

NBER WORKING PAPER SERIES

ANALYZING THE SPECTRUM OF ASSET RETURNS:  
JUMP AND VOLATILITY COMPONENTS IN HIGH FREQUENCY DATA

Yacine Aït-Sahalia  
Jean Jacod

Working Paper 15808  
<http://www.nber.org/papers/w15808>

NATIONAL BUREAU OF ECONOMIC RESEARCH  
1050 Massachusetts Avenue  
Cambridge, MA 02138  
March 2010

We are very grateful to Larry Harris and Joel Hasbrouck for helpful discussions regarding market microstructure issues. We are also grateful for the comments of the Editor and two anonymous referees. This research was partly funded by the NSF under grants DMS-0532370 and SES-0850533. MATLAB code to implement the methods described in this paper and sample data files can be downloaded from the authors' web pages. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

NBER working papers are circulated for discussion and comment purposes. They have not been peer-reviewed or been subject to the review by the NBER Board of Directors that accompanies official NBER publications.

© 2010 by Yacine Aït-Sahalia and Jean Jacod. All rights reserved. Short sections of text, not to exceed two paragraphs, may be quoted without explicit permission provided that full credit, including © notice, is given to the source.

Analyzing the Spectrum of Asset Returns: Jump and Volatility Components in High Frequency Data

Yacine Aït-Sahalia and Jean Jacod

NBER Working Paper No. 15808

March 2010

JEL No. C12,C14,C22,G12

**ABSTRACT**

This paper describes a simple yet powerful methodology to decompose asset returns sampled at high frequency into their base components (continuous, small jumps, large jumps), determine the relative magnitude of the components, and analyze the finer characteristics of these components such as the degree of activity of the jumps. We extend the existing theory to incorporate the effect of market microstructure noise on the test statistics, apply the methodology to high frequency individual stock returns, transactions and quotes, stock index returns and compare the qualitative features of the estimated process for these different data and discuss the economic implications of the results.

Yacine Aït-Sahalia

Department of Economics

Fisher Hall

Princeton University

Princeton, NJ 08544-1021

and NBER

yacine@princeton.edu

Jean Jacod

Institut de Mathématiques de Jussieu

CNRS UMR 7586

Université P. et M. Curie (Paris-6)

jean.jacod@upmc.fr

# I. Introduction

We present in this paper econometric methods designed to analyze the workhorse model of modern asset pricing:  $X$ , typically the log of an asset price, is assumed to follow an Itô semimartingale. As is well known, for an asset pricing model to avoid arbitrage opportunities, asset prices must follow semimartingales (see Harrison and Pliska (1981), Delbaen and Schachermayer (1994)). Semimartingales are very general models that nest most if not all continuous-time models used in asset pricing. A semimartingale can be decomposed into the sum of a drift, a continuous Brownian-driven part and a discontinuous, or jump, part. The jump part can in turn be decomposed into a sum of small jumps and big jumps. The continuous part can be scaled by a stochastic volatility process, which may be correlated with the asset price, may jump in conjunction or independently of the asset price, and in fact be a semimartingale itself.

This paper is devoted to analyzing the specification of semimartingales on the basis of high frequency financial returns. We wish to decide on the basis of statistical tests which component(s) need to be included in the model (jumps, finite or infinite activity, continuous component, etc.) and determine their relative magnitude. We may then magnify specific components of the model if they are present, so that we can analyze their finer characteristics such as the degree of activity of jumps.<sup>1</sup> While the underlying mathematical

---

<sup>1</sup>Alternative methodologies exist for some of the questions we consider when taken individually. For example, tests for the presence of jumps have been proposed by Aït-Sahalia (2002), Carr and Wu (2003b), Barndorff-Nielsen and Shephard (2004), Huang and Tauchen (2005), Andersen, Bollerslev, and Diebold (2007), Jiang and Oomen (2008), Lee and Mykland (2008), Aït-Sahalia and Jacod (2009b) and Lee and Hannig (2010), among others. And some of these methods are applicable (in fact, designed for) splitting the quadratic variation into continuous and discontinuous proportions, another of the issues of interest. To study the finer characteristics of jumps, Todorov and Tauchen (2010) use the test statistics of Aït-Sahalia and Jacod (2009b), study its logarithm for different values of the power argument and contrast the behavior of the plot above two and below two in order to identify the presence of a Brownian component. Cont and Mancini (2009) use threshold or truncation-based estimators of the continuous component of the quadratic variation, originally proposed in Mancini (2001), in order to test for the presence of a continuous component in the price process. The resulting test is applicable when the jump component of the process has finite variation, and a test for whether the jump component indeed has finite variation is also proposed. Belomestny (2009) proposes a different estimator of the index of jump activity based on low frequency data.

However, to the best of our knowledge, none of the alternative methods are able to address all the questions we consider here in a common framework. In fact, for some of the issues addressed in this paper, there exist

tools are heavily technical, and are developed elsewhere<sup>2</sup>, the end result happens to be straightforward from the point of view of applications. It requires little more than the recording of asset returns at high frequency, and the computation of a few key quantities which we call truncated power variations.

Relative to the existing literature, this paper makes three separate contributions. First, we show that seemingly disparate test statistics developed individually can in fact be understood as part of a common framework, relying on an analogy with spectrography, which we will carry through the entire paper, hence its title. This contribution is primarily expositional but hopefully has the benefit of facilitating the application of all these methods. Second, we provide new theoretical results regarding the asymptotic behavior of these test statistics in situations where market microstructure noise is present, a contribution which is essential for practical applications to high frequency financial data. Third, we compare the empirical results obtained on asset returns measured in different complementary ways, comparing the results obtained from transaction prices and quotes, but also by constructing from quotes the National Best Bid and Offer (NBBO) prices at each point time in order to filter the transactions into different liquidity categories, and by comparing the results obtained on the Dow Jones Industrial Average (DJIA) index from those obtained on its thirty individual constituents. We then provide economic interpretations and implications of the results for option pricing, risk management and the distinction between systematic and idiosyncratic risk in the individual stocks and the index.

To describe the methodology, it can be helpful to proceed by analogy with a spectrographic analysis. We observe a time series of high frequency returns, that is a single path, over a finite length of time  $[0, T]$ . Using that time series as input, we will then design a set of statistical tools that can tell us something about specific components of the process that produced the observations. These tools play the role of the measurement devices used in astrophysics to analyze the light emanating from a star, for instance. Our observations are as of this writing no other alternative(s), hence our focus on a unified approach to address all these issues together.

---

<sup>2</sup>In earlier work, we developed tests to determine on the basis of the observed log-returns whether a jump part was present (Aït-Sahalia and Jacod (2009b)), whether the jumps had finite or infinite activity (Aït-Sahalia and Jacod (2008b)), in the latter situation proposed a definition and an estimator of a degree of jump activity parameter (Aït-Sahalia and Jacod (2009a)), and finally whether a Brownian continuous component was needed when infinite activity jumps are included (Aït-Sahalia and Jacod (2008a)).

the high frequency returns; in astrophysics it would be the light, visible or not. Here, the data generating mechanism is assumed to be a semimartingale; in astrophysics it would be whatever nuclear reactions inside the star are producing the light that is collected. Astrophysicists can look at a specific range of the light spectrum to learn about specific chemical elements present in the star. Here, we design statistics that focus on specific parts of the distribution of high frequency returns in order to learn about the different components of the semimartingale that produced those returns.

From the time series of returns, we can get the distribution of returns at time interval  $\Delta_n$ . Based on the information contained in that distribution, we would like to figure out which components should be included in the model (continuous? jumps? which types of jumps?) and in what proportions. That is, we would like to deconstruct the observed series of returns back into its original components, continuous and jumps, as described in Figure 1. Figure 1 cannot be produced by visual inspection alone of either the time series of returns or its distribution. We need to run the raw data through some devices that will emphasize certain components to the exclusion of others, magnify certain aspects of the model, etc.

Similarly to what is done in spectrographic analysis, we will emphasize visual tools in this paper. In spectrography, one needs to be able to recognize the visual signature of certain chemical elements. Here, we need to know what to expect to see if a certain component of the model is present or not in the observed data. This means that we will need to have a law of large numbers, obtained by imagining that we had collected a large number of sample paths instead of a single one. This allows us to determine the visual signature of specific components of the model. We will not attempt here to measure the dispersion around the expected pattern, and instead refer to the papers in the reference list for the corresponding central limit theorems, the formal derivations of the results including regularity conditions, as well as simulation evidence on the adequacy of the asymptotics. Those papers are technically demanding because of the very nature of semimartingales, but also because depending upon which component is included or not in the model – precisely the questions we wish to answer – the asymptotics are driven by components with very different characteristics. By contrast, the intuition is fairly clear and this is what this paper focuses on, with the objective of facilitating applications of the results rather than their derivation, with the exception of the new results concerning the limits of all the test statistics when market microstructure noise is present.

The methodology helps determine which components should be included in a given semimartingale model of asset returns. This knowledge has various economic implications for asset pricing. Many high frequency trading strategies rely on specific components of the model being present or absent. If jumps need to be included in the model, then the familiar consequences of market completeness for contingent claims valuation typically no longer hold. And changes of measure will vary depending upon the type of jumps that are included. Optimal portfolios will vary depending upon the nature of the underlying asset dynamics. Risk management is also heavily dependent upon the underlying dynamics: clearly, a model with only a continuous component will yield very different risk measures than one with jump components also present, and different types of jumps aggregate separately over longer horizons. And in derivative pricing, the type of components included change the nature of observed prices: see for example the analysis of Carr and Wu (2003b) which shows how to distinguish between option prices when the price jumps and when it does not, based on their asymptotic behavior for short times to expiration. We will discuss these economic implications in more detail below.

A word on data considerations before we proceed: when implementing the method on returns data, we will rely on ultra-high frequencies, meaning that the sampling intervals we use are typically of the order of a few seconds to a few minutes. This has two consequences. First, obviously, it limits the analysis to data series for which such sampling frequencies are available. This is becoming less and less of a restriction as such data are rapidly becoming more readily available, but it does limit our ability to use long historical series, or returns data from less liquid assets. Second, this means that even for liquid assets market microstructure noise is going to be at least potentially a concern. Continuing with the spectrography analogy, market microstructure noise plays the same role as the blurring of astronomical images due to the Earth's atmosphere or light pollution. And we do not have the equivalent of a space-based telescope enabling the direct observation of the true or fundamental asset price. We will in the course of our analysis examine the consequences of this noise on the various statistics. From the mathematical standpoint, the new theoretical results in this paper are the development of the various asymptotic behaviors of all the test statistics under consideration when market microstructure noise is present. We then proceed to analyze the data in light not only of the idealized no-noise limits but also of these new limits, and contrast the first order asymptotic behavior identified at different sampling

frequencies where the impact of the noise can be expected to be more or less significant.

The paper is organized as follows. Section II presents the common measurement device we designed to answer the various specification questions. In latter sections, we analyze these questions one by one: which components are present (Section III), in what relative proportions (Section IV), and some of the finer characteristics of the jump component (Section V). Section VII analyzes theoretically the impact of the noise on the various statistics under consideration. Section VIII describes the data and the transformation and filtering algorithms we employ to transactions, quotes, and transactions filtered by quotes for the DJIA and its individual components. Section IX reports the results of applying the analysis to the data, analyzes the patterns that emerge in terms of liquidity and discuss the economic implications of the results for option pricing, risk management and the distinction between systematic and idiosyncratic risk in the individual DJ components vs. the DJ index. Section X concludes.

## II. The Measurement Device

The log-price  $X_t$  follows an Itô semimartingale, a hypothesis maintained throughout, and formally stated as

$$X_t = X_0 + \underbrace{\int_0^t b_s ds}_{\text{drift}} + \underbrace{\int_0^t \sigma_s dW_s}_{\text{continuous part}} + \text{JUMPS} \quad (1)$$

$$\text{JUMPS} = \underbrace{\int_0^t \int_{\{|x| \leq \varepsilon\}} x(\mu - \nu)(ds, dx)}_{\text{small jumps}} + \underbrace{\int_0^t \int_{\{|x| > \varepsilon\}} x\mu(ds, dx)}_{\text{big jumps}} \quad (2)$$

where as usual  $W$  denotes a standard Brownian motion, and  $\mu$  is the jump measure of  $X$ , and its predictable compensator is the Lévy measure  $\nu$  (both  $\mu$  and  $\nu$  are random positive measures on  $\mathbb{R}_+ \times \mathbb{R}$ , and further  $\nu$  factorizes as  $\nu(\omega, dt, dx) = dt F_t(\omega, dx)$ ). In the perhaps more familiar differential form,

$$dX_t = b_t dt + \sigma_t dW_t + dJ_t \quad (3)$$

where  $J_t$  is the jump term.

The distinction between small and big jumps is based on a cutoff level  $\varepsilon > 0$  in (2) that is arbitrary. What is important is that  $\varepsilon > 0$  is fixed. A semimartingale will always generate a finite number of big jumps on  $[0, T]$ . But it may give rise to either a finite or infinite number of small jumps. For any measurable subset  $A$  of  $\mathbb{R}$  at a positive distance of the origin, the increasing process  $\nu([0, t] \times A)$  is increasing and “compensates” the number of jumps of  $X$  whose size is in  $A$ , in the sense that the difference of these two processes is a (local) martingale. Therefore,  $\nu([0, t] \times (-\infty, -\varepsilon) \cup (\varepsilon, +\infty)) < \infty$ , whereas  $\nu([0, t] \times [-\varepsilon, \varepsilon])$  may be finite or infinite, although we must have  $\int_{\{|x| \leq \varepsilon\}} x^2 \nu([0, t], dx) < \infty$ .

In economic terms, each component of the model can be fairly naturally mapped into an economic source of risk in the underlying asset: the continuous part of the model captures the normal risk of the asset, which is hedgeable using standard differential methods; the big jumps component which can capture default risk, or more generally big news-related events; and the small jumps component can represent price moves which are large on a time scale of a few seconds, but generally not significant on a daily and below sampling frequency. Such jumps may result for example from the limited ability of the marketplace to absorb large transactions without a price impact. That component represents risk that is relevant in particular for trading strategies that are executed at high frequency.

Note that we have compensated the small jumps part, but not the big jumps one. Compensating the big jumps part is not always possible because the moments may not exist, whereas summing small jumps without compensation may lead to a divergent sum. However, when jumps have finite activity, or more generally when they are summable, that is  $\sum_{s \leq t} |\Delta X_s| < \infty$  for all  $t$ , where

$$\Delta X_s = X_s - X_{s-}, \quad (4)$$

is the size of the jump at time  $s$ , it turns out that  $\int_{\{|x| \leq \varepsilon\}} |x| \nu([0, t], dx) < \infty$ . Then compensating the small jumps is not necessary, and we may rewrite (1) as follows:

$$X_t = X_0 + \underbrace{\int_0^t b'_s ds}_{\text{drift}} + \underbrace{\int_0^t \sigma_s dW_s}_{\text{continuous part}} + \sum_{s \leq t} \Delta X_s \quad (5)$$

with a different drift term: namely  $b'_s = b_s - \int_{\{|x| \leq \varepsilon\}} x F_s(dx)$ .

We will assume that the model produces observations that are collected at a discrete sampling interval  $\Delta_n$ : this means in particular that only “regular” sampling schemes are

considered below, although the methodology can be extended to some non-regular sampling scheme, at the expense of – significantly more – mathematical sophistication. There are  $[T/\Delta_n]$  (where  $[x]$  denotes the integer part of the positive real  $x$ ) observed increments of  $X$  on  $[0, T]$ , which are

$$\Delta_i^n X = X_{i\Delta_n} - X_{(i-1)\Delta_n}, \quad (6)$$

to be contrasted with the actual (unobservable) jumps  $\Delta X_s$  of  $X$ , as described in Figure 2.

Our basic methodology consists in constructing realized power variations of these increments, suitably truncated and/or sampled at different frequencies. These realized power variations are defined as follows, where  $p \geq 0$  is any nonnegative real and  $u_n > 0$  is a sequence of truncation levels:

$$B(p, u_n, \Delta_n) = \sum_{i=1}^{[T/\Delta_n]} |\Delta_i^n X|^p 1_{\{|\Delta_i^n X| \leq u_n\}} \quad (7)$$

Throughout,  $T$  is fixed, and asymptotics are all with respect to  $\Delta_n \rightarrow 0$ . Typically the truncation levels  $u_n$  go to 0, and this is usually achieved by taking  $u_n = \alpha \Delta_n^\varpi$  for some constants  $\varpi \in (0, 1/2)$  and  $\alpha > 0$ . Setting  $\varpi < 1/2$  allows us to keep all the increments which mainly contain a Brownian contribution. There will be further restrictions on the rate at which  $u_n \rightarrow 0$ , expressed in the form of restrictions on the choice of  $\varpi$ . In some instances, we do not want to truncate at all and we then write  $B(p, \infty, \Delta_n)$ . Sometimes we will truncate in the other direction, that is retain only the increments larger than  $u$  :

$$U(p, u_n, \Delta_n) = \sum_{i=1}^{[T/\Delta_n]} |\Delta_i^n X|^p 1_{\{|\Delta_i^n X| > u_n\}}. \quad (8)$$

With  $u_n = \alpha \Delta_n^\varpi$  as above, that can allow us to eliminate all the increments from the continuous part of the model. Then obviously

$$U(p, u_n, \Delta_n) = B(p, \infty, \Delta_n) - B(p, u_n, \Delta_n). \quad (9)$$

Finally, we sometimes simply count the number of increments of  $X$ , that is, take the power  $p = 0$

$$U(0, u_n, \Delta_n) = \sum_{i=1}^{[T/\Delta_n]} 1_{\{|\Delta_i^n X| > u_n\}}. \quad (10)$$

We exploit the different asymptotic behavior of the variations  $B(p, u_n, \Delta_n)$  and/or  $U(p, u_n, \Delta_n)$  as we vary: the power  $p$ , the truncation level  $u_n$  and the sampling frequency

$\Delta_n$ . This gives us three degrees of freedom, or tuning parameters, with enough flexibility to isolate what we are looking for. Having these three parameters to play with,  $p$ ,  $u_n$  and  $\Delta_n$ , is akin to having three knobs to adjust in the measurement device.

