NONPARAMETRIC NEURAL ESTIMATION OF LYAPUNOV EXPONENTS AND A DIRECT TEST FOR CHAOS

by

Mototsugu Shintani
Vanderbilt University, Nashville

Oliver Linton
London School of Economics and Political Science

Contents:

Abstract
1. Introduction
2. Model and Assumptions
3. Theoretical results
4. Discussions
5. Simulation results
6. Application to financial data
7. Conclusion
Appendix
References
Tables 1 – 4
Figures 1 and 2

Discussion Paper
No.EM/03/455
May 2003

The Suntory Centre
Suntory and Toyota International Centres for
Economics and Related Disciplines
London School of Economics and Political Science
Houghton Street
London WC2A 2AE
Tel.: 020 7955 6698

* We thank Doug Nychka and Barb Bailey for providing their computer program. We also thank an associate editor, two referees, Xiaohong Chen, Rainer Dahlhaus, Yanqin Fan, Ron Gallant, Joon Park, Peter Phillips, and Yoon Wang for comments, and the National Science Foundation for financial support.

Abstract

This paper derives the asymptotic distribution of the nonparametric neural network estimator of the Lyapunov exponent in a noisy system. Positivity of the Lyapunov exponent is an operational definition of chaos. We introduce a statistical framework for testing the chaotic hypothesis based on the estimated Lyapunov exponents and a consistent variance estimator. A simulation study to evaluate small sample performance is reported. We also apply our procedures to daily stock return data. In most cases, the hypothesis of chaos in the stock return series is rejected at the 1% level with an exception in some higher power transformed absolute returns.

Keywords: Artificial neural networks; nonlinear dynamics; nonlinear time series; nonparametric regression; sieve estimation.

JEL Nos.: C14, C22.

Contact addresses:
Professor Oliver Linton, Department of Economics, London School of Economics and Political Science, Houghton Street, London WC2A 2AE, UK. Email: o.linton@lse.ac.uk

Mototsugu Shintani, Assistant Professor, Department of Economics, Vanderbilt University, Calhoun Hall, Nashville, TN 37235, USA. Email: mototsugu.shintani@vanderbilt.edu

1 Introduction

The Lyapunov exponent, which measures the average rate of divergence or convergence of two nearby trajectories, is a useful measure of the stability of a dynamic system. To obtain the Lyapunov exponent from observed data, Eckmann and Ruelle (1985) and Eckmann, Kamphorst, Ruelle, and Ciliberto (1986) proposed a method based on nonparametric regression which is known as the Jacobian method. While any nonparametric regression estimator can be employed in the Jacobian method, one of the most widely used approaches in applications is the Lyapunov exponent estimator based on neural networks proposed by Nychka, Ellner, Gallant, and McCaffrey (1992).[1] For example, applications using this approach in economics include: Dechert and Gençay's (1992) analysis of foreign exchange rates; studies on monetary aggregates by Serletis (1995) and Barnett, Gallant, Hinich, Jungeilges, Kaplan and Jensen (1995); and the analysis of stock return series by Abhyankar, Copeland and Wong (1997). However, despite the popularity of this Jacobian method using neural networks, empirical researchers have been confined to reporting only the point estimates of the Lyapunov exponents, as the distributional theory is not known.

This paper first derives the asymptotic distribution of the neural network estimator of the Lyapunov exponent. A formal statistical framework regarding the sign of the Lyapunov exponent is then introduced, based on a consistent estimator of the asymptotic variance. In a recent paper by Whang and Linton (1999), the asymptotic normality of the Jacobian-based estimator using a kernel-type nonparametric regression was derived. The basic idea of our approach is to combine the theoretical result of Whang and Linton (1999) and the recent results on neural network asymptotics obtained by Chen and White (1999) and others. The conditions for asymptotic normality of the estimator, in terms of the number of hidden units in neural network models as well as the block length, are derived for both one-dimensional and multidimensional

[1] A similar procedure was independently proposed by Gençay and Dechert (1992) with more emphasis on embedded dynamics.

cases. The required growth rate of block length and the convergence rate of neural network estimator of Lyapunov exponent are compared to those based on kernel estimators.

The positivity of the Lyapunov exponent in a bounded dissipative nonlinear system is a widely used formal definition of chaos (Eckmann and Ruelle, 1985). Furthermore, chaos can be defined not only in a deterministic system but also in a stochastic system using the same definition (such a generalization of the notion of chaos is sometimes referred to as noisy chaos as opposed to deterministic chaos). Since we allow the presence of stochastic noise in the system, the consistent estimation of the asymptotic variance of the Lyapunov exponent estimator offers a formal statistical framework for testing the hypothesis of positive Lyapunov exponent in a stochastic environment. In other words, we can construct a direct test for chaos using the consistent standard error proposed in this paper.[2]

Following former theoretical studies on the statistical properties of the neural network estimator of the Lyapunov exponent, including McCaffrey (1991), McCaffrey, Ellner, Gallant and Nychka (1992), and Nychka et al. (1992), we focus on a class of single hidden layer feedforward artificial neural networks. The most notable theoretical advantage of using neural networks seems to be their universal approximation property. Theoretically, neural networks are expected to perform better than other approximation methods at least within the confines of the particular class of functions considered. Especially with high-dimensional models, "[the neural net form,] compared to the preceding functional approximations, ... is not sensitive to increasing d [dimension] (McCaffrey et al., 1992, p. 689)." This universal approximation property also applies to the derivatives (Gallant and White, 1992). Since the nonparametric estimation of the first derivative is required in the Jacobian method, this fact is useful in the context of Lyapunov exponent estimation. In contrast, as Ellner, Gallant, McCaffrey and Nychka (1991, p.362) pointed out, kernel

[2]The well-known BDS test proposed by Brock, Dechert, Scheinkman, LeBaron (1996) should be viewed as a test for i.i.d. against general dependence which include chaos rather than a direct test for chaos.

methods often provide poor derivative estimates.

On the whole, simulation results available in the literature are favorable to the neural network method. With respect to the flexibility of neural networks, the simulation results reported in Kuan and White (1994) show the near-exact approximation property of neural networks, even if the nonlinear function is complex enough to generate chaos. The robustness of neural networks to the choice of the number of hidden units is reported in a simulation in Gallant and White (1992), while the advantage of using BIC in selecting the number of hidden units and dimension is reported in Nychka et al. (1992). The reliability of the Jacobian method based on neural networks was recently reaffirmed by a single-blind controlled competition conducted by Barnett, Gallant, Hinich, Jungeilges, Kaplan, and Jensen (1997). In our paper, after presenting the theoretical results, the small sample properties of our procedure are examined using the data set used in the competition of Barnett et al. (1997) as well as using the artificially generated chaotic data in a noisy system. Finally, as an empirical application, we apply our procedure to the analysis of daily stock return series. This application is well-motivated since a certain type of economic model predicts chaos as a source of fluctuation in stock prices.

The remainder of the paper is organized as follows: Definitions of the Lyapunov exponent and the neural network estimator are presented in Section 2. Section 3 derives asymptotic properties of the Lyapunov exponent estimators based on neural networks and proposes test statistics. Some additional discussion is given in Section 4. Monte Carlo evidence is presented in Section 5. An empirical application is reported in Section 6. Some concluding remarks are made in Section 7. All proofs are given in the Appendix.

We will use the following notation throughout the paper. When $|\cdot|$ is applied to a $d \times 1$ vector $x = (x_1, \ldots, x_d)'$, it denotes a vector norm defined by $|x| \equiv \sum_{i=1}^{d} |x_i|$. Let $\mu = (\mu_1, \ldots, \mu_d)'$ denote a $d \times 1$ vector of non-negative integer constants; we denote $x^{\mu} = \prod_{i=1}^{d} x_i^{\mu_i}$ and

$$D^{\mu}g(x) = \frac{\partial^{|\mu|}g(x)}{\partial x_1^{\mu_1}, \ldots, \partial x_d^{\mu_d}},$$

for any real function $g(x)$ on $\mathbb{R}^d$. When μ is a scalar constant, as is the case when $d = 1$, we define $D^{\mu}g(x)$ to be the μ-th order derivative of $g(\cdot)$ evaluated at x with the convention that $D^0 g(x) = g(x)$ and $D^1 g(x) = Dg(x)$. We use $\mathcal{B}_d^m$ to denote a weighted Sobolev space of all functions on $\mathbb{R}^d$ that have continuous and uniformly bounded (partial) derivative up to order m. For $g \in \mathcal{B}_d^m$, the norm is defined by

$$\|g\|_{\mathcal{B}_d^m} \equiv \max_{0\leq|\mu|\leq m} \sup_{x\in\mathbb{R}^d} |D^{\mu}g(x)| < \infty$$

and the associated metric is defined with this norm. The symbols "$\Rightarrow$" and "$\xrightarrow{p}$" are used to signify convergence in distribution and convergence in probability, respectively. All the limits in the paper are taken as the sample size $T \to \infty$ unless noted otherwise.

2 Model and Assumptions

Let $\{x_t\}_{t=1}^T$ be a random scalar sequence generated by the following nonlinear autoregressive model

$$x_t = \theta_0(x_{t-1}, \ldots, x_{t-d}) + u_t, \tag{1}$$

where θ_0: $\mathbb{R}^d \to \mathbb{R}$ is a nonlinear dynamic map and $\{u_t\}$ is a sequence of random variables. The model (1) can be expressed in terms of a map with an error vector $U_t = (u_t, 0, \ldots, 0)'$ and the map function

F: $\mathbb{R}^d \to \mathbb{R}^d$ such that

$$Z_t = F(Z_{t-1}) + U_t \tag{2}$$

where $Z_t = (x_t, \ldots, x_{t-d+1})' \in \mathbb{R}^d$. Let J_t be the Jacobian of the map F in (2) evaluated at Z_t. Specifically, we define

$$J_t = \begin{bmatrix} \Delta\theta_{01t} & \Delta\theta_{02t} & \cdots & \Delta\theta_{0,d-1,t} & \Delta\theta_{0dt} \\ 1 & 0 & \cdots & 0 & 0 \\ 0 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & 0 \end{bmatrix} \tag{3}$$

for $t = 0, 1, \ldots, T-1$, where $\Delta\theta_{0jt} = D^{e_j}\theta_0(Z_t)$ for $j = 1, \ldots, d$ in which $e_j = (0, \ldots, 1, \ldots, 0)' \in \mathbb{R}^d$ denotes the j-th elementary vector.

Let $\widehat{\theta}$ be the nonparametric neural network estimator of the target function θ_0 in (1). In this paper, we consider the feed-forward single hidden layer networks with a single output. Following Chen and Shen (1998) and Chen and White (1999), we view this neural network estimator as a special case of the sieve extremum estimator. To be more specific, we view it as a problem of maximizing an empirical criterion, $L_T(\theta)$, over the neural network sieve, Θ_T, which is a sequence of approximating parameter spaces that is dense in the infinite dimensional parameter space, Θ, as $T \to \infty$.

The basic idea of the Jacobian method is to obtain $\widehat{J}_t$ by substituting $\widehat{\theta}$ in the Jacobian formula (3) and construct a sample analogue estimator of the Lyapunov exponent. Following the convention of neural network estimation of the Lyapunov exponent, we distinguish between the "sample size" T used for estimating Jacobian $\widehat{J}_t$ and the "block length" M which is the number of evaluation points used for estimating the Lyapunov exponent. Since the number of evaluation points is less than or equal to T, M

can be also understood as the sample size of a subsample. The neural network estimator of i-th largest Lyapunov exponent is given by

$$\widehat{\lambda}_{iM} = \frac{1}{2M} \ln \nu_i \left(\widehat{\mathbf{T}}'_M \widehat{\mathbf{T}}_M \right), \qquad \widehat{\mathbf{T}}_M = \prod_{t=1}^{M} \widehat{J}_{M-t} = \widehat{J}_{M-1} \cdot \widehat{J}_{M-2} \cdot \cdots \cdot \widehat{J}_0, \tag{4}$$

for $1 \leq i \leq d$, where $\nu_i (A)$ is i-th largest eigenvalue of a matrix A,

$$\widehat{J}_t = \begin{bmatrix} \Delta\widehat{\theta}_{1t} & \Delta\widehat{\theta}_{2t} & \cdots & \Delta\widehat{\theta}_{d-1,t} & \Delta\widehat{\theta}_{dt} \\ 1 & 0 & \cdots & 0 & 0 \\ 0 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & 0 \end{bmatrix}, \tag{5}$$

and $\Delta\widehat{\theta}_{jt} = D^{e_j}\widehat{\theta}(Z_t)$ for $t = 0,\ 1, \ldots, M-1$. For notational convenience we have just taken the first M observations. However, in practice, there are several alternative choices of subsample, a matter that will be discussed in subsection 4.2.

Below, we introduce two groups of assumptions. One is on the dynamics and the other is on the neural networks.

Assumptions on Dynamics

A1. (a) $\{Z_t\}_{t=1}^{T}$ *is a strictly stationary* β*-mixing sequence with a mixing coefficient satisfying* $\beta(j) \leq \beta_0 j^{-\zeta}$ *for some* $\beta_0 > 0$, $\zeta > 2$, *where the* β*-mixing coefficient is given by*

$$\beta(j) = E \sup \left\{ \left| P(B|\mathcal{F}_{-\infty}^{0}) - P(B) \right| : B \in \mathcal{F}_{j}^{\infty} \right\},$$

where $\mathcal{F}_s^t$ *is the* σ*-field generated by* $(Z_s, \ldots, Z_t)$.

