ELSEVIER

Journal of Econometrics 91 (1999) 1–42

JOURNAL OF
Econometrics

The asymptotic distribution of nonparametric estimates of the Lyapunov exponent for stochastic time series

Yoon-Jae Whang[a], Oliver Linton[b,*]

[a] *Department of Economics, Ewha University, Seoul, 120-750, South Korea*
[b] *Department of Economics, Yale University, Yale Station Box 208281, New Haven, CT 06520, USA*

Received 1 November 1996; received in revised form 1 March 1998

Abstract

This paper derives the asymptotic distribution of a smoothing-based estimator of the Lyapunov exponent for a stochastic time series under two general scenarios. In the first case, we are able to establish root-T consistency and asymptotic normality, while in the second case, which is more relevant for chaotic processes, we are only able to establish asymptotic normality at a slower rate of convergence. We provide consistent confidence intervals for both cases. We apply our procedures to simulated data. © 1999 Elsevier Science S.A. All rights reserved.

JEL classification: C13, C14, C22

Keywords: Chaos; Kernel; Nonlinear dynamics; Nonparametric regression; Semiparametric

1. Introduction

There has been much interest in nonlinear chaotic dynamics in economics following the early work of Brock (1986). That deterministic systems can lead to very complex and essentially unpredictable behaviour is an important finding for economics and raises many philosophical as well as procedural questions. See the recent special issues of the Journal of Applied Econometrics (1994) (Eds. Pesaran and Potter) and the Journal of the Royal Statistical Society,

* Corresponding author. E-mail: linton@econ.yale.edu; http: \\www.econ.yale.edu\~ linton

PII: S0304-4076(98)00047-5

Series B (1992) for excellent discussion, and LeBaron (1997) for a recent review of applications in social sciences. We address purely the practical question about how to apply statistical inference to a commonly used measure of stability when the data are generated by a stochastic mechanism.

The Lyapunov exponent provides some information on the stability (i.e., chaotic) properties of a dynamic system. Indeed, Nychka et al. (1992) reported that

> A bounded system with a positive Lyapunov exponent is one operational definition of chaotic behaviour.

Nychka et al. (1992) suggest a method, improving on an idea of Eckmann et al. (1986), of testing for chaos based on estimating the Lyapunov exponent λ using nonparametric regression methods (see also McCaffrey et al., 1992). Of course, for purely chaotic series there is no way of evaluating the uncertainty surrounding that point estimate. Many authors, however, have advocated including stochastic disturbances along with the chaotic 'skeleton' to reflect measurement error, unobservable quantities, and so on (Cheng and Tong, 1992). With this addition, the usual rules of inference can in principle be applied. McCaffrey et al. (1992) established the consistency of their nonparametric estimates of λ for a stochastic system; however, they did not establish the asymptotic distribution.

In this paper, we examine the asymptotic properties of a smoothing-based estimate of the Lyapunov exponent for stochastic time series. Our focus is primarily on kernel estimates, because of their theoretical tractability, but some of our results require less detailed specification of the smoothing method used – in fact, only knowledge of its uniform rate of convergence is needed. We establish the asymptotic distribution of our estimator of λ under two sets of regularity conditions. The first set of conditions could be considered fairly standard for stochastic time series models, at least in the one-dimensional case. In this case, we obtain root-T consistency (where T is the full sample size) and asymptotic normality. Unfortunately, these conditions exclude many chaotic processes (as defined below) because they essentially exclude the first derivative of the regression function from passing through the origin, a feature which must hold for a bounded but explosive process. We, therefore, give a second set of regularity conditions that are considerably weaker in this regard and in fact can be verified for the univariate Feingenbaum mapping, a standard example of a chaotic process. The price we have to pay is that we can obtain only root-n convergence, for a procedure that averages over only a subsample of size n (the magnitude of n is determined not only by the rate of convergence of the nonparametric estimation but also by the properties of the process near the interior maxima or minima of the regression function). The subsample results require only that the smoothing procedure converge in probability uniformly with a certain rate, and thus can be applied more generally to the neural net and spline methods used elsewhere in for example Nychka et al. (1992) and Gençay and Dechert (1992) . We give results both for the univariate and multivariate case, although an additional complication occurs in the proofs for the multivariate case, which necessitates less transparent regularity conditions.

As in other semiparametric estimation problems, we expect that the first-order asymptotic variance of the estimate is independent of the smoothing method, see Newey (1994), and that choice of smoothing method is only a second-order issue. Nevertheless, it can be extremely important in practice, especially when the full sample is used.

The Lyapunov exponent can be interpreted within the standard nonlinear time series framework as a measure of local stability, see Tong (1990, p. 309), and is of interest even outside of any direct connection with deterministic chaos theory per se. We would like to make clear that our work does not bear on how λ should be interpreted. We merely provide rules of inference that apply under certain probabilistic regularity conditions. Whether chaotic dynamics exist in economics data is an empirical question, which we hope our methods can help answer.

Notation. Let $\mu = (\mu_1, \dots, \mu_k)'$ denote a k-vector of non-negative integer constants. For such a vector, define $|\mu| = \sum_{i=1}^{k} \mu_i$, $z^\mu = \prod_{i=1}^{k} z_i^{\mu_i}$ for $z = (z_1, \dots, z_k)' \in \mathbb{R}^k$ and

$$D^\mu c(z) = \frac{\partial^{|\mu|} c(z)}{\partial z_1^{\mu_1}, \dots, \partial z_k^{\mu_k}},$$

for any real function $c(z)$ on $\mathbb{R}^k$. When μ denotes a scalar constant, as is the case when $k = 1$, we define $D^\mu c(z)$ to be the μth order derivative of $c(\cdot)$ evaluated at z with the convention that $D^0 c(z) = c(z)$ and $D^1 c(z) = Dc(z)$.

2. Estimation

Consider the following difference equation

$$X_t = m_0(X_{t-1}, \dots, X_{t-k}), \quad t = 1, \dots, T, \tag{1}$$

where $m_0 : \mathbb{R}^k \to \mathbb{R}$ is a nonlinear dynamic mapping. The model (1) can be expressed in terms of a state vector $Z_t = (X_t, \dots, X_{t-k+1})' \in \mathbb{R}^k$, and a function $M: \mathbb{R}^k \to \mathbb{R}^k$ such that

$$Z_t = M(Z_{t-1}) \quad \text{for } t = 1, \dots, T. \tag{2}$$

Let J_t be the Jacobian of the map M in Eq. (2) evaluated at Z_t. Specifically, we define

$$J_t = \begin{bmatrix} \Delta m_{1t} & \Delta m_{2t} & \cdots & \Delta m_{k-1,t} & \Delta m_{kt} \\ 1 & 0 & \cdots & 0 & 0 \\ 0 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & 0 \end{bmatrix} \tag{3}$$

for $t = 0, 1, \ldots, T-1$, where $\Delta m_{jt} = D^{e_j} m_0(Z_t)$ for $j = 1, \ldots, k$ in which $e_j = (0, \ldots, 1, \ldots, 0)' \in \mathbb{R}^k$ denotes the jth elementary vector. The dominant Lyapunov exponent is defined as

$$\lambda \overset{\text{a.s.}}{=} \lim_{T \to \infty} \frac{1}{T} \ln \left\| \prod_{t=1}^{T} J_{T-t} \right\|, \tag{4}$$

where $\|\cdot\|$ is any matrix norm. This definition is now standard in the literature, see Nychka et al. (1992). When $\lambda > 0$, the series appears to the eye to be essentially random, and is very sensitive to initial conditions – even a very tiny difference in initial conditions lead to very large differences in the subsequent trajectory.

In earlier work this definition was applied to purely deterministic systems and is still used in this way for much work in the physical sciences. For many time series, especially perhaps those of interest to economists, it may be desirable to include a stochastic disturbance because of measurement error, unobserved quantities and so on. Thus we consider the extension to allow for error as in Cheng and Tong (1992), for example. In the first case, called system noise, we have

$$X_t = m_0(X_{t-1}, \ldots, X_{t-k}) + \varepsilon_t, \quad t = 1, \ldots, T, \tag{5}$$

where $\{\varepsilon_t\}$ is a sequence of random variables with mean zero and variance σ^2, and m_0 is now an unspecified regression function. Model (5) can be expressed in state space form as in Eq. (2) with the addition of the error vector $U_t = (\varepsilon_t, 0, \ldots, 0)' \in \mathbb{R}^k$. An alternative model here, called measurement noise, is that X_t is defined in Eq. (1), but that we observe

$$Y_t = X_t + \varepsilon_t, \quad t = 1, \ldots, T,$$

with ε_t as in Eq. (5).[1] In this case, we can find an m_0^* for which Eq. (5) holds with X_s replaced by the observed data Y_s, so there is not really any loss of generality, for our asymptotic results, in restricting attention to Eq. (5). Furthermore, Nychka et al. (1992) argue in favour of the system noise specification for economics data, and we therefore concentrate on system noise in the sequel.

The stochastic system can exhibit chaotic behaviour, i.e., sensitive dependence to initial conditions; obviously so when $\sigma \to 0$, but true even for moderate σ.[2] For stationary linear autoregressions $\lambda < 0$, while for the unit root process $\lambda = 0$. For explosive autoregressions $\lambda > 0$, but these series torque off to infinity. Chaotic series are bounded processes with $\lambda > 0$. Their explosiveness is local in nature, and the series is kept bounded and even stationary.

Let $\hat{m}(\cdot)$ be a consistent smoothing-based estimate of $m_0(\cdot)$. Suitable estimation methods include: kernels, nearest neighbours, splines, series, local polynomials, neural nets, see Härdle and Linton (1994) for a general discussion, and Nychka et al. (1992) and McCaffrey et al. (1992) for considerable practical advice on how best to implement some of these methods for the specific purpose of estimating λ. We now define our estimator of λ. We first select a subsample of values at which to evaluate $\hat{m}(\cdot)$ (the estimate itself will be computed with the full sample). We assume that Z_t is supported on a bounded set $\mathscr{Z}$ and that $\partial\mathscr{Z}$ is its topological boundary, and trim out observations too close to this boundary; specifically, we keep only the T^* sample values that satisfy

$$\inf_{z_0 \in \partial\mathscr{Z}} d(Z_t, z_0) \geqslant d_T, \tag{6}$$

for some data-dependent distance function $d(.,.)$ and trimming sequence $d_T \geqslant 0$. From this trimmed sample we select a further subsample of evaluation points of size $n(T^*) \leqslant T^*$, and for notational simplicity we let these observations be Z_t, $t = 1, \ldots, n$.[3] We now let

$$\hat{\lambda} = \frac{1}{n} \ln \left\| \prod_{t=1}^{n} \hat{J}_{n-t} \right\|, \tag{7}$$

[1] See Gençay and Dechert (1992) for a slightly more general set-up.

[2] As $\sigma \to 0$, $\lambda_\sigma \to \lambda_0$.

[3] Under our assumptions, we have $T^*/T \to 1$ and $n(T^*)/n(T) \to 1$; we will not distinguish between quantities based on T and those based on T^* unless necessary.

where

$$\hat{J}_t = \begin{bmatrix} \Delta\hat{m}_{1t} & \Delta\hat{m}_{2t} & \cdots & \Delta\hat{m}_{k-1,t} & \Delta\hat{m}_{kt} \\ 1 & 0 & \cdots & 0 & 0 \\ 0 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & 0 \end{bmatrix}, \tag{8}$$

where $\Delta\hat{m}_{jt} = D^{e_j}\hat{m}(Z_t)$ for $t = 0, 1, \dots, n-1$. This estimator can generally be computed fairly cheaply for moderate k and small n. For notational convenience we have just taken the first n observations, but more general 'blocking' schemes (for selecting a subsample of size n) could be employed; in fact, it is generally advisable to take equally spaced observations. See McCaffrey et al. (1992) for a discussion on the 'optimal' choice of n in practice.

For some of our results we use a kernel estimator $\hat{m}(\cdot)$ of $m_0(\cdot)$, which is defined as follows:

$$\hat{m}(z) = \hat{g}(z)/\hat{f}(z), \tag{9}$$

where

$$\hat{g}(z) = \frac{1}{T\hat{b}_T^k}\sum_{t=1}^{T} \hat{K}\left(\frac{z - Z_{t-1}}{\hat{b}_T}\right)X_t; \qquad \hat{f}(z) = \frac{1}{T\hat{b}_T^k}\sum_{t=1}^{T} \hat{K}\left(\frac{z - Z_{t-1}}{\hat{b}_T}\right),$$

$$\hat{K}(z) = [\det(\hat{\Omega})]^{-1/2}K(\hat{\Omega}^{-1/2}z); \qquad \hat{\Omega} = \frac{1}{T}\sum_{t=1}^{T} (Z_{t-1} - \bar{Z}_T)(Z_{t-1} - \bar{Z}_T)',$$

$$\bar{Z}_T = \frac{1}{T}\sum_{t=1}^{T} Z_{t-1}. \tag{10}$$

Derivatives of $\hat{m}(z)$ and $\hat{f}(z)$ provide suitable estimates of the corresponding derivatives of $m(z)$ and $f(z)$. The kernel $K(\cdot)$ is a nonrandom real function on $\mathbb{R}^k$, and the bandwidth parameter $\hat{b}_T$ is a positive constant or scalar random variable satisfying the assumptions below. The data-dependent scale matrix $\hat{\Omega}$ is used to obtain invariance of the estimator under one-to-one transformations of the form $Z_{t-1} \to BZ_{t-1} + \gamma$, where B is a nonsingular $k \times k$ matrix and γ is a k-vector, see Bierens (1987) for further motivation. We use the class of kernels, denoted $\mathscr{K}_{k,\delta,v}$ defined for non-negative integers, k, δ, and v, introduced in Andrews (1995). A complete definition is given in the appendix, and here it suffices to note that k is the dimensionality, the vector δ is the order of derivative being estimated, and v ($\geqslant \delta$) the number of zero moments of K, i.e., its order.

See Andrews (1995) for more discussion and examples of these kernels. The main reason for using this particular smoothing method is theoretical, some useful results we need are only available for this method, as far as we are aware.

3. Asymptotic properties

3.1. The case $k = 1$

Our main result is for the scalar case, $k = 1$. The asymptotic behavior of $\hat{\lambda}$ for general $k \geqslant 2$ will be discussed later.

Note that when $k = 1$ we have $Z_t = X_t$, $J_t = Dm_0(X_t)$ (defined in Eq. (3)), and $\hat{J}_t = D\hat{m}(X_t)$ (defined in Eq. (8)) for all t. In this case, we replace $\mathscr{Z} = \mathscr{X}$, and take $d(x, y) = |x - y|$. Hence expressions (4) and (7) of Section 2 simplify as follows:

$$\lambda \overset{\text{a.s.}}{=} \lim_{T \to \infty} \frac{1}{2T} \sum_{t=1}^{T} \ln\,[Dm_0(X_{t-1})^2], \tag{11}$$

$$\hat{\lambda} = \frac{1}{2n} \sum_{t=1}^{n} \ln[D\hat{m}(X_{t-1})^2]\mathbf{1}_t, \tag{12}$$

where $\mathbf{1}_t = 1$ if $\inf_{x_0 \in \partial \mathscr{X}} |X_{t-1} - x_0| \geqslant d_T$, and $\mathbf{1}_t = 0$ otherwise.[4] This representation of λ and $\hat{\lambda}$ emphasizes the mathematical connection with the problem of estimating average derivatives (Härdle and Stoker, 1989).