#### A. *The First Knob: Varying the Power*

The role of the power variable is to isolate either the continuous or jump components, or to keep them both present. As illustrated in Figure 3, powers  $p < 2$  will emphasize the continuous component of the underlying sampled process while powers  $p > 2$  will conversely accentuate its jump component. The power  $p = 2$  (which receives much attention in the form of measuring realized volatility) puts them on an equal footing, which turns out to be useful here only when we seek to measure the relative magnitude of the components.

#### B. *The Second Knob: Varying the Truncation Rate*

Truncating the large increments at a suitably selected cutoff level can eliminate the big jumps when needed. The key is that there is a finite number of large jumps. Asymptotically, as the sampling frequency increases, the cutoff level gets smaller. But the large jumps have a fixed size, so at some point along the asymptotics the cutoff level becomes smaller than the large jumps, which are thus no longer part of the realized power variation  $B(p, u_n, \Delta_n)$ , as illustrated in Figure 4.

Alternatively, we can truncate to eliminate the Brownian component if we use the upwards power variation  $U(p, u_n, \Delta_n)$ , since the continuous component is only capable of generating increments that are smaller than  $u_n = \alpha \Delta_n^\varpi$  when  $\varpi < 1/2$ .

#### C. *The Third Knob: Varying the Sampling Frequency*

Sampling at different frequencies can let us distinguish between the three situations where the variations converge to a finite limit, converge to zero or diverge to infinity. We will achieve this by computing the ratio of two  $B$ 's evaluated at the biggest available frequency  $\Delta_n$  and at the same time at some lower frequency  $k\Delta_n$  where  $k \geq 2$  is an integer. Sampling at frequency  $k\Delta_n$  is obtained from the same data series, simply retaining one out of every

$k$  data points in Figure 2. As described in Figure 5, the limiting behavior of the ratio (1, less than 1 or greater than 1) will identify the underlying limiting behavior of  $B$ .

As we will see, the various limiting behaviors of the variations are indicative of which component of the model dominates at a particular power and in a certain range of returns (by truncation), just like certain chemical elements have a very specific spectrographic signature. So they will effectively allow us to distinguish between all manners of null and alternative hypotheses if we can identify which situation corresponds to which of the spectrographic signatures of  $B$ .

### III. Which Component(s) Are Present

Leaving aside the drift, which is effectively invisible at high frequency, the model (1)-(2) has three components: a continuous part, a small jumps part and a big jumps part. The analogy with spectrography would be that we are looking for the signature of three possible chemical elements (say, hydrogen, helium and everything else) in the light being recorded. Here, based on the observed log-returns, what can we tell about which component(s) of the model are present?

Consider the following sets defined pathwise on  $[0, T]$ :

$$\begin{aligned}
 \Omega_T^c &= \{X \text{ is continuous in } [0, T]\} \\
 \Omega_T^j &= \{X \text{ has jumps in } [0, T]\} \\
 \Omega_T^f &= \{X \text{ has finitely many jumps in } [0, T]\} \\
 \Omega_T^i &= \{X \text{ has infinitely many jumps in } [0, T]\} \\
 \Omega_T^W &= \{X \text{ has a Wiener component in } [0, T]\} \\
 \Omega_T^{\text{no}W} &= \{X \text{ has no Wiener component in } [0, T]\}
 \end{aligned} \tag{11}$$

Formally,  $\Omega_T^W = \left\{ \int_0^T \sigma_s^2 ds > 0 \right\}$  and  $\Omega_T^{\text{no}W} = \left\{ \int_0^T \sigma_s^2 ds = 0 \right\}$ , and the definition of the four other sets is clear.

We observe a time series originating in a given unobserved path, and wish to determine in which set(s) the path is. At any given – fixed – frequency this is a theoretically unanswerable question since for example any such time series can be obtained by discretization of a continuous path, and also of a discontinuous one. However we wish to construct

test statistics that behave well asymptotically, as  $\Delta_n \rightarrow 0$ , and if possible under the only structural assumptions (1)–(2). That is, they should be model-free in the sense that their implementation *and their asymptotic properties* do not require that we specify or calibrate the model, which can potentially be quite complicated (stochastic volatility, jumps, jumps in volatility, jumps in jump intensity, etc.).

It turns out that this aim is achievable, using the power variations introduced above, for some of the problems. For others we need some additional structural assumptions, to be explained later when needed. Let us also mention that for all results one also needs some weak boundedness-like or smoothness-like assumptions on the coefficients, such as the process  $b_t$  should be (locally) bounded: as a rule, these assumptions are not explicitly stated here, and we refer to the original papers for the mathematically precise statements.

#### A. *Jumps: Present or Not*

The first question we address is whether the path of  $X$  contains jumps or not. As discussed in the Introduction, there is by now a vast literature concerned with detecting jumps but we will focus on the approach which lends itself to answering the full range of specification questions listed for semimartingales.

Using the methodology of power variations, we start with two processes which measure some kind of variability of  $X$  and depend on the whole (unobserved) path of  $X$ :

$$A(p) = \int_0^T |\sigma_s|^p ds, \quad B(p) = \sum_{s \leq T} |\Delta X_s|^p \quad (12)$$

where  $p > 0$ . The variable  $A(p)$  is finite for all  $p > 0$ , and positive on the set  $\Omega_T^W$ . The variable  $B(p)$  is finite if  $p \geq 2$  but often not when  $p < 2$ . The quadratic variation of  $X$  is  $[X, X]_T = A(2) + B(2)$ .

Of course, hoping to estimate  $B(p)$  using  $B(p, u_n, \Delta_n)$  is too naive in general, but it works in specific cases. Namely, we have the following behavior of  $B(p, \infty, \Delta_n)$

$$\begin{cases} p > 2, \text{ all } X & \Rightarrow B(p, \infty, \Delta_n) \xrightarrow{\mathbb{P}} B(p) \\ \text{all } p, \text{ on } \Omega_T^c & \Rightarrow \frac{\Delta_n^{1-p/2}}{m_p} B(p, \infty, \Delta_n) \xrightarrow{\mathbb{P}} A(p) \end{cases} \quad (13)$$

where  $m_p$  denotes the  $p$ th absolute moment of the standard normal variable.

So we see that, when  $p > 2$ ,  $B(p, \infty, \Delta_n)$  tends to  $B(p)$ : the jump component dominates. If there are jumps, the limit  $B(p)_t > 0$  is finite. On the other hand when  $X$  is continuous, then the limit is  $B(p) = 0$  and  $B(p, \infty, \Delta_n)_t$  converges to 0 at rate  $\Delta_n^{p/2-1}$ .

These considerations lead us to pick a value of  $p > 2$  and compare  $B(p, \infty, \Delta_n)_t$  on two different sampling frequencies. Specifically, for an integer  $k$ , consider the test statistic  $S_J$ :

$$S_J(p, k, \Delta_n) = \frac{B(p, \infty, k\Delta_n)}{B(p, \infty, \Delta_n)}. \quad (14)$$

The ratio in  $S_J$  exhibits a markedly different behavior depending upon whether  $X$  has jumps or not:

$$S_J(p, k, \Delta_n) \rightarrow \begin{cases} 1 & \text{on } \Omega_T^j \\ k^{p/2-1} & \text{on } \Omega_T^c \cap \Omega_T^W \end{cases} \quad (15)$$

That is, in the context of Figure 5, under  $\Omega_T^j$  the variation converges to a finite limit and so the ratio tends to 1 (the middle situation depicted in the figure) while under  $\Omega_T^c \cap \Omega_T^W$  the variation converges to 0 and the ratio tends to a limit greater than 1, with value specifically depending upon the rate at which the variation tends to 0 (the lower situation depicted in the figure). The notion of a set  $\Omega_T^c \cap \Omega_T^W$  may seem curious at first, but it is possible for a process to have continuous paths without a Brownian component if the process consists only of a pure drift. Because this would be an unrealistic model for financial data, we are excluding the set  $\Omega_T^c \cap \Omega_T^{noW}$  from consideration.

If one desires a formal statistical test of  $\Omega_T^c \cap \Omega_T^W$  vs.  $\Omega_T^j$ , with a prescribed asymptotic level  $\alpha \in (0, 1)$ , one can use a CLT under  $\Omega_T^c \cap \Omega_T^W$  and one under  $\Omega_T^j$ , such CLT being available again in a model-free situation, apart from some additional smoothness assumptions: so one can in fact test either  $H_0 : \Omega_T^c \cap \Omega_T^W$  vs.  $H_1 : \Omega_T^j$  or the reverse  $H_0 : \Omega_T^j$  vs.  $H_1 : \Omega_T^c \cap \Omega_T^W$ . Note that the first limit in (15) is valid on  $\Omega_T^j$  whether the jump component includes finite or infinite components, or both. It is not designed to disentangle the two types of jumps. How to do this is the question we now turn to.

## B. Jumps: Finite or Infinite Activity

Many models in mathematical finance do not include jumps. But among those that do, the framework most often adopted consists of a jump-diffusion: these models include a

drift term, a Brownian-driven continuous part, and a finite activity jump part (compound Poisson process): early examples include Merton (1976), Ball and Torous (1983), Bates (1991) and Duffie, Pan, and Singleton (2000).

Other models are based on infinite activity jumps: see for example Madan and Seneta (1990), Madan and Milne (1991), Eberlein and Keller (1995), Barndorff-Nielsen (1997), Barndorff-Nielsen (1998), Carr, Geman, Madan, and Yor (2002), Carr and Wu (2003a), Carr and Wu (2004) and Schoutens (2003), although with the exception of Carr, Geman, Madan, and Yor (2002) models of this type are justified primarily by their ability to produce interesting pricing formulae rather than necessarily an attempt at empirical realism.

So, which is it, based on the data? Our objective is now to discriminate between finite and infinite activity jumps using again the same set of tools.

### B.1. Null Hypothesis: Finite Activity

We first set the null hypothesis to be finite activity, that is  $H_0 : \Omega_T^f \cap \Omega_T^W$ , whereas the alternative is  $H_1 : \Omega_T^i$ . As in the previous subsection, we rule out the set  $\Omega_T^f \cap \Omega_T^{\text{no}W}$  which, for all models in use in finance, is empty. We choose an integer  $k \geq 2$  and a real  $p > 2$ . The only difference with testing for jumps using  $S_J$  is that we now truncate

$$S_{FA}(p, u_n, k, \Delta_n) = \frac{B(p, u_n, k\Delta_n)}{B(p, u_n, \Delta_n)}. \quad (16)$$

Without truncation, as in  $S_J$ , we could discriminate between jumps and no jumps, but not among different types of jumps. Like before, we set  $p > 2$  to magnify the jump component at the expense of the continuous component. But since we want to separate big and small jumps, we now truncate as a means of eliminating the large jumps. Since the large jumps are of finite size (independent of  $\Delta_n$ ), at some point in the asymptotics the truncation level  $u_n = \alpha\Delta_n^{\varpi}$  will have eliminated all the large jumps: see Figure 4 earlier. Then if there are only big jumps and the Brownian component, the two truncated power variations  $B(p, u_n, k\Delta_n)$  and  $B(p, u_n, \Delta_n)$  will behave as if there were no jumps, leaving only the Brownian component. The limit of the ratio will be  $k^{p/2-1}$  as in the test for jumps when there are no jumps.

But if there are infinitely many jumps, which are necessarily small, then the truncation

cannot eliminate them. This is because however small  $u_n$  is, there are still infinitely many jumps in each  $\Delta_n$ -increment. The Brownian component is dominated in every increment by the small jumps because  $p > 2$ . Both  $B(p, u_n, k\Delta_n)$  and  $B(p, u_n, \Delta_n)$  behave like the sum of the  $p^{\text{th}}$  power of the jumps that are smaller than  $u_n$ , and although they both go to 0, their ratio tends to 1. In the context of Figure 5, we are in the limiting case where both  $B$ 's go to zero but at the same rate: hence the ratio is 1.

That is, we have:

$$S_{FA}(p, u_n, k, \Delta_n) \xrightarrow{\mathbb{P}} \begin{cases} k^{p/2-1} & \text{on } \Omega_T^f \cap \Omega_T^W. \\ 1 & \text{on } \Omega_T^i. \end{cases} \quad (17)$$

## B.2. Null Hypothesis: Infinite Activity

We next set the null hypothesis to be infinite activity, that is  $H_0 : \Omega_T^i$ , whereas the alternative is  $H_1 : \Omega_T^f \cap \Omega_T^W$ . We need a different statistic,  $S_{IA}$ , because although  $S_{FA}$  goes to 1, the distribution of  $S_{FA}$  is not model-free under  $\Omega_T^i$ . The problem comes from the fact that the behavior of the truncated power variations  $B(p, u_n, \Delta_n)$  depend on the degree of activity of the jumps when there are infinitely many jumps. So we need to specify what we precisely mean by “degree of activity”.

To this end, recalling the definition of  $B(p)$  given in (12), we consider now the set  $I_T = \{p \geq 0 : B(p) < \infty\}$ . This (random) set  $I_T$  is of the form  $[\beta_T, \infty)$  or  $(\beta_T, \infty)$  for some  $\beta_T(\omega) \in [0, 2]$ , and  $2 \in I_T$  always. It turns out that  $\beta_T(\omega)$ , the lower bound of the set  $I_T$ , is a sensible measure of jump activity for the path  $t \mapsto X_t(\omega)$  at time  $T$ . In the special case where  $X$  is a Lévy process, then  $\beta_T(\omega) = \beta$  does not depend on  $(\omega, T)$ , and it is also the infimum of all  $r \geq 0$  such that  $\int_{\{|x| \leq 1\}} |x|^r F(dx) < \infty$ , where  $F$  is the Lévy measure, and this number has been introduced by Blumenthal and Gettoor (1961) and by extension we call  $\beta_T$  the (generalized) Blumenthal-Gettoor index, or degree of jump activity, of the process.

In other words the degree of jump activity measure the rate at which the jump measure diverge near 0, so it characterizes the concentration of small jumps. Many examples of models proposed in finance for asset returns fall in this category, with either fixed values of  $\beta$  or  $\beta$  being a free parameter. (We will discuss estimating  $\beta$  below.) Examples are included

in Figure 7. They include compound Poisson-based models starting with Merton (1976), the variance gamma model of Madan and Seneta (1990) and Madan, Carr, and Chang (1998) ( $\beta = 0$ ), the Normal Inverse Gaussian model of Barndorff-Nielsen (1998) ( $\beta = 1$ ), the hyperbolic model of Eberlein and Keller (1995), the generalized hyperbolic model of Barndorff-Nielsen (1977) and the CGMY model of Carr, Geman, Madan, and Yor (2002) (in which  $\beta$  is a free parameter).

A priori the degree of jump activity can be random and depend on time, but we assume for tractability that this index is in fact constant in time and non-random as is the case in all known examples. More precisely, we assume that the Lévy measure  $\nu$  in (2) is of the form

$$\nu(dt, dx) = \frac{1}{|x|^{1+\beta}} \left( a_t^+ 1_{(0, z_t^+]}(x) + a_t^- 1_{[-z_t^-, 0)}(x) \right) + \nu'(dt, dx), \quad (18)$$

where  $a_t^\pm$  are nonnegative and  $z_t^\pm$  are positive stochastic processes, and  $\nu'$  is another Lévy measure whose index is smaller than  $\beta$ . Note that the assumption (18) is only about the local behavior of the jump measure  $\nu$  near 0, that is, only about the behavior of the small jumps. The big jumps, controlled by  $\nu'$ , are unrestricted. The processes  $a_t^\pm$  are intensity parameters: as they go up, there are more and more small jumps. The processes  $z_t^\pm$  control the range of returns over which the behavior of the overall jump measure is stable-like with index  $\beta$ .

Note that, necessarily,  $\beta \in (0, 2)$  here, otherwise (18) would not be a Lévy measure. Then, if further  $\int_0^T (a_s^+ + a_s^-) ds > 0$ , the number  $\beta$  is the index of  $X$  on the full interval  $[0, T]$ . Note that when  $X$  is a (possibly asymmetric) stable process, that is a process whose jump measure is proportional to  $1/|x|^{1+\beta}$ , then it satisfies this assumption,  $\beta$  being the index of the stable process. In fact, this assumption amounts to saying that the small jumps of  $X$  behave like the small jumps of a stable or tempered stable process, or more accurately as those of a process which is a stochastic integral with respect to a stable or tempered stable process, whereas the big jumps are governed by  $\nu'$ . We call processes which satisfy (18) “proto-stable” processes. Most models in finance which exhibit jumps of infinite activity are proto-stable. While we will propose estimators of  $\beta$  below, the true  $\beta$  is of course unknown, and our model-free requirement means here that we wish to construct a test which does not depend upon  $\beta$ , the processes  $a_t^\pm$  or  $z_t^\pm$ , nor the residual jump measure  $\nu'$ .

Coming back to our problem, we consider the set

$$\Omega_T^{i\beta} = \left\{ \int_0^T (a_s^+ + a_s^-) ds > 0 \right\}$$

on which the jump activity index of  $X$  equals  $\beta$ . Note that  $\Omega_T^{i\beta} \subset \Omega_T^i$ , the inclusion possibly being strict. However testing the null being  $\Omega_T^i$  is impossible without further restriction, and so we set the null to be  $\Omega_T^{i\beta}$ .

We choose three reals  $\gamma > 1$  and  $p' > p > 2$  and define a family of test statistics as follows:

$$S_{IA}(p, u_n, \gamma, \Delta_n) = \frac{B(p', \gamma u_n, \Delta_n) B(p, u_n, \Delta_n)}{B(p', u_n, \Delta_n) B(p, \gamma u_n, \Delta_n)} \quad (19)$$

which has the following limits:

$$S_{IA}(p, u_n, \gamma, \Delta_n) \xrightarrow{\mathbb{P}} \begin{cases} \gamma^{p'-p} & \text{on } \Omega_T^{i\beta} \\ 1 & \text{on } \Omega_T^f \cap \Omega_T^W \end{cases} \quad (20)$$

Intuitively, under the alternative of finite jump activity, the behavior of each one of the four truncated power variations in (19) is driven by the continuous part of the semimartingale. The truncation level is such that essentially all the Brownian increments are kept. Then the truncated power variations all tend to zero at rates  $\Delta_n^{p/2-1}$  and  $\Delta_n^{p'/2-1}$  respectively and by construction the (random) constants of proportionality cancel out in the ratios, producing a limit 1 given under  $H_1$  in (20).

