answer (the proof of this fact, not given below, is a simple extension of the proof of the second claim of the theorem). In practice, however, the solution to this stopping problem is not quite clear, since in particular the estimated sequence $\widetilde{\beta}_i$ is not necessarily decreasing, although it is so asymptotically.

Problem 2 above is clearly more annoying. We have to admit that, in the setting presented here, we have no theoretical solution for solving it. A possible way out would be to make the estimation with several values of ε, going downward, until the estimated differences $\widetilde{\beta}_i - \widetilde{\beta}_{i-1}$ all become significantly bigger than the chosen ε, but no mathematical result so far is available in this direction. In addition, since rates are very slow, the probability that such a difference is bigger than ε when the true values satisfy the same inequality may be not close to 1 (for finite, but even large, samples).

Nonetheless, bad as it looks, this condition is probably relatively innocuous in practice: indeed, when two successive indices are very close to each other, they are obviously very difficult to tell apart. So the problem is practically meaningful only if the indices are a small number (as 2, 3 or perhaps 4) and reasonably well separated. Hence taking $\varepsilon = 0.1$ for instance, as in Remark 5, seems to be safe enough.

5. Discretely observed semimartingales: An improved method. The observation scheme is the same as in the previous section: X is observed at the times $i\Delta_n$ smaller or equal to some fixed terminal time T.

As already mentioned, the previous estimators converge at a *very* slow rate, especially for higher order indices; see Remark 8. So, in order to implement the estimation with any kind of reasonable accuracy, it is absolutely necessary to come up with better estimators.

This is the aim of this section. Assuming Assumption 2, we also suppose that we can construct preliminary estimators, such as in the previous section. Exactly as there, we must know the number j of BG indices that are to be estimated.

The method consists in minimizing, at each stage n, a suitably chosen contrast function Φ_n. First we take an integer $L \geq 2j$ and numbers $1 = v_1 < v_2 < \cdots < v_L$. We also choose positive weights w_k (typically $w_k = 1$, but any choice is indeed possible), and we pick truncation levels u_n satisfying (16). We also let D be the set of all $(x_i, \gamma_i)_{1 \leq i \leq j}$ with $0 \leq x_j \leq x_{j-1} \leq \cdots \leq x_1 \leq 2$ and $\gamma_i \geq 0$. Then the contrast function is defined on D by

$$(23) \qquad \Phi_n(x_1, \gamma_1, \ldots, x_j, \gamma_j) = \sum_{l=1}^{L} w_l \left(U(v_l u_n, \Delta_n)_T - \sum_{i=1}^{j} \frac{\gamma_i}{(v_l u_n)^{x_i}} \right)^2,$$

where the sequence u_n satisfies (16). Then the estimation goes as follows:

STEP 1. We construct preliminary estimators $\widetilde{\beta}_i^n$ (decreasing in i) and $\widetilde{\Gamma}_i^n$ (nonnegative) for β_i and A_T^i for $i = 1, \ldots, j$, such that $(\widetilde{\beta}_i - \beta_i)/u_n^\eta$ and $(\widetilde{\Gamma}_i - A_T^i)/u_n^\eta$ go to 0 in probability for some $\eta > 0$. For example, we may choose those described in the previous section (see Remark 6): the consistency requirement is fulfilled for any $\eta < (\varepsilon/2)^j$.

STEP 2. We denote by D_n the (compact and nonempty) random subset of D defined by $D_n = \{(x_i, \gamma_i) \in D : |x_i - \widetilde{\beta}_i^n| \leq \alpha u_n^\eta, |\gamma_i - \widetilde{\Gamma}_i^n| \leq \alpha u_n^\eta, \forall i = 1, \ldots, j\}$, for some arbitrary (fixed) $\alpha > 0$. Then the final estimators $\overline{\beta}_i^n$ and $\overline{\Gamma}_i^n$ will be

$$(24) \qquad (\overline{\beta}_i^n, \overline{\Gamma}_i^n)_{1 \leq i \leq j} = \arg\min_{D_n} \Phi_n(x_1, \gamma_1, \ldots, x_j, \gamma_j).$$

THEOREM 3. *Under Assumption 2, and for all choice of $v_2, \ldots, v_L$ outside a λ_{L-1}-null set (depending on the β_i's; λ_l is the l-dimensional Lebesgue measure), the sequences*

$$(25) \qquad \frac{\overline{\beta}_i^n - \beta_i}{u_n^{\beta_i - \beta_1/2 - \mu}}, \qquad \frac{\overline{\Gamma}_i^n - \Gamma_i}{u_n^{\beta_i - \beta_1/2 - \mu}}$$

are bounded in probability for all $i = 1, \ldots, j$ and all $\mu > 0$.

The rates obtained here are much faster than in Theorem 2: we replace $u_{n,i}^{\beta_i - \beta_{i+1} \vee (\beta_1/2)}$ by $u_n^{\beta_i - \beta_1/2}$, for two reasons: the exponent $\beta_i - \beta_{i+1}$ is bigger than $\beta_i - \beta_{i+1} \vee (\beta_1/2)$, unless $i = j$; more importantly, we replace the auxiliary truncation levels $u_{n,i}$ of (18) by the original sequence u_n, which is much smaller when $i \geq 2$, and only subject to (16). We will examine in the next section how far from optimality those rates are.

REMARK 9. As stated, and as seen from the proof, we only need $L = 2j$, and choosing $L > 2j$ does not improve the asymptotic properties. However, from a practical viewpoint, it is probably wise to take L bigger than $2j$ in order to smooth out the contrast function somehow, especially for (relatively) small samples. A choice of the weights $w_l > 0$ other than $w_l = 1$, such as w_l decreasing in l, may serve to put less emphasis on the large truncation values $u_n v_l$ for which less data are effectively used.

REMARK 10. The result does not hold (or at least we could not prove it) for *all* choices of the v_l's, but only when $(v_2, \ldots, v_L)$ (recall $v_1 = 1$) does not belong to some Lebesgue-null set $G(\beta_1, \ldots, \beta_j)$. This seems a priori a serious restriction, because $(\beta_1, \ldots, \beta_j)$ is unknown. In practice, we choose a priori $(v_2, \ldots, v_L)$, so we may have bad luck, just as we may have bad luck for the outcome ω which is drawn. . . .

We may also do the estimation for a number of different choices for the weights and/or values of $L \geq 2j$ and compare or average the results. This should contribute to weaken the numerical instability inherent to minimization problems such as (24). This numerical instability is similar to the one occurring in nonlinear regression problems.

We have to state, however, that these problems, just as those stated in the "practical considerations" of the previous section, are not fully addressed in this paper, and they are probably quite difficult to overcome. Our emphasis here is more on theoretical results, and on the possibility of performing the estimation with reasonable rates (see, however, Section 7 below, to see how the problem of finding a "good" ε and doing preliminary estimation in our simulation study is skipped, without affecting the quality of the procedure in any noticeable way).

6. Optimality in a special case.

6.1. *Why the convergence rates are necessarily slow.* Intuitively, the fact that we are right at the boundary between identifiability and lack thereof suggests that we should expect the rate, as we approach the loss of identifiability boundary, to deteriorate all the way to zero. In order to quantify precisely how slow the rates of convergence for the estimators of the second (and higher) index must be, even in ideal circumstances, we study a simple parametric model of the following form.

Let W be a Brownian motion and Y^1, Y^2 be two independent standard symmetric stable processes, and set

$$X_t = bt + \sigma W_t + Y_t^1 + Y_t^2. \tag{26}$$

Each Y^i depends on two parameters, the index β_i and a scale parameter a_i, the latter being characterized by the fact that the Lévy measure of Y^i is

$$F^j(dx) = \frac{a_j \beta_j}{|x|^{1+\beta_j}} \, dx. \tag{27}$$

We have six parameters,

$$b \in \mathbb{R}, \qquad c = \sigma^2 > 0, \qquad a_1, a_2 > 0, \qquad 0 < \beta_2 < \beta_1 < 2, \tag{28}$$

among which b is not identifiable, and c, β_1, a_1 are identifiable, and (β_2, a_2) are identifiable if and only if $\beta_2 \geq \beta_1/2$. In what follows, we restrict our attention to the four parameters $\beta_1, \beta_2, a_1, a_2$.

In order to find at which rate it is possible to estimate these four parameters, when X is observed at the discrete times $(i\Delta_n : i = 0, 1, \ldots, [T/\Delta_n])$ and $\Delta_n \to 0$, we study the behavior of the Fisher information matrix. Due to the fact that X is a Lévy process, the information matrix at stage n is $[T/\Delta_n]$ times the information matrix obtained when we observe only the variable X_{Δ_n}; since the variable X_Δ admits a density $x \mapsto p(_\Delta(x|c, \beta_1, a_1, \beta_2, a_2)$ which is C^∞ in x, and also in $(c, \beta_1, a_1, \beta_2, a_2)$ on the domain defined by (28), it is no wonder that Fisher's information I_Δ for a single observation X_Δ (recall $X_0 = 0$) exists, and we can study its behavior as $\Delta \to 0$.

