Statistical tests for Lyapunov exponents of deterministic systems

Rodney C. Wolff, Qiwei Yao and Howell Tong

Statistical Tests for Lyapunov Exponents of Deterministic Systems

Rodney C. Wolff
School of Economics and Finance,
Queensland University of Technology,
GPO Box 2434, Brisbane QLD 4001, Australia
r.wolff@qut.edu.au

Qiwei Yao
Department of Statistics,
The London School of Economics and Political Science,
Houghton Street, London, WC2A 2AE, U.K.
q.yao@lse.ac.uk

Howell Tong
Department of Statistics,
The London School of Economics,
and
Department of Statistics and Actuarial Science,
Hong Kong University, Hong Kong
htong@hku.hk

Abstract

In order to develop statistical tests for the Lyapunov exponents of deterministic dynamical systems, we develop bootstrap tests based on empirical likelihood for percentiles and expectiles of strictly stationary processes. The percentiles and expectiles are estimated in terms of asymmetric least deviations and asymmetric least squares methods. Asymptotic distributional properties of the estimators are established.

Key words: Bootstrap, chaos, empirical likelihood, expectile, percentile.

1 Introduction

The goal of this paper is to develop tools for statistical inference for the Lyapunov exponents of deterministic chaos. The need for such statistical techniques is as follows.

Essentially, the Lyapunov exponent of a one-dimensional deterministic chaotic system measures the exponential rate of divergence of pairs of trajectories which are initially close. We discuss details in Section 3, and the reader is referred to Ruelle (1989) for an elegant introduction to analytical and geometric properties of chaotic systems, and to Tong (1990) for an elucidation in the context of non-linear time series. For one-dimensional chaotic systems, Lyapunov exponents have an important practical use: it is a necessary condition for the existence of chaos that the Lyapunov exponent be positive. However, for a large class of dynamical systems (the so-called 'Axiom A' systems), it is generally accepted as being a sufficient condition. The presence of a positive Lyapunov exponent indicates a phenomenon known as sensitive dependence upon initial conditions. (The problem may similarly be expressed for higher-dimensional dynamical systems.)

While there exists a variety of point estimators for Lyapunov exponents, little has been done to capture their second-order properties, essential for testing hypotheses on them. In particular, to test for the presence of sensitive dependence upon initial conditions, it must be that a Lyapunov exponent is positive. (Strictly speaking, in the simplest scenario, one would pose a null hypothesis that the Lyapunov exponent is zero and test against a one-sided alternative.) Bailey *et al.* (1997) derived a central limit theorem for local Lyapunov exponents, a quantity equivalent to a short-term version of the (global) Lyapunov exponent and localised in the state space. Lu and Smith (1997) approached the same problem using local regression methods. Both papers provide asymptotic second order properties of the estimators. See also Nychka *et al.* (1992). Since a collection of local Lyapunov exponents may be used to construct the global Lyapunov exponent, their methods have implications for computing confidence intervals for, and tests of, hypotheses on Lyapunov exponents. As illustrated in Wolff (1995), small-sample properties of estimators of invariants of chaotic dynamical systems may differ markedly from the asymptotics. In the present context, finite sample properties of gradient estimators, such as are required to compute sample Lyapunov exponents, are impossible to obtain analytically.

To this end, we first develop bootstrap tests, based on empirical likelihood, for percentiles and expectiles of general strictly stationary processes. We then apply the methods to test the Lyapunov exponents and other indices of one-dimensional deterministic systems.

In the same spirit as the broadly non-parametric approach which we take here, we note that Shintani and Linton (2003) obtained a test for chaos using a neural network methodology to estimate Lyapunov exponents, whereas Whang and Linton (1999) obtained distributional properties of Lyapunov exponents using non-parametric regression. Further, time domain resampling methods for chaotic time series and Lyapunov exponent estimation have been presented by Giannerini and Rosa (2001) and Golia and Sandri (2001).

The rest of the paper is organised as follows. In Section 2, we propose some bootstrap tests for percentiles and expectiles based on asymmetric least absolute deviations (ALAD) estimators and the asymmetric least squares (ALS) estimators, for a probability measure on R^1 . The bootstrap methods involve utilising constraints: see also Hall and Presnell (1999a, 1999b). Asymptotic properties of those estimators are established. We then apply the tests in Section 3 for testing hypotheses on the Lyapunov exponent of a one-dimensional deterministic chaotic system. All technical proofs are given in Section 4.

2 Bootstrap Tests for Percentiles and Expectiles

Let $\{Y_i, -\infty < i < \infty\}$ be a strictly stationary process. We assume that $E\left(Y_1^2\right) < \infty$, and Y_1 has the smooth probability density function $f(.)$.

2.1 Percentiles and expectiles

For $\alpha \in (0,1)$, the 100α-th percentile of Y_1 is a constant, say ξ_α, for which $P\left(Y_1 \leq \xi_\alpha\right) = \alpha$. Obviously, $\xi_{0.5}$ is the median of Y_1. We exclude the cases that $\alpha = 0$ or $\alpha = 1$ because the percentiles (also expectiles) are not uniquely determined and are of little practical interest there.

It is easy to see that

$$\xi_\alpha = \operatorname*{argmin}_{|b|<\infty} E\{R_\alpha(Y_1 - b)\}, \tag{2.1}$$

where

$$R_\alpha(y) = \begin{cases} \alpha|y| & y > 0, \\ (1-\alpha)|y| & y \leq 0. \end{cases}$$

In fact, for small x, we have the asymptotic approximation

$$E\{R_\alpha(Y_1 - \xi_\alpha + x)\} = E\{R_\alpha(Y_1 - \xi_\alpha)\} + \frac{1}{2} f(\xi_\alpha) x^2 + o(x^2). \tag{2.2}$$

If we define a loss function as

$$Q_\alpha(y) = \begin{cases} \alpha y^2 & (y > 0) \\ (1-\alpha) y^2 & (y \leq 0), \end{cases}$$

the 100α-th expectile of Y_1 is defined as

$$\tau_\alpha = \operatorname*{argmin}_{|b|<\infty} E\{Q_\alpha(Y_1 - b)\}. \tag{2.3}$$