If, on the other hand, jumps have infinite activity, then the small jumps are the ones that matter and the truncation level becomes material, producing four terms that all tend to zero but at the different orders in probability  $u_n^{p-\beta}$ ,  $u_n^{p'-\beta}$ ,  $(\gamma u_n)^{p-\beta}$  and  $(\gamma u_n)^{p'-\beta}$  respectively, resulting in the limit  $\gamma^{p'-p}$  given under  $H_0$  in (20). By design, that limit in  $S_{IA}$  is independent of  $\beta$ .

### C. *Brownian Motion: Present or Not*

We now would like to construct procedures which allow to decide whether the Brownian motion is really there, or if it can be forgone with in favor of a pure jump process with infinite activity. When infinitely many jumps are included, there are a number of models in the literature which dispense with the Brownian motion altogether. The log-price process

is then a purely discontinuous Lévy process with infinite activity jumps, or more generally is driven by such a process. Is this a realistic model in light of the data?

### C.1. Null Hypothesis: Brownian Motion Present

In order to construct a test, we seek a statistic with markedly different behavior under the null and alternative. Using the same class of tools, the idea is now to consider powers  $p$  less than 2, since in the presence of Brownian motion the power variation would be dominated by it while in its absence it would behave quite differently. Specifically, the large number of small increments generated by a continuous component would cause a power variation of order less than 2 to diverge to infinity: recall Figure 5.

Without the Brownian motion, however, and when  $p$  is bigger than the Blumenthal-Gettoor index  $\beta_T = \beta$ , assuming the structural assumption (18), the power variation converges to 0 at exactly the same rate for the two sampling frequencies  $\Delta_n$  and  $k\Delta_n$ , whereas with a Brownian motion the choice of sampling frequency will influence the magnitude of the divergence. Taking a ratio will eliminate all unnecessary aspects of the problem and focus on that key aspect.

So we choose an integer  $k \geq 2$  and a real  $p < 2$  and propose the test statistic

$$S_W(p, u_n, k, \Delta_n) = \frac{B(p, u_n, \Delta_n)}{B(p, u_n, k\Delta_n)} \quad (21)$$

which has the limits

$$S_W(p, u_n, k, \Delta_n) \xrightarrow{\mathbb{P}} \begin{cases} k^{1-p/2} & \text{on } \Omega_T^W \\ 1 & \text{on } \Omega_T^{\text{no}W} \cap \Omega_T^\beta, p > \beta \end{cases} \quad (22)$$

Note that the first convergence above, on  $\Omega_T^W$ , does not require any specific assumptions on the jumps, only the second convergence requires (18).

### C.2. Null Hypothesis: No Brownian Motion

When there are no jumps, or finitely many jumps, and no Brownian motion,  $X$  reduces to a pure drift plus occasional jumps, and such a model is fairly unrealistic in the context of most financial data series. But one can certainly consider models that consist only of a

jump component, plus perhaps a drift, if that jump component is allowed to be infinitely active. If one wishes to set the null model to be a pure jump model (plus perhaps a drift), then the issue becomes to design a test statistic using power variations whose behavior is independent of the specific nature of the infinitely active pure jump process. In other words, we again assume (18), but we do not know  $\beta$  and wish to design a test that remains model-free in the sense that it does not depend on  $\beta$ ,  $a_t^\pm$  or  $z_t^\pm$  in (18).

We choose a real  $\gamma > 1$  to define two different truncation ratios and define a family of test statistics as follows:

$$S_{\text{no}W}(p, u_n, \gamma, \Delta_n) = \frac{B(2, \gamma u_n, \Delta_n) U(0, u_n, \Delta_n)}{B(2, u_n, \Delta_n) U(0, \gamma u_n, \Delta_n)}. \quad (23)$$

To understand the construction of this test statistic, recall that in a power variation of order 2 the contributions from the Brownian and jump components are of the same order. If the Brownian motion is present ( $H_1 : \Omega_T^W$ ) then once that power variation is properly truncated, the Brownian motion will dominate it if it is present. And the truncation can be chosen to be sufficiently loose that it retains essentially all the increments of the Brownian motion at cutoff level  $u_n$  and a fortiori  $\gamma u_n$ , thereby making the ratio of the two truncated quadratic variations converge to 1 under the alternative hypothesis.

If on the other hand the Brownian motion is not present ( $H_0 : \Omega_T^{\text{no}W} \cap \Omega_T^\beta$ ), then the nature of the tail of jump distributions is such that the difference in cutoff levels between  $u_n$  and  $\gamma u_n$  remains material no matter how far we go in the tail and the limit of the ratio  $B(2, \gamma u_n, \Delta_n)/B(2, u_n, \Delta_n)$  in (23) will reflect it: it will now be  $\gamma^{2-\beta}$ . But since absence of a Brownian motion is now the null hypothesis, the issue for constructing a test is that this limit depends on the unknown  $\beta$ .

Canceling out that dependence is the role devoted to the ratio  $U(0, u_n, \Delta_n)/U(0, \gamma u_n, \Delta_n)$  of the number of large increments. The  $U$ 's are always dominated by the jump components of the model whether the Brownian motion is present or not. Their inclusion in the statistic is merely to ensure that the statistic is model-free, by effectively canceling out the dependence on the jump characteristics that emerges from the ratio of the truncated quadratic variations.

Indeed, the limit of the ratio of the  $U$ 's is  $\gamma^\beta$  under both the null and alternative hypotheses. As a result, the probability limit of  $S_{\text{no}W}$  will be  $\gamma^2$  under the null, independent

of  $\beta$ :

$$S_{\text{no}W}(p, u_n, \gamma, \Delta_n) \xrightarrow{\mathbb{P}} \begin{cases} \gamma^2 & \text{on } \Omega_T^{\text{no}W} \cap \Omega_T^\beta \\ \gamma^\beta & \text{on } \Omega_T^W \end{cases} \quad (24)$$

Generally speaking, the statistic  $S_W$  is more robust than  $S_{\text{no}W}$ ; similarly  $S_{FA}$  is more robust than  $S_{IA}$ . This is due to their simpler design, and the lesser reliance on subtle cancellations to achieve their respective objectives. As a result, we recommend using  $S_{FA}$  and  $S_W$  in practical applications.

## IV. The Relative Magnitude of the Components

A typical “main sequence” star might be made of 90% hydrogen, 10% helium and 0.1% everything else. In astrophysics, a natural metric to compare different atoms and address the question of percentages of various components is atomic mass. Here, what is the relative magnitude of the two jump and continuous components? We can answer this question using the same power variation devices. The natural metric is now to consider  $p = 2$  since this is the power where all the components are present together, instead of powers  $p > 2$  or  $p < 2$  that eliminate one or the other of the components, and ask the question of percentages of total quadratic variation (QV) attributable to each component.

As illustrated in Figure 6, by using truncations at the right rate we can split the QV into its continuous and jump components, and not truncate to estimate the full QV:

$$\begin{cases} \frac{B(2, u_n, \Delta_n)}{B(2, \infty, \Delta_n)} = \% \text{QV due to the continuous component} \\ 1 - \frac{B(2, u_n, \Delta_n)}{B(2, \infty, \Delta_n)} = \% \text{QV due to the jump component} \end{cases} \quad (25)$$

The use of truncation to estimate the continuous part of the quadratic variation when there are jumps was proposed by Mancini (2001), who relied on the law of the iterated logarithm for that purpose. Alternatively, one can split the QV based on bipower variations instead of truncating: see Barndorff-Nielsen and Shephard (2004), Huang and Tauchen (2005) and Andersen, Bollerslev, and Diebold (2007).

Note that (25) suggests that an alternative test for the presence of jumps can be constructed based on the ratio  $B(2, u_n, \Delta_n)/B(2, \infty, \Delta_n)$ . However, this would work only if the

null hypothesis is that no jumps are present, and the null hypothesis is that the ratio is 1. With jumps under the null, one would have to specify exogenously as part of the null hypothesis how large the fraction of QV due to jumps is.

We can split the rest of the QV, which by construction is attributable to jumps, into a small jumps and a big jumps component. This depends on the cutoff level  $\varepsilon$  selected to distinguish big and small jumps:

$$\left\{ \begin{array}{l} \frac{U(2,\varepsilon,\Delta_n)}{B(2,\infty,\Delta_n)} = \%QV \text{ due to big jumps} \\ \frac{B(2,\infty,\Delta_n) - B(2,u_n,\Delta_n) - U(2,\varepsilon,\Delta_n)}{B(2,\infty,\Delta_n)} = \%QV \text{ due to small jumps} \end{array} \right. \quad (26)$$

We can then obtain a plot that looks like Figure 8 and provides a split of the QV into the various components.

## V. Estimating the Degree of Jump Activity

The method described in Section B is able to tell finite activity jumps from infinite activity ones. Among jump processes, however, finite activity are the exception rather than the norm. And “infinite activity” can mean quite different things depending upon “how infinite” that infinite jump activity is. In fact, the degree of activity is accurately measured by the Blumenthal-Gettoor index  $\beta_T$  introduced earlier: at one end of the spectrum, infinite activity jump processes such as the Gamma process, whose jump measure diverge at a sub-polynomial rate, can look like Poisson jumps; at the other end, they can look almost like Brownian motion, which is to say extremely active. So it seems natural to try to estimate the index  $\beta_T$ . As discussed above, specific models in finance correspond either to fixed values of  $\beta$  (such as  $\beta = 0$  for the Gamma and Variance Gamma models,  $\beta = 1/2$  for the Lévy model and the Inverse Gaussian model,  $\beta = 1$  for the Cauchy model and the Normal Inverse Gaussian Process) or  $\beta$  is a free parameter (as in the stable model, the Generalized Hyperbolic model and the CGMY model).

The next issue is then to estimate  $\beta_T$ , or rather  $\beta$  under the somewhat restricted assumption (18). The problem is made more challenging by the potential presence in  $X$  of a continuous, or Brownian, martingale part.  $\beta$  characterizes the behavior of  $\nu$  near 0. Hence it is natural to expect that the small increments of the process are going to be the ones

that are most informative about  $\beta$ . But that is where the contribution from the continuous martingale part of the process is inexorably mixed with the contribution from the small jumps. In other words, we need to see through the continuous part of the semimartingale in order to say something about the number and concentration of small jumps.

So we are now looking in a different range of the spectrum of returns, namely by considering only returns that are larger than the cutoff  $u_n = \alpha \Delta_n^\varpi$  for some  $\varpi \in (0, 1/2)$ , as opposed to those that are smaller than the cutoff. This allows us to eliminate the increments due to the continuous component. We can then use all values of  $p$ , not just those  $p > 2$ , despite the fact that we wish to concentrate on jumps: see Figure 6. In fact, we will simply use the power  $p = 0$ .

We propose two estimators of  $\beta$  based on counting the number of increments greater than the cutoff  $u_n$ . The first one is based on varying the actual cutoff level: fix  $0 < \alpha < \alpha'$  and consider two cutoffs  $u_n = \alpha \Delta_n^\varpi$  and  $u'_n = \alpha' \Delta_n^\varpi$  with  $\gamma = \alpha'/\alpha$ :

$$\widehat{\beta}_n(\varpi, \alpha, \alpha') = \frac{\log(U(0, u_n, \Delta_n)/U(0, \gamma u_n, \Delta_n))}{\log(\gamma)}, \quad (27)$$

The second one is based on varying the sampling frequency: sample at two time scales,  $\Delta_n$  and  $2\Delta_n$ :

$$\widehat{\beta}'_n(\varpi, \alpha, k) = \frac{\log(U(0, u_n, \Delta_n)/U(0, u_n, k\Delta_n))}{\varpi \log k}. \quad (28)$$

These estimators are consistent for  $\beta$ , and we have derived CLTs for them.

These basic estimators are based on the first-order asymptotics

$$U(0, u_n, \Delta_n) \sim b_0 \Delta_n^{-\varpi\beta} \frac{1}{\alpha^\beta} \quad (29)$$

where  $b_0$  is independent of  $u_n$  and  $\Delta_n$ . In small samples, a bias corrected procedure is based on the second-order asymptotics

$$U(0, u_n, \Delta_n) \sim b_0 \Delta_n^{-\varpi\beta} \frac{1}{\alpha^\beta} + b_1 \Delta_n^{1-2\varpi\beta} \frac{1}{\alpha^{2\beta}} \quad (30)$$

works as follows: we can estimate  $\beta$ , along with the unknown coefficients  $b_0$  and  $b_1$  in (30) by a straightforward nonlinear regression of  $U(0, u_n, \Delta_n)$  on  $\alpha$  by varying  $\alpha$  in the cutoff  $u_n = \alpha \Delta_n^\varpi$ .

One can then test various hypotheses involving  $\beta$ . The approach described in this Section to estimate  $\beta$  is due to Aït-Sahalia and Jacod (2009a). Related approaches include Woerner (2006), who proposes an estimator of the jump activity index in the case of fractionally integrated processes, Cont and Mancini (2009), who are testing whether  $\beta > 1$  or  $\beta < 1$ , which correspond to finite or infinite variation for  $X$ , and Todorov and Tauchen (2010), who provide a graphical method to determine whether  $\beta = 2$  or  $\beta < 2$  using the test statistic of Aït-Sahalia and Jacod (2009b), and Belomestny (2009) who proposes a method based on low frequency historical and options data.

## **VI. Summary of the Spectrogram Methodology: Tuning Power, Truncation and Sampling Frequency**

We have seen that setting the three knobs of power, truncation level and sampling frequency in various combinations allowed us to determine which component of the model was likely to be present, in what proportion, and estimate the degree of activity of the jumps. Table I summarizes the choice of the three tuning parameters  $(p, u, \Delta)$  for the corresponding tasks under consideration.

In a nutshell, we address specification questions that require an emphasis on the jump component of the model with powers  $p > 2$ , those that require an emphasis on the continuous component with powers  $p < 2$ , and those that require them on an equal footing with the singular power  $p = 2$ . Truncating makes it possible to eliminate either the big jumps or the Brownian component, as necessary. And finally sampling at different frequencies allows us to identify the asymptotic behavior of the relevant power variations, thereby discriminating between components of the model that are present or absent in the sampled data.

## **VII. Theoretical Limits When Market Microstructure Noise Dominates**

We consider in the empirical analysis that follows sampling frequencies up to 5 seconds. In different applications, this selection is going to be asset-dependent, as a function of the

assets' liquidity and other trading characteristics. But in any event, real data observations of the process  $X$  at such ultra high frequencies are blurred by market microstructure noise, which has the potential to change the asymptotic behavior of many statistics at very high frequency, and can force us to downsample as is often done in the classical volatility estimation setting.

When observations are affected by an additive noise, then instead of  $X_{i\Delta_n}$  we observe  $Y_{i\Delta_n} = X_{i\Delta_n} + \varepsilon_i$ , and the  $\varepsilon_i$  are i.i.d. with  $E(\varepsilon_i^2)$  and  $E(\varepsilon_i^4)$  finite, and not depending of the observation frequency. When rounding is introduced, we observe  $Y_{i\Delta_n} = [X_{i\Delta_n}]_a$  which is  $X$  rounded to the nearest multiple of  $a$ , say 1 cent for a decimalized asset, or for many bond markets,  $a = 1/32^{\text{nd}}$  of a dollar: again the rounding level  $a$  does not depend on  $n$ . As a matter of fact, the real microstructure noise is probably a mixture of the two types above, first an additive noise (or perhaps a “colored” additive noise) which may account for some bouncebacks, and then the noisy price is rounded at the level  $a$ .<sup>3</sup>

The power variations that form the building blocks of our methodology are affected by either type of noise, in a rather drastic way, since the presence of noise modifies the limit in probability of most of our statistics, not to speak about their second order behavior like CLTs. In order to be able to interpret the empirical results, we need to extend the existing theory by explicitly incorporating the noise into the probability limits of the various statistics. As discussed, we will consider in turn the two polar cases of a pure additive noise, and of a pure rounding noise.

---

<sup>3</sup>While far from being complete descriptions of the reality, these two specifications for the noise can be thought of as proxies for some of the main features identified as relevant in the market microstructure literature: see e.g., Hasbrouck (1993), who discusses the theoretical market microstructure underpinnings of an additive noise model and argues that the standard deviation of the noise,  $E(\varepsilon_i^2)^{1/2}$ , is a summary measure of market quality. In the Roll (1984) model, the noise is due entirely to the bid-ask spread. Harris (1990b) considers additional sources of noise and their impact on the Roll model and its estimators. More complex structural models, such as Madhavan, Richardson, and Roomans (1997), also give rise to reduced forms where the observed transaction price takes the form of an unobserved price plus noise. Adverse selection effects are considered in Glosten (1987) and Glosten and Harris (1988), where the spread has different components. Especially when asymmetric information is involved, the noise term may no longer satisfy the basic assumptions here (such as i.i.d. or uncorrelatedness with the price process). The second case we consider, where the noise is due to rounding, has been analyzed in the market microstructure literature (see e.g., Gottlieb and Kalay (1985)). The specification of the model in Harris (1990a) combines both rounding and bid-ask effects as the dual sources of noise.