Only the diagonal entries are important for the various rates of convergence, so we only need to focus on the following diagonal entries of this matrix:

$$I_\Delta^{\beta_1\beta_1}, I_\Delta^{a_1 a_1}, I_\Delta^{\beta_2\beta_2}, I_\Delta^{a_2 a_2}.$$

The main result of this section follows, giving the asymptotic order of the relevant terms in Fisher's information:

THEOREM 4. *We have the following equivalences, as $\Delta \to 0$:*

$$I_\Delta^{\beta_1\beta_1} \sim \frac{a_1}{2(2-\beta_1)^{\beta_1/2} c^{\beta_1/2}} \Delta^{1-\beta_1/2} (\log(1/\Delta))^{2-\beta_1/2},$$

$$I_\Delta^{a_1 a_1} \sim \frac{2\beta_1 c_{\beta_1} a_1^{\beta_1}}{(2-\beta_1)^{\beta_1/2} \sigma^{\beta_1} a_1^2} \frac{\Delta^{1-\beta_1/2}}{(\log(1/\Delta))^{\beta_1/2}}$$

and also, provided $\beta_2 > \beta_1/2$,

$$I_\Delta^{\beta_2\beta_2} \sim \frac{a_2^2 \beta_2^2}{2a_1 \beta_1 (2\beta_2 - \beta_1)(2-\beta_1)^{\beta_2 - \beta_1/2} c^{\beta_2 - \beta_1/2}}$$
$$\times \Delta^{1-\beta_2+\beta_1/2} (\log(1/\Delta))^{2-\beta_2+\beta_1/2},$$

$$I_\Delta^{a_2 a_2} \sim \frac{2\beta_2^2}{a_1 \beta_1 (2\beta_2 - \beta_1)(2-\beta_1)^{\beta_2 - \beta_1/2} c^{\beta_2 - \beta_1/2}} \frac{\Delta^{1-\beta_2+\beta_1/2}}{(\log(1/\Delta))^{\beta_2 - \beta_1/2}}.$$

REMARK 11.   We are not concerned here with the identification and estimation of the volatility parameter c; the term I_Δ^{cc} in a simpler model has been studied in Aït-Sahalia and Jacod (2008), as well as $I_\Delta^{a_1 a_1}$ when $a_2 = 0$ (i.e., when there is only one stable process on top of the Brownian motion). The asymptotic equivalent for the term $I_\Delta^{a_1 a_1}$ of course reduces to (4.11) of that paper, with $\alpha = \beta_1$, $\beta = 2$, $\theta = a_1$, up to a change of parametrization for a_1, since here we use the parametrization (27) which corresponds to the notation of Assumption 1, which is fulfilled here.

Coming back to the original problem, we deduce that it should be possible in principle to find estimators $\widehat{\beta}_i^n$ and $\widehat{a}_i^n$ having the following properties:

$$
\begin{aligned}
&\frac{(\log(1/\Delta_n))^{1-\beta_1/4}}{\Delta_n^{\beta_1/4}}(\widehat{\beta}_1^n - \beta_1) \xrightarrow{\mathcal{L}} \mathcal{N}(0, 1/T\mathcal{I}^{\beta_1\beta_1}), \\[2mm]
&\frac{1}{\Delta_n^{\beta_1/4}(\log(1/\Delta_n))^{\beta_1/4}}(\widehat{a}_1^n - a_1) \xrightarrow{\mathcal{L}} \mathcal{N}(0, 1/T\mathcal{I}^{a_1 a_1}), \\[2mm]
&\frac{(\log(1/\Delta_n))^{1-\beta_2/2+\beta_1/4}}{\Delta_n^{\beta_2/2-\beta_1/4}}(\widehat{\beta}_2^n - \beta_2) \xrightarrow{\mathcal{L}} \mathcal{N}(0, 1/T\mathcal{I}^{\beta_2\beta_2}), \\[2mm]
&\frac{1}{\Delta_n^{\beta_2/2-\beta_1/4}(\log(1/\Delta_n))^{\beta_2/2-\beta_1/4}}(\widehat{a}_2^n - a_2) \xrightarrow{\mathcal{L}} \mathcal{N}(0, 1/T\mathcal{I}^{a_2 a_2}),
\end{aligned}
\tag{29}
$$

where $\mathcal{I}^{\beta_1\beta_1}$, $\mathcal{I}^{a_1 a_1}$, $\mathcal{I}^{\beta_2\beta_2}$ and $\mathcal{I}^{a_2 a_2}$ are the constants in front of the term involving Δ in the equivalences above, for $I_\Delta^{\beta_1\beta_1}$, $I_\Delta^{a_1 a_1}$, $I_\Delta^{\beta_2\beta_2}$ and $I_\Delta^{a_2 a_2}$, respectively. Conversely, by the Cramér–Rao lower bound, Theorem 4 also implies that it will be impossible to find consistent estimators with faster rates of convergence, or smaller asymptotic variance, that those exhibited in (29).

Note that these rates are consistent with the results of Theorem 1. The first two convergences above shows that it is always possible to estimate consistently β_1 and a_1, the third one implies consistency for β_2 only if $\beta_2 \geq \beta_1/2$, and the last one implies consistency for a_2 only if $\beta_2 > \beta_1/2$. The last statement seems contradictory with Theorem 1 when $\beta_2 = \beta_1/2$, but of course it is possible to have a (somewhat irregular) statistical model for which consistency holds even though the Fisher information does not go to infinity.

6.2. *Comparison of rates.*   Now, we can compare these optimal rates with the rates obtained in Theorem 3. Doing as such, we compare a semiparametric model with a parametric sub-model. However, a minimax rate for a given parameter in a semiparametric model cannot be faster than the rate obtained for any parametric sub-model, hence the previous results are bounds for the rates in the general model considered in this paper.

Neglecting the logarithmic terms, and considering only the estimation of β_i for $i = 1, 2$, the rates above are $\Delta_n^{\gamma_i}$, whereas in Theorem 3, and upon choosing u_n optimally [i.e., ρ as large as possible in (16)], they are $\Delta_n^{\gamma_i'}$, where

$$\gamma_i = \frac{2\beta_i - \beta_1}{4}, \qquad \gamma_i' = \begin{cases} \gamma_i \dfrac{2}{2 + \beta_1} - \varepsilon, & \text{if } \beta_1 \leq (\sqrt{97} - 1)/6 \approx 1.475, \\[2mm] \gamma_i \dfrac{8}{5\beta_1 + 3\beta_1^2} - \varepsilon, & \text{if } \beta_1 \geq (\sqrt{97} - 1)/6, \end{cases}$$

and $\varepsilon > 0$ arbitrarily small (and if $\beta_i > \beta_1/2$ when $i = 2$).

As it should be, we have $\gamma_i \leq \gamma_i'$, and if equality were holding we would conclude that our estimators achieve the minimax rate (up to $\Delta_n^{-\varepsilon}$, of course, but ε is arbitrarily small). What one can say is that the actual minimax rate lies somewhere in between these two values, and the ratio γ_i/γ_i' is a kind of (imperfect) measure of the quality of the estimators proposed in Section 5: the closest to 1, the closest to optimality. Then we can conclude the following:

(a) This ratio is the same for $j = 1, 2$, which is an a priori surprising result: the quality of our estimator for β_2, relative to the optimal estimators in the stable sub-model, is the same as for β_1.

(b) This ratio is close to 1 (near optimality) when β_1 is small, and decreases down to $4/11$ as β_1 increases up to 2. The worst value is small, but not catastrophically such, especially in light of the fact that we are considering semiparametric estimators whereas the rates are optimal in the parametric context (i.e., assuming additional structure).

7. Simulation results. We now provide some simulation evidence regarding the estimators in the case where $j = 2$; we are attempting to estimate the first two jump activity indices of the process β_1 and β_2. The data generating process is a stochastic volatility model for X_t with jumps driven by two stable processes Y^1 and Y^2, with W, Y^1, Y^2 independent below:

$$(30) \qquad\qquad dX_t = \sigma_t \, dW_t + \theta_1 \, dY_t^1 + \theta_2 \, dY_t^2$$

with $\sigma_t = v_t^{1/2}$, $dv_t = \kappa(\eta - v_t) \, dt + \gamma v_t^{1/2} \, dB_t + dJ_t$, $\mathbb{E}[dW_t \, dB_t] = \rho \, dt$, $\eta^{1/2} = 0.25$, $\gamma = 0.5$, $\kappa = 5$, $\rho = -0.5$, the volatility jump term J is a compound Poisson jump process with jumps that are uniformly distributed on $[-0.3, 0.3]$ and intensity $\lambda = 10$ and $X_0 = 1$. Recall that the second component can be identified only if $\beta_2 > \beta_1/2$. We consider the situation where $(\beta_1, \beta_2) = (1.00, 0.75)$.

Given η, each scale parameter θ_i (or equivalently A_T^i) of the stable process in simulations is calibrated to deliver different various values of the tail probability $P_i = \mathbb{P}(|\Delta Y_t^i| \geq 4\eta^{1/2}\Delta_n^{1/2})$. In the various simulations' design, we hold η fixed and consider the cases where $P_1 = 0.05$ and $P_2 = 0.005$. We sample the process X over $T = 21$ days (6.5 hours per day) every $\Delta_n = 0.01$ second. This results of course in a number of observations (nearly 5×10^7) that is unrealistically high

for most high-frequency financial data series, at least presently, but extremely large numbers of observations are needed if we are going to be able to see the component β_2 of the model "behind" the two components with indices of activity 2 (the continuous component) and β_1 (the most active jump component). Of course, much smaller datasets would be sufficient in the absence of a continuous component.

Note that in general, and besides the preliminary estimators $\widetilde{\beta}_i^n$ and $\widetilde{\Gamma}_i^n$, we need to choose the number $\alpha > 0$ coming in the definition of the set D_n. Since in practice n (or Δ_n) is given, we need to choose in fact the number αu_n^η. So in concrete situations one probably can forget about the preliminary estimators and take a domain D_n which is the set of all (x_i, γ_i) in D with $\gamma_i \leq A$ for some "reasonably chosen" A, or even $A = \infty$.