Obviously, $\tau_{0.5} = E(Y_1)$. It can be shown that

$$\alpha = \frac{E\{|Y_1 - \tau_\alpha| \mathcal{I}\left(Y_1 \leq \tau_\alpha\right)\}}{E\left(|Y_1 - \tau_\alpha|\right)}, \tag{2.4}$$

which resamples the property of the percentile that

$$\alpha = P\left(Y_1 \leq \xi_\alpha\right) = \frac{E\{1 \mathcal{I}\left(Y_1 \leq \xi_\alpha\right)\}}{E(1)},$$

where $\mathcal{I}(A)$ denotes the indicator function of the event A. Hence, in a similar way that the percentile ξ_α specifies the position below which $100\alpha\%$ of the probability mass of Y_1 lies, the expectile τ_α determines the point such that $100\alpha\%$ of the mean distance between it and Y_1 comes from the mass below it. Note that function Q_α has continuous derivative. This can make theoretical exploration for expectiles slightly easier than that for percentiles. Similarly to (2.2), we have the approximation

$$\begin{aligned}&E\{Q_\alpha(Y_1-\tau_\alpha+x)\}\\ &\quad= \quad E\{Q_\alpha(Y_1-\tau_\alpha)\}+\{\alpha P(Y_1>\tau_\alpha)+(1-\alpha)P(Y_1\le\tau_\alpha)\}x^2+o\left(x^2\right).\end{aligned}$$

For more information about expectiles, we cite Neway and Powell (1987), Efron (1991), and Yao and Tong (1996).

From (2.1), an ALAD estimate for the percentile from sample $\{Y_i, 1\le i\le n\}$ can be defined as

$$\widehat{\xi}_\alpha = \underset{b}{\operatorname{argmin}} \sum_{i=1}^n R_\alpha(Y_i-b). \tag{2.5}$$

Similarly, from (2.3) an ALS estimate for the expectile is defined as

$$\widehat{\tau}_\alpha = \underset{b}{\operatorname{argmin}} \sum_{i=1}^n Q_\alpha(Y_i-b). \tag{2.6}$$

The above estimates will play a key role in constructing the bootstrap tests. Theorems 1 and 2 below show that both estimates are consistent and asymptotically normal under some mixing conditions. The stationary process $\{Y_i\}$ is said to be ρ-mixing if

$$\rho_j \equiv \sup_{i\ge1}\{\sup_{U\in\mathcal{F}_1^i, V\in\mathcal{F}_{i+j}^\infty}|\mathrm{Corr}(U,\ V)|\}\to 0,$$

where $\mathcal{F}_i^j$ is the σ-field generated by $\{Y_i,\dots,Y_j\}$ $(j\ge i)$.

Theorem 1. Suppose that the stochastic process $\{Y_i\}$ is strictly stationary and ergodic. We also assume that $E(Y_i^2)<\infty$. Then, for $\alpha\in(0,1)$, $\widehat{\xi}_\alpha \xrightarrow{P} \xi_\alpha$ and $\widehat{\tau}_\alpha \xrightarrow{P} \tau_\alpha$ as $n\to\infty$.

Theorem 2. Suppose that the stochastic process $\{Y_i\}$ is strictly stationary and ρ-mixing. Further, we assume that $\limsup_{n\to\infty}\sum_{i=1}^n(1-i/n)\rho_i<\infty$. We also assume that $E(Y_i^4)<\infty$. Then, as $n\to\infty$, for $\alpha\in(0,1)$,

(i) $\sqrt{n}(\widehat{\xi}_\alpha-\xi_\alpha)\xrightarrow{d} N(0,\ \alpha(1-\alpha)\{f(\xi_\alpha)\}^{-2})$,

(ii) $\sqrt{n}(\widehat{\tau}_\alpha-\tau_\alpha)\xrightarrow{d} N(0,\ \sigma_\alpha^2))$, where

$$\sigma_\alpha^2=\frac{\alpha^2E\{(Y_1-\tau_\alpha)^2\mathcal{I}(Y_1>\tau_\alpha)\}+(1-\alpha)^2E\{(Y_1-\tau_\alpha)^2\mathcal{I}(Y_1\le\tau_\alpha)\}}{\{\alpha P(Y_1>\tau_\alpha)+(1-\alpha)P(Y_1\le\tau_\alpha)\}^2}.$$

The assumption of ρ-mixing is for the brevity of proofs. Actually Theorem 2 still holds under the assumption of strong mixing (α-mixing). On the other hand, a linear or non-linear (stochastic) autoregressive process satisfying some mild conditions is ρ-mixing. However, any sequence generated by a deterministic equation such as $X_{t+1} = m(X_t)$ is not α-mixing, therefore is also not ρ-mixing. See Section 2.6 of Fan and Yao (2003) for further discussion on different mixing conditions and their properties.

The proofs of the above theorems are given in Section 4.

2.2 Bootstrap tests

We start with simple null hypotheses $H_{01}:\ \xi_\alpha = a$ and $H_{02}: \tau_\alpha = a$, for both $\alpha \in (0,1)$ and $a \in (-\infty, \infty)$ given. The method can be easily adjusted to test the hypotheses that $\xi_\alpha \geq a$ or $\xi_\alpha \leq a$, and so on. We construct the tests along the lines of Efron and Tibshirani (1993, Ch. 6). Note that when $\alpha = 0.5$, we test the hypotheses on the median (H_{01}) and the mean (H_{02}) of the distribution of Y_t.

Bootstrap test for $\mathbf{H}_{01}$.

1. Estimate ξ_α by $\widehat{\xi}_\alpha$ of (2.5).
2. Construct an estimator for the density function f, say $\widehat{f}_n$, in such a way that $\int_{-\infty}^a \widehat{f}_n(y)\mathrm{d}y = \alpha$.
3. Draw an independent sample $Y_1^*, \ldots, Y_n^*$ from $\widehat{f}_n$, and calculate $\widehat{\xi}_\alpha^*$ using (2.5) with the sample $\{Y_1^*, \ldots, Y_n^*\}$ instead of $\{Y_1, \ldots, Y_n\}$.
4. Repeat Step 3 B times, and the *achieved significance level* is defined as the relative frequency of the occurrence of the event $\{|\widehat{\xi}_\alpha^* - a| \geq |\widehat{\xi}_\alpha - a|\}$ among the B repetitions. See Efron and Tibshirani (1993, p. 232).

Remark 1. For testing the hypothesis that $\xi_\alpha \geq a$, the above test will be changed as follows: the achieved significance level is defined as the relative frequency of the occurrence of the event $\{\widehat{\xi}_\alpha^* \leq \widehat{\xi}_\alpha\}$.

Bootstrap test for $\mathbf{H}_{02}$.