When we take $n = T$, we impose the following assumptions:

Assumption A

(1) (a) $\{X_t\colon t \geqslant 1\}$ is a sequence of strictly stationary strong mixing random variables with mixing numbers of size $-2r/(r-2)$ for some $r > 2$.[5]
(b) For all $t \geqslant 1$, X_t lies in an open bounded set $\mathscr{X}$ ($\subset \mathbb{R}$) with minimally smooth boundary denoted $\partial \mathscr{X}$.
(c) $\{\varepsilon_t\colon t \geqslant 1\}$ is a sequence of martingale differences with $E(\varepsilon_t \mid \mathscr{F}_{-\infty}^{t-1}) = 0$ a.s. and $E(|\varepsilon_t|^{2r}) < \infty$ for all $t \geqslant 1$, where $\mathscr{F}_s^t = \sigma(X_s, \ldots, X_t)$ is the σ-algebra generated by $(X_s, \ldots, X_t)$.

---

[4] Note also that

$$\lambda \overset{\text{a.s.}}{=} \lim_{T \to \infty} 1/2T \sum_{t=1}^{T} \ln[Dm_0(X_{t-1})^2]\mathbf{1}_t,$$

i.e., trimming has an asymptotically negligible effect.

[5] See Appendix for the definition of strong mixing random variables.

(2) (a) The distribution of X_t is absolutely continuous with respect to Lebesgue measure with density $f(x)$ for all $t \geqslant 1$.
(b) $D^{\mu}f(x)$ exists and is continuous on $\mathscr{X}$ and $\sup_{x\in\mathscr{X}}|D^{\mu}f(x)| < \infty$ for all μ with $\mu \leqslant \max\{\omega, q+1\}$ where $\omega \geqslant 2$ and $q \geqslant 1$ are positive integers that also appear in the other assumptions below.
(c) (i) The density $f(x) > 0$ for all $x \in \mathscr{X}$. Furthermore, there exist strictly positive constants $\underline{A}$, $\overline{A}$, ϱ, and Δ such that $\underline{A} \leqslant |x - x_0|^{-\varrho} f(x) \leqslant \overline{A}$ for all $x_0 \in \partial\mathscr{X}$ and all x such that $|x - x_0| \leqslant \Delta$.
(ii) $\int |v(x)|^{2r/(r-1)} f(x)\mathrm{d}x < \infty$, where $v(x) = D^2 m_0(x)/\{Dm_0(x)\}^2 - Df(x)/\{Dm_0(x)\}f(x)$.
(3) (a) $D^{\mu}m_0(x)$ exists and is continuous on $\mathscr{X}$ and $\sup_{x\in\mathscr{X}}|D^{\mu}m_0(x)| < \infty$ for all μ with $\mu \leqslant \max\{\omega, q+1\}$.
(b) $\mathrm{E}(|\ln|Dm_0(X_{t-1})||^{2r}) < \infty$.
(c) $(\min_{1\leqslant t\leqslant T}|Dm_0(X_{t-1})|)^{-1} = \mathrm{O_p}(1)$, $|Dm_0(x_0)| > 0$ for all $x_0 \in \partial\mathscr{X}$, and $\mathrm{E}[|Dm_0(X_{t-1})|^{-4r}] < \infty$.
(4) $K(\cdot) \in \mathscr{K}_{1,1,\omega} \cap \mathscr{K}_{1,q+1,q+1}$.
(5) The data-dependent bandwidth parameters $\{\hat{b}_T : T \geqslant 1\}$ satisfy $C_1 b_{1T} \leqslant \hat{b}_T \leqslant C_2 b_{2T}$ with probability tending to one for some sequences of bounded positive constants $\{b_{1T}: T \geqslant 1\}$ and $\{b_{2T}: T \geqslant 1\}$ and some positive finite constants C_1, C_2 that satisfy (a) $T^{1/2} b_{1T}^{\max\{4,q+2\}} d_T^{\max\{6,q+3\}\varrho} \to \infty$; (b) $b_{2T}^{\omega-(q+1)} d_T^{-(q+3)\varrho} \to 0$; (c) $T^{1/2} b_{2T}^{\omega} d_T^{-6\varrho} \to 0$; (d) $b_{1T}^{-1} b_{2T}^{\omega} d_T^{-4\varrho} \to 0$; (e) the trimming sequence d_T satisfies $d_T \geqslant C_2 b_{2T}$, $T d_T^{2+2\varrho} \to 0$, and $d_T^{2\varrho}/b_{1T} \to 0$.

The martingale difference assumption for the noise (Assumption 1(c)) is more general than the i.i.d. assumption that has been assumed in much of the previous chaos literature, e.g., Nychka et al. (1992) and McCaffrey et al. (1992). This assumption allows the distribution of the noise to be state dependent. In this case, it is possible for a noisy chaotic process to be both stationary and stay inside its basin of attraction, provided the support of noise is sufficiently small (see Chan and Tong (1994) among others for this point).

Our Assumption A2(c) (i) is that the marginal density $f(x)$ is zero at, and only at, the boundary points, and that the rate of convergence to zero is controlled by the parameter ϱ. This type of assumption has been used before in the context of average derivative estimation (Härdle et al., 1992, p. 223). The bounded support assumption is quite strong. However, the formula for the asymptotic variance of our estimator is finite for positively dependent Gaussian autoregressions, so we expect that our result holds more generally. Assumption A2(c) (ii) requires that $\varrho + 1 > 2r/(r-1)$.

The main substantive restriction is given in A3(c), which is slightly weaker than the requirement that Dm_0 be bounded strictly away from zero. This condition is not satisfied by any univariate chaotic process that we are aware of. However, it does include a large range of other processes such as first-order linear autoregressions with a nonzero coefficient, and other nonlinear processes.

In Assumption A1(b) the set $\mathscr{X}$ is assumed to be some open bounded subset of $\mathbb{R}$ with minimally smooth boundary. Examples of sets with minimally smooth boundaries include open bounded set that are convex or whose boundaries are C^1-embedded in $\mathbb{R}$. Finite unions of the aforementioned type whose closures are disjoint also have minimally smooth boundaries. Assumptions A2(b) and A3(a) imposes smoothness on $f(\cdot)$ and $m_0(\cdot)$. These conditions are needed to ensure that the realizations of $\hat{m}(\cdot)$ are smooth with probability tending to one and therefore the stochastic equicontinuity condition can be verified. This condition is used in the stochastic equicontinuity-based proof considered here (viz., Theorem 7 of Andrews (1989)). The use of higher-order kernels $K(\cdot)$ (Assumption A4) is due to the need to establish the T^{κ} convergence of the kernel estimators (see Lemma A.1(a) and (b) in Appendix) for some sufficiently large $\kappa \geqslant 1/4$, and in particular to reduce bias. Assumption A5 allows for data-dependent methods of choosing bandwidth parameters including cross-validation, generalized cross-validation, and plug-in procedures. The assumptions for the bandwidth and trimming parameters are compatible if ω is sufficiently large.

On the other hand, when we use a subsample of size n ($< T$) to estimate $\hat{\lambda}$, we use the following alternative assumptions:

A2(b)* $D^{\mu}f(x)$ exists and is continuous on $\mathscr{X}$ and $\sup_{x\in\mathscr{X}}|D^{\mu}f(x)| < \infty$ for all μ with $\mu \leqslant \omega$ where $\omega \geqslant 2$ is a positive integer.

A2(c)* The marginal density f satisfies $\inf_{x\in\mathscr{X}} f(x) > 0$.

A3(a)* $D^{\mu}m_0(x)$ exists and is continuous on $\mathscr{X}$ and $\sup_{x\in\mathscr{X}}|D^{\mu}m_0(x)| < \infty$ for all μ with $\mu \leqslant \omega$.

A3(c)* $M_n = \max_{1\leqslant t\leqslant n}(|Dm_0(X_{t-1})|^{-1}) = (\min_{1\leqslant t\leqslant n}|Dm_0(X_{t-1})|)^{-1} = O_p(n^{\phi})$ for some $\phi \geqslant 0$.

A(4)* $K(\cdot)\in\mathscr{K}_{1,1,\omega}$.

A(5)* The data-dependent bandwidth parameters $\{\hat{b}_T\colon T \geqslant 1\}$ satisfy $C_1 b_{1T} \leqslant \hat{b}_T \leqslant C_2 b_{2T}$ with probability tending to one for some sequences of bounded positive constants $\{b_{1T}\colon T \geqslant 1\}$ and $\{b_{2T}\colon T \geqslant 1\}$ and some positive finite constants C_1, C_2 that satisfy: (a) $T^{1/2-\kappa} b_{1T}^2 d^{3\rho} \to \infty$; (b) $T^{\kappa} b_{2T}^{\omega-1} d_T^{-3\varrho} \to 0$; (c) the trimming sequence d_T satisfies $d_T \geqslant C_2 b_{2T}$, $T d_T^{2+2\varrho} \to 0$ for some $0 < \kappa < \frac{1}{2}$.

A(6)* $n \to \infty$ and $n = O(T^{2\kappa/(1+2\phi)})$.

Our new Assumption A2(c)* that X_t has positive density on its bounded support strengthens Assumption A2(c). It is perhaps strong in relation to some previous work on nonparametric and semiparametric estimation in time series, see for example Robinson (1983) and Andrews (1995). We have made it here because the boundedness assumption is frequently assumed in the chaos literature, see Nychka et al. (1992), see below for a chaotic process that satisfies this condition.

Assumption 3(c)* is considerably weaker than Assumption 3(c). It is one of the subjects of extreme value theory whose mathematical study was initiated by

Gnedenko (1943), and can be easily understood in the context of i.i.d. random variables. Suppose that $Y_1,\dots,Y_n$ are i.i.d., then what magnitude can we expect for the random sequence $(\min_{1\leqslant i\leqslant n}|Y_i|)^{-1}$? We have for any $0<\varepsilon<\infty$,

$$\Pr\left(\frac{1}{n^{\phi}\min_{1\leqslant i\leqslant n}|Y_i|}\leqslant\varepsilon\right)=\{1-F_Y(\varepsilon^{-1}n^{-\phi})+F_Y(-\varepsilon^{-1}n^{-\phi}\}^n$$
$$=\left\{1-\frac{\xi_n(\varepsilon)}{n}\right\}^n\to \mathrm{e}^{-\xi(\varepsilon)}, \tag{13}$$

where by Taylor expansion

$$\xi_n(\varepsilon)=n\{F_Y(\varepsilon^{-1}n^{-\phi})-F_Y(-\varepsilon^{-1}n^{-\phi})\}=2f_Y(0)\varepsilon^{-1}+\mathrm{o}(1)$$

provided $\phi=1$ and $0<f_Y(0)<\infty$. Here, F_Y and f_Y are the cumulative distribution function and density of Y_i. Therefore, $M_n=(n\min_{1\leqslant i\leqslant n}|Y_i|)^{-1}=\mathrm{O_p}(1)$ and in fact converges in law to an extreme value distribution. In the dependent (strong mixing) case, the same rates and indeed limiting distributions apply under a certain additional condition on the tails of the joint distributions, called Watson's condition after Watson (1954), see the recent review paper by Leadbetter and Rootzén (1988). Specifically, if convergence holds for the associated i.i.d. sequence $\hat{Y}_i$ that has the same marginal distribution, and if the Watson condition holds, then the rates of convergence and asymptotic distributions of M_n and $\hat{M}_n$, where $\hat{M}_n=(n\min_{1\leqslant i\leqslant n}|\hat{Y}_i|)^{-1}$, are the same.[6]

In our case, we associate Y with $Dm_0(X)$. We, therefore, expect that Assumption 3(c)* holds with $\phi=1$ for chaotic Markov processes, since, as we have assumed, they typically have positive density throughout its support. For example, consider the (zero noise) univariate Feingenbaum mapping

$$X_t=4X_{t-1}(1-X_{t-1}). \tag{14}$$

The stationary ergodic density for this process is the well-known arcsine law $f(x)=1/\sqrt{\pi^2x(1-x)}$ on [0, 1] (Tong, 1990, p. 60).[7] This is strictly positive at the point $x=1/2$ at which $|Dm_0(x)|=0$, so the 'associated' i.i.d. sequence satisfies Assumption 3(c)* with $\phi=1$.

---

[6] Watson (1954) showed this result for m-dependent sequences, while Loynes (1965) extended this to strong mixing sequences.

[7] Although this process is deterministic, expectations can still be given meaning by interpreting them as integrals with respect to f. Furthermore, the process with a small amount of system noise is genuinely random but can be expected to have stationary density close to f.

When $\phi = 1$, the rate of growth of n compatible with Assumption A6* is $O(T^{2\kappa/3})$, where κ is the rate of convergence of the nonparametric estimator. The best one can hope for here, therefore, is an expansion rate of n close to $O(T^{1/3})$, which means that $\hat{\lambda}$ would be (close to) $T^{1/6}$ consistent. The rates improve when ϕ is smaller.

For the process (14), Assumption 3(b) is satisfied for any r. For $r = 3$, for example, we have

$$E(|\ln|Dm_0(X_{t-1})||^{2r}) = \int_0^1 (\ln|4 - 8x|)^6 \frac{dx}{\sqrt{\pi^2 x(1-x)}} = 117.04.$$

One last issue, which we partially address, is that our Assumption 2(b) can exclude some chaotic processes. Namely, in the zero noise Feingenbaum mapping $f(0) = f(1) = \infty$. This boundedness assumption for the density has been widely made in previous work in this field (see, for example, Cheng and Tong, 1992, Appendix A), but we believe it is not crucial. The main use of this assumption in our analysis is to establish the uniform consistency of the nonparametric estimates, for which purpose this assumption has been universally made (see Andrews (1995) and Masry (1996) among others).[8] Intuitively, however, an infinite density can be expected to improve the rate of convergence for a regression function estimator, because inside a given window width, there will be more observations at such poles. Recent work by Hengartner and Linton (1996) has established the superior pointwise rates of convergence in such problems.

Let

$$\Phi = \lim_{n\to\infty} \operatorname{var}\left[\frac{1}{\sqrt{n}} \sum_{t=1}^{n} \eta_t\right],$$

$$\Phi_j = \lim_{n\to\infty} \operatorname{var}\left[\frac{1}{\sqrt{n}} \sum_{t=1}^{n} \eta_{jt}\right], \quad j = 1, 2, \tag{15}$$

where

$$\eta_t = \eta_{1t} + \eta_{2t}, \qquad \eta_{1t} = v(X_{t-1})\varepsilon_t \quad \text{and} \quad \eta_{2t} = \ln|Dm_0(X_{t-1})| - \lambda.$$

We now have the main result of this paper.

---

[8] Uniform convergence results for series estimators, like Newey (1995), typically have been shown using 'higher-level' conditions that depend on the marginal density in a complicated way, and which most likely require this assumption also.

Theorem 1. (a) *Suppose that Assumptions A1–A5 hold with $n = T$. Then, we have*

$$\sqrt{T}(\hat{\lambda} - \lambda) \Rightarrow \mathrm{N}(0, \Phi).$$

(b) *Suppose that Assumptions A1, A2(a), (b), (c)*, A3(a)*, (b), (c)*, A4*, and A5* hold and n satisfies A6*. Then, we have*

$$\sqrt{n}(\hat{\lambda} - \lambda) \Rightarrow \mathrm{N}(0, \Phi_2).$$

Remark

1. This answers Conjectures 3.2 and 3.3 of McCaffrey et al. (1992) in the affirmative, namely that root-n convergence is to be expected.
2. The result of Theorem 1(b) can be straightforwardly extended to the cases in which $\hat{\lambda}$ is based on nonparametric estimates other than the kernel estimate. This is because an inspection of the proof of Theorem 1(b) shows that the argument goes through for any nonparametric estimator $D\hat{m}(\cdot)$ of $Dm_0(\cdot)$ provided it is uniformly consistent at the T^{κ} rate for some $0 < \kappa < 1/2$ (see Eq. (A.28) in Appendix). In these cases, one can replace the assumptions that we use to ensure the uniform convergence of the kernel estimates (see Lemma A.1 in Appendix for the list of such assumptions) by alternative assumptions depending on the choice of smoothing method. For local polynomial and series estimators one can find such conditions in Masry (1996) and Newey (1995), respectively.
3. It is evident from Eq. (15) that the choice of which particular size n triangular array subset affects the resulting asymptotic variance. In practice, it is advisable to take observations equally spaced so as to minimize their mutual dependence.