This is what we do below, by taking the estimators to be

(31)
$$(\overline{\beta}_1'^n, \overline{\beta}_2'^n, \overline{\Gamma}_1'^n, \overline{\Gamma}_2'^n)$$

$$= \underset{(x_1, \gamma_1, x_2, \gamma_2)}{\arg\min} \sum_{l=1}^{L} \left(U(v_l u_n, \Delta_n)_T - \frac{\gamma_1}{(v_l u_n)^{x_1}} - \frac{\gamma_2}{(v_l u_n)^{x_2}} \right)^2,$$

where the cutoff levels $v_l u_n$ are chosen in terms of the number α_l of the long-term standard deviation $\sqrt{\eta \Delta_n}$ over a time lag Δ_n of the continuous martingale part of the process: we take α_l to be $\{7, 10, 15, 20\}$ and multiples $\{2, 4, 6\}$ thereof (giving all together $L = 10$ distinct values). Here we know η: we could also estimate for each path the average volatility, using truncated estimators for the integrated volatility [see, e.g., Mancini (2004) and Aït-Sahalia and Jacod (2009b)].

The optimization problem (31) is a quadratic problem similar to classical nonlinear least squares minimization. In situations where the parameter space is high dimensional, the objective function can exhibit local extrema, which can make the search for the optimal solution time-consuming as many starting values must be employed to validate the solution. In the case of the application here, we are only including 4 parameters, and for this small dimension, this is not causing many difficulties. In any case, it is unlikely, given the slow rates of convergence, that one would want to go beyond the second index β_2 in practice.

The results in Figure 2 are obtained with $M = 1000$ simulations: the estimators appear to be reasonably good, but then again this is for an unrealistically large number of observations, at least from the point of view of financial applications; it is perhaps feasible in other applications, such as Internet data traffic or wind measurement.

8. Conclusions. This paper determined theoretically what the successive BG indices are and how they are identified, including the perhaps surprising theoretical bound on the identification of the successive indices as a function of the previous ones. This result clarifies the border between the aspects of the jump measure which are identifiable from those which are not on the basis of discrete observations on a finite time horizon. Beyond the leading index, the identification requires
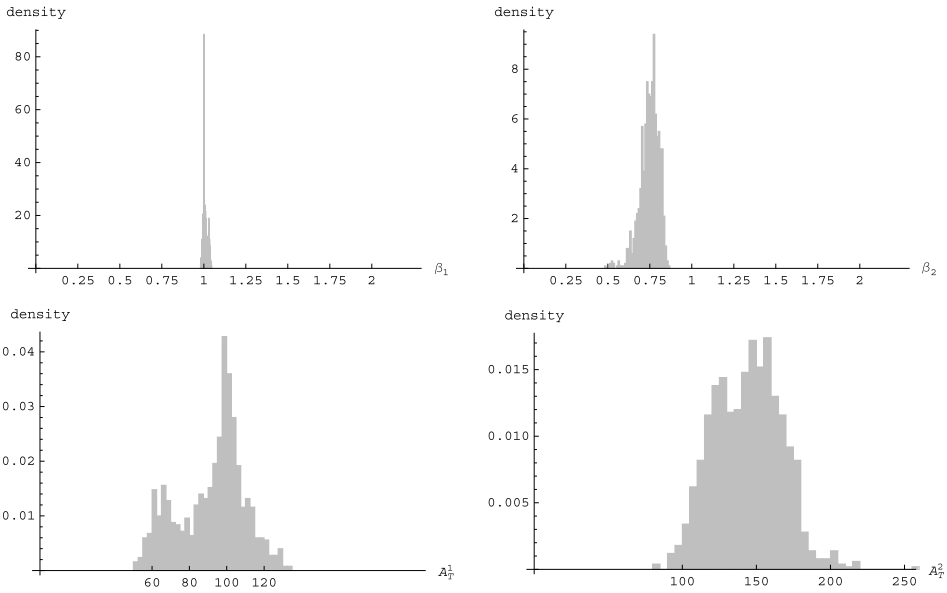
FIG. 2. *Monte Carlo simulation results*: *estimators* $\widehat{\beta}_1^n$ (*upper left graph*), $\widehat{\beta}_2^n$ (*upper right graph*), $\widehat{A}_T^{1,n}$ (*lower left graph*), $\widehat{A}_T^{2,n}$ (*lower right graph*).

in practice vast quantities of data which are out of reach of financial applications at present but may be relevant in other fields (such as the study of turbulence data, or Internet traffic). We showed through explicit calculations of Fisher's information that this limitation is a genuine, inescapable feature of the problem. There are a number of important questions that this paper does not touch upon: central limit theorems for the estimators, estimators that achieve the optimal rates of convergence, estimators that are robust to microstructure noise, estimators that are applicable with random sampling intervals, among others. The issue of the optimality of the rates in general remains an open question.

APPENDIX: PROOFS

We use the following notation throughout the Appendix. First, K denotes a constant which may change from line to line, and may depend on the characteristics or the law of the processes at hand. It never depends on n, and it is denoted as K_p if it depends on an additional parameter p. Second, for any sequence Z_n of variables and any sequence v_n of positive numbers,

$$(32) \qquad Z_n = \begin{cases} O_P(v_n), & \text{if } Z_n/v_n \text{ is bounded in probability,} \\ o_P(v_n), & \text{if } Z_n/v_n \xrightarrow{\mathbb{P}} 0. \end{cases}$$

APPENDIX A: PROOF OF THEOREM 1

(1) We fix $F \in \mathcal{T}_3^{(1)}$, with F given by (11). We also consider another $F' \in \mathcal{T}_3^{(1)}$, with F' given by (11) with β_i', a_i' and $\widetilde{F}'$. As said before, it is not a restriction to assume the representation (11) with the same $\eta > 0$ for both F and F'. Set

$$(33) \qquad\qquad j = \inf(1 \leq 1 : (\beta_i, a_i) \neq (\beta_i', a_i')).$$

The result amounts to proving the following two properties, with j as above and $b, b' \in \mathbb{R}$ and $c, c' \geq 0$:

$$(34) \quad \beta_j \geq \frac{\beta_1}{2} \quad \Rightarrow \quad Q_{b,c,F} \perp Q_{b',c',F'},$$

$$(35) \quad \beta_j < \frac{\beta_1}{2} \quad \Rightarrow \quad \begin{cases} \exists b'' \in \mathbb{R}, \exists F'' \in \mathcal{T}_3^{(1)} \\ \qquad \text{with } F'' = F' \text{ on } [-\eta, \eta) \text{ and } Q_{b,c,F} \not\perp Q_{b'',c,F'}. \end{cases}$$

These conditions being symmetrical in F and F', in both (34) and (35) we may assume

$$(36) \qquad\qquad \text{either } \beta_j > \beta_j' \quad \text{or} \quad \beta_j = \beta_j' \quad \text{and} \quad a_j > a_j'.$$

(2) In this step we assume (36). We set

$$\widehat{F}(dx) = \sum_{i \geq 1} \frac{a_i \beta_i}{|x|^{1+\beta_i}} 1_{[-\eta,\eta]}(x)\,dx, \qquad \widehat{F}'(dx) = \sum_{i \geq 1} \frac{a_i' \beta_i'}{|x|^{1+\beta_i'}} 1_{[-\eta,\eta]}(x)\,dx.$$

Then $\widehat{F}' = f \bullet \widehat{F}$, where $f = \frac{g'}{g}$ (with $\frac{0}{0} = 1$) and $g = H + G$ and $g' = H + G'$ and

$$H(x) = \sum_{i=1}^{j-1} \frac{a_i \beta_i}{|x|^{\beta_i}} 1_{[-\eta,\eta]}(x), \qquad G(x) = \sum_{i \geq j} \frac{a_i \beta_i}{|x|^{\beta_i}} 1_{[-\eta,\eta]}(x),$$

$$G'(x) = \sum_{i \geq j} \frac{a_i' \beta_i'}{|x|^{\beta_i'}} 1_{[-\eta,\eta]}(x).$$

On $[-\eta, \eta]$ we have $f - 1 = \frac{G'-G}{H+G}$ and

$$G(x) - G'(x)$$
$$= \frac{a_j \beta_j}{|x|^{1+\beta_j}} \left(1 - \frac{a_j' \beta_j'}{a_j \beta_j} |x|^{\beta_j - \beta_j'} + \sum_{i \geq j+1} \frac{a_i \beta_i}{a_j \beta_j} x^{\beta_j - \beta_i} - \sum_{i \geq j+1} \frac{a_i' \beta_i'}{a_j \beta_j} |x|^{\beta_j - \beta_i'} \right).$$

By virtue of (ii), (iii) and (iv) of (11), and of (36), we then deduce that

$$(37) \qquad x \in (-\varepsilon, \varepsilon) \quad \Rightarrow \quad \begin{cases} A_- |x|^{\beta_1 - \beta_j} \leq |f(x) - 1| \leq A_+ |x|^{\beta_1 - \beta_j}, \\ \dfrac{A_-}{|x|^{1+\beta_1}} \leq g(x) \leq \dfrac{A_+}{|x|^{1+\beta_1}}, \end{cases}$$

for three constants $A_+ > A_- > 0$ and $\varepsilon \in (0, \eta)$, depending on the two sequences (β_i, a_i) and (β_i', a_i').

