

COWLES FOUNDATION FOR RESEARCH IN ECONOMICS
AT YALE UNIVERSITY

Box 2125, Yale Station
New Haven, Connecticut 06520

COWLES FOUNDATION DISCUSSION PAPER NO. 1130R

Note: Cowles Foundation Discussion Papers are preliminary materials circulated to stimulate discussion and critical comment. Requests for single copies of a Paper will be filled by the Cowles Foundation within the limits of the supply. References in publications to Discussion Papers (other than mere acknowledgment by a writer that he has access to such unpublished material) should be cleared with the author to protect the tentative character of these papers.

THE ASYMPTOTIC DISTRIBUTION OF NONPARAMETRIC ESTIMATES
OF THE LYAPUNOV EXPONENT FOR STOCHASTIC TIME SERIES

Yoon-Jae Whang and Oliver Linton

Revised October 1997

THE ASYMPTOTIC DISTRIBUTION OF NONPARAMETRIC ESTIMATES OF THE LYAPUNOV EXPONENT FOR STOCHASTIC TIME SERIES

YOON-JAE WHANG AND OLIVER LINTON

EWHA UNIVERSITY AND YALE UNIVERSITY

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.

Key Words: Chaos; Kernel; Nonlinear dynamics; Nonparametric regression; Semiparametric.

JEL Classification Codes: C13, C14, C22.

SEPTEMBER 25TH, 1997

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, 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, as Nychka, Ellner, Gallant and McCaffrey (1992) say

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, Kamphorst, Ruelle, and Ciliberto (1986), of testing for chaos based on estimating the Lyapunov exponent λ using nonpara-

metric regression methods [see also McCaffrey, Ellner, Gallant, and Nychka (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, see 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 Feigenbaum 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 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, p309), 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 μ -th 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

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

where $m_0: \mathbb{R}^k \rightarrow \mathbb{R}$ 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 $M: \mathbb{R}^k \rightarrow \mathbb{R}^k$ such that

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

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

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

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

$$(4) \quad \lambda \stackrel{\text{a.s.}}{=} \lim_{T \rightarrow \infty} \frac{1}{T} \ln \left\| \prod_{t=1}^T J_{T-t} \right\|,$$

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

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

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. The model (5) can be expressed in state space form as in (2) with the addition of the error vector $U_t = (\varepsilon_t, 0, \dots, 0)' \in \mathbb{R}^k$. An alternative model here, called measurement noise, is that X_t is defined in (1), but that we observe

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

with ε_t as in (5).¹ In this case, we can find an m_0^* for which (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 (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 $\sigma \rightarrow 0$, but true even for moderate σ .² For stationary linear autoregressions

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

² As $\sigma \rightarrow 0$, $\lambda_\sigma \rightarrow \lambda_0$.

$\lambda < 0$, 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 $\widehat{m}(\cdot)$ be a consistent smoothing-based estimate of $m_0(\cdot)$. Suitable estimation methods include: kernels, nearest neighbors, 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 $\widehat{m}(\cdot)$ [the estimate itself will be computed with the full sample]. We assume that Z_t is supported on a bounded set \mathcal{Z} and that $\partial\mathcal{Z}$ is its topological boundary, and trim out observations too close to this boundary; specifically, we keep only the T^* sample values that satisfy

$$(6) \quad \inf_{z_0 \in \partial\mathcal{Z}} d(Z_t, z_0) \geq d_T,$$

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

$$(7) \quad \widehat{\lambda} = \frac{1}{n} \ln \left\| \prod_{t=1}^n \widehat{J}_{n-t} \right\|, \text{ where}$$

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

where, $\Delta\widehat{m}_{jt} = D^{e_j}\widehat{m}(Z_t)$ 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

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

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 $\widehat{m}(\cdot)$ of $m_0(\cdot)$, which is defined as follows

$$(9) \quad \widehat{m}(z) = \frac{\widehat{g}(z)}{\widehat{f}(z)}, \quad \text{where}$$

$$(10) \quad \begin{aligned} \widehat{g}(z) &= \frac{1}{T\widehat{b}_T^k} \sum_{t=1}^T \widehat{K} \left(\frac{z - Z_{t-1}}{\widehat{b}_T} \right) X_t \quad ; \quad \widehat{f}(z) = \frac{1}{T\widehat{b}_T^k} \sum_{t=1}^T \widehat{K} \left(\frac{z - Z_{t-1}}{\widehat{b}_T} \right), \\ \widehat{K}(z) &= \left[\det(\widehat{\Omega}) \right]^{-1/2} K(\widehat{\Omega}^{-1/2}z) \quad ; \quad \widehat{\Omega} = \frac{1}{T} \sum_{t=1}^T (Z_{t-1} - \overline{Z}_T)(Z_{t-1} - \overline{Z}_T)', \\ \overline{Z}_T &= \frac{1}{T} \sum_{t=1}^T Z_{t-1}. \end{aligned}$$

Derivatives of $\widehat{m}(z)$ and $\widehat{f}(z)$ provide suitable estimates of the corresponding derivatives of $m(z)$ and $f(z)$. The kernel $K(\cdot)$ is a non-random real function on \mathbb{R}^k , and the bandwidth parameter \widehat{b}_T is a positive constant or scalar random variable satisfying the assumptions below. The data dependent scale matrix $\widehat{\Omega}$ is used to obtain invariance of the estimator under one-to-one transformations of the form $Z_{t-1} \rightarrow BZ_{t-1} + \gamma$, 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 $\mathcal{K}_{k,\delta,\nu}$ defined for non-negative integers, k , δ , and ν , 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 $\nu(\geq \delta)$ 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 $\widehat{\lambda}$ for general $k \geq 2$ will be discussed later.

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

$$(11) \quad \lambda \stackrel{\text{a.s.}}{=} \lim_{T \rightarrow \infty} \frac{1}{2T} \sum_{t=1}^T \ln [Dm_0(X_{t-1})^2] \quad \text{and}$$

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

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

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

ASSUMPTION A.

- (1) (a) $\{X_t : t \geq 1\}$ 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 $t \geq 1$, X_t lies in an open bounded set $\mathcal{X} (\subset \mathbb{R})$ with minimally smooth boundary denoted $\partial\mathcal{X}$.
- (c) $\{\varepsilon_t : t \geq 1\}$ is a sequence of martingale differences with $E(\varepsilon_t | \mathcal{F}_{-\infty}^{t-1}) = 0$ a.s. and $E(|\varepsilon_t|^{2r}) < \infty$ for all $t \geq 1$, where $\mathcal{F}_s^t = \sigma(X_s, \dots, X_t)$ is the σ -algebra generated by (X_s, \dots, X_t) .

⁴ Note also that

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

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

The martingale difference assumption for the noise (Assumption 1(c)) is more general than the i.i.d. assumption that 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, see Härdle, Hart, Marron, and Tsybakov (1992, p223). 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 non-zero coefficient, and other nonlinear processes.

In Assumption A1(b) the set \mathcal{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 1(a)-(b) in Appendix] for some sufficiently large $\kappa \geq 1/4$, 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 \mathcal{X} and $\sup_{x \in \mathcal{X}} |D^\mu f(x)| < \infty$ for all μ with $\mu \leq \omega$ where $\omega \geq 2$ is a positive integer.

