ARTICLE IN PRESS

Journal of Econometrics ▮ (▮▮▮▮) ▮▮▮–▮▮▮

Contents lists available at ScienceDirect

Journal of Econometrics

journal homepage: www.elsevier.com/locate/jeconom

# Activity signature functions for high-frequency data analysis☆

Viktor Todorov[a], George Tauchen[b,*]

[a] Northwestern University, USA
[b] Duke University, USA

ARTICLE INFO

*Article history:*
Received 2 June 2009
Received in revised form
2 June 2009
Accepted 22 June 2009
Available online xxxx

*JEL classification:*
C01
C14
C32
G1

*Keywords:*
Activity index
Blumenthal–Getoor index
Jumps
Lévy process
Realized power variation

ABSTRACT

We define a new concept termed activity signature function, which is constructed from discrete observations of a continuous-time process, and derive its asymptotic properties as the sampling frequency increases. We show that the function is a useful device for estimating the activity level of the underlying process and in particular for deciding whether the process contains a continuous martingale. An application to $/*DM* exchange rate over 1986–1999 indicates that a jump-diffusion model is more plausible than a pure-jump model. A second application to internet traffic at NASA servers shows that an infinite variation pure-jump model is appropriate for its modeling.



## 1. Introduction

We consider measuring and estimating the activity level of an Itô semimartingale, which is an analytically convenient probability model for many stochastic processes evolving in continuous time. The index of activity of an Itô semimartingale is an extension of the classical Blumenthal–Getoor index (and its generalization proposed in Ait-Sahalia and Jacod (2009b)) of a pure-jump process, where the leading example is a stable process. The (generalized) Blumenthal–Getoor index for a jump process lies in the interval [0, 2], and it indicates whether the jumps are relatively quiescent or highly vibrant. For example, when the jump process is finitely active, i.e. with only a finite number of jumps in any finite time interval, the Blumenthal–Getoor index is zero. Jump processes with Blumenthal–Getoor indices above zero are infinitely active, with paths of finite variation or infinite variation depending upon where the index lies relative to unity. The extension of the (generalized) Blumenthal–Getoor index to Itô semimartingales considered in this paper also lies in the interval [0, 2], and it is also a measure for the vibrancy of the process. Furthermore, it applies also to continuous path processes. The activity index of a continuous (local) martingale is always 2, i.e. the highest possible value, and thus it dominates that of all jump processes. The index allows us to classify the different processes used in continuous-time modeling in the following order from low to high activity: finite activity jumps, finite variation (but infinite activity) jumps, absolutely continuous processes, jumps of infinite variation and continuous martingales. If the underlying process is the superposition of several different processes from the above list, then the measure will equal that of the most active component. Thus, for example if the activity index takes a value of 2, the process contains a continuous martingale, while otherwise it is of a pure-jump type.

The focus of the paper is on the *activity signature function*, defined below, that proves useful for making non-parametric inference about the generalized activity index from a finite sample. The estimation strategy is simple. We first compute for an arbitrary interval of time the realized power variation over two different time scales, a strategy derived from Ait-Sahalia and Jacod (2009a) and Ait-Sahalia and Jacod (2009b) (see also Zhang et al. (2005)). The realized power variation is just the sum of the absolute values

☆ We thank Cecilia Mancini, Per Mykland, Neil Shephard, Jeannette Woerner as well as seminar participants at Duke University, University of Chicago workshop on finance and statistics, Imperial College 2008 Financial Econometrics Conference, 2008 Oxford Vast Data Conference and Purdue University Financial Mathematics seminar for many helpful comments.

* Corresponding address: Box 90097, Duke University, Durham, NC, 27708, USA.
*E-mail addresses:* v-todorov@kellogg.northwestern.edu (V. Todorov), george.tauchen@duke.edu (G. Tauchen).

0304-4076/$ – see front matter © 2009 Elsevier B.V. All rights reserved.
doi:10.1016/j.jeconom.2009.06.009

of the increments of the process within the interval raised to some power $p$. We then compute a nonlinear transformation of the ratio of realized power variations (over the two time scales) and plot it as a function of the exponent over the range $p \in (0, 4]$. This function is the activity signature function.

The rate at which the realized power variation converges, as we sample more frequently, depends solely on the activity of the process. Therefore, the ratio of the realized power variations computed over different scales, and hence the activity signature function, identifies the activity of the observed process. Asymptotically, the activity signature function stays flat at the activity level until the power reaches it. For powers above the activity level, the activity signature function will either increase linearly or will continue to stay flat, depending on whether the process contains jumps or not. Thus, the overall behavior of the activity signature function reveals the appropriate model for both the "small" and "big" moves in the discretely-observed process. We synthesize the information in the activity signature function in a point estimator for the activity level and a formal test for the presence of continuous martingale, both of which are tested on simulated data.

There are several antecedents in the literature pertaining to the results in this paper. Woerner (2006) initially had the insight of examination of plots of a different function of the level of realized power variation in order to make inference about the Blumenthal–Getoor index of a jump process; but this approach entails a significant bias, as documented in a supplementary web-available appendix (available at http://www.econ.duke.edu/~get/AC_appendix/appendix.pdf). The two-scale approach removes this bias as seen below. Also, the underlying results of Woerner (2006) are derived under the assumption that the driving jump measure is independent from the time-varying intensity or jump size, and this rules out some interesting models, in particular many of the stochastic volatility models used in finance. Second, for the case of $p = 4$ our statistic is a nonlinear transformation of the test statistic for jumps proposed by Ait-Sahalia and Jacod (2009a). Finally, our focus on the activity level of the entire process differs from that of Ait-Sahalia and Jacod (2009b), who consider the challenging problem of estimating the generalized Blumenthal–Getoor index of the jumps in the presence of a continuous martingale. Intuitively, while our estimation strategy identifies the most active part of the discretely-observed process, the estimator of Ait-Sahalia and Jacod (2009b) recovers the least active component of the sum of two processes, given that the index of the dominant component equals 2.

Our estimation strategy can be useful for discriminating nonparametrically across classes of models of stochastic processes. For instance, in financial economics one needs to model both the asset price and the stochastic volatility. An alternative to the classical model, which has a continuous martingale component, is a pure-jump model for which the price or volatility is comprised solely of jumps. The idea behind the pure-jump modeling is that small jumps can eliminate the need for a continuous martingale. The class of pure-jump models is extremely wide. It includes the normal inverse Gaussian (Rydberg, 1997; Barndorff-Nielsen, 1997, 1998), the variance gamma (Madan et al., 1998), the CGMY model of Carr et al. (2002), the time-changed Lévy models of Carr et al. (2003), the COGARCH model of Klüppelberg et al. (2004) for the financial prices as well as the non-Gaussian Ornstein–Uhlenbeck based models of Barndorff-Nielsen and Shephard (2001) and the Lévy-driven continuous-time moving average (CARMA) models of Brockwell (2001) for the stochastic volatility. Pure-jump models have been extensively considered and used for general options pricing (Huang and Wu, 2004; Broadie and Detemple, 2004; Levendorskii, 2004; Schoutens, 2006; Ivanov, 2007), and for foreign exchange options pricing (Huang and Hung, 2005; Daal and Madan, 2005; Carr and Wu, 2007). Other applications of pure-jump models include reliability theory (Drosen, 1986), insurance valuation (Ballotta, 2005), and financial equilibrium analysis (Madan, 2005).

The generalized index of activity of a pure-jump model is strictly less than 2, while the index of the classical jump-diffusion model of finance equals 2. The two classes of models are disjoint, and one does not know *a priori* which is empirically more plausible. The non-parametric evidence in Section 6 comes down on the side of the classical model for the spot $/DM exchange rate over 1986–1999.

A second application uses high-frequency internet traffic data on downloads from NASA servers. The outcome is quite different. We find evidence that these data are best described by a model with jumps, no continuous component, and paths of infinite variation. The sharp contrasts between the outcomes obtained by using high-frequency financial data or internet traffic data illustrate the capability of the estimation strategy to differentiate across the classes of models.

The remainder of the paper is organized as follows. Sections 2 and 3 are theoretical, and a reader only interested in understanding the computations and how to apply them can skip immediately to Section 4. In the theoretical sections we introduce the different types of continuous-time models that our analysis encompasses and the assumptions needed for the main results in the paper. We also define an activity index for a continuous-time process and introduce the activity signature function and characterize its asymptotic behavior using fill-in asymptotics. In the more practical Section 4, we describe the basics of the computations and illustrate how to apply them graphically. In Section 5 we propose a very simple point estimator of the activity index, and we summarize an extensive Monte Carlo study of this estimator and a statistical test for the presence of a continuous component. In Section 6 we undertake two empirical applications, and in Section 7 we summarize with concluding remarks.

## 2. Setup

Our goal in this paper is to measure the "activity" (defined formally in the next section) of a continuous-time one-dimensional stochastic process from discrete observations. We start our analysis with introducing the different models for the discretely-observed process, whose activity we investigate, and stating the assumptions needed for our asymptotic results. Throughout, we always assume implicitly that each of the processes given below is defined on some probability space $(\Omega, \mathcal{F}, \mathbb{P})$. Further, we equip this probability space with some filtration $\mathbf{F}$, with respect to which all the processes used in the definitions of the different models below are adapted.

In all models the underlying process is an *Itô semimartingale*, i.e., a semimartingale whose characteristics, drift, diffusion and jump compensator, are absolutely continuous with respect to time (Jacod and Shiryaev, 2003). The different models differ in whether the stochastic process contains continuous martingale and/or jumps. In what follows we will always assume that the processes in each of the models are well defined, for "classical" conditions (Cont and Tankov, 2003; Jacod and Shiryaev, 2003).

*Continuous model*:

$$Y_t = \int_0^t b_{1s}\mathrm{d}s + \int_0^t \sigma_{1s}\mathrm{d}W_s, \tag{1}$$

and $W_t$ is a standard Brownian motion.

*Pure-jump model*:

$$X_t = \int_0^t b_{2s}\mathrm{d}s + \int_0^t \int_{\mathbb{R}} \sigma_{2s-}\kappa(x)\tilde{\mu}(\mathrm{d}s, \mathrm{d}x) + \int_0^t \int_{\mathbb{R}} \sigma_{2s-}\kappa'(x)\mu(\mathrm{d}s, \mathrm{d}x), \tag{2}$$

Please cite this article in press as: Todorov, V., Tauchen, G., Activity signature functions for high-frequency data analysis. Journal of Econometrics (2009), doi:10.1016/j.jeconom.2009.06.009

where $\mu$ is a jump measure on $\mathbb{R}_+ \times \mathbb{R}$ with compensator $a_s\mathrm{d}s \otimes \nu(x)\mathrm{d}x$, $\tilde{\mu}(\mathrm{d}s, \mathrm{d}x) := \mu(\mathrm{d}s, \mathrm{d}x) - a_{s-}\mathrm{d}s\nu(x)\mathrm{d}x$; $\kappa'(x) = x - \kappa(x)$ and $\kappa(x)$ is a continuous truncation function, i.e. a continuous function with bounded support and which coincides with the identity around the origin; $a_t$ is some non-negative process. The two processes $\sigma_{2t}$ and $a_t$ in Eq. (2) generate stochastic volatility in $X_t$, but in two different ways. $\sigma_{2t}$ generates stochastic volatility through time-varying jump size, while $a_t$ generates stochastic volatility through time-varying the intensity of the jumps. We refer to the absolutely continuous processes $\int_0^t b_{1s}\mathrm{d}s$ and $\int_0^t b_{2s}\mathrm{d}s$ in (1) and (2) as the drift terms in $Y$ and $X$ respectively.

*Continuous plus jumps model*:

$$Z_t = X_t + Y_t. \tag{3}$$

We next state the assumptions needed for the asymptotic results in the paper.