(3) Now we prove (34). Since $\alpha(F, F') \geq \alpha(\widehat{F}, \widehat{F}')$, it is enough to show that $\alpha(\widehat{F}, \widehat{F}') = \infty$. By (37), $|f(x) - 1| \leq 1$ when $x \in (-\varepsilon', \varepsilon')$ for some $\varepsilon' \in (0, \varepsilon]$. Thus

$$\alpha(\widehat{F}, \widehat{F}') \geq \int_{-\varepsilon}^{\varepsilon'} |f(x) - 1|^2 g(x)\, dx \geq A_-^3 \int_{-\varepsilon'}^{\varepsilon'} |x|^{\beta_1 - 2\beta_j - 1}\, dx.$$

The last integral is infinite when $\beta_j \geq \beta_1/2$, and (35) follows by (9).

(4) Finally we prove (35). Recall that $F = \widehat{F} + \widetilde{F}$ and $F' = \widehat{F}' + \widetilde{F}'$. The measure $F'' = \widehat{F}' + \widetilde{F}$ is obviously in $\mathcal{T}_3^{(1)}$ and satisfies $F'' = f \bullet F$. Since $f(x) = 1$ outside $[-\eta, \eta]$, the quantity $\alpha'(F, F'')$ introduced in (9) is

$$\alpha'(F, F'') = \int_{-\eta}^{\eta} x(f(x) - 1)g(x)\, dx$$

$$\leq A_+^2 \int_{-\varepsilon}^{\varepsilon} |x|^{-\beta_j}\, dx + \left(\int_{\varepsilon}^{\eta} + \int_{-\eta}^{-\varepsilon}\right) |x||f(x) - 1|g(x)\, dx,$$

which is finite by (37) (because $\beta_j < \beta_1/2 < 1$) and because f and g are bounded on $[\varepsilon, \eta] \cup [-\eta, -\varepsilon]$. Therefore the number $b'' = b - \int_0^{\eta \wedge 1} x(f(x) - 1)g(x)\, dx$ is well defined. Now we consider the two triples (b, c, F) and (b'', c, F''). From what precedes they satisfy the first and the last three properties in (9). We also have by (37)

$$\alpha(F, F'') = \int_{-\eta}^{\eta} \left(|f(x) - 1|^2 \wedge |f(x) - 1|\right) g(x)\, dx$$

$$\leq A_+^2 \int_{-\varepsilon}^{\varepsilon} \left(|x|^{-\beta_j - 1} \wedge (A_+ |x|^{\beta_1 - 2\beta_j - 1})\right) dx$$

$$+ \left(\int_{\varepsilon}^{\eta} + \int_{-\eta}^{-\varepsilon}\right) \left(|f(x) - 1|^2 \wedge |f(x) - 1|\right) g(x)\, dx.$$

Since $\beta_j < \beta_1/2$ and that f and g are bounded on $[\varepsilon, \eta] \cup [-\eta, -\varepsilon]$, we deduce $\alpha(F, F'') < \infty$. So all conditions in (9) are satisfied, and we have proved (35).

APPENDIX B: COMPARING BIG JUMPS AND BIG INCREMENTS

Before starting, let us mention that for the proofs of Theorems 2 and 3 one may use a localization argument which allows us to replace Assumption 2 by the so-called "strengthened Assumption 2," which is the same except that all processes b_t, c_t, a_t^i are bounded, as well as the process A_t^{j+1} and X_t itself.

In this section we compare the number of "large" increments of X with the number of correspondingly large jumps, that is, the numbers

$$(38) \qquad V(u)_t = \sum_{s \leq t} 1_{\{|\Delta X_s| > u\}}.$$

We will indeed show that the difference $U(u_n, \Delta_n)_T - V(u_n)_T$ is negligible for our purposes, when the sequence u_n satisfies (16). The reason for doing this is that the analysis of the processes $V(u_n)$ is an easy task. Indeed, as soon as $u_n \to 0$,

$$(39) \qquad V(u_n)_T - \overline{A}(u_n)_T = O_P(u_n^{-\beta_1/2}).$$

To see this, we observe that each process $M^n = u_n^{\beta_1/2}(V(u_n) - \overline{A}(u_n))$ is a quasi-left continuous, purely discontinuous, martingale with jumps smaller than $u_n^{\beta_1/2}$, which goes to 0. Its predictable quadratic variation is $\langle M^n, M^n \rangle = u_n^{\beta_1}\overline{A}(u_n)$, which by (6) converges for each t to A_t^1. Since further A^1 is a continuous process, it follows from Theorem VI.4.13 of Jacod and Shiryaev (2003), for example, that the sequence M^n is C-tight (and even converges in law), so a fortiori, (39) holds.

The main result of this section is the next proposition:

PROPOSITION 1. *Under the strengthened Assumption 2 and if the sequence u_n satisfies* (16), *we have*

$$(40) \qquad U(u_n, \Delta_n)_T - V(u_n)_T = \frac{1}{u_n^{\beta_1}} O_P(u_n^{\beta_1 - \beta_{j+1}} + u_n^{\beta_1/2}).$$

The proof is based on a series of lemmas. The constant K may depend on an implicit way on the bounds in this strengthened assumption, but not on the two numbers $u, r \in (0, 1)$ which are fixed in most of this section.

With any càdlàg process Y and $u \in (0, 1]$, we associate the process and the variables

$$(41) \qquad Y(u)_t = \sum_{s \leq t} \Delta Y_s 1_{\{|\Delta Y_s| > u\}}, \qquad \zeta(Y, u)_i^n = 1_{\{|\Delta_i^n Y| > u\}}.$$

For simpler notation, we denote by $\mathbb{E}_{i-1}^n$ and $\mathbb{P}_{i-1}^n$, respectively, the conditional expectation and conditional probability, with respect to $\mathcal{F}_{(i-1)\Delta_n}$.

LEMMA 1. *For all $u, r \in (0, 1]$ with $u^r < 1/3$, all $w \in (0, 1/3)$ and all $k \geq 1$, we have*

$$(42) \quad \mathbb{P}_{i-1}^n(\Delta_i^n V(u) \geq k) \leq (K \Delta_n u^{-\beta_1})^k,$$

$$(43) \quad \begin{aligned} &\mathbb{P}_{i-1}^n(u(1 - w) < \Delta_i^n X(u^{1+r}) \leq u(1 + w)) \\ &\qquad \leq K(\Delta_n u^{-\beta_1} w + \Delta_n u^{-\beta_{j+1}} + \Delta_n^2 u^{-\beta_1(2+r)} + \Delta_n^3 u^{-\beta_1(3+3r)}). \end{aligned}$$

Moreover there is a $\gamma > 0$ such that, if

$$(44) \qquad \Delta_n \leq \gamma u^{\beta_1(1+r)},$$

we have for all $u \in (0, 1]$

$$(45) \quad \mathbb{E}_{i-1}^n(|\zeta(X(u^{1+r}), u)_i^n - \Delta_i^n V(u)|) \leq K(\Delta_n^2 u^{-\beta_1(2+r)} + \Delta_n^3 u^{-\beta_1(3+3r)}).$$

PROOF. If $D \subset \mathbb{R}$ the compensator of the process $N(D)_t = \sum_{s \leq t} 1_D(\Delta X_s)$ is $\int_0^t F_s(D) \, ds$. Our strengthened assumption implies the existence of a constant θ such that $F_s(D) \leq \phi(D)$, where

$$\phi(D) = \begin{cases} \theta u^{-\beta_1}, & \text{if } D \subset [-u, u]^c, \\ \theta(u^{-\beta_1} w + u^{-\beta_{j+1}}), & \text{if } D = [-u(1+w), -u) \cup (u, u(1+w)], \\ & 0 < w \leq 1. \end{cases}$$

Then for any finite stopping time S we have

$$\mathbb{E}(N(D)_{S+t} - N(D)_S \mid \mathcal{F}_S) \leq t \phi(D).$$

Let $S(D)_0 = (i-1)\Delta_n$ and $S(D)_1, S(D)_2, \ldots$ be the successive jump times of $N(D)$ after time $(i-1)\Delta_n$. What precedes implies that for $k \geq 1$ and on the set $\{S(D)_{k-1} < i\Delta_n\}$,

$$\mathbb{P}(S(D)_j \leq i\Delta_n \mid \mathcal{F}_{S(D)_{k-1}}) \leq \mathbb{E}(N(D)_{i\Delta_n} - N(D)_{S(D)_{k-1}} \mid \mathcal{F}_{S(D)_{k-1}}) \leq \Delta_n \phi(D).$$

An induction on k yields the following, which gives us the first part of (42):

(46) $$\mathbb{P}_{i-1}^n(\Delta_i^n N(D) \geq k) = \mathbb{P}_{i-1}^n(S(D)_k \leq i\Delta_n) \leq (\Delta_n \gamma(D))^k.$$

In the same way, if $D \cap D' = \varnothing$, the set $\{\Delta_i^n N(D) \geq k, \Delta_i^n N(D') \geq 1\}$ is the union for $l = 1, \ldots, k+1$ of the sets $\Gamma_l = \{S(D)_{l-1} < S(D')_1 < S(D)_l \leq i\Delta_n\}$, whereas

$$\begin{aligned}
\mathbb{P}_{i-1}^n &(S(D)_{l-1} < S(D')_1 < S(D)_l \leq i\Delta_n) \\
&= \mathbb{E}_{i-1}^n(1_{S(D)_{l-1} < S(D')_1 < i\Delta_n} \mathbb{P}(N(D)_{i\Delta_n} - N(D)_{S(D')_1} \geq k-l+1 \mid \mathcal{F}_{S(D')_1})) \\
&\leq (\Delta_n \phi(D))^{k-l+1} \mathbb{P}_{i-1}^n(S(D)_{l-1} < S(D')_1 < i\Delta_n) \\
&= (\Delta_n \phi(D))^{k-l+1} \\
&\quad \times \mathbb{E}_{i-1}^n(1_{S(D)_{l-1} < i\Delta_n} \mathbb{P}(N(D')_{i\Delta_n} - N(D')_{S(D'')_{l-1}} \geq 1 \mid \mathcal{F}_{S(D)_1})) \\
&\leq (\Delta_n \phi(D))^{k-l+1} \Delta_n \phi(D') \mathbb{P}_{i-1}^n(S(D)_{l-1} < i\Delta_n),
\end{aligned}$$

where (46) has been applied twice. Another application of the same then yields

(47) $$\begin{aligned}
D \cap D' = \varnothing \Rightarrow \mathbb{P}_{i-1}^n(\Delta_i^n N(D) \geq k, \Delta_i^n N(D') \geq 1) \\
\leq (k+1)\Delta_n^{k+1} \phi(D)^k \phi(D').
\end{aligned}$$