1. Estimate τ_α by $\widehat{\tau}_\alpha$ of (2.6).
2. Construct an estimator for the density function f, say $\widehat{f}_n$, in such a way that
$$\frac{\int_{-\infty}^a (a-y)\widehat{f}_n(y)\mathrm{d}y}{\int_{-\infty}^\infty |y-a|\widehat{f}_n(y)\mathrm{d}y} = \alpha.$$
3. Draw an independent sample $Y_1^*, \ldots, Y_n^*$ from $\widehat{f}_n$, and calculate $\widehat{\tau}_\alpha^*$ using (2.6) with the sample $\{Y_1^*, \ldots, Y_n^*\}$ instead of $\{Y_1, \ldots, Y_n\}$.
4. Repeat Step 3 B times, and the achieved significance level is the relative frequency of the occurrence of the event $\{|\widehat{\tau}_\alpha^* - a| \geq |\widehat{\tau}_\alpha - a|\}$.

We adapt the smoothed empirical likelihood method (Chen and Hall, 1993) to construct the density estimators required in the above bootstrap tests.

The estimator for testing hypothesis H_{0j} $(j=1,2)$ can be defined as

$$\widehat{f}_n(y) = \frac{1}{h}\sum_{i=1}^n \widehat{p}_i K\left(\frac{y-Y_i}{h}\right), \tag{2.7}$$

where

$$(\widehat{p}_1,\ldots,\widehat{p}_n) = \underset{p\in\mathcal{A}_j}{\operatorname{argmax}} \prod_{i=1}^n p_i, \tag{2.8}$$

and

$$\begin{aligned}
\mathcal{A}_1 &= \left\{(p_1,\ldots,p_n)\middle| p_i\ge 0, \sum_{i=1}^n p_i = 1, \int_{-\infty}^a \frac{1}{h}\sum_{i=1}^n p_i K\left(\frac{y-Y_i}{h}\right) \mathrm{d}y = \alpha\right\},\\
\mathcal{A}_2 &= \left\{(p_1,\ldots,p_n)\middle| p_i\ge 0, \sum_{i=1}^n p_i = 1,\right.\\
&\qquad \left.\frac{\int_{-\infty}^a \frac{1}{h}\sum_{i=1}^n p_i(a-y)K\left(\frac{y-Y_i}{h}\right)\mathrm{d}y}{\int_{-\infty}^{\infty} \frac{1}{h}\sum_{i=1}^n p_i|a-y|K\left(\frac{y-Y_i}{h}\right)\mathrm{d}y} = \alpha\right\}.
\end{aligned}$$

In the above expressions, $h>0$ is a bandwidth and K is a density function.

It is easy to see that estimator (2.7) is the conventional kernel density estimator except that we use the weights $\{\widehat{p}_i\}$ instead of the uniform weight n^{-1} to ensure hypothesis H_{01} or H_{02} holds under $\widehat{f}_n$. The constrained optimization problem (2.7) can be solved using the standard procedure for empirical likelihood for means; see Owen (2001). For example, for testing hypothesis H_{01},

$$\widehat{p}_i = \frac{1}{n-\lambda(U_i-\alpha)}, \qquad i=1,\cdots,n,$$

where $U_i = \int_{-\infty}^a \frac{1}{h}K(\frac{y-Y_i}{h})dy$, and λ is the solution of the equation

$$\sum_{i=1}^n \frac{U_i-\alpha}{n-\lambda(U_i-\alpha)} = 0.$$

For testing hypothesis H_{02},

$$\widehat{p}_i = \frac{1}{n-\lambda V_i}, \qquad i=1,\cdots,n,$$

where $V_i = \int_{-\infty}^a \frac{1}{h}(a-y)K(\frac{y-Y_i}{h})dy - \alpha\int_{-\infty}^{\infty}\frac{1}{h}|a-y|K(\frac{y-Y_i}{h})dy$, and λ is the solution of the equation

$$\sum_{i=1}^n \frac{V_i}{n-\lambda V_i} = 0.$$

If we choose K as Gaussian kernel, $\mathcal{A}_j$ $(j=1,2)$ can be simplified as

$$\begin{aligned}
\mathcal{A}_1 &= \left\{(p_1,\ldots,p_n)\middle| p_i\ge 0, \sum_{i=1}^n p_i = 1, \sum_{i=1}^n p_i\Phi\left(\frac{a-Y_i}{h}\right) = \alpha\right\},\\
\mathcal{A}_2 &= \left\{(p_1,\ldots,p_n)\middle| p_i\ge 0, \sum_{i=1}^n p_i = 1,\right.\\
&\qquad \left.\frac{\sum_{i=1}^n p_i\left\{(a-Y_i)\Phi\left(\frac{a-Y_i}{h}\right)+h\varphi\left(\frac{a-Y_i}{h}\right)\right\}}{\sum_{i=1}^n p_iY_i - a + 2\sum_{i=1}^n p_i\left\{(a-Y_i)\Phi\left(\frac{a-Y_i}{h}\right)+h\varphi\left(\frac{a-Y_i}{h}\right)\right\}} = \alpha\right\},
\end{aligned}$$

where φ and Φ denote, respectively, the standard normal density function and its distribution function. Further, the bootstrap sample in Step 3 of the above tests can be equivalently drawn by taking $Y_i^* = Z_i^* + h\epsilon_i$ $(i = 1, \ldots, n)$, where $\{\epsilon_i\}$ are independent standard normal random variables, and $\{Z_i^*\}$ are independent samples from the discrete distribution given by

$$\begin{array}{c|cccc} & Y_1 & Y_2 & \ldots & Y_n \\ \hline \text{probability} & \widehat{p}_1 & \widehat{p}_2 & \ldots & \widehat{p}_n \end{array} \tag{2.9}$$

Alternatively we can use an unsmoothed empirical likelihood procedure to estimate the density f (Owen, 2001), which corresponds to the limit of the above procedure as $h \to 0$. In practice, this implies drawing the bootstrap sample $(Y_1^*, \ldots Y_n^*)$ directly from distribution (2.9), in which $(\widehat{p}_1, \ldots, \widehat{p}_n)$ is determined by (2.8) with $\mathcal{A}_j$ replaced by $\mathcal{B}_j$ for testing hypothesis H_{0j} $(j = 1, 2)$, where

$$\mathcal{B}_1 = \left\{ (p_1, \ldots, p_n) \,\middle|\, p_i \geq 0, \sum_{i=1}^n p_i = 1, \sum_{i:\, Y_i \leq a} p_i = \alpha \right\},$$

$$\mathcal{B}_2 = \left\{ (p_1, \ldots, p_n) \,\middle|\, p_i \geq 0, \sum_{i=1}^n p_i = 1, \frac{\sum_{i:\, Y_i \leq a} |Y_i - a| p_i}{\sum_{i=1}^n |Y_i - a| p_i} = \alpha \right\}.$$