### A. Additive noise

Pure additive noise has been considered by many authors for the purpose of estimating the volatility, but its effect on power variations, truncated or not, has not been thoroughly analyzed so far, to the best of our knowledge. So, in contrast with the style of the rest of the paper, we now present theorems. To avoid intricate statements, we make two basic, and mild, assumptions on the noise, which is a sequence  $(\varepsilon_i)_{i \geq 0}$  of i.i.d. variables, independent of the underlying process. Namely we assume that

$$m'_p = \mathbb{E}(|\varepsilon_i|^p) \text{ is finite for all } p > 0, \text{ and the noise is centered: } \mathbb{E}(\varepsilon_i) = 0. \quad (31)$$

and

$$\text{the variables } \varepsilon_i - \varepsilon_{i-1} \text{ have a density } f \text{ which is continuous and positive at } 0. \quad (32)$$

We also write  $B'(p, u_n, \Delta_n)$  and  $U'(p, u_n, \Delta_n)$  for the variables introduced in (7), (8) and (10), where the increments  $\Delta_i^n X$  of (6) are substituted with

$$(X_{i\Delta_n} + \varepsilon_i) - (X_{(i-1)\Delta_n} + \varepsilon_{i-1}). \quad (33)$$

In other words, in the case of additive noise, the observations on the process,  $X_{i\Delta_n}$ , are replaced by  $X_{i\Delta_n} + \varepsilon_i$ . The limits that are relevant for the spectrogram analysis are given by the following result (recall that  $B'(p, \infty, \Delta_n) = U'(p, 0, \Delta_n)$ ):

**Theorem 1.** *Assume (31) and (32).*

a) *If  $p \geq 0$ , then for any sequence  $u_n$  of nonnegative numbers going to 0 we have*

$$\Delta_n U'(p, u_n, \Delta_n) \xrightarrow{\mathbb{P}} m'_p T. \quad (34)$$

b) *If  $p > 0$  and if further*

$$\frac{\Delta_n}{u_n^r} \rightarrow 0, \quad \text{where } r = \begin{cases} \frac{2p+2}{p} & \text{if } p \leq 1 \\ 4 & \text{if } 1 \leq p \leq 3 \\ p+1 & \text{if } p \geq 3 \end{cases} \quad (35)$$

*then we have*

$$\frac{\Delta_n}{u_n^{p+1}} B'(p, u_n, \Delta_n) \xrightarrow{\mathbb{P}} \frac{2f(0)}{p+1} T. \quad (36)$$

The proof is given in the appendix.

## B. Rounding noise

The situation of rounding noise is much more delicate to analyze. There, instead of observing  $X_{i\Delta_n}$ , we observe  $[X_{i\Delta_n}]_\alpha$ , that is  $X_{i\Delta_n}$  rounded to the nearest multiple of  $\alpha$ . For instance, if the market under consideration is decimalized, then  $\alpha = 1$  cent. For many bond markets,  $\alpha = 1/32^{\text{nd}}$  of a dollar.

If  $X$  is continuous, then most of the increments  $\Delta_i^n X$  vanish, and all are integral multiples of the rounding value  $\alpha$ . So the key role is played by the upcrossings and downcrossings of the levels  $(q + 1/2)\alpha$  by  $X$ , for all integers  $q$ . So, not surprisingly, the limit of the variations involve the local times of  $X$  at those levels, that is the amount of time that the process spends in a neighborhood of the crossing levels. The same is true when  $X$  jumps, except that in this case the theory of local times is not even well established unless the Blumenthal-Gettoor index of the jump is smaller than 1.

More specifically, only the case when  $X$  is continuous is completely determined, and the following result is due to Delattre (1997). Below,  $L_t^a$  denotes the local time of  $X$  at level  $a \in \mathbb{R}$  (see the Appendix for a precise definition of the local time), and as before we denote by  $B''(p, u_n, \Delta_n)$  and  $U''(p, u_n, \Delta_n)$  for the variables introduced in (7), (8) and (10), where the increments  $\Delta_i^n X$  of (6) are substituted with

$$[X_{i\Delta_n}]_\alpha - [X_{(i-1)\Delta_n}]_\alpha. \quad (37)$$

**Theorem 2.** *If  $X$  is a continuous semimartingale with the drift process locally bounded, as well as the volatility process  $\sigma_t$  and its inverse  $1/\sigma_t$ , then*

$$\sqrt{\Delta_n} B''(p, \infty, \Delta_n) \xrightarrow{\mathbb{P}} \alpha^p \sqrt{\frac{2}{\pi}} \sum_{q \in \mathbb{Z}} \int_0^T \frac{1}{\sigma_s} dL_s^{(q+1/2)\alpha}. \quad (38)$$

The same still holds if the semimartingale is discontinuous, but with finite activity: we will show this in the Appendix. Otherwise, no precise mathematical result is known so far in this domain.

The behavior of the truncated power variations is quite different in the presence of rounding. Indeed, if  $u_n \rightarrow 0$ , for  $n$  large enough we have  $u_n < \alpha$  since the rounding level is fixed. In this case all increments of the noisy process are either 0 or bigger than  $u_n$ . In

other words, we have for  $n$  large enough:

$$B''(p, u_n, \Delta_n) = 0, \quad U''(p, u_n, \Delta_n) = B''(p, \infty, \Delta_n). \quad (39)$$

### C. Limit of our statistics

If we gather the previous results together with those for the observations without noise, we see that  $S_J$  has four possible limits:

$$\left\{ \begin{array}{ll} 1/k & : \quad \text{additive noise dominates} \\ 1/k^{1/2} & : \quad \text{rounding error dominates (and jumps have finite activity)} \\ 1 & : \quad \text{jumps present and no significant noise} \\ k^{p/2-1} & : \quad \text{no jumps present and no significant noise} \end{array} \right. . \quad (40)$$

The impact of the noise on  $S_{FA}$  is given by (where “no limit” means that for  $n$  large the statistics is the ratio 0/0):

$$\left\{ \begin{array}{ll} 1/k & : \quad \text{additive noise dominates} \\ \text{no limit} & : \quad \text{rounding error dominates} \\ 1 & : \quad \text{infinite activity jumps and no significant noise} \\ k^{p/2-1} & : \quad \text{finite activity jumps and no significant noise} \end{array} \right. . \quad (41)$$

$S_W$ , taking market microstructure noise into account, has four possible limits:

$$\left\{ \begin{array}{ll} 1/k & : \quad \text{additive noise dominates} \\ \text{no limit} & : \quad \text{rounding error dominates} \\ 1 & : \quad \text{No Brownian motion and no significant noise} \\ k^{1-p/2} & : \quad \text{Brownian motion present and no significant noise} \end{array} \right. . \quad (42)$$

Taking microstructure noise into account yields the following limits for our measure of the fraction of QV due to the continuous component:

$$\left\{ \begin{array}{ll} 0 & : \quad \text{additive noise dominates} \\ 0 & : \quad \text{rounding error dominates} \\ \text{actual fraction of QV} & : \quad \text{no significant noise} \end{array} \right. . \quad (43)$$

We will now analyze the data in light of these predictions of the theory.

## VIII. The Data

### A. *The Starting Data*

We use a dataset consisting of all transactions and quotes on all thirty individual components of the Dow Jones Industrial Average in 2006. The data source is the TAQ database. Using the correction variables in the dataset, we retain only transactions that are labeled “good trades” by the exchanges: regular trades that were not corrected, changed, or signified as canceled or in error; and original trades which were later corrected, in which case the trade record contains the corrected data for the trade. Beyond that, no further adjustment to the raw data are made to produce what we call the “unfiltered transactions.” We also consider and use the corresponding quotes, as we explain below.

We sample the price series at up to the 5 second frequency, and then lower frequency multiples. The most liquid stocks in the DJIA trade on average more than once every second, so we are not retaining every transaction by doing so, which avoids incorporating every bid-ask bounce. For the less liquid stocks, some smoothing is involved. When no transaction is available at the exact time stamp, we use the closest one available. When more than one transaction is available at the same time stamp, we average them. We do not include the overnight returns.

### B. *Bouncebacks and National Best Bid and Offer (NBBO) Filter*

Different measurements of the stock returns lead to different properties of the constructed price process. One particular issue that deserves careful attention in the data is that of “bouncebacks.” We call “bouncebacks” price observations that are either higher or lower than the sequences of prices that both immediately precede and follow it. Such prices generate a log-return from one transaction to the next that is large in magnitude, and is followed immediately by a log-return of the same magnitude but of the opposite sign, so that the price returns to its starting level before that particular transaction. To the extent that we have no reason to believe that those transactions did not actually take place, as we already eliminate transactions known to TAQ to be incorrect, we start with the premise that bouncebacks should not be arbitrarily removed from the sample. However one may

think that bouncebacks, although significant in a sense, should not be incorporated in the model for the “true” latent price process, whatever this “true” might mean.

The prevalence of bouncebacks can lead to a large number of relatively small jumps in the raw data, and bias the empirical results towards finding more small jumps than actually happen if the data were correctly measured, or biasing the estimated degree of jump activity. By contrast, a true jump can be followed by another jump (due to the prevalence of jump clustering in the data), but these successive jumps will not necessarily be of the same magnitude and of the opposite sign.

One straightforward approach to eliminate bouncebacks would be to eliminate all log-returns that are followed immediately by another log-return of the opposite sign, when they are both greater than a pre-determined magnitude, such as some number of ticks. There is however typically in the data a continuum of bouncebacks in terms of such magnitude, so this approach ends up creating a discontinuity at the arbitrary pre-determined magnitude selected: many of them of size less than that level and then none. On the other hand, setting that level within one tick would be extreme and would change the nature of the observed prices.

To deal with bouncebacks endogenously, we will instead make use of the matched quotes data. Transactions that take place outside the currently prevailing quotes are known as “out-trades.” In our context, a single out-trade will generate a bounceback. As discussed by Stoll and Schenzler (2006), they tend to occur on Nasdaq because of delays in reporting trades; because of the ability of dealers to delay the execution of trades, thereby creating a look-back option which when exercised results in out-trades; and because large trades can take place at prices outside the quotes. Out-trades are less frequently observed on the NYSE because the market is more centralized. Of the 30 DJIA components in 2006, Microsoft and Intel trade on Nasdaq while the other components all trade on the NYSE.

We will use the quotes data in order to reduce the incidence rate of bouncebacks in the transactions data, in a manner that is compatible with market rules. SEC regulations require brokers to guarantee customers the best available ask price when buying securities, and the best available bid price when selling securities. These are called the National Best Bid and Offer (NBBO) prices.

We compute the NBBO at each point in time as follows. We collect all quotes at each instant retaining only those with positive sizes and prices, and eliminating crossed quotes, those where the bid price is greater than the ask price. We then compute trimmed size-weighted means (eliminating the extreme 10% of quotes at each end, in price levels and sizes) of the bid and ask prices in effect at that time. Eliminating quotes with small depths or extreme prices is important: bad or erroneous quotes tends to have smaller size, hence the need to trim the weighted average to smooth them away. Quotes with small depths or extreme bid and ask prices (“off-market”) are often used by market makers to effectively pull out of the market, usually temporarily, while still posting quotes. Incidentally, bouncebacks can happen in quotes data as well. But they tend to appear when there is only a very small number of quotes at that point in time, with one or more that is off-market for the reasons just described. Quote bouncebacks seem to be unrelated to transaction bouncebacks.

Given our computed NBBO at each point in time, we then take a moving window of 90 seconds. For instance, a block trade might have been arranged by a manual broker and its reporting delayed. Or opening trades done manually can be delayed, for instance, even at small sizes. The 90-second window we employ is set to reflect the SEC rules that specify that exchanges must report trades within 90 seconds. Trades that are delayed beyond 90 seconds are marked as “late,” and already excluded from the starting data by our TAQ filters. We use this 90-second moving window to construct a running minimum of the national best bid prices and a running maximum of the national best offer prices over the time window. This NBBO bid-offer moving envelope is then used as our filter for transactions: we retain only transactions that take place inside the envelope.

There are many reasons for trades to be delayed, especially when some form of manual execution is involved. For example, in the case of negotiated trades, brokers might work the order over time, leading to a sequence of smaller trades reports. Or the broker might (acting as principal) sell the whole amount to the customer, in which case we would see a single trade report. Another practice involves “stopping” the order: the broker does not execute the order immediately, but does guarantee the buyer a price. The broker can then work the order, deferring any trade reports until the execution is complete, at which time an average price is reported. The average price can appear out of line with the prices around it and lead to a bounceback.

The NBBO filter we employ tends to be conservative – erring on the side of retaining transactions – since there is no guarantee that an out-trade is actually necessarily due to a time delay, or that even if delayed it took place at a time when the bid-ask spread was less than the maximal width of the spread over the 90-second time window. On the other hand, block (negotiated) trades usually carry a price concession and so could be executed outside this envelope and thereby be wrongly excluded. To the extent that, for it to lead to a bounceback, this is by a definition a single isolated transaction, it did not have a permanent price impact, but rather was associated with a transient liquidity effect.

We do not make use of the NBBO quotes depth: from the list of prevailing quotes by exchange, we already determine the best (maximum) bid price, and the best (minimum) ask price. Using the prevailing quote list, we could sum the bid sizes for quotes at the best bid price, and sum the ask sizes for the quotes at the best ask price. This would produce the NBBO sizes available to trade at each point of time and one could consider filtering out transactions with size that exceeds it. However, this would also eliminate legitimate transactions, such as block trades, or for that matter any trade that has a price impact. In any event, we find empirically that filtering the transactions by the NBBO filter as described above reduces drastically the number of bouncebacks in the data.

We started with the unfiltered transactions data, and this procedure results in a time series of NBBO-filtered transactions. We also produce a series of the midpoint of the just computed NBBO bid and ask prices and use this as our measurement of the quote price at that point in time. We therefore end up comparing three different measurements of the “price” series, for each stock:

- The unfiltered transactions;
- The NBBO-filtered transactions;
- The NBBO quotes midpoint.

We will apply the spectrogram methodology to each price series for each one of the 30 individual DJIA components and compare the results.

### C. *Individual Stocks vs. Indices*

The DJIA is defined as a simple average of the stock prices of 30 of the largest and most widely held public companies in the U.S., as determined by the Dow Jones Co. The index is adjusted by a divisor, which changes whenever one of the component stocks has a stock split or stock dividend, or one of the component is replaced by another stock, so as to generate a continuous value for the index at the time of the split, dividend or change of component. We collect high frequency data (sampled every 5 seconds, down from an average sampling interval of 3 seconds) on the cash value of the DJIA index from Tick Data, Inc. The data provider uses proprietary data filters.

2006 was a rather unremarkable, or quiet, year from the perspective of systematic jumps (i.e., jumps that would be visible at the level of the index), as perusal of financial news for that period would confirm. In what follows, we will compare the results obtained from the index to those on its individual components.

## IX. Empirical Results and Economic Implications

### A. *Jumps: Present or Not*

The empirical values of  $S_J$  are shown in the form of a histogram in Figure 9 for each of the three possible measurements of the data. The data for the histogram are produced by computing  $S_J$  for the four quarters of the year, the thirty stocks, and for a range of values of  $p$  from 3 to 6 (in increments of 0.25),  $\Delta_n$  from 5 seconds to 2 minutes (with values 5, 15, 30, 60, 90 and 120 seconds), and  $k = 2, 3$ . The top left histogram corresponds to the unfiltered transactions, the top right one to the NBBO-filtered transactions and the lower left one to the NBBO midpoint quotes.

As indicated in (40), values below 1 are indicative of noise of one form or another dominating. We find that this is the case only for the unfiltered transactions data, and only at the highest sampling frequencies, the histogram then displaying a left tail. For the other data measurements, the histograms display very little mass in the regions where the noise dominates. The conclusion from  $S_J$  is that the noise is not a major concern, except

for the unfiltered transactions at the ultra high frequencies, but once past this domain, the evidence points towards the presence of jumps with the histograms centered around 1.

This conclusion is not surprising per se, even for the index and a fortiori for the individual assets. The fact that the tails of high-frequencies log-return distributions exhibit a power decay, which points towards a model including jumps, has been recognized a long time ago, see for example Eberlein and Keller (1995). A power law fits the data quite well and a continuous component alone (with typically exponentially decaying tails) would be rather unlikely to generate such returns.

The middle right panel in Figure 9 displays the median value of  $S_J$  (across values of  $p$  and  $k$ , and the four quarters) as a function of the sampling interval  $\Delta_n$ . Unfiltered transactions are marked “U”, filtered transactions are marked “F” and the NBBO midpoint quotes are marked “M.” In all cases, the median value of  $S_J$  starts around 1 at the highest frequencies and then rises. Comparing different data measurements, we find that  $S_J$  is generally highest for “F” (meaning less evidence of jumps there), then for “U” and then for “M”. Similar results obtain if the mean is employed instead of the median.

But as the sampling frequency increases, the evidence in favor of the presence of jumps becomes more mixed, when the 30 components of the DJIA are taken in the aggregate. When implemented on the DJIA index itself, we find values of  $S_J$  that range between 1.5 and 2.5, providing between less evidence in favor of jumps and evidence against them for the index.

However, two points should be emphasized here. First, the histogram is more spread out when frequency decreases because less data are used and the statistical error increases, so that the procedure becomes less conclusive. Second, when the sampling frequency is low, the statistic  $S_J$  should not necessarily be expected to be close to its limiting value, 1 or 2, which is the theoretical limit when the frequency  $1/\Delta_n$  goes to infinity.

Finally, we check if any cross-sectional differences in  $S_J$  can be explained by cross-sectional differences in liquidity among the thirty stocks. To this aim, the lower two panels on Figure 9 show the results of a nonlinear regression of the statistic  $S_J$  on two stock-level measures of liquidity, the average time between successive trades, and the average size of the transactions. Both regressions show a slight decrease in  $S_J$  values as the asset becomes less

liquid, meaning higher time between trades and lower transaction size, but neither result is strong enough to be noticeable.

### *B. Jumps: Finite or Infinite Activity*

$S_J$  tells us that jumps are likely to be present, but it cannot distinguish between finite and infinite activity jumps. For this, we turn to the statistic  $S_{FA}$  which is like  $S_J$  with the addition of truncation.

Whenever we need to truncate, we express the truncation cutoff level  $u_n$  in terms of a number  $\alpha$  of standard deviations of the continuous part of the semimartingale. That initial standard deviation estimate is obtained by using  $B(2, 4\sigma\Delta_n^{1/2}, \Delta_n)$  where  $\sigma$  is a fixed realistic value for the asset under consideration; we use  $\sigma = 0.25$ . The specific value of this number serves only to identify a reasonable range of values; it does not matter asymptotically. It does of course matter in small samples, and so we suggest picking a realistic value for the asset and time period. We then use for the truncation level  $u_n$  different multiples of it. Our view of the joint choice of  $(\varpi, \alpha)$  is that they are not independent parameters in finite sample: they are different parameters for asymptotic purposes but in finite samples the only relevant quantity is the actual resulting cutoff size  $u_n$ . This is why we are reporting the values of the cutoffs  $u_n$  in the form of the  $\alpha$  that would correspond to  $\varpi = 1/2$ . This has the advantage of providing an easily interpretable size of the cutoff compared to the size of the increments that would be expected from the Brownian component of the process: we can then think in terms of truncating at a level that corresponds to  $\alpha = 4, 6$ , etc., standard deviations of the continuous part of the model. Since the ultimate purpose of the truncation is either to eliminate or conserve that part, it provides an immediate and intuitively clear reference point. Given  $u_n$  and this  $\alpha$ , it is possible to back this into the value of the  $\alpha$  corresponding to any  $\varpi$ , for that given sample size, including the value(s) of  $\varpi$  that satisfy the required inequalities imposed by the asymptotic results. This approach would lose its effectiveness if we were primarily interested in testing the validity of the asymptotic approximation as the sample size varies, but for applications, by definition on a finite sample, it seems to us that the interpretative advantage outweighs this disadvantage.