Next, let $w \in (0, 1/3]$. By convention $(a, b] = \varnothing$ when $a \geq b$ below. If $u(1-w) < \Delta_i^n X(u^{1+r}) \leq u(1+w)$ we have four (nonexclusive) possibilities: either $\Delta_i^n N((u^{1+r}, \infty)) \geq 3$, or $\Delta_i^n N((u^{1+r}, u/3]) = \Delta_i^n N((u/3, \infty)) = 1$, or $\Delta_i^n N((u/3, \infty)) = 2$, or $\Delta_i^n N((u(1-w), u(1+w)]) = 1$. We an analogous implication if $-u(1+w) < \Delta_i^n X(u^{1+r}) \leq -u(1-w)$. Then (43) easily follows from (46) applied with $D = [-u^{1+r}, u^{1+r}]^c$, with $D = [-u/3, u/3]^c$ and with

$D = [-u(1 + w), -u(1 - w)) \cup (u(1 - w), u(1 + w)]$, and from (47) applied with
$D = (-u/3, -u^{1+r}) \cup (u^{1+r}, u/3]$ and $D' = [-u/3, u/3]^c$.

Finally we prove (45). Let $H = |\zeta(X(u^{1+r}), u)_i^n - \Delta_i^n V(u)|$ and $D = [-u/2, -u^{1+r}) \cup (u^{1+r}, u/2]$ and $D' = [-u/2, u/2]^c$ and $D'' = D \cup D'$. From what precedes, we have

$$\mathbb{P}_{i-1}^n\big(\Delta_i^n N(D'') \geq k\big) \leq \big(\theta \Delta_n u^{-\beta_1(1+r)}\big)^k,$$

(48)
$$\mathbb{P}_{i-1}^n\big(\Delta_i^n N(D') = 2\big) \leq \theta^2 \Delta_n^2 u^{-2\beta_1},$$

$$\mathbb{P}_{i-1}^n\big(\Delta_i^n N(D) = \Delta_i^n N(D') = 1\big) \leq \theta^2 \Delta_n^2 u^{-\beta_1(2+r)}.$$

We have $H = 0$ on the sets $\{\Delta_i^n N(D'') \leq 1\}$ and $\{\Delta_i^n N(D'') = \Delta_i^n N(D) = 2\}$, and $H \leq k - 1$ on the set $\{\Delta_i^n N(D'') = k\}$, for all $k \geq 2$. Thus if $v = \theta \Delta_n u^{-\beta_1(1+r)}$,

$$\mathbb{E}_{i-1}^n(H) \leq \sum_{k=3}^\infty k \mathbb{P}_{i-1}^n\big(\Delta_i^n N(D'') \geq k\big) + \mathbb{P}_{i-1}^n\big(\Delta_i^n N(D') = 2\big)$$

$$+ \mathbb{P}_{i-1}^n\big(\Delta_i^n N(D) = \Delta_i^n N(D') = 1\big)$$

$$\leq \sum_{k=3}^\infty k v^k + \theta^2 \Delta_n^2 u^{-2\beta_1} + \theta^2 \Delta_n^2 u^{-\beta_1(2+r)}$$

by (48). When $v \leq 1/2$, that is, when $\Delta_n \leq \gamma u^{\beta_1(1+r)}$ for $\gamma = 1/2\theta$, we have $\sum_{k=3}^\infty k v^k \leq K v^3$, and the above is smaller than the right-hand side of (45). □

LEMMA 2. *Let $q \geq 2$ and $u, r \in (0, 1)$. As soon as (44) holds for some constant $\gamma > 0$, we have*

(49) $$\mathbb{E}\big(|\Delta_i^n (X - X(u^{1+r}))|^q\big) \leq K_{\gamma,q} \Delta_n\big(\Delta_n^{q/2-1} + u^{(q-\beta_1)(1+r)}\big).$$

PROOF. Letting X^c and μ be the continuous martingale part and the jump measure of X, we have $X - X(u^{1+r}) = B + B' + X^c + M$, where

$$B_t' = -\int_0^t ds \int_{\{u^{1+r} < |x| \leq 1\}} x F_s(dx), \qquad M_t = \int_0^t \int_{\{0 < |x| \leq u^{1+r}\}} x(\mu - \nu)(ds, dx).$$

By the strengthened Assumption 2, for any $y > 0$ the integral $\int_{\{|x|>y\}} |x| F(dx)$ is smaller than K when $\beta_1 < 1$, than $K \log \frac{1}{y}$ when $\beta_1 = 1$, and than $K y^{1-\beta_1}$ when $\beta_1 > 1$. Therefore, since (44) implies $2\beta_1(1 + r) > (\beta_1 - 1)^+$ we have $|\Delta_i^n B'| \leq K_\gamma \sqrt{\Delta_n}$. The strengthened Assumption 2 also implies $|\Delta_i^n B| \leq K \Delta_n$ and, by well-known estimates about continuous and purely discontinuous martingales [see, e.g., Aït-Sahalia and Jacod (2011)], we also deduce that

$$\mathbb{E}(|\Delta_i^n M'|^q) \leq K_q \Delta_n u^{(q-\beta_1)(1+r)}, \qquad \mathbb{E}(|\Delta_i^n X^c|^q) \leq K_q \Delta_n^{q/2}.$$

All these estimates readily give (49). □

PROOF OF PROPOSITION 1.   (a) It follows from (16) that $u_n^{2\beta_1}/\Delta_n \to \infty$, so for any $\gamma > 0$ (44) is satisfied for all $r \in (0, 1)$ and all n large enough. Hence, both estimates (45) and (49) hold, with constants K and $K_{\gamma,q}$ independent of r, for all n large enough.

The following inequality, where $u, w \in (0, 1)$ and $x, y \in \mathbb{R}$, is elementary:

$$\big|1_{\{x+y>u\}} - 1_{\{x>u\}}\big| \le 1_{\{|y|\ge uw\}} + 1_{\{u(1-w)<|x|\le u(1+w)\}}.$$

We apply this with $x = \Delta_i^n X(u^{1+r})$ and $x + y = \Delta_i^n X$ and $u = u_n$, and with $w \le 1/3$ to be chosen later. In order to evaluate the probabilities for having $|y| \ge u_n w$, respectively, $u_n(1 - w) < |x| \le u_n(1 + w)$, we use (49) and Markov's inequality, respectively, (43). This gives that $\mathbb{E}(|\zeta(X(u_n^{1+r}), u_n)_i^n - \zeta(X, u_n)_i^n|)$ is smaller, for all $q \ge 2$, than

$$\frac{K_q \Delta_n}{u_n^{\beta_1}}\left(\frac{\Delta_n^{q/2-1}}{w^q u_n^{q-\beta_1}} + \frac{u_n^{(q-\beta_1)r}}{w^q} + w + \frac{\Delta_n}{u_n^{\beta_1(1+r)}} + \frac{\Delta_n^2}{u_n^{\beta_1(2+3r)}} + u_n^{\beta_1-\beta_{j+1}}\right).$$

Optimizing over w leads to take $w = w_n$ such that $w_n^{q+1} = u_n^{(q-\beta_1)r} + \Delta_n^{q/2-1}/u_n^{q-\beta_1}$, which is indeed smaller than $1/3$ for all n large. Thus, putting the above together with (45), and recalling that $\Delta_n \le K u_n^{1/\rho}$, we end up with

$$(50) \qquad \mathbb{E}\big(|\zeta(X, u_n)_i^n - \Delta_i^n V(u_n)|\big) \le \frac{K_q \Delta_n}{u_n^{\beta_1}} \sum_{k=1}^5 u_n^{x_k},$$

where $x_k = x_k(q, r)$ are given by

$$x_1 = \frac{qr - \beta_1 r}{q+1}, \qquad x_2 = \frac{q(1-2\rho) - 2 + 2\beta_1\rho}{2\rho(q+1)},$$

$$x_3 = \frac{1}{\rho} - \beta_1(1+r), \qquad x_4 = \frac{2}{\rho} - \beta_1(2+3r), \qquad x_5 = \beta_1 - \beta_{j+1}.$$

(b) Now, for proving (40), it clearly follows from (50) that it suffices to show that one can choose q and r in such a way that $x_k \ge \beta_1/2$ for $k = 1, 2, 3, 4$. When $q \to \infty$ we see that $x(1) \to x'(1) = r$ and $x(2) \to x'(2) = \frac{1-2\rho}{2\rho}$, so it remains to show that one can choose $r \in (0, 1)$ such that $x'(k) \ge \beta_1/2$ for $k = 1, 2$ and $x_k \ge \beta_1/2$ for $k = 3, 4$. Letting r be bigger than but as close as possible to $\beta_1/2$, we deduce from (16) that such a choice or r is possible, and the proof is complete. $\square$