(b) *The distribution of* Z_t *is absolutely continuous with respect to Lebesgue measure with marginal density function* f *with a compact support* $\mathcal{Z}$ *in* $\mathbb{R}^d$. *The initial condition* Z_0 *is a random variable generated from the same distribution.*

A2. $\{u_t\}_{t=1}^T$ *is a random sequence of either:* (a) *i.i.d. with* $E(u_t) = 0$ *and* $E(u_t^2) = \sigma^2 < \infty$, *or*

(b) *martingale difference with* $E(u_t|\mathcal{F}_{-\infty}^{t-1}) = 0$ *and* $E(u_t^2|\mathcal{F}_{-\infty}^{t-1}) = \sigma_t^2 \in [\varepsilon, \varepsilon^{-1}]$ *for some* $\varepsilon > 0$.

A3.

$$\theta_0 \in \Theta = \left\{ \theta : \theta(z) \ = \int \exp(ia'z) d\mu_\theta(a), \quad \|\mu_\theta\|_3 \equiv \int l(a)^3 d|\mu_\theta|(a) \leq C < \infty \right\},$$

where μ_θ *is a complex-valued measure on* $\mathbb{R}^d$, $|\mu_\theta|$ *denotes total variation of* μ_θ , $l(a) = \max\left[(a'a)^{1/2}, 1\right]$ *and* $a' = (a_1, \ldots, a_d) \in \mathbb{R}^d$.

A4. *The system (1) has distinct Lyapunov exponents defined by*

$$\lambda_i \equiv \lim_{M \to \infty} \frac{1}{2M} \ln \nu_i \left(\mathbf{T}_M' \mathbf{T}_M\right) < \infty, \qquad \mathbf{T}_M = \prod_{t=1}^{M} J_{M-t} = J_{M-1} \cdot J_{M-2} \cdot \cdots \cdot J_0,$$

for $1 \leq i \leq d$.

A5. *For* $1 \leq i \leq d$ *and some* $\phi \geq 0$,

$$\max_{1 \leq t \leq M} |F_{i,t-1}(J_{M-1}, \ldots, J_0)| = O_p(M^\phi),$$

where

$$F_{i,t-1}(J_{M-1}, \ldots, J_0) = \frac{\partial \ln \nu_i \left(\mathbf{T}_M' \mathbf{T}_M\right)}{\partial \Delta\theta(Z_{t-1})} \text{ and } \Delta\theta(Z_t) = (\Delta\theta_{1,t}, \Delta\theta_{2,t}, \ldots, \Delta\theta_{d,t})'.$$

A6. *For* $1 \leq i \leq d$,

$$\Phi_i \equiv \lim_{M\to\infty} \text{var}\left[\frac{1}{\sqrt{M}}\sum_{t=1}^{M}\eta_{it}\right]$$

is positive and finite, where

$$\eta_{it} = \xi_{it} - \lambda_i \textit{ with } \xi_{it} = \frac{1}{2}\ln\left(\frac{\nu_i\left(\mathbf{T}_t'\mathbf{T}_t\right)}{\nu_i\left(\mathbf{T}_{t-1}'\mathbf{T}_{t-1}\right)}\right) \textit{ for } \ t \geq 2 \textit{ and } \xi_{i1} = \frac{1}{2}\ln\nu_i\left(\mathbf{T}_1'\mathbf{T}_1\right).$$

Assumptions on Neural Networks

B1. *The neural network estimator* $\widehat{\theta}_T$ *is an extremum sieve estimator that satisfies*

$$L_T(\widehat{\theta}_T) \geq \sup_{\theta\in\Theta_T} L_T(\theta) - O(\varepsilon_T^2)$$

with $\varepsilon_T \to 0$ *as* $T \to \infty$, *where* $L_T(\theta)$ *is a least square criterion*

$$L_T(\theta) = \frac{1}{T}\sum_{t=1}^{T} l(\theta, x_t, Z_{t-1}) = -\frac{1}{T}\sum_{t=1}^{T}\frac{1}{2}(x_t - \theta(Z_{t-1}))^2.$$

B2. *The neural network sieve* θ_T: $\mathbb{R}^d \to \mathbb{R}$ *is an approximation function in the parameter space* Θ_T *satisfying*

$$\theta_T(z) \ = \beta_0 + \sum_{j=1}^{2^k r(T)} \beta_j l(a_j)^{-2}\psi(a_j'z + b_j)$$

with

$$\max_{1\leq j\leq 2^k r(T)} |a_j| \leq C_T, \quad \sum_{j=0}^{2^k r(T)} |\beta_j| \leq B_T,$$

where ψ *is an activation function,* $a_j \in\mathbb{R}^d$, $b_j, \beta_j \in\mathbb{R}$ *are parameters, and* k *is the number related to the class of activation function defined in B3 below.*

B3. *The activation function* ψ *is a possibly nonsigmoid function satisfying* $\psi \in \mathcal{B}_1^2$ *and is k-finite for some* $k \geq 2$*, namely,*

$$0 < \int_{\mathbb{R}} |D^k \psi(u)| du < \infty.$$

B4. *For any* $(a', b), (a_1', b_1) \in \mathbb{R}^d \times \mathbb{R}$*, there exists an* $\alpha \in (0, 1]$ *associated with* $\psi \in \mathcal{B}_1^2$ *such that for all* z *in the compact support* S*,*

$$\left\| \psi_{a,b} - \psi_{a_1,b_1} \right\|_{\mathcal{B}_1^2} \leq const. \times \left[\left((a - a_1)'(a - a_1) \right)^{1/2} + |b - b_1| \right]^{\alpha},$$

where $\psi_{a,b}(z)$ *is the rescaled activation function defined by* $\psi_{a,b}(z) = l(a)^{-2}\psi(a'z + b)$*.*

Remarks on Assumptions on Dynamics

A1, A2, and A3 are conditions on the data, the error term and the class of nonlinear function, respectively, required to obtain the convergence rate of the neural network estimator. While many nonlinear Markov processes are known to be stationary β-mixing, A1 is slightly stronger than the condition used in Whang and Linton's (1999) study on the kernel estimator that allows α-mixing (strong mixing). A1 can be also replaced by imposing some additional conditions on u_t and θ_0 as discussed in Chen, Racine and Swanson (2001, Lemma 2.1). A3 implies that we consider the class of functions that have finite third absolute moments of the Fourier magnitude distributions. This type of smoothness condition was used by Barron (1993) when he showed that the rate of neural network approximation does not depend on the input dimension d. Since the Jacobian method requires estimation of partial derivatives, a convergence result in a stronger norm is required. For this purpose, we follow Hornik, Stinchcombe, White and Auer (1994) and use the scaling factor $l(a)^3$ to derive the approximation rate in Sobolev norm of higher order. In contrast to Barron's original condition that requires only the first derivatives to be bounded, A3 requires the boundedness of the third derivatives ($\theta_0 \in \mathcal{B}_d^3$). However, requirement of third derivatives for any dimension d is much weaker than Whang and Linton's (1999) case since kernel regression essentially requires higher order differentiability for higher dimensional model to maintain the same rate of convergence.

A4 defines the Lyapunov exponents of the system (1). Since the largest (or dominant) Lyapunov exponent λ_1 has often been of main interest in the literature, we mainly focus our analysis on the largest Lyapunov exponent and simply use notation λ to denote λ_1. However, it should be noted that other exponents λ_i for $2 \leq i \leq d$ also contain some important information related to the stability of the system, including the directions of divergence and contraction of trajectories (see Nychka et al., 1992) and the types of non-chaotic attractors (see Dechert and Gençay, 1992). Necessary conditions for the existence of λ have been discussed in the literature (e.g., see Nychka et al., 1992, p. 406). It is known that, if J_t is ergodic and stationary and if $\max\{\ln \nu_1(J_t' J_t), 0\}$ has a finite expectation, then the limit in A4 almost surely exists and will be a constant, irrespective of the initial condition. When $\sigma^2 = 0$, the system (1) reduces to a deterministic system and the interpretation of $\lambda > 0$ is identical to the definition of deterministic chaos. For moderate σ^2, the stochastic system generated by (1) can also have sensitive dependence to initial conditions, and noisy chaos with $\lambda > 0$ can be also defined. For example, a stationary linear autoregressive process has $\lambda < 0$, while the unit root and the explosive autoregressive process imply $\lambda \geq 0$. One interesting question here is whether the Lyapunov exponent is continuous in the amount of noise for small amounts of noise. Specifically, let λ_σ denote the Lyapunov exponent for a noisy system with error variance σ^2 and let λ_0 be the Lyapunov exponent for the deterministic skeleton with $\sigma^2 = 0$. We suspect that $\lim_{\sigma \to 0} \lambda_\sigma = \lambda_0$. This is certainly the case for a large class of processes including the linear autoregressive processes, but we do not have a proof that works under general conditions. Under further continuity properties, our distributional theory in the next section can also be extended to 'small sigma' asymptotics, i.e., to work under the condition that $\sigma \to 0$.

A5 and A6 are assumptions identical to the ones used by Whang and Linton (1999) in their B(6)* and B(7)*. The role of these assumptions can be better understood by considering the one-dimensional case. When $d = 1$, A5 and A6 simplify to

A5*. *For some* $\phi \geq 0$,

$$\max_{1\leq t\leq M}\left(|D\theta_0(x_{t-1})|^{-1}\right) = \left(\min_{1\leq t\leq M}|D\theta_0(x_{t-1})|\right)^{-1} = O_p(M^{\phi}).$$

A6*.

$$\Phi \equiv \lim_{M\to\infty} \text{var}\left[\frac{1}{\sqrt{M}}\sum_{t=1}^{M}\eta_t\right]$$

is positive and finite, where $\eta_t = \ln|D\theta_0(x_{t-1})| - \lambda$.

A5* is a condition on the properties of the data around first derivative being zero and is closely related to extreme value theory for stochastic processes (see Whang and Linton, 1999, p. 9). With this assumption, we have a valid Taylor series expansion of the estimator of Lyapunov exponent. The condition is weak and is expected to hold for many chaotic processes including the well-known logistic map (with $\phi = 1$). A6* provides the asymptotic variance of the local Lyapunov exponent that will be introduced in the next section. In general, Φ is the long-run variance of η_t and differs from the variance of η_t. However, since η_t is a weakly dependent process, if we take an equally spaced subsample of size M (instead of block), η_t becomes an asymptotically independent sequence with Φ being its variance.

Remarks on Assumptions on Neural Networks

In this paper, the nonparametric neural network estimator $\widehat{\theta}$ for the unknown functional form θ_0 in the nonlinear autoregressive model (1) is obtained by the least squares method. B1 allows an approximate maximization problem where exact maximization is included as a special case when $\varepsilon_T = 0$. Similar to the case shown in Chen and Shen (1998), our asymptotics in the next section are valid as long as ε_T converges to zero faster than the theoretical rate of convergence of the estimator. B2 implies that the neural network sieve consists of $2^k r(T)$ number of hidden units with common activation function ψ. Since the asymptotic theory will only depend on the increasing rate rather than the exact number of hidden units, we simply refer $r(T)$ as the number of hidden unit. Typically, ψ is a sigmoid function defined by a bounded measurable function on $\mathbb{R}$ with $\psi(u) \to 1$ as $u \to \infty$, and $\psi(u) \to 0$ as $u \to -\infty$. However, it is known that the universal approximation property of the neural networks is not confined to the ones with sigmoid activation functions. Indeed, the neural networks with nonsigmoid activation functions such as radial basis activation functions are becoming more popular in applications. B3 is from Hornik et al. (1994) and allows nonsigmoid as well as sigmoid activation functions. B4 is a Hölder condition on the activation function used in Chen and White (1999) and is stronger than B3. While B3 is sufficient to derive our main theoretical result, B4 will be later used to investigate full sample asymptotics ($M = T$) in subsection 4.2 since it requires the improved rate for the derivative estimator (Lemma 2).

3 Theoretical results

3.1 Uniform convergence rate of the derivative estimator

The Jacobian-based estimator of the Lyapunov exponent requires the estimation of the first derivative at Z_t, namely, $\Delta\theta_0(Z_t) = (\Delta\theta_{01,t}, \Delta\theta_{02,t}, \ldots, \Delta\theta_{0d,t})'$. Since the neural network estimator $\widehat{\theta}$ is obtained by selecting values for the parameters a_j's, b_j's, and β_j's in B2 by minimizing the least square criterion in B1, the derivative estimator $\Delta\widehat{\theta}(Z_t) = (\Delta\widehat{\theta}_{1,t}, \Delta\widehat{\theta}_{2,t}, \ldots, \Delta\widehat{\theta}_{d,t})'$ can be obtained by using an analytical derivative of the neural network sieve in B2 evaluated at selected values of the parameters and Z_t. We first provide the uniform convergence rate for this derivative estimator.

