Computational Statistics & Data Analysis 37 (2001) 419–431

COMPUTATIONAL STATISTICS & DATA ANALYSIS

www.elsevier.com/locate/csda

Asymptotic distribution of time-series intermittency estimates: applications to economic and clinical data

David R. Bickel[a,*], Dejian Lai[b]

[a] *Office of Biostatistics and Bioinformatics, Medical College of Georgia, 1120 Fifteenth St., AE-3037, Augusta, GA 30912-4900, USA*

[b] *Department of Biometry, The University of Texas-Houston Health Science Center, School of Public Health, Houston, TX 77030, USA*

Received 1 August 2000; received in revised form 1 March 2001; accepted 1 March 2001

Abstract

The intermittency of a time series can be defined as its normalized difference in scaling parameters. We establish the central limit theorem for the estimates of intermittency under the null hypothesis of a random walk. Simulations of random walks indicate that the distribution of intermittency estimates is slightly negatively skewed and positively biased, but that the skewness and bias approach zero as the length n of the random walks increases. We provide a formula by which the sample variance of the intermittency estimates of these simulations can be used to approximate the standard error of the intermittency for any large n. These results can be used to test whether the intermittency estimate of an observed long time series is significantly greater than zero, the intermittency of a random walk. This test reveals that the intermittency estimates of the S&P 500 index and of the heart rate of a human adult are significantly positive. The hypothesis testing proposed in this paper can also be applied to other observed time series to determine whether their intermittency estimates are sufficiently high for the series to be considered intermittent, or whether their estimates are small enough to be consistent with a random walk.

Keywords: Time series analysis; Random walk; Multifractal scaling; Intermittency; Stock market data; Heart rate data

* Corresponding author.

E-mail addresses: dbickel@mail.mcg.edu; bickel@mailaps.org (D.R. Bickel).

0167-9473/01/$ - see front matter
PII: S 0167-9473(01)00021-4

1. Introduction

Many observed time series are *intermittent* in the sense that observations differ dramatically from previous observations from time to time. The *intermittency* of a stationary time series $\{\Delta x_i\}_{i=1}^{n-1}$ is informally defined as the tendency of its absolute values, $\{|\Delta x_i|\}_{i=1}^{n-1}$, to be much greater than the most probable absolute values. Thus, intermittency increases with the skewness of $|\Delta x_i|$ for single-modal distributions. The intermittency of a nonstationary time series $\{x_i\}_{i=1}^{n}$ with stationary increments can be quantified by the intermittency of the first difference, $\{\Delta x_i \equiv x_{i+1} - x_i\}_{i=1}^{n-1}$ (Davis et al., 1994). Although intermittency can easily be confused with corruption by outliers, intermittency is an essential property of the system studied, whereas corrupt data is contaminated by factors external to the system of interest. For example, unpredictable, fast increases in the heart rate resulting from physiological activity exhibit intermittency, but isolated jumps in heart rate data due to equipment failure are considered to be outliers. Thus, intermittency is estimated to characterize data, but outliers can impede the estimation of parameters of interest. Contamination by outliers can falsely elevate estimates of intermittency, requiring robust methods of intermittency estimation (Bickel, 2001b). Since intermittency generally does not depend on the mean or variance of a time series, it can be used as another summary statistic to describe data and compare data to models (Davis et al., 1994; Bickel, 1999b). Several models of intermittency are available, such as multiplicative-rate point processes, fractal renewal point processes (Bickel, 1999a), and multiplicative cascades (Davis et al., 1994), including the multifractal α-model (Schertzer and Lovejoy, 1987). There are multiple formal definitions of intermittency, some of which are based on a particular multifractal model. For example, Holley and Waymire (1992) defined an intermittency parameter to characterize a multifractal cascade. Herein we instead study the asymptotic distribution of an estimator using a more general definition of intermittency, based on the theory of multifractal scaling (Bickel, 1999b). Rényi (1959) introduced the concept of multifractal scaling by generalizing the concept of dimension in the context of probability theory. Multifractal statistical analysis has since successfully described various data sets (Falconer, 1994), including those of turbulence and the spatial distribution of rainfall (Holley and Waymire, 1992).

We assume that $x(t)$ is a *scaling process* in the sense that it satisfies the relation $s(q;\theta T) = \theta^{H(q)} s(q;T)$, the solution of which is

$$s(q;T) \propto T^{H(q)}, \tag{1}$$

where the *structure function*, $s(q;T)$, is

$$s(q;T) \equiv \{E[|x(t+T) - x(t)|^q]\}^{1/q}. \tag{2}$$

Here $\theta \in \Re^+$, $x(t) \in \Re$, $t \in \Re$, $q \in \Re^+$, and $E[\bullet]$ denotes the expectation of its argument. Note that $s^2(2;T)$ is the variogram of $x(t)$, and that $s(2;T)$ is the standard deviation of $x(t+T) - x(t)$ and $s(q;T)$ is the qth-power deviation of $x(t+T) - x(t)$ when $E[x(t+T) - x(t)] = 0$. Bickel and Lehmann (1976) define the qth-power deviation and give the requirements for a measure of dispersion, which are satisfied by $s(q;T)$. The class of scaling processes is broadened by including all $x(t)$ for which