APPENDIX C: PROOF OF THEOREM 2

(1) In addition to the strengthened Assumption 2, we assume (15) for some $\varepsilon > 0$. Theorem 2 says something about the estimators of β_i and A_T^i only when $\beta_i > \frac{\beta_1}{2}$. Moreover, if (6) holds for the sequence $\beta_1, \ldots, \beta_{j+1}$, it also holds for

the sequence $\beta'_1, \ldots, \beta'_{j'+1}$, where $j' = j$ if $\beta_{j+1} \geq \frac{\beta_1}{2}$ and $j' = \sup(i : \beta_i > \frac{\beta_1}{2})$ otherwise, and where $\beta'_i = \beta_i$ when $i \leq j'$ and $\beta'_{j'+1} = \beta_{j'+1} \vee \frac{\beta_1}{2}$. Henceforth, upon discarding the indices such that $\beta_i \leq \frac{\beta_1}{2}$, we can assume without loss of generality that

$$(51) \qquad\qquad \beta_1 > \cdots > \beta_j > \beta_{j+1} = \frac{\beta_1}{2}.$$

Under this additional assumption, we have $\beta_1 - \beta_j < 1$, and (18) yields

$$(52) \qquad 1 \leq i \leq k < j \quad \Rightarrow \quad u_{n,i}^{\beta_i - \beta_{i+1}} \log \frac{1}{u_{n,i}} = o(u_{n,k}^{\beta_i - \beta_{k+1}}).$$

Moreover, combining (6), (39) and (40), we deduce that

$$(53) \qquad\qquad U(v_n, \Delta_n)_T = \sum_{i=1}^{j} \frac{A_T^i}{v_n^{\beta_i}} + O_P(v_n^{-\beta/2})$$

for any sequence v_n such that $v_n \leq u_n$, and in particular for the sequences $v_n = u_{n,i}$. All of the proof will rely on this, and below H_i is always given by (20).

(2) We first consider the case $i = 1$ when $j > 1$. A simple calculation, based on (53) applied with $v_n = u_n$ and $v_n = \gamma u_n$, yields that in restriction to the set Ω_T,

$$\log\bigl(U(v_n, \Delta_n)_T / U(\gamma v_n, \Delta_n)_T\bigr) = (\beta_1 - H_1 u_n^{\beta_1 - \beta_2}) \log \gamma + o_P(u_n^{\beta_1 - \beta_2}).$$

This gives the first part of (21). It also implies that

$$u_n^{\widetilde{\beta}_1^n} = u_n^{\beta_1} e^{-(\widetilde{\beta}_1^n - \beta_1) \log(1/u_n)}$$
$$= u_n^{\beta_1} \bigl(1 + H_1 u_n^{\beta_1 - \beta_2} \log(1/u_n) + o_P\bigl(u_n^{\beta_1 - \beta_2} \log(1/u_n)\bigr)\bigr).$$

This and (53) yield the second part of (21).

(3) Now we suppose that (21) holds for all $i \leq k - 1$, for some $k \in \{2, \ldots, j - 1\}$. We observe that we have the following identities, for all $y = (y_1, \ldots, y_{k+1})$ and $r = 1, \ldots, k + 1$:

$$\sum_{l=0}^{k-1} (-1)^l \gamma^{-l y_r} \sum_{J \in I(k-1,l)} \gamma^{\sum_{j \in J} y_j}$$
$$= \prod_{l=1}^{k-1} (1 - \gamma^{y_i - y_r}) = \begin{cases} 0, & \text{if } r \leq k - 1, \\ G(k, y, \gamma), & \text{if } r = k, \\ G'(k, y, \gamma), & \text{if } r = k + 1, \end{cases}$$

where $G(k, y, \gamma) = \prod_{i=1}^{k-1} (1 - \gamma^{y_i - y_k})$ and $G'(k, y, \gamma) = \prod_{i=1}^{k-1} (1 - \gamma^{y_i - y_{k+1}})$. Therefore, (53) applied to $v_n = x \gamma^l u_{n,k}$ and the definition of $U^n(k, x)$ yield for

all $x \geq 1$ fixed, and with $\beta = (\beta_1, \ldots, \beta_{k+1})$,

$$
U^n(k, x) = \sum_{r=1}^{k-1} \frac{A_T^r}{x^{\beta_r} u_{n,k}^{\beta_r}} \sum_{l=0}^{k-1} (-1)^l (\gamma^{-l\beta_r} - \gamma^{-l\widetilde{\beta}_r^n}) \sum_{J \in I(k-1,l)} \gamma^{\sum_{j \in J} \widehat{\beta}_j^n}
$$

$$
+ \sum_{r=k}^{k+1} \frac{A_T^r}{x^{\beta_r} u_{n,k}^{\beta_r}} \sum_{l=0}^{k-1} (-1)^l \gamma^{-l\beta_r} \sum_{J \in I(k-1,l)} (\gamma^{\sum_{j \in J} \widetilde{\beta}_j^n} - \gamma^{\sum_{j \in J} \beta_j})
$$

$$
+ \frac{A_T^k}{x^{\beta_k} u_{n,k}^{\beta_k}} G(k, \beta, \gamma) + \frac{A_T^{k+1}}{x^{\beta_{k+1}} u_{n,k}^{\beta_{k+1}}} G'(k, \beta, \gamma) + o_P(u_{n,k}^{-\beta_{k+1}}).
$$

The functions $z \mapsto \gamma^{-lz}$ are C^∞. The induction hypothesis gives $\widetilde{\beta}_i^n - \beta_i = O_P(u_{n,i}^{\beta_i - \beta_{i+1}})$ for $i = 1, \ldots, k-1$. Then (52) and $\beta_i - \beta_{i+1} > \varepsilon$ allow us to deduce

$$
0 \leq l \leq k-1, J \in I(k-1,l) \quad \Rightarrow \quad \gamma^{\sum_{j \in J} \widetilde{\beta}_j^n} - \gamma^{\sum_{j \in J} \beta_j} = o_P(u_{n,k}^{\beta_{k-1} - \beta_{k+1}}),
$$

$$
1 \leq r \leq k-1 \quad \Rightarrow \quad \gamma^{-l\beta_r} - \gamma^{-l\widetilde{\beta}_r^n} = o_P(u_{n,i}^{\beta_r - \beta_{r+1}}) = o_P(u_{n,k}^{\beta_i - \beta_{k+1}}).
$$

Therefore we finally obtain

$$
U^n(k, x) = \frac{A_T^k G(k, \beta, \gamma)}{x^{\beta_k} u_{n,k}^{\beta_k}} + \frac{A_T^{k+1} G'(k, \beta, \gamma)}{x^{\beta_{k+1}} u_{n,k}^{\beta_{k+1}}} + o_P(u_{n,k}^{-\beta_{k+1}})
$$

(54)

$$
= \frac{A_T^k G(k, \beta, \gamma)}{x^{\beta_k} u_{n,k}^{\beta_k}} \left(1 + \frac{H_k \log \gamma}{\gamma^{\beta_k - \beta_{k+1}} - 1} (x u_{n,k})^{\beta_k - \beta_{k+1}} + o_P(u_{n,k}^{\beta_k - \beta_{k+1}})\right),
$$

where the last equality comes from the definition of H_k in (20). Then exactly as in Step 2, a simple calculation shows the first half of (21) for $i = k$.

For the second part of (21), and as in Step 2, we first deduce from the above that

$$
u_{n,k}^{\widetilde{\beta}_k^n} = u_{n,k}^{\beta_k} (1 + H_k u_{n,k}^{\beta_k - \beta_{k-1}} \log(1/u_{n,k}) + o_P(u_{n,k}^{\beta_k - \beta_{k-1}} \log(1/u_{n,k}))).
$$

Therefore it is enough to show that

$$
u_{n,k}^{\beta_k} \left(U(u_{n,k})_T - \sum_{i=1}^{k-1} \widetilde{\Gamma}_i^n u_{n,k}^{-\widetilde{\beta}_i^n}\right) = A_T^k + o_P(u_{n,k}^{\beta_k - \beta_{k-1}} \log(1/u_{n,k})).
$$

In view of (53) with $v_n = u_{n,k}$ this amounts to proving for $i = 1, \ldots, k-1$,

(55)
$$
\widetilde{\Gamma}_i^n u_{n,k}^{\beta_k - \widetilde{\beta}_i^n} - A_T^i u_{n,k}^{\beta_k - \beta_i} = o_P(u_{n,k}^{\beta_k - \beta_{k-1}} \log(1/u_{n,k})).
$$

The induction hypothesis yields that

$$
u_{n,k}^{\beta_k - \widetilde{\beta}_i^n} = u_{n,k}^{\beta_k - \beta_i} (1 + O_P(u_{n,i}^{\beta_i - \beta_{i+1}} \log(1/u_{n,k}))),
$$

$$
\widetilde{\Gamma}_i^n = A_T^i + O_P(u_{n,i}^{\beta_i - \beta_{i+1}} \log(1/u_{n,i})).
$$

Then (21) readily follows from (18).

(4) It remains to prove that the variables in (22) are tight. The difference with the previous case is that (54) no longer holds when $i = j = 1$ or $k = j > 1$, but it can be replaced by

$$U^n(j, x) = \frac{A_T^j G(j, \beta, \gamma)}{x^{\beta_j} u_{n,j}^{\beta_j}} (1 + O_P(u_{n,j}^{\beta_j - \beta/2})).$$

The rest of the proof goes unchanged [note that η in (22) is $\beta_j - \beta/2$ here].