We next discuss estimation of the asymptotic variance. The proof of Theorem 1(a) consists in establishing that

$$\sqrt{T}(\hat{\lambda} - \lambda) = \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \eta_t + \mathrm{o_p}(1),$$

and $T^{-1/2}\sum_{t=1}^{T}\eta_t$ satisfies a central limit theorem. In fact, η_t is a stationary and mixing sequence with mean zero. Likewise, in the proof of Theorem 1(b) we show that

$$\sqrt{n}(\hat{\lambda} - \lambda) = \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \eta_{2t} + \mathrm{o_p}(1).$$

This structure can be used now to apply standard methods for estimating asymptotic covariance matrices, which have been developed in, inter alia: White

(1984), Gallant (1987), Newey and West (1987), Andrews (1991), Andrews and Monahan (1992), Hansen (1992), and DeJong and Davidson (1996).

Define

$$\hat{\Phi} = \sum_{j=-n+1}^{n-1} k\left(\frac{j}{S_n}\right)\hat{\gamma}(j) \quad \text{and} \quad \hat{\Phi}_\ell = \sum_{j=-n+1}^{n-1} k\left(\frac{j}{S_n}\right)\hat{\gamma}_\ell(j), \quad \ell = 1,2,$$

where

$$\hat{\gamma}(j) = \frac{1}{n}\sum_{t=|j|+1}^{n} \hat{\eta}_t\hat{\eta}_{t-|j|}; \qquad \hat{\gamma}_\ell(j) = \frac{1}{n}\sum_{t=|j|+1}^{n} \hat{\eta}_{\ell t}\hat{\eta}_{\ell, t-|j|}, \tag{16}$$

$$\hat{\eta}_{1t} = \{X_t - \hat{m}(X_{t-1})\}\left\{\frac{D^2\hat{m}(X_{t-1})}{[D\hat{m}(X_{t-1})]^2} - \frac{D\hat{f}(X_{t-1})}{[D\hat{m}(X_{t-1})]\hat{f}(X_{t-1})}\right\}\mathbf{1}_t,$$

$$\hat{\eta}_{2t} = \{\ln|D\hat{m}(X_{t-1})| - \hat{\lambda}\}\mathbf{1}_t \quad \text{and} \quad \hat{\eta}_t = \hat{\eta}_{1t} + \hat{\eta}_{2t}.$$

Here, $k(\cdot)$ and S_n denote a kernel function and a lag truncation parameter, respectively. Under the conditions of Theorem 1(a), we use $\hat{\Phi}$ with $n = T$, whereas we use $\hat{\Phi}_2$ with $n < T$ under the conditions of Theorem 1(b).

For the consistency of the estimators $\hat{\Phi}$ and $\hat{\Phi}_2$, we impose the following assumptions on the kernel function and the lag truncation parameter:

Assumption S

(1) $k(\cdot)$ belongs to the following class:

$$\mathcal{K}_1 = \{k:\mathbb{R} \to [-1,1] \,|\, k(0) = 1,$$

$$k(-x) = k(x)\ \forall x \in \mathbb{R}, \int_{\infty}^{\infty} k^2(x)\,\mathrm{d}x < \infty,$$

$k(\cdot)$ is continuous at 0 and at all but a finite number of points$\}$.

(2) S_n satisfies $S_n \to \infty$ and $S_n/n \to 0$.

We also use the following bandwidth conditions:

A5** The data-dependent bandwidth parameters $\{\hat{b}_T: T \geqslant 1\}$ satisfy $C_1 b_{1T} \leqslant \hat{b}_T \leqslant C_2 b_{2T}$ with probability tending to one for some sequences of bounded positive constants $\{b_{1T}: T \geqslant 1\}$ and $\{b_{2T}: T \geqslant 1\}$ and some positive finite constants C_1, C_2 that satisfy (a) $T^{1/2} b_{1T}^{\max\{3,q+3\}} d_T^{\{7,q+4\}\varrho} \to \infty$; (b) $b_{2T}^{\omega-q-2} d_T^{-\max\{7,q+4\}\rho} \to 0$; (c) the trimming sequence d_T satisfies $d_T \to 0$.

Corollary 1. (a) *Suppose that Assumptions S, A1–A4, and A5** hold with* $2r/(r-2)$ *in A1*(*a*), $\max\{\omega,q+1\}$ *in A2*(*b*) *and A3*(*a*), *and* $\mathscr{K}_{1,1,\omega}\cap\mathscr{K}_{1,q+1,q+1}$ *in A4 replaced by* $3r/(r-2)$, $\max\{\omega,q+2\}$, *and* $\mathscr{K}_{1,q+2,q+2}$ *respectively. Then,*

$$\hat{\Phi}\rightarrow_{\mathrm{p}}\Phi.$$

(b) *Suppose that Assumptions S, A1, A2*(*a*), (*b*), (*c*)*, *A3*(*a*)*, (*b*), (*c*)*, *A4*, and A5* hold with* $2r/(r-2)$ *in* A1(a) *replaced by* $3r/(r-2)$ *and n satisfies A6*. Then,*

$$\hat{\Phi}_2\rightarrow_{\mathrm{p}}\Phi_2.$$

Remark

1. Compared to Theorem 1(a), Corollary 1(a) imposes weaker conditions for the kernel function. In particular, no higher-order kernel is now needed for consistent estimation of Φ.
2. The proof of Corollary 1 requires an extension of Andrews (1991, Proposition 1) to allow for residuals derived from nonparametric smoothing procedures (and to deal with certain specific features of our data generating process in part (b)).[9]
3. Regarding practical implementation of the above procedure we refer the reader to the above cited papers, and Andrews (1991) in particular.[10]

3.2. The general case $k\geqslant 1$

We now discuss the extension of the results in Section 3.1 to the case in which k may be greater than one. The central technical problem here is that

$$\ln\left\|\prod_{t=1}^{n}J_{n-t}\right\|^{2}\neq\sum_{t=1}^{n}\ln\|J_{n-t}\|^{2} \tag{17}$$

as was the case for $k=1$. This makes the argument much less transparent. We shall consider two cases: (a) when n is small relative to T and (b) when $n=T$. In the latter case, stronger unverifiable assumptions are required, while the former case permits weaker standard conditions although results in slower rates of convergence. The rate of convergence to a normal distribution is $n^{1/2}$ as is to be expected from Section 3.1.

[9] An alternative bootstrap approach to computing standard errors has been proposed in Gençay (1996), but nothing has been established about its asymptotic properties.

[10] There is a commercially available sweet of programs in the Gauss language, see Ouliaris and Phillips (1994), that can perform many of the bandwidth selection and other procedures defined in Andrews (1991) and the other references.

We first introduce more notation. Define

$$F_t(J_{n-1},\dots,J_0)=\frac{\partial \ln \|\prod_{t=1}^{n} J_{n-t}\|}{\partial \Delta m(Z_t)'}, \tag{18}$$

$$F_n(z)=\mathrm{E}\{F_{t-1}(J_{n-1},\dots,J_0)\,|\,Z_t=z\}=(F_{n1}(z),\dots,F_{nk}(z))', \tag{19}$$

$$Q_{ts}(J_{n-1},\dots,J_0)=\frac{\partial^2 \ln \|\prod_{t=1}^{n} J_{n-t}\|}{\partial \Delta m(Z_t)\partial \Delta m(Z_s)'}, \tag{20}$$

$$\Delta m(Z_t)=(\Delta m_{1t},\Delta m_{2t},\dots,\Delta m_{kt})'. \tag{21}$$

Define also

$$G_{T1}=-\frac{1}{\sqrt{T}}\sum_{t=1}^{T} v(Z_{t-1})\varepsilon_t, \tag{22}$$

$$G_{T2}=\sqrt{T}\left[\frac{1}{T}\ln\left\|\prod_{t=1}^{T} J_{T-t}\right\|-\lambda\right], \tag{23}$$

where $v(Z_{t-1})=\sum_{j=1}^{k} D^{e_j}\{F_{Tj}(Z_{t-1})f(Z_{t-1})\}/f(Z_{t-1})$. To determine the trimming, we take $d(x,y)=\{x'\hat{\Omega}^{-1}y\}^{1/2}$.

We make the following assumptions.

Assumption B

(1) (a) $\{X_t\colon t\geqslant 1\}$ is a sequence of strictly stationary strong mixing random variables with mixing numbers of size $-2r/(r-2)$ for some $r>2$.
 (b) For all $t\geqslant 1$, X_t lies in an open bounded set $\mathscr{X}$ ($\subset\mathbb{R}$) with minimally smooth boundary denoted $\partial\mathscr{X}$.
 (c) $\{\varepsilon_t\colon t\geqslant 1\}$ are i.i.d. with $\mathrm{E}(\varepsilon_t\,|\,\mathscr{G}_{-\infty}^{t-1})=0$ and $\mathrm{E}(|\varepsilon_t|^{2r})<\infty$ for all $t\geqslant 1$, where $\mathscr{G}_s^t=\sigma(Z_s,\dots,Z_t)$ is the σ-algebra generated by $(Z_s,\dots,Z_t)$ and $Z_t=(X_t,\dots,X_{t-k+1})'$.

(2) (a) The joint distribution of $Z_t\in\mathbb{R}^k$ is absolutely continuous with respect to Lebesgue measure with density $f(z)$.
 (b) $D^{\mu}f(z)$ exists and is continuous on $\mathscr{Z}=\mathscr{X}\times\cdots\times\mathscr{X}$ and $\sup_{z\in\mathscr{Z}}|D^{\mu}f(z)|<\infty$ for all μ with $|\mu|\leqslant\max\{\omega,q+1\}$, where $\omega\geqslant 2$ and $q\geqslant k/2$ are positive integers that also appear in the other assumptions below.
 (c) (i) The density $f(z)>0$ for all $z\in\mathscr{Z}$. Furthermore, there exist strictly positive constants $\underline{A}$, $\overline{A}$, ϱ, and Δ such that $\underline{A}\leqslant\|z-z_0\|^{-\varrho}f(z)\leqslant\overline{A}$ for all $z_0\in\partial\mathscr{Z}$ and all z such that $\|z-z_0\|\leqslant\Delta$.
 (ii) $\int|v(z)|^{2r/(r-1)}f(z)\,\mathrm{d}z<\infty$.

(3) $D^{\mu}m_0(z)$ exists and is continuous on $\mathscr{Z}$ and $\sup_{z\in\mathscr{Z}}|D^{\mu}m_0(z)| < \infty$ for all μ with $|\mu| \leqslant \max\{\omega, q+1\}$.

(4) (a) $K(\cdot)\in\mathscr{K}_{k,1,\omega}\cap\mathscr{K}_{k,q+1,q+1}$.
(b) $\lambda_{\min}\{\mathrm{E}(Z_{t-1}-\mathrm{E}Z_{t-1})(Z_{t-1}-\mathrm{E}Z_{t-1})'\} > 0$.

(5) The data-dependent bandwidth parameters $\{\hat{b}_T\colon T \geqslant 1\}$ satisfy $C_1 b_{1T} \leqslant \hat{b}_T \leqslant C_2 b_{2T}$ with probability tending to one for some sequences of bounded positive constants $\{b_{1T}\colon T\geqslant 1\}$ and $\{b_{2T}\colon T\geqslant 1\}$ and some positive finite constants C_1, C_2 that satisfy (a) $T^{1/2} b_{1T}^{\max\{2(k+1),\, q+k+1\}} d_T^{\max\{6,q+3\}\varrho} \to \infty$; (b) $b_{2T}^{\omega-(q+1)} d_T^{-(q+3)\varrho} \to 0$; (c) $T^{1/2} b_{2T}^{\omega} d_T^{-6\varrho} \to 0$; (d) $b_{1T}^{-k} b_{2T}^{\omega} d_T^{-4\varrho} \to 0$; (e) the trimming sequence d_T satisfies $d_T \geqslant C_2 b_{2T}$, $T d_T^{2+2\varrho} \to 0$, and $d_T^{2\varrho}/b_{1T} \to 0$.

(6) (a) $(1/T)\sum_{t=1}^{T} \|F_{t-1}(J_{T-1},\dots,J_0) - \mathrm{E}[F_{t-1}(J_{T-1},\dots,J_0)|\mathscr{G}_{t-l-1}^{t+m-1}]\| = \mathrm{O_p}(T^{-1/4})$ for some non-negative integers $l,m < \infty$.
(b) $\sup_{t\leqslant T, T\geqslant 1}\mathrm{E}(\|F_{t-1}(J_{T-1},\dots,J_0)\|^r) < \infty$ and $\sup_{T\geqslant 1}|F_T(z_0)\| < \infty$ for all $z_0\in\partial\mathscr{Z}$.
(c) $(1/T)\sum_{t=1}^{T}\sum_{s=1}^{T} \|Q_{ts}(J_{T-1},\dots,J_0)\| = \mathrm{O_p}(1)$.

(7) For some 2×2 positive semi-definite matrix Ξ, $G_T = (G'_{T1}, G'_{T2})' \Rightarrow \mathrm{N}(0,\Xi)$.

Assumption B6(a) is trivially satisfied for the case $k=1$, since in this case $F_{t-1}(J_{T-1},\dots,J_0) = 1/Dm_0(X_{t-1})$ and hence one can take $l=m=0$ and $\mathscr{G}_{t-l-1}^{t+m-1} = \mathscr{G}_{t-1}^{t-1} = \sigma(Z_{t-1}) = \sigma(X_{t-1})$. Assumption B6(b) and (c) are also satisfied for the case $k=1$ under Assumption A3(c), because in that case we have

$$\sup_{t\leqslant T,\, T\geqslant 1} \mathrm{E}[\|F_{t-1}(J_{T-1},\dots,J_0)|^r] = \mathrm{E}[|Dm_0(X_{t-1})|^{-r}],$$

$$\sup_{T\geqslant 1}|F_T(z_0)| = |Dm_0(x_0)|^{-1},$$

and

$$\frac{1}{T}\sum_{t=1}^{T}\sum_{s=1}^{T}\|Q_{ts}(J_{T-1},\dots,J_0)\| = \frac{1}{T}\sum_{t=1}^{T}\frac{1}{|Dm_0(X_{t-1})|^2}.$$

However, for the case $k\geqslant 2$, the Assumption B6 can be restrictive when $n=T$, at least it is hard to find more primitive conditions that would justify this assumption, except when m_0 is linear – in this case, F_t is constant with respect to t and $Q_{ts}=0$.