Each one of the statistic below is computed separately for each quarter of 2006 and for each asset. The data for the histogram in Figure 10 are produced by computing for the

four quarters of the year and each stock the value of  $S_{FA}$  for a range of values of  $p$  from 3 to 6,  $\alpha$  from 5 to 10 standard deviations,  $\Delta_n$  from 5 seconds to 2 minutes, and  $k = 2, 3$ . We find that the empirical values of  $S_{FA}$  are distributed around 1, which is indicative of infinite activity jumps. That is, even as we truncate, the statistic continues to behave as if jumps are present. If only a finite number of jumps had been present, then the statistic should have behaved as if the process were continuous. But the histograms do display a fat right tail, indicative of finite activity jumps for at least some of the DJIA components. The histograms are quite similar for all three data measurements, suggesting that they tend to differ only because of the larger increments: those are indeed the ones that are filtered in “F” compared to “U”, but since they are truncated away by  $S_{FA}$  anyway, then for the purpose of computing  $S_{FA}$  the two data measurements produce close results.

The middle right panel in Figure 10 displays the mean value of  $S_{FA}$  (across the four quarters, two stocks, and values of  $p, \alpha$  and  $k$ ) as a function of  $\Delta_n$ . A pattern similar to the corresponding plot in Figure 9 emerges. Even for very small values of  $\Delta_n$ , the noise does not dominate (limits below 1), instead the limit is around 1 as  $\Delta_n$  increases away from the frequencies where the noise would have been expected to dominate. Unless we start downsampling more (reaching 5 to 10 minutes), the limit does not get close to  $k^{p/2-1}$ . The lower panels examine any patterns linking  $S_{FA}$  to stock-level measures of liquidity; no strong cross-sectional pattern emerges.

Overall, the evidence suggests the presence of infinite activity jumps in the DJIA 30 components. To the extent that jumps are present in the DJIA index itself, the evidence is in favor of finite activity jumps: we find values of  $S_{FA}$  ranging from 1.7 to 2.2 for the index.

### *C. Brownian Motion: Present or Not*

In light of the likely presence of infinite activity jumps identified by  $S_{FA}$ , it makes sense to ask empirically whether a Brownian component is needed at all. For this purpose, we turn to the statistic  $S_W$ .

Figure 11 displays histograms of the distribution of  $S_W$  obtained by computing its value for the four quarters of the year for a range of values of  $p$  from 1 to 1.75,  $\alpha$  from 5 to 10 standard deviations,  $\Delta_n$  from 5 seconds to 2 minutes, and  $k = 2, 3$ . The empirical estimates

are always on the side of the limit arising in the presence a continuous component. Even as the sampling frequency increases, the noise is not a dominant factor, although as usual, lower values of  $S_W$  below 1 are now obtained and for very high sampling frequencies, the results are consistent with some mixture of the noise driving the asymptotics.

This is confirmed by the middle right panel in Figure 11 which displays the mean value of  $S_W$  (across the four quarters, two stocks, and values of  $p$ ,  $\alpha$  and  $k$ ) as a function of  $\Delta_n$ . As we downsample away from the noise-dominated frequencies, the average value of the statistic settles down towards the one indicating presence of a Brownian motion.

Because values of  $p$  less than 2 are employed by the statistic  $S_W$ , we find relatively small differences between the results for filtered and unfiltered transactions: since they differ mainly by a few of their large increments, but values of  $p < 2$  tend to under-emphasize large increments, we obtain similar results for “F” and “U”. The lower panels look at the relationship between  $S_W$  and the underlying asset’s liquidity. Like  $S_J$ , we find that there is a very slight increase in the value of the statistic as the asset becomes more liquid.

In the case of the DJIA index, we find that a Brownian motion is likely present, except at the highest frequencies where an alternative made of a pure jump infinite activity process would be plausible. Indeed, high frequency increments of the index tend to be very smooth owing to the nature of the index as an average.

#### *D. QV Relative Magnitude*

The previous empirical results indicate that are likely in the presence of a jump as well as a continuous component in the DJIA 30 stocks. We can then ask what fraction of the QV is attributable to the continuous and jump components.

The histograms in Figure 12 are obtained from computing the fraction of QV from the Brownian component using the four quarters, two stocks, values of  $\alpha$  ranging from 2 to 5 standard deviations, in increments of 1, and  $\Delta_n$  from 5 seconds to 2 minutes in the same increments as earlier. We find values around 75% for “F” and “U”, and somewhat lower for “M”, around 60% with some stock/quarter samples leading to values that are in fact indicative of an almost pure jump process in the quotes data.

In the middle right panel (similar but as a function of  $\Delta_n$ ), we see that the estimated fraction is fairly stable as we vary the sampling frequency. It is also quite stable for the two different measurements of the transactions data, “F” and “U”, and the quotes data “M”, going up slightly as the sampling frequency decreases. The lower panels show a more pronounced increase in the Brownian-driven of QV as a function of the asset’s liquidity: using both measures, we find that more liquid assets are associated with a higher proportion of Brownian-driven QV.

In the case of the DJIA index, we find values that range from 85% to 95%, suggesting in line with the previous evidence that jumps are less of a factor for the index. Incidentally, one could imagine measuring the proportion of jumps that are systematic vs. those that are idiosyncratic on the basis of the comparison between %QV estimated for the index and for its components.

### *E. Estimating the Degree of Jump Activity*

Finally, we estimate the degree of jump activity  $\beta$ . We found above that infinite activity jumps were likely present in the data. We are now asking how active are those jumps among all infinitely active jumps.

Of all the empirical methods employed in the paper, estimating  $\beta$  is the one that requires the largest sample size due to its reliance on truncating from the right in the power variations  $U$ . That is, the estimators of  $\beta$  discard by construction a large fraction of the original sample and to retain a sufficient number of observations to the right of a cutoff  $u_n$  given by 5 or more standard deviations of the continuous part, we need to have a large sample to begin with. So we will estimate  $\beta$  using only the first two sampling frequencies, 5 and 10 seconds on a quarterly basis. In the case of the previous statistics, we noted that these sampling frequencies were subject to market microstructure noise. Here, however, because we are only retaining the increments larger than the cutoff  $u_n$  instead of those smaller than the cutoff, this could be less of a concern despite the ultra high sampling frequencies.

We find estimated  $\beta$ ’s in the range from 1.5 to 1.8, indicating a very high degree of jump activity, in effect much closer to Brownian motion than to compound Poisson. The filtered transactions produce the highest estimates of  $\beta$ , leading on average to a process

that effectively looks like Brownian motion. Figure 13 reports the values of the estimator  $\hat{\beta}$  computed for the four quarters of the year, the two stocks, a range of values of  $\alpha$  from 5 to 10 standard deviations, and  $\Delta_n$  from 5 to 10 seconds. The middle right plot reports the corresponding data against the limited range of  $\Delta_n$  employed. Within the context of their parametric model, Carr, Geman, Madan, and Yor (2002) also found estimates of  $\beta$  on most stock return series that were greater than zero. (By contrast, the estimates here do not rely on a parametric structure for the rest of the model, so they can be thought of as semiparametric.) The lower panels, relating the estimated values of  $\beta$  to stock-level liquidity, do not display strong patterns.

Looking at the DJIA index itself, to the extent that an infinite activity component is present, we find that it is less active, with estimated values of  $\beta$  ranging from 0.9 to 1.4. But in light of the results of the test of finite vs. infinite jump activity, this component is likely small in the index relative to the finite activity component.

## *F. Economic Implications of the Empirical Results*

### *F.1. Implications for Option Pricing*

We have documented some of the salient features of asset returns viewed through the prism of a semimartingale. The analysis in this paper is, and can only be, entirely conducted under the physical probability measure,  $\mathbb{P}$ . As such, it does not lend itself directly to conclusions for option or other derivative prices, which depend upon the properties of the semimartingale under an equivalent martingale measure,  $\mathbb{Q}$ . However, important features of the process identified under  $\mathbb{P}$  do translate into similar features under  $\mathbb{Q}$ .

Specifically, economic theory dictates that  $\mathbb{P}$  and  $\mathbb{Q}$  must be absolutely continuous with respect to each other,  $\mathbb{P} \sim \mathbb{Q}$ , meaning that the two probability measures must have the same null sets. This requirement limits the possible ways in which the features of the process under  $\mathbb{Q}$  can differ from those already identified empirically under  $\mathbb{P}$ . As is well known, if the process is continuous under  $\mathbb{P}$ , it must be continuous under  $\mathbb{Q}$ , and vice versa. Hence, if jumps are possible under one measure they must be possible under the other as well. Thus the evidence in favor of jumps in the dynamics of the underlying asset therefore argues in favor of models with jumps in option pricing. This is of course well-accepted in

the empirical option pricing, see, e.g., Bates (1991) for some early evidence.

But more subtle findings of our analysis can also be translated into similar features under the pricing measure  $\mathbb{Q}$ . Two types of quantities are indeed almost surely identical pathwise, that is, are the same under both  $\mathbb{P}$  and  $\mathbb{Q}$ : one is the quadratic variation of the process, hence in particular its continuous part, that is the process  $\int_0^t \sigma_s^2 ds$ . The other is the set  $I_T$  of nonnegative reals  $p$  for which the sum  $\sum_{s \leq T} |\Delta X_s|^p$  is almost surely finite. In other words, the process  $\int_0^t \sigma_s^2 ds$  for all  $t \in [0, T]$  and the set  $I_T$  are the same under  $\mathbb{P}$  and under  $\mathbb{Q}$ .

As a consequence, the proportion of QV estimated in the empirical analysis under  $\mathbb{P}$  must be the same under  $\mathbb{Q}$ . Also, the absence of Brownian motion (tested under  $\mathbb{P}$  using  $S_W$ ), which amounts to saying that  $\sigma_t = 0$  for Lebesgue-almost all  $t$  in  $[0, T]$ , holds under  $\mathbb{P}$  if and only if it holds under  $\mathbb{Q}$ . The degree of jump activity  $\beta$  (the Blumenthal-Gettoor index, estimated above), which is the infimum of all  $p$  in the set  $I_T$ , is also the same under the two measures, that is

$$\beta_{\mathbb{P}} = \beta_{\mathbb{Q}}. \tag{44}$$

And the fact that jumps have finite activity (tested under  $\mathbb{P}$  using  $S_{FA}$ ) which amounts to saying that 0 belongs to  $I_T$ , is again true under  $\mathbb{P}$  if and only if it is true under  $\mathbb{Q}$ .

On the other hand, some features of the model may hold under  $\mathbb{P}$  and not under  $\mathbb{Q}$ , or vice-versa: typically, these concern the drift, which changes more or less arbitrarily when we change the measure, or the law of the big jumps, that is the tails of  $\nu$ . In other words, the components of the model that are identified at high frequency under  $\mathbb{P}$  by the spectrogram analysis (volatility, presence or not of jumps, finite or infinite activity, degree of jump activity, presence or not of the Brownian motion, proportion of continuous QV) happen to be the ones that are also restricted by the change of measure, while those that cannot be identified at high frequency (drift, law of the big jumps) are not.

The consequence of the previous discussion, and in particular of the restriction (44), combined with the empirical evidence regarding the estimated value of  $\beta_{\mathbb{P}}$  is to make some option pricing models more likely than others to be in agreement with the empirical evidence on the dynamics of the underlying asset price process. There are many option pricing models in the literature which include jumps, and quite often rely on specific Lévy processes, with specific values of  $\beta_{\mathbb{Q}}$ . Examples include compound Poisson-based models starting with

Merton (1976), the variance gamma model of Madan and Seneta (1990) and Madan, Carr, and Chang (1998) ( $\beta_{\mathbb{Q}} = 0$ ), the Normal Inverse Gaussian model of Barndorff-Nielsen (1998) ( $\beta_{\mathbb{Q}} = 1$ ), the hyperbolic model of Eberlein and Keller (1995), the generalized hyperbolic model of Barndorff-Nielsen (1977) and the CGMY model of Carr, Geman, Madan, and Yor (2002) (in which  $\beta_{\mathbb{Q}}$  is a free parameter). In the case of the data studied in this paper, models which allow for unrestricted values of  $\beta_{\mathbb{Q}}$  are likely to be more plausible candidates than those that restrict  $\beta_{\mathbb{Q}}$  to be small, 0 in particular. Indeed we tend to find empirically relatively high degrees of jump activity in  $\beta_{\mathbb{P}}$ . This is compatible with the findings from the literature on empirical option pricing with Lévy processes, where models with significant infinite-activity return innovations, that is a component with a  $\beta_{\mathbb{Q}} > 0$ , tend to be more accurate in their pricing of equity options than those with finite activity or sub-polynomial infinite activity,  $\beta_{\mathbb{Q}} = 0$  (see e.g., Carr, Geman, Madan, and Yor (2002) and Huang and Wu (2004).)

Models with jumps of finite or infinite activity also have differential implications for the prices of options at short and long horizons. As option maturities approach zero, out-of-the-money option values tend to zero. The speed of decay is exponential when the process is continuous, but linear in time-to-maturity  $T$  in the presence of jumps (see Carr and Wu (2003b)). In the presence of jumps, the implied volatility for an out-of-the-money option diverges as  $T \rightarrow 0$ . This holds in the presence of small and big jumps alike.

As option maturity increases, on the other hand, the big jumps component becomes a key element. From the Black-Scholes scenario, one would expect the implied volatility smile to flatten as a result of the Central Limit Theorem. The empirical evidence suggests, on the other hand, that the smile steepens instead (see, e.g., Foresi and Wu (2005)). The big jumps component can explain this fact. In a model where the variance of returns is infinite, a situation which can only arise because of the big jumps, the Central Limit Theorem ceases to apply for long-maturity asymptotics. But on the other hand, if most big jumps are negative, then the option would still retain a finite value, as observed empirically. In other words, the combination of components identified here under  $\mathbb{P}$  results in a plausible configuration for the model in light of the available evidence under  $\mathbb{Q}$  from the empirical option pricing literature.

## *F.2. Implications for Risk Management*

Risk management is primarily concerned with the tails of the returns distributions over different horizons, from the very short term horizons (a few seconds or minutes for high frequency trading applications), to the typical 10 days characteristic of VaR calculations, to the longer horizons relevant in portfolio management.

On a micro time scale, the main risk faced by a market participant practicing high frequency trading comes from the small jumps component, since the continuous component is small (of order  $\Delta_n$ ), and the big jumps component is very unlikely to give rise to realization over a very short interval. As discussed, we find the small jumps risk to be prevalent in individual equities we studied. The finer properties of the small jump component become an important consideration: as  $\beta$  increases, small jumps become more and more likely, and anecdotal evidence from the rapidly growing practice of high frequency trading suggests that the small jumps component is indeed a key consideration in the implementation of some of these strategies.

On a macro time scale, the implications are much more tenuous. The specification of the price process derived from the spectrogram methodology is nonparametric in nature, and relies on high frequency observations. As such, it fully identifies the behavior of the jump measure near 0. It is common in finance to rely on parametric models. One can imagine the nonparametric methodology employed here serving as a guide to pick one parametric model over another. Most parametric models rely on extrapolation in that the tails of the jump measure (i.e., the big jumps driver) is linked through parametric assumptions to the behavior of the jump measure near 0; within that framework, identifying the behavior of  $\nu$  near 0 is sufficient to characterize fully the jump measure. It is of course difficult in any case to pin down precisely how heavy the tail distributions are, since rare events require extremely large samples, that is, long time periods, to get realized in sufficient quantities to characterize precisely the big jumps component (see e.g., Heyde and Kou (2004)). Long sample sizes in turn require an implicit assumption of model stability over the observation horizon. Alternatively, by time aggregation, one can derive the implications for long horizons of the continuous-time model identified over short horizons, typically by simulations. But, overall, drawing long-horizon implications from this analysis entails successive leaps of faith that make it potentially fraught with dangers.

### *F.3. Individual Assets vs. Stock Indices, Factor Models and Systematic vs. Idiosyncratic Risk*

As part of the empirical analysis above, we compared the result of the spectrogram methodology on the DJIA and its components. We found that while the individual components seemed to contain an infinite activity jump component with a fairly high degree of jump activity, the evidence for the index was pointing more towards a finite activity component only.

This empirical evidence can be interpreted in light of a factor model with systematic and idiosyncratic components. If we further decompose the two jump components in the model (2) for individual stocks into a common component and an idiosyncratic component, then when an idiosyncratic jump occurs the other stock prices do not jump, so the influence of this particular jump on the index, which is an average, is essentially wiped out: the index will not exhibit a jump of significant size. In contrast, a systematic jump will typically occur at the same time (and often with the same sign) for most or all stocks, resulting in a jump of the index. Therefore, the absence of empirical evidence in favor of infinite activity jumps in the index, combined with their presence in the individual components, point towards purely idiosyncratic small jump components in the individual assets.

This makes sense if we think of most systematic, i.e., market-wide, price moves as driven by important macroeconomic news and therefore likely to be large. On the other hand, small infinite activity jumps are more likely to reflect individual stock-level considerations, such as stock-specific information and its immediate trading environment, and therefore be idiosyncratic.

Consistent with this, we also found that the proportion of quadratic variation attributable to jumps is lower for the index than for its individual components. One could conceivably measure the proportion of jumps that are systematic vs. those that are idiosyncratic on the basis of the comparison between the proportions of quadratic variation estimated for the index and for its components. Doing this using the empirical results above would suggest a proportion of systematic jumps representing about 10% of the total quadratic variation, and a proportion of idiosyncratic jumps representing about 15% of total quadratic variation.

## X. Conclusions

The empirical results on the DJIA 30 components appear to indicate that jumps are present in the data; point towards the presence of infinite activity jumps; of degree of jump activity that is around 1.5 or higher; indicate that, on top of the jump components, a continuous component is present; which represents approximately 3/4 of the total QV.