Lemma 1. *Suppose that assumptions A1 to A4 and B1 to B3 hold,* $B_T \geq const. \times \|\mu_\theta\|_3$*,* $C_T = const.$ *and* $r(T)$ *satisfies* $r^2 \ln r = O(T)$*. Then*

$$\sup_{z \in \mathcal{Z}} \left| \Delta\widehat{\theta}(z) - \Delta\theta_0(z) \right| = O_p([T/\ln T]^{-1/4}).$$

The improved rate for the derivative estimator can be further obtained by employing a Hölder condition B4 on the activation function. See Makovoz (1996) and Chen and White (1999) for the relation between this condition and the source of improvement in the rate of approximation.

Lemma 2. *Suppose that assumptions in Lemma 1 and B4 hold,* $B_T \geq const. \times \|\mu_\theta\|_3$*,* $C_T = const.$ *and* $r(T)$ *satisfies* $r^{2(1+\alpha/d^*)} \ln r = O(T)$*, where* $d^* = d$ *if* ψ *is homogeneous* ($\psi(cz) = c\psi(z)$)*, and* $d^* = d + 1$ *otherwise. Then*

$$\sup_{z \in \mathcal{Z}} \left| \Delta\widehat{\theta}(z) - \Delta\theta_0(z) \right| = o_p(T^{-1/4}).$$

3.2 Asymptotic distribution of Lyapunov exponent estimator

We begin with investigating the asymptotic behavior of $\widehat{\lambda}$ for the scalar case ($d = 1$), mainly for the purpose of illustration, followed by the general results for the multidimensional case ($d \geq 2$). When $d = 1$, since $Z_t = x_t$, $\mathcal{Z} = \chi$, and $J_t = D\theta_0(x_t)$, the Lyapunov exponent estimator in (4) simplifies to

$$\widehat{\lambda}_M = \frac{1}{2M} \sum_{t=1}^{M} \ln \left[D\widehat{\theta}(x_{t-1})^2 \right].$$

To investigate the asymptotic properties of the estimator, it is convenient to introduce the notion of the local Lyapunov exponent defined by

$$\lambda_M = \frac{1}{2M}\sum_{t=1}^{M} \ln\left[D\theta(x_{t-1})^2\right].$$

Unlike the "global" Lyapunov exponent λ, the local Lyapunov exponent with finite M measures the short-term rate of divergence. It should also be noted that λ_M is a random variable in general. From the definition in A4, λ can be seen as a limit of λ_M with $M \to \infty$. Using λ_M, the total estimation error, $\widehat{\lambda}_M - \lambda$, with the normalizer $\sqrt{M}$ can be decomposed as

$$\sqrt{M}(\widehat{\lambda}_M - \lambda) = \sqrt{M}(\widehat{\lambda}_M - \lambda_M) + \sqrt{M}(\lambda_M - \lambda). \tag{6}$$

The second term represents the asymptotic behavior of the local Lyapunov exponent which is common to all Jacobian methods irrespective of the choice of the nonparametric estimator. The $\sqrt{M}$ rate of convergence and asymptotic normality for this term were derived by McCaffrey et al. (1992) and Bailey (1996), respectively.[3] The first term can be understood as the estimation error for the local Lyapunov exponent. In contrast to the second term, the asymptotic behavior of the first term depends on the estimation method. Whang and Linton (1999) employed kernel regression methods and showed that the asymptotic behavior of (6) is dominated by the second term under some conditions. For the neural network estimator, we can introduce new conditions on the rate of block length along with assumptions introduced in the previous section so that the first term has a negligible effect on the asymptotic behavior of (6).

Theorem 1. *Suppose that the assumptions in Lemma 1, A5* and A6* hold,* $M \to \infty$ *and* $M =$

[3] To be more specific, McCaffrey et al. (1992) decomposed the second term into block bias $E(\lambda_M) - \lambda$ and the block error $\lambda_M - E(\lambda_M)$ with the order of the block error being $1/\sqrt{M}$. See also Ellner et al. (1991).

$o([T/\ln T]^{1/(2+4\phi)})$. *Then,*

$$\sqrt{M}(\widehat{\lambda}_M - \lambda) \Rightarrow N(0, \Phi).$$

The multidimensional case ($d \geq 2$) can be also considered by applying similar arguments to all the i-th largest Lyapunov exponents for $1 \leq i \leq d$. Below, we have the main theoretical result of this paper.

Theorem 2. *Suppose that the assumptions in Lemma 1, A5 and A6 hold,* $M \to \infty$ *and* $M = o([T/\ln T]^{1/(2+4\phi)})$. *Then, for* $1 \leq i \leq d$,

$$\sqrt{M}(\widehat{\lambda}_{iM} - \lambda_i) \Rightarrow N(0, \Phi_i).$$

Remarks. The results show the asymptotic normality of Lyapunov exponent estimators that can be used in the inference. The convergence rate of Lyapunov exponent estimator depends on the growth rate of block length M and thus depends on ϕ with smaller ϕ implying faster convergence. When $\phi = 1$, which is satisfied by the logistic map (Whang and Linton, 1999), the Lyapunov exponent estimator converges at the rate $(T/\ln T)^{1/12-\varepsilon}$ where $\varepsilon > 0$ is an arbitrary small number.

It should be noted that both one-dimensional and multidimensional results are obtained using the same smoothness condition in A3 and same growth rate of block length. This contrasts to the results based on kernel smoothing methods. For example, by modifying the result of Whang and Linton (1999), Shintani and Linton (2003) showed that, with an optimal choice of the rate of bandwidth, the Lyapunov exponent estimator based on local quadratic smoother was $\sqrt{M}$ consistent with $M = o([T/\ln T]^{4/\{(d+6)(1+2\phi)\}})$. Thus, the convergence rate of the kernel-based Lyapunov exponent estimator becomes slower in the higher dimensional case.[4] Simple comparison with neural network case reveals that the two estimators have

[4] Lu and Smith (1997) also used the local quadratic regression method to estimate the local Lyapunov exponent λ_M for finite M.

the same rate when $d = 2$ but the rate for the kernel-based estimator is always slower than that of the neural network estimator for $d > 2$ for any value of ϕ. This advantage of the neural network approach comes from the powerful approximation properties of neural networks given in Lemma 1. In general, other nonparametric approaches yield slower rate of convergence or require stronger smoothness conditions when the dimension increases.

3.3 Test statistics

In this subsection, feasible test statistics are introduced and a one-sided test is proposed for the purpose of testing chaotic behavior of time series. First, we construct the test statistics based on the asymptotic results on Lyapunov exponent estimators obtained in the previous subsection. Suppose $\widehat{\Phi}$ is a consistent estimator of Φ in Theorem 1. Our primary interest is to test the null hypothesis $H_0 : \lambda \geq 0$ $(\lambda \leq 0)$ against the alternative of $H_1 : \lambda < 0$ $(\lambda > 0)$. Our test statistic is

$$\widehat{t} = \frac{\widehat{\lambda}_M}{\sqrt{\widehat{\Phi}/M}} \ . \tag{7}$$

We reject the null hypothesis if $\widehat{t} \leq -z_\alpha$ $(\widehat{t} \geq z_\alpha)$ where z_α is the critical value that satisfies $\Pr[Z \geq z_\alpha] = \alpha$ with Z being a standard normal random variable.

Next, we consider consistent estimation of Φ. In general, a heteroskedasticity and autocorrelation consistent (HAC) covariance matrix estimator (see e.g., Andrews, 1991) for Φ is required, since η_t's are serially dependent and not identically distributed.[5] For the one-dimensional case, the covariance estimator

[5] The resampling method may be an alternative way to compute the standard error of the estimated Lyapunov exponent. This line of research is pursued by Gençay (1996) and Giannerini and Rosa (2001). However, the computing burden of resampling neural network estimator seems to be the main problem in practice.

$\widehat{\Phi}$ is defined as:

$$\widehat{\Phi} = \sum_{j=-M+1}^{M-1} w(j/S_M)\widehat{\gamma}(j) \text{ and } \widehat{\gamma}(j) = \frac{1}{M}\sum_{t=|j|+1}^{M} \widehat{\eta}_t \widehat{\eta}_{t-|j|},$$

where $\widehat{\eta}_t = \ln|D\widehat{\theta}(x_{t-1})| - \widehat{\lambda}_M$ and where $w(x)$ and S_M denote a kernel function and a lag truncation parameter, respectively. For the multidimensional case, the test statistic $\widehat{t}_i = \widehat{\lambda}_{iM}/\sqrt{\widehat{\Phi}_i/M}$ with the covariance estimators $\widehat{\Phi}_i$ can be similarly constructed by replacing $\widehat{\eta}_t$ by

$$\widehat{\eta}_{it} = \widehat{\xi}_{it} - \widehat{\lambda}_{iM} \text{ with } \widehat{\xi}_{it} = \frac{1}{2}\ln\left(\frac{\nu_i\left(\widehat{\mathbf{T}}_t'\widehat{\mathbf{T}}_t\right)}{\nu_i\left(\widehat{\mathbf{T}}_{t-1}'\widehat{\mathbf{T}}_{t-1}\right)}\right) \text{ for } t \geq 2 \text{ and } \widehat{\xi}_{i1} = \frac{1}{2}\ln\nu_i\left(\widehat{\mathbf{T}}_1'\widehat{\mathbf{T}}_1\right).$$

For the covariance matrix estimation, we employ the following class of kernel functions $w : \mathbb{R} \to [-1, 1]$ similar to that used in Andrews (1991).

C1.

$$w \in \mathcal{W} = \Big\{ w : w(0) = 1, \quad w(-x) = w(x)\ \forall x \in \mathbb{R}, \quad \int_{-\infty}^{\infty} |w(x)|dx < \infty,$$
$$w(x) \text{ is continuous at 0 and at all but a finite number of other points}\Big\}.$$

Corollary 1. *Suppose that assumptions in Theorem 2 and C1 hold, $S_M \to \infty$ and $S_M = o(M^{1/2})$. Then, for $1 \leq i \leq d$, $\widehat{\Phi}_i \xrightarrow{p} \Phi_i$.*

Remarks. This result shows that the HAC estimation with given growth rate of bandwidth can be used to construct the standard error for Lyapunov exponents. Since the infeasible statistic $\widetilde{t}_i = (\widehat{\lambda}_{iM} - \lambda_i)/\sqrt{\widehat{\Phi}_i/M} \Rightarrow N(0,1)$, $\widehat{t}_i = \widetilde{t}_i + \lambda_i/\sqrt{\widehat{\Phi}_i/M}$ diverges to $-\infty$ (∞) for any λ_i under $H_1 : \lambda_i < 0$ $(\lambda_i > 0)$. Therefore, the test is consistent under reasonable conditions.

4 Discussions

4.1 Optimal block length and optimal subsampling scheme

It should be noted that the asymptotic results in the previous section required that the number of products of the Jacobian in the Lyapunov exponent estimate (M) be less than the sample size of data used in Jacobian estimation (T). Therefore, the choice of block length M is an important issue in practice. McCaffrey et al. (1992) discussed the optimal choice of block length by decomposing the local Lyapunov exponent asymptotics [the second term in (6)] into a bias term and a variance term. Furthermore, they suggested that averaging the Lyapunov exponent estimators from the nonoverlapping T/M blocks might reduce the overall bias (see also Ellner et al., 1991, and Nychka et al., 1992). However, it should be noted that such an estimate in the one-dimensional case is identical to the estimate based on a full sample ($M = T$).

Whang and Linton (1999) pointed out that the valid asymptotic results for the Lyapunov exponent estimators can be derived not only from the blocking method but also from any other subsampling method. This fact also raised a question of the optimal choice of subsampling scheme for a given number of M. Suppose the optimal choice is made on the grounds that it minimizes the variance Φ_i in A6 (or A6*). The comparison between the blocking scheme and the equally spaced subsampling scheme can then be understood from the following simple example.

Suppose we have three observations of the time series data (y_1, y_2, y_3) generated from the autoregressive (AR) process of order one. If we want to estimate the mean of the process using two observations out of three, we have only two alternatives; using the adjacent sample $[(y_1, y_2)$ or $(y_2, y_3)]$ or using the skipped sample $[(y_1, y_3)]$. The variance of such an estimate depends on the AR parameter. A simple calculation implies that the first scheme is more efficient when the parameter is negative and the second scheme is more efficient when the parameter is positive. Similarly, when the data are generated by the moving average (MA) process of order one, the first scheme is better when the MA parameter is positive and the second scheme is better when the parameter is negative.

This simple example shows that the optimal subsample for the Lyapunov exponent estimation depends on the data generating process. Therefore, we may use either the blocking scheme or equally spaced subsample scheme as a choice of subsample. For this reason, in this paper, we report the results based on equally spaced subsamples in addition to the results based on the commonly used blocking method in the simulation and empirical analysis.

4.2 Full sample estimation

As discussed by Ellner et al. (1991), it has been questioned whether the requirement of block length (M) less than full sample (T) is necessary in the theoretical analysis of asymptotic behavior of the neural network approach. When the Jacobians from the whole sample points are used for Lyapunov exponent calculation ($M = T$), the first term in (6) now enters the asymptotic behavior of the overall estimation error. Therefore, we can expect the full sample estimator to have a different asymptotic distribution from the one based on subsamples. Whang and Linton (1999) showed that the asymptotic distribution for a full sample estimator, based on kernel regression, can be derived if one employs stronger assumptions on the functional form. The purpose of this subsection is to illustrate that it is also possible in the neural network approach to derive the asymptotic results if the similar assumptions are employed. To simplify the argument, we only consider the one-dimensional case.