Assumption B7 is satisfied for G_{T1} by a central limit theorem for strong mixing random variables for any k. In the case $k=1$, it is satisfied for G_{T2}, and in fact jointly for (G_{T1}, G_{T2}), because then

$$\sqrt{T}\left[\frac{1}{T}\ln\left\|\prod_{t=1}^{T} J_{T-t}\right\| - \lambda\right] = \frac{1}{\sqrt{T}}\sum_{t=1}^{T}[\ln|Dm_0(X_{t-1})| - \lambda].$$

For $k > 1$, the asymptotic normality of G_{T2} follows by some results of Furstenberg and Kesten (1960). Letting

$$\xi_s = \ln\left(\left\|\prod_{t=s-1}^{T} J_{T-t}\right\| \Big/ \left\|\prod_{t=s}^{T} J_{T-t}\right\|\right),$$

where $\{\xi_s\}_{s=1}^{T}$ is a sequence of stationary and asymptotically independent scalar random variables, we can write

$$\sqrt{T}\left[\frac{1}{T}\ln\left\|\prod_{t=1}^{T} J_{T-t}\right\| - \lambda\right] = \frac{1}{\sqrt{T}}\sum_{t=1}^{T}[\xi_s - \lambda],$$

and apply their central limit theorem.[11]

On the other hand, when we use a subsample of size $n(<T)$ to estimate $\hat{\lambda}$, we use the following alternative assumptions:

B(2) (a)* $D^{\mu}f(z)$ exists and is continuous on $\mathscr{Z}$ and $\sup_{z\in\mathscr{Z}}|D^{\mu}f(z)| < \infty$ for all μ with $\mu \leqslant \omega$ where $\omega \geqslant 2$ is a positive integer.
(b)* The marginal density f satisfies $\inf_{z\in\mathscr{Z}} f(z) > 0$.

B(3)* $D^{\mu}m_0(z)$ exists and is continuous on $\mathscr{Z}$ and $\sup_{z\in\mathscr{Z}}|D^{\mu}m_0(z)| < \infty$ for all μ with $|\mu| \leqslant \omega$.

B(4)* $K(\cdot) \in \mathscr{K}_{1,1,\omega}$.

B(5)* The data-dependent bandwidth parameters $\{\hat{b}_T: T \geqslant 1\}$ satisfy $C_1 b_{1T} \leqslant \hat{b}_T \leqslant C_2 b_{2T}$ with probability tending to one for some sequences of bounded positive constants $\{b_{1T}: T \geqslant 1\}$ and $\{b_{2T}: T \geqslant 1\}$ and some positive finite constants C_1, C_2 that satisfy (a) $T^{(1/2)-\kappa} b_{1T}^{k+1} d^{3\rho} \to \infty$; (b) $T^{\kappa} b_{2T}^{\omega-1} d_T^{-3\varrho} \to 0$; (c) the trimming sequence d_T satisfies $d_T \geqslant C_2 b_{2T}$, $T d_T^{2+2\varrho} \to 0$ for some $0 < \kappa < \frac{1}{2}$.

B(6)* $\max_{1\leqslant t\leqslant n}\|F_{t-1}(J_{n-1}, \ldots, J_0)\| = O_p(n^{\phi})$ for some $\phi \geqslant 0$.

B(7)* For some $\Xi_{22} > 0$, $G_{n2} \Rightarrow N(0, \Xi_{22})$.

B(8)* $n \to \infty$ and $n = O(T^{2\kappa/(1+2\phi)})$.

[11] See the remark after Theorem 3 in Furstenberg and Kesten (1960).

Theorem 2. (a) *Suppose that Assumptions B1–B8 hold. Then,*

$$\sqrt{T}(\hat{\lambda} - \lambda) \Rightarrow \mathrm{N}(0, a'\Xi a),$$

where $a = (1, 1)'$.

(b) *Suppose that Assumptions B1, B2(a), (b)*, (c)*, B3*, B4(a)*, (b), and B5*–B7* hold and n satisfies B8*. Then,*

$$\sqrt{n}(\hat{\lambda} - \lambda) \Rightarrow \mathrm{N}(0, \Xi_{22}).$$

When we use the full sample result of Theorem 2(a) (i.e., $n = T$), we use the same estimate $\hat{\Phi}_2$ of the asymptotic covariance matrix defined in Eq. (16) above, except that

$$\hat{\eta}_{1t} = \left[\frac{1}{\hat{f}(Z_{t-1})}\sum_{j=1}^{k} D^{e_j}\{\hat{F}_j(Z_{t-1})\hat{f}(Z_{t-1})\}\right]\hat{\varepsilon}_t, \qquad \hat{\eta}_{2t} = \hat{\xi}_s - \hat{\lambda}, \tag{24}$$

where

$$\hat{\xi}_s = \ln\left(\left\|\prod_{t=s-1}^{n} \hat{J}_{n-t}\right\| \Big/ \left\|\prod_{t=s}^{n} \hat{J}_{n-t}\right\|\right); \qquad \hat{F}_{j,t} = \frac{\partial \ln\|\prod_{t=1}^{n}\hat{J}_{n-t}\|}{\partial \Delta \hat{m}_{jt}}.^{12}$$

Here, $\hat{F}_j(\cdot)$ is the kernel estimate of the regression function $F_{nj}(\cdot)$ obtained by smoothing $\hat{F}_{j,t}$ against Z_{t-1}. If, for example, $\|\cdot\|$ is the Euclidean matrix norm, i.e., $\|C\| = \mathrm{tr}(C'C)^{1/2}$ for any matrix C, we have

$$F_{j,t} = \frac{\partial \ln\|\prod_{t=1}^{n} J_{n-t}\|}{\partial \Delta m_{jt}} = \frac{\mathrm{tr}\{A_n'\partial A_n/\partial \Delta m_{jt} + \partial A_n'/\partial \Delta m_{jt} A_n\}}{2\|\prod_{t=1}^{n} J_{n-t}\|^2}, \quad j = 1, \ldots, k,$$

where $A_n = \prod_{t=1}^{n} J_{n-t}$, and

$$\frac{\partial A_n}{\partial \Delta m_{jt}} = \left(\prod_{s=1}^{n-t-1} J_{n-t}\right)\frac{\partial J_t}{\partial \Delta m_{jt}}\left(\prod_{s=0}^{t} J_{t-s}\right).$$

The matrix $\partial J_s/\partial \Delta m_{jt}$ consists of either zeros or ones. Therefore, $\hat{F}_{j,t}$ can be calculated analytically using these expressions, although it is considerably easier to use numerical derivatives. On the other hand, when $n < T$, we can drop the first term in Eq. (24) and use $\hat{\eta}_{2t} = \hat{\xi}_s - \hat{\lambda}$.

[12] Note that the trimming is implicit here, and is reflected in the smaller sample size T^*.

4. Numerical results

We first analyse how our procedure works on the following AR(1) process

$$X_t = \rho X_{t-1} + u_t; \quad u_t \sim \mathrm{N}(0,1),$$

for which $\lambda = \ln|\rho|$ and $\Phi = (1-\rho^2)/\rho^2$. We conducted simulations of our procedures for estimating λ and Φ when $k = 1$.[13] We used the analytic first derivative of estimator (9) with a quartic kernel $K(u) = 15(1-u^2)^2\mathbf{1}(|u| \leqslant 1)/16$ and bandwidth $h = \gamma * \mathrm{range}\{X_t\}_{t=1}^{T}$, where $\gamma \in \{0.2, 0.3, 0.4\}$. No trimming at all was used. A total of 1000 replications were used for each experiment. We take $\rho \in \{0.5, 0.7, 0.9, 0.95, 1.0\}$, $n = T \in \{100, 200, 300, 500\}$, and first report the mean, median, and standard deviation of $\hat{\lambda}$ (Tables 1–3).

There does appear to be a substantial small sample upward bias; although this improves with bandwidth and sample size, so that when $n = 500$ and $h = 0.4 \times$ range the estimated λ lies very close to the truth. For this configuration, the standard deviation of the estimator is very close to the asymptotic standard deviation predicted by Theorem 1. In Fig. 1 we compare the finite sample density (as computed by kernel method with Silverman's rule of thumb bandwidth) with the normal density that has the same mean and variance for one particular configuration. The distribution has some skewness, but generally is close to normality.

We now turn to the standard errors. The second derivatives estimates appearing in the residuals $\hat{\eta}_{1t}$ were computed using analytic second derivatives of estimator (9) using the same bandwidth used in estimating the first derivative. The covariance functions were weighted with Bartlett's kernel $k(t) = 1 - |t|$. We just report the results for $n = 500$ and lag length $S_n = 5$. We also tried other lag lengths but the results were very similar. We report two different estimates of $\Phi^{1/2}$, $\hat{\Phi}^{1/2}$ and $\hat{\Phi}_1^{1/2}$ – we know that $\eta_{2t} = 0$ here, because the mean is linear. (See Table 4).

There appears to be considerable dependence on h, but in any case small h does quite well for large $\rho = 0.9$, 0.95. It seems that the infeasible $\hat{\Phi}_1^{1/2}$ does much better for small ρ.

We next examined a univariate chaotic process, the Feingenbaum map with system noise:

$$X_t = 4X_{t-1}(1 - X_{t-1}) + \sigma\varepsilon_t,$$

[13] Although this d.g.p. violates the assumption of bounded support, the asymptotic variance formula for this design is finite. We have used exactly the procedure described in Section 2, that is, we did not trim out regions where the marginal density was small.

Table 1
Gaussian autoregression for the full sample estimator (i.e., $n = T$), with $\gamma = 0.2$

		$n = 100$	$n = 200$	$n = 300$	$n = 500$
$\rho = 0.5$	mean	− 0.937	− 0.844	− 0.793	− 0.748
($\lambda = -0.693, \Phi = 3.0$)	median	− 0.887	− 0.801	− 0.769	− 0.739
	sd	0.292	0.207	0.146	0.094
	asd[a]	0.173	0.122	0.100	0.078
$\rho = 0.7$	mean	− 0.552	− 0.447	− 0.414	− 0.388
($\lambda = -0.357, \Phi = 1.04$)	median	− 0.508	− 0.429	− 0.405	− 0.383
	sd	0.208	0.109	0.076	0.053
	asd	0.104	0.072	0.059	0.046
$\rho = 0.9$	mean	− 0.269	− 0.179	− 0.152	− 0.132
($\lambda = -0.105, \Phi = 0.24$)	median	− 0.242	− 0.170	− 0.146	− 0.129
	sd	0.123	0.058	0.041	0.027
	asd	0.050	0.0346	0.028	0.022
$\rho = 0.95$	mean	− 0.215	− 0.127	− 0.099	− 0.079
($\lambda = -0.051, \Phi = 0.11$)	median	− 0.198	− 0.119	− 0.096	− 0.077
	sd	0.101	0.049	0.032	0.021
	asd	0.033	0.024	0.019	0.015
$\rho = 1.0$	mean	− 0.167	− 0.080	− 0.053	− 0.032
($\lambda = 0, \Phi = 0$)	median	− 0.151	− 0.073	− 0.048	− 0.030
	sd	0.091	0.037	0.026	0.015

[a]Value predicted by the theory, for reference.

Table 2
Gaussian autoregression for the full sample estimator (i.e., $n = T$), with $\gamma = 0.3$

		$n = 100$	$n = 200$	$n = 300$	$n = 500$
$\rho = 0.5$	mean	− 0.887	− 0.776	− 0.742	− 0.720
($\lambda = -0.693, \Phi = 3.0$)	median	− 0.831	− 0.751	− 0.731	− 0.712
	sd	0.292	0.162	0.116	0.084
$\rho = 0.7$	mean	− 0.481	− 0.410	− 0.388	− 0.374
($\lambda = -0.357, \Phi = 1.04$)	median	− 0.449	− 0.398	− 0.383	− 0.369
	sd	0.162	0.093	0.066	0.049
$\rho = 0.9$	mean	− 0.212	− 0.152	− 0.134	− 0.121
($\lambda = -0.105, \Phi = 0.24$)	median	− 0.193	− 0.142	− 0.129	− 0.118
	sd	0.092	0.049	0.036	0.025
$\rho = 0.95$	mean	− 0.162	− 0.100	− 0.082	− 0.069
($\lambda = -0.051, \Phi = 0.11$)	median	− 0.147	− 0.094	− 0.078	− 0.066
	sd	0.077	0.039	0.028	0.019
$\rho = 1.0$	mean	− 0.117	− 0.056	− 0.037	− 0.022
($\lambda = 0, \Phi = 0$)	median	− 0.102	− 0.051	− 0.034	− 0.020
	sd	0.069	0.029	0.021	0.012

Table 3
Gaussian autoregression for the full sample estimator (i.e., $n = T$)

		$n = 100$	$n = 200$	$n = 300$	$n = 500$
$\rho = 0.5$	mean	-0.834	-0.747	-0.724	-0.710
$(\lambda = -0.693, \Phi = 3.0)$	median	-0.786	-0.725	-0.716	-0.703
	sd	0.260	0.147	0.110	0.083
$\rho = 0.7$	mean	-0.447	-0.393	-0.378	-0.369
$(\lambda = -0.357, \Phi = 1.04)$	median	-0.420	-0.382	-0.371	-0.364
	sd	0.143	0.087	0.065	0.049
$\rho = 0.9$	mean	-0.183	-0.138	-0.126	-0.117
$(\lambda = -0.105, \Phi = 0.24)$	median	-0.169	-0.130	-0.121	-0.114
	sd	0.079	0.046	0.034	0.025
$\rho = 0.95$	mean	-0.134	-0.087	-0.073	-0.064
$(\lambda = -0.051, \Phi = 0.11)$	median	-0.122	-0.080	-0.070	-0.062
	sd	0.066	0.036	0.026	0.018
$\rho = 1.0$	mean	-0.090	-0.043	-0.029	-0.017
$(\lambda = 0, \Phi = 0)$	median	-0.080	-0.039	-0.026	-0.016
	sd	0.057	0.025	0.019	0.011

where $\varepsilon_t/v_t \sim \mathrm{U}(-1,1)$ independent of X_t, and

$$v_t = \min\{4X_{t-1}(1-X_{t-1}), 1-4X_{t-1}(1-X_{t-1})\}.$$

This particular form of heteroskedasticity ensures that the process X_t is restricted to the unit interval. The parameter σ was chosen to make the noise/signal ratio, as defined in Dechert and Gençay (1992), lie in $\{0.005, 0.007, 0.010\}$.[14] When $\sigma = 0$, $\lambda = \ln 2$. We take $\gamma \in \{0.2, 0.3, 0.4\}$ and $T \in \{100, 200, 300, 500\}$ as before, and considered full sample estimation $n = T$ and subsamples $n = cT^{1/6}$, $n = cT^{1/3}$, and $n = cT^{1/2}$ with $c = 4.31$ (the subsamples were chosen equally spaced). A local quadratic smoother (see Masry, 1996) with a quartic kernel was used in the estimation of λ and in constructing $\hat{\eta}_{2t}$. We again used the Bartlett kernel to weight the autocovariances. Again no trimming was used. Our results do not change much with bandwidth or with signal-to-noise ratio, so we only report the results for $\gamma = 0.2$ and $\sigma = 0.005$ for which $\lambda = 0.692$. (see Table 5).

The estimates are close to the truth, although are rather noisy, especially for the small subsamples (note that when $T = 100$ and $n = cT^{1/6}$, then n is actually

[14] Specifically, the unconditional variance σ^2 of the innovation was calculated by simulating the series many times. The signal to noise ratio is $\sigma_x/\sigma\sigma_\varepsilon$, where σ_x is the standard deviation of the system data.