The analysis of different measurement mechanisms suggests that the results are fairly consistent when various ways of measuring the data (“F”, “U” or “M”) are employed. The DJIA index itself looks quite different, when viewed through this prism, than its individual components. In particular, although market microstructure noise is less a factor for all DJIA 30 stocks than it actually is for its two Nasdaq components (Microsoft and Intel), there is very little evidence of noise even at the highest frequencies for the index itself. And there is less evidence for the presence of jumps in the index than in its components, and less evidence for infinite activity jumps than in the components, or at least less active small jumps.

Of course, we do not claim on the basis of this limited evidence that these empirical results are in any way “universal”: they are likely in general to depend upon the assets under consideration, the time period, what type of data are used (transactions, quotes, etc.), among other considerations.

In terms of methodology, our assessment at present is the following. The pros of the approach are: we have a unified methodology to address all these seemingly disparate specification questions in a common framework; the method allows for a symmetric treatment of null and alternative hypotheses in each case, including the full distribution theory; all the test statistics are model-free; they are extremely simple to implement in practical applications; and, finally, we are able to characterize the impact of two important types of market microstructure noise, additive and rounding, on the various statistics.

The cons of the approach, at present, are the flip side of the commonality of technique: if we take individually each one of the specification questions we have identified, then the statistic we have proposed for that problem, based on that common approach, is not necessarily the optimal approach for that individual specification question. Also, the practical implementation of this method does require high frequency data (particularly the estima-

tion of the degree of jump activity  $\beta$ , which requires ultra-high frequency data). Finally, while we understand what happens to each one of the statistics when microstructure noise dominates, we do not yet have noise-robust statistics that would be fully immune to the presence of the noise.

To conclude, we certainly expect to see refinements of our approach (see e.g., Fan and Fan (2008)) or perhaps entirely different techniques proposed in the future for each one of these problems, some already existing. However, we think that the main appeal of the approach comes from the commonality of technique, as well as the resulting simplicity of implementation.

## REFERENCES

- Aït-Sahalia, Yacine, 2002, Telling from discrete data whether the underlying continuous-time model is a diffusion, *Journal of Finance* 57, 2075–2112.
- Aït-Sahalia, Yacine, and Jean Jacod, 2008a, Is Brownian motion necessary to model high frequency data?, *Annals of Statistics*, *forthcoming*.
- Aït-Sahalia, Yacine, and Jean Jacod, 2008b, Testing whether jumps have finite or infinite activity, Working paper, Princeton University and Université de Paris-6.
- Aït-Sahalia, Yacine, and Jean Jacod, 2009a, Estimating the degree of activity of jumps in high frequency financial data, *Annals of Statistics* 37, 2202–2244.
- Aït-Sahalia, Yacine, and Jean Jacod, 2009b, Testing for jumps in a discretely observed process, *Annals of Statistics* 37, 184–222.
- Andersen, Torben G., Tim Bollerslev, and Francis X. Diebold, 2007, Roughing it up: Including jump components in the measurement, modeling, and forecasting of return volatility, *Review of Economics and Statistics* 89, 701–720.
- Azaïs, Jean-Marc, 1989, Approximation des trajectoires et temps local des diffusions, *Annales de l'Institut Henri Poincaré* 25, 175–194.
- Ball, Clifford A., and Walter N. Torous, 1983, A simplified jump process for common stock returns, *Journal of Financial and Quantitative Analysis* 18, 53–65.
- Barndorff-Nielsen, Ole E., 1977, Exponentially decreasing distributions for the logarithm of particle size, *Proceedings of the Royal Society of London A* 353, 401–419.
- Barndorff-Nielsen, Ole E., 1997, Normal inverse Gaussian distributions and stochastic volatility modelling, *Scandinavian Journal of Statistics* 24, 1–13.
- Barndorff-Nielsen, Ole E., 1998, Processes of normal inverse Gaussian type, *Finance and Stochastics* 2, 41–68.
- Barndorff-Nielsen, Ole E., and Neil Shephard, 2004, Power and bipower variation with stochastic volatility and jumps (with discussion), *Journal of Financial Econometrics* 2, 1–48.
- Bates, David S., 1991, The crash of '87: Was it expected? The evidence from options markets, *Journal of Finance* 46, 1009–1044.
- Belomestny, Denis, 2009, Spectral estimation of the fractional order of a Lévy process, *Annals of Statistics*, *forthcoming*.
- Blumenthal, R.M., and R.K. Gettoor, 1961, Sample functions of stochastic processes with stationary independent increments, *Journal of Mathematics and Mechanics* 10, 493–516.

- Carr, Peter, Hélyette Geman, Dilip B. Madan, and Marc Yor, 2002, The fine structure of asset returns: An empirical investigation, *Journal of Business* 75, 305–332.
- Carr, Peter, and Liuren Wu, 2003a, The finite moment log stable process and option pricing, *Journal of Finance* 58, 753–777.
- Carr, Peter, and Liuren Wu, 2003b, What type of process underlies options? A simple robust test, *Journal of Finance* 58, 2581–2610.
- Carr, Peter, and Liuren Wu, 2004, Time-changed Lévy processes and option pricing, *Journal of Financial Economics* 71, 113–141.
- Cont, Rama, and Cecilia Mancini, 2009, Nonparametric tests for probing the nature of asset price processes, Working paper, Università di Firenze.
- Delattre, Sylvain, 1997, Estimation du coefficient de diffusion avec erreur d’arrondi, Ph.D. thesis Université Paris-6.
- Delbaen, Freddy, and Walter Schachermayer, 1994, A general version of the fundamental theorem of asset pricing, *Mathematische Annalen* 300, 463–520.
- Duffie, Darrell, Jun Pan, and Kenneth J. Singleton, 2000, Transform analysis and asset pricing for affine jump-diffusions, *Econometrica* 68, 1343–1376.
- Eberlein, Ernst, and Ulrich Keller, 1995, Hyperbolic distributions in finance, *Bernoulli* 1, 281–299.
- Fan, Yingying, and Jianqing Fan, 2008, Testing and detecting jumps based on a discretely observed process, Working paper, Princeton University.
- Foresi, Silverio, and Liuren Wu, 2005, Crash-o-phobia: A domestic fear or a worldwide concern?, *Journal of Derivatives* 13, 8–21.
- Glosten, Lawrence R., 1987, Components of the bid-ask spread and the statistical properties of transaction prices, *Journal of Finance* 42, 1293–1307.
- Glosten, Lawrence R., and Lawrence E. Harris, 1988, Estimating the components of the bid/ask spread, *Journal of Financial Economics* 21, 123–142.
- Gottlieb, Gary, and Avner Kalay, 1985, Implications of the discreteness of observed stock prices, *Journal of Finance* 40, 135–153.
- Harris, Lawrence, 1990a, Estimation of stock price variances and serial covariances from discrete observations, *Journal of Financial and Quantitative Analysis* 25, 291–306.
- Harris, Lawrence, 1990b, Statistical properties of the Roll serial covariance bid/ask spread estimator, *Journal of Finance* 45, 579–590.

- Harrison, Michael, and Stanley Pliska, 1981, Martingales and stochastic integrals in the theory of continuous trading, *Stochastic Processes and Their Applications* 11, 215–260.
- Hasbrouck, Joel, 1993, Assessing the quality of a security market: A new approach to transaction-cost measurement, *Review of Financial Studies* 6, 191–212.
- Heyde, Chris C., and Steven G. Kou, 2004, On the controversy over tailweight of distributions, *Operations Research Letters* 32, 399–408.
- Huang, Jingzhi, and Liuren Wu, 2004, Specification analysis of option pricing models based on time-changed Lévy processes, *Journal of Finance* pp. 1405–1440.
- Huang, Xin, and George T. Tauchen, 2005, The relative contribution of jumps to total price variance, *Journal of Financial Econometrics* 4, 456–499.
- Jiang, George J., and Roel C.A. Oomen, 2008, Testing for jumps when asset prices are observed with noise - A "swap variance" approach, *Journal of Econometrics* 144, 352–370.
- Lee, Suzanne, and Per A. Mykland, 2008, Jumps in financial markets: A new nonparametric test and jump dynamics, *Review of Financial Studies* 21, 2535–2563.
- Lee, Suzanne S., and Jan Hannig, 2010, Detecting jumps from Lévy jump diffusion processes, *Journal of Financial Economics*, forthcoming.
- Madan, D. B., P. P. Carr, and E. E. Chang, 1998, The Variance Gamma process and option pricing, *European Finance Review* 2, 79–105.
- Madan, Dilip B., and Frank Milne, 1991, Option pricing with V.G. martingale components, *Mathematical Finance* 1, 39–56.
- Madan, Dilip B., and Eugene Seneta, 1990, The Variance Gamma (V.G.) model for share market returns, *Journal of Business* 63, 511–524.
- Madhavan, Ananth, Matthew Richardson, and Mark Roomans, 1997, Why do security prices change?, *Review of Financial Studies* 10, 1035–1064.
- Mancini, Cecilia, 2001, Disentangling the jumps of the diffusion in a geometric jumping Brownian motion, *Giornale dell'Istituto Italiano degli Attuari* LXIV, 19–47.
- Merton, Robert C., 1976, Option pricing when underlying stock returns are discontinuous, *Journal of Financial Economics* 3, 125–144.
- Roll, Richard, 1984, A simple model of the implicit bid-ask spread in an efficient market, *Journal of Finance* 39, 1127–1139.
- Schoutens, Win, 2003, *Lévy Processes in Finance: Pricing Financial Derivatives*. (Wiley U.K.).

Stoll, Hans R., and Christoph Schenzler, 2006, Trades outside the quotes: Reporting delay, trading option, or trade size?, *Journal of Financial Economics* 79, 615–653.

Todorov, Viktor, and George T. Tauchen, 2010, Activity signature functions for high-frequency data analysis, *Journal of Econometrics* 154, 125–138.

Woerner, Jeannette H.C., 2006, Analyzing the fine structure of continuous-time stochastic processes, Working paper, University of Göttingen.

## APPENDIX: PROOFS

### Appendix 1. Proof of Theorem 1

*Proof.* We assume (31) and (32). We start with the behavior of the power variations when there is only noise, that is we consider the variables

$$\overline{B}(p, u_n, \Delta_n) = \sum_{i=1}^{\lfloor T/\Delta_n \rfloor} |\varepsilon_i^n - \varepsilon_{i-1}^n|^p \mathbf{1}_{\{|\varepsilon_i^n - \varepsilon_{i-1}^n| \leq u_n\}}. \quad (\text{A.1})$$

$$\overline{U}(p, u_n, \Delta_n) = \sum_{i=1}^{\lfloor T/\Delta_n \rfloor} |\varepsilon_i^n - \varepsilon_{i-1}^n|^p \mathbf{1}_{\{|\varepsilon_i^n - \varepsilon_{i-1}^n| > u_n\}}. \quad (\text{A.2})$$

We also write  $G_p(x) = \mathbb{E}(|\varepsilon_1 - \varepsilon_0|^p \mathbf{1}_{\{|\varepsilon_1 - \varepsilon_0| \leq x\}})$ . Then (32) readily implies

$$\frac{G_p(x)}{x^{p+1}} \rightarrow \frac{2f(0)}{p+1}. \quad (\text{A.3})$$

□

**Lemma 1.** *Let  $p \geq 0$ . Then*

$$\Delta_n \overline{B}'(p, \infty, \Delta_n) \xrightarrow{\mathbb{P}} m'_p T. \quad (\text{A.4})$$

*If further  $u_n > 0$  with  $u_n \rightarrow 0$  and  $\Delta_n u_n^p / G_p(u_n) \rightarrow 0$ , we have*

$$\frac{\Delta_n}{G_p(u_n)} \overline{B}(p, u_n, \Delta_n) \xrightarrow{\mathbb{P}} T. \quad (\text{A.5})$$

*Proof.* We denote by  $\overline{B}_+(p, u_n, \Delta_n)$  and  $\overline{B}_-(p, u_n, \Delta_n)$  the quantities defined by the right side of (A.1), but where the sum is extended over all indices  $i$  that are even, resp. odd. It is then sufficient to prove that, suitably normalized, both these variables converge to half the limit in (A.4) or (A.5), and it suffices to prove the result for, say,  $\overline{B}_+(p, u_n, \Delta_n)$ .

For (A.4) this is immediate: indeed  $\overline{B}_+(p, u_n, \Delta_n)$  is the sum of  $\lfloor [T/\Delta_n]/2 \rfloor \sim T/2\Delta_n$  i.i.d. terms with mean  $m'_p$  and a law independent of  $n$ , so the law of large numbers applies.

For (A.5) the summands  $\zeta_i^n = |\varepsilon_i^n - \varepsilon_{i-1}^n|^p \mathbf{1}_{\{|\varepsilon_i^n - \varepsilon_{i-1}^n| \leq u_n\}}$  are still i.i.d. for each  $n$  (recall that  $i$  is even), but their law depends on  $n$ . We have  $\mathbb{E}(\zeta_i^n) = G_p(u_n)$  and  $\mathbb{E}((\zeta_i^n)^2) \leq u_n^p G_p(u_n)$ . The result then readily follows from standard properties of i.i.d. triangular arrays, because  $\Delta_n u_n / G_p(u_n) \rightarrow 0$ .

We are now ready to prove Theorem 1. Below,  $K$  denotes a constant which may change from line to line.

a) In view of the previous lemma, it is enough to prove that

$$\Delta_n |U'(p, u_n, \Delta_n) - \overline{B}(p, \infty, \Delta_n)| \xrightarrow{\mathbb{P}} 0. \quad (\text{A.6})$$

For this, we recall the following properties, under our standing assumptions on  $X$ :

$$\begin{cases} p \geq 2 & \Rightarrow \text{the sequence } B(p, \infty, \Delta_n) \text{ is bounded in probability} \\ p < 2 & \Rightarrow \Delta_n^{1-p/2} B(p, \infty, \Delta_n) \xrightarrow{\mathbb{P}} m_p \int_0^T |\sigma_s|^p ds \end{cases} \quad (\text{A.7})$$

where  $m_p$  is the  $p$ th absolute moment of the law  $\mathcal{N}(0, 1)$ . We observe that for all reals  $x, y$ ,

$$||x + y|^p 1_{\{|x+y| > u_n\}} - |y|^p| \leq \begin{cases} u_n & \text{if } p = 0 \\ K(|x|^p + u_n^{p/2}|y|^{p/2}) & \text{if } p \in (0, 1] \\ K(|x|^p + u_n^{p/2}|y|^{p/2} + |x||y|^{p-1}) & \text{if } p > 1. \end{cases}$$

Then if  $p = 0$  the left side of (A.6) is smaller than  $u_n T$ , hence the result. If  $p \in (0, 1]$  it is smaller than  $\Delta_n B(p, \infty, \Delta_n) + u_n^{p/2} \Delta_n \overline{B}(p, \infty, \Delta_n)$ , and the result follows from (A.7) and from (A.4) applied with  $p/2$ . If  $p > 1$  it is smaller than

$K \Delta_n (B(p, \infty, \Delta_n) + K u_n^{p/2} \Delta_n \overline{B}(p, \infty, \Delta_n) + K (\Delta_n B(p, \infty, \Delta_n))^{1/p} (\Delta_n \overline{B}(p, \infty, \Delta_n))^{1-1/p})$  and (A.4) and (A.7) again yield (A.6).

b) The proof is more complicated. We assume now  $p > 0$ . A standard localization procedure yields that one may assume that

$$\mathbb{E}(|\Delta_i^n X|^2) \leq K \Delta_n. \quad (\text{A.8})$$

In view of (A.3), and again because of the previous lemma, it is enough to show

$$\frac{\Delta_n}{u_n^{p+1}} |B'(p, u_n, \Delta_n) - \overline{B}(p, u_n, \Delta_n)| \xrightarrow{\mathbb{P}} 0. \quad (\text{A.9})$$

For all reals  $x, y$  we have, with the notation  $A_n = \{(x, y) : |x + y| \leq u_n < |y|\} \cup \{|y| \leq u_n < |x + y|\}$ :

$$||x + y|^p 1_{\{|x+y| \leq u_n\}} - |y|^p 1_{\{|y| \leq u_n\}}| \leq \begin{cases} K(|x|^p + u_n^p 1_{A_n}(x, y)) & \text{if } p \leq 1 \\ K(|x|^p + u_n^p 1_{A_n}((x, y) + u_n^{p-1}|x|) & \text{if } p > 1. \end{cases}$$

Then if we set

$$H_n = \frac{\Delta_n}{u_n} \sum_{i=1}^{\lceil T/\Delta_n \rceil} 1_{A_n}(\Delta_i^n X, \varepsilon_i^n - \varepsilon_{i-1}^n). \quad (\text{A.10})$$

we see that when  $p \leq 1$ , and by Hölder's inequality, the left side of (A.9) is smaller than  $\frac{K \Delta_n}{u_n^{p+1}} B(p, \infty, \Delta_n) + K H_n$ , whereas when  $p > 1$  it is smaller than

$$\frac{K \Delta_n}{u_n^{p+1}} B(p, \infty, \Delta_n) + \frac{K \Delta_n}{u_n^2} B(1, \infty, \Delta_n) + K H_n. \quad (\text{A.11})$$

(35) and (A.7) imply that  $\frac{\Delta_n}{u_n^{p+1}} B(p, \infty, \Delta_n) \xrightarrow{\mathbb{P}} 0$ , and also  $\frac{K\Delta_n}{u_n^2} B(1, \infty, \Delta_n) \xrightarrow{\mathbb{P}} 0$  when  $p \geq 1$ . Hence it remains to show that

$$\mathbb{E}(H_n) \rightarrow 0.$$

To see this we need to evaluate  $\mathbb{P}(\tilde{A}_i^n)$ , where  $\tilde{A}_i^n = \{(\Delta^n i X, \varepsilon_i^n - \varepsilon_{i-1}^n) \in A_n\}$ . By (A.8) and Markov's inequality and the independence of the  $\varepsilon_i$ 's from the process  $X$ , and since the density  $f$  is an even function, we deduce for any  $q \in (0, 1)$  that

$$\mathbb{P}(\tilde{A}_i^n) \leq K\Delta_n^q \left( \frac{1}{u_n^q} + \int_0^{2u_n} \frac{f(y)}{|y - u_n|^q} dy \right) \leq \frac{K\Delta_n^q}{u_n^q}. \quad (\text{A.12})$$