Corollary 2. *Suppose that assumptions in Lemma 2, A5* with $\phi = 0$ hold, η_t in A6* is replaced by*

$\eta_t = v(x_{t-1})u_t + \ln|D\theta_0(x_{t-1})| - \lambda$, *where*

$$v(x) = \frac{D^2\theta_0(x)}{\{D\theta_0(x)\}^2} - \frac{Df(x)}{\{D\theta_0(x)\}\, f(x)}.$$

Further assume that $f(x)/D\theta_0(x) = 0$ *at the boundary points* $\underline{x}$ *and* $\overline{x}$. *Then we have the asymptotic normality result in Theorem 1 with* $M = T$.

Remarks. To derive this result, stronger conditions for both activation function and target function need to be employed. Among all additional conditions, $\phi = 0$ is the most difficult requirement since it "is not satisfied by any univariate chaotic process that we are aware of (Whang and Linton, 1999, p.8)." The consistent estimator of Φ can be constructed by using the sample analogue of η_t, which requires a second derivative estimation of target function as well as density and density derivative estimation.

4.3 Upper bound estimation

The definition of ξ_{it} in Theorem 2 does not have a simple form as ξ_t in Theorem 1 since $\ln \nu_i\left(\mathbf{T}_M'\mathbf{T}_M\right) \neq \sum_{t=1}^{M} \ln \nu_i\left(J_{M-t}'J_{M-t}\right)$ for the multivariate case. However, for the largest Lyapunov exponent ($i = 1$), we have the following relation between the two quantities:

$$\sum_{t=1}^{M} \ln \nu_1\left(J_{M-t}'J_{M-t}\right) = \ln \prod_{t=1}^{M} \nu_1\left(J_{M-t}'J_{M-t}\right) \geq \ln \nu_1\left((\Pi_{i=1}^{M} J_{M-t})'(\Pi_{i=1}^{M} J_{M-t})\right) = \ln \nu_i\left(\mathbf{T}_M'\mathbf{T}_M\right).$$

Here, we used the matrix norm inequality $|\nu_1(A'A)|\,|\nu_1(B'B)| \geq |\nu_1((AB)'(AB))|$. Using this relationship, we can bound the largest Lyapunov exponent from above by $\overline{\lambda} \equiv \lim_{M\to\infty} \frac{1}{2M} \sum_{t=1}^{M} \ln \nu_1\left(J_{M-t}'J_{M-t}\right)$.

We can consistently estimate this quantity, using its sample analogue,

$$\widehat{\overline{\lambda}}_M \equiv \frac{1}{2M}\sum_{t=1}^{M}\ln \nu_1\left(\widehat{J}'_{M-t}\widehat{J}_{M-t}\right).$$

Corollary 3. *Suppose that assumptions in Lemma 1 hold, $F_{i,t-1}$ in A5 is replaced by $F_t = \partial \ln \nu_1 \left(J_t' J_t\right)/\partial\Delta\theta(Z_t)$, η_{it} in A6 is replaced by $\eta_t = \frac{1}{2}\ln \nu_1\left(J'_{t-1}J_{t-1}\right) - \overline{\lambda}$. If $M = O([T/\ln T]^{1/(2+4\phi)})$, then*

$$\sqrt{M}(\widehat{\overline{\lambda}}_M - \overline{\lambda}) \Rightarrow N(0, \Phi).$$

Remarks. For the multidimensional case, $\overline{\lambda}$ is always positive. This implies that the asymptotic distribution of the upper bound estimator seems to be useful only if the data is generated from a chaotic process (with positive λ). For example, when some specific positive value of the Lyapunov exponent is predicted by a theory, upper bound estimates below this value provide strong evidence against the hypothesis.

5 Simulation results

5.1 Logistic map

Since the testing procedure proposed in the previous section is based on asymptotic theory, it is of interest to examine its performance with sample sizes that are typical for economic time series. This section reports the result of the Monte Carlo experiments designed to assess the small sample performance of neural network estimates of Lyapunov exponent with various data generating processes.

We first examine the logistic map with system noise:

$$x_t = ax_{t-1}(1 - x_{t-1}) + \sigma\varepsilon_t,$$

where $\varepsilon_t / v_t \sim U(-1, 1)$ independent of x_t, and

$$v_t = \min\{ax_{t-1}(1 - x_{t-1}), 1 - ax_{t-1}(1 - x_{t-1})\}.$$

This particular form of heteroskedasticity ensures that the process x_t is restricted to the unit interval. It is interesting to note that this simple one-dimensional model contains both a globally stable case ($0 < a < 3$) and a chaotic case ($3.57 < a \leq 4$) depending on the parameter a. We use $a = 1.5$ as an example of a system with a negative Lyapunov exponent ($\lambda = -\ln 2$ when $\sigma = 0$) and $a = 4$ as that with a positive Lyapunov exponent ($\lambda = \ln 2$ when $\sigma = 0$).

For the neural network estimation, we use FUNFITS program developed by Nychka, Bailey, Ellner, Haaland and O'Connell (1996). As an activation function ψ, this program uses a type of sigmoid function

$$\psi(u) = \frac{u(1 + |u/2|)}{2 + |u| + u^2/2},$$

which was also employed by Nychka et al. (1992). For the estimation of Φ, Bartlett's kernel $w(u) = 1 - |u|$ with one lag is employed. We use the block subsample and equally spaced subsample in addition to the entire sample. To see how the results differ with the choice of the lags of the autoregression, we consider the cases with lag length d varying from 1 to 4. The results are based on the parameters $r = 4$, $\sigma = 0.25$, $T = 200$ with 1000 replications. For subsample estimation, we use $M = 66$ giving three blocks and estimates for each replication.[6] The results are reported in Table 1. When correct lag length is chosen ($d = 1$), the mean and the median of Lyapunov exponent estimates appeared close to the true value for both stable ($a = 1.5$) and chaotic ($a = 4$) cases. This outcome suggests that our method works well even in the small sample environment. When d increases, the number of estimates with incorrect sign increases for the stable case, while the estimates is robust to the additional lag lengths for the chaotic case.[7] One important implication of this observation is that we should be careful about the selection of lag length in the system since such information is usually not provided in practice. If the specification of the system is completely known, as in this subsection, a parametric approach such as the one employed by Bask and de Luna (2002) should yield a more efficient estimator as well as a powerful test. While our theory of nonparametric approach is designed for an unknown system with a given lag length, we expect information criteria such as BIC to provide a consistent lag selection procedure. For this reason, we utilize BIC to select lag length (as well as the number of hidden units) in the next subsection and the empirical section.

For the standard errors in Table 1, there is a systematic downward bias for the stable case, but those for the chaotic case are in close agreement with actual standard deviations. Figures 1 and 2 show the finite sample densities of the Lyapunov exponent estimates standardized by the mean and variance superimposed on the standard normal densities. The distribution shows some skewness, but with this small sample situation, it is close enough to normality predicted by the theory.[8]

---

[6] For the block length in the simulation and empirical analysis of this paper, we use $M = \text{int}[c \times (T/\ln T)^{1/6}]$ with $c = 36.2$ where $\text{int}[A]$ signifies the integer part of A.

[7] Gençay and Dechert (1996) have pointed out the possibility of obtaining spurious Lyapunov exponents which can be larger than the true largest Lyapunov exponent when embedded dynamics are used.

[8] We also conducted a simulation with a Henon map as an example of higher-dimensional chaotic process. Our approach worked as well as a logistic case provided sufficient lag length was used.

5.2 Barnett competition data

Powerful properties of the neural network approach were confirmed by the successful results in the single-blind controlled competition conducted by William Barnett. Detail of the competition design and the results can be found in Barnett et al. (1997). However, since they used only point estimates of the neural network approach, it is of interest to examine how statistical procedure in this paper works for the same data used in the competition.[9]

The competition used two different sample sizes, 380 and 2000. Both small sample data and large sample data are taken from a single observation generated from the following five different models with u_t being an i.i.d. standard normal random variable.

- Model I (Logistic map): $x_t = 3.57x_{t-1}(1 - x_{t-1})$ with $x_0 = 0.7$.
- Model II (GARCH): $x_t = h_t^{1/2}u_t$ where $h_t = 1 + 0.1x_{t-1}^2 + 0.8h_{t-1}$ with $h_0 = 1$ and $x_0 = 0$.
- Model III (NLMA): $x_t = u_t + 0.8u_{t-1}u_{t-2}$.
- Model IV (ARCH): $x_t = (1 + 0.5x_{t-1}^2)^{1/2}u_t$ with $x_0 = 0$.
- Model V (ARMA): $x_t = 0.8x_{t-1} + 0.15x_{t-2} + u_t + 0.3u_{t-1}$ with $x_0 = 1$ and $x_1 = 0.7$.

Of the five models described above, only Model I has a positive Lyapunov exponent. For this subsection and the empirical part of this paper, the number of lag length (d) and the number of hidden units (r) will be jointly determined by minimizing the BIC criterion defined by

$$BIC(d, r) = \ln \widehat{\sigma}^2 + \frac{\ln T}{T}\left[1 + r(d + 2)\right]$$

[9] The data is downloaded from the archive given in Barnett et al. (1997, footnote 2).

where $\widehat{\sigma}^2 = T^{-1} \sum_{t=1}^{T} \left(x_t - \widehat{\theta}(x_{t-1}, \ldots, x_{t-d}) \right)^2$. For the HAC estimation required for the standard error, we employ the QS kernel with optimal bandwidth selection method developed in Andrews (1991). The employed block length (M) for the small sample data ($T = 380$) is 72 giving a total of 5 blocks, while that for the large sample data ($T = 2000$) is 91 giving a total of 21 blocks.

The results for Barnett competition data are presented in Table 2. For the subsample estimates, the median values are reported. The results can be summarized as follows. First, the signs of all point estimates correspond to the true signs of the processes. Second, for models II to V, the positivity hypothesis is rejected at a 1% level of significance based on both full sample and subsample estimation. These results confirm the validity of the neural network approach and our testing procedure. Third, positivity of the Lyapunov exponent in model I is not rejected for both full sample and subsample cases. At the same time, it did not provide strong evidence against the negativity.

6 Application to financial data

Over the past decades, numerous models that can generate chaos in economic variables have been developed. For example, Brock and Hommes (1998) showed that chaos in stock price was possible if heterogeneous beliefs of agents were introduced in a traditional asset pricing model.[10] In this section, we apply our proposed procedure to investigate the possibility of chaos in the U.S. financial market using stock price series.[11]

We use daily observations on the Dow Jones Industrial Average (DJIA), P_t. The sample period extends from January 3, 1928, to October 18, 2000, providing a total of 18,490 observations. It should be noted

[10] See Abhyankar, Copeland and Wong (1997) for a survey of previous results of analyses of chaos using financial data.

[11] Other economic theories predict chaos in real aggregate series. The method proposed in this paper is also applied to international real output series by Shintani and Linton (2003).

that the period of the stock market crash of 1987 is included in the sample period. The stock return is simply defined as the difference of log of the stock price index ($R_t = \Delta \ln P_t$). Following Taylor's (1986) finding, it is now well-known that the volatility measures such as the absolute return ($|R_t|$) have higher autocorrelation compared to the return series (R_t). Ding, Granger and Engle (1993) also examined the correlation of power transformation of the absolute return ($|R_t|^k$) and found quite high autocorrelations. Extending this line of approach, we estimate the Lyapunov exponent of various power transformed absolute return series. Table 3 shows the sample autocorrelations of the transformed absolute DJIA stock returns $|R_t|^k$ for $k = 0.5, 1, 1.5, 2, 2.5$ in addition to those of the untransformed return series. The return series has small positive first order autocorrelation and small negative second order autocorrelation, while the transformed absolute return has much higher autocorrelations with $k = 1$ being the highest. These results are very similar to those of Ding, Granger and Engle (1993) based on S&P 500 series with a number of observations close to that of our data.

The estimated Lyapunov exponents for each series is presented in Table 4 along with the t statistics and p-values for the null hypothesis of positive Lyapunov exponent ($H_0 : \lambda \geq 0$). The block length (M) and the number of blocks used for subsampling estimates are 127 and 145, respectively. The number of hidden units (r) are selected using BIC. For all cases, the Lyapunov exponents from full sample estimation are negative, and the positivity hypothesis is significantly rejected at the 1% level with the exception of transformed series with $k = 2.5$. Similar strong evidence is obtained from subsample estimation except for the same series. Another interesting observation is that the Lyapunov exponents are larger for the transformed absolute returns than for the level of returns, suggesting less stability in volatility (or absolute values) than in returns themselves. These results from various transformed data offer strong statistical evidence against the chaotic explanation in stock returns. This strengthens the results in Abhyankar, Copeland and Wong (1997) who obtained negative Lyapunov exponent point estimates for both S&P500

cash and futures series with 5-minute and 1-minute frequencies.

7 Conclusion

This paper has derived the asymptotic distribution of the neural network Lyapunov exponent estimator proposed by Nychka et al. (1992) and introduced a formal statistical framework of testing hypotheses concerning the sign of the Lyapunov exponent. Such a procedure offers a useful empirical tool for detecting chaos in a noisy system. The small sample properties of the new procedure were examined in simulations, which indicate that the performance of the procedure is satisfactory in moderate-sized samples. The procedure was applied to investigate chaotic behavior of financial market. In most cases, we strongly rejected the hypothesis of chaos in the stock return series, with one mild exception in some higher power transformed absolute returns.