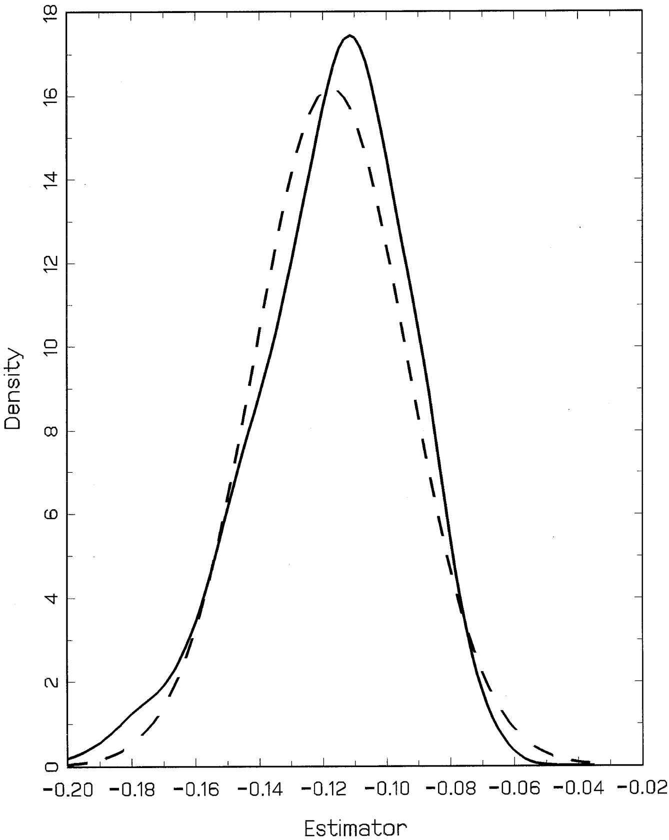

Fig. 1. Solid line is empirical distribution of estimator $\hat{\lambda}$, while dashed line is a normal distribution with same mean and variance. Parameter values were: $\rho = 0.9$, $T = 500$, and $\gamma = 0.4$.

9). The actual standard deviations are in quite close agreement with the values predicted by the asymptotic theory. One surprise here is how well the procedure works for the full sample case, which is not covered by our theory. The standard errors, $\hat{\Phi}_2^{1/2}$, were less successful when n was small, but appeared to do pretty well in the larger subsample cases. In Fig. 2 we give the finite sample distribution of $\hat{\lambda}$ for one configuration.

5. Conclusions

Our results have provided the basic tools needed to apply statistical inference to the Lyapunov exponent computed from stochastic data. The main result,

Table 4

		$\rho =$	0.5	0.7	0.9	0.95[a]
	$\gamma = 0.2$	median	0.431	0.079	0.022	0.014
		iqr	4.213	0.328	0.012	0.005
$\hat{\Phi}^{1/2}$	$\gamma = 0.3$	median	0.080	0.034	0.015	0.010
		iqr	0.154	0.010	0.003	0.002
	$\gamma = 0.4$	median	0.047	0.024	0.012	0.008
		iqr	0.019	0.006	0.002	0.002
	$\gamma = 0.2$	median	0.092	0.044	0.020	0.014
		iqr	0.104	0.018	0.004	0.003
$\hat{\Phi}_1^{1/2}$	$\gamma = 0.3$	median	0.057	0.032	0.015	0.011
		iqr	0.027	0.006	0.003	0.002
	$\gamma = 0.4$	median	0.042	0.024	0.012	0.009
		iqr	0.010	0.004	0.002	0.002

[a] $n = T = 500$.

Note: The (corresponding) true values of $1/\hat{2}$ for $\bar{.}$,0.7,0.9, and 0.95 are, respectively, 0.077, 0.046, 0.022, and 0.015.

Table 5

Results for the Feingenbaum mapping with system noise with $T = 500$ and various subsamples $n = T$,$n = cT^{1/2}$,$n = cT^{1/3}$, and $cT^{1/6}$

		$T = 100$	$T = 200$	$T = 300$	$T = 500$
$\hat{\lambda}$ $(n = T)$	mean	0.693	0.693	0.693	0.693
	median	0.693	0.693	0.693	0.693
	sd	0.013	0.007	0.004	0.003
se($\hat{\lambda}$)	median	0.049	0.036	0.030	0.023
	iqr	0.019	0.011	0.011	0.007
$\hat{\lambda}$ $(n = cT^{1/6})$	mean	0.730	0.689	0.696	0.687
	median	0.767	0.728	0.732	0.706
	sd	0.283	0.289	0.273	0.259
	asd	0.302	0.286	0.273	0.262
se($\hat{\lambda}$)	median	0.205	0.213	0.210	0.191
	iqr	0.141	0.142	0.129	0.124
$\hat{\lambda}$ $(n = cT^{1/3})$	mean	0.694	0.696	0.689	0.686
	median	0.707	0.717	0.696	0.699
	sd	0.203	0.184	0.173	0.154
	asd	0.203	0.181	0.171	0.155
se($\hat{\lambda}$)	median	0.180	0.158	0.153	0.140
	iqr	0.085	0.070	0.062	0.053
$\hat{\lambda}$ $(n = cT^{1/2})$	mean	0.692	0.691	0.693	0.697
	median	0.691	0.690	0.693	0.696
	sd	0.111	0.110	0.103	0.089
	asd	0.138	0.117	0.105	0.092
se($\hat{\lambda}$)	median	0.120	0.109	0.100	0.088
	iqr	0.032	0.024	0.022	0.020

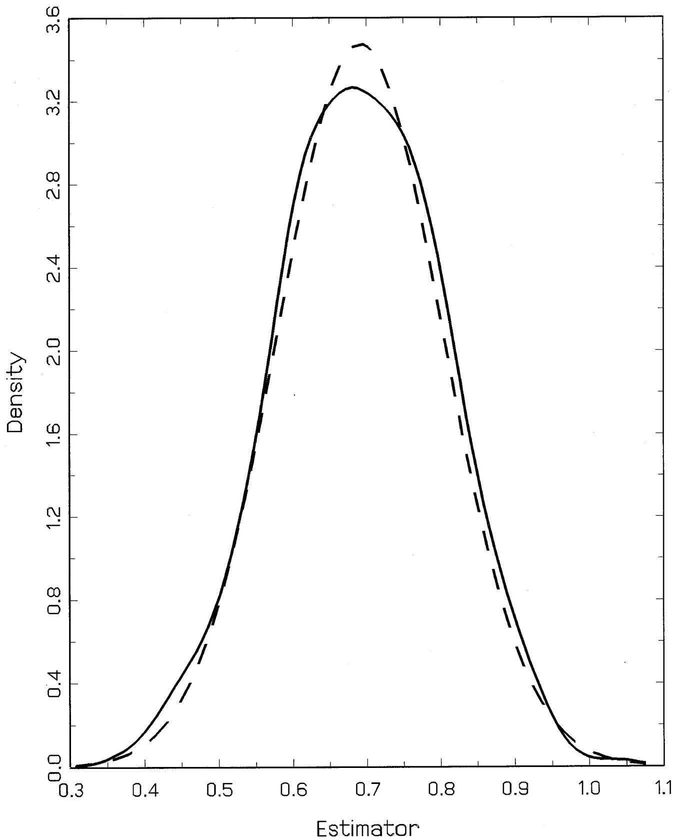

Fig. 2. Solid line is empirical distribution of estimator $\hat{\lambda}$, while dashed line is a normal distribution with same mean and variance. Parameter values were $n = 96$, $T = 500$, and $\gamma = 0.4$.

Theorem 1, confirms that the subsample smoothing-based Lyapunov exponent estimate is asymptotically normally distributed at rate root-n. The best magnitude we are able to allow for n is (close to) $T^{1/3}$ for processes with chaotic-like behaviour. This clearly leads to very large data requirements in order to achieve desirable accuracy in the estimation, requirements which are only likely to be met by certain financial time series. However, when these data are available, our results permit standard rules of inference to be applied, in particular the huge array of methods for estimating asymptotic covariance matrices can be employed, and the practical advice given in, for example, Andrews (1991) can be used to decide the implementation issues. In terms of the smoothing method used to estimate the regression function, our results also leave much room for choice. At present, we are inclined to recommend the procedures developed in

Nychka et al. (1992) and Dechert and Gençay (1992), since they have been well tuned to this particular problem.

Acknowledgements

We would like to thank two referees, Don Andrews, Ramo Gençay, A. Ron Gallant, and Peter Phillips for helpful comments, Moto Shintani for research assistance, and the National Science Foundation for financial support. The first author would like to thank the Cowles Foundation for its generous hospitality during his recent visit.

Appendix A.

Let $\{X_t\colon t \geqslant 1\}$ be a sequence of rv's and let $\mathscr{F}_s^t$ denote the σ-field generated by $(X_s, \dots, X_t)$. Define

$$\alpha(s) = \sup_{t \geqslant 1} \sup_{A \in \mathscr{F}^t{}_1,\, B \in \mathscr{F}^\infty{}_{t+s}} |\Pr(A \cap B) - \Pr(A)\Pr(B)| \quad \text{for } s \geqslant 1.$$

Then, $\{X_t\colon t \geqslant 1\}$ is defined to be *strong mixing of size* $-\beta$ if $\alpha(s) = O(s^{-\beta-\varepsilon})$ for some $\varepsilon > 0$.

The class of kernels is defined as follows:

$$\begin{aligned}\mathscr{K}_{k,\delta,v} = \Big\{ & K(\cdot)\colon[-1,1]^k \to \mathbb{R} \,\Big|\, \int K(z)\,\mathrm{d}z = 1, \int z^\mu K(z)\,\mathrm{d}z = 0 \\ & \forall 1 \leqslant |\mu| \leqslant v - \delta - 1, \int |z^\mu K(z)|\,\mathrm{d}z < \infty \quad \forall |\mu| = v - \delta, \\ & D^\mu K(z) \to 0 \quad \text{as } \|z\| \to \infty \quad \forall \mu \text{ with } |\mu| < \delta, D^\mu K(z) \text{ is} \\ & \text{absolutely integrable and has Fourier transform} \\ & \Psi_\mu(r) = (2\pi)^k \int \exp(\mathrm{i}r'z) D^\mu K(z)\,\mathrm{d}z \text{ that satisfies} \\ & \int (1 + \|r\|) \sup_{b \geqslant 1} \|\Psi_\mu(br)|\,\mathrm{d}r < \infty \quad \forall \mu \text{ with } |\mu| \leqslant \delta, \text{ and} \\ & \sup_{z \in \mathbb{R}^k} |D^{\mu + e_j} K(z)|(\|z\| \vee 1) < \infty \quad \forall \mu \text{ with } |\mu| \leqslant \delta \ \forall j = 1, \dots, k, \\ & \text{where } \vee \text{ denotes the maximum operator and} \mathrm{i} = \sqrt{-1} \Big\}.\end{aligned}$$

Below, for notational simplicity, we let C_j for some integer $j \geqslant 1$ denote a generic constant. (It is not meant to be equal in any two places it appears.)

Proof of Theorem 1. Define the interior region

$$\mathscr{X}_T = \left\{ x \in \mathscr{X} \colon \inf_{x_0 \in \partial \mathscr{X}} |x - x_0| \geqslant d_T \right\}$$

and its topological boundary $\partial \mathscr{X}_T$, and note that $\partial \mathscr{X}_T$ is a finite set. Furthermore, $\inf_{x_0 \in \partial \mathscr{X}} |x - x_0| \geqslant d_T \Rightarrow f(x) \geqslant \underline{A} d_T^{\varrho}$, i.e., $\mathscr{X}_T \subseteq \{x \in \mathscr{X} \colon f(x) \geqslant \underline{A} d_T^{\varrho}\}$. We first present a lemma that is used to prove our main result.

Lemma A.1. Suppose that Assumptions A1, A2(a)–(c), A3(a), A4, and A5 hold. Then, we have for $\mu \leqslant \omega$,

(a) $\sup_{x \in \mathscr{X}_T} |D^{\mu} \hat{f}(x) - D^{\mu} f(x)| = O_p(T^{-1/2} b_{1T}^{-(\mu+1)}) + O_p(b_{2T}^{\omega-\mu})$.
(b) $\sup_{x \in \mathscr{X}_T} |D^{\mu} \hat{m}(x) - D^{\mu} m_0(x)| = O_p(T^{-1/2} b_{1T}^{-(\mu+1)} d_T^{-(2+\mu)\varrho}) + O_p(b_{2T}^{\omega-\mu} d_T^{-(2+\mu)\varrho})$.

Proof. The results of Lemma A.1 follow directly from Theorem 1 of Andrews (1995). It suffices to verify (Assumptions) NP1–NP5 of the latter paper. Note that NP1, NP2, and NP3 are implied by (Assumptions) A1, A2(a)–(b), and A3(a), respectively, with $\eta = \infty$, $\beta = r$, and $|\lambda| = \mu$ and (Y_{Tt}, X_{Tt}), $f_{Tt}(x)$ and $g_T(x)$ given by (X_t, X_{t-1}), $f(x)$, and $m_0(x)$, respectively. NP4(a) and (b) hold by A4. NP4(c) holds with $\hat{\Omega}$ defined in Eq. (10) with Z_{t-1} replaced by X_{t-1} and $\Omega_T = \mathrm{var}(X_{t-1})$ using A1 and a central limit theorem (CLT) of (Herrndorf, 1984, Corollary 1) applied to the strong mixing random variables $\{X_{t-1} \colon t \geqslant 1\}$. Finally NP5 is satisfied by A5. □

Proof of Theorem 1 (*Conclusion*). (a) By adding and subtracting terms, we have

$$\sqrt{T}(\hat{\lambda} - \lambda) = \frac{1}{2\sqrt{T}} \sum_{t=1}^{T} [\ln (D\hat{m}(X_{t-1}))^2 - \ln (Dm_0(X_{t-1}))^2] \mathbf{1}_t$$

$$+ \frac{1}{2\sqrt{T}} \sum_{t=1}^{T} [\ln (Dm_0(X_{t-1}))^2 - 2\lambda] \mathbf{1}_t - \frac{\lambda}{\sqrt{T}} \sum_{t=1}^{T} \{1 - \mathbf{1}_t\} \tag{A.1}$$

$$\equiv A_T + B_T - C_T, \text{ say.} \tag{A.2}$$

Below we establish the following results:

$$A_T = \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \eta_{1t} + o_p(1), \tag{A.3}$$

$$C_T = o_p(1), \tag{A.4}$$

$$B_T = \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \eta_{2t} + o_p(1). \tag{A.5}$$

Results (A.1)–(A.5) and an application of the CLT of (Herrndorf (1984), Corollary 1) for strong mixing random variables give the desired result.