Therefore  $\mathbb{E}(H_n) \leq KT\Delta_n^q/u_n^{1+q}$ , which goes to 0 by (A.12) if we take  $q = 1/3$ . This finishes the proof.  $\square$

## Appendix 2. Proof of Theorem 2

*Proof.* Theorem 2 is a rather complicated result, whose main elements of proof are due to Delattre (1997). We will just sketch here the reason why such a result holds. The key points are as follows:

a) When  $X$  is continuous, or when it has jumps with finite activity (or more generally summable jumps), the local time  $L^a$  at level  $a$  is the unique continuous increasing adapted process starting at 0, and such that

$$\begin{aligned} |X_t - a| &= |X_0 - a| + \int_0^t \text{sign}(X_{s-} - a) dX_s + L_t^a \\ &\quad + \sum_{s \leq t} (|X_s - a| - |X_{s-} - a| - \text{sign}(X_{s-} - a) \Delta X_s) \end{aligned} \quad (\text{A.13})$$

(Tanaka's formula; the function  $x \mapsto \text{sign}(x)$  takes the values  $-1, 0$  and  $+1$ , according to whether  $x < 0$  or  $x = 0$  or  $x > 0$ ).

b) When  $X$  is continuous, any increment

$$\chi_i^n = [X_{i\Delta_n}]_\alpha - [X_{(i-1)\Delta_n}]_\alpha$$

is (for all  $n$  large enough and all  $i \leq T/\Delta_n$  equal to  $-\alpha$  or  $0$  or  $\alpha$ . In other words  $B^n(p, \infty, \Delta_n)$  can be written as follows:

$$B^n(p, \infty, \Delta_n) = \alpha^p \sum_{q \in \mathbb{Z}} V^n(q), \quad (\text{A.14})$$

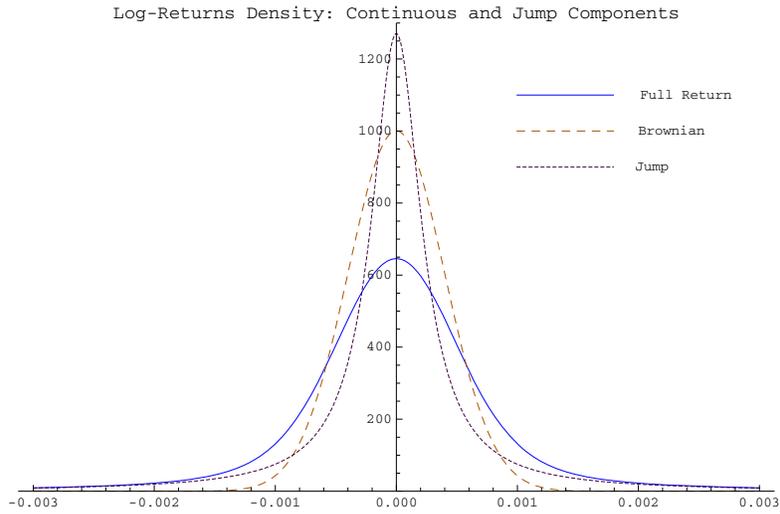
where  $V^n(q)$  is the number of integers  $i$  less than  $T/\Delta_n$  and such that the product  $([X_{i\Delta_n}]_\alpha - (q + 1/2)\alpha) ([X_{(i-1)\Delta_n}]_\alpha - (q + 1/2)\alpha)$  is negative (that is, the path of  $X$  crosses the level  $(q + 1/2)\alpha$  between times  $(i - 1)\Delta_n$  and  $i\Delta_n$ ).

c) Suppose that  $X = \sigma W$  is a Wiener process with unit variance  $\sigma^2$ . On the one hand Rice's formula connects the number of crossings of a given level  $a$  by  $X$ , suitably integrated in  $a$ , with the occupation times, which themselves are integrals over  $a$  of the local time  $L_T^a$ . On the other hand these properties can be discretized in time and "localized" around any given level  $a$ , according to Azaïs (1989) for example: this gives that  $\sqrt{\Delta_n} V^n(q)$  converges in probability to  $\frac{\sqrt{\pi/2}}{\sigma} L_T^{(q+1/2)\alpha}$ .

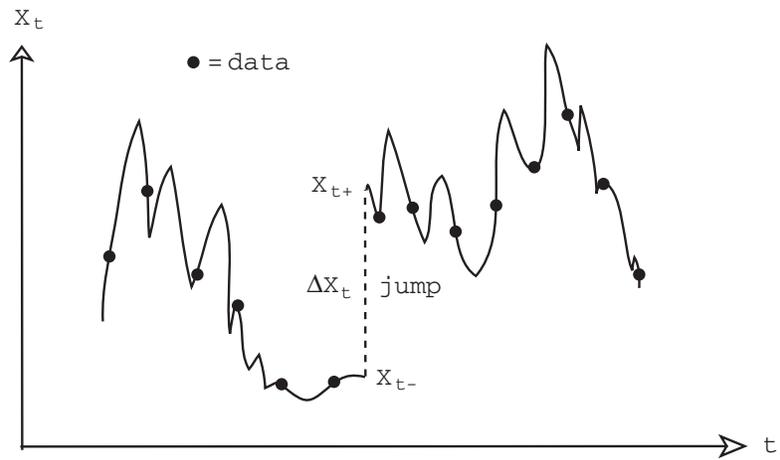
d) Coming back to a continuous Itô semimartingale, we can argue as if the process  $\sigma_t$  were piecewise constant in time, whereas the drift plays no role: so we obtain that  $\sqrt{\Delta_n} V^n(q)$  converges in probability to  $\sqrt{\pi/2} \int_0^T \frac{1}{\sigma_s} dL_s^{(q+1/2)\alpha}$ .

e) Observing that the sum in (A.14) is indeed a finite sum (for any fixed  $\omega$ ), because  $V^n(q) = 0$  for any  $q$  bigger than  $1 + 2 \sup_{s \leq T} |X_s|/\alpha$ , we deduce (38) when  $X$  is continuous.

f) When  $X$  has jumps with finite activity, the same argument holds provided we discard the (finitely many) increments over the time intervals containing a jump. Those increments are bounded (for  $n$  large enough) by the size of the jump, plus  $2\alpha$ . Since we normalize by  $\sqrt{\Delta_n}$ , they are wiped out, and we get the result.  $\square$

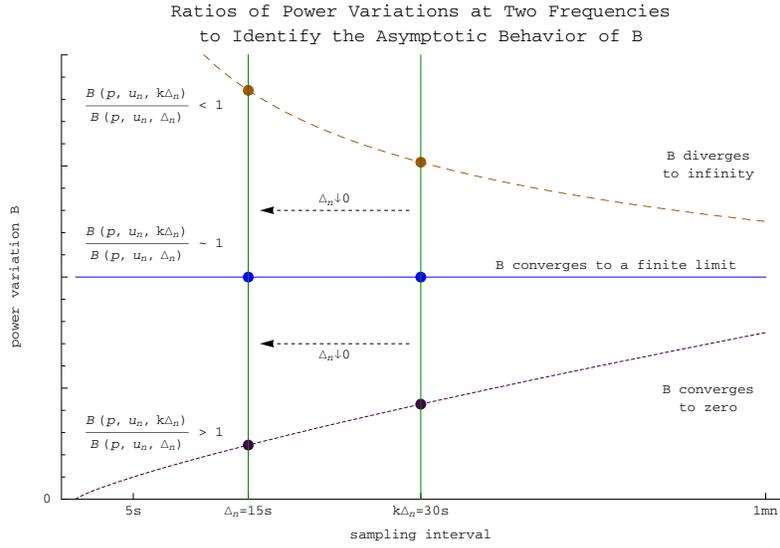


**Figure 1**  
**Deconvoluting the log-returns distribution into continuous and jump components.**

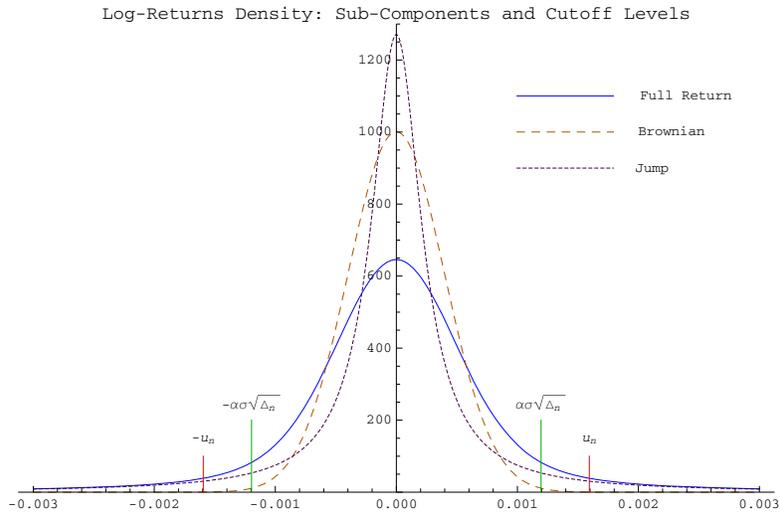


**Figure 2**  
**Discretely sampled data at interval  $\Delta_n$  vs. continuous-time sample path, and difference between increments and jumps.**

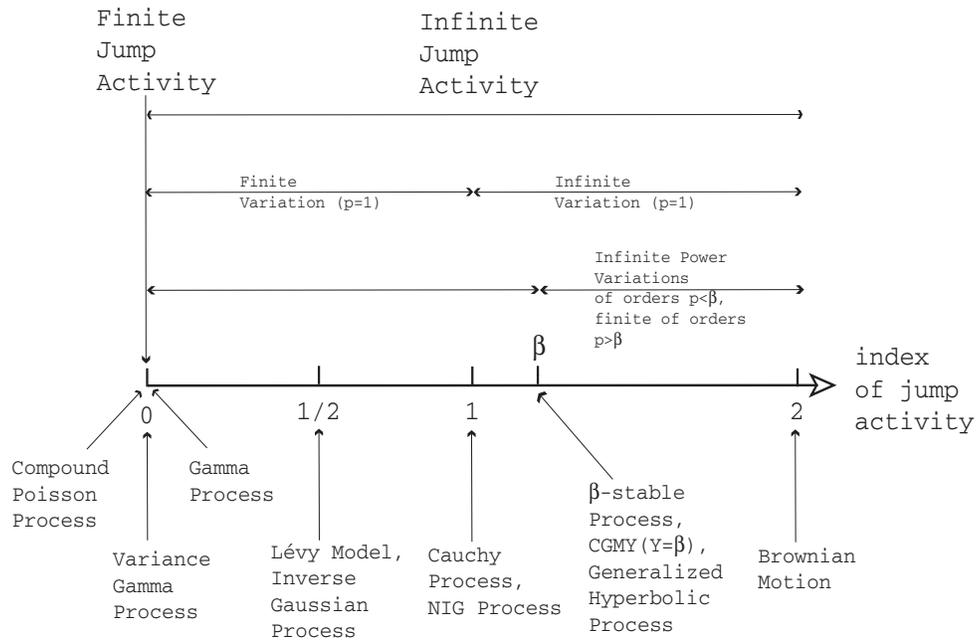




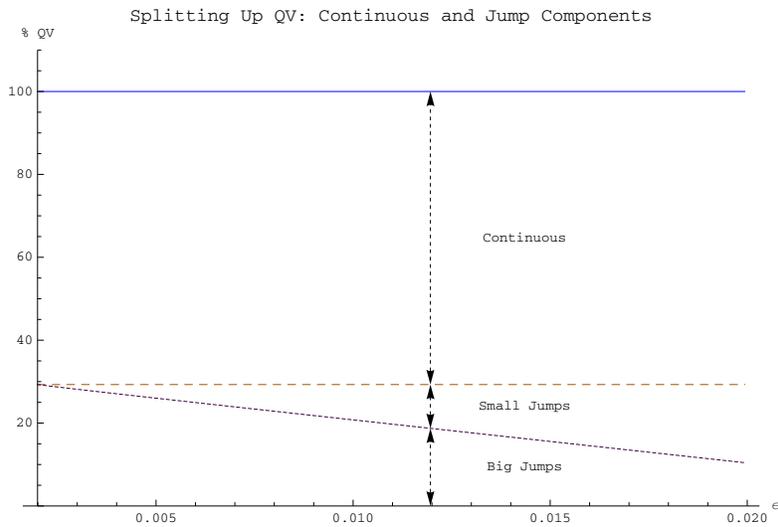
**Figure 5**  
**Three possible asymptotic behaviors of the power variation (diverge to infinity, converge to a finite limit, converge to zero) and means of identifying them.**



**Figure 6**  
**Truncating to retain or avoid the contribution from the Brownian component of the model.**

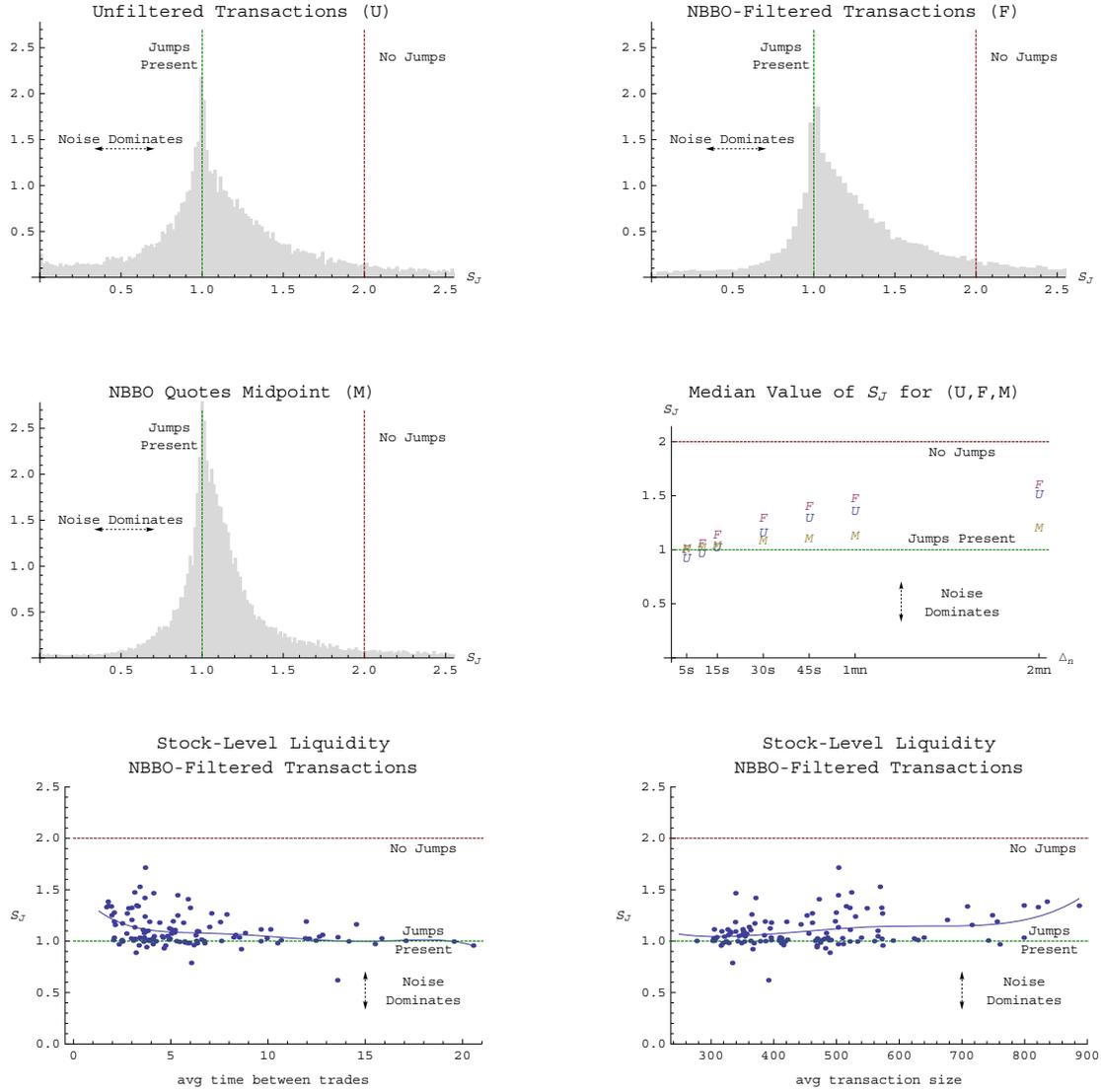


**Figure 7**  
**Index of jump activity  $\beta$  : Examples of processes.**



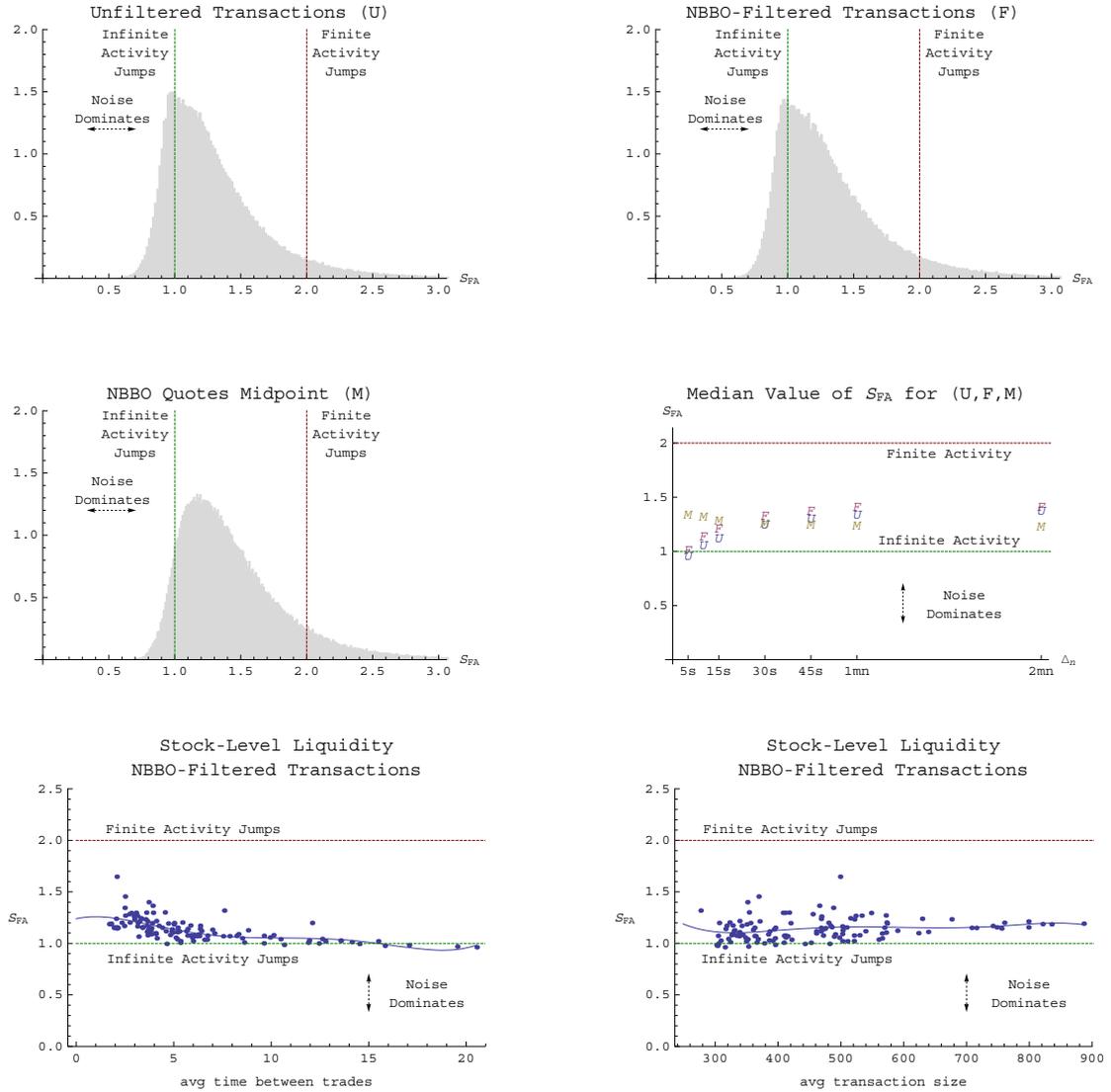
**Figure 8**  
**Splitting up the QV into continuous and jump components, and into small and big jumps as a function of the jump size cutoff  $\epsilon$ .**