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

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

$$\mu \leq \omega.$$

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

$$\text{A(4)}^* \quad K(\cdot) \in \mathcal{K}_{1,1,\omega}.$$

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

$$\text{A(6)}^* \quad n \rightarrow \infty \text{ and } n = O(T^{\frac{2\kappa}{1+2\phi}}).$$

Our new assumption A2(c)^* that X_t has positive density on its bounded support strengthens 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), 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 $(\min_{1 \leq i \leq n} |Y_i|)^{-1}$? We have for any $0 < \epsilon < \infty$,

$$\begin{aligned} \text{Pr} \left(\frac{1}{n^\phi \min_{1 \leq i \leq n} |Y_i|} \leq \epsilon \right) &= \{1 - F_Y(\epsilon^{-1} n^{-\phi}) + F_Y(-\epsilon^{-1} n^{-\phi})\}^n \\ &= \left\{ 1 - \frac{\xi_n(\epsilon)}{n} \right\}^n \\ (13) \quad &\rightarrow e^{-\xi(\epsilon)}, \end{aligned}$$

where by Taylor expansion

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

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, $M_n = (n \min_{1 \leq i \leq n} |Y_i|)^{-1} = O_p(1)$ 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 \widehat{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 \widehat{M}_n , where $\widehat{M}_n = \left(n \min_{1 \leq i \leq n} |\widehat{Y}_i|\right)^{-1}$, are the same.⁶

In our case, we associate Y with $Dm_0(X)$. We therefore expect that assumption 3c* 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 Feigenbaum mapping

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

The stationary ergodic density for this process is the well-known arcsine law $f(x) = 1 / \sqrt{\pi^2 x(1-x)}$ on $[0, 1]$, see Tong (1990, p60).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 $\phi = 1$.

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 $\widehat{\lambda}$ 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 Feigenbaum mapping $f(0) = f(1) = \infty$. This boundedness

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 .

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

$$\begin{aligned}
 \Phi &= \lim_{n \rightarrow \infty} \text{var} \left[\frac{1}{\sqrt{n}} \sum_{t=1}^n \eta_t \right] \text{ and} \\
 \Phi_j &= \lim_{n \rightarrow \infty} \text{var} \left[\frac{1}{\sqrt{n}} \sum_{t=1}^n \eta_{jt} \right], \quad j = 1, 2, \text{ where} \\
 \eta_t &= \eta_{1t} + \eta_{2t}; \quad \eta_{1t} = v(X_{t-1})\varepsilon_t; \text{ and } \eta_{2t} = \ln |Dm_0(X_{t-1})| - \lambda.
 \end{aligned}
 \tag{15}$$

We now have the main result of this paper.

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

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

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

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

REMARKS:

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.

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.

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 equation (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 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 (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

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

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

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

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

$$\begin{aligned} \hat{\Phi} &= \sum_{j=-n+1}^{n-1} k\left(\frac{j}{S_n}\right) \hat{\gamma}(j) \text{ and } \hat{\Phi}_\ell = \sum_{j=-n+1}^{n-1} k\left(\frac{j}{S_n}\right) \hat{\gamma}_\ell(j), \ell = 1, 2, \text{ where} \\ (16) \quad \hat{\gamma}(j) &= \frac{1}{n} \sum_{t=|j|+1}^n \hat{\eta}_t \hat{\eta}_{t-|j|}; \quad \hat{\gamma}_\ell(j) = \frac{1}{n} \sum_{t=|j|+1}^n \hat{\eta}_{\ell t} \hat{\eta}_{\ell, t-|j|} \end{aligned}$$

$$\begin{aligned}\widehat{\eta}_{1t} &= \{X_t - \widehat{m}(X_{t-1})\} \left\{ \frac{D^2 \widehat{m}(X_{t-1})}{[D\widehat{m}(X_{t-1})]^2} - \frac{Df(X_{t-1})}{[D\widehat{m}(X_{t-1})]f(X_{t-1})} \right\} \mathbf{1}_t; \\ \widehat{\eta}_{2t} &= \left\{ \ln |D\widehat{m}(X_{t-1})| - \widehat{\lambda} \right\} \mathbf{1}_t; \text{ and } \widehat{\eta}_t = \widehat{\eta}_{1t} + \widehat{\eta}_{2t}.\end{aligned}$$

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 $\widehat{\Phi}$ with $n = T$, whereas we use $\widehat{\Phi}_2$ with $n < T$ under the conditions of Theorem 1(b).

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

ASSUMPTION S.

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

$$\begin{aligned}\mathcal{K}_1 &= \left\{ k : \mathbb{R} \rightarrow [-1, 1] \mid k(0) = 1, k(-x) = k(x) \forall x \in \mathbb{R}, \int_{-\infty}^{\infty} k^2(x) dx < \infty, \right. \\ &\quad \left. k(\cdot) \text{ is continuous at } 0 \text{ and at all but a finite number of points} \right\}.\end{aligned}$$

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

We also use the following bandwidth conditions:

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

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 $\mathcal{K}_{1,1,\omega} \cap \mathcal{K}_{1,q+1,q+1}$ in A4 replaced by $3r/(r-2), \max\{\omega, q+2\}$, and $\mathcal{K}_{1,q+2,q+2}$ respectively. Then,

$$\widehat{\Phi} \rightarrow_p \Phi.$$

(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,

$$\widehat{\Phi}_2 \rightarrow_p \Phi_2.$$

REMARKS:

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 $k \geq 1$.

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

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

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

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

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

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

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

Define also

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

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

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 $d(x, y) =$

$$\left\{ x' \widehat{\Omega}^{-1} y \right\}^{1/2}.$$

We make the following assumptions.

ASSUMPTION B.

- (1) (a) $\{X_t : t \geq 1\}$ 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 $t \geq 1$, X_t lies in an open bounded set $\mathcal{X} (\subset \mathbb{R}^k)$ with minimally smooth boundary denoted $\partial \mathcal{X}$.
- (c) $\{\varepsilon_t : t \geq 1\}$ are i.i.d. with $E(\varepsilon_t | \mathcal{G}_{-\infty}^{t-1}) = 0$ and $E(|\varepsilon_t|^{2r}) < \infty$ for all $t \geq 1$, where $\mathcal{G}_s^t = \sigma(Z_s, \dots, Z_t)$ 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 $\mathcal{Z} = \mathcal{X} \times \dots \times \mathcal{X}$ and $\sup_{z \in \mathcal{Z}} |D^\mu f(z)| < \infty$ for all μ with $|\mu| \leq \max\{\omega, q+1\}$, where $\omega \geq 2$ and $q \geq k/2$ are positive integers that also appear in the other assumptions below.