We now verify Eq. (A.3). A two-term Taylor expansion of A_T (defined in Eq. (A.1) about $Dm_0(X_{t-1})$ gives

$$\begin{aligned} A_T &= \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \frac{1}{Dm_0(X_{t-1})} [D\hat{m}(X_{t-1}) - Dm_0(X_{t-1})] \mathbf{1}_t \\ &\quad - \frac{1}{2\sqrt{T}} \sum_{t=1}^{T} \frac{1}{[Dm^*(X_{t-1})]^2} [D\hat{m}(X_{t-1}) - Dm_0(X_{t-1})]^2 \mathbf{1}_t \\ &= A_{1T} + A_{2T}, \text{ say,} \end{aligned} \tag{A.6}$$

where $Dm^*(X_{t-1})$ lies between $Dm_0(X_{t-1})$ and $D\hat{m}(X_{t-1})$. We first consider A_{2T}. We have

$$\begin{aligned} |A_{2T}| &\leq \frac{1}{2} T^{1/2} \left[\sup_{x \in \mathscr{X}_T} |D\hat{m}(x) - Dm_0(x)| \right]^2 \frac{1}{T} \sum_{t=1}^{T} \frac{\mathbf{1}_t}{[Dm^*(X_{t-1})]^2} \\ &\leq o_p(1) \times \left(\frac{1}{\min_{\{t:\ X_{t-1} \in \mathscr{X}_T\}} |Dm^*(X_{t-1})|} \right)^2 \xrightarrow{p} 0, \end{aligned} \tag{A.7}$$

where the second inequality follows by the uniform consistently results in Lemma A.1(b) and the bandwidth conditions in Assumption A5 and the last convergence to zero holds because $(\min_{\{t:\ X_{t-1} \in \mathscr{X}_T\}} |Dm^*(X_{t-1})|)^{-2} = O_p(1)$ by

Assumption A3(c) and Lemma A.1(b). To see the latter, it suffices to show that $(\min_{\{t:\ X_{t-1}\in\mathscr{X}_T\}}|Dm^*(X_{t-1})|)^{-1} = O_p(1)$. For this purpose, choose first some $\varepsilon\in(0,1)$. Note that there exists $0 < M_\varepsilon < \infty$ and $T_{1\varepsilon}$ such that

$$\Pr\left(\frac{1}{\min_{\{t:X_{t-1}\in\mathscr{X}_T\}}|Dm_0(X_{t-1})|} \leqslant \frac{M_\varepsilon}{2}\right) \geqslant 1-\frac{\varepsilon}{2}, \quad \forall T \geqslant T_{1\varepsilon}, \tag{A.8}$$

by Assumption A3(c). There also exists $T_{2\varepsilon}$ such that

$$\Pr\left(\max_{\{t:\ X_{t-1}\in\mathscr{X}_T\}}|D\hat{m}(X_{t-1}) - Dm_0(X_{t-1})| \geqslant \frac{1}{M_\varepsilon}\right) \leqslant \frac{\varepsilon}{2}, \quad \forall T \geqslant T_{2\varepsilon}, \tag{A.9}$$

by Lemma A.1(b). Therefore, for $T \geqslant \max\{T_{1\varepsilon}, T_{2\varepsilon}\}$, we have

$$\begin{aligned}
&\Pr\left(\frac{1}{\min_{\{t:\ X_{t-1}\in\mathscr{X}_T\}}|Dm^*(X_{t-1})|} \leqslant M_\varepsilon\right)\\
&\geqslant \Pr\left(\frac{1}{\min_{\{t:\ X_{t-1}\in\mathscr{X}_T\}}|Dm_0(X_{t-1})|} \leqslant \frac{M_\varepsilon}{2}\right)\\
&\quad - \Pr\left(\frac{1}{\min_{\{t:\ X_{t-1}\in\mathscr{X}_T\}}|Dm_0(X_{t-1})|}\right.\\
&\left.\leqslant \frac{M_\varepsilon}{2} \text{ and } \frac{1}{\min_{\{t:\ X_{t-1}\in\mathscr{X}_T\}}|Dm^*(X_{t-1})|} \geqslant M_\varepsilon\right)\\
&\geqslant 1-\frac{\varepsilon}{2} - \Pr\left(\min_{\{t:\ X_{t-1}\in\mathscr{X}_T\}}|Dm_0(X_{t-1})|\right.\\
&\left.\geqslant \frac{2}{M_\varepsilon} \text{ and } \min_{\{t:\ X_{t-1}\in\mathscr{X}_T\}}|Dm^*(X_{t-1})| \leqslant \frac{1}{M_\varepsilon}\right)\\
&\geqslant 1-\frac{\varepsilon}{2} - \Pr\left(\max_{\{t:\ X_{t-1}\in\mathscr{X}_T\}}|Dm^*(X_{t-1}) - Dm_0(X_{t-1})|\right.\\
&\quad \left. + \min_{\{t:\ X_{t-1}\in\mathscr{X}_T\}}|Dm^*(X_{t-1})| \geqslant \frac{2}{M_\varepsilon} \text{and} \left(\min_{\{t:\ X_{t-1}\in\mathscr{X}_T\}}|Dm^*(X_{t-1})| \leqslant \frac{1}{M_\varepsilon}\right)\right.
\end{aligned}$$

$$\geqslant 1 - \frac{\varepsilon}{2} - \Pr\left(\max_{\{t:\ X_{t-1} \in \mathscr{X}_T\}} |D\hat{m}(X_{t-1}) - Dm_0(X_{t-1})| \geqslant \frac{1}{M_\varepsilon} \right)$$

$$\geqslant 1 - \frac{\varepsilon}{2} - \frac{\varepsilon}{2} = 1 - \varepsilon. \tag{A.10}$$

Since the choice of ε can be arbitrary, we have the desired result.

We next consider A_{1T}. We verify the following result:

$$A_{1T} = \sqrt{T} \int_{\mathscr{X}_T} \frac{1}{Dm_0(x)} [D\hat{m}(x) - Dm_0(x)] f(x)\,\mathrm{d}x + \mathrm{o_p}(1). \tag{A.11}$$

To show Eq. (A.11), define an empirical process $v_T(\cdot)$ by

$$v_T(\tau) = \frac{1}{\sqrt{T}} \sum_{t=1}^{T} [r(X_{t-1}, \tau) - Er(X_{t-1}, \tau)], \tag{A.12}$$

where

$$r(X_{t-1}, \tau) = \frac{\mathbf{1}_t}{Dm_0(X_{t-1})} \tau(X_{t-1}),$$

and $\tau \in \mathscr{T}$ for some pseudo-metric space $\mathscr{T}$ with pseudo-metric $\rho_{\mathscr{T}}(\cdot, \cdot)$ defined by

$$\rho_{\mathscr{T}}(\tau_1, \tau_2) = \left[\int_{\mathscr{X}_T} \{r(x, \tau_1) - r(x, \tau_2)\}^2 \,\mathrm{d}x \right]^{1/2}. \tag{A.13}$$

Suppose $\hat{\tau}$ is an estimator of $\tau_0 \in \mathscr{T}$. It is well known that (see, for example, (Andrews (1994), p. 2257))

$$v_T(\hat{\tau}) - v_T(\tau_0) \xrightarrow{\mathrm{p}} 0 \tag{A.14}$$

if (i) $\Pr(\hat{\tau} \in \mathscr{T}) \to 1$, (ii) $\rho_{\mathscr{T}}(\hat{\tau}, \tau_0) \xrightarrow{\mathrm{p}} 0$, and (iii) $\{v_T(\cdot):\ T \geqslant 1\}$ is stochastically equicontinuous at τ_0. Therefore, the result in Eq. (A.11) holds by taking $\hat{\tau}(\cdot) = D\hat{m}(\cdot)$ and $\tau_0(\cdot) = Dm_0(\cdot)$ if we verify the conditions (i)–(iii) above, as is done below. Let $\mathscr{T}$ be a class of smooth functions; specifically, for large $C < \infty$, let

$$\mathscr{T} = \left\{ \tau(\cdot): \left(\sum_{|\alpha| \leqslant q} \int_{\mathscr{X}_T} (D^\alpha \tau(x))^{1/2} \, dx \right)^{1/2} \leqslant C \right\}, \tag{A.15}$$

where $q \geqslant 1$ is a positive integer that appears in Assumptions A2–A5. Note that $\tau_0(\cdot) = Dm_0(\cdot)$ lies in $\mathscr{T}$ by Assumption A3(a). To show that (i) $\Pr(\hat{\tau} \in \mathscr{T}) \to 1$, it suffices to show that $\hat{\tau}(\cdot) = D\hat{m}(\cdot)$ has partial derivatives of order q on $\mathscr{X}_T$ that are bounded uniformly over $\mathscr{X}_T$ with probability tending to one. Note that the latter holds by the uniform consistency result in Lemma A.1(b) and Assumption A3(a). With the pseudo-metric defined in Eq. (A.13). The condition (ii) $\rho_{\mathscr{T}}(\hat{\tau}, \tau_0) \xrightarrow{p} 0$ holds since, for some positive constant $C_1 < \infty$,

$$\begin{aligned}
\rho^2_{\mathscr{T}}(\hat{\tau}, \tau_0) &= \int_{\mathscr{X}_T} \frac{1}{[Dm_0(x)]^2} [D\hat{m}(x) - Dm_0(x)]^2 \, dx \\
&\leqslant \left[\sup_{x \in \mathscr{X}_T} |D\hat{m}(x) - Dm_0(x)| \right]^2 (\mathrm{E}|Dm_0(X_{t-1})|^{-2}) \\
&\quad \times \inf_{x \in \mathscr{X}_T} f^{-1}(x) \leqslant (\mathrm{O_p}(T^{-1/2} b_{1T}^{-2} d_T^{-3\varrho}) + \mathrm{O_p}(b_{2T}^{(\omega-1)} d_T^{-3\varrho}))^2 \\
&\quad \times \mathrm{O}(d_T^{-\rho}) \\
&\xrightarrow{p} 0,
\end{aligned} \tag{A.16}$$

where the second inequality holds by Assumptions A2(c) and A3(c) and Lemma A.1(b). Now, the condition (iii) of $\{v_T(\cdot): T \geqslant 1\}$ holds by the stochastic equicontinuity results of (Andrews, 1989, Theorem 7) that are applicable to classes of functions that are products of smooth functions from an infinite dimensional class and a single unbounded function. It suffices, therefore, to verify (Assumption) E of the latter paper. E(i) holds by taking W_{aTt}, W_{bTt}, $\tau_a(\cdot)$,$\tau_b(\cdot)$, $m_a(W_{aTt}, \tau_a)$, and $m_b(W_{bTt}, \tau_b)$ to be X_{t-1}, X_{t-1}, $\mathbf{1}_t/Dm_0(\cdot)$, $\tau(\cdot)$, $\mathbf{1}_t/Dm_0(X_{t-1})$, and $\tau(X_{t-1})$, respectively. E(ii) holds by Assumption A1(b) with $\mathscr{W}_a^*$ given by $\mathscr{X}_T$. E(iii) follows by the definition of $\mathscr{T}$ in Eq. (A.15). E(iv) is irrelevant to our case. E (v) holds since $\sup_{t \geqslant 1} \mathrm{E}|\mathbf{1}_t/Dm_0(X_{t-1})|^r < \infty$ for $r > 2$ by Assumption A3(c). Finally, E(vi) holds by Assumption A1(a).

We now consider the first term on the right-hand side of Eq. (A.11). We have

$$\sqrt{T}\int_{\mathscr{X}_T} \frac{1}{Dm_0(x)}[D\hat{m}(x) - Dm_0(x)]f(x)\,\mathrm{d}x$$

$$= \sqrt{T}\int_{\mathscr{X}_T} \left[\frac{f(x)}{Dm_0(x)}\right][D\hat{m}(x) - Dm_0(x)]\,\mathrm{d}x$$

$$= -\sqrt{T}\int_{\mathscr{X}_T} \left[D\left(\frac{f(x)}{Dm_0(x)}\right)\right][\hat{m}(x) - m_0(x)]\,\mathrm{d}x$$

$$+ \sqrt{T}\sum_{x\in\partial\mathscr{X}_T} \left(\frac{f(x)}{Dm_0(x)}\right)[\hat{m}(x) - m_0(x)] \tag{A.17}$$

$$= -\sqrt{T}\int_{\mathscr{X}_T} \left[D\left(\frac{f(x)}{Dm_0(x)}\right)\frac{1}{f(x)}\right][\{\hat{m}(x) - m_0(x)\}\hat{f}(x)]\,\mathrm{d}x$$

$$+ \sqrt{T}\int_{\mathscr{X}_T} \left[D\left(\frac{f(x)}{Dm_0(x)}\right)\right]\left[\frac{1}{\hat{f}(x)} - \frac{1}{f(x)}\right][\{\hat{m}(x) - m_0(x)\}\hat{f}(x)]\mathrm{d}x$$

$$+ \sqrt{T}\sum_{x\in\partial\mathscr{X}_T} \left(\frac{f(x)}{Dm_0(x)}\right)[\hat{m}(x) - m_0(x)]$$

$$= A^{*}_{1T} + A^{**}_{1T} + A^{***}_{1T}, \text{ say,} \tag{A.18}$$

where the second equality (A.17) holds by integration by parts and the third equality (A.18) is obtained by adding in and subtracting out a term. We first show that A^{**}_{1T} is $\mathrm{o_p}(1)$. We have, for some positive constant $C_2, C_3 < \infty$

$$|A^{**}_{1T}| \leqslant T^{1/2}\left(\sup_{x\in\mathscr{X}_T} |\hat{m}(x) - m_0(x)|\right)\left(\sup_{x\in\mathscr{X}_T} |\hat{f}(x) - f(x)|\right)$$

$$\times\left(\inf_{x\in\mathscr{X}_T} f^{-2}(x)\right) \times C_1(\mathrm{E}|Dm_0(X_{t-1})|^{-2})$$

$$\leqslant T^{1/2}(\mathrm{O_p}(T^{-1/2}b_{1T}^{-1}d_T^{-2\varrho}) + \mathrm{O_p}(b_{2T}^{\omega}d_T^{-2\varrho}))(\mathrm{O_p}(T^{-1/2}b_{1T}^{-1})$$

$$+ \mathrm{O_p}(b_{2T}^{\omega})) \times \mathrm{O}(d_T^{-2\rho}) \xrightarrow{\mathrm{p}} 0, \tag{A.19}$$

where the second inequality holds by Lemmas A.1–A.3 and the last convergence to zero holds using the bandwidth conditions in Assumption A5.

We next consider A_{1T}^{*}. We write

$$\{\hat{m}(x) - m_0(x)\}\hat{f}(x) = \frac{1}{T\hat{b}_T}\sum_{t=1}^{T} \hat{K}\left(\frac{x - X_{t-1}}{\hat{b}_T}\right)\{X_t - m_0(X_{t-1})\} + \frac{1}{T\hat{b}_T}\sum_{t=1}^{T} \hat{K}\left(\frac{x - X_{t-1}}{\hat{b}_T}\right)\{m_0(X_{t-1}) - m_0(x)\},$$

and, letting $v(x) = [D(f(x)/Dm_0(x))/f(x)]$, we have that

$$\sqrt{T}\int_{\mathscr{X}_T} v(x)\left[\frac{1}{T\hat{b}_T}\sum_{t=1}^{T} \hat{K}\left(\frac{x - X_{t-1}}{\hat{b}_T}\right)\{X_t - m_0(X_{t-1})\}\right]\mathrm{d}x$$

$$= \frac{1}{\sqrt{T}}\sum_{t=1}^{T} \{X_t - m_0(X_{t-1})\}\int_{\mathscr{X}_T} v(x)\frac{1}{\hat{b}_T}\hat{K}\left(\frac{x - X_{t-1}}{\hat{b}_T}\right)\mathrm{d}x$$

$$= \frac{1}{\sqrt{T}}\sum_{t=1}^{T} \{X_t - m_0(X_{t-1})\}\left[v(X_{t-1}) + \int_{\mathscr{X}_T} \{v(X_{t-1} + u\hat{b}_T) - v(X_{t-1})\}\hat{K}(u)\,\mathrm{d}u\right] = \frac{1}{\sqrt{T}}\sum_{t=1}^{T} \{X_t - m_0(X_{t-1})\}v(X_{t-1})\{1 + o_p(1)\}, \tag{A.20}$$

where the second equality follows by a change of variables $x \to u = (x - X_{t-1})/\hat{b}_T$, while the third equality (A.20) is true by the following argument. First consider this term with deterministic sequences b_T and Ω_T with $C_1 b_{1T} \leqslant b_T \leqslant C_2 b_{2T}$ and $|\Omega_T - \Omega| \leqslant c/\sqrt{T}$ for any $c < \infty$. Then, since $X_t - m_0(X_{t-1})$ is a martingale difference sequence,

$$\mathrm{E}\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T} \{X_t - m_0(X_{t-1})\}\left[\int_{\mathscr{X}_T} \{v(X_{t-1} + ub_T) - v(X_{t-1})\}K_T(u)\,\mathrm{d}u\right]\right)^2$$

$$= \mathrm{E}\left(\{X_t - m_0(X_{t-1})\}^2\left[\int_{\mathscr{X}_T} \{v(X_{t-1} + ub_T) - v(X_{t-1})\}K_T(u)\,\mathrm{d}u\right]^2\right)$$

$$\leqslant \{\mathrm{E}(|X_t - m_0(X_{t-1})|^{2r})\}^{1/r}\left\{\mathrm{E}\left(\left[\int_{\mathscr{X}_T} \{v(X_{t-1} + ub_T) - v(X_{t-1})\} \times K_T(u)\,\mathrm{d}u\right]^{2r/(r-1)}\right)\right\}^{(r-1)/r} \to 0,$$

where $K_T(u) = \det(\Omega_T)^{-1/2}K(\Omega_T^{-1/2}u)$. The inequality is due to Hölder, and convergence to zero follows by Assumption A2(c)(ii) using a well-known result for convolutions of functions in an L_p-space $[p = 2r/(r-1)]$ (see Theorem 8.14 of Folland (1984)). Using stochastic equicontinuity arguments as above we can extend the result to bounded stochastic sequences $\hat{b}_T$ and $\hat{\Omega}_T$, where $\hat{\Omega}_T$ satisfies $|\hat{\Omega}_T - \Omega| \leqslant c/\sqrt{T}$ for some constant c. The result follows because $\Pr[|\hat{\Omega} - \Omega| \leqslant c/\sqrt{T}]$ can be made arbitrarily close to unity by taking c large.