3 Tests for Lyapunov Exponents of Deterministic Systems

3.1 Lyapunov exponents

We consider the one-dimensional discrete time dynamical system

$$X_{t+1} = F(X_t), \tag{3.1}$$

where F is a differentiable and bounded function. When the system is chaotic (Chan and Tong, 2001, Ch. 2), we need to take account of the sensitivity to the initial condition. To quantify the sensitivity, let X_0 and X_0' denote two nearby initial values. Then, after n iterates,

$$\begin{aligned} X_n - X_n' &= F^{(n)}(X_0) - F^{(n)}(X_0') \approx \left\{ \frac{\mathrm{d}}{\mathrm{d}x} F^{(n)}(X_0) \right\} (X_0 - X_0') \\ &= \left\{ \prod_{t=0}^{n-1} \dot{F}(X_t) \right\} (X_0 - X_0') = \pm \exp\left\{ \frac{1}{n} \sum_{t=0}^{n-1} \log |\dot{F}(X_t)| \right\} (X_0 - X_0'), \end{aligned} \tag{3.2}$$

where $F^{(n)}$ denotes the n-fold composition of F, and $\dot{F}$ denotes the derivative of F. If the limit $\lambda(X_0) \equiv \lim_{n\to\infty} \frac{1}{n} \sum_{t=0}^{n-1} \log |\dot{F}(X_t)|$ exists, we have a simple approximation $|X_n - X_n'| \approx \exp\{n\lambda(X_0)\} |X_0 - X_0'|$. When $\lambda(X_0)$ is a constant over the attractor of F, $\lambda \equiv \lambda(X_0)$ is called the Lyapunov exponent. The existence of one positive Lyapunov exponent is a necessary condition for the system being chaotic. See Eckmann and Ruelle (1985), and Chan and Tong (2001).

For the dynamical system in (3.1), its attracting set gives a global picture of its long-term behaviour. A more detailed picture is presented by the invariant probability measures with supports in this attracting set. Among those invariant measures, a class of *ergodic* (also called *indecomposable*) measures are of particular interest. Essentially, an ergodic measure gives the proportions of time, in the long term, that the system spends on different parts of an attractor. Therefore, for any P being such an invariant ergodic measure and A a measurable set,

$$P(A) = \lim_{n\to\infty} \frac{1}{n} \sum_{t=0}^{n-1} \mathcal{I}\{F^{(t)}(X_0) \in A\} \quad (a.s. - P). \tag{3.3}$$

This is an ergodic theorem for deterministic chaos; see, for example, Eckmann and Ruelle (1985). The notation '$a.s. - P$' implies that the equality holds for all the initial values X_0 in a set, with probability 1, with respect to the measure P. A dynamical system can carry uncountably many distinct ergodic measures, and all of them are mutually singular (that is, for any two ergodic measures, P_1 and P_2, there exists a set A for which $P_1(A) = 1$ and $P_2(A) = 0$). Of practical interest are the ergodic measures which are defined on sets with positive Lebesgue measures. Those measures are associated with the attractors with positive Lebesgue measures. For more discussion on attractors and invariant measures, see Eckmann and Ruelle (1985).

Suppose that P is an ergodic invariant probability measure of system (3.1). Suppose further that P has a density function p (with respect to Lebesgue measure). By (3.3), for this measure, the Lyapunov exponent can be expressed as

$$\begin{aligned} \lambda &= \int \log|\dot{F}(x)|P(\mathrm{d}x) = \int \log|\dot{F}(x)|p(x)\mathrm{d}x = E\{\log|\dot{F}(X_t)|\}, \\ &= \lim_{n\to\infty} \frac{1}{n} \sum_{t=0}^{n-1} \log|\dot{F}\{F^{(t)}(X_0)\}| \quad (a.s. - P). \end{aligned} \tag{3.4}$$

As in the above expression, we can formally treat $\{X_t\}$ as a strictly stationary stochastic process, and X_t has the marginal distribution P. The relation (3.3) ensures that this process is ergodic.

3.2 Statistical tests

Based on a finite observed time series $\{X_t,\ 1 \le t \le n\}$ from (3.1), a full recovery of an attracting set and its associated invariant measures is a formidable task. Instead, our primary interest is to detect whether the Lyapunov exponent in the attractor in which trajectory $\{X_t,\ t = 1, 2, \ldots\}$ (discarding transient period) lives is positive. We assume that P is an ergodic invariant probability measure on this attractor and P has a density function. Hence the Lyapunov exponent can be expressed as in (3.4). Using the bootstrap tests developed in Section 2.2, we are able to accomplish our goal. Furthermore, a scrutiny of a relevant distribution helps us to describe the initial-value sensitivity of the system in terms of other indices. Technically, the ALAD estimate (2.5) and the ALS estimate (2.6) are the building blocks of the proposed tests. Under the ergodicity of (3.3), those estimates are still consistent (see Theorem 1). However, the asymptotic Normality stated in Theorem 2 is no longer relevant, since the time series $\{X_t\}$ generated from deterministic model (3.1) is not even α-mixing. The findings on the auto-dependence of logistic maps are reported in Hall and Wolff (1995).

For a one-dimensional model (3.1), it is of little difficulty to estimate function F and its derivative $\dot{F}$ numerically within the range of the given data. Therefore, we assume that the data

$$Y_t \equiv \log|\dot{F}(X_t)| \quad (t = 1, \ldots, n)$$

are known. Since we may treat $\{X_t\}$ as a strictly stationary process with a marginal distribution continuous with respect to Lebesgue measure, the process $\{Y_t\}$ is also strictly stationary, and its marginal distribution has a density function. It follows from (3.3) that $\lambda = E(Y_1)$, which is the 50% expectile of Y_1. In order to find evidence of chaos or, more weakly, operational determinism, as in Yao and Tong (1998), we test a necessary but not sufficient condition, $\lambda > 0$, *via* the hypothesis $H : \tau_{0.5} > 0$, where τ_α is the 100α-th expectile of distribution Y_1. The bootstrap tests for hypothesis H_{02} in Section 2.2 are readily applicable here (see also Remark 1). Note that the test in this case is slightly easier to implement since $\widehat{\tau}_{0.5} = n^{-1}\sum_{t=1}^n Y_t$ (refer to (2.3)). Tests for hypotheses that λ exceeds, equals, or is less than a, where a is a constant, can be carried out in the similar way.