- (c) (i) The density $f(z) > 0$ for all $z \in \mathcal{Z}$. Furthermore, there exist strictly positive constants $\underline{A}, \bar{A}, \varrho$, and Δ such that $\underline{A} \leq \|z - z_o\|^{-\varrho} f(z) \leq \bar{A}$ for all $z_o \in \partial\mathcal{Z}$ and all z such that $\|z - z_o\| \leq \Delta$.
- (ii) $\int |v(z)|^{2r/(r-1)} f(z) dz < \infty$.
- (3) $D^\mu m_0(z)$ exists and is continuous on \mathcal{Z} and $\sup_{z \in \mathcal{Z}} |D^\mu m_0(z)| < \infty$ for all μ with $|\mu| \leq \max\{\omega, q+1\}$.
- (4) (a) $K(\cdot) \in \mathcal{K}_{k,1,\omega} \cap \mathcal{K}_{k,q+1,q+1}$.
- (b) $\lambda_{\min} \{E(Z_{t-1} - EZ_{t-1})(Z_{t-1} - EZ_{t-1})'\} > 0$.
- (5) The data-dependent bandwidth parameters $\{\hat{b}_T : T \geq 1\}$ satisfy $C_1 b_{1T} \leq \hat{b}_T \leq C_2 b_{2T}$ with probability tending to one for some sequences of bounded positive constants $\{b_{1T} : T \geq 1\}$ and $\{b_{2T} : T \geq 1\}$ and some positive finite constants C_1, C_2 that satisfy: (a) $T^{\frac{1}{2}} b_{1T}^{\max\{2(k+1), q+k+1\}} d_T^{\max\{6, q+3\}\varrho} \rightarrow \infty$; (b) $b_{2T}^{\omega-(q+1)} d_T^{-(q+3)\varrho} \rightarrow 0$; (c) $T^{\frac{1}{2}} b_{2T}^\omega d_T^{-6\varrho} \rightarrow 0$; (d) $b_{1T}^{-k} b_{2T}^\omega d_T^{-4\varrho} \rightarrow 0$; (e) the trimming sequence d_T satisfies $d_T \geq C_2 b_{2T}$, $T d_T^{2+2\varrho} \rightarrow 0$, and $d_T^{2\varrho}/b_{1T} \rightarrow 0$.
- (6) (a) $\frac{1}{T} \sum_{t=1}^T \|F_{t-1}(J_{T-1}, \dots, J_0) - E[F_{t-1}(J_{T-1}, \dots, J_0) | \mathcal{G}_{t-l-1}^{t+m-1}]\| = O_p(T^{-\frac{1}{4}})$ for some non-negative integers $l, m < \infty$.
- (b) $\sup_{t \leq T, T \geq 1} E(\|F_{t-1}(J_{T-1}, \dots, J_0)\|^r) < \infty$ and $\sup_{T \geq 1} |F_T(z_0)| < \infty$ for all $z_0 \in \partial\mathcal{Z}$.
- (c) $\frac{1}{T} \sum_{t=1}^T \sum_{s=1}^T \|Q_{ts}(J_{T-1}, \dots, J_0)\| = O_p(1)$.
- (7) For some 2×2 positive semi-definite matrix Ξ , $G_T = (G'_{T1}, G'_{T2})' \Rightarrow N(0, \Xi)$.

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 $\mathcal{G}_{t-l-1}^{t+m-1} = \mathcal{G}_{t-1}^{t-1} = \sigma(Z_{t-1}) = \sigma(X_{t-1})$. Assumption B6(b) and (c) are also satisfied for the case $k = 1$ under Assumption A3(c), because in that case we have

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

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

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 $k \geq 2$, 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

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

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 \mathcal{Z} and $\sup_{z \in \mathcal{Z}} |D^\mu f(z)| < \infty$ for all μ with $\mu \leq \omega$ where $\omega \geq 2$ is a positive integer.

(b)* The marginal density f satisfies $\inf_{z \in \mathcal{Z}} f(z) > 0$.

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

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

B(5)* The data-dependent bandwidth parameters $\{\hat{b}_T : T \geq 1\}$ satisfy $C_1 b_{1T} \leq \hat{b}_T \leq C_2 b_{2T}$ with probability tending to one for some sequences of bounded positive constants $\{b_{1T} : T \geq 1\}$ and $\{b_{2T} : T \geq 1\}$ and some positive finite constants C_1, C_2 that satisfy: (a)

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

$T^{\frac{1}{2}-\kappa} b_{1T}^{k+1} d^{3\rho} \rightarrow \infty$; (b) $T^\kappa b_{2T}^{\omega-1} d_T^{-3\varrho} \rightarrow 0$; (c) the trimming sequence d_T satisfies $d_T \geq C_2 b_{2T}$, $T d_T^{2+2\varrho} \rightarrow 0$ for some $0 < \kappa < \frac{1}{2}$.

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

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

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

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

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

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,*

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

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

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

where

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

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

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

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

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, $\widehat{F}_{j,t}$ 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 (24) and use $\widehat{\eta}_{2t} = \widehat{\xi}_s - \widehat{\lambda}$.

4. Numerical Results. We first analyze how our procedure works on the following AR(1) process,

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

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 the estimator (9) with a quartic kernel $K(u) = 15(1 - u^2)^2 \mathbf{1}(|u| \leq 1) / 16$ and bandwidth $h = \gamma * \text{range} \{X_t\}_{t=1}^T$, 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 $\widehat{\lambda}$.

*** TABLES 1,2,3 HERE ***

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 $h = 0.4 \times \text{range}$ 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 Figure 1 below 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.

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.

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

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 the 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.

TABLE 4 HERE

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 Feigenbaum map with system noise:

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

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

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

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 $\lambda = 0.692$.

*** TABLES 5 HERE ***

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 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 Figure 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, 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 behavior. 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

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

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

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.

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

$$(25) \quad \alpha(s) = \sup_{t \geq 1} \sup_{A \in \mathcal{F}_1^t, B \in \mathcal{F}_{t+s}^\infty} |\Pr(A \cap B) - \Pr(A) \Pr(B)| \text{ for } s \geq 1.$$

Then, $\{X_t : t \geq 1\}$ 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:

$$\mathcal{K}_{k,\delta,\nu} = \{K(\cdot) : [-1, 1]^k \rightarrow \mathbb{R} \mid \int K(z) dz = 1, \int z^\mu K(z) dz = 0$$

$$\forall 1 \leq |\mu| \leq \nu - \delta - 1, \int |z^\mu K(z)| dz < \infty \forall |\mu| = \nu - \delta,$$

$$D^\mu K(z) \rightarrow 0 \text{ as } \|z\| \rightarrow \infty \forall \mu \text{ with } |\mu| < \delta,$$

$D^\mu K(z)$ is absolutely integrable and has Fourier transform

$$\Psi_\mu(r) = (2\pi)^k \int \exp(ir'z) D^\mu K(z) dz \text{ that satisfies}$$

$$\int (1 + \|r\|) \sup_{b \geq 1} \|\Psi_\mu(br)\| dr < \infty \forall \mu \text{ with } |\mu| \leq \delta, \text{ and}$$

$$\sup_{z \in \mathbb{R}^k} |D^{\mu+e_j} K(z)| (\|z\| \vee 1) < \infty \forall \mu \text{ with } |\mu| \leq \delta \forall j = 1, \dots, k,$$

where \vee denotes the maximum operator and $i = \sqrt{-1}$.