Furthermore,

$$\left|\sqrt{T}\int_{\mathscr{X}_T} v(x)\frac{1}{T\hat{b}_T}\sum_{t=1}^{T} \hat{K}\left(\frac{x - X_{t-1}}{\hat{b}_T}\right)\{m_0(X_{t-1}) - m_0(x)\}\mathrm{d}x\right|$$

$$= \left|\sqrt{T}\int_{\mathscr{X}_T} v(x)f^{1/2}(x)\frac{1}{T\hat{b}_T}\sum_{t=1}^{T} \hat{K}\left(\frac{x - X_{t-1}}{\hat{b}_T}\right)\{m_0(X_{t-1}) - m_0(x)\}f^{-1/2}(x)\,\mathrm{d}x\right|$$

$$\leqslant \sqrt{T}\int_{\mathscr{X}_T} |v(x)|^2 f(x)\,\mathrm{d}x \times \inf_{x\in\mathscr{X}_T} f^{-1}(x) \times \sup_{x\in\mathscr{X}_T}\left|\frac{1}{T\hat{b}_T}\sum_{t=1}^{T} \times \hat{K}\left(\frac{x - X_{t-1}}{\hat{b}_T}\right)\{m_0(X_{t-1}) - m_0(x)\}\right| \leqslant \sqrt{T}\mathrm{O}0(d_T^{-\varrho}) \times \mathrm{O_p}(b_{2T}^{\omega}) \xrightarrow{\mathrm{p}} 0, \tag{A.21}$$

where the first inequality uses the Cauchy–Schwarz inequality, while the second inequality in Eq. (A.21) using similar arguments to Lemma A.1 and the fact that $\int_{\mathscr{X}}|v(x)|^2 f(x)\,\mathrm{d}x < \infty$, while the last convergence to zero holds by the bandwidth conditions in Assumption A5.

We now turn to the boundary term A_{1T}^{***}. By the pointwise properties of kernel estimators (Hengartner and Linton, 1996),

$$A_{1T}^{***} = \sqrt{T}\sum_{x\in\partial\mathscr{X}_T}\left(\frac{f(x)}{Dm_0(x)}\right)[\hat{m}(x) - m_0(x)]$$

$$= \sqrt{T}\mathrm{O_p}(d_T^{\varrho})\{\mathrm{O_p}(b_{2T}^{\omega}) + \mathrm{O_p}(T^{-1/2}b_{1T}^{-1/2}d_T^{-\varrho/2})\} \xrightarrow{\mathrm{p}} 0, \tag{A.22}$$

by Assumptions A2 and A5.

We next prove Eq. (A.6). Let $\varepsilon > 0$ be given. Since each term in C_T is positive, we have by the Markov inequality,

$$\Pr[C_T > \varepsilon] \leqslant \mathrm{E}(C_T)/\varepsilon = \frac{\sqrt{T}}{2\varepsilon}\{1 - \Pr(X_{t-1} \in \mathscr{X}_T)\} \leqslant \frac{\sqrt{T}}{2\varepsilon} A d_T^{1+\varrho} \to 0, \tag{A.23}$$

provided that $T d_T^{2+2\varrho} \to 0$ which holds under Assumption A5. The last inequality follows by our Assumption A2(c). In conclusion, $C_T = \mathrm{o_p}(1)$ as required.

Proof of Eq. (A.5). Let $\delta_t = [\ln(Dm_0(X_{t-1}))^2 - 2\lambda](1 - \mathbf{1}_t)$, then by Assumption A5,

$$\sqrt{T}\mathrm{E}(\delta_t) = \mathrm{o}(1),$$

while

$$\mathrm{var}\left[\frac{1}{\sqrt{T}}\sum_{t=1}^{T} \delta_t\right] = \frac{1}{T}\sum_{t=1}^{T} \mathrm{var}(\delta_t) + \frac{2}{T} \sum_{t=1}^{T} \sum_{\substack{s=1 \\ s \neq t}}^{T} \mathrm{cov}(\delta_t, \delta_s)$$

$$\leqslant \frac{1}{T}\sum_{t=1}^{T} \mathrm{var}(\delta_t) + 16\left\{\sum_{s=1}^{T} \alpha^{1/a}(s)\mathrm{E}^{2/b}[|\delta_t|^b]\right\} \to 0. \tag{A.24}$$

The inequality follows by Theorem 3, part 1 of Doukhan (1994) provided $1/a + 2/b = 1$. Convergence to zero then follows taking $b = 2r - \varepsilon'$, for some $\varepsilon' > 0$, since

$$\mathrm{E}[|\delta_t|^b] \leqslant \mathrm{E}[|\ln(Dm_0(X_{t-1}))^2 - 2\lambda|^{b+\varepsilon'}]\mathrm{E}[(1 - \mathbf{1}_t)] = \mathrm{O}(d_T^{1+\varrho}),$$

by the Hölder inequality and Assumption A3(b), while

$$\sum_{s=1}^{T} \alpha^{(2r-\varepsilon'-2)/(2r-\varepsilon')}(s) = \sum_{s=1}^{T} s^{-(2r-\varepsilon'-2)/(r-2)} \to 0,$$

taking $\varepsilon' = \varepsilon(r-2)$ in the definition of mixing below (25). Note that $\mathrm{var}(\delta_t) \to 0$ by dominated convergence.

(b) By rearranging terms,

$$\sqrt{n}(\hat{\lambda} - \lambda) = \frac{1}{2\sqrt{n}}\sum_{t=1}^{n} [\ln(D\hat{m}(X_{t-1}))^2 - \ln(Dm_0(X_{t-1}))^2]\mathbf{1}_t$$

$$+\frac{1}{2\sqrt{n}}\sum_{t=1}^{n}[\ln(Dm_0(X_{t-1}))^2-2\lambda]\mathbf{1}_t-\frac{\lambda}{\sqrt{n}}\sum_{t=1}^{n}\{1-\mathbf{1}_t\}$$

$$\equiv A_n+B_n-C_n,\text{ say.} \tag{A.25}$$

The arguments in Eq. (A.23) can be used to show that

$$C_n=\mathrm{o_p}(1), \tag{A.26}$$

$$B_n=\frac{1}{2\sqrt{n}}\sum_{t=1}^{n}[\ln(Dm_0(X_{t-1}))^2-2\lambda]+\mathrm{o_p}(1). \tag{A.27}$$

Therefore, Assumption A1(a) along with Assumption A3(b) gives

$$B_n\Rightarrow \mathrm{N}(0,\Phi_2)$$

by the CLT of (Herrndorf (1984), Corollary 1). Now the proof of part (b) of Theorem 1 is complete because

$$|A_n|=\left|\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\frac{\mathbf{1}_t}{Dm^*(X_{t-1})}[D\hat{m}(X_{t-1})-Dm_0(X_{t-1})]\right|$$

$$\leqslant T^{-\kappa}n^{1/2}+\phi[T^{\kappa}\sup_{x\in\mathscr{X}_T}|D\hat{m}(x)-Dm_0(x)|]$$

$$\times\left(\frac{1}{n^{\phi}\min_{\{t:\ X_{t-1}\in\mathscr{X}_T\}}|Dm^*(X_{t-1})|}\right)\xrightarrow{\mathrm{p}}0, \tag{A.28}$$

where the equality holds by a one-term Taylor expansion of A_n about $Dm_0(X_{t-1})$ (with $Dm^*(X_{t-1})$ lying between $Dm_0(X_{t-1})$ and $D\hat{m}(X_{t-1})$) and the inequality and the convergence to zero holds by Lemma A.1(b) and Assumptions A6* and A3(c)* because we then have $T^{-\kappa}n^{1/2+\phi}=\mathrm{O}(1)$ and $(n^{\phi}\min_{1\leqslant t\leqslant n}|Dm^*(X_{t-1})|)^{-1}=\mathrm{O_p}(1)$, respectively (with the latter being verified using the similar arguments to those given in Eqs. (A.8), (A.9) and (A.10) above). □

Proof of Corollary 1. (a) Define

$$\tilde{\Phi}=\sum_{j=-T+1}^{T-1}k\left(\frac{j}{S_T}\right)\tilde{\gamma}(j),\qquad \tilde{\gamma}(j)=\frac{1}{T}\sum_{t=|j|+1}^{T}\eta_t\eta_{t-|j|},$$

$$\bar{\Phi}=\sum_{j=-T+1}^{T-1}k\left(\frac{j}{S_T}\right)\bar{\gamma}(j),\qquad \bar{\gamma}(j)=\frac{1}{T}\sum_{t=|j|+1}^{T}\bar{\eta}_t\bar{\eta}_{t-|j|}, \tag{A.29}$$

where $\bar{\eta}_t = \eta_t \mathbf{1}_t$ and η_t is defined in Eq. (15). Note that $\tilde{\Phi} \xrightarrow{p} \Phi$ by the result of Andrews (1991), Proposition 1) because Assumption A of the latter paper holds by A1(a) and Lemma 1 of Andrews (1991). Furthermore, $\tilde{\Phi} - \bar{\Phi} \xrightarrow{p} 0$ by dominated convergence. By the triangular inequality, therefore, it suffices to show that $\hat{\Phi} - \bar{\Phi} \xrightarrow{p} 0$.

Define the process

$$\eta_t(\tau) = \left(\frac{\tau_1(X_{t-1})}{\tau_2(X_{t-1})^2} - \frac{\tau_3(X_{t-1})}{\tau_2(X_{t-1})\tau_4(X_{t-1})}\right)(X_t - \tau_5(X_{t-1}))\mathbf{1}_t + (\ln|\tau_2(X_{t-1})| - \lambda)\mathbf{1}_t \tag{A.30}$$

for $\tau = (\tau_1, \tau_2, \tau_3, \tau_4, \tau_5) \in \mathscr{T}^* = \mathscr{T}_1 \times \cdots \times \mathscr{T}_5$, where $\mathscr{T}_i = \mathscr{T}$ (defined in Eq. (A.15) for $i = 1,2,3,5$ and $\mathscr{T}_4 \subset \{\tau_4 : \inf_{x \in \mathscr{X}_T} |\tau_4(x)| \geqslant A d_T^{\varrho}\}$. Let $\hat{\tau}(\cdot) = (D^2\hat{m}(\cdot), D\hat{m}(\cdot), D\hat{f}(\cdot), \hat{f}(\cdot), \hat{m}(\cdot))'$ and $\tau_0(\cdot) = (D^2 m_0(\cdot), Dm_0(\cdot), Df(\cdot), f(\cdot), m_0(\cdot))'$. Note that, with this definition, $\hat{\eta}_t = \eta_t(\hat{\tau})$ and $\bar{\eta}_t = \eta_t(\tau_0)$.

We now have

$$\frac{1}{S_T}|\hat{\Phi} - \bar{\Phi}| = \frac{1}{S_T}\left|\sum_{j=-T+1}^{T-1} k\left(\frac{j}{S_T}\right)\{\hat{\gamma}(j) - \bar{\gamma}(j)\}\right| \leqslant \sup_j|\hat{\gamma}(j) - \bar{\gamma}(j)|\left(\frac{1}{S_T}\sum_{j=-T+1}^{T-1}\left|k\left(\frac{j}{S_T}\right)\right|\right) \xrightarrow{p} 0 \tag{A.31}$$

because $(1/S_T)\sum_{j=-T+1}^{T-1}|k(j/S_T)| \to \int_{-\infty}^{\infty}|k(x)|\mathrm{d}x < \infty$ and

$$\sup_j |\hat{\gamma}(j) - \bar{\gamma}(j)| = \sup_j \left|\frac{1}{T}\sum_{t=|j|+1}^{T} (\eta_t(\hat{\tau})\eta_{t-|j|}(\hat{\tau}) - \eta_t(\tau_0)\eta_{t-|j|}(\tau_0))\right|$$
$$\leqslant 2\left(\sup_{x\in\mathscr{X}_T} \|\hat{\tau}(x) - \tau_0(x)\|\right)\left[\frac{1}{T}\sum_{t=1}^{T} \sup_{\tau\in\mathscr{T}^*}\left\|\frac{\partial}{\partial\tau}\eta_t(\tau)\right\|^2\right]^{1/2} \times \left[\frac{1}{T}\sum_{t=1}^{T} \sup_{\tau\in\mathscr{T}^*} |\eta_t(\tau)|^2\right]^{1/2} \tag{A.32}$$
$$+ \left(\sup_{x\in\mathscr{X}_T} \|\hat{\tau}(x) - \tau_0(x)\|^2\right)\left[\frac{1}{T}\sum_{t=1}^{T} \sup_{\tau\in\mathscr{T}^*}\left\|\frac{\partial}{\partial\tau}\eta_t(\tau)\right\|^2\right]$$
$$= O_p(T^{-1/2}b_{1T}^{-3}d_T^{-7\rho}) + O_p(b_{2T}^{\omega-2}d_T^{-7\rho}) \tag{A.33}$$
$$\xrightarrow{p} 0, \tag{A.34}$$

as is required, where the first inequality (A.32) holds with probability that goes to one using the fact that $\Pr(\hat{\tau}\in\mathscr{T}^*)\to 1$ (see the arguments following Eq. (A.15)). The second equality (A.33), on the other hand, uses Lemma A.1 and the results that $(1/T)\sum_{t=1}^{T}\sup_{\tau\in\mathscr{T}^*}\|(\partial/\partial\tau)\eta_t(\tau)\|^2 = O_p(d_T^{-4\varrho})$ and $(1/T)\sum_{t=1}^{T}\sup_{\tau\in\mathscr{T}^*}\|\eta_t(\tau)\|^2 = O_p(d_T^{-2\varrho})$ which hold by Assumption A3(c) and the definition of $\mathscr{T}^*$. Finally, the convergence to zero Assumption (A.34) holds by Assumption A5.

(b) The proof of part (b) of Corollary is similar to that of part (a) using the arguments with T replaced by n in appropriate places and the following result:

$$\sup_j |\hat{\gamma}_2(j) - \bar{\gamma}_2(j)| \leqslant \left\{ 2\frac{T^{\kappa}\sup_{x\in\mathscr{X}_T}|D\hat{m}(x) - Dm_0(x)|}{n^{\phi}\min_{\{t:X_{t-1}\in\mathscr{X}_T\}}|Dm^*(X_{t-1})|} \times\left[\frac{1}{n}\sum_{t=1}^{n}(\ln|Dm_0(X_{t-1})| - \lambda)^2\mathbf{1}_t\right]^{1/2} + \frac{T^{2\kappa}\sup_{x\in\mathscr{X}_T}|D\hat{m}(x) - Dm_0(x)|^2}{n^{2\phi}\min_{\{t:X_{t-1}\in\mathscr{X}_T\}}|Dm^*(X_{t-1})|^2}\right\} O(1) \xrightarrow{p} 0, \quad \text{(A.35)}$$

where the convergence to zero holds by Lemma A.1, Assumptions A3(b), A3(c)*, and A6*. □

Appendix B. Proof of Theorem 2.