Acknowledgments

We thank Doug Nychka and Barb Bailey for providing their computer program. We also thank the associate editor, two anonymous referees, Xiaohong Chen, Rainer Dahlhaus, Yanqin Fan, Ron Gallant, Joon Park, Peter Phillips, and Yoon Whang for comments, and the National Science Foundation for financial support.

Appendix

Proof of Lemma 1

In this proof, we define the (weighted) L_p space (with distribution function $F(x)$) by the set of L_p-integrable functions with norm $\|g\|_p = \left\{\int_\chi |g(x)|^p \, dF(x)\right\}^{1/p}$ and associated metric from this norm. For $p = \infty$, we use $\|g\|_\infty = \sup_{x\in\chi} |g(x)|$. We will denote L_2 norm $\|g\|_2$ simply by $\|g\|$. Similarly, we define the (weighted) Sobolev W_p^m space with a set of functions with L_p-integrable (partial) derivatives up to order m with norm $\|g\|_{m,p} = \left\{\sum_{|\mu|=0}^m \int_\chi |D^\mu g(x)|^p \, dF(x)\right\}^{1/p}$ and associated metric from this norm. For $p = \infty$, we use $\|g\|_{m,\infty} = \max_{0\le|\mu|\le m} \sup_{x\in\chi} |D^\mu g(x)|$.

(a) To simplify the argument, we first derived the result for one-dimensional case, and then extend the result to the multidimensional case. For $d = 1$, we denote $\mathcal{Z} = \chi$ and our goal is to obtain the convergence rate for

$$\sup_{x\in\chi} \left|D\widehat{\theta}(x) - D\theta_0(x)\right|.$$

Note that interpolation inequality (See Gabushin, 1967, and Shen and Wong, 1994) implies

$$\|g(x) - g_0(x)\|_\infty \le K \|g(x) - g_0(x)\|^{(2m-1)/2m} \|D^m g(x) - D^m g_0(x)\|^{1/2m}.$$

where K is a fixed constant. Substituting $g(x) = D\widehat{\theta}(x)$, $g_0(x) = D\theta_0(x)$, $m = 1$ yields

$$\left\|D\widehat{\theta}(x) - D\theta_0(x)\right\|_\infty \le K \left\|D\widehat{\theta}(x) - D\theta_0(x)\right\|^{1/2} \left\|D^2\widehat{\theta}(x) - D^2\theta_0(x)\right\|^{1/2}.$$

If we use that

$$\left\|D\widehat{\theta}(x) - D\theta_0(x)\right\| \le \left\|\widehat{\theta}(x) - \theta_0(x)\right\|_{1,2} \le \left\|\widehat{\theta}(x) - \theta_0(x)\right\|_{2,2}$$

and

$$\left\|D^2\widehat{\theta}(x) - D^2\theta_0(x)\right\| \le \left\|\widehat{\theta}(x) - \theta_0(x)\right\|_{2,2},$$

the $\left\|D\widehat{\theta}(x) - D\theta_0(x)\right\|_\infty$ term is bounded by $K\left\|\widehat{\theta}(x) - \theta_0(x)\right\|_{2,2}$. Therefore, it suffices to show the convergence rate of $\left\|\widehat{\theta}(x) - \theta_0(x)\right\|_{2,2}$.

Approximation rate in Sobolev norm is derived in Hornik et al. (1994). Convergence rate of the estimator in L_2 norm is derived in Chen and Shen (1998) and Chen and White (1999). We will combine their results to derive the convergence rate of the estimator in Sobolev norm. From the definition of criterion in B1, we have

$$E\left[l(Z_t, \theta) - l(Z_t, \theta_0)\right] = \frac{1}{2}\|\theta - \theta_0\|^2.$$

Since A3 implies the boundedness of the third derivatives, the equivalence of L_2 norm and Sobolev norm with second derivatives holds and there exist two constants c_1 and c_2 satisfying

$$c_1 \|\theta - \theta_0\|_{2,2}^2 \le E\left[l(Z_t, \theta) - l(Z_t, \theta_0)\right] \le c_2 \|\theta - \theta_0\|_{2,2}^2$$

which is required for Theorem 1 in Chen and Shen (1998). Further, Condition A1 in Chen and Shen can be replaced by our class of mixing condition in A1(a) which is shown by Chen and White (1999). Conditions A2 and A4 in Chen and Shen (Assumptions 3.4 (a) and (b) in Chen and White) follows from the proof of Proposition 1 in Chen and Shen. Therefore, from Theorem 1 of Chen and Shen (1998), we have

$$\left\|\widehat{\theta}_T-\theta_0\right\|_{2,2}=O_p\left(\max\left(\delta_T,\|\theta_0-\pi_T\theta_0\|_{2,2}\right)\right)$$

where $\pi_T\theta_0\in\Theta_T$ and

$$\delta_T=\inf\left\{\delta>0:\delta^{-2}\int_{\delta^2}^{\delta}[H(\varepsilon,\mathcal{F}_T)]^{1/2}\,d\varepsilon\leq const.\times n^{1/2}\right\}$$

where $H(\varepsilon,\mathcal{F}_T)$ is the L_2 metric entropy with bracketing which controls the size of the space of criterion differences induced by $\theta\in\Theta_T$ (See Chen and Shen, 1998, for the definition. Formally, the bracketing L_2 metric entropy of the space of the L_2 measurable functions indexed by Θ_T given by $\mathcal{F}_T=\{h(\theta,z)=l(\theta,z)-l(\theta_0,z):\theta\in\Theta_T\}$ is defined as follows. For any given ε, if there exists $S(\varepsilon,N)=\{h_1^l,h_1^u,,\ldots h_N^l,h_N^u\}\subset L_2$ with $\max_{1\leq j\leq N}\left\|h_j^u-h_j^l\right\|\leq\varepsilon$ such that for any $h\in\mathcal{F}_T$ there exists a j with $h_j^l\leq h\leq-h_j^u$ a.e., then $S(\varepsilon,N)$ is called a bracketing ε-covering of $\mathcal{F}_T$ with respect to $\|\cdot\|$. We define $H(\varepsilon,\mathcal{F}_T)$ by $\ln(\min\{N:S(\varepsilon,N)\})$.)

Using the result in the proof of Theorem 3.1 in Chen and White (1999), we have

$$H(\varepsilon,\mathcal{F}_T)\leq 2^k rB_T(d+1)\ln(2^k rB_T(d+1)/\varepsilon)$$

and

$$\delta_T=const.\times[r\ln(r)]^{1/2}\,T^{-1/2}.$$

From Hornik et al. (1994), the approximation rate in Sobolev W_2^2 norm is given by

$$\|\theta_0-\pi_T\theta_0\|_{2,2}\leq const.\times r^{-1/2}.$$

By choosing $\delta_T=\|\theta_0-\pi_T\theta_0\|_{2,2}$, we have

$$r^2\ln r=O(T)$$

and

$$\left\|\widehat{\theta}_T-\theta_0\right\|_{2,2}=O_p([T/\ln T]^{-1/4})$$

as required.

(b) For the multidimensional case, from Gabushin's interpolation inequality, we have

$$\left\|\Delta\widehat{\theta}_i(z)-\Delta\theta_{0i}(z)\right\|_{\infty}\leq K\left\|\Delta\widehat{\theta}_i(z)-\Delta\theta_{0i}(z)\right\|^{1/2}\left\|\Delta^2\widehat{\theta}_i(z)-\Delta^2\theta_{0i}(z)\right\|^{1/2}$$

for each $i=1,\ldots,d$ with $|\cdot|$ here being absolute value. If we use that

$$\sum_{i=1}^{d}\left\|\Delta\widehat{\theta}_i(z)-\Delta\theta_{0i}(z)\right\|\leq\left\|\widehat{\theta}(z)-\theta_0(z)\right\|_{1,2}\leq\left\|\widehat{\theta}(z)-\theta_0(z)\right\|_{2,2},$$

and

$$\sum_{i=1}^{d} \left\| \Delta^2 \widehat{\theta}_i(z) - \Delta^2 \theta_{0i}(z) \right\| \leq \left\| \widehat{\theta}(z) - \theta_0(z) \right\|_{2,2},$$

then,

$$\begin{aligned}
\sup_{z \in \mathcal{Z}} \left| \Delta\widehat{\theta}(z) - \Delta\theta_0(z) \right| &= \sup \sum_{i=1}^{d} \left| \Delta\widehat{\theta}_i(z) - \Delta\theta_{0i}(z) \right| \\
&\leq \sum_{i=1}^{d} \sup \left| \Delta\widehat{\theta}_i(z) - \Delta\theta_{0i}(z) \right| \\
&= \sum_{i=1}^{d} \left\| \Delta\widehat{\theta}_i(z) - \Delta\theta_{0i}(z) \right\|_{\infty} \\
&\leq K \sum_{i=1}^{d} \left(\left\| \Delta\widehat{\theta}_i(z) - \Delta\theta_{0i}(z) \right\|^{1/2} \left\| \Delta^2\widehat{\theta}_i(z) - \Delta^2\theta_{0i}(z) \right\|^{1/2} \right) \\
&\leq K \left(\sum_{i=1}^{d} \left\| \Delta\widehat{\theta}_i(z) - \Delta\theta_{0i}(z) \right\| \right)^{1/2} \left(\sum_{i=1}^{d} \left\| \Delta^2\widehat{\theta}_i(z) - \Delta^2\theta_{0i}(z) \right\| \right)^{1/2} \\
&\leq K \left\| \widehat{\theta}(z) - \theta_0(z) \right\|_{2,2}
\end{aligned}$$

where the second inequality follows from Cauchy-Schwarz's inequality. Therefore, it again suffices to show the convergence rate of $\left\| \widehat{\theta}(z) - \theta_0(z) \right\|_{2,2}$. Since the convergence rate of neural network estimator does not depend on d, the same argument for the one-dimensional case can be directly applied and the result follows. □

Proof of Lemma 2

As in the proof of Lemma 1 it suffices to show the convergence rate of $\left\| \widehat{\theta}(x) - \theta_0(x) \right\|_{2,2}$ for the one-dimensional case. Since additional assumption B4 is identical to assumption H in Chen and White, the result for the improved rate in Sobolev norm in Theorem 2.1 of Chen and White can be used. The improved approximation rate in Sobolev W_2^2 norm is now given by

$$\|\theta_0 - \pi_T\theta_0\|_{2,2} \leq const. \times r^{-1/2-\alpha/d^*}.$$

From

$$\delta_T = const. \times [r \ln(r)]^{1/2} T^{-1/2}$$

with choice of $\delta_T = \|\theta_0 - \pi_T\theta_0\|_{2,2}$, we have

$$r^{2(1+\alpha/d^*)} \ln r = O(T)$$

and

$$\left\| \widehat{\theta}_T - \theta_0 \right\|_{2,2} = O_p([T/\ln T]^{-\frac{1+(2\alpha/d^*)}{4(1+(\alpha/d^*))}}) = o_p(T^{-1/4})$$

as required. The same argument can be used for multidimensional case as in the proof of Lemma 1. □

Proof of Theorem 1

By rearranging terms,

$$\sqrt{M}(\widehat{\lambda}_M - \lambda) = \sqrt{M}(\widehat{\lambda}_M - \lambda_M) + \sqrt{M}(\lambda_M - \lambda).$$

For the second term, we have

$$\sqrt{M}(\lambda_M - \lambda) = \frac{1}{2\sqrt{M}} \sum_{t=1}^{M} \left[\ln(D\theta_0(x_{t-1}))^2 - 2\lambda \right] \Rightarrow N(0, \Phi)$$

by the central limit theorem (CLT) of Herrndorf (1984, Corollary 1) and A6*.