It is easy to see from (3.2) that the Lyapunov exponent offers a rough approximation of the divergence of trajectories, and is not necessarily to be a unique exponent to utilise. The quantity

$$\alpha_0 \equiv P\{\log|\dot{F}(X_t)| > 0\}$$

may be another important index, along with the Lyapunov exponent, to pronounce chaotic behaviour of the system. For example, in the event that $\alpha_0 \approx 1$ the system would appear to diverge in every iterations. With the given data, a natural estimate for α_0 is

$$\widehat{\alpha}_0 = \#\{\log|\dot{F}(X_t)| > 0\}/n.$$

Since, it is an estimate from a *finite* sample, we would like to test whether the above estimate is reliable. Mathematically, we need to test $H : \xi_{\alpha_1} = 0$ with $\alpha_1 = 1 - \widehat{\alpha}_0$, which is a special case of hypothesis H_{01} in Section 2.2. Now ξ_α denotes the percentile of random variable $Y_t = \log|\dot{F}(X_t)|$. A general form of hypothesis H_{01} corresponds to test whether $P\{\log|\dot{F}(X_t)| > a\} = \alpha$.

On the other hand, is an expectile τ_α with $\alpha \neq 0.5$ of any meaningful implication here? From (2.4), we have that

$$\frac{1-\alpha}{\alpha} = \frac{E\left[\log\left|\dot{F}(X_t)\right| - \tau_\alpha \mathcal{I}\left\{\log\left|\dot{F}(X_t)\right| > \tau_\alpha\right\}\right]}{E\left\{\tau_\alpha - \log\left|\dot{F}(X_t)\right| \mathcal{I}\log\left|\dot{F}(X_t)\right| \leq \tau_\alpha\right\}}.$$

Therefore, τ_α is the level for which the ratio of the average 'overshoot' in the attractor of $\log|\dot{F}(X_t)|$, to this level, to the average undershoot (or trajectory convergence) in the attractor is $1-\alpha$ to α. Further hypotheses can be considered in terms of the above interpretation.

3.3 Numerical examples

Figures 3.1 and 3.2 are based on tests for expectiles for simulated standard Normal data. In both cases, 100 experiments were performed for each value of a in a test of $H_0 : \tau_{0.5} = a$ *versus* $H_1 : \tau_{0.5} \neq a$, each time simulating an independent sample of size 30, and applying 500 bootstrap replications to compute the achieved significance level for the test on each sample.

Figure 3.1 displays boxplots of each of the 100 p-values for the test, where $a = \Phi^{-1}(\alpha)$, for $\alpha = 0.50, 0.55, \ldots, 0.95$, and where Φ is the standard Normal cumulative distribution function. (The p-values are, of course, the achieved significance levels as discussed in Section 2.2.) In other words, we test the "true" value of a, and we should each time retain H_0. The horizontal dotted lines at 0.05 and 0.10 are for reference. It is clear from this plot that the test performs very well for these data in terms of achieved significance.

Figure 3.2 displays the mean power for each of the 100 experiments, along with experimental error limits at two standard errors, in tests of $H_0 : \tau_{0.5} = 0$, when the true value of the expectile is $\Phi^{-1}(\gamma)$, $\gamma = 0.5, 0.51, \ldots, 0.95$. The resulting power is quite acceptable for a non-parametric test, though with the curious effect of diminishing at extreme values of the expectile.

Figure 3.1

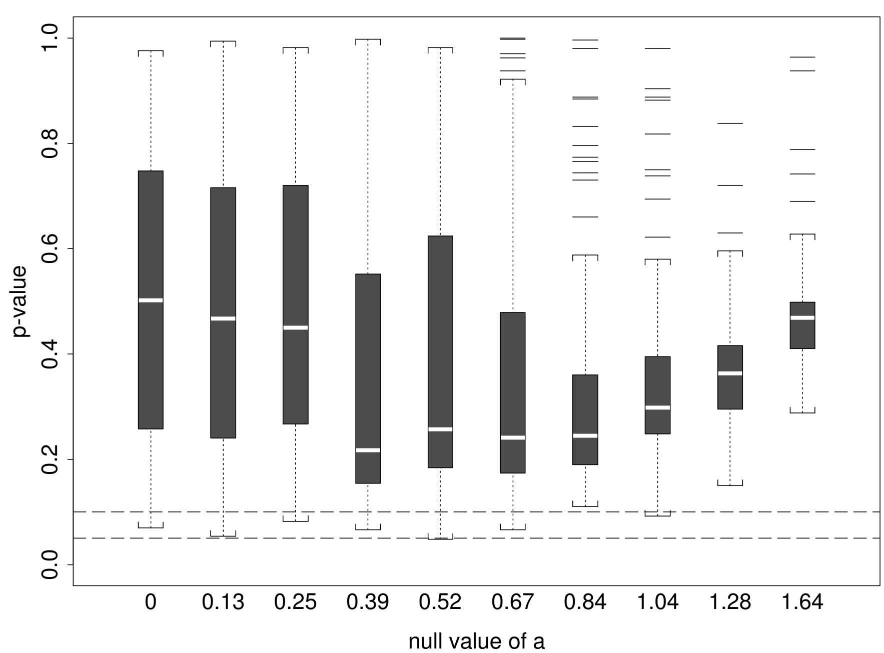
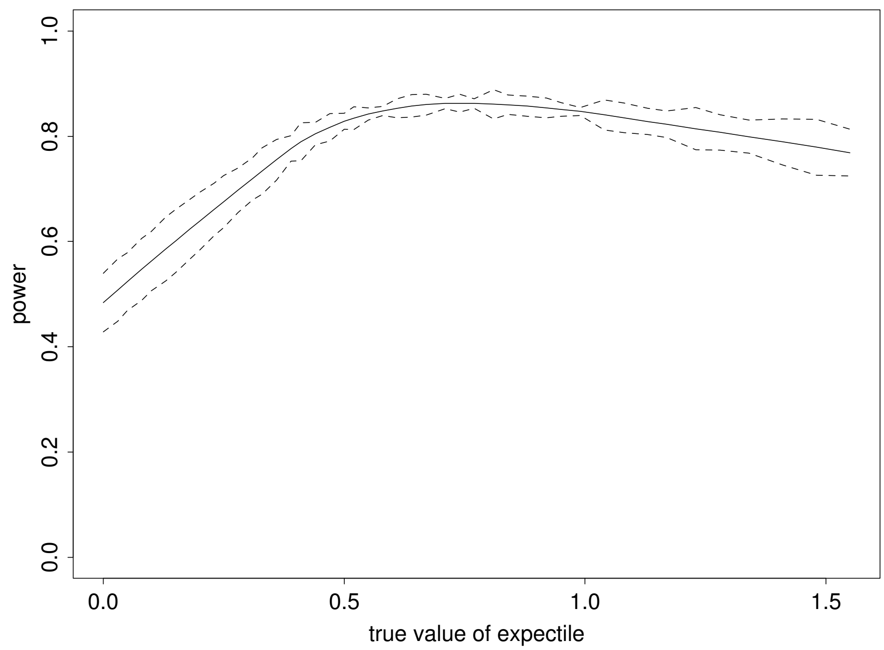