We first note that the results of Lemma A.1 can be extended to cover the general case $k \geqslant 1$ Specifically, one can establish the results of Lemma A.1 with x and $\mathscr{X}_T$ replaced by z and $\mathscr{Z}_T = \mathscr{X}_T\times\cdots\times\mathscr{X}_T$, respectively, and $\mu+1$ that appears in the first terms of the right-hand sides of Lemma A.1(a) and (b) replaced by $\mu+k$. In the remaining proof, we use this modified version of Lemma A.1.

(a) Define

$$\lambda_T = \frac{1}{T}\ln\left\|\prod_{t=1}^{T^*} J_{T-t}\right\|. \quad \text{(B.1)}$$

By rearranging terms we have,

$$\sqrt{T}(\hat{\lambda} - \lambda) = \sqrt{T}(\hat{\lambda} - \lambda_T) + \sqrt{T}(\lambda_T - \lambda)$$

$$= \frac{1}{\sqrt{T}} \left[\ln \left\| \prod_{t=1}^{T^*} \hat{J}_{T-t} \right\| - \ln \left\| \prod_{t=1}^{T^*} J_{T-t} \right\| \right]$$

$$+ \sqrt{T} \left[\frac{1}{T} \ln \left\| \prod_{t=1}^{T^*} J_{T-t} \right\| - \lambda \right] \equiv E_T + G^*_{T2}, \quad say. \tag{B.2}$$

Below we establish the following result:

$$E_T = G_{T1} + o_p(1), \tag{B.3}$$

where G_{T1} is as defined in Eq. (22). This result and Assumption B7 give the desired result using the result that $G^*_{T2} - G_{T2} = o_p(1)$, which follows by similar arguments using Eq. (A.24).

We now verify Eq. (B.3). A two-term Taylor expansion of A_T (defined in Eq. (B.2))about Δm_t gives

$$|E_T| = \frac{1}{\sqrt{T}} \left[\ln \left\| \prod_{t=1}^{T^*} \hat{J}_{T-t} \right\| - \ln \left\| \prod_{t=1}^{T^*} J_{T-t} \right\| \right]$$

$$= \frac{1}{\sqrt{T}} \sum_{t=1}^{T^*} F_{t-1}(J_{T-1}, \dots, J_0)[\Delta\hat{m}(Z_t) - \Delta m(Z_t)]$$

$$+ \frac{1}{\sqrt{T}} \sum_{t=1}^{T^*} \sum_{s=1}^{T^*} [\Delta\hat{m}(Z_t) - \Delta m(Z_t)]' Q_{ts}(J^*_{T-1}, \dots, J^*_0)$$

$$\times [\Delta\hat{m}(Z_s) - \Delta m(Z_s)] \equiv E_{1T} + E_{2T}, \quad \text{say}, \tag{B.4}$$

where $\Delta\hat{m}(Z_t) = (\Delta\hat{m}_{1t}, \Delta\hat{m}_{2t}, \dots, \Delta\hat{m}_{kt})'$ and the elements of J^*_t lie between those of $\hat{J}_t$ and J_t for $t = 0, \dots, T^* - 1$.

Consider E_{2T} first. Note that, for any $\varepsilon > 0$,

$$|E_{2T}| \leqslant T^{1/2} \left[\sup_{z \in \mathscr{Z}_T} \|\Delta\hat{m}(z) - \Delta m(z)\| \right]^2 \frac{1}{T} \sum_{t=1}^{T^*} \sum_{s=1}^{T^*} \|Q_{ts}(J^*_{T-1}, \dots, J^*_0)\| \xrightarrow{p} 0, \tag{B.5}$$

where the convergence to zero holds by Lemma A.1, and Assumptions B5 and B6(c).

Now consider E_{1T}. By adding and subtracting terms, we have

$$E_{1T} = \frac{1}{\sqrt{T}} \sum_{t=1}^{T^*} F_{t-1}(J_{T-1}, \dots, J_0)[\Delta\hat{m}(Z_t) - \Delta m(Z_t)]$$

$$= \frac{1}{\sqrt{T}} \sum_{t=1}^{T^*} (F_{t-1}(J_{T-1}, \ldots, J_0) - \mathrm{E}[F_{t-1}(J_{T-1}, \ldots, J_0) | \mathscr{G}_{t-l}^{t+m}])$$

$$\times [\Delta\hat{m}(Z_t) - \Delta m(Z_t)] + \frac{1}{\sqrt{T}} \sum_{t=1}^{T^*} \mathrm{E}[F_{t-1}(J_{T-1}, \ldots, J_0) | \mathscr{G}_{t-l}^{t+m}]$$

$$\times [\Delta\hat{m}(Z_t) - \Delta m(Z_t)] = E_{1T}^* + E_{1T}^{**}, \quad \text{say.} \tag{B.6}$$

Note that E_{1T}^* is $\mathrm{o_p}(1)$ because

$$|E_{1T}^*| \leqslant T^{1/4} \left[\sup_{z \in \mathscr{Z}_T} \| \Delta\hat{m}(z) - \Delta m(z) \| \right]$$

$$\times T^{-1/4} \left[\frac{1}{T} \sum_{t=1}^{T^*} \| F_{t-1}(J_{T-1}, \ldots, J_0) \right.$$

$$\left. - \mathrm{E}[F_{t-1}(J_{T-1}, \ldots, J_0) | \mathscr{G}_{t-l-1}^{t+m}] \| \right] \tag{B.7}$$

and the first term on the rhs of Eq. (B.7) is $\mathrm{o_p}(1)$ by Lemma A.1(b) (modified as described above) and the second term is $\mathrm{O_p}(1)$ by Assumption B6(a). We next consider E_{1T}^{**}. Define $F_T(z) = \mathrm{E}[F_{t-1}(J_{T-1}, \ldots, J_0) \,|\, Z_t = z]$ for $z \in \mathbb{R}^k$. We have

$$E_{1T}^{**} = \frac{1}{\sqrt{T}} \sum_{t=1}^{T^*} \mathrm{E}[F_{t-1}(J_{T-1}, \ldots, J_0) | \mathscr{G}_{t-l-1}^{t+m}][\Delta\hat{m}(Z_t) - \Delta m(Z_t)]$$

$$= v_T(\hat{\tau}) - v_T(\tau_0) + E_{1T}^{***},$$

where

$$v_T(\tau) = \frac{1}{\sqrt{T}} \sum_{t=1}^{T} [r(W_t, \tau) - Er(W_t, \tau)], \tag{B.8}$$

$$\mathrm{r}(W_t, \tau) = 1(z \in \mathscr{Z}_T) \mathrm{E}[F_{t-1}(J_{T-1}, \ldots, J_0) | \mathscr{G}_{t-l-1}^{t+m}] \tau(Z_t),$$

$$E_{1T}^{***} = \sqrt{T} \int_{\mathscr{Z}_T} F_T(z)[\Delta\hat{m}(z) - \Delta m(z)] f(z) \mathrm{d}z,$$

$W_t = (Z_{t-l}, \ldots, Z_{t+m})$, $\hat{\tau}(\cdot) = \Delta\hat{m}(\cdot)$, and $\tau_0(\cdot) = \Delta m(\cdot)$. By arguments similar to those which verified the stochastic equicontinuity results in Eqs. (A.12), (A.13), (A.14), (A.15) and (A.16), we can verify $v_T(\hat{\tau}) - v_T(\tau_0) = \mathrm{o_p}(1)$.

We finally consider E_{1T}^{***}. We have

$$E_{1T}^{***} = \sqrt{T}\int_{\mathscr{Z}_T} F_T(z)[\Delta\hat{m}(z) - \Delta m(z)]f(z)\,\mathrm{d}z$$

$$= -\sqrt{T}\int_{\mathscr{Z}_T}\left[\sum_{j=1}^{k} D^{e_j}\{F_{Tj}(z)f(z)\}\right][\hat{m}(z) - m(z)]\,\mathrm{d}z$$

$$+ \sqrt{T}\sum_{x\in\partial\mathscr{Z}_T} F_T(z)f(z)[\hat{m}(x) - m_0(x)]$$

$$= -\sqrt{T}\int_{\mathscr{Z}_T}\left[\frac{1}{f(z)}\sum_{j=1}^{k} D^{e_j}\{F_{Tj}(z)f(z)\}\right][\{\hat{m}(z) - m_0(z)\}\hat{f}(z)]\mathrm{d}z$$

$$+ \sqrt{T}\int_{\mathscr{Z}_T}\left[\sum_{j=1}^{k} D^{e_j}\{F_{Tj}(z)f(z)\}\right]\left[\frac{1}{\hat{f}(z)} - \frac{1}{f(z)}\right]$$

$$\times [\{\hat{m}(z) - m_0(z)\}\hat{f}(z)]\mathrm{d}z + \sqrt{T}\sum_{x\in\partial\mathscr{Z}_T} F_T(z)f(z)[\hat{m}(x) - m_0(x)]$$

$$= -\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\left[\frac{1}{f(Z_{t-1})}\sum_{j=1}^{k} D^{e_j}\{F_{Tj}(Z_{t-1})f(Z_{t-1})\}\right][X_t - m_0(Z_{t-1})]$$

$$+ \mathrm{o_p}(1), \tag{B.9}$$

where the second equality holds by integration by parts, the third equality follows by rearranging terms, and the last equality holds using the similar arguments as in Eqs. (A.19), (A.20), (A.21) and (A.22). □

(b) Consider expression (B.2) with T and T^* replaced by n wherever it appears. It suffices to show that E_T (with T replaced by n) is $\mathrm{o_p}(1)$. A one-term Taylor expansion gives

$$|E_T| = \frac{1}{\sqrt{n}}\left|\ln\left\|\prod_{t=1}^{n}\hat{J}_{n-t}\right\| - \ln\left\|\prod_{t=1}^{n} J_{n-t}\right\|\right|$$

$$= \left|\frac{1}{\sqrt{n}}\sum_{t=1}^{n} F_{t-1}(J_{n-1}^*,\dots,J_0^*)[\Delta\hat{m}(Z_t) - \Delta m(Z_t)]\right| \tag{B.10}$$

$$\leqslant T^{-\kappa} n^{1/2+\phi}\left[T^{\kappa} \sup_{z \in \mathscr{Z}_T} \| \Delta\hat{m}(z) - \Delta m(z) \| \right]$$

$$\times \left(n^{-\phi} \max_{1 \leqslant t \leqslant n} \| F_{t-1}(J^*_{n-1}, \dots, J^*_0) \| \right) \xrightarrow{p} 0,$$

where $\Delta\hat{m}(Z_t) = (\Delta\hat{m}_{1t}, \Delta\hat{m}_{2t}, \dots, \Delta\hat{m}_{kt})'$ and the elements of J^*_t lie between those of $\hat{J}_t$ and J_t for $t = 0, \dots, n-1$ and the convergence to zero holds by Lemma A.1 and Assumption B6(c)*. □

References

Andrews, D.W.K., 1989. Stochastic equicontinuity for semiparametric models. Unpublished manuscript, Cowles Foundation, Yale University.

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

Andrews, D.W.K., 1994. Asymptotics for semiparametric econometric models via stochastic equicontinuity. Econometrica 62, 43–72.

Andrews, D.W.K., 1995. Nonparametric Kernel estimation for semiparametric models. Econometric Theory 11, 560–596.

Andrews, D.W.K., Monahan, J.C., 1992. An improved heteroskedasticity and autocorrelation consistent covariance matrix estimator. Econometrica 60, 953–966.

Bierens, H., 1987. Kernel estimators of regression functions. In: Bewley, T.F. (Ed.), Advances in Econometrics: Fifth World Congress, vol 1. Cambridge University Press, Cambridge.

Brock, W.A., 1986. Distinguishing random and deterministic systems. Journal of Economic Theory 40, 168–195.

Chan, K.S., Tong, H., 1994. A note on noisy chaos. Journal of the Royal Statistical Society Series B 56, 301–311.

Cheng, B., Tong, H., 1992. On consistent nonparametric order determination and chaos. Journal of the Royal Statistical Society, Series B 54, 427–449.

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

De Jong, R.M., Davidson, J., 1996. Consistency of kernel estimators of heteroskedastic and autocorrelated covariance matrices. CentER Discussion paper no. 9652, Tilburg University.

Doukhan, P., 1994. Mixing: Properties and Examples. Springer, Berlin.

Eckmann, J.P., Kamphorst, S.O., Ruelle, D., Ciliberto, S., 1986. Lyapunov exponents from time series. Physics Review A 34, 4971–4979.

Folland, G.B., 1984, Real Analysis: Modern Techniques and their Applications. Wiley, New York.

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

Gallant, A.R., 1987. Nonlinear Statistical Models. Wiley, New York.

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.

Gnedenko, B.V., 1943. Sur la distribution limite du term maximum d'une série aléatoire. Annals of Mathematical Statistics 44, 423–453.

Hansen, B.E., 1992. Consistent covariance matrix estimation for dependent heterogeneous processes. Econometrica 60, 967–972.

Härdle, W., Hart, J., Marron, J.S., Tsybakov, A.B., 1992. Bandwidth choice for average derivative estimation. Journal of the American Statistical Association 87, 218–226.

Härdle, W., Stoker, T., 1989. Estimating smooth multiple regression by the method of average derivatives. Journal of the American Statistical Association 84, 986–995.

Hengartner, N.W., Linton, O.B., 1996. Nonparametric regression estimation at design poles and zeros. Canadian Journal of Statistics 24, 583–591.

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

Leadbetter, M.R., Rootzén, H., 1988. Extremal theory for stochastic processes. Annals of Probability 16, 431–478.

LeBaron B., 1997. Has chaos theory found any useful application in the social sciences. Manuscript, University of Wisconsin, http://www.ssc.wisc.edu/~blebaron.

Loynes, R.M., 1965. Extreme values in uniformly mixing stationary stochastic processes. Annals of Mathematical Statistics 66, 993–999.

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

Masry, E., 1996. Multivariate local polynomial regression for time series: Uniform strong consistency and rates. Journal of Time Series Analysis 17, 571–599.

Newey, W.K., 1994. The asymptotic variance of semiparametric estimators. Econometrica 62, 1349–1382.

Newey, W.K., 1995. Convergence rates for series estimators. in: Maddala, G.S., Phillips, P.C.B., Srinivasan, T.N. (Eds.), Advances in Econometrics and Quantitative Econometrics. Basil Blackwell, Oxford.

Newey, W.K., West, K.D., 1987. A simple positive-definite, heteroskedastic and autocorrelation consistent covariance matrix. Econometrica 55, 703–708.

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

Ouliaris, S., Phillips, P.C.B., 1994. COINT 2.0: Gauss procedures for cointegrated regressions. Predicta Software Inc, 133 Concord Drive, Madison CT 06443.

Robinson, P.M., 1983. Nonparametric estimators for time series. Journal of Time Series Analysis 4, 185–208.

Tong, H., 1990. Non-linear Time Series: A Dynamical Systems Approach. Clarendon Press, Oxford.

Watson, G.S., 1954. Extreme values in samples from m-dependent stationary stochastic processes. The Annals of Mathematical Statistics 25, 798–800.

White, H., 1984. Asymptotic Theory for Econometricians. Academic Press, New York.