APPENDIX D: PROOF OF THEOREM 3

We use simplifying notation: a point in D is $\theta = (x_i, \gamma_i)_{1 \leq i \leq j}$, and we define the functions $F_{n,l}(\theta) = \sum_{i=1}^j \gamma_i / (v_l u_n)^{x_i}$. The "true value" of the parameter is $\theta_0 = (\beta_i, \Gamma_i)_{1 \leq i \leq j}$, the preliminary estimators are $\widetilde{\theta}_n = (\widetilde{\beta}_i^n, \widetilde{\Gamma}_i^n)_{1 \leq i \leq j}$, and the final estimators are $\overline{\theta}_n = (\overline{\beta}_i^n, \overline{\Gamma}_i^n)_{1 \leq i \leq j}$. We set $h_n = \log(1/u_n)$, and as in the previous proof we can assume (51).

(1) We introduce some specific notation. For $m \geq 2$ we set $G_m = (1, \infty)^{m-1}$, a point in G_m being denoted as $\overline{v} = (v_2, \ldots, v_m)$. For $1 \leq k \leq j$ and $\overline{v} \in G_{2k}$, and with the convention $v_1 = 1$, we let $\Sigma(\overline{v})$ be the $2k \times 2k$ matrix with entries

$$(56) \qquad \Sigma(\overline{v})_{l,i} = \begin{cases} v_l^{-\beta_i}, & \text{if } 1 \leq i \leq k, \\ v_l^{-\beta_{i-k}} \log v_l, & \text{if } k+1 \leq i \leq 2k. \end{cases}$$

The aim of this step is to show that the set Z_k of all $\overline{v} \in G_{2k}$ for which the matrix $\Sigma(\overline{v})$ is invertible satisfies $\lambda_{2k}((Z_k)^c) = 0$, where λ_r is the Lebesgue measure on G_r.

When $1 \leq m \leq 2k$ and $\overline{v} \in G_{2k}$, we denote by $\mathcal{M}_m(\overline{v})$ the family of all $m \times m$ sub-matrices of the $m \times 2k$ matrix $(\Sigma(\overline{v})_{l,r} : 1 \leq l \leq m, 1 \leq r \leq 2k)$. A key fact is that $\mathcal{M}_m(\overline{v}) = \mathcal{M}_m(\overline{v}_m)$ only depends on the restriction $\overline{v}_m = (v_2, \ldots, v_m)$ of $\overline{v}$ to its first $m - 1$ coordinates. Moreover, $\Sigma(\overline{v})_{1i}$ equals 1 if $i \leq k$ and 0 otherwise: so the entries of the first column of any $M \in \mathcal{M}_m(\overline{v})$ are 0 or 1, and $\mathcal{M}_m'(\overline{v})$ denotes the subset of all $M \in \mathcal{M}_m(\overline{v})$ for which $M_{1,i} = 1$ for at least one value of i. Finally, H_m stands for the set of all $\overline{v}_m \in G_m$ such that all $M \in \mathcal{M}_m'(\overline{v}_m)$ are invertible. Since $\mathcal{M}_{2k}'(\overline{v})$ is the singleton $\{\Sigma(\overline{v})\}$, we have $Z_k = H_{2k}$.

If $m \geq 2$ and $\overline{v}_m = (v_2, \ldots, v_m) \in G_m$ and $M \in \mathcal{M}_m'(\overline{v}_m)$, by expanding along the last column, we see that

$$(57) \qquad \det(M) = \sum_{i=1}^k v_m^{\beta_i} (a_i + a_{k+i} \log v_m),$$

where each a_r is of the form: either (i) a_r is plus or minus $\det(M_r)$ for some $M_r \in \mathcal{M}_{m-1}(\overline{v}_m)$ (for m values of r) or (ii) $a_r = 0$ (for the other $2k - m$ values of r). Note that we can also have $a_r = 0$ in case (i), and since $M \in \mathcal{M}_m'(\overline{v}_m)$ there is at least one a_r of type (i) with $M_r \in \mathcal{M}_{m-1}'(\overline{v}_m)$.

When at least one a_r in (57) is not 0, the right-hand side of this expression, as a function of v_m, has finitely many roots only, because all β_i's are distinct. Observing that $\mathcal{M}'_1(\overline{v})$ is the 1×1 matrix equal to 1, it follows that, with $(\overline{v}_{m-1}, v_m) = (v_2, \ldots, v_{m-1}, v_m)$ when $\overline{v}_{m-1} = (v_2, \ldots, v_{m-1})$, and recalling that with our standing notation λ_2 is the Lebesgue measure on $(1, \infty)$,

$$m = 2 \quad \Rightarrow \quad \lambda_2((H_2)^c) = 0,$$

(58)

$$m \geq 3, \qquad \overline{v}_{m-1} \in H_{m-1} \quad \Rightarrow \quad \lambda_2(v_m : (\overline{v}a_{m-1}, v_m) \notin H_m) = 0.$$

Since

$$\lambda_m((H_m)^c) = \int_{G_{m-1}} \lambda_2(v_m : (\overline{v}_{m-1}, v_m) \notin H_m) \lambda_{m-1}(d\overline{v}_{m-1}),$$

which equals $\int_{H_{m-1}} \lambda_1(v_m : (\overline{v}_{m-1}, v_m) \notin H_m) \lambda_{m-1}(d\overline{v}_{m-1})$ if $\lambda_{m-1}((H_{m-1})^c) = 0$, when $m \geq 3$, we deduce from (58), by induction on m, that indeed $\lambda_m((H_m)^c) = 0$ for all $m = 2, \ldots, 2k$. Recalling $Z_k = H_{2k}$, the result follows.

Since the claim of the theorem holds for all $(v_2, \ldots, v_L)$ outside a λ_L-null set only, and $L \geq 2k$, we thus can and will assume below that the numbers v_l are such that $\overline{v}_{2k} = (v_2, \ldots, v_{2k}) \in Z_k$, hence $\Sigma(\overline{v}_{2k})$ is invertible, for all $k = 1, \ldots, j$.

(2) Our assumptions on the preliminary estimators yield that the set Ω_n on which $\|\widetilde{\theta}_i^n - \theta_0\| \leq 1/u_n^\eta$ satisfies $\mathbb{P}(\Omega_n) \to 1$. So below we argue on the set Ω_n, or equivalently (and more conveniently) we suppose $\Omega_n = \Omega$. Then $\overline{\theta}_n$ converges pointwise to θ_0, which belongs to all the sets D_n. Set

$$y_i^n = A_T^i(\overline{\beta}_i^n - \beta_i), \qquad z_i^n = \overline{\Gamma}_i^n - A_T^i + y_i^n h_n, \qquad a_i^n = |y_i^n| h_n + |z_i^n|.$$

We have $a_i^n \leq 2u_n^{-\eta} h_n$ because $\Omega_n = \Omega$. Then an expansion of $(x_i, \gamma_i) \mapsto \gamma_i/(v_l u_n)^{x_i}$ around (β_i, A_T^i) yields for all l,

(59)

$$\frac{\overline{\Gamma}_i}{(v_l u_n)^{\overline{\beta}_i}} - \frac{A_T^i}{(v_l u_n)^{\beta_i}} = \frac{1}{(v_l u_n)^{\beta_i}}(z_i^n - y_i^n \log v_l + x_{i,l}^n),$$

where

$$|x_{i,l}^n| \leq K|y_i^n| h_n(|z_i^n| + |y_i^n|) \leq K|y_i^n| h_n a_i^n \leq K(a_i^n)^2.$$

Combining (6), (39) and (40), we see that

$$U(v_l u_n, \Delta_n)_T - F_{n,l}(\theta_0) = O_P(u_n^{-\beta_1/2}).$$

Since $\Phi_n(\theta) = \sum_{l=1}^L w_l(U(v_l u_n, \Delta_n)_T - F_{n,l}(\theta))^2$, we deduce

$$\Phi_n(\theta_0) = O_P(u_n^{-\beta_1}).$$

Since $\theta_0 \in D_n$ and $\overline{\theta}_n$ minimizes Φ_n over D_n, we also have $\Phi_n(\overline{\theta}_n) = O_P(u_n^{-\beta_1})$, hence $F_{n,l}(\theta_0) - F_{n,j}(\overline{\theta}_n) = O_P(u_n^{-\beta_1/2})$ for all l. Using (59), this can be rewritten as

(60)

$$\sum_{i=1}^j \frac{1}{(v_p u_n)^{\beta_i}}(z_i^n - y_i^n \log v_l + x_{i,l}^n) = O_P(u_n^{-\beta_1/2}).$$

(3) Taking k between 1 and j, we consider the $2k$-dimensional vectors $\zeta(k, n)$ and $\xi(n)$ with components (for $l = 1, \ldots, 2k$),

$$\zeta(k, n)_l = \sum_{i=1}^{k} \frac{1}{(v_p u_n)^{\beta_i}} (z_i^n - y_i^n \log v_l),$$

$$\xi(k, n)_i = \begin{cases} z_i^n u_n^{-\beta_i}, & \text{if } 1 \le i \le k, \\ -y_{i-k}^n u_n^{-\beta_{i-k}}, & \text{if } k + 1 \le i \le 2k. \end{cases}$$

With matrix notation, and (56), we have $\zeta(k, n) = \Sigma(\overline{v}_{2k})\xi(k, n)$, hence