Below, for notational simplicity, we let C_j for some integer $j \geq 1$ 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

$$\mathcal{X}_T = \left\{ x \in \mathcal{X} : \inf_{x_0 \in \partial \mathcal{X}} |x - x_0| \geq d_T \right\}$$

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

LEMMA 1. *Suppose that assumptions A1, A2(a)-(c), A3(a), A4, and A5 hold. Then, we have:*
for $\mu \leq \omega$,

$$(a) \sup_{x \in \mathcal{X}_T} \left| D^\mu \widehat{f}(x) - D^\mu f(x) \right| = O_p(T^{-1/2} b_{1T}^{-(\mu+1)}) + O_p(b_{2T}^{\omega-\mu}).$$

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

PROOF OF LEMMA 1. The results of Lemma 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 $\widehat{\Omega}$ defined in (10) with Z_{t-1} replaced by X_{t-1} and $\Omega_T = \text{var}(X_{t-1})$ using A1 and a central limit theorem (CLT) of Herrndorf (1984, Corollary 1) applied to the strong mixing random variables $\{X_{t-1} : t \geq 1\}$. Finally NP5 is satisfied by A5. \square

(a) By adding and subtracting terms, we have

$$\begin{aligned}
 \text{(A.1)} \quad \sqrt{T}(\widehat{\lambda} - \lambda) &= \frac{1}{2\sqrt{T}} \sum_{t=1}^T [\ln(D\widehat{m}(X_{t-1}))^2 - \ln(Dm_0(X_{t-1}))^2] \mathbf{1}_t \\
 &\quad + \frac{1}{2\sqrt{T}} \sum_{t=1}^T [\ln(Dm_0(X_{t-1}))^2 - 2\lambda] \mathbf{1}_t \\
 &\quad - \frac{\lambda}{\sqrt{T}} \sum_{t=1}^T \{1 - \mathbf{1}_t\} \\
 \text{(A.2)} \quad &\equiv A_T + B_T - C_T, \text{ say.}
 \end{aligned}$$

Below we establish the following results:

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

$$\text{(A.4)} \quad C_T = o_p(1)$$

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

The 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 (A.3). A two-term Taylor expansion of A_T (defined in (A.1)) about $Dm_0(X_{t-1})$ gives:

$$A_T = \frac{1}{\sqrt{T}} \sum_{t=1}^T \frac{1}{Dm_0(X_{t-1})} [D\widehat{m}(X_{t-1}) - Dm_0(X_{t-1})] \mathbf{1}_t$$

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

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

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

where the second inequality follows by the uniform consistency results in Lemma 1(b) and the bandwidth conditions in A5 and the last convergence to zero holds because $(\min_{\{t: X_{t-1} \in \mathcal{X}_T\}} |Dm^*(X_{t-1})|)^{-2} = O_p(1)$ by A3(c) and Lemma 1(b). To see the latter, it suffice to show that $(\min_{\{t: X_{t-1} \in \mathcal{X}_T\}} |Dm^*(X_{t-1})|)^{-1} = O_p(1)$. For this purpose, choose first some $\varepsilon \in (0, 1)$. Note that there exists $0 < M_\varepsilon < \infty$ and $T_{1\varepsilon}$ such that

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

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

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

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

$$\begin{aligned}
& \Pr \left(\frac{1}{\min_{\{t: X_{t-1} \in \mathcal{X}_T\}} |Dm^*(X_{t-1})|} \leq M_\varepsilon \right) \\
\text{(A.10)} \geq & \Pr \left(\frac{1}{\min_{\{t: X_{t-1} \in \mathcal{X}_T\}} |Dm_0(X_{t-1})|} \leq \frac{M_\varepsilon}{2} \right) \\
& - \Pr \left(\frac{1}{\min_{\{t: X_{t-1} \in \mathcal{X}_T\}} |Dm_0(X_{t-1})|} \leq \frac{M_\varepsilon}{2} \text{ and } \frac{1}{\min_{\{t: X_{t-1} \in \mathcal{X}_T\}} |Dm^*(X_{t-1})|} \geq M_\varepsilon \right) \\
& \geq 1 - \frac{\varepsilon}{2} - \Pr \left(\min_{\{t: X_{t-1} \in \mathcal{X}_T\}} |Dm_0(X_{t-1})| \geq \frac{2}{M_\varepsilon} \text{ and } \min_{\{t: X_{t-1} \in \mathcal{X}_T\}} |Dm^*(X_{t-1})| \leq \frac{1}{M_\varepsilon} \right) \\
& \geq 1 - \frac{\varepsilon}{2} - \Pr \left(\max_{\{t: X_{t-1} \in \mathcal{X}_T\}} |Dm^*(X_{t-1}) - Dm_0(X_{t-1})| + \min_{\{t: X_{t-1} \in \mathcal{X}_T\}} |Dm^*(X_{t-1})| \geq \frac{2}{M_\varepsilon} \right. \\
& \quad \left. \text{and } \min_{\{t: X_{t-1} \in \mathcal{X}_T\}} |Dm^*(X_{t-1})| \leq \frac{1}{M_\varepsilon} \right)
\end{aligned}$$

$$\begin{aligned}
&\geq 1 - \frac{\varepsilon}{2} - \Pr\left(\max_{\{t: X_{t-1} \in \mathcal{X}_T\}} |D\hat{m}(X_{t-1}) - Dm_0(X_{t-1})| \geq \frac{1}{M_\varepsilon}\right) \\
&\geq 1 - \frac{\varepsilon}{2} - \frac{\varepsilon}{2} = 1 - \varepsilon.
\end{aligned}$$

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

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

$$(A.11) \quad A_{1T} = \sqrt{T} \int_{\mathcal{X}_T} \frac{1}{Dm_0(x)} [D\hat{m}(x) - Dm_0(x)] f(x) dx + o_p(1).$$

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

$$(A.12) \quad \begin{aligned} \nu_T(\tau) &= \frac{1}{\sqrt{T}} \sum_{t=1}^T [r(X_{t-1}, \tau) - Er(X_{t-1}, \tau)], \text{ where} \\ r(X_{t-1}, \tau) &= \frac{\mathbf{1}_t}{Dm_0(X_{t-1})} \tau(X_{t-1}), \end{aligned}$$

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

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

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

$$(A.14) \quad \nu_T(\hat{\tau}) - \nu_T(\tau_0) \xrightarrow{p} 0$$

if (i) $\Pr(\hat{\tau} \in \mathcal{T}) \rightarrow 1$, (ii) $\rho_{\mathcal{T}}(\hat{\tau}, \tau_0) \xrightarrow{p} 0$, and (iii) $\{\nu_T(\cdot) : T \geq 1\}$ is stochastically equicontinuous at τ_0 . Therefore, the result in (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 \mathcal{T} be a class of smooth functions; specifically, for large $C < \infty$, let

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

where $q \geq 1$ is a positive integer that appears in A2-A5. Note that $\tau_0(\cdot) = Dm_0(\cdot)$ lies in \mathcal{T} by A3(a). To show that (i) $\Pr(\hat{\tau} \in \mathcal{T}) \rightarrow 1$, it suffices to show that $\hat{\tau}(\cdot) = D\hat{m}(\cdot)$ has partial derivatives of order q on \mathcal{X}_T that are bounded uniformly over \mathcal{X}_T with probability tending to one. Note that the

latter holds by the uniform consistency result in Lemma 1(b) and A3(a). With the pseudo-metric defined in (A.13), the condition (ii) $\rho_T(\widehat{\tau}, \tau_0) \xrightarrow{p} 0$ holds since, for some positive constant $C_1 < \infty$,

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

where the second inequality holds by A2(c), A3(c) and Lemma 1(b). Now, the condition (iii) of $\{\nu_T(\cdot) : T \geq 1\}$ 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 A1(b) with \mathcal{W}_a^* given by \mathcal{X}_T . E(iii) follows by the definition of \mathcal{T} in (A.15). E(iv) is irrelevant to our case. E(v) holds since $\sup_{t \geq 1} E |\mathbf{1}_t/Dm_0(X_{t-1})|^r < \infty$ for $r > 2$ by A3(c). Finally, E(vi) holds by A1(a).