For the first term,

$$\begin{aligned}
\left| \sqrt{M}(\widehat{\lambda}_M - \lambda_M) \right| &= \left| \frac{1}{2\sqrt{M}} \sum_{t=1}^{M} \left[\ln(D\widehat{\theta}(x_{t-1}))^2 - \ln(D\theta_0(x_{t-1}))^2 \right] \right| \\
&= \left| \frac{1}{\sqrt{M}} \sum_{t=1}^{M} \frac{1}{D\theta^*(x_{t-1})} \left[D\widehat{\theta}(x_{t-1}) - D\theta_0(x_{t-1}) \right] \right| \\
&\leq [T/\ln T]^{-\frac{1}{4}} M^{\frac{1}{2}+\phi} \left[[T/\ln T]^{\frac{1}{4}} \sup_{x \in \chi} \left| D\widehat{\theta}(x) - D\theta_0(x) \right| \right] \\
&\quad \times \left(\frac{1}{M^{\phi} \min_{1 \leq t \leq M} |D\theta^*(x_{t-1})|} \right) = o_p(1)
\end{aligned}$$

where the second equality holds by a one-term Taylor expansion about $D\theta_0(x_{t-1})$ with $D\theta^*(x_{t-1})$ lying between $D\theta_0(x_{t-1})$ and $D\widehat{\theta}(x_{t-1})$. The convergence to zero holds because of $[T/\ln T]^{-\frac{1}{4}} M^{\frac{1}{2}+\phi} = o(1)$ from the growth rate of block length, uniform convergence from Lemma 1 and $\left(M^{\phi} \min_{1 \leq t \leq M} |D\theta^*(x_{t-1})|\right)^{-1} = O_p(1)$ from A5*, respectively. The latter can be verified by using the argument given in the proof of Theorem 1 in Whang and Linton (1999). □

Proof of Theorem 2

By rearranging terms,

$$\sqrt{M}(\widehat{\lambda}_{iM} - \lambda_i) = \sqrt{M}(\widehat{\lambda}_{iM} - \lambda_{iM}) + \sqrt{M}(\lambda_{iM} - \lambda_i)$$

where

$$\lambda_{iM} = \frac{1}{2M} \ln \nu_i \left((\Pi_{t=1}^{M} J_{M-t})' (\Pi_{t=1}^{M} J_{M-t}) \right).$$

For the second term, we have

$$\begin{aligned}
\sqrt{M}(\lambda_{iM} - \lambda_i) &= \sqrt{M} \left[\frac{1}{2M} \ln \nu_i \left((\Pi_{t=1}^{M} J_{M-t})' (\Pi_{t=1}^{M} J_{M-t}) \right) - \lambda_i \right] \\
&= \sqrt{M} \left[\frac{1}{2M} \ln \nu_i \left(\mathbf{T}_M' \mathbf{T}_M \right) - \lambda_i \right]
\end{aligned}$$

$$
\begin{aligned}
&= \sqrt{M}\left[\frac{1}{2M}\ln\left(\frac{\nu_i\left(\mathbf{T}_M'\mathbf{T}_M\right)}{\nu_i\left(\mathbf{T}_{M-1}'\mathbf{T}_{M-1}\right)}\right)+\frac{1}{2M}\ln\nu_i\left(\mathbf{T}_{M-1}'\mathbf{T}_{M-1}\right)-\lambda_i\right] \\
&= \sqrt{M}\left[\sum_{k=1}^{M-1}\frac{1}{2M}\ln\left(\frac{\nu_i\left(\mathbf{T}_{M-k+1}'\mathbf{T}_{M-k+1}\right)}{\nu_i\left(\mathbf{T}_{M-k}'\mathbf{T}_{M-k}\right)}\right)+\frac{1}{2M}\ln\nu_i\left(\mathbf{T}_1'\mathbf{T}_1\right)-\lambda_i\right] \\
&= \sqrt{M}\left[\frac{1}{M}\sum_{k=1}^{M}\xi_{i,M-k+1}-\lambda_i\right] \\
&= \frac{1}{\sqrt{M}}\sum_{t=1}^{M}\left[\xi_{it}-\lambda_i\right]\Rightarrow N(0,\Phi_i)
\end{aligned}
$$

by the CLT of Herrndorf (1984, Corollary 1) and results of Furstenberg and Kesten (1960, Theorem 3) and A6.

For the first term,

$$
\begin{aligned}
\left|\sqrt{M}(\widehat{\lambda}_i-\lambda_{iM})\right| &= \frac{1}{2\sqrt{M}}\left|\ln\nu_i\left((\Pi_{t=1}^M\widehat{J}_{M-t})'(\Pi_{t=1}^M\widehat{J}_{M-t})\right)-\ln\nu_i\left((\Pi_{t=1}^M J_{M-t})'(\Pi_{t=1}^M J_{M-t})\right)\right| \\
&= \left|\frac{1}{\sqrt{M}}\sum_{t=1}^{M}F_{i,t-1}(J_{M-1}^*,\ldots,J_0^*)'\left[\Delta\widehat{\theta}(Z_{t-1})-\Delta\theta_0(Z_{t-1})\right]\right| \\
&\leq [T/\ln T]^{-\frac{1}{4}}M^{\frac{1}{2}+\phi}\left[[T/\ln T]^{\frac{1}{4}}\sup_{z\in\mathcal{Z}}\left|\Delta\widehat{\theta}(z)-\Delta\theta_0(z)\right|\right] \\
&\quad\times M^{-\phi}\max_{1\leq t\leq M}\left|F_{i,t-1}(J_{M-1}^*,\ldots,J_0^*)\right| = o_p(1)
\end{aligned}
$$

where the second equality follows from a one-term Taylor expansion

$$
\begin{aligned}
&\ln\nu_i\left((\Pi_{t=1}^M\widehat{J}_{M-t})'(\Pi_{t=1}^M\widehat{J}_{M-t})\right) \\
= \;&\ln\nu_i\left((\Pi_{t=1}^M J_{M-t})'(\Pi_{t=1}^M J_{M-t})\right)+\frac{\partial\ln\nu_i\left((\Pi_{t=1}^M J_{M-t}^*)'(\Pi_{t=1}^M J_{M-t}^*)\right)}{\partial\Delta\theta_0(Z_{t-1})'}\left[\Delta\widehat{\theta}(Z_{t-1})-\Delta\theta_0(Z_{t-1})\right] \\
= \;&\ln\nu_i\left((\Pi_{t=1}^M J_{M-t})'(\Pi_{t=1}^M J_{M-t})\right)+F_{i,t-1}(J_{M-1}^*,\ldots,J_0^*)'\left[\Delta\widehat{\theta}(Z_{t-1})-\Delta\theta_0(Z_{t-1})\right]
\end{aligned}
$$

where the elements of J_t^* lie between those of $\widehat{J}_t$ and J_t for $t=0,\ldots,M-1$. Analogous to the proof of Theorem 1, the convergence to zero holds because of $[T/\ln T]^{-\frac{1}{4}}M^{\frac{1}{2}+\phi}=o(1)$ from the growth rate of block length, uniform convergence from Lemma 1 and $M^{-\phi}\max_{1\leq t\leq M}\left|F_{i,t-1}(J_{M-1}^*,\ldots,J_0^*)\right|=O_p(1)$ from A5, respectively. □

Proof of Corollary 1

We only prove the one-dimensional case since the multidimensional case can be obtained using the similar argument. First define

$$
\widetilde{\Phi}=\sum_{j=-M+1}^{M-1}w(j/S_M)\widetilde{\gamma}(j) \text{ and } \widetilde{\gamma}(j)=\frac{1}{M}\sum_{t=|j|+1}^{M}\eta_t\eta_{t-|j|}
$$

where $\eta_t = \ln|D\theta_0(x_{t-1})| - \lambda$. From Proposition 1 of Andrews (1991), $\widetilde{\Phi} \xrightarrow{p} \Phi$. Therefore, it suffices to show that $\widehat{\Phi} \xrightarrow{p} \widetilde{\Phi}$. Since $\sqrt{M}/S_M \to \infty$, the result follows by showing

$$\begin{aligned}
\frac{\sqrt{M}}{S_M}\left|\widehat{\Phi} - \widetilde{\Phi}\right| &= \frac{\sqrt{M}}{S_M}\left|\sum_{j=-M+1}^{M-1} w\left(\frac{j}{S_M}\right)\{\widehat{\gamma}(j) - \widetilde{\gamma}(j)\}\right| \\
&\leq \sqrt{M} \sup_{j\in[-M+1,M-1]} |\widehat{\gamma}(j) - \widetilde{\gamma}(j)| \left(\frac{1}{S_M}\sum_{j=-M+1}^{M-1}\left|w\left(\frac{j}{S_M}\right)\right|\right) = O_p(1).
\end{aligned}$$

The second element is bounded since $(1/S_M)\sum_{j=-M+1}^{M-1}|w(j/S_M)| \to \int_{-\infty}^{\infty}|w(x)|\,dx < \infty$. For the first element, we have

$$\begin{aligned}
\sqrt{M}\sup_j |\widehat{\gamma}(j) - \gamma(j)| &= \sqrt{M}\sup_j\left|\frac{1}{M}\sum_{t=|j|+1}^{M}\left(\widehat{\eta}_t\widehat{\eta}_{t-|j|} - \eta_t\eta_{t-|j|}\right)\right| \\
&\leq \sqrt{M}\sup_j\left|\frac{1}{M}\sum_{t=|j|+1}^{M}(\widehat{\eta}_t - \eta_t)\eta_{t-|j|}\right| + \sqrt{M}\sup_j\left|\frac{1}{M}\sum_{t=|j|+1}^{M}(\widehat{\eta}_{t-|j|} - \eta_{t-|j|})\eta_t\right| \\
&\quad + \sqrt{M}\sup_j\left|\frac{1}{M}\sum_{t=|j|+1}^{M}\left((\widehat{\eta}_t - \eta_t)(\widehat{\eta}_{t-|j|} - \eta_{t-|j|})\right)\right|.
\end{aligned}$$

Since

$$\begin{aligned}
&\frac{1}{M}\sum_{t=|j|+1}^{M}(\widehat{\eta}_t - \eta_t)\eta_{t-|j|} \\
&= \frac{1}{M}\sum_{t=|j|+1}^{M}\frac{1}{D\theta^*(x_{t-1})}\left[D\widehat{\theta}(x_{t-1}) - D\theta_0(x_{t-1})\right]\eta_{t-|j|} - \sqrt{M}(\widehat{\lambda}_M - \lambda)\frac{1}{\sqrt{M}}\frac{1}{M}\sum_{t=|j|+1}^{M}\eta_{t-|j|} \\
&\leq \left(\frac{\sup_{x\in\chi}\left|D\widehat{\theta}(x) - D\theta_0(x)\right|}{\min_{1\leq t\leq M}|D\theta^*(x_{t-1})|} + O_p(M^{-1/2})\right)\left[\frac{1}{M}\sum_{t=|j|+1}^{M}\eta_{t-|j|}^2\right]^{1/2}
\end{aligned}$$

we have

$$\begin{aligned}
&\sqrt{M}\sup_j\left|\frac{1}{M}\sum_{t=|j|+1}^{M}(\widehat{\eta}_t - \eta_t)\eta_{t-|j|}\right| \\
&\leq \left([T/\ln T]^{-\frac{1}{4}}M^{\frac{1}{2}+\phi}\frac{[T/\ln T]^{\frac{1}{4}}\sup_{x\in\chi}\left|D\widehat{\theta}(x) - D\theta_0(x)\right|}{M^{\phi}\min_{1\leq t\leq M}|D\theta^*(x_{t-1})|} + O_p(1)\right)\left[\frac{1}{M}\sum_{t=1}^{M}\eta_t^2\right]^{1/2} \\
&= O_p(1)
\end{aligned}$$

using $[T/\ln T]^{-\frac{1}{4}} M^{\frac{1}{2}+\phi} = o(1)$ from condition in Theorem 2, uniform convergence from Lemma 1 and $\left(M^{\phi} \min_{1\le t\le M} |D\theta^*(x_{t-1})|\right)^{-1} = O_p(1)$ from A5*, respectively. Boundedness for the other two terms can be obtained using the same argument. □

Proof of Corollary 2

Since the proof is similar to the one for Theorem 1(a) in Whang and Linton (1999), we only provide a sketch of the proof. By rearranging terms,

$$\sqrt{T}(\widehat{\lambda}_T - \lambda) = \sqrt{T}(\widehat{\lambda}_T - \lambda_T) + \sqrt{T}(\lambda_T - \lambda).$$

For the second term, we have asymptotics identical to those in Theorem 1. For the first term,

$$\begin{aligned}
&\sqrt{T}(\widehat{\lambda}_T - \lambda_T)\\
=\;& \frac{1}{\sqrt{T}}\sum_{t=1}^{T} \frac{1}{D\theta_0(x_{t-1})}\left[D\widehat{\theta}(x_{t-1}) - D\theta_0(x_{t-1})\right]\\
&-\frac{1}{2\sqrt{T}}\sum_{t=1}^{T} \frac{1}{[D\theta^*(x_{t-1})]^2}\left[D\widehat{\theta}(x_{t-1}) - D\theta_0(x_{t-1})\right]^2\\
=\;& \frac{1}{\sqrt{T}}\sum_{t=1}^{T} \frac{1}{D\theta_0(x_{t-1})}\left[D\widehat{\theta}(x_{t-1}) - D\theta_0(x_{t-1})\right] + o_p(1)\\
=\;& \sqrt{T}\int_{\chi} \frac{1}{D\theta_0(x)}\left[D\widehat{\theta}(x) - D\theta_0(x)\right] f(x)dx + o_p(1)\\
=\;& -\sqrt{T}\int_{\chi} \left[D\left(\frac{f(x)}{D\theta_0(x)}\right)\frac{1}{f(x)}\right]\left\{\widehat{\theta}(x) - \theta_0(x)\right\} f(x)dx + o_p(1)\\
=\;& \sqrt{T}\int_{\chi} v(x)\left\{\widehat{\theta}(x) - \theta_0(x)\right\} f(x)dx + o_p(1)\\
=\;& \sqrt{T}\left\langle v(x), \widehat{\theta} - \theta_0\right\rangle + o_p(1)\\
=\;& \frac{1}{\sqrt{T}}\sum_{t=1}^{T} l'_{\theta_0}[v(x), x_{t-1}] + o_p(1)\\
=\;& \frac{1}{\sqrt{T}}\sum_{t=1}^{T} v(x_{t-1})u_t + o_p(1).
\end{aligned}$$