Figure 3.2



Figures 3.3 and 3.4 are based on tests for percentiles and expectiles, respectively, for simulated realisations of the logistic map, given by

$$X_{t+1} = F(X_t) \equiv \theta X_t (1 - X_t), \tag{3.5}$$

with $\theta = 4$. For this value of θ alone in (3.5) can the Lyapunov exponent be obtained analytically, and is known to be $\lambda = \log 2 > 0$. (In fact, while other values of $\theta < 4$ appear to produce chaotic behaviour, it can not be shown theoretically that the behaviour is indeed chaotic: see Hall and Wolff, (1995).) In both cases, 100 experiments were performed, based on independent series of length 30, and applying 1000 bootstrap replications to compute the achieved significance level for the test on each sample. The reason why a larger number of bootstrap replications were used here than in the first two experiments was because of the issue over mixing, as discussed in Section 2.1. In the case of percentiles, we test the hypothesis $\xi_{\alpha_1} = 0$, and in the case of expectiles, we test the hypothesis $\tau_{0.5} > 0$, as discussed in Section 3.2, and in which α_1 is also defined.

Figure 3.3 displays a boxplot of each of the 100 p-values for the test based on the percentile. The horizontal dotted lines at 0.05 and 0.10 are for reference. It can be seen that there is clear evidence of tendency for divergence of trajectories in each experiment.

Figure 3.4 displays a boxplot of each of the 100 p-values for the test based on the expectile. The horizontal dotted lines at 0.05 and 0.10 are once again for reference. It can be seen that five of the 100 experiments render a p-value of less than 0.05, in which cases the hypothesis of a positive Lyapunov exponent would be rejected (wrongly).

Figure 3.3

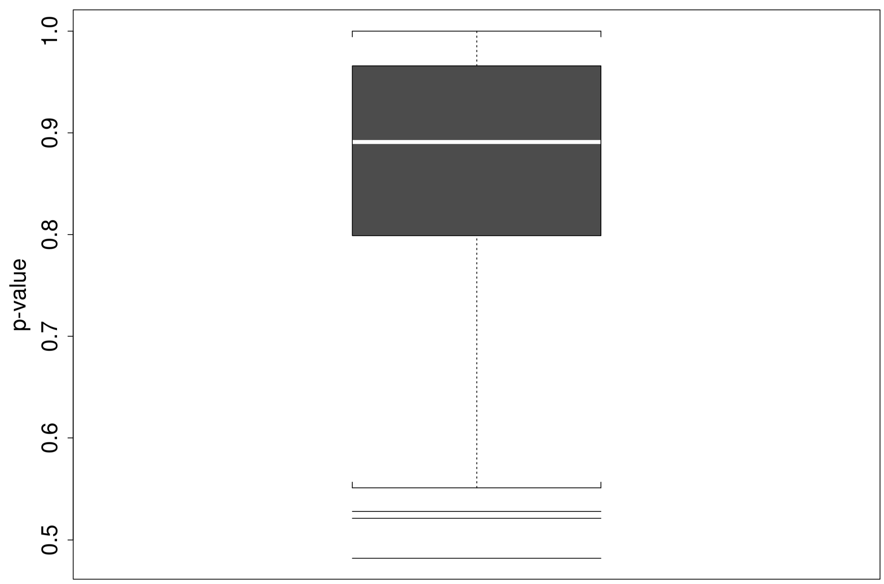
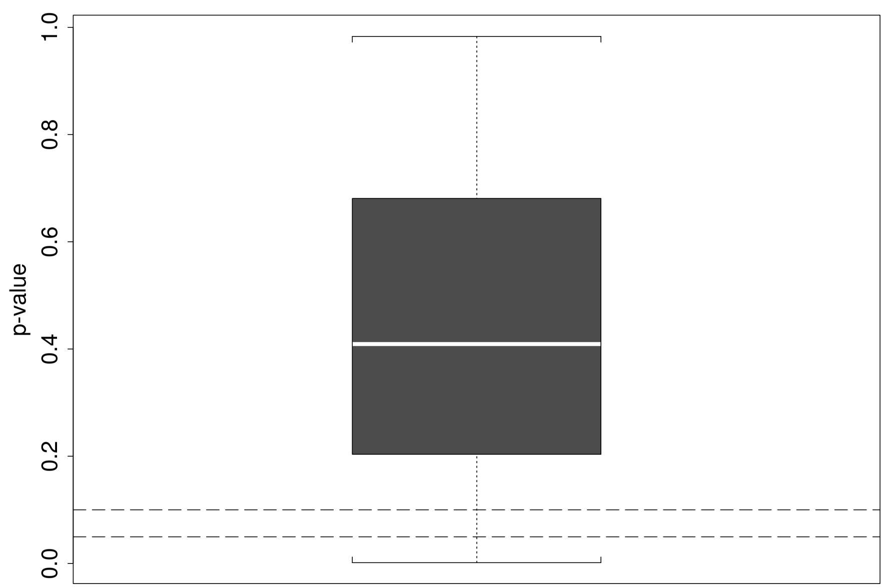


Figure 3.4



3.4 Some consequential issues

The present results are for one-dimensional, discrete time systems. Continuous chaotic processes must evolve in at least three dimensions (so that trajectories do not cross, and thus enable unique images of the chaotic function at each point). To generalise our method would require an estimate of the three-dimensional map — or, more specifically, its derivatives — and the method of Whang and Linton (1999) might be of use in this regard. Data requirements for higher dimensional systems can be formidable, and it is not clear what sample sizes might be required to give a bootstrap method integrity. In the alternative, dimension reduction, such as a principal components approach, may help, particularly if the direction of greatest separation of trajectories can be identified: see Broomhead *et al.* (1992) for a possible solution to this problem.