Test Statistic  $S_J$  to Test for the Presence of Jumps  
 DJIA 30 Components



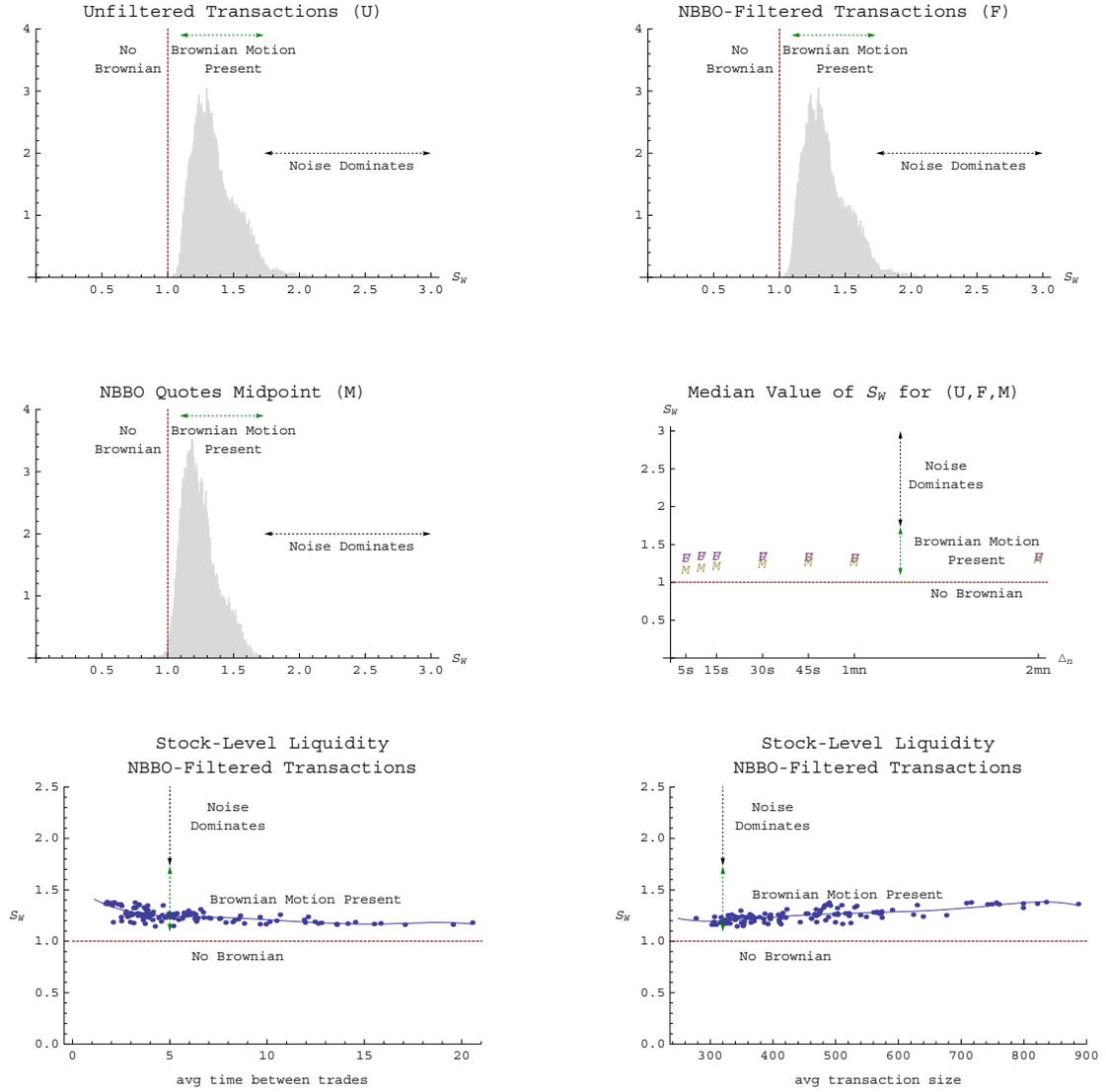
**Figure 9**  
 Empirical distribution of  $S_J$  for all 30 DJIA components, 2006, measured using transactions (unfiltered and NBBO-filtered) and NBBO quotes midpoint, median value of  $S_J$  as a function of the sampling interval  $\Delta_n$ , and nonlinear regression of  $S_J$  against stock-level liquidity measures.

Test Statistic  $S_{FA}$  to Test Whether Jumps Have Finite or Infinite Activity  
 DJIA 30 Components



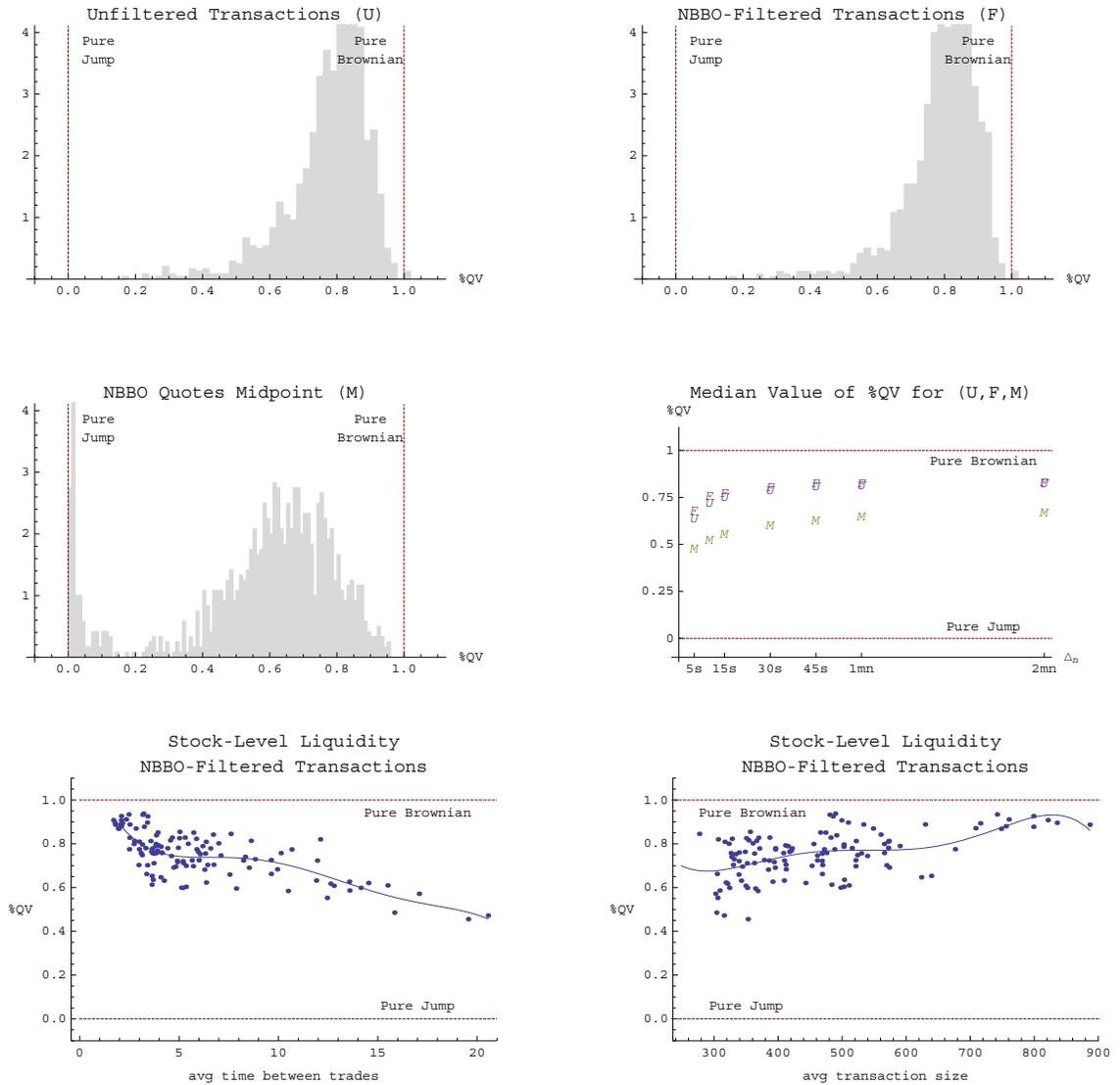
**Figure 10**  
 Empirical distribution of  $S_{FA}$  for all 30 DJIA components, 2006, measured using transactions (unfiltered and NBBO-filtered) and NBBO quotes midpoint, median value of  $S_{FA}$  as a function of the sampling interval  $\Delta_n$ , and nonlinear regression of  $S_{FA}$  against stock-level liquidity measures.

Test Statistic  $S_W$  to Test Whether Brownian Motion is Present  
DJIA 30 Components



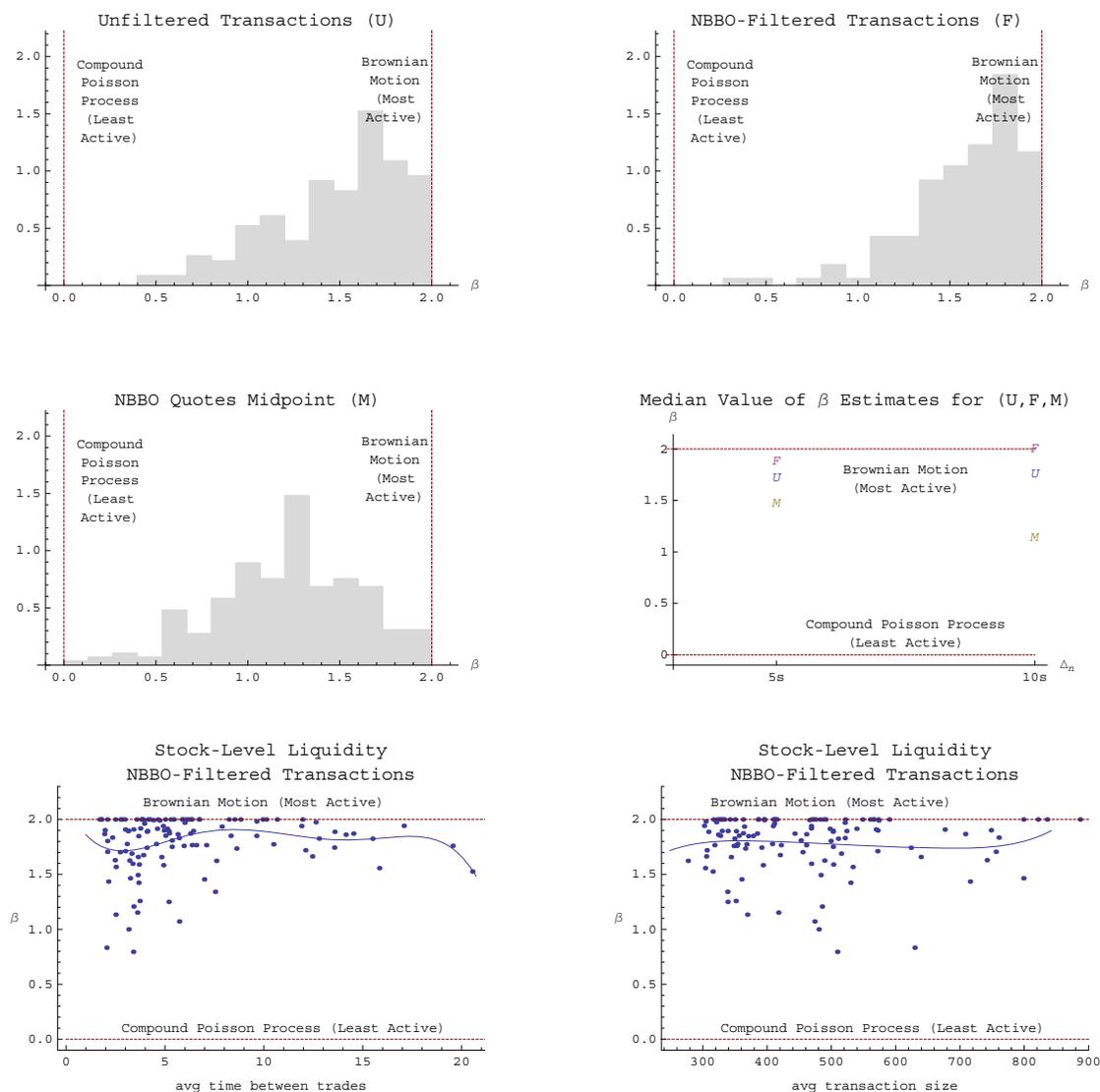
**Figure 11**  
Empirical distribution of  $S_W$  for all 30 DJIA components, 2006, measured using transactions (unfiltered and NBBO-filtered) and NBBO quotes midpoint, median value of  $S_W$  as a function of the sampling interval  $\Delta_n$ , and nonlinear regression of  $S_W$  against stock-level liquidity measures.

Proportion of Quadratic Variation Attributable to the Continuous Component  
DJIA 30 Components



**Figure 12**  
Empirical distribution of the proportion of  $QV$  attributable to the continuous component for all 30 DJIA components, 2006, measured using transactions (unfiltered and NBBO-filtered) and NBBO quotes midpoint, median value of  $\%QV$  as a function of the sampling interval  $\Delta_n$ , and nonlinear regression of  $\%QV$  against stock-level liquidity measures.

Estimate of the Degree of Jump Activity  $\beta$   
 DJIA 30 Components



**Figure 13**  
 Empirical distribution of the index of jump activity  $\beta$  for all 30 DJIA components, 2006, measured using transactions (unfiltered and NBBO-filtered) and NBBO quotes midpoint, and median value of the estimated  $\beta$  as a function of the sampling interval  $\Delta_n$ , and nonlinear regression of the estimated  $\beta$  against stock-level liquidity measures.

|  |   |  |
|--|---|--|
| Jumps: Present or Not                          | $H_0 : \Omega_T^c$<br>$H_1 : \Omega_T^j$<br>$S_J : \begin{pmatrix} p > 2 \\ \infty \\ \Delta_n, k\Delta_n \end{pmatrix}$  | $H_0 : \Omega_T^j$<br>$H_1 : \Omega_T^c$<br>$S_J : \begin{pmatrix} p > 2 \\ \infty \\ \Delta_n, k\Delta_n \end{pmatrix}$               |
| Jumps: Finite or Infinite Activity             | $H_0 : \Omega_T^f$<br>$H_1 : \Omega_T^i$<br>$S_{FA} : \begin{pmatrix} p > 2 \\ u_n \\ \Delta_n, k\Delta_n \end{pmatrix}$  | $H_0 : \Omega_T^i$<br>$H_1 : \Omega_T^f$<br>$S_{IA} : \begin{pmatrix} p > 2, p' > 2 \\ u_n, \gamma u_n \\ \Delta_n \end{pmatrix}$      |
| Brownian Motion: Present or Not                | $H_0 : \Omega_T^W$<br>$H_1 : \Omega_T^{noW}$<br>$S_W : \begin{pmatrix} p < 2 \\ u_n \\ \Delta_n, k\Delta_n \end{pmatrix}$ | $H_0 : \Omega_T^{noW}$<br>$H_1 : \Omega_T^W$<br>$S_{noW} : \begin{pmatrix} p = 0, p' = 2 \\ u_n, \gamma u_n \\ \Delta_n \end{pmatrix}$ |
| Relative Magnitude of the Components           | $\%QV : \begin{pmatrix} p = 2 \\ u_n \\ \Delta_n \end{pmatrix}$   |  |
| Estimating the Degree of Jump Activity $\beta$ | $\hat{\beta} : \begin{pmatrix} p = 0 \\ U : u_n, \gamma u_n \\ \Delta_n \end{pmatrix}$                                    | $\hat{\beta}' : \begin{pmatrix} p = 0 \\ U : u_n \\ \Delta_n, k\Delta_n \end{pmatrix}$   |

**Table I**

**Combinations of  $(p, u, \Delta)$  employed to implement the complete spectrogram methodology: test for the presence of jumps, test whether jumps have finite or infinite activity, test whether a continuous component is present, estimate the relative magnitude of the components, and estimate the degree of jump activity.**