**Assumption A1.** (a) The processes $\sigma_{1s}$ and $b_{1s}$ have càdlàg paths; the process $\sigma_{1s}$ is everywhere different from zero.
(b) The processes $\sigma_{2s}$, $a_s$ and $b_{2s}$ have càdlàg paths; both of the processes $\sigma_{2s}$ and $a_s$ are everywhere different from zero.

**Assumption A2.** The Lévy density $\nu(x)$ can be decomposed as

$$\nu(x) = \nu_1(x) + \nu_2(x), \tag{4}$$

$$\nu_1(x) = \frac{A}{x^{\beta+1}}1_{\{x>0\}} + \frac{B}{|x|^{\beta+1}}1_{\{x<0\}}, \quad |\nu_2(x)| \le \frac{\phi(x)}{|x|^{\beta'+1}}, \tag{5}$$

where $A$ and $B$ are some non-negative constants with $A + B > 0$; $0 \le \beta' < \beta < 2$; and $\phi(x)$ is some non-negative, slowly-varying at zero function which is bounded at zero.

**Assumption A3.** In addition to Assumption A2, assume that

(a) If $\beta < 1$, then

$$\int_0^t b_{2s}\mathrm{d}s - \int_0^t \int_{\mathbb{R}} a_s\sigma_{2s}\kappa(x)\mathrm{d}s\nu(x)\mathrm{d}x \equiv 0, \quad \text{for every } t > 0. \tag{6}$$

(b) If $\beta = 1$, then $\nu(x)$ and $\kappa(x)$ are symmetric and $b_{2t} \equiv 0$ for every $t > 0$.

Assumption A1 is a very mild regularity assumption and it is satisfied by most parametric applications. On the other hand Assumption A2 is a non-trivial assumption that we impose on the behavior of $\nu(x)$ around the origin. In the Lévy case it essentially amounts to restricting the Lévy density to be a power function around zero like in the (tempered) stable processes. In other words we control the way in which $\nu(x)$ increases to infinity as $x$ approaches zero. This is closely related with the Blumenthal–Getoor index which we explain in the next section. An assumption similar to A2 has been made also in Jacod (2004), Woerner (2006) and Ait-Sahalia and Jacod (2009b). Finally, Assumption A3 guarantees that when the jumps are not very active (in a sense defined in the next section), the drift term resulting from the compensation of the small jumps and/or the process $b_{2s}$ is not dominating the activity in the pure-jump model. Under Assumption A3, the process $X_t$ is a sum of its jumps. We will need A3 in order to discriminate the activity behavior of the different pure-jump processes.

## 3. Asymptotic results

For all our theoretical results we will fix the time interval to be $[0, T]$ and we will suppose that we observe the process $\Upsilon$ (where $\Upsilon \equiv X, Y$ or $Z$) at the equidistant times $0, \Delta_n, 2\Delta_n, \ldots, [T/\Delta_n]\Delta_n$ within this interval. We will think of $\Delta_n$ as being "small", i.e. our asymptotic results will be for $\Delta_n \to 0$, and in this section we will derive statistics for the activity of the discretely-observed process, which we now formally define. We first introduce the realized power variation constructed from the discrete observations of the process in the following way

$$V(p, \Upsilon, \Delta_n)_t = \sum_{i=1}^{[t/\Delta_n]} |\Delta_i^n\Upsilon|^p, \quad p > 0, t \in (0, T], \tag{7}$$

where $\Delta_i^n\Upsilon := \Upsilon_{i\Delta_n} - \Upsilon_{(i-1)\Delta_n}$. Then we associate (index) the activity of the *observed path* of $\Upsilon$ with

$$\beta_{\Upsilon,T} := \inf\left\{r > 0 : \operatorname*{plim}_{\Delta_n\to 0} V(r, \Upsilon, \Delta_n)_T < \infty\right\}. \tag{8}$$

There are few things to note about this index of the activity. First, it is defined pathwise. Second, it is always in the interval $[0, 2]$. When $\Upsilon \equiv Y$, it is well known that the activity index equals 2 on every path (up to paths with measure zero), see e.g. Revuz and Yor (1999). For an absolutely continuous process the index is always 1. When $\Upsilon$ is the pure-jump process $X$, things are more complicated. As a simple example we can consider the case when $\Upsilon$ is a pure-jump Lévy process. Then the index in (8) will be the same on every observed path (up to paths with measure 0) and it will coincide with the Blumenthal–Getoor index of the Lévy measure, see (10) below. Thus for a stable process with index $\alpha$, our activity index in (8) will coincide with $\alpha$ and therefore can take all the values in the range $[0, 2]$. Another example is the compound Poisson process often used in empirical applications; its activity index is 0. Finally, the process $Z$ defined in Eq. (3) is a sum of continuous martingale (driven by Brownian motion) and jumps, therefore its activity is determined by the most "active" part of it which is the continuous martingale part.

Our goal will be to measure non-parametrically the activity of the discretely-observed continuous-time process. It is evident from the discussion above that the activity index will allow us to discriminate between the different classes of models defined in the previous section. For example, if the estimated activity index is in the interval $(0, 1)$, that means that the observed process is from a pure-jump model with no drift and with jumps having finite variation. If the estimated index is in the interval $(1, 2)$, then the appropriate model for the observed process is again a pure-jump model, but now with jumps exhibiting infinite variation. Finally, if the estimated activity is 2, then to model appropriately the very small moves in the discretely-observed process, we need a continuous martingale.

### 3.1. Asymptotics of realized power variation

To determine the activity index and develop a strategy for its estimation from discrete observations of the process $\Upsilon$, we need first to know the asymptotic behavior of the realized power variation for the different models and different powers. In what follows $\xrightarrow{u.c.p.}$ denotes convergence in probability, locally uniformly in time.

#### 3.1.1. When $\Upsilon \equiv Y$

When the (partially) observed process is a continuous martingale (plus a drift), then it follows from Barndorff-Nielsen et al. (2005) that under Assumption A1(a) and for $p > 0$

$$\Delta_n^{1-p/2}V(p, Y, \Delta_n)_t \xrightarrow{u.c.p.} A_p\int_0^t |\sigma_{1u}|^p\mathrm{d}u, \tag{9}$$

where $A_p$ is some constant depending only on $p$.

Please cite this article in press as: Todorov, V., Tauchen, G., Activity signature functions for high-frequency data analysis. Journal of Econometrics (2009), doi:10.1016/j.jeconom.2009.06.009

*3.1.2. When $\Upsilon \equiv X$*

To avoid trivial situations we will assume that on the observed path $X$ contains at least one jump. When jumps dominate the activity of the discretely-observed process, our activity index can be expressed directly as a function of the jumps on the (partially) observed path. In other words, when $\Upsilon \equiv X$ (and Assumption A3 holds in the case of finite variation jumps), the activity index in (8) coincides with the following generalization of the Blumenthal–Getoor index due to Ait-Sahalia and Jacod (2009b)

$$BG(X)_T := \inf\left\{r > 0 : \sum_{0\le s\le T} |\Delta X_s|^r < \infty\right\}, \tag{10}$$

where $\Delta X_s := X_s - X_{s-}$. The Blumenthal–Getoor index was originally defined in Blumenthal and Getoor (1961) only for pure-jump Lévy processes. The definition in (10) extends it to an arbitrary jump semimartingale and was proposed in Ait-Sahalia and Jacod (2009b). It is implicit in the definition that $BG_T(X)$ depends on the time interval $[0, T]$ and also on the particular realization of the process, i.e. the index is defined *pathwise*. The behavior of the index is determined by the behavior of the small jumps (jumps bigger than any *fixed* size are always finite on a finite time interval and therefore are absolutely summable for any power $r$). So, what is important for our determination of the activity is essentially the very small increments $\Delta_i^n X$. This is unlike the problem of detecting jumps in discretely-sampled process, where we look to discriminate the very "small" moves (associated with the continuous part of the process) from the very "big" ones (associated with the discontinuous part of the process). Finally, under Assumption A2, the generalized Blumenthal–Getoor index $BG_T(X)$ will coincide with $\beta$ in Eq. (5). Therefore, under our Assumption A2, the generalized Blumenthal–Getoor index will be constant for each different realization of the process.

Turning to the behavior of the realized power variation, we can summarize it as follows. First, if $BG(X)_T > 1$, then on the interval $[0, T]$ we have

$$\begin{cases} \Delta_n^{1-p/\beta} V(p, X, \Delta_n)_t \xrightarrow{u.c.p.} \int_0^t |\sigma_{2u}|^p g_p(a_s)\mathrm{d}u, \\ \quad \text{if } p < BG(X)_T, \text{ under A1(b) and A2} \\ V(p, X, \Delta_n)_t \xrightarrow{u.c.p.} \sum_{s\le t} |\Delta X_s|^p, \quad \text{if } p > BG(X)_T, \text{ under A1(b),} \end{cases} \tag{11}$$

where $g_p(\cdot)$ is a known function related with the absolute moments of the stable process, defined in the web-available appendix. When $BG(X)_T < 1$, we have to discriminate between the case when a drift term (or compensator for the jumps) is present or not. Thus, when $BG(X)_T < 1$ on the interval $[0, T]$ we have

$$\begin{cases} \Delta_n^{1-p} V(p, X, \Delta_n)_t \xrightarrow{u.c.p.} \int_0^t \left| b_{2s} - a_s\sigma_{2s}\int_{\mathbb{R}} \kappa(x)\nu(x)\mathrm{d}x \right|^p \mathrm{d}s, \\ \quad \text{if } p < 1, \text{ under A1(b) and A2} \\ V(p, X, \Delta_n)_t \xrightarrow{u.c.p.} \int_0^t \left| b_{2s} - a_s\sigma_{2s}\int_{\mathbb{R}} \kappa(x)\nu(x)\mathrm{d}x \right| \mathrm{d}s \\ \quad + \sum_{s\le t} |\Delta X_s|, \quad \text{if } p = 1, \text{ under A1(b)} \\ V(p, X, \Delta_n)_t \xrightarrow{u.c.p.} \sum_{s\le t} |\Delta X_s|^p, \quad \text{if } p > 1, \text{ under A1(b),} \end{cases} \tag{12}$$

and

$$\begin{cases} \Delta_n^{1-p/\beta} V(p, X, \Delta_n)_t \xrightarrow{u.c.p.} \int_0^t |\sigma_{2u}|^p g_p(a_s)\mathrm{d}u, \\ \quad \text{if } p < BG(X)_T, \text{ under A1(b), A2 and A3} \\ V(p, X, \Delta_n)_t \xrightarrow{u.c.p.} \sum_{s\le t} |\Delta X_s|^p, \\ \quad \text{if } p > BG(X)_T, \text{ under A1(b), A2 and A3,} \end{cases} \tag{13}$$

That is, when jumps are of finite variation and a drift term is present, the latter will "dominate" the jumps and will determine the activity of $X$. The limiting behavior for the realized power variation when $p > BG(X)_T$ was derived in Jacod (2008), see also Lepingle (1976). It holds under fairly weak conditions. The case $p < BG(X)_T$ is more complicated and we need to assume the stronger Assumption A2 (under which $BG_T(X) \equiv \beta$ and is therefore nonrandom). The limiting behavior of the realized power variation in this case follows from Theorem 1 in the web-available appendix. A precursor special case of this theorem has been proved in Woerner (2003), but only for the case when $a_s \equiv 1$ and $\sigma_{2s}$ is independent from the jump measure $\mu$ (and more restrictive specification for $\nu(x)$). Also related is a result in Ait-Sahalia and Jacod (2009b) on the asymptotic behavior of the count of the increments bigger than a decreasing threshold.