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

$$\begin{aligned}
&\sqrt{T} \int_{\mathcal{X}_T} \frac{1}{Dm_0(x)} [D\widehat{m}(x) - Dm_0(x)] f(x) dx \\
&= \sqrt{T} \int_{\mathcal{X}_T} \left[\frac{f(x)}{Dm_0(x)} \right] [D\widehat{m}(x) - Dm_0(x)] dx \\
\text{(A.17)} &= -\sqrt{T} \int_{\mathcal{X}_T} \left[D \left(\frac{f(x)}{Dm_0(x)} \right) \right] [\widehat{m}(x) - m_0(x)] dx + \sqrt{T} \sum_{x \in \partial \mathcal{X}_T} \left(\frac{f(x)}{Dm_0(x)} \right) [\widehat{m}(x) - m_0(x)]
\end{aligned}$$

$$\begin{aligned}
\text{(A.18)} &= -\sqrt{T} \int_{\mathcal{X}_T} \left[D \left(\frac{f(x)}{Dm_0(x)} \right) \frac{1}{f(x)} \right] [\{\widehat{m}(x) - m_0(x)\} \widehat{f}(x)] dx \\
&\quad + \sqrt{T} \int_{\mathcal{X}_T} \left[D \left(\frac{f(x)}{Dm_0(x)} \right) \right] \left[\frac{1}{\widehat{f}(x)} - \frac{1}{f(x)} \right] [\{\widehat{m}(x) - m_0(x)\} \widehat{f}(x)] dx \\
&\quad + \sqrt{T} \sum_{x \in \partial \mathcal{X}_T} \left(\frac{f(x)}{Dm_0(x)} \right) [\widehat{m}(x) - m_0(x)] \\
&= A_{1T}^* + A_{1T}^{**} + A_{1T}^{***}, \text{ say,}
\end{aligned}$$

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 $o_p(1)$. We have, for some positive constant $C_2, C_3 < \infty$

$$\begin{aligned}
|A_{1T}^{**}| &\leq T^{\frac{1}{2}} \left(\sup_{x \in \mathcal{X}_T} |\widehat{m}(x) - m_0(x)| \right) \left(\sup_{x \in \mathcal{X}_T} |\widehat{f}(x) - f(x)| \right) \\
&\quad \times \left(\inf_{x \in \mathcal{X}_T} f^{-2}(x) \right) \times C_1 (E |Dm_0(X_{t-1})|^{-2}) \\
\text{(A.19)} \quad &\leq T^{\frac{1}{2}} (O_p(T^{-1/2} b_{1T}^{-1} d_T^{-2\varrho}) + O_p(b_{2T}^\omega d_T^{-2\varrho})) (O_p(T^{-1/2} b_{1T}^{-1}) + O_p(b_{2T}^\omega)) \\
&\quad \times O(d_T^{-2\rho}) \\
&\quad \xrightarrow{p} 0,
\end{aligned}$$

where the second inequality holds by Lemma 1, A2, and A3 and the last convergence to zero holds using the bandwidth conditions in A5.

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

$$\begin{aligned}
\{\widehat{m}(x) - m_0(x)\} \widehat{f}(x) &= \frac{1}{T \widehat{b}_T} \sum_{t=1}^T \widehat{K} \left(\frac{x - X_{t-1}}{\widehat{b}_T} \right) \{X_t - m_0(X_{t-1})\} + \\
&\quad \frac{1}{T \widehat{b}_T} \sum_{t=1}^T \widehat{K} \left(\frac{x - X_{t-1}}{\widehat{b}_T} \right) \{m_0(X_{t-1}) - m_0(x)\},
\end{aligned}$$

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

$$\begin{aligned}
&\sqrt{T} \int_{\mathcal{X}_T} v(x) \left[\frac{1}{T \widehat{b}_T} \sum_{t=1}^T \widehat{K} \left(\frac{x - X_{t-1}}{\widehat{b}_T} \right) \{X_t - m_0(X_{t-1})\} \right] dx \\
&= \frac{1}{\sqrt{T}} \sum_{t=1}^T \{X_t - m_0(X_{t-1})\} \int_{\mathcal{X}_T} v(x) \frac{1}{\widehat{b}_T} \widehat{K} \left(\frac{x - X_{t-1}}{\widehat{b}_T} \right) dx \\
&= \frac{1}{\sqrt{T}} \sum_{t=1}^T \{X_t - m_0(X_{t-1})\} \left[v(X_{t-1}) + \int_{\mathcal{X}_T} \left\{ v(X_{t-1} + u \widehat{b}_T) - v(X_{t-1}) \right\} \widehat{K}(u) du \right] \\
\text{(A.20)} \quad &= \frac{1}{\sqrt{T}} \sum_{t=1}^T \{X_t - m_0(X_{t-1})\} v(X_{t-1}) \{1 + o_p(1)\},
\end{aligned}$$

where the second equality follows by a change of variables $x \rightarrow u = (x - X_{t-1})/\widehat{b}_T$, 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 $C_1 b_{1T} \leq b_T \leq C_2 b_{2T}$ and $|\Omega_T - \Omega| \leq c/\sqrt{T}$ for any $c < \infty$. Then, since $X_t - m_0(X_{t-1})$ is a martingale difference sequence,

$$\begin{aligned}
& E \left(\frac{1}{\sqrt{T}} \sum_{t=1}^T \{X_t - m_0(X_{t-1})\} \left[\int_{\mathcal{X}_T} \{v(X_{t-1} + ub_T) - v(X_{t-1})\} K_T(u) du \right] \right)^2 \\
&= E \left(\{X_t - m_0(X_{t-1})\}^2 \left[\int_{\mathcal{X}_T} \{v(X_{t-1} + ub_T) - v(X_{t-1})\} K_T(u) du \right]^2 \right) \\
&\leq \{E(|X_t - m_0(X_{t-1})|^{2r})\}^{1/r} \left\{ E \left(\left[\int_{\mathcal{X}_T} \{v(X_{t-1} + ub_T) - v(X_{t-1})\} K_T(u) du \right]^{2r/(r-1)} \right) \right\}^{(r-1)/r} \\
&\rightarrow 0,
\end{aligned}$$

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 \widehat{b}_T and $\widehat{\Omega}_T$, where $\widehat{\Omega}_T$ satisfies $|\widehat{\Omega}_T - \Omega| \leq c/\sqrt{T}$ for some constant c . The result follows because $\Pr \left[|\widehat{\Omega}_T - \Omega| \leq c/\sqrt{T} \right]$ can be made arbitrarily close to unity by taking c large.