An issue of broad interest is the effect of dynamic noise on a chaotic system; that is, adjusting (3.1) to be

$$X_{t+1} = F(X_t) + \omega_{t+1},$$

for some sequence of independent and identically distribution noise terms $\{\omega_t\}$. We are cautious about treatment of such a situation which uses a gradient measure for a Lyapunov exponent: apart from the technical issue of the system noise possibly leading a trajectory outside the dynamic range of the map F and giving rise to an explosive time series, the main issue is in interpretation of the Lyapunov exponent. For deterministic systems, it has a geometrical interpretation, in the context of the system's *deterministic* attractor. In the presence of noise, this interpretation is lost. It is for this reason that Yao and Tong (1994) and Fan *et al.* (1996) chose to generalise this concept to a probabilistic sensitivity measure.

Finally, a natural question is to consider the power of the present test. While results may be readily available for stochastic systems, it would be a very inexact study for chaotic maps. Consider the possibly best understood chaotic map, the logistic map, given by (3.5). It is tempting to consider values of θ such that the Lyapunov exponent, λ, is close to zero, and thereby study the power of the test. It is known that $\lambda < 0$ and $\lambda > 0$ for explicit values of θ. Plots of λ *versus* θ abound, and one is given in Hall and Wolff (1995), who further note that such a plot may be highly discontinuous and that computational foibles most likely mask the true character of the plot. Specifically, they show increasing numerical accuracy leads to greater detail in such a plot, particularly identifying regions where $\lambda < 0$ may exist but where less accurate computations fail to reveal such facts. Therefore, choosing values of θ known to have values of λ close to zero is fraught with possibility for error, and is likely to confound a power study.

4 Proofs

We now prove Theorems 1 and 2. We use the same notation as in Section 2.1. Basically, the proofs are the application of the Convexity Lemma (Pollard, 1991). The proof of Theorem 1 is straightforward. The main idea of the proof of Theorem 2 is to approximate the objective functions in (2.5) (or (2.6)) by a quadratic function whose minimiser is asymptotically Normal, and then to show that $\widehat{\xi}_\alpha$ (or $\widehat{\tau}_\alpha$) lies close enough to the minimiser to share the latter's asymptotic behaviour. The Convexity Lemma plays a key role in the above approximation.

Proof of Theorem 1. Let $H_n(b) = n^{-1}\sum_{i=1}^n R_\alpha(Y_i - b)$, and $H(b) = E\{R_\alpha(Y_1 - b)\}$. Then both H_n and H are convex functions. Since the process $\{Y_i\}$ is ergodic, $H_n(b) \xrightarrow{P} H(b)$ for any $b \in R^1$.

For any $\varepsilon > 0$, let

$$A(\varepsilon) = \{\, x \in R^1 \,|\, \varepsilon \le |x - \xi_\alpha| \le 1\}.$$

Since ξ_α is the unique minimiser of H, there exists a constant $\eta > 0$ for which $H(b) > H(\xi_\alpha) + \eta$ for all $b \in A(\varepsilon)$. By the Convexity Lemma (Pollard, 1991), $H_n(b) \xrightarrow{P} H(b)$ uniformly for $b \in A(\varepsilon)$. Therefore, for all sufficiently large n,

$$\min_{b\in A(\varepsilon)} H_n(b) > H(\xi_\alpha) + \eta/2 > H_n(\xi_\alpha).$$

This implies that for all sufficiently large n, the function H_n has a local minimiser in the interval $(\xi_\alpha - \varepsilon, \xi_\alpha + \varepsilon)$. Since H_n is convex, its local minimiser is also the global minimiser $\widehat{\xi}_\alpha$. Hence, $P\{|\widehat{\xi}_\alpha - \xi_\alpha| < \varepsilon\} \to 1$. The proof for the consistency of $\widehat{\tau}_\alpha$ is similar and omitted here.

Proof of Theorem 2. The proofs for (i) and (ii) are similar. We only prove (i) since it is technically more involved.

For $\theta \in R^1$, we define

$$G_n(\theta) = \sum_{i=1}^n \left\{ R_\alpha\left(Y_i - \xi_\alpha - \frac{\theta}{\sqrt{n}}\right) - R_\alpha(Y_i - \xi_\alpha)\right\}.$$

We express the function G_n as

$$G_n(\theta) = E\{G_n(\theta)\} + \frac{\theta}{\sqrt{n}}\sum_{i=1}^n D(Y_i - \xi_\alpha) + RR_n(\theta), \tag{4.1}$$

where

$$D(x) = \begin{cases} \alpha & (x > 0) \\ -(1-\alpha) & (x \le 0), \end{cases}$$

$$\begin{aligned} RR_n(\theta) &= \sum_{i=1}^n \left\{ R_\alpha\left(Y_i - \xi_\alpha - \frac{\theta}{\sqrt{n}}\right) - R_\alpha(Y_i - \xi_\alpha) - D(Y_i - \xi_\alpha)\frac{\theta}{\sqrt{n}}\right\} \\ &\quad - \sum_{i=1}^n E\left\{ R_\alpha\left(Y_i - \xi_\alpha - \frac{\theta}{\sqrt{n}}\right) - R_\alpha(Y_i - \xi_\alpha)\right\}. \end{aligned}$$

Note that $E\{D(Y_i - \xi_\alpha)\} = 0$, therefore $E\{RR_n(\theta)\} = 0$ also. The function D plays the role of the derivative of R_α in the sense that for and any $x, y \in R^1$,

$$|R_\alpha(x+y) - R_\alpha(x) - D(x)y| \le |y|\mathcal{I}\left(|x| \le |y|\right).$$

Consequently,

$$
\begin{aligned}
&\operatorname{Var}(RR_n(\theta)) \\
&\leq \sum_{i=1}^{n} E\left\{R_\alpha\left(Y_i-\xi_\alpha-\frac{\theta}{\sqrt{n}}\right)-R_\alpha(Y_i-\xi_\alpha)-D(Y_i-\xi_\alpha)\frac{\theta}{\sqrt{n}}\right\}^2 \\
&\quad +2\sum_{1\leq i<j\leq n}\left|\operatorname{Cov}\left\{R_\alpha\left(Y_i-\xi_\alpha-\frac{\theta}{\sqrt{n}}\right)-R_\alpha(Y_i-\xi_\alpha)-D(Y_i-\xi_\alpha)\frac{\theta}{\sqrt{n}},\right.\right. \\
&\qquad \left.\left. R_\alpha\left(Y_j-\xi_\alpha-\frac{\theta}{\sqrt{n}}\right)-R_\alpha(Y_j-\xi_\alpha)-D(Y_j-\xi_\alpha)\frac{\theta}{\sqrt{n}}\right\}\right| \\
&\leq \theta^2 P\{|Y_1-\xi_\alpha|<\theta/\sqrt{n}\}\left\{1+2\sum_{i=1}^{n-1}(1-i/n)\rho_i\right\},
\end{aligned}
$$

which converges to 0. Therefore, $RR_n(\theta)=o_p(1)$. It follows from (2.2) that $E\{G_n(\theta)\}=\frac{\theta^2}{2}f(\xi_\alpha)+o(1)$. From (4.1), we have

$$
G_n(\theta)=\frac{1}{2}f(\xi_\alpha)\theta^2+\frac{\theta}{\sqrt{n}}\sum_{i=1}^{n}D(Y_i-\xi_\alpha)+o_p(1). \tag{4.2}
$$