*3.1.3. When $\Upsilon \equiv Z$*

We are finally left with the case when the observed process is a superposition of a continuous martingale, jumps and drift. We will assume that $Z$ contains at least one jump on the observed path, otherwise the behavior of the realized power variation is as in the pure-continuous model. When $p < 2$, the limiting behavior of the realized power variation is determined by the most active component in $Z$, which is the continuous martingale. When $p = 2$, then both the jumps and the continuous martingale determine the behavior of the realized power variation. Finally, when $p > 2$, the limiting behavior of the realized power variation is governed by the jumps. The precise results are as follows. Under Assumption A1 we have

$$\begin{cases} \Delta_n^{1-p/2} V(p, Z, \Delta_n)_t \xrightarrow{u.c.p.} A_p \int_0^t |\sigma_{1u}|^p \mathrm{d}u, & \text{if } 0 < p < 2 \\ V(2, Z, \Delta_n)_t \xrightarrow{u.c.p.} \int_0^t \sigma_{1u}^2 \mathrm{d}u + \sum_{s\le t} |\Delta Z_s|^2, & \text{if } p = 2 \\ V(p, Z, \Delta_n)_t \xrightarrow{u.c.p.} \sum_{s\le t} |\Delta Z_s|^p, & \text{if } p > 2, \end{cases} \tag{14}$$

where $A_p$ is the same constant that appears in Eq. (9). These results are trivial consequences of (or follow directly from) the results in Barndorff-Nielsen et al. (2005), Jacod (2008) and Barndorff-Nielsen et al. (2006).

*3.2. Measuring activity via the activity signature function*

Having characterized the asymptotic behavior of the realized power variation for the different models, we are now ready to develop strategy for inferring the activity index from the discrete observations of the process. The idea is to compute the realized power variation at two different frequencies and use the fact that the scaling factors at the two different frequencies will differ, provided the activity index is above the power that is used. This two-scale approach was first proposed in Ait-Sahalia and Jacod (2009a) for designing tests for presence of jumps and further used in Ait-Sahalia and Jacod (2009b) for estimation of the generalized Blumenthal–Getoor index of the jumps. To this end set $0 < p < \beta_{\Upsilon,T}$, then for arbitrary integer $k \ge 1$ and under A1–A3 we have

$$k^{1-p/\beta_{\Upsilon,T}} \Delta_n^{1-p/\beta_{\Upsilon,T}} V(p, \Upsilon, k\Delta_n)_T \xrightarrow{\mathbb{P}} C_T(p),$$

where $C_t(p)$ is some stochastic process, depending on $\Upsilon$. This suggests that if we fix $p$ sufficiently low and compute the $p$th realized power variation over different sampling frequencies we can recover $\beta_{\Upsilon,T}$ from the slope coefficient in a regression of $\{\ln(V(p, \Upsilon, \Delta_n)_T)\}_k$ on a constant and $\{\ln(k\Delta_n)\}_k$. For example, in

Please cite this article in press as: Todorov, V., Tauchen, G., Activity signature functions for high-frequency data analysis. Journal of Econometrics (2009), doi:10.1016/j.jeconom.2009.06.009

the case when we use only the two sampling frequencies $\Delta_n$ and $k\Delta_n$ we can determine $\beta_{\Upsilon,T}$ using

$$\widehat{\beta}_{(0,T]}(\Upsilon, p) = \frac{\ln(k)\, p}{\ln(k) + \ln(V(p, \Upsilon, k\Delta_n)_T) - \ln(V(p, \Upsilon, \Delta_n)_T)}. \tag{15}$$

We note that our sample function $\widehat{\beta}_{(0,T]}(\Upsilon, p)$ makes use of all increments of the process $\Upsilon$. In fact the small increments carry the most important information about the activity index. Our measure is thus different from the statistic proposed in Ait-Sahalia and Jacod (2009b), which is essentially based on the count of the increments of the discretely-observed process that are bigger than a threshold level (the threshold level decreases to zero at a given rate). While (15) is estimating the activity index in (8), their estimator is designed to infer the generalized Blumenthal–Getoor index of the jumps in the (possible) presence of a continuous component (i.e. they estimate (10) when $\Upsilon \equiv Z$). Thus, since the continuous martingale dominates the activity of the process $Z = X + Y$, Ait-Sahalia and Jacod (2009b) need to discard the very small increments which are dominated by the continuous martingale. In contrast, when $\Upsilon \equiv Z$ (15) uses the very small increments and estimates the activity of the dominating component in the observed process, i.e. that of the continuous martingale. Of course, our estimator and the one in Ait-Sahalia and Jacod (2009b) will have the same asymptotic limit only in the pure-jump case, i.e. when $\Upsilon \equiv X$.

We can look at $\widehat{\beta}_{(0,T]}(\cdot, p)$ as a function of $p$ and analyze the behavior of this function under the three different models introduced in Section 2. In the following theorem by local uniform convergence on a given (open) set we mean convergence that is uniform on each compact subset of that set.

**Theorem 1.** (a) *Continuous semimartingale: Suppose $\Upsilon \equiv Y$ and Assumption A1(a) holds. Then for fixed $T > 0$ we have*

$$\widehat{\beta}_{(0,T]}(\Upsilon, p) \xrightarrow{\mathbb{P}} \begin{cases} 2 & \text{if } p \le 2 \\ 2 & \text{if } p > 2, \end{cases} \tag{16}$$

*where the convergence is locally uniform in $p$ on $(0, \infty)$.*

(b) *Pure-jump semimartingale: Suppose $\Upsilon \equiv X$ and Assumptions A1(b), A2 and A3 hold. Then for fixed $T > 0$ we have*

$$\widehat{\beta}_{(0,T]}(\Upsilon, p) \xrightarrow{\mathbb{P}} \begin{cases} \beta & \text{if } p < \beta \\ p & \text{if } p > \beta, \end{cases} \tag{17}$$

*where the convergence is locally uniform in $p$ on $(0, \beta) \cup (\beta, \infty)$.*

(c) *Continuous plus jumps semimartingale: Suppose $\Upsilon \equiv Z$, Assumption A1 holds and there is at least one jump on the observed path. Then for fixed $T > 0$ we have*

$$\widehat{\beta}_{(0,T]}(\Upsilon, p) \xrightarrow{\mathbb{P}} \begin{cases} 2 & \text{if } p \le 2 \\ p & \text{if } p > 2, \end{cases} \tag{18}$$

*where the convergence is locally uniform in $p$ on $(0, 2) \cup (2, \infty)$.*

From this theorem it follows that the activity of the discretely-observed process can be inferred from the overall behavior of $\widehat{\beta}_{(0,T]}(\cdot, p)$. We therefore refer to $\widehat{\beta}_{(0,T]}(\cdot, p)$ as the *activity signature function*. The function identifies the activity through: (1) its level in the flat part and (2) the range of powers $p \in (0, 2]$ for which it stays flat. On the other hand, higher values of $p$ can allow us to determine whether there is a kink at 2 in the activity signature function $\widehat{\beta}_{(0,T]}(\cdot, p)$ or its stays flat, which in turn determines whether the process contains jumps in addition to the continuous martingale (the test for jumps of Ait-Sahalia and Jacod (2009a) is a monotone transformation of $\widehat{\beta}_{(0,T]}(\cdot, p)$ for $p = 4$). This means that using the activity signature function we can discriminate between the three types of models defined in Section 2 for the discretely-observed process.

The asymptotic behavior of the activity signature function can be used to derive point estimators of the unknown activity level of the discretely-observed process. Such estimators will rely on the flat part of the function, which in turn can be determined in some adaptive way. Key for that is that the convergence of the activity signature function in Theorem 1 is (locally) uniform (in $p$). In Section 5.1 we will illustrate with a particular activity estimator.

### 3.3. Testing jump diffusions versus pure-jump models using the activity signature function

An activity level of 2 has a special meaning since it separates pure-jump models (part (b) of Theorem 1) from models containing continuous martingale (parts (a) and (c) of Theorem 1). As noted in the Introduction, pure-jump models have been proposed as an alternative to standard jump diffusions in fields such as finance and insurance. Thus, it is useful to have a formal test for the presence of a continuous component against the alternative of a pure-jump process. We construct such a test using the activity signature function evaluated at a *fixed* power. The next theorem gives the relevant null distribution for this test. In the theorem, $\xrightarrow{\mathcal{L}-s}$ denotes stable convergence in law.

**Theorem 2.** *Set $\Upsilon \equiv Y$ or $\Upsilon \equiv Z$. Assume that A1 holds and that $BG_T(X) < 1$ and $p \in \left(\frac{BG_T(X)}{2-BG_T(X)}, 1\right)$, where $BG_T(X)$ is the Blumenthal–Getoor index of the jumps in $Z$ defined in (10) (which can be random). Then, if the process $\sigma_1$ is an Itô semimartingale with locally bounded coefficients, we have*

(a)

$$\Delta_n^{-1/2}\left(\log(\widehat{\beta}_{(0,T]}(\Upsilon, p)) - \log(2)\right) \xrightarrow{\mathcal{L}-s} K_p \times \epsilon, \tag{19}$$

$$K_p = \frac{2}{p\mu_p \ln k} \frac{\sqrt{\int_0^T |\sigma_{1u}|^{2p} \mathrm{d}u}}{\int_0^T |\sigma_{1u}|^p \mathrm{d}u} \times \sqrt{(k+1)\mu_{2p} - 2k^{1-p/2}\mu_p(k) + (k-1)\mu_p^2}, \tag{20}$$

*where $\epsilon$ is standard normal defined on an extension of the original probability space; for the definition of $\mu_p$ and $\mu_p(k)$ introduce $u_1$ and $u_2$ two independent standard normal variables, then we set $\mu_p = \mathbb{E}|u_1|^p$ and $\mu_p(k) = \mathbb{E}\left(|u_1|^p |u_1 + \sqrt{k-1}u_2|^p\right)$.*

(b) *A consistent estimator for $K_p$ in (20) under the conditions of the Theorem is given by*

$$\widehat{K}_p = \frac{2\Delta_n^{-1/2}}{p\mu_p \ln k} \frac{\sqrt{\sum_{i=4}^{[T/\Delta_n]} |\Delta_i^n \Upsilon|^{p/2} |\Delta_{i-1}^n \Upsilon|^{p/2} |\Delta_{i-2}^n \Upsilon|^{p/2} |\Delta_{i-3}^n \Upsilon|^{p/2}}}{\sum_{i=2}^{[T/\Delta_n]} |\Delta_i^n \Upsilon|^{p/2} |\Delta_{i-1}^n \Upsilon|^{p/2}} \times \sqrt{(k+1)\mu_{2p} - 2k^{1-p/2}\mu_p(k) + (k-1)\mu_p^2}. \tag{21}$$

The above theorem can be used to conduct a one-sided test of the statistical significance of the discrepancy between $\log[\widehat{\beta}_{(0,T]}(\Upsilon, p)]$ and the null value of $\log(2)$; negative values of the studentized left-hand side of (19) discredit the null hypothesis of the presence of a continuous component. For practical purposes a value of $p$ very close to 1 is desirable, so that the requirement for $p$ in the theorem is satisfied, when $Z$ contains jumps of finite variation. When $Z$ contains jumps of infinite variation (19) does not hold, since the jumps slow down the rate of convergence. Therefore in this case the one-sided test based on Theorem 2 will be rather conservative. In Section 5.2 we provide Monte Carlo evidence on the performance of the test.