The first equality follows from a two-term Taylor expansion about $D\theta_0(x_{t-1})$ with $D\theta^*(x_{t-1})$ lying between $D\theta_0(x_{t-1})$ and $D\widehat{\theta}(x_{t-1})$. The second equality follows from the fact that the second term is bounded by

$$\begin{aligned}
&\frac{1}{2}\left[T^{\frac{1}{4}}\sup_{x\in\chi}|D\widehat{\theta}(x) - D\theta_0(x)|\right]^2 \frac{1}{T}\sum_{t=1}^{T}\frac{1}{[D\theta^*(x_{t-1})]^2}\\
\le\;& o_p(1)\times\left(\frac{1}{\min_{1\le t\le T}|D\theta^*(x_{t-1})|}\right)^2 = o_p(1)
\end{aligned}$$

where the inequality follows by the uniform consistency results in Lemma 2, the last convergence to zero holds because $(\min_{1\leq t\leq T}|D\theta^*(x_{t-1})|)^{-2} = O_p(1)$ by A5*. The third equality follows from the stochastic equicontinuity argument employed in Whang and Linton (1999). The fourth equality follows from integration by parts with the zero boundary condition. The last three equalities follows from the definition of the linear functional $l'_{\theta_0}[\widehat{\theta}-\theta_0, x_{t-1}]$ and inner product $\langle .,.\rangle$ used in Shen (1997), Chen and Shen (1998) and Chen and White (1999), and

$$l'_{\theta_0}[\widehat{\theta}-\theta_0, x_{t-1}] = [\widehat{\theta}-\theta_0]u_t$$

from our criterion function given in A3(a). □

Proof of Corollary 3

We use a one-term Taylor expansion

$$\begin{aligned}
&\ln\nu_1\left(\widehat{J}'_{t-1}\widehat{J}_{t-1}\right)\\
=\;&\ln\nu_1\left(J'_{t-1}J_{t-1}\right) + \frac{\partial\nu_1\left(J^{*\prime}_{t-1}J^*_{t-1}\right)}{\partial\Delta\theta_0(Z_{t-1})'}\left[\Delta\widehat{\theta}(Z_{t-1}) - \Delta\theta_0(Z_{t-1})\right]\\
=\;&\ln\nu_1\left(J'_{t-1}J_{t-1}\right) + F'_{t-1}\left[\Delta\widehat{\theta}(Z_{t-1}) - \Delta\theta_0(Z_{t-1})\right]
\end{aligned}$$

where the elements of J^*_{t-1} lie between those of $\widehat{J}_{t-1}$ and J_{t-1}. The result follows from the argument similar (but simpler) to the one used in the proof of Theorem 2. □

References

Abhyankar, A., Copeland, L.S., Wong, W., 1997. Uncovering nonlinear structure in real-time stock-market indexes: The S&P 500, the DAX, the Nikkei 225, and the FTSE-100. Journal of Business and Economic Statistics 15, 1-14.

Andrews, D.W.K., 1991. Heteroskedasticity and autocorrelation consistent covariance matrix estimation. Econometrica 59, 817-858.

Bailey, B.A., 1996. Local Lyapunov exponents: Predictability depends on where you are. In: Barnett, W.A., Kirman, A.P., Salmon, M. (Eds.), Nonlinear Dynamics and Economics, Cambridge University Press, Cambridge, pp. 345-359.

Barnett, W.A., Gallant, A.R., Hinich, M.J., Jungeilges, J.A., Kaplan, D.T., Jensen, M.J., 1995. Robustness of nonlinearity and chaos tests to measurement error, inference method, and sample size. Journal of Economic Behavior and Organization 27, 301-320.

Barnett, W.A., Gallant, A.R., Hinich, M.J., Jungeilges, J.A., Kaplan, D.T., Jensen, M.J., 1997. A single-blind controlled competition among tests for nonlinearity and chaos. Journal of Econometrics 82, 157-192.

Barron, A.R., 1993. Universal approximation bounds for superpositions of a sigmoidal function. IEEE Transactions on Information Theory 39, 930-945.

Bask, M., de Luna, X., 2002. Characterizing the degree of stability of non-linear dynamic models. Studies in Nonlinear Dynamics and Econometrics 6, Article 3.

Brock, W.A., Dechert, W.D., Scheinkman, J.A., LeBaron, B., 1996. A test for independence based on the correlation dimension. Econometric Reviews 15, 197-235.

Brock, W.A., Hommes, C.H., 1998. Heterogeneous beliefs and routes to chaos in a simple asset pricing model. Journal of Economic Dynamics and Control 22, 1235-1274.

Chen, X., Racine, J., Swanson, N.R., 2001. Semiparametric ARX neural-network models with an application to forecasting inflation. IEEE Transactions on Neural Networks 12, 674-683.

Chen, X., Shen, X., 1998. Sieve extremum estimates for weakly dependent data. Econometrica 66, 289-314.

Chen, X., White, H., 1999. Improved rates and asymptotic normality for nonparametric neural network estimators. IEEE Transactions on Information Theory 45, 682-691.

Dechert, W.D., Gençay, R., 1992. Lyapunov exponents as a nonparametric diagnostic for stability analysis. Journal of Applied Econometrics 7, S41-S60.

Ding, Z., Granger, C.W.J., Engle, R.F., 1993. A long memory property of stock market returns and a new model. Journal of Empirical Finance 1, 83-106.

Eckmann, J.-P., Kamphorst, S.O., Ruelle, D., Ciliberto, S., 1986. Liapunov exponents from time series. Physical Review A 34, 4971-4979.

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

Ellner, S., Gallant, A.R., McCaffrey, D., Nychka, D., 1991. Convergence rates and data requirements for Jacobian-based estimates of Lyapunov exponents from data. Physics Letter A 153, 357-363.

Furstenberg, H., Kesten, H., 1960. Products of random matrices. The Annals of Mathematical Statistics 31, 457-469.

Gabushin, V. N., 1967. Inequalities for the norms of a function and its derivatives in metric Lp. Matematicheskie Zametki 1, 291-298.

Gallant, A.R., White, H., 1992. On learning the derivatives of an unknown mapping with multilayer feedforward networks. Neural Networks 5, 129-138.

Gençay, R., 1996. A statistical framework for testing chaotic dynamics via Lyapunov exponents. Physica D 89, 261-266.

Gençay, R., Dechert, W.D., 1992. An algorithm for the n Lyapunov exponents of an n-dimensional unknown dynamical system. Physica D 59, 142-157.

Gençay, R., Dechert, W.D., 1996. The identification of spurious Lyapunov exponents in Jacobian algorithms, Studies in Nonlinear Dynamics and Econometrics 1, 145-154.

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

Herrndorf, N., 1984. An invariance principle for weakly dependent sequences of random variables. Annals of Probability 12, 141-153.

Hornik, K., Stinchcombe, M., White, H., Auer, P., 1994. Degree of approximation results for feedforward networks approximating unknown mappings and their derivatives. Neural Computation 6, 1262-1275.

Kuan, C.-M., White, H., 1994. Artificial neural networks: An econometric perspective. Econometric Reviews 13, 1-91.

Lu, Z.-Q., Smith, R.L., 1997. Estimating local Lyapunov exponents. Fields Institute Communications 11, 135-151.

Makovoz, Y., 1996. Random approximants and neural networks. Journal of Approximation Theory 85, 98-109.

McCaffrey, D.F., 1991. Estimating Lyapunov exponents with nonparametric regression and convergence rates for feedforward single hidden layer networks. Ph.D. Dissertation. Department of Statistics. Raleigh, North Carolina State University.

McCaffrey, D.F., Ellner, S., Gallant, A.R., Nychka, D.W., 1992. Estimating the Lyapunov exponent of a chaotic system with nonparametric regression. Journal of the American Statistical Association 87, 682-695.

Nychka, D., Bailey, B., Ellner, S., Haaland, P., O'Connell, M., 1996. FUNFITS: data analysis and statistical tools for estimating functions. North Carolina Institute of Statistics Mimeoseries No. 2289.

Nychka, D., Ellner, S., Gallant, A.R., McCaffrey, D., 1992. Finding chaos in noisy system. Journal of the Royal Statistical Society, Series B 54, 399-426.

Serletis, A., 1995. Random walks, breaking trend functions, and the chaotic structure of the velocity of money. Journal of Business and Economic Statistics 13, 453-458.

Shen, X., 1997. On methods of sieves and penalization. Annals of Statistics 25, 2555-2591.

Shen, X., Wong, W.H., 1994. Convergence rate of sieve estimates. Annals of Statistics 22, 580-615.

Shintani, M., Linton, O., 2003. Is there chaos in the world economy? A nonparametric test using consistent standard errors. International Economic Review 44, 331-358.

Taylor, S.J., 1986. Modelling Financial Time Series. Wiley, New York.

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

Table 1
Logistic Map

(1) Stable System with $a = 1.5$ $(\lambda = -0.693)$						
	$d = 1$			$d = 2$		
	Full	Block	ES	Full	Block	ES
mean $(\widehat{\lambda})$	-0.729	-0.729	-0.729	-0.291	-0.283	-0.280
median $(\widehat{\lambda})$	-0.710	-0.705	-0.707	-0.276	-0.271	-0.271
std $(\widehat{\lambda})$	0.312	0.333	0.326	0.182	0.205	0.194
mean (se)	0.069	0.118	0.114	0.064	0.108	0.107
median (se)	0.068	0.114	0.112	0.062	0.105	0.104
lower 5%	0.060	0.060	0.060	0.050	0.060	0.060
upper 5%	0.040	0.040	0.040	0.040	0.030	0.040
	$d = 3$			$d = 4$		
	Full	Block	ES	Full	Block	ES
mean $(\widehat{\lambda})$	-0.101	-0.089	-0.082	0.009	0.022	0.027
median $(\widehat{\lambda})$	-0.091	-0.079	-0.076	0.014	0.027	0.031
std $(\widehat{\lambda})$	0.124	0.147	0.140	0.094	0.116	0.111
mean (se)	0.054	0.093	0.091	0.048	0.082	0.081
median (se)	0.053	0.090	0.089	0.047	0.080	0.079
lower 5%	0.060	0.060	0.060	0.060	0.060	0.050
upper 5%	0.030	0.030	0.040	0.040	0.040	0.050
(2) Chaotic System with $a = 4$ $(\lambda = 0.693)$						
	$d = 1$			$d = 2$		
	Full	Block	ES	Full	Block	ES
mean $(\widehat{\lambda})$	0.689	0.689	0.689	0.664	0.667	0.669
median $(\widehat{\lambda})$	0.689	0.691	0.691	0.679	0.681	0.674
std $(\widehat{\lambda})$	0.019	0.031	0.100	0.059	0.066	0.112
mean (se)	0.054	0.092	0.102	0.051	0.087	0.098
median (se)	0.053	0.090	0.101	0.050	0.085	0.097
lower 5%	0.050	0.050	0.050	0.070	0.070	0.060
upper 5%	0.050	0.040	0.050	0.000	0.010	0.040
	$d = 3$			$d = 4$		
	Full	Block	ES	Full	Block	ES
mean $(\widehat{\lambda})$	0.662	0.666	0.668	0.662	0.667	0.669
median $(\widehat{\lambda})$	0.673	0.676	0.675	0.670	0.675	0.671
std $(\widehat{\lambda})$	0.054	0.061	0.112	0.046	0.054	0.107
mean (se)	0.050	0.086	0.098	0.050	0.086	0.097
median (se)	0.050	0.085	0.097	0.050	0.085	0.097
lower 5%	0.050	0.060	0.060	0.050	0.050	0.050
upper 5%	0.000	0.010	0.040	0.000	0.010	0.040

Note: Sample size $(T) = 200$. Number of hidden units $(r) = 4$. Number of replications $= 1000$. Jacobians are evaluated using full sample (Full) as well as blocks (Block) and equally spaced subsamples (ES) with block length $(M) = 66$. Lower 5% and upper 5% are tail frequencies of normalized Lyapunov exponent estimates using standard normal critical values.

Table 2
Barnett Competition Data

$T = 380$				$T = 2000$			
	Sample				Sample		
(d, r)	Full	Block	ES	(d, r)	Full	Block	ES
			(I) Logistic map				
$(2, 4)$	0.015	0.019	0.028	$(1, 4)$	0.012	0.008	0.014
	(0.396)	(0.207)	(0.442)		(1.190)	(0.119)	(0.210)
	[0.654]	[0.582]	[0.671]		[0.883]	[0.547]	[0.583]
			(II) GARCH				
$(1, 1)$	-4.260	-4.219	-4.323	$(1, 1)$	-5.017	-5.034	-5.043
	(-56.00)	(-26.20)	(-24.58)		(-215.1)	(-49.09)	(-45.87)
	[<0.001]	[<0.001]	[<0.001]		[<0.001]	[<0.001]	[<0.001]
			(III) NLMA				
$(2, 3)$	-0.435	-0.400	-0.430	$(3, 4)$	-0.360	-0.354	-0.323
	(-15.66)	(-6.345)	(-7.371)		(-43.93)	(-8.792)	(-8.145)
	[<0.001]	[<0.001]	[<0.001]		[<0.001]	[<0.001]	[<0.001]
			(IV) ARCH				
$(1, 1)$	-3.925	-3.875	-3.939	$(1, 1)$	-3.606	-3.607	-3.606
	(-69.56)	(-28.76)	(-31.18)		(-1324)	(-302.5)	(-278.1)
	[<0.001]	[<0.001]	[<0.001]		[<0.001]	[<0.001]	[<0.001]
			(V) ARMA				
$(1, 1)$	-0.049	-0.048	-0.051	$(3, 1)$	-0.041	-0.028	-0.034
	(-4.843)	(-3.832)	(-4.659)		(-8.116)	(-2.496)	(-3.559)
	[<0.001]	[<0.001]	[<0.001]		[<0.001]	[0.006]	[<0.001]

Note: For the full sample estimation (Full), the largest Lyapunov exponent estimates are presented with t statistics in parentheses and p-value for $H_0 : \lambda \geq 0$ in brackets. For the estimation based on blocks (Block) and equally spaced subsamples (ES), median values are presented. The block length (M) for subsample is 72 for $T = 380$ and 91 for $T = 2000$, respectively. The lag length (d) and the number of hidden units (r) are jointly selected based on BIC. QS kernel with optimal bandwidth (Andrews, 1991) is used for the heterosckedasticity and autocorrelation consistent covariance estimation.