Furthermore,

$$\begin{aligned}
& \left| \sqrt{T} \int_{\mathcal{X}_T} v(x) \frac{1}{T\widehat{b}_T} \sum_{t=1}^T \widehat{K} \left(\frac{x - X_{t-1}}{\widehat{b}_T} \right) \{m_0(X_{t-1}) - m_0(x)\} dx \right| \\
&= \left| \sqrt{T} \int_{\mathcal{X}_T} v(x) f^{1/2}(x) \frac{1}{T\widehat{b}_T} \sum_{t=1}^T \widehat{K} \left(\frac{x - X_{t-1}}{\widehat{b}_T} \right) \{m_0(X_{t-1}) - m_0(x)\} f^{-1/2}(x) dx \right| \\
&\leq \sqrt{T} \int_{\mathcal{X}_T} |v(x)|^2 f(x) dx \times \inf_{x \in \mathcal{X}_T} f^{-1}(x) \\
&\quad \times \sup_{x \in \mathcal{X}_T} \left| \frac{1}{T\widehat{b}_T} \sum_{t=1}^T \widehat{K} \left(\frac{x - X_{t-1}}{\widehat{b}_T} \right) \{m_0(X_{t-1}) - m_0(x)\} \right| \\
\text{(A.21)} \quad &\leq \sqrt{T} O(d_T^{-\varrho}) \times O_p(b_{2T}^\omega) \\
&\xrightarrow{p} 0,
\end{aligned}$$

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

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

$$\begin{aligned}
A_{1T}^{***} &= \sqrt{T} \sum_{x \in \partial \mathcal{X}_T} \left(\frac{f(x)}{Dm_0(x)} \right) [\hat{m}(x) - m_0(x)] \\
&= \sqrt{T} O_p(d_T^\varrho) \left\{ O_p(b_{2T}^\omega) + O_p(T^{-1/2} b_{1T}^{-1/2} d_T^{-\varrho/2}) \right\} \\
&\xrightarrow{p} 0,
\end{aligned}
\tag{A.22}$$

by A2 and A5.

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

$$\begin{aligned}
\Pr [C_T > \epsilon] &\leq \frac{E(C_T)}{\epsilon} \\
&= \frac{\sqrt{T}}{2\epsilon} \{1 - \Pr(X_{t-1} \in \mathcal{X}_T)\} \\
&\leq \frac{\sqrt{T}}{2\epsilon} \bar{A} d_T^{1+\varrho} \rightarrow 0,
\end{aligned}
\tag{A.23}$$

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

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

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

while

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

$$\leq \frac{1}{T} \sum_{t=1}^T \text{var}(\delta_t) + 16 \left\{ \sum_{s=1}^T \alpha^{1/a}(s) E^{2/b} [|\delta_t|^b] \right\}$$

$$(A.24) \quad \rightarrow 0.$$

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

$$\begin{aligned} E [|\delta_t|^b] &\leq E \left[|\ln(Dm_0(X_{t-1}))^2 - 2\lambda|^{b+\epsilon'} \right] E[(1 - \mathbf{1}_t)] \\ &= O(d_T^{1+\epsilon}), \end{aligned}$$

by the Hölder inequality and A3(b), while

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

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

(b) By rearranging terms,

$$\begin{aligned} \sqrt{n}(\hat{\lambda} - \lambda) &= \frac{1}{2\sqrt{n}} \sum_{t=1}^n [\ln(D\hat{m}(X_{t-1}))^2 - \ln(Dm_0(X_{t-1}))^2] \mathbf{1}_t \\ (A.25) \quad &+ \frac{1}{2\sqrt{n}} \sum_{t=1}^n [\ln(Dm_0(X_{t-1}))^2 - 2\lambda] \mathbf{1}_t \\ &- \frac{\lambda}{\sqrt{n}} \sum_{t=1}^n \{1 - \mathbf{1}_t\} \\ &\equiv A_n + B_n - C_n, \text{ say.} \end{aligned}$$

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

$$(A.26) \quad C_n = o_p(1)$$

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

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

$$B_n \Rightarrow N(0, \Phi_2)$$

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

$$(A.28) \quad \begin{aligned} |A_n| &= \left| \frac{1}{\sqrt{n}} \sum_{t=1}^n \frac{\mathbf{1}_t}{Dm^*(X_{t-1})} [D\widehat{m}(X_{t-1}) - Dm_0(X_{t-1})] \right| \\ &\leq T^{-\kappa} n^{\frac{1}{2} + \phi} \left[T^\kappa \sup_{x \in \mathcal{X}_T} |D\widehat{m}(x) - Dm_0(x)| \right] \\ &\quad \times \left(\frac{1}{n^\phi \min_{\{t: X_{t-1} \in \mathcal{X}_T\}} |Dm^*(X_{t-1})|} \right) \\ &\xrightarrow{p} 0, \end{aligned}$$

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 $D\widehat{m}(X_{t-1})$) and the inequality and the convergence to zero holds by Lemma 1(b) and A6* and A3(c)* because we then have $T^{-\kappa} n^{\frac{1}{2} + \phi} = O(1)$ and $(n^\phi \min_{1 \leq t \leq n} |Dm^*(X_{t-1})|)^{-1} = O_p(1)$ respectively (with the latter being verified using the similar arguments to those given in (A.8)-(A.10) above). \square

PROOF OF COROLLARY 1. (a) Define

$$(A.29) \quad \begin{aligned} \widetilde{\Phi} &= \sum_{j=-T+1}^{T-1} k \left(\frac{j}{S_T} \right) \widetilde{\gamma}(j) \quad ; \quad \widetilde{\gamma}(j) = \frac{1}{T} \sum_{t=|j|+1}^T \eta_t \eta_{t-|j|}, \\ \overline{\Phi} &= \sum_{j=-T+1}^{T-1} k \left(\frac{j}{S_T} \right) \overline{\gamma}(j) \quad ; \quad \overline{\gamma}(j) = \frac{1}{T} \sum_{t=|j|+1}^T \overline{\eta}_t \overline{\eta}_{t-|j|}, \end{aligned}$$

where $\overline{\eta}_t = \eta_t \mathbf{1}_t$ and η_t is defined in (15). Note that $\widetilde{\Phi} \xrightarrow{p} \Phi$ 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, $\widetilde{\Phi} - \overline{\Phi} \xrightarrow{p} 0$ by dominated convergence. By the triangular inequality, therefore, it suffices to show that $\widehat{\Phi} - \overline{\Phi} \xrightarrow{p} 0$.