Please cite this article in press as: Todorov, V., Tauchen, G., Activity signature functions for high-frequency data analysis. Journal of Econometrics (2009), doi:10.1016/j.jeconom.2009.06.009

## 4. Computations and illustrations

### 4.1. Practical aspects

We turn now to the practical application of our theoretical results stated in the previous sections. We start by explaining the basic mechanics of the computations involved in the analysis. Suppose we have equi-spaced data for a single unit interval, typically a day in financial econometrics or an hour in our internet example below. Let $x_i$, $i = 1, 2, \ldots, n$, denote the $i$th increment in the process over the period, which would be the log-return, or geometric return, in financial econometrics. Define the realized power variation

$$V(p) = \sum_{i=1}^{n} |x_i|^p \tag{22}$$

and think of it as function of $p \geq 0$. We always find in practice that the using domain $p \in (0, 4]$ exhausts the relevant information in the observed $x_i$. Now define

$$\tilde{V}(p) = \sum_{i=1}^{[n/2]} |x_{2i-1} + x_{2i}|^p \tag{23}$$

which is the realized power using a sampling frequency half that of the original data. The activity signature function is

$$b(p) = \frac{\ln(2)p}{\ln(2) + \ln[\tilde{V}(p)] - \ln[V(p)]} \tag{24}$$

viewed as a function $p$ on the domain $(0, 4]$.[1]

In practice, we have high-frequency data for $N$ intervals, where $N$ can be very large, e.g., 3000 trading days. We can compute for each of the $N$ intervals the signature function (24) (over a fine grid of $p$), yielding $\{\{b_s(p)\}_{p\in(0,4]}\}_{s=1}^{N}$, where $s$ indexes intervals. It is impossible to report all $N$ signature functions, so we look for measures of central tendency and dispersion of the signature functions. Since we are handling data with jumps and outliers, we use a statistically robust approach. For this purpose, define the *Quantile Activity Signature Function* (QASF):

$$b_\alpha(p) = \text{Quantile}_\alpha\left[\{b_s(p)\}_{s=1}^{N}\right], \quad \text{for each } p \in (0, 4] \tag{25}$$

where, in a self-evident notation, $\text{Quantile}_\alpha$ selects the $\alpha$th quantile of a sequence of numbers.[2] In practice we find the quartiles, 0.25, 0.50, 0.75, which are commonly used in statistics, to be very informative.

### 4.2. Signature functions on simulated data

We now show the general methodology by computing signature functions on simulated data using the following model for the discretely-observed process $\Upsilon$:

$$\mathrm{d}\Upsilon_t = \rho\Upsilon_t + \sigma_1\,\mathrm{d}W_t + \sigma_2\,\mathrm{d}\mathcal{L}_t \tag{26}$$

where the first term represents the drift, which is absent if $\rho = 0$, the second is the Brownian component, absent if $\sigma_1 = 0$, and the third represents the increment of a pure-jump Lévy process $\mathcal{L}$, which is absent if $\sigma_2 = 0$. The process $\mathcal{L}$ has a zero drift, a zero truncation function, and its compensator for the jumps is either of these two:

$$A\frac{\mathrm{e}^{-\lambda x}}{|x|^{\beta+1}}\mathrm{d}x\mathrm{d}s, \quad \beta \in [0, 2) \quad \text{or,} \quad \lambda_J\delta_{\{x=\pm\tau\}}\mathrm{d}x\mathrm{d}s, \tag{27}$$

where $\delta_\epsilon$ denotes the Dirac point mass.[3] The left compensator in (27) corresponds to a symmetric tempered stable process (Rosiński, 2007; Carr et al., 2002, 2003), also called the CGMY process, and $\beta$ is the Blumenthal–Getoor index. The right-hand compensator in (27) corresponds to a compound Poisson process with intensity $\lambda_J$ and jumps $\tau$ or $-\tau$ with equal probability, frequently termed "rare-jumps". If jumps and a continuous component are present, we fix the proportion of the jumps in the total variance of $\Upsilon$ to be 20%, an upper limit found empirically in finance. If $\rho < 0$, the process $\Upsilon$ is an Ornstein–Uhlenbeck (**OU**) process driven by Brownian motion or jumps. We use $\rho = -0.0693$ corresponding to half-life of 10 units of time; the results are insensitive to the value of $\rho$.

Table 1 contains full details on the simulations. In each scenario we simulate a realization of $N = 3000$ units of time and 288 increments within each unit of time, which correspond to 5 min sampling over a 24 h day. In the first three scenarios the simulated process is the infinitely active tempered stable distribution with parameter $\beta = 0.50,\ 1.0,\ 1.50$ for the stable part.[4] The parameter $\beta$ must be between 0.00 and 2.00. For small values of $\beta$ the process is infinitely active but relatively quiescent, while for larger values above unity it is infinitely active and very vibrant. In the fourth scenario, labeled **C**, the process is Brownian motion and $\beta = 2$. In the next three it is a classical jump diffusion, without drift, and rare jumps that follow a compound Poisson process with three values for the intensity parameter $\lambda_J$ corresponding to low, medium, and high intensity. For the last four cases in Table 1, the prefix **OU** means the process contains a drift, with the background driving Lévy process being the first four Lévy processes of the table. Only in the last of these four cases, labeled **OU-C**, is the process continuous ($\sigma_1 > 0, \sigma_2 = 0$), and it is otherwise purely discontinuous.[5]

Fig. 1 shows the QASFs for a symmetric tempered stable process with activity levels 0.50 and 1.00. The two left-hand panels pertain to the tempered stable only, while in the right-hand panels the tempered stable is the background driving Lévy processes for an **OU** process, so the increments exhibit temporal dependence, and the drift promotes the activity level to at least unity. (See Table 1.) From Theorem 1, the activity signature function of (24) satisfies $b(p) \xrightarrow{\mathbb{P}} \max\{\beta, p\}$ as the sampling frequency increases, uniformly in $p$, for $\forall t$. As one expects from the asymptotic theory, in Fig. 1 the ordinates of the median QASFs about equal the true (population) $\beta$ for abscissa $p < \beta$, and the ordinates equal $p$ for abscissa above $\beta$. Not surprisingly, the QASFs are curvilinear at abscissa $p$ close to $\beta$, since the plots are based on simulations of finite length and the asymptotic limit has not been reached.

In Fig. 2, the simulation settings correspond to a symmetric tempered stable process with activity 1.50 and Brownian motion, which is of activity level 2.00. The two left-hand panels pertain directly to these two processes, while in the right-hand panel

[1] In terms of the theory in Section 3, $x_i = \Delta_i^n\Upsilon$ in Eq. (7), and the activity signature function shown in (24) is (15) with $k = 2$, implying a sampling frequency half that of the original data.

[2] The subscript $\alpha$ in (25) should not be confused with the time index $s$ (which takes integer values). We make this slight abuse of notation in an effort to keep the notation as simple as possible. For the same reason, from now on we use $\beta$ as a short for the activity index $\beta_{\Upsilon,T}$ (with $\beta = 2$ indicating presence of continuous martingale).

[3] Note that because of the symmetry of the jump compensator, we do not need to compensate the small jumps, i.e. we use zero truncation function, even in the infinite variation case when $\beta > 1$.

[4] The tempered stable is simply the classical stable distribution with the tails of the Lévy density "tempered" to attenuate the impact of large jumps and thereby make all moments exist, regardless of the value of $\beta \in [0, 2]$. The tempering function is $\mathrm{e}^{-\lambda|x|}$ and $\lambda = 0.05$ (very gentle tempering) in all scenarios in Table 1.

[5] See Todorov and Tauchen (2006) and references therein for details on discontinuous processes with drift in continuous time.

Please cite this article in press as: Todorov, V., Tauchen, G., Activity signature functions for high-frequency data analysis. Journal of Econometrics (2009), doi:10.1016/j.jeconom.2009.06.009

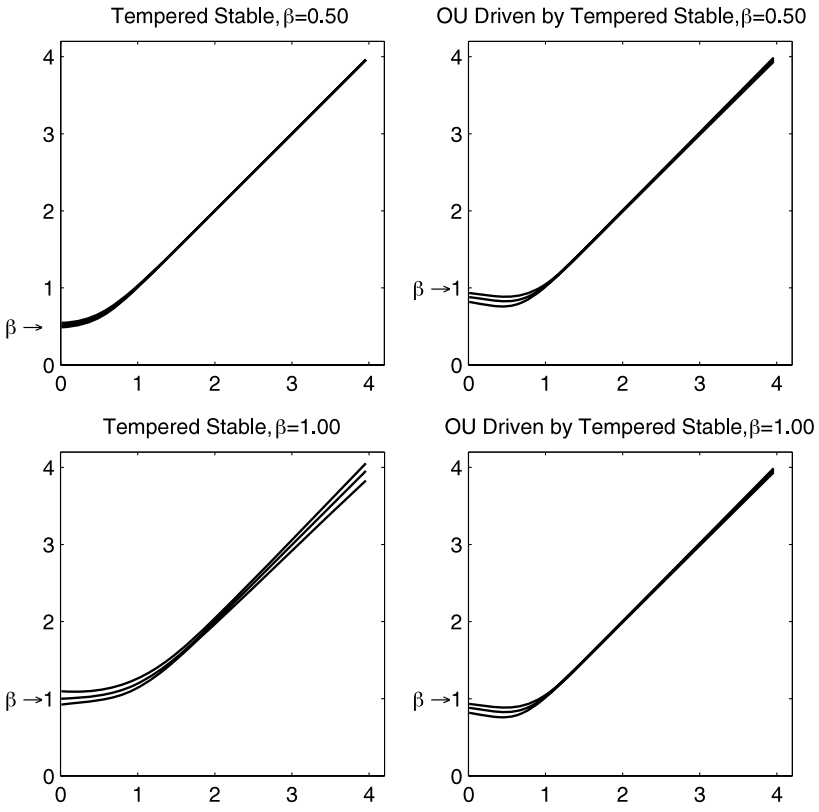


**Fig. 1.** Quantile activity signature plots for processes driven by the symmetric tempered stable process of activity less than or equal to unity. The figures show the QASFs $b_\alpha(p)$ defined in (25) as a function of $p$ for the quantiles $\alpha = 0.25$, $\alpha = 0.50$, and $\alpha = 0.75$. The left-side panels correspond to values of $\beta$ and other parameters given by **Cases TS0.5** and **TS1.0** of Table 1; right-side panels pertain to **Cases OU-TS0.5** and **OU-TS1.0** of Table 1.

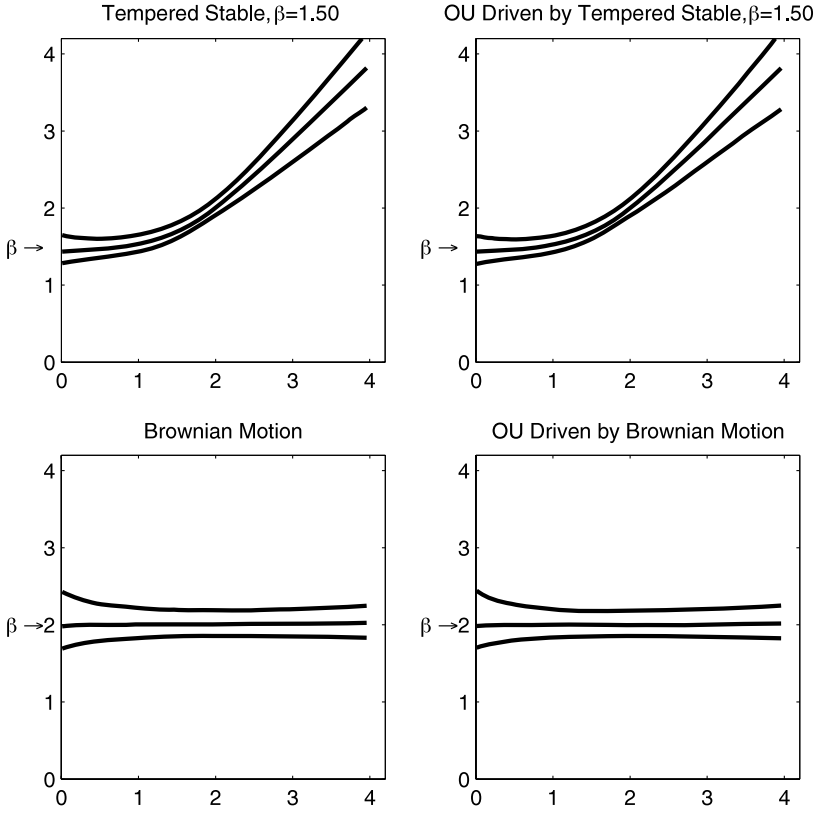


**Fig. 2.** Quantile activity signature plots for processes driven by Lévy processes of infinite variation: the symmetric tempered stable process of activity 1.50 and Brownian motion. The figures show the QASFs $b_\alpha(p)$ defined in (25) as a function of $p$ for the quantiles $\alpha = 0.25$, $\alpha = 0.50$, and $\alpha = 0.75$. The left-side panels correspond to values of $\beta$ and other parameters given by **Cases TS1.5** and **C** of Table 1; the right-side panels pertain to **Cases OU-TS1.5** and **C** of Table 1.