Table 3
Autocorrelations of Stock Return Series

x_t	$\widehat{\rho}(1)$	$\widehat{\rho}(2)$	$\widehat{\rho}(3)$	$\widehat{\rho}(4)$	$\widehat{\rho}(5)$	$\widehat{\rho}(10)$
(1) R_t	0.029	-0.022	0.005	0.018	0.019	0.007
	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)
(2) $\|R_t\|^{0.5}$	0.233	0.242	0.245	0.251	0.260	0.236
	(0.007)	(0.007)	(0.008)	(0.008)	(0.008)	(0.010)
(3) $\|R_t\|^{1.0}$	0.295	0.314	0.308	0.300	0.311	0.266
	(0.007)	(0.007)	(0.008)	(0.009)	(0.009)	(0.011)
(4) $\|R_t\|^{1.5}$	0.280	0.294	0.269	0.243	0.271	0.198
	(0.007)	(0.007)	(0.008)	(0.008)	(0.009)	(0.010)
(5) $\|R_t\|^{2.0}$	0.202	0.211	0.160	0.131	0.177	0.095
	(0.007)	(0.007)	(0.007)	(0.008)	(0.008)	(0.008)
(6) $\|R_t\|^{2.5}$	0.117	0.129	0.072	0.054	0.098	0.034
	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)

Note: Numbers in parentheses are standard errors.

Table 4
Lyapunov Exponents of Stock Return Series

	NLAR lag (d)					
	1	2	3	4	5	6
			(1) $x_t = R_t$			
Full	-2.685	-1.539	-1.355	-0.820	-0.562	-0.503
	(-262.1)	(-347.7)	(-721.6)	(-228.5)	(-322.7)	(-455.81)
	[<0.001]	[<0.001]	[<0.001]	[<0.001]	[<0.001]	[<0.001]
Block	-2.689	-1.538	-1.339	-0.800	-0.546	-0.487
	(-24.31)	(-30.49)	(-44.93)	(-18.21)	(-13.36)	(-14.70)
	[<0.001]	[<0.001]	[<0.001]	[<0.001]	[<0.001]	[<0.001]
ES	-2.684	-1.540	-1.330	-0.799	-0.541	-0.490
	(-23.62)	(-30.35)	(-45.40)	(-17.64)	(-13.40)	(-14.71)
	[<0.001]	[<0.001]	[<0.001]	[<0.001]	[<0.001]	[<0.001]
BIC	-8.944(2)	-8.953(2)	-8.951(3)	-8.953(2)	-8.949(3)	-8.958(3)
			(2) $x_t = \|R_t\|^{0.5}$			
Full	-1.876	-0.985	-0.568	-0.364	-0.260	-0.194
	(-306.9)	(-189.7)	(-191.3)	(-130.3)	(-113.8)	(-129.7)
	[<0.001]	[<0.001]	[<0.001]	[<0.001]	[<0.001]	[<0.001]
Block	-1.921	-1.017	-0.582	-0.372	-0.264	-0.195
	(-49.65)	(-24.24)	(-20.99)	(-18.01)	(-16.27)	(-16.29)
	[<0.001]	[<0.001]	[<0.001]	[<0.001]	[<0.001]	[<0.001]
ES	-1.874	-0.960	-0.549	-0.352	-0.250	-0.188
	(-38.53)	(-19.61)	(-19.05)	(-16.51)	(-14.27)	(-14.77)
	[<0.001]	[<0.001]	[<0.001]	[<0.001]	[<0.001]	[<0.001]
BIC(r)	-6.459(1)	-6.508(2)	-6.536(3)	-6.554(3)	-6.572(3)	-6.576(3)
			(3) $x_t = \|R_t\|^{1.0}$			
Full	-1.424	-0.677	-0.476	-0.304	-0.211	-0.173
	(-939.3)	(-233.6)	(-153.1)	(-220.5)	(-177.8)	(-180.2)
	[<0.001]	[<0.001]	[<0.001]	[<0.001]	[<0.001]	[<0.001]
Block	-1.437	-0.693	-0.488	-0.308	-0.213	-0.173
	(-209.2)	(-41.18)	(-25.88)	(-27.07)	(-22.45)	(-19.80)
	[<0.001]	[<0.001]	[<0.001]	[<0.001]	[<0.001]	[<0.001]
ES	-1.424	-0.669	-0.460	-0.298	-0.204	-0.166
	(-128.2)	(-36.19)	(-23.27)	(-24.52)	(-21.08)	(-20.04)
	[<0.001]	[<0.001]	[<0.001]	[<0.001]	[<0.001]	[<0.001]
BIC(r)	-9.554(1)	-9.619(2)	-9.660(3)	-9.688(3)	-9.711(3)	-9.716(3)

Table 4 (Continued)

	NLAR lag (d)					
	1	2	3	4	5	6
			(4) $x_t = \|R_t\|^{1.5}$			
Full	-1.196	-0.452	-0.216	-0.136	-0.071	-0.111
	(-2056)	(-525.0)	(-804.9)	(-329.5)	(-75.29)	(-110.4)
	[<0.001]	[<0.001]	[<0.001]	[<0.001]	[<0.001]	[<0.001]
Block	-1.196	-0.454	-0.216	-0.131	-0.060	-0.114
	(-311.0)	(-66.31)	(-88.06)	(-51.85)	(-14.79)	(-17.93)
	[<0.001]	[<0.001]	[<0.001]	[<0.001]	[<0.001]	[<0.001]
ES	-1.195	-0.449	-0.215	-0.135	-0.066	-0.108
	(-203.9)	(-62.53)	(-48.19)	(-31.11)	(-8.660)	(-16.64)
	[<0.001]	[<0.001]	[<0.001]	[<0.001]	[<0.001]	[<0.001]
BIC(r)	-12.33(3)	-12.38(2)	-12.42(3)	-12.45(3)	-12.46(3)	-12.47(3)
			(5) $x_t = \|R_t\|^{2.0}$			
Full	-1.218	-0.111	-0.018	-0.014	-0.123	-0.088
	(-909.6)	(-38.94)	(-13.24)	(-22.28)	(-104.3)	(-106.5)
	[<0.001]	[<0.001]	[<0.001]	[<0.001]	[<0.001]	[<0.001]
Block	-1.232	-0.088	-0.005	-0.009	-0.129	-0.090
	(-148.7)	(-13.10)	(-1.943)	(-3.994)	(-32.84)	(-21.73)
	[<0.001]	[<0.001]	[0.026]	[<0.001]	[<0.001]	[<0.001]
ES	-1.220	-0.108	-0.015	-0.013	-0.124	-0.086
	(-102.2)	(-6.911)	(-2.159)	(-2.974)	(-23.80)	(-15.65)
	[<0.001]	[<0.001]	[0.015]	[0.001]	[<0.001]	[<0.001]
BIC(r)	-14.53(2)	-14.56(2)	-14.59(3)	-14.63(3)	-14.68(3)	-14.65(3)
			(6) $x_t = \|R_t\|^{2.5}$			
Full	-0.040	0.078	-0.172	0.087	-0.380	-0.292
	(-13.14)	(23.99)	(-160.6)	(67.01)	(-126.6)	(-68.38)
	[<0.001]	[1.000]	[<0.001]	[1.000]	[<0.001]	[<0.001]
Block	-0.008	0.103	-0.180	0.093	-0.407	-0.328
	(-1.085)	(20.53)	(-93.92)	(25.69)	(-36.20)	(-20.15)
	[0.139]	[1.000]	[<0.001]	[1.000]	[<0.001]	[<0.001]
ES	-0.039	0.082	-0.170	0.089	-0.375	-0.269
	(-1.918)	(4.918)	(-14.63)	(8.333)	(-15.49)	(-7.245)
	[0.028]	[1.000]	[<0.001]	[1.000]	[<0.001]	[<0.001]
BIC(r)	-16.30(3)	-16.31(3)	-16.34(3)	-16.38(3)	-17.45(3)	-16.46(2)

Note: For the full sample estimation (Full), the largest Lyapunov exponent estimates are presented with t statistics in parentheses and p-value for $H_0 : \lambda \geq 0$ in brackets. For the estimation based on blocks (Block) and equally spaced subsamples (ES), median values are presented. The block length (M) for subsample is 127. For each lag (d), the number of hidden units (r) are selected based on BIC. QS kernel with optimal bandwidth (Andrews, 1991) is used for the heterosckedasticity and autocorrelation consistent covariance estimation.

Figure 1

Logistic Map (a = 1.5, d = 1, T = 200)

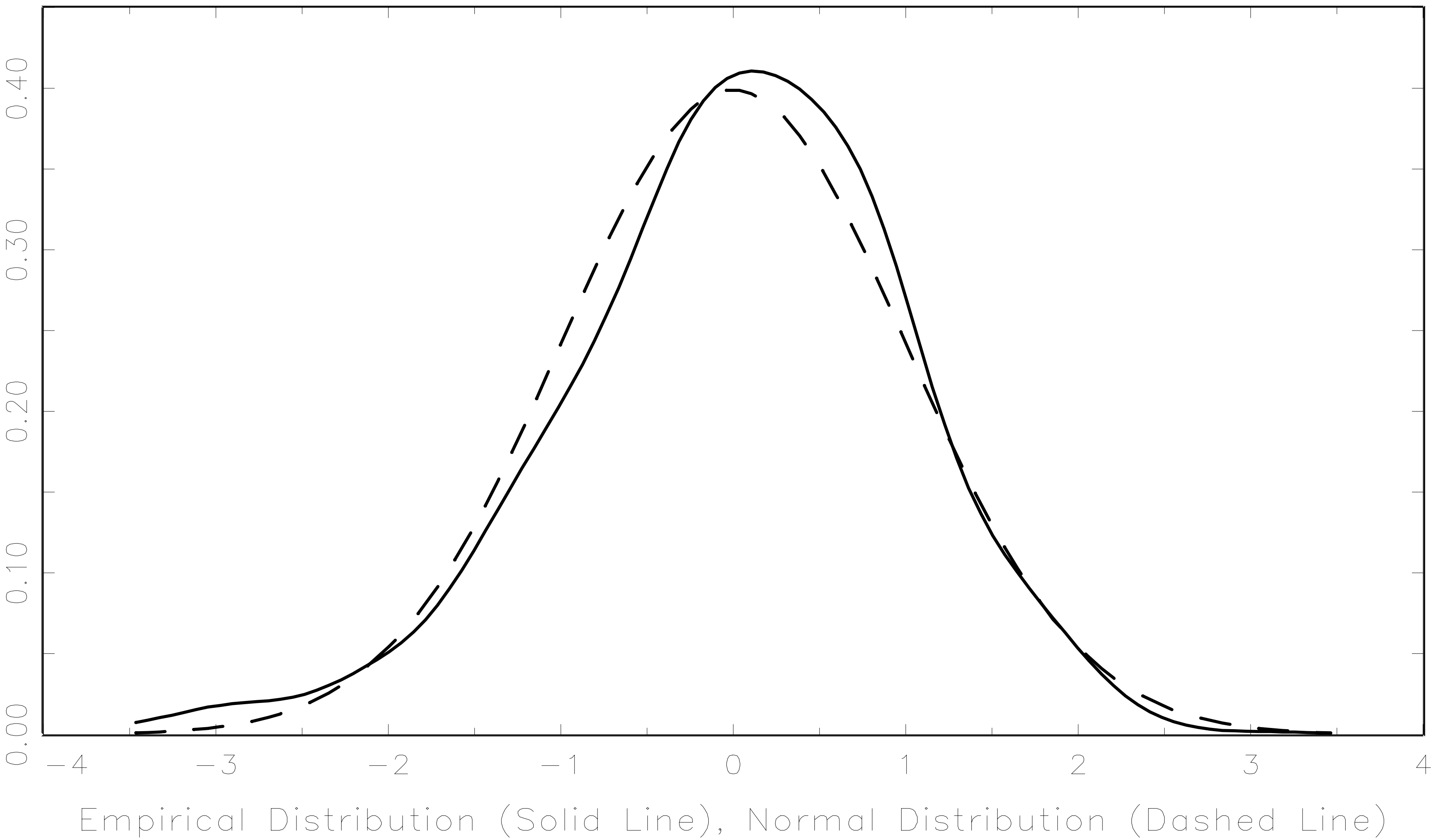

Figure 2
Logistic Map (a = 4, d = 1, T = 200)