Define the process

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

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

We now have

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

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

$$(A.32) \quad \sup_j |\hat{\gamma}(j) - \bar{\gamma}(j)| \\ = \sup_j \left| \frac{1}{T} \sum_{t=|j|+1}^T (\eta_t(\hat{\tau})\eta_{t-|j|}(\hat{\tau}) - \eta_t(\tau_0)\eta_{t-|j|}(\tau_0)) \right| \\ \leq 2 \left(\sup_{x \in \mathcal{X}_T} \|\hat{\tau}(x) - \tau_0(x)\| \right) \left[\frac{1}{T} \sum_{t=1}^T \sup_{\tau \in \mathcal{T}^*} \left\| \frac{\partial}{\partial \tau} \eta_t(\tau) \right\|^2 \right]^{1/2} \left[\frac{1}{T} \sum_{t=1}^T \sup_{\tau \in \mathcal{T}^*} \|\eta_t(\tau)\|^2 \right]^{1/2} \\ + \left(\sup_{x \in \mathcal{X}_T} \|\hat{\tau}(x) - \tau_0(x)\|^2 \right) \left[\frac{1}{T} \sum_{t=1}^T \sup_{\tau \in \mathcal{T}^*} \left\| \frac{\partial}{\partial \tau} \eta_t(\tau) \right\|^2 \right]$$

$$(A.33) \quad = O_p(T^{-1/2} b_{1T}^{-3} d_T^{-7\rho}) + O_p(b_{2T}^{\omega-2} d_T^{-7\rho})$$

$$(A.34) \quad \xrightarrow{p} 0,$$

as is required, where the first inequality (A.32) holds with probability that goes to one using the fact that $\Pr(\hat{\tau} \in \mathcal{T}^*) \rightarrow 1$ [see the arguments following equation (A.15)]. The second equality

(A.33), on the other hand, uses Lemma 1 and the results that $\frac{1}{T} \sum_{t=1}^T \sup_{\tau \in \mathcal{T}^*} \left\| \frac{\partial}{\partial \tau} \eta_t(\tau) \right\|^2 = O_p(d_T^{-4\varrho})$ and $\frac{1}{T} \sum_{t=1}^T \sup_{\tau \in \mathcal{T}^*} \|\eta_t(\tau)\|^2 = O_p(d_T^{-2\varrho})$ which hold by A3(c) and the definition of \mathcal{T}^* . Finally, the convergence to zero (A.34) holds by 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:

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

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

□

PROOF OF THEOREM 2. We first note that the results of Lemma 1 can be extended to cover the general case $k \geq 1$. Specifically, one can establish the results of Lemma 1 with x and \mathcal{X}_T replaced by z and $\mathcal{Z}_T = \mathcal{X}_T \times \cdots \times \mathcal{X}_T$ respectively and $\mu + 1$ that appears in the first terms of the right hand sides of Lemma 1(a) and (b) replaced by $\mu + k$. In the remaining proof, we use this modified version of Lemma 1.

(a) Define

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

By rearranging terms we have,

$$\begin{aligned}
\sqrt{T}(\widehat{\lambda} - \lambda) &= \sqrt{T}(\widehat{\lambda} - \lambda_T) + \sqrt{T}(\lambda_T - \lambda) \\
&= \frac{1}{\sqrt{T}} \left[\ln \left\| \prod_{t=1}^{T^*} \widehat{J}_{T-t} \right\| - \ln \left\| \prod_{t=1}^{T^*} J_{T-t} \right\| \right] \\
& \quad + \sqrt{T} \left[\frac{1}{T} \ln \left\| \prod_{t=1}^{T^*} J_{T-t} \right\| - \lambda \right] \\
\text{(B.2)} \quad &\equiv E_T + G_{T2}^*, \quad \text{say.}
\end{aligned}$$

Below we establish the following result:

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

where G_{T1} is as defined in (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 (A.24).

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

$$(B.4) \quad \begin{aligned} |E_T| &= \frac{1}{\sqrt{T}} \left[\ln \left\| \prod_{t=1}^{T^*} \widehat{J}_{T-t} \right\| - \ln \left\| \prod_{t=1}^{T^*} J_{T-t} \right\| \right] \\ &= \frac{1}{\sqrt{T}} \sum_{t=1}^{T^*} F_{t-1}(J_{T-1}, \dots, J_0) [\Delta \widehat{m}(Z_t) - \Delta m(Z_t)] \\ &\quad + \frac{1}{\sqrt{T}} \sum_{t=1}^{T^*} \sum_{s=1}^{T^*} [\Delta \widehat{m}(Z_t) - \Delta m(Z_t)]' Q_{ts}(J_{T-1}^*, \dots, J_0^*) [\Delta \widehat{m}(Z_s) - \Delta m(Z_s)] \\ &\equiv E_{1T} + E_{2T}, \quad \text{say,} \end{aligned}$$

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

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

$$(B.5) \quad \begin{aligned} |E_{2T}| &\leq T^{\frac{1}{2}} \left[\sup_{z \in \mathcal{Z}_T} \|\Delta \widehat{m}(z) - \Delta m(z)\| \right]^2 \frac{1}{T} \sum_{t=1}^{T^*} \sum_{s=1}^{T^*} \|Q_{ts}(J_{T-1}^*, \dots, J_0^*)\| \\ &\xrightarrow{p} 0, \end{aligned}$$

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

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

$$(B.6) \quad \begin{aligned} E_{1T} &= \frac{1}{\sqrt{T}} \sum_{t=1}^{T^*} F_{t-1}(J_{T-1}, \dots, J_0) [\Delta \widehat{m}(Z_t) - \Delta m(Z_t)] \\ &= \frac{1}{\sqrt{T}} \sum_{t=1}^{T^*} (F_{t-1}(J_{T-1}, \dots, J_0) - E[F_{t-1}(J_{T-1}, \dots, J_0) | \mathcal{G}_{t-l}^{t+m}]) \\ &\quad \times [\Delta \widehat{m}(Z_t) - \Delta m(Z_t)] \\ &\quad + \frac{1}{\sqrt{T}} \sum_{t=1}^{T^*} E[F_{t-1}(J_{T-1}, \dots, J_0) | \mathcal{G}_{t-l}^{t+m}] [\Delta \widehat{m}(Z_t) - \Delta m(Z_t)] \\ &= E_{1T}^* + E_{1T}^{**}, \quad \text{say.} \end{aligned}$$

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

$$(B.7) \quad |E_{1T}^*| \leq T^{\frac{1}{4}} \left[\sup_{z \in \mathcal{Z}_T} \|\Delta \widehat{m}(z) - \Delta m(z)\| \right] \\ \times T^{-\frac{1}{4}} \left[\frac{1}{T} \sum_{t=1}^{T^*} \|F_{t-1}(J_{T-1}, \dots, J_0) - E[F_{t-1}(J_{T-1}, \dots, J_0) | \mathcal{G}_{t-l-1}^{t+m}]\| \right]$$

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

$$(B.8) \quad E_{1T}^{**} = \frac{1}{\sqrt{T}} \sum_{t=1}^{T^*} E[F_{t-1}(J_{T-1}, \dots, J_0) | \mathcal{G}_{t-l-1}^{t+m}] [\Delta \widehat{m}(Z_t) - \Delta m(Z_t)] \\ = \nu_T(\widehat{\tau}) - \nu_T(\tau_0) + E_{1T}^{***}, \text{ where} \\ \nu_T(\tau) = \frac{1}{\sqrt{T}} \sum_{t=1}^T [r(W_t, \tau) - Er(W_t, \tau)], \\ r(W_t, \tau) = 1(z \in \mathcal{Z}_T) E[F_{t-1}(J_{T-1}, \dots, J_0) | \mathcal{G}_{t-l-1}^{t+m}] \tau(Z_t), \\ E_{1T}^{***} = \sqrt{T} \int_{\mathcal{Z}_T} F_T(z) [\Delta \widehat{m}(z) - \Delta m(z)] f(z) dz,$$

$W_t = (Z_{t-l}, \dots, Z_{t+m})$, $\widehat{\tau}(\cdot) = \Delta \widehat{m}(\cdot)$, and $\tau_0(\cdot) = \Delta m(\cdot)$. By similar arguments as those to verify the stochastic equicontinuity results in (A.12) - (A.16), we can verify $\nu_T(\widehat{\tau}) - \nu_T(\tau_0) = o_p(1)$.