Please cite this article in press as: Todorov, V., Tauchen, G., Activity signature functions for high-frequency data analysis. Journal of Econometrics (2009), doi:10.1016/j.jeconom.2009.06.009

**Table 1**
Parameter setting for the various scenarios.

| Case | $\rho$ | $\sigma_1^2$ | $\sigma_2^2$ | Jump specification |
|---|---|---|---|---|
| **TS0.5** | 0.00 | 0.0 | 1.0 | Tempered stable, $A = 1$, $\beta = 0.5$ and $\lambda = 0.05$ |
| **TS1.0** | 0.00 | 0.0 | 1.0 | Tempered stable, $A = 1$, $\beta = 1.0$ and $\lambda = 0.05$ |
| **TS1.5** | 0.00 | 0.0 | 1.0 | Tempered stable, $A = 1$, $\beta = 1.5$ and $\lambda = 0.05$ |
| **C** | 0.00 | 0.8 | 0.0 | None |
| **C-RJL** | 0.00 | 0.8 | 1.0 | Rare-jump, $\lambda_J = 0.0400$, $\tau = 2.2361$ |
| **C-RJM** | 0.00 | 0.8 | 1.0 | Rare-jump, $\lambda_J = 0.3333$, $\tau = 0.7746$ |
| **C-RJH** | 0.00 | 0.8 | 1.0 | Rare-jump, $\lambda_J = 2.0000$, $\tau = 0.3162$ |
| **OU-TS0.5** | −0.0693 | 0.0 | 1.0 | Tempered stable, $A = 1$, $\beta = 0.5$ and $\lambda = 0.05$ |
| **OU-TS1.0** | −0.0693 | 0.0 | 1.0 | Tempered stable, $A = 1$, $\beta = 1.0$ and $\lambda = 0.05$ |
| **OU-TS1.5** | −0.0693 | 0.0 | 1.0 | Tempered stable, $A = 1$, $\beta = 1.5$ and $\lambda = 0.05$ |
| **OU-C** | −0.0693 | 0.8 | 0.0 | None |

these processes are the background driving Lévy processes for an **OU** process. Again, as expected from the asymptotic theory, the median QASFs appear generally flat with ordinates about the level of $\beta$ for the lower values of the abscissa $p$ and, if there are jumps as in the top two panels, the functions are near $p$ for larger values, especially for abscissa above 2.00, where only jumps matter asymptotically. In the upper two panels the functions are curvilinear around the actual value of $\beta$, since the asymptotic limit is not reached with a finite stretch of simulated data. Interestingly, in the lower two panels the functions appear generally flat over the entire domain $p \in (0, 4]$ because the underlying process is continuous. The contrast between the top and bottom two panels of Fig. 2 reveals the impact of jumps.

The important implication of Figs. 1 and 2 is that the behavior of the entire signature function over the domain $p \in (0, 4]$ conveys information about the level of activity and the presence of jumps. We can expect to see a flat function about equal to the level of activity $\beta$ for the smaller values of abscissa, and, if there are jumps, a rising value of $p$ close to a 45 degree line for higher values of $p$. Also, if there are jumps, the function is curvilinear around the actual value of $\beta$, and the sharpness of the curvature is less at higher levels of $\beta$. In practice, then, the flat region on the left, the region of curvature, and the increase on the right, all taken together guide inference about $\beta$ and jumps. The sharpness of the region of curvature and the gaps between inter-quartile functions provides insight into the extent to which the particular data set is informative about the activity level. Of course, in the absence of jumps, one can expect a flat line around 2.00 over the entire domain $p \in (0, 4]$.

We now consider the standard model consisting a Brownian motion plus rare jumps. For the three cases shown in Fig. 3 we add to Brownian motion "rare" jumps, the cases **C-RJL, C-RJM,** and **C-RJH** of Table 1. The intensities correspond to a jump every 25, 3, and 0.5 units of time respectively, on average. Now the Brownian motion is the most active component. For the sake of reference, the bottom right panel shows the asymptotic limit of the signature function. For a time interval on which there is a jump $b(p) \xrightarrow{\mathbb{P}} \max\{2, p\}$, while if no jump $b(p) \xrightarrow{\mathbb{P}} 2.00$, a flat line at 2.00. For the low intensity jump case in the top left panel of Fig. 3, the three QASF functions are nearly flat because the jumps are so rare. For the medium intensity case in the top right panel of Fig. 3, we see that for $\alpha = 0.75$ the function $b_\alpha(p)$ increases after $p = 2$, but for the other two quantiles it stays flat, again due to the relative infrequency of the jumps. In the high intensity case in the bottom left panel of Fig. 3, the three QASFs increase after $p = 2$, reflecting the high intensity of the jumps. Overall, the figure suggests that unless the jumps are of moderate to high intensity, the QASFs can be expected to be flat at a level of around 2.00 for $p < 2$ and to increase at best gently for $p \geq 2$, reflecting the very rare nature of the jumps.

Another gauge of the sampling variability of the activity signature function is seen in Fig. 4, which shows 1000 replicates of the median signature function for the first four cases of Table 1. The up–down spread in the simulated functions is not large, although the bias, i.e., the location of the bend in the curve relative to the population value, is larger. We return to this in Section 5 below, which summarizes an extensive Monte Carlo study of point estimates based on the signature function.

## 5. Estimating the activity level

### 5.1. Point estimator based on the activity signature function

In the above, the approach is graphical but based on rigorous theory; there are many reasons, however, why one would seek a point estimate to succinctly summarize empirical findings. Towards this end, note that there is information about the activity level embedded in the entire signature function, especially for ordinates (heights) of $b(p)$ evaluated at the lower values of $p$. This observation suggests generating a consistent estimator of activity $\beta$ by averaging via integration the signature function over a suitable domain. From the asymptotic theory, this domain would be some interval in $(0, \beta]$, but of course $\beta$ is the unknown parameter to be estimated. Thus, we have to use a domain of integration with an upper limit given by some initial consistent estimator of $\beta$, which is available by evaluating signature function at a small positive value $\tau$, yielding $b(\tau)$. The proposed estimator of $\beta$ is

$$\hat{\beta} = \frac{1}{b(\tau) - \tau} \int_{p=\tau}^{p=b(\tau)} b(p)\, \mathrm{d}p, \tag{28}$$

where simple Riemann sums over a fine grid can be used to compute the integral in (28).[6] The right-side of (28) is clearly a statistic – a function only of the data – along with the user-selected tuning parameter $\tau$, and such tuning parameters are ubiquitous in non-parametric estimation. We suggest using $\tau = 0.10$.[7]

There are several comments regarding our proposed estimator. First, provided $\beta > \tau$, the estimator $\hat{\beta}$ is consistent and this is a simple consequence of the uniform convergence in Theorem 1. When $\beta < \tau$, the estimator will be asymptotically upward biased with a bias of at most $\tau$. From a theoretical point of view the researcher has often prior information that restricts $\beta$ from below. This will be the case for example when the discretely-observed process is modeling a process with high temporal dependence. In this case, a drift term will be typically needed to capture the mean reversion, and its presence will guarantee that $\beta \geq 1$. Similarly, the presence of risk premia in traded assets would guarantee a drift

[6] Computing $\hat{\beta}$ for each of large number of data intervals conveys bootstrap-type information on sampling fluctuations.

[7] As later illustrated in the Monte Carlo, for the frequencies that we use here, $\beta$ is estimated with a precision level of about 0.10.

Please cite this article in press as: Todorov, V., Tauchen, G., Activity signature functions for high-frequency data analysis. Journal of Econometrics (2009), doi:10.1016/j.jeconom.2009.06.009

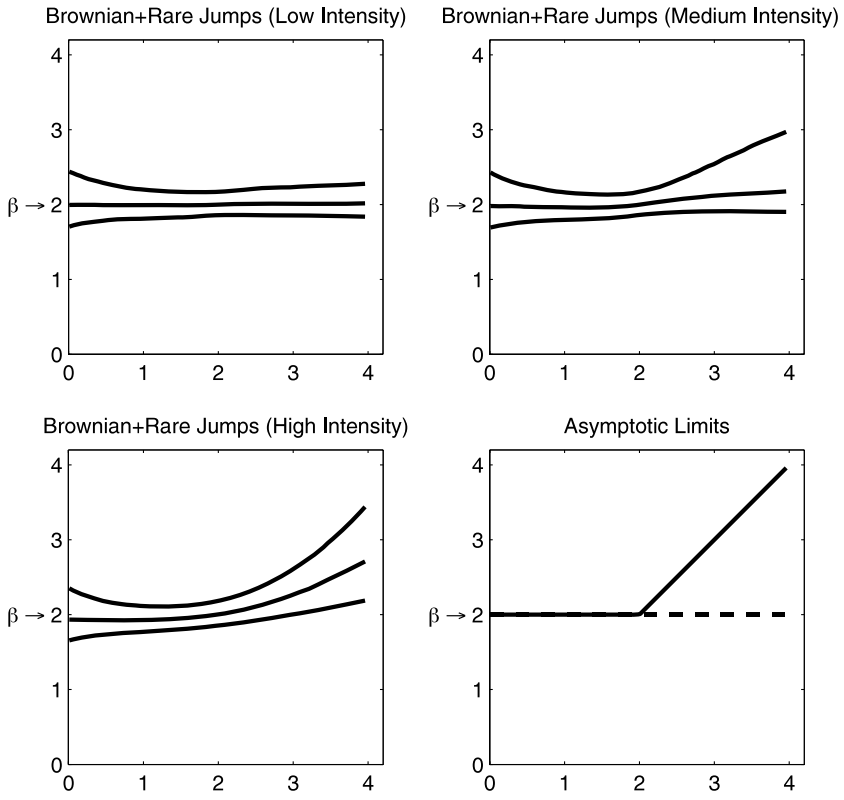


**Fig. 3.** Quantile activity signature plots for Brownian Motion plus rare jumps. The rare jumps are a compound Poisson of three different intensity levels and parameters given by **Cases C-RJL** (top left panel), **C-RJM** (top right panel), and **C-RJH** (lower left panel), where the parameter settings are given in Table 1; these panels show the QASFs $b_\alpha(p)$ defined in (25) as a function of $p$ for the quantiles $\alpha = 0.25$, $\alpha = 0.50$, and $\alpha = 0.75$. The lower right panel, shown for the purpose of reference, is the asymptotic limit as the sampling interval goes to zero.