It is easy to see that the minimiser of the first two terms on the RHS of the above expression is

$$
\widehat{\theta}=\frac{1}{\sqrt{n}f(\xi_\alpha)}\sum_{i=1}^{n}D(Y_i-\xi_\alpha),
$$

which is asymptotically Normal with mean 0 and variance $\alpha(1-\alpha)/\{f(\xi_\alpha)\}^2$ (see Theorem 2.4 of Peligrad, 1986). By the Convexity Lemma, the convergence of (4.2) is uniform on any compact set in R^1. Using the same arguments as in Pollard (1991, p. 193), we can show that

$$
\sqrt{n}(\widehat{\xi}_\alpha-\xi_\alpha)=\widehat{\theta}+o_p(1).
$$

Therefore, (i) holds. The proof is completed.

Acknowledgments

The authors thank an anonymous referee for helpful comments regarding generalisations of our work, and for pointing out some important relevant references. Rodney Wolff thanks the Principal and Fellows of Jesus College, Oxford, for the generous provision of a Visiting Senior Research Fellowship, during the tenure of which this work was completed. Qiwei Yao's research was partially supported by an EPSRC (UK) research grant. Howell Tong's research was partially supported by the Research Grants Council of Hong Kong and the Distinguished Research Achievement Award from the University of Hong Kong.

References

Bailey, B.A., S. Ellner and D.W. Nychka (1997): "Chaos with confidence: asymptotics and applications of local Lyapunov exponents," *Fields Institute Communications*, 11, 115-133.

Bradley, R.C. (1986): "Basic properties of strong mixing conditions," in E. Eberlein and M.S. Taqqu, eds, *Dependence in Probability and Statistics*, Boston: Birkhäuser, 165-192.

Broomhead, D.S., J.P. Huke and M.R. Muldoon (1992): "Linear filters and non-linear systems," *Journal of the Royal Statistical Society Series B*, 54, 373-382.

Chan, K.S., and H. Tong (2001): *Chaos: a Statistical Perspective*, New York: Springer.

Chen, S.X., and P. Hall (1993): "Smoothed empirical likelihood confidence intervals for quantiles," *Annals of Statistics*, 21, 1166-1181.

Eckmann, J.P., and D. Ruelle (1985): "Ergodic theory of chaos and strange attractors," *Reviews of Modern Physics*, 57, 617-656.

Efron, B. (1991): "Regression percentiles using asymmetric squared error loss," *Statistica Sinica*, 1, 93-125.

Efron, B., and R.J. Tibshirani (1993): *An Introduction to the Bootstrap*, New York: Chapman and Hall.

Fan, J., and Q. Yao (2003): *Nonlinear Time Series: Nonparametric and Parametric Methods*, New York: Springer.

Fan, J., Q. Yao and H. Tong (1996): "Estimation of conditional densities and sensitivity measures in nonlinear dynamical systems," *Biometrika*, 83, 189-206.

Giannerini, S., and R. Rosa (2001): "New resampling method to assess the accuracy of the maximal Lyapunov exponent estimation," *Physica D*, 155, 101-111.

Golia, S., and M. Sandri (2001): "A resampling algorithm for chaotic time series," *Statistics and Computing*, 11, 241-255.

Győrfi, L., W. Härdle, P. Sarda and P. Vieu (1989): *Non-parametric Curve Estimation from Time Series*, Berlin: Springer-Verlag.

Hall, P., and B. Presnell (1999a): "Density estimation under constraints," *Journal of Computation and Graphical Statistics*, 8, 259-277.

Hall, P., and B. Presnell (1999b): "Intentionally biased bootstrap methods," *Journal of the Royal Statistical Society Series B*, 61, 143-158.

Hall, P., and R.C. Wolff (1995): "Properties of invariant distributions and Lyapunov exponents for chaotic logistic maps," *Journal of the Royal Statistical Society Series B*, 57, 439-452.

Lu, Z.Q., and R.L. Smith (1997): "Estimating local Lyapunov exponents," *Fields Institute Communications*, 11, 135-151.

Neway, W.K., and J.K. Powell (1987): "Asymmetric least squares estimation and testing," *Econometrica*, 55, 819-847.

Nychka, D., S. Ellner, D. McCaffrey and A.R. Gallant (1992): "Finding chaos in noisy systems," *Journal of the Royal Statistical Society Series B*, 54, 399-426.

Owen, A. (2001): *Empirical Likelihood*, Boca Raton: Chapman and Hall / CRC.

Peligrad, M. (1986): "Recent advances in the central limit theorem and its weak invariance principle for mixing sequences of random variables," in E. Eberlein and M.S. Taqqu, eds, *Dependence in Probability and Statistics*, Boston: Birkhäuser, 193-223.

Pollard, D. (1991): "Asymptotic for least absolute deviation regression estimators," *Econometric Theory*, 7, 186-198.

Ruelle, D. (1989): *Chaos and Strange Attractors*, Cambridge: Cambridge University Press.

Shintani, M., and O. Linton (2003): "Nonparametric neural network estimation of Lyapunov exponents and a direct test for chaos," *Journal of Econometrics*, in press.

Tong, H. (1990): *Non-linear Time Series: a Dynamical Systems Approach*, Oxford: Oxford University Press.

Whang, Y.J., and O. Linton (1999): "The asymptotic distribution of nonparametric estimates of the Lyapunov exponent for stochastic time series," *Journal of Econometrics*, 91, 1-42.

Wolff, R.C. (1994): "Independence in time series: another look at the BDS test (with Discussion)," *Philosophical Transactions of the Royal Society of London Series A*, 348, 383-395.

Yao, Q., and H. Tong (1994): "Quantifying the influence of initial values on non-linear prediction," *Journal of the Royal Statistical Society Series B*, 56, 701-725.

Yao, Q., and H. Tong (1996): "Asymmetric least squares regression estimation: a nonparametric approach," *Journal of Nonparametric Statistics*, 6, 273-292.

Yao, Q., and H. Tong (1998): "A bootstrap detection for operational determinism," *Physica D*, 115, 49-55.