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

$$(B.9) \quad E_{1T}^{***} = \sqrt{T} \int_{\mathcal{Z}_T} F_T(z) [\Delta \widehat{m}(z) - \Delta m(z)] f(z) dz \\ = -\sqrt{T} \int_{\mathcal{Z}_T} \left[\sum_{j=1}^k D^{e_j} \{F_{Tj}(z) f(z)\} \right] [\widehat{m}(z) - m(z)] dz \\ + \sqrt{T} \sum_{x \in \partial \mathcal{Z}_T} F_T(z) f(z) [\widehat{m}(x) - m_0(x)] \\ = -\sqrt{T} \int_{\mathcal{Z}_T} \left[\frac{1}{f(z)} \sum_{j=1}^k D^{e_j} \{F_{Tj}(z) f(z)\} \right] [\{\widehat{m}(z) - m_0(z)\} \widehat{f}(z)] dz \\ + \sqrt{T} \int_{\mathcal{Z}_T} \left[\sum_{j=1}^k D^{e_j} \{F_{Tj}(z) f(z)\} \right] \left[\frac{1}{\widehat{f}(z)} - \frac{1}{f(z)} \right] [\{\widehat{m}(z) - m_0(z)\} \widehat{f}(z)] dz \\ + \sqrt{T} \sum_{x \in \partial \mathcal{Z}_T} F_T(z) f(z) [\widehat{m}(x) - m_0(x)]$$

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

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 (A.19) - (A.22). \square

(b) Consider the 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 $o_p(1)$. A one-term Taylor expansion gives:

$$\begin{aligned} |E_T| &= \frac{1}{\sqrt{n}} \left| \ln \left\| \prod_{t=1}^n \widehat{J}_{n-t} \right\| - \ln \left\| \prod_{t=1}^n J_{n-t} \right\| \right| \\ (B.10) \quad &= \left| \frac{1}{\sqrt{n}} \sum_{t=1}^n F_{t-1}(J_{n-1}^*, \dots, J_0^*) [\Delta \widehat{m}(Z_t) - \Delta m(Z_t)] \right| \\ &\leq T^{-\kappa} n^{\frac{1}{2} + \phi} \left[T^\kappa \sup_{z \in \mathcal{Z}_T} \|\Delta \widehat{m}(z) - \Delta m(z)\| \right] \\ &\quad \times \left(n^{-\phi} \max_{1 \leq t \leq n} \|F_{t-1}(J_{n-1}^*, \dots, J_0^*)\| \right) \\ &\xrightarrow{p} 0, \end{aligned}$$

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

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.

REFERENCES

- [1] ANDREWS, D.W.K. (1989): "Stochastic Equicontinuity for Semiparametric Models." Unpublished manuscript, Cowles Foundation, Yale University.

- [2] ANDREWS, D.W.K. (1991): "Heteroskedasticity and autocorrelation consistent covariance matrix estimation." *Econometrica* **59**, 817-858.
- [3] ANDREWS, D.W.K. (1994): "Asymptotics for Semiparametric Econometric Models via Stochastic Equicontinuity." *Econometrica* **62**, 43-72.
- [4] ANDREWS, D.W.K. (1995): "Nonparametric Kernel Estimation for Semiparametric Models." *Econometric Theory* **11**, 560-596..
- [5] ANDREWS, D.W.K., AND J.C. MONAHAN (1992): "An improved heteroskedasticity and autocorrelation consistent covariance matrix estimator." *Econometrica* **60**, 953-966.
- [6] BIERENS, H. (1987): "Kernel estimators of regression functions," in *Advances in Econometrics: Fifth World Congress*, Vol 1, ed. by T.F. Bewley. Cambridge University Press.
- [7] BROCK, W.A. (1986): "Distinguishing random and deterministic systems," *Journal of Economic Theory*, **40**, 168-195.
- [8] BROCK, W.A., D. HSIEH, AND B. LEBARON (1991): *Nonlinear Dynamics, Chaos and Instability*, MIT Press, Cambridge, MA.
- [9] CHAN, K. S. AND H. TONG (1994): "A note on noisy chaos." *Journal of the Royal Statistical Society, Series B* **56**, 301-311.
- [10] CHENG, B. AND H. TONG (1992): "On consistent nonparametric order determination and chaos." *Journal of the Royal Statistical Society, Series B* **54**, 427-449.
- [11] CHERNICK, M.R. (1981): "A limit theorem for the maximum of autoregressive processes with uniform marginal distributions," *The Annals of Probability* **9**, 145-149.
- [12] DAVIES, R.A. (1979): "Maxima and minima of stationary processes," *The Annals of Probability* **7**, 453-460.
- [13] DECHERT, W.D. AND R. GENÇAY (1992): "Lyapunov Exponents as a Nonparametric Diagnostic for Stability Analysis," *Journal of Applied Econometrics* **7**, S41-S60.
- [14] DE JONG, R.M., AND J. DAVIDSON (1996): "Consistency of kernel estimators of heteroskedastic and autocorrelated covariance matrices," CentER Discussion paper no. 9652, Tilburg University.
- [15] DOUKHAN, P., (1994): *Mixing: Properties and Examples*. Springer-Verlag: Berlin.
- [16] ECKMANN, J.P., S.O. KAMPHORST, D. RUELLE, AND S. CILIBERTO (1986): "Lyapunov exponents from time series," *Physics Review A*, **34**, 4971-4979.
- [17] FOLLAND, G.B. (1984): *Real Analysis: Modern Techniques and their Applications*, John Wiley: New York.
- [18] FURSTENBERG, H., AND H. KESTEN (1960): "Products of random matrices," *The Annals of Mathematical Statistics*, **31**, 457-469.
- [19] GALLANT, A.R. (1987): *Nonlinear statistical models*. New York: Wiley.
- [20] GENÇAY, R., (1996): "A statistical framework for testing chaotic dynamics via Lyapunov exponents," *Physica D* **89**, 261-266.
- [21] GENÇAY, R., AND W.D. DECHERT (1992): "An algorithm for the n Lyapunov exponents of an n -dimensional unknown dynamical system," *Physica D* **59**, 142-157.

- [22] GNEDENKO, B.V. (1943): “Sur la distribution limite du term maximum d’une série aléatoire,” *Annals of Mathematical Statistics* **44**, 423-453.
- [23] HANSEN, B.E., (1992): “Consistent covariance matrix estimation for dependent heterogeneous processes,” *Econometrica* **60**, 967-972.
- [24] HÄRDLE, W., J. HART, J. S. MARRON, AND A. B. TSYBAKOV (1992): “Bandwidth Choice for Average Derivative Estimation,” *Journal of the American Statistical Association*, **87**, 218-226.
- [25] HÄRDLE, W., AND T. STOKER (1989): “Estimating smooth multiple regression by the method of average derivatives,” *Journal of the American Statistical Association* **84**, 986-995.
- [26] HENGARTNER, N.W. AND O.B. LINTON (1996): “Nonparametric regression estimation at