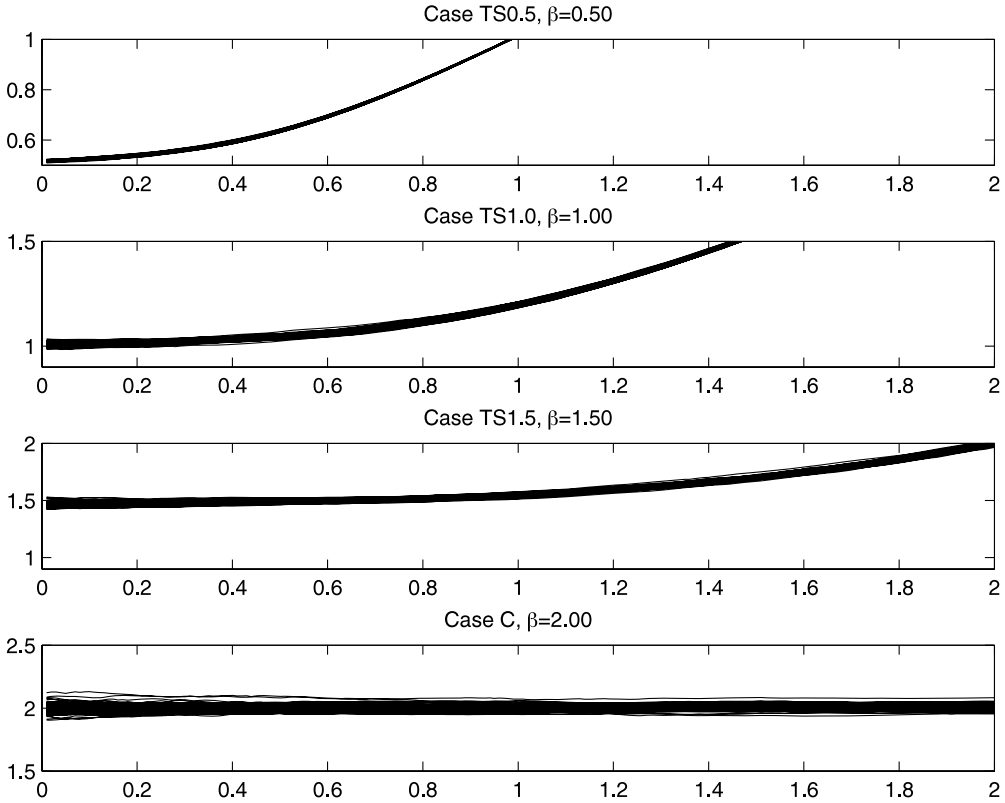


**Fig. 4.** Over-plots of 1000 Monte Carlo Replicates of the Median Activity Signature Function for Cases **TS0.5**, **TS1.0**, **TS1.5**, and **C** of Table 1 with 5 min sampling and 1500 periods (days) of data.

Please cite this article in press as: Todorov, V., Tauchen, G., Activity signature functions for high-frequency data analysis. Journal of Econometrics (2009), doi:10.1016/j.jeconom.2009.06.009

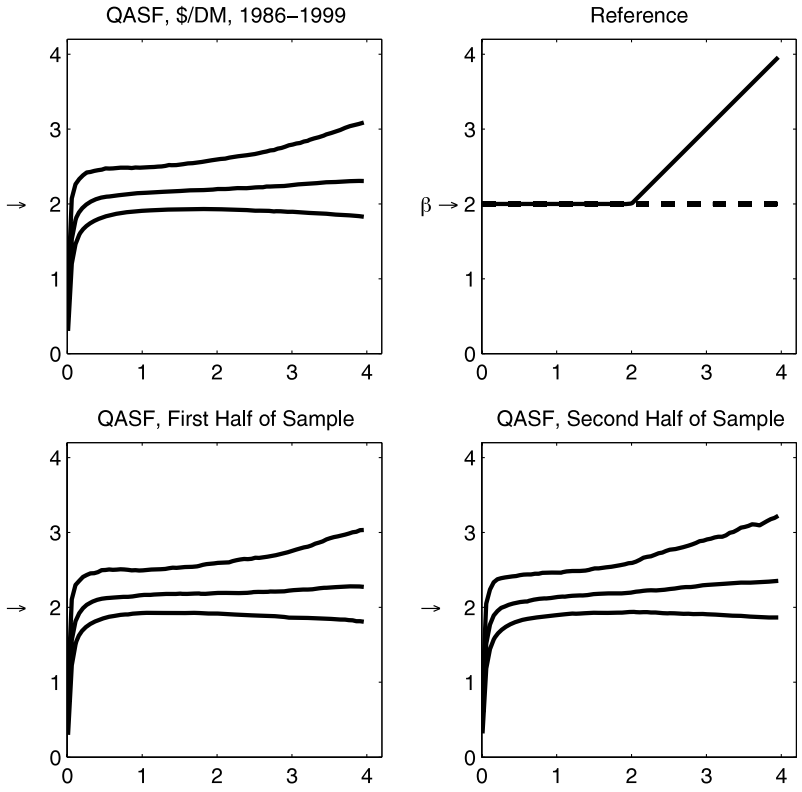


**Fig. 5.** Quantile Activity Signature Functions for the $ /DM exchange rate, 1986–1999. The top left panel shows the QASF $b_\alpha(p)$ defined in (25) as a function of $p$ for the quantiles $\alpha = 0.25$, $\alpha = 0.50$, and $\alpha = 0.75$, using the full sample, while the top right shows a theoretical reference for a continuous plus jumps process. The bottom two panels show the QASFs using evenly split (first half, second half) sample subperiods to compute the QASFs. The → in the top left and bottom two panels points to 2.0; the value of $\beta$ itself is an unknown parameter.

term and thus again $\beta \geq 1$. Therefore, from a theoretical point of view, having a prior information that restricts $\beta > \tau$ for some small value of $\tau$ is not uncommon. Indeed, in practice the data sets that we have worked with have activity well above 1.[8]

Second, from a practical point of view starting the integration in (28) from some $\tau$ away from zero is very important if the data are grainy, say reported to four or five digits. The effect of rounding causes a sharp drop of the activity signature function near the origin, as seen from Fig. 5 for our empirical application, and this limits the usefulness of the function for those very small values of the power. An extremely simple technique to eliminate the drop is to jitter the data by adding a very slight amount of random noise small enough to be of no practical significance but large enough to knock the data off the grid, and the drop-off problem goes away.

Finally, in designing $\hat{\beta}$ we used the activity signature function for all powers that carry information about the object of estimation $\beta$. It is clear though that the different powers will differ in their efficiency in estimating $\beta$—for example powers that are relatively close to the true value of $\beta$ will be relatively noisier estimates of the activity. Thus, a refinement of the current estimator $\hat{\beta}$ would be to weigh appropriately the different powers used in the estimation. We are currently developing such an estimator.

### 5.2. Monte Carlo assessment

We next conduct an extensive Monte Carlo analysis of $\hat{\beta}$ in (28) and the statistical test for $\beta = 2$ given in Theorem 2. The study covers a much wider range of scenarios than Table 1 along with a broad range of sampling frequencies and spans of the data set. The voluminous Monte Carlo output is relegated to the supplementary web-available appendix, and here we report a subset of the Monte Carlo analysis that is quite representative of the full analysis.

Table 2 shows the Monte Carlo evidence on the accuracy of $\hat{\beta}$ in (28) as an estimator of $\beta$ for the first four scenarios of Table 1 in a financial econometrics setting. The spans of the data are either one year ($N = 250$) or twelve years ($N = 3000$), and the sampling frequencies are 1 min, 5 min, or 10 min over a 24 h day. As seen from the table, the key determinant of accuracy is the sampling frequency, since it controls the bias.[9] At 1 min sampling, the mean absolute deviations and error rates indicate that the estimator $\hat{\beta}$ is very accurate, but this needs to be qualified since microstructure noise is ignored. At 5 min sampling, which is generally robust to microstructure noise, $\hat{\beta}$ always estimates $\beta$ to within 0.10. At 10 min sampling, however, the estimator can be well off the mark relative to the 5 min sampling case, especially with a short span of one year's worth of data. Interestingly, the contrast between the rightmost two columns of Table 2 suggests the bias of the estimator is between 0.05 and 0.10. As the span ($N$) increases with the sampling frequency held fixed, the estimator concentrates around a biased value, but if that bias is less than 0.10, then the error rate (for precision of up to 0.10) will be smaller with a longer span, which is evident in the table. On the other hand, if the bias is less than 0.10 but exceeds 0.05, then a longer span can be expected to provide no help in reducing the error rate (for precision of up to 0.05), which is also evident in the table. Overall, the estimator

[8] An alternative simple strategy that will turn $\hat{\beta}$ consistent even when $\beta < \tau$ is the following. Starting from $\tau$, we move it to the left till the slope $b'(p)$ is "sufficiently" different from 1 which can be determined by an additional tuning parameter. Then our estimator is (28) for this (potentially) shifted to the left, and further trimmed, level of $\tau$.

[9] This of course is not surprising as our asymptotic theory is of fill-in type, while keeping the time-span fixed.

Please cite this article in press as: Todorov, V., Tauchen, G., Activity signature functions for high-frequency data analysis. Journal of Econometrics (2009), doi:10.1016/j.jeconom.2009.06.009

**Table 2**
Monte Carlo Results for $\hat{\beta}$.

| $N$ | $M$ | $\beta_0$ | MED | IQR | MAD | Error rates | | |
|---|---|---|---|---|---|---|---|---|
| | | | | | | $\eta = 0.15$ | $\eta = 0.10$ | $\eta = 0.05$ |
| **Case TS0.5** | | | | | | | | |
| 250 | 1440 | 0.50 | 0.54 | 0.002 | 0.038 | 0.00 | 0.00 | 0.00 |
| 250 | 288 | 0.50 | 0.57 | 0.005 | 0.071 | 0.00 | 0.00 | 1.00 |
| 250 | 144 | 0.50 | 0.60 | 0.006 | 0.095 | 0.00 | 0.17 | 1.00 |
| 3000 | 1440 | 0.50 | 0.54 | 0.001 | 0.038 | 0.00 | 0.00 | 0.00 |
| 3000 | 288 | 0.50 | 0.57 | 0.001 | 0.071 | 0.00 | 0.00 | 1.00 |
| 3000 | 144 | 0.50 | 0.60 | 0.001 | 0.095 | 0.00 | 0.00 | 1.00 |
| **Case TS1.0** | | | | | | | | |
| 250 | 1440 | 1.00 | 1.04 | 0.005 | 0.040 | 0.00 | 0.00 | 0.00 |
| 250 | 288 | 1.00 | 1.07 | 0.012 | 0.071 | 0.00 | 0.00 | 0.99 |
| 250 | 144 | 1.00 | 1.09 | 0.018 | 0.092 | 0.00 | 0.27 | 1.00 |
| 3000 | 1440 | 1.00 | 1.04 | 0.001 | 0.040 | 0.00 | 0.00 | 0.00 |
| 3000 | 288 | 1.00 | 1.07 | 0.003 | 0.071 | 0.00 | 0.00 | 1.00 |
| 3000 | 144 | 1.00 | 1.09 | 0.004 | 0.092 | 0.00 | 0.00 | 1.00 |
| **Case TS1.5** | | | | | | | | |
| 250 | 1440 | 1.50 | 1.49 | 0.009 | 0.009 | 0.00 | 0.00 | 0.00 |
| 250 | 288 | 1.50 | 1.53 | 0.021 | 0.031 | 0.00 | 0.00 | 0.10 |
| 250 | 144 | 1.50 | 1.56 | 0.034 | 0.057 | 0.00 | 0.02 | 0.60 |
| 3000 | 1440 | 1.50 | 1.49 | 0.002 | 0.009 | 0.00 | 0.00 | 0.00 |
| 3000 | 288 | 1.50 | 1.53 | 0.005 | 0.031 | 0.00 | 0.00 | 0.00 |
| 3000 | 144 | 1.50 | 1.55 | 0.011 | 0.054 | 0.00 | 0.00 | 0.71 |
| **Case C** | | | | | | | | |
| 250 | 1440 | 2.00 | 2.00 | 0.014 | 0.007 | 0.00 | 0.00 | 0.00 |
| 250 | 288 | 2.00 | 2.00 | 0.032 | 0.016 | 0.00 | 0.00 | 0.03 |
| 250 | 144 | 2.00 | 1.99 | 0.045 | 0.025 | 0.00 | 0.01 | 0.16 |
| 3000 | 1440 | 2.00 | 2.00 | 0.004 | 0.002 | 0.00 | 0.00 | 0.00 |
| 3000 | 288 | 2.00 | 2.00 | 0.009 | 0.006 | 0.00 | 0.00 | 0.00 |
| 3000 | 144 | 2.00 | 1.99 | 0.012 | 0.013 | 0.00 | 0.00 | 0.00 |

The estimator $\hat{\beta}$ is defined in (28) of Section 5. The parameter settings for the cases above are given in Table 1, and the number of replicates is 1000. $N$ is the number of units of time in a simulation; $M$ is the number of high-frequency observations per unit of time. $\beta_0$ is the population value. MED is the median, IQR the inter-quartile range, MAD the median absolute deviation, each computed over the 1000 replications. Error rates are the proportion of replicates where $|\hat{\beta} - \beta_0| > \eta$.

appears to be reasonably accurate using the large dense data sets typical in financial econometrics.