(61) $$\xi(k, n) = \Sigma(\overline{v}_{2k})^{-1}\zeta(k, n).$$

Next, we have

$$\frac{1}{(v_l u_n)^{\beta_i}} |z_i^n + (v_g'(\beta_i) + v_g(\beta_i) \log \delta_l) y_i^n + x_{i,l}^n| \le \frac{K a_i^n}{u_n^{\beta_i}}, \qquad \frac{|x_{i,l}^n|}{(v_l u_n)^{\beta_i}} \le \frac{K (a_i^n)^2}{u_n^{\beta_i}},$$

and hence (60) and $a_i^n \le K u_n^\eta h_n \le K / h_n^2 \le K$ yield

$$|\zeta(k, n)_l| \le K \left(\sum_{i=1}^{k-1} (a_i^n)^2 u_n^{-\beta_i} + \frac{a_k^n}{h_n^2} u_n^{-\beta_k} + \sum_{i=k+1}^{j} a_i^n u_n^{-\beta_i} \right) + O_P(u_n^{-\beta_{j+1}}).$$

By (61) the variables $\xi(k, n)_l$ satisfy the same estimate. Since $a_k^n \le (|\xi(k, n)_k| + |\xi(k, n)_{2k}| h_n) u_n^{\beta_k}$,

$$a_k^n \le C h_n \left(\sum_{i=1}^{k-1} (a_i^n)^2 u_n^{\beta_k - \beta_i} + \frac{a_k^n}{h_n^2} + \sum_{i=k+1}^{j} a_i^n u_n^{\beta_k - \beta_i} \right) + O_P(h_n u_n^{\beta_k - \beta_{j+1}})$$

for some constant C. When n is large enough, $C / h_n \le \frac{1}{2}$, and we deduce

(62) $$a_k^n \le 2 C h_n \left(\sum_{i=1}^{k-1} (a_i^n)^2 u_n^{\beta_k - \beta_i} + \sum_{i=k+1}^{j} a_i^n u_n^{\beta_k - \beta_i} \right) + O_P(h_n u_n^{\beta_k - \beta_{j+1}}).$$

(4) In view of the definition of y_i^n and z_i^n, to get the result, and recalling that we assume $\beta_{j+1} = \beta_1/2$, it is clearly enough to prove the existence of a number $\nu > 0$ such that, for all $i = 1, \ldots, j$, we have

(63) $$a_i^n = O_P(h_n^\nu u_n^{\beta_i - \beta_{j+1}}).$$

To this aim, we introduce the following property, denoted $(P_{m,q,r})$, where r runs through $\{1, \ldots, j\}$ and $m, q \ge 1$, and where we use the notation $\zeta_r = \beta_r - \beta_{r+1}$:

(64) $$i = 1, \ldots, r \quad \Rightarrow \quad a_i^n = O_P(h_n^m (u_n^{\beta_i - \beta_r + q\zeta_r} + u_n^{\beta_i - \beta_{r+1}})).$$

Since $a_i^n \le K$, applying (62) with $k = 1$ yields $a_1^n = O_P(h_n u_n^{\beta_1 - \beta_2})$, which is $(P_{1,1,1})$.

Next, we suppose that $(P_{m,q,r})$ holds for some $r < j$, and for some $m, q \geq 1$. Letting first $k = r + 1$, we deduce from (62) that, since again $a_i^n \leq K$,

$$a_k^n = O_P\left(h_n^{1+2m} \sum_{i=1}^{k-1} (u_n^{\beta_k - \beta_i + 2(\beta_i - \beta_r + q\zeta_r)} + u_n^{\beta_i - \beta_{r+1}})\right.$$

$$(65) \qquad\qquad \left. + h_n \sum_{i=k+1}^{j} u_n^{\beta_k - \beta_i} + h_n u_n^{\beta_k - \beta_{j+1}}\right)$$

$$= O_P\left(h_n^{1+2m} (u_n^{\beta_k - \beta_r + 2q\zeta_r} + u_n^{\zeta_r} + u_n^{\beta_k - \beta_{r+2}})\right),$$

where the last line holds because $k = r + 1$ and $h_n > 1$ for n large enough and the sequence β_i is decreasing. This in turn implies, for $k = r + 1$ again,

$$(66) \qquad a_k^n = O_P\left(h_n^{r+2-k+2m} (u_n^{\beta_k - \beta_r + 2q\zeta_r} + u_n^{\beta_k - \beta_{r+1}})\right).$$

Then, exactly as above, we apply (62) with $k = r$, and (64) and also (66) with $k = r + 1$, to get that (66) holds for $k = r$ as well. Repeating the argument, a downward induction yields that indeed (66) holds for all k between 1 and $r + 1$. Thus (64) holds with q and m substituted with $2q$ and $r + 1 + 2m$. Hence $(P_{m,q,r})$ implies $(P_{r+1+2m,2q,r})$. Since obviously $(P_{m,q,r}) \Rightarrow (P_{m,q',r})$ for any $q' \in [1, q]$, by a repeated use of the previous argument we deduce that if $(P_{m,1,r})$ holds for some $m \geq 1$, then for any $q' \geq 1$ we can find $m(q') \geq 1$ such that $(P_{m(q'),q',r})$ holds as well.

Now, assuming $(P_{m,q,r})$ for some m, q, r, we take $q' = \frac{\zeta_{r+1}}{2\zeta_r} \vee 1$ and $m' = m(q')$. What precedes yields $(P_{m',q',r})$, hence (65) holds for all $k \leq r + 1$, with q' and m'. In view of our choice of q', this implies that $(P_{r+1+m',1,r+1})$ holds. Since $(P_{1,1,1})$ holds, we see by induction that for any $r \leq j$ there exists $m_r \geq 1$ such that $(P_{m_r,1,r})$ holds.

It remains to apply (64) with $r = j$ and $m = m_r$ and $q = 1$, and we get (63) with $\nu = m_j$. This completes the proof.

APPENDIX E: PROOF OF THEOREM 4

The proof of Theorem 4 is contained in the supplemental article [Aït-Sahalia and Jacod (2012)].

SUPPLEMENTARY MATERIAL

Supplement to "Identifying the successive Blumenthal–Getoor indices of a discretely observed process" (DOI: 10.1214/12-AOS976SUPP; .pdf). This supplement contains the proof of Theorem 4.

REFERENCES

AÏT-SAHALIA, Y. and JACOD, J. (2008). Fisher's information for discretely sampled Lévy processes. *Econometrica* **76** 727–761. MR2433480

AÏT-SAHALIA, Y. and JACOD, J. (2009a). Estimating the degree of activity of jumps in high frequency financial data. *Ann. Statist.* **37** 2202–2244.

AÏT-SAHALIA, Y. and JACOD, J. (2009b). Testing for jumps in a discretely observed process. *Ann. Statist.* **37** 184–222. MR2488349

AÏT-SAHALIA, Y. and JACOD, J. (2011). Testing whether jumps have finite or infinite activity. *Ann. Statist.* **39** 1689–1719. MR2850217

AÏT-SAHALIA, Y. and JACOD, J. (2012). Supplement to "Identifying the successive Blumenthal–Getoor indices of a discretely observed process." DOI:10.1214/12-AOS976SUPP.

BASAWA, I. V. and BROCKWELL, P. J. (1982). Nonparametric estimation for nondecreasing Lévy processes. *J. Roy. Statist. Soc. Ser. B* **44** 262–269. MR0676217

BELOMESTNY, D. (2010). Spectral estimation of the fractional order of a Lévy process. *Ann. Statist.* **38** 317–351. MR2589324

BLUMENTHAL, R. M. and GETOOR, R. K. (1961). Sample functions of stochastic processes with stationary independent increments. *J. Math. Mech.* **10** 493–516. MR0123362

COMTE, F. and GENON-CATALOT, V. (2009). Nonparametric estimation for pure jump Lévy processes based on high frequency data. *Stochastic Process. Appl.* **119** 4088–4123. MR2565560

CONT, R. and MANCINI, C. (2011). Nonparametric tests for pathwise properties of semimartingales. *Bernoulli* **17** 781–813. MR2787615

FIGUEROA-LÓPEZ, J. E. and HOUDRÉ, C. (2006). Risk bounds for the non-parametric estimation of Lévy processes. In *High Dimensional Probability* (E. Giné, V. Koltchinskii, W. Li and J. Zinn, eds.). *Institute of Mathematical Statistics Lecture Notes—Monograph Series* **51** 96–116. IMS, Beachwood, OH. MR2387763

JACOD, J. and SHIRYAEV, A. N. (2003). *Limit Theorems for Stochastic Processes*, 2nd ed. *Grundlehren der Mathematischen Wissenschaften* [*Fundamental Principles of Mathematical Sciences*] **288**. Springer, Berlin. MR1943877

MANCINI, C. (2004). Estimating the integrated volatility in stochastic volatility models with Lévy type jumps. Technical report, Univ. Firenze.

NEUMANN, M. H. and REISS, M. (2009). Nonparametric estimation for Lévy processes from low-frequency observations. *Bernoulli* **15** 223–248. MR2546805

NISHIYAMA, Y. (2008). Nonparametric estimation and testing time-homogeneity for processes with independent increments. *Stochastic Process. Appl.* **118** 1043–1055. MR2418257

ROSIŃSKI, J. (2007). Tempering stable processes. *Stochastic Process. Appl.* **117** 677–707. MR2327834

TODOROV, V. and TAUCHEN, G. (2010). Activity signature functions for high-frequency data analysis. *J. Econometrics* **154** 125–138. MR2558956

ZOLOTAREV, V. M. (1995). On representation of densities of stable laws by special functions. *Theory Probab. Appl.* **39** 354–362.

DEPARTMENT OF ECONOMICS
PRINCETON UNIVERSITY AND NBER
PRINCETON, NEW JERSEY 08544-1021
USA
E-MAIL: yacine@princeton.edu

INSTITUT DE MATHÉMATIQUES DE JUSSIEU
CNRS UMR 7586
UNIVERSITÈ PIERRE AND MARIE CURIE
75252 PARIS CÉDEX 05
FRANCE
E-MAIL: jean.jacod@upmc.fr