Next, Table 3 shows the rejection probabilities under six scenarios of the test based on Theorem 2 above of the null hypothesis that a continuous martingale component is present in the process, i.e. that its activity index equals 2. In the first three blocks, the null hypothesis is true and the test is generally slightly undersized and thereby conservative. An exception is the case **C-RJH**, which is Brownian motion plus rare jumps at high intensity, where the test is oversized, but only seriously so at the 10 percent significance level, which is not commonly used in practice. In the lower three blocks the null hypothesis is false, and the process is pure-jump. The test appears reasonably powerful, with the exception of very course sampling, 10 min, for the case **TS1.5** where the rejection rates are rather low. Taken together, the Monte Carlo results in Table 3 indicate that the test for presence of a continuous component, which we apply later in our empirical application, is well behaved without serious size distortions and reasonably high power.

**Table 3**
Size and power of test for presence of continuous martingale.

| Sampling frequency | Proportion of rejections of the null hypothesis at conventional significance levels | | |
|---|---|---|---|
| | 0.01 | 0.05 | 0.10 |
| **Case C** ($H_0$ **true**) | | | |
| $M = 1440$ (1 min) | 0.0052 | 0.0407 | 0.0913 |
| $M = 288$ (5 min) | 0.0012 | 0.0248 | 0.0743 |
| $M = 144$ (10 min) | 0.0002 | 0.0185 | 0.0623 |
| **Case C-RJH** ($H_0$ **true**) | | | |
| $M = 1440$ (1 min) | 0.0132 | 0.0820 | 0.1596 |
| $M = 288$ (5 min) | 0.0029 | 0.0497 | 0.1227 |
| $M = 144$ (10 min) | 0.0005 | 0.0274 | 0.0919 |
| **Case OU-C** ($H_0$ **true**) | | | |
| $M = 1440$ (1 min) | 0.0050 | 0.0406 | 0.0912 |
| $M = 288$ (5 min) | 0.0020 | 0.0258 | 0.0735 |
| $M = 144$ (10 min) | 0.0006 | 0.0193 | 0.0640 |
| **Case TS0.5** ($H_0$ **false**) | | | |
| $M = 1440$ (1 min) | 0.9998 | 0.9998 | 0.9998 |
| $M = 288$ (5 min) | 0.9942 | 0.9981 | 0.9990 |
| $M = 144$ (10 min) | 0.9735 | 0.9914 | 0.9942 |
| **Case TS1.0** ($H_0$ **false**) | | | |
| $M = 1440$ (1 min) | 0.9999 | 0.9999 | 0.9999 |
| $M = 288$ (5 min) | 0.9823 | 0.9964 | 0.9977 |
| $M = 144$ (10 min) | 0.7121 | 0.9321 | 0.9700 |
| **Case TS1.5** ($H_0$ **false**) | | | |
| $M = 1440$ (1 min) | 0.9967 | 0.9995 | 0.9998 |
| $M = 288$ (5 min) | 0.2634 | 0.6207 | 0.7736 |
| $M = 144$ (10 min) | 0.0428 | 0.2942 | 0.4920 |

The test is one-sided and is based on asymptotic distribution under the null using $p = 0.9$ in Theorem 2 in Section 3. The parameter settings for the cases above are given in Table 1, and the number of replicates is 10,000 units of time.

## 6. Empirical applications

### 6.1. Exchange rates

In the first application we use high-frequency data on the log of the spot $/*DM* exchange rate for the period 1986:12–1999:06, 3045 trading days. Each trading day is 24 h and sampling every five-minutes gives 288 log-returns (increments). For each day we calculate the activity signature function $b_s(p)$ where $s = 1, \ldots, 3045$ and then evaluate the quantile activity function QASF as $b_\alpha(p)$ defined in (25). The top left panel of Fig. 5 shows the QASFs for $\alpha = 0.25, 0.50, 0.75$ computed on the observed data; the sharp drop near the origin is due to the effects of rounding, since

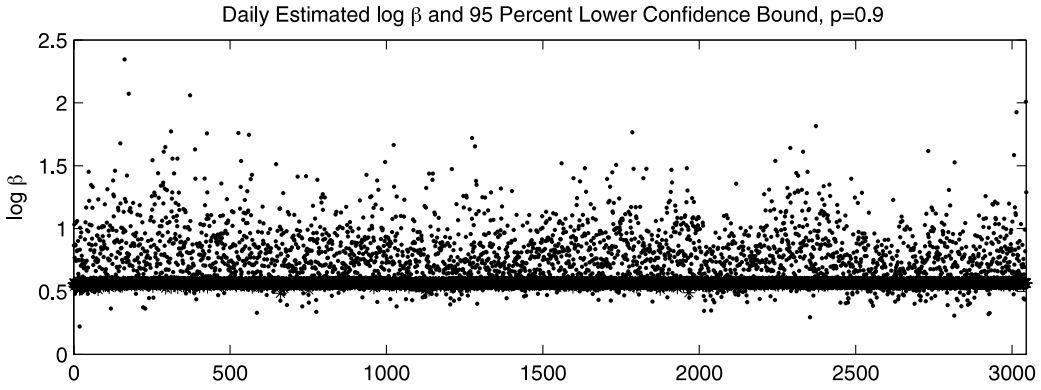


**Fig. 6.** Daily point estimates $\log(b_s(p))$ marked by • for $p = 0.90$ and 95% lower confidence bounds. A point estimate below its corresponding confidence bound would be considered statistically significantly different from the null value of $\log(2) = 0.69$.

the exchange rate is quoted to five digits, whereas simulations are computed to machine precision.[10] The top right panel of Fig. 5 shows an asymptotic reference signature function, while the bottom two panels show the QASFs computed on the first and second halfs of the sample, respectively.

From Fig. 5, the evidence suggests that the exchange rate process contains a continuous martingale component, which has activity level of 2.00. The bottom two panels of Fig. 5 indicate that the empirical finding is robust to using either the first or second half of the sample. Additional support comes also from our point estimator of the activity given in (28)—its median value over the entire sample equals 2.12.

To test formally for the presence of a continuous martingale in the spot $/*DM* exchange rate, we implement the test given in Theorem 2. Fig. 6 is a plot of the day-by-day values of $\log(b_s(p))$ for $p = 0.90$ along with a 95% lower confidence bound based on the null of a continuous martingale being present. Only 1.5 percent of the point estimates lie below their corresponding 95 percent confidence bound and thereby are statistically significantly different from the null value of log(2). Consistent with contemporaneous evidence obtained using a different test (Cont and Mancini, 2007), Fig. 6 indicates very little statistical evidence against a model with a continuous component.

Overall, the non-parametric analysis of the activity level of the exchange rate data shows that pure-jump models for asset prices as proposed in Carr et al. (2002, 2003) and elsewhere are probably not good descriptions of the high-frequency exchange rate data. Instead, an appropriate model for the exchange rate (at least for the time period used in this study) is of the form of a continuous semimartingale plus jumps as in (3), which has many small diffusive moves along with jumps.

The quantile activity signature plots also provide evidence on the intensity of the jumps present in the discretely-observed process in addition to the continuous martingale. As seen in Fig. 5, only the 75th quantile of QASFs computed on observed data increases for $p > 2$. By way of comparison, the cases **C-RJL**, **C-RJM**, and **C-RJH** in Table 1 are realistically calibrated so that the jump part accounts for 20% of the total variance of the process. Only with intensities of one jump every three days, or two jumps every day, can activity signature plots computed from simulations as in Fig. 3 be consistent with the signature plots computed from observed data in Fig. 5. However, all empirical evidence we know of reports a very low intensity level of between five to twenty jumps per year, if the data are presumed to follow a (time-changed) Brownian motion plus a compound Poisson process. As seen in the top left panel of Fig. 3, at such low intensity levels the process would generate quantile activity signature plots very much unlike those of Fig. 5 computed on the data. Our evidence thus suggest that a model of a time-changed Brownian motion and rare jumps is misspecified.[11]

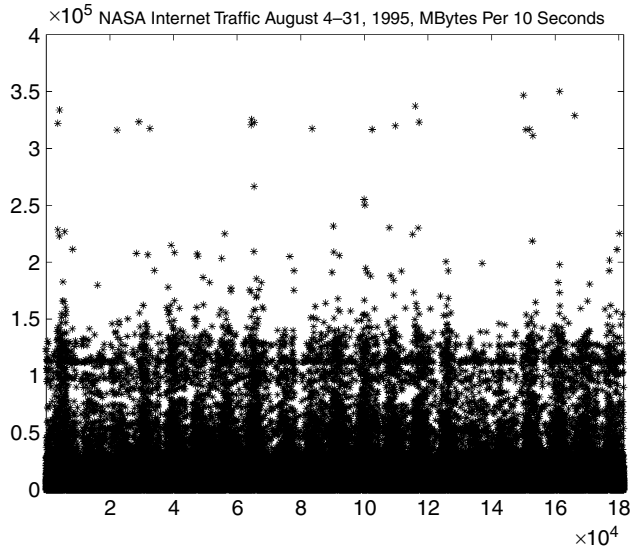


**Fig. 7.** Raw data: The number of megabytes downloaded from NASA servers over ten second intervals for the period August 4–31, 1995.

### 6.2. Internet traffic

Under a so-called "slow-growth" condition, processes related to the classical $\alpha$-stable process are important for modeling internet traffic (Mikosch et al., 2002, Theorem 1, p. 33). Parameter estimates of the index obtained on the maintained hypothesis of a classical $\alpha$-stable model can range from 0.70 to 1.67 depending upon the data set (Xiaohu et al., 2004, Table 1, p. 450).

Fig. 7 shows the number of megabytes downloaded from NASA servers over ten second intervals for the period August 4–31, 1995. Raw data are from the file `NASA_access_log_Aug95` in the public domain. NASA servers experience heavy demand,

[10] Rounding simulated data has exactly same effect near the origin on activity signature plots as that seen in Fig. 5.

[11] Note that our estimation is not designed to infer the Blumenthal–Getoor of the jumps *when* a continuous-martingale is known to be present. To assess formally the activity of the jumps in the exchange rate series, *in the presence of a continuous martingale*, we need much higher frequencies than the ones used here and the formal tools developed in Ait-Sahalia and Jacod (2009b,c). Such analysis is beyond the scope of the current paper.

Please cite this article in press as: Todorov, V., Tauchen, G., Activity signature functions for high-frequency data analysis. Journal of Econometrics (2009), doi:10.1016/j.jeconom.2009.06.009

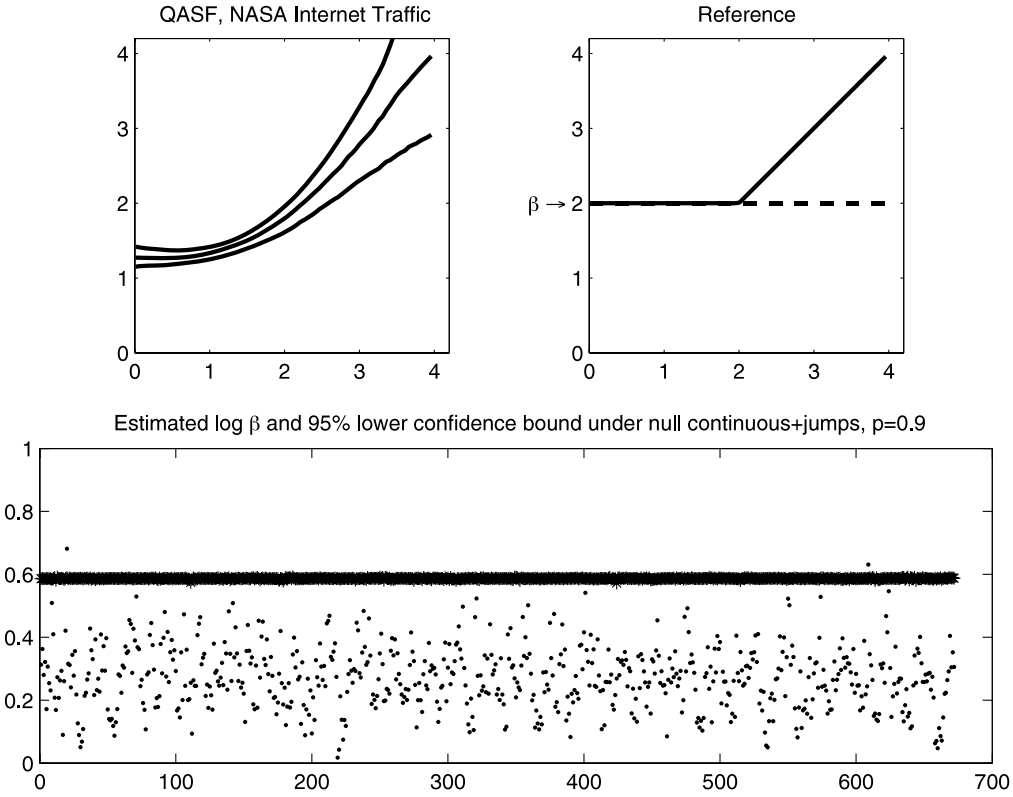


**Fig. 8.** Top left: Quantile Activity Signature Functions computed from the NASA download data shown in Fig. 7. Top right: Asymptotic function *if* the process were a continuous plus jumps. Bottom panel: Formal test for the presence of a continuous component, point estimates $\log(b_s(p))$ marked by • for $p = 0.90$ and 95% lower confidence bounds; a point estimate below its corresponding confidence bound would be considered statistically significantly different from the null value of $\log(2) = 0.69$.

sometimes between two to three large requests per second at peak periods. The time stamps show the hour, minute, and second of the data-request, with multiple records per second in the data file. We aggregated to ten second intervals as is common practice. We exclude August 1–3, 1995, because of various anomalies and we use segments of length one hour, reflecting the slowly-varying overall level of traffic revealed in initial analysis. There are 360 observations per segment, and 24 (h) $\times$ 28 (days) $=$ 672 segments. The data are centered using the full sample mean.

The top left panel of Fig. 8 shows the quantile signature functions for these NASA internet traffic data; the top right panel shows, for sake of reference, what would be the asymptotic limit of the signature function *if* it were a continuous process plus jumps. The contrast between the top left and right panels suggests that the process generating the internet traffic data lacks a continuous component. This is further confirmed from the bottom panel of Fig. 8, which shows the results from the test for presence of continuous martingale component. For almost all periods in the sample (constituting 97% of the sample) this null hypothesis can be easily rejected. Finally, the median estimate of the activity level based on (28) equals 1.30. This non-parametric evidence suggests that the activity index of this NASA internet traffic series is in the range of 1.10–1.50, which is slightly more active than the Cauchy process and of infinite variation.

The analysis here provides a useful starting point for building a fully parametric pure-jump model of the time series. The close resemblance between the signature plots from the NASA series in Fig. 8 and those obtained from the tempered stable process in Fig. 2 is interesting, and suggests the tempered stable appears an excellent candidate for an initial parametric model that could be refined under the guidance of various specification tests. This task is beyond the scope of this paper.

## 7. Conclusion

Our non-parametric strategy for inference about the activity level of a discretely-observed semimartingale is intuitive and graphical, and it gives a non-parametric test for the presence of continuous martingale. In a finance application to the $ /DM exchange rate, the findings indicate that pure-jump models are empirically less plausible than the classical model of Brownian Motion plus jumps. In the second application, the findings suggest that NASA internet traffic follows a pure-jump model without a continuous component and of infinite variation. The two disparate findings indicate the range of applicability of the methods.

## References

Ait-Sahalia, Y., Jacod, J., 2009a. Testing for jumps in a discretely observed process. Annals of Statistics 37, 184–222.
Ait-Sahalia, Y., Jacod, J., 2009b. Estimating the degree of activity of jumps in high frequency financial data. Annals of Statistics 37, 2202–2244.
Ait-Sahalia, Y., Jacod, J., 2009c. Testing whether jumps have finite or infinite activity. Working paper. Princeton University and Université de Paris-6.
Ballotta, L., 2005. A Lévy process-based framework for the fair valuation of participating life insurance contracts. Insurance Mathematics and Economics 37, 173–196.
Barndorff-Nielsen, O., Graversen, S., Jacod, J., Podolskij, M., Shephard, N., 2005. A central limit theorem for realised power and bipower variations of continuous semimartingales. In: Kabanov, Y., Lipster, R. (Eds.), From Stochastic Analysis to Mathematical Finance, Festschrift for Albert Shiryaev. Springer.
Barndorff-Nielsen, O., Shephard, N., 2001. Non-Gaussian Ornstein–Uhlenbeck-based models and some of their uses in financial economics. Journal of the Royal Statistical Society Series B, 63, 167–241.
Barndorff-Nielsen, O., Shephard, N., Winkel, M., 2006. Limit theorems for multipower variation in the presence of jumps in financial econometrics. Stochastic Processes and their Applications 116, 796–806.
Barndorff-Nielsen, O.E., 1997. Normal inverse gaussian distributions and stochastic volatility modeling. Scandinavian Journal of Statistics 24, 1–13.
Barndorff-Nielsen, O.E., 1998. Processes of normal inverse gaussian type. Finance and Stochastics 41–419.

Please cite this article in press as: Todorov, V., Tauchen, G., Activity signature functions for high-frequency data analysis. Journal of Econometrics (2009), doi:10.1016/j.jeconom.2009.06.009

Blumenthal, R., Getoor, R., 1961. Sample functions of stochastic processes with independent increments. Journal of Mathematics and Mechanics 10, 493–516.
Broadie, M., Detemple, J.B., 2004. Option pricing: Valuation models and applications. Management Science 50, 1145–1177.
Brockwell, P., 2001. In: Shanbhag, D., Rao, C. (Eds.), Continuous-Time ARMA Processes. In: Handbook of Statistics, vol. 19. North-Holland.
Carr, P., Geman, H., Madan, D., Yor, M., 2002. The fine structure of asset returns: An empirical investigation. Journal of Business 75, 305–332.
Carr, P., Geman, H., Madan, D., Yor, M., 2003. Stochastic Volatility for Lévy processes. Mathematical Finance 13, 345–382.
Carr, P., Wu, L., 2007. Stochastic skew for currency options. Journal of Financial Economics 86, 213–247.
Cont, R., Mancini, C., 2007. Nonparametric tests for analyzing the fine structure of price fluctuations. Working paper. November 2007, Columbia University Center for Financial Engineering.
Cont, R., Tankov, P., 2003. Financial Modelling with Jump Processes. Chapman and Hall, Boca Raton, Florida, USA.
Daal, E., Madan, D.B., 2005. An empirical examination of the variance-gamma model for foreign currency options. Journal Of Business 78, 2121–2152.
Drosen, J.W., 1986. Pure jump shock models in reliability. Advances in Applied Probability 18, 423–440.
Huang, J.Z., Wu, L., 2004. Specification analysis of option pricing models based on time-changed Lévy processes. Journal of Finance 59, 1405–1439.
Huang, S., Hung, M.W., 2005. Pricing foreign equity options under Lévy processes. Journal Of Futures Markets 25, 917–944.
Ivanov, R.V., 2007. Specification analysis of option pricing models based on time-changed Lévy processes. Journal of Applied Probability 44, 409–419.
Jacod, J., 2004. The euler scheme for Levy driven stochastic differential equations: Limit theorems. Annals of Probability 32, 1830–1872.
Jacod, J., 2008. Asymptotic properties of power variations and associated functionals of semimartingales. Stochastic Processes and their Applications 118, 517–559.
Jacod, J., Shiryaev, A.N., 2003. Limit Theorems For Stochastic Processes, 2nd ed.. Springer-Verlag, Berlin.
Klüppelberg, C., Lindner, A., Maller, R., 2004. A continuous time GARCH process driven by a Lévy process: Stationarity and second order behavior. Journal of Applied Probability 41, 601–622.
Lepingle, D., 1976. La variation d'ordre p des semi-martingales. Zeitschrift fur Wahrscheinlichkeitstheorie und Verwandte Gebiete 36, 295–316.
Levendorskii, S.Z., 2004. Early exercise boundary and option prices in Lévy driven models. Quantitative Finance 4, 525–547.
Madan, D., Carr, P., Chang, E., 1998. The of variance gamma process and option pricing. European Finance Review 2, 79–105.
Madan, D.B., 2005. Equilibrium asset pricing: With non-Gaussian factors and exponential utilities. Insurance Mathematics and Economics 37, 173–196.
Mikosch, T., Resnick, S., Rootzèn, H., Stegeman, A., 2002. Approximated by stable Lévy motion or fractional Brownian motion? The Annals of Applied Probability 12, 23–68.
Revuz, D., Yor, M., 1999. Continuous Martingales and Brownian Motion. Springer.
Rosiński, J., 2007. Tempering stable processes. Stochastic Processes and their Applications 117, 677–707.
Rydberg, T.H., 1997. The normal inverse gaussian and Lévy process: Simulation and approximation. Communications Statistics: Stochastic Models 13, 887–910.
Schoutens, W., 2006. Exotic options under Lévy models: An overview. Journal of Computational and Applied Mathematics 189, 526–538.
Todorov, V., Tauchen, G., 2006. Simulation methods for Lévy -driven CARMA stochastic volatility models. Journal of Business and Economic Statistics 24, 455–469.
Woerner, J., 2003. Purely discontinuous Lévy processes and power variation: Inference for the integrated volatility and the scale parameter. Working paper. University of Oxford.
Woerner, J., 2006. Analyzing the fine structure of continuous-time stochastic processes. Working paper.
Xiaohu, G., Guangxi, Z., Yaoting, Z., 2004. On the testing for alpha-stable distributions of network traffic. Computer Communications 5, 447–457.
Zhang, L., Mykland, P., Ait-Sahalia, 2005. A Tale of two time scales: Determining integrated volatility with noisy high-frequency data. Journal of the American Statistical Association 100, 1394–1411.

Please cite this article in press as: Todorov, V., Tauchen, G., Activity signature functions for high-frequency data analysis. Journal of Econometrics (2009), doi:10.1016/j.jeconom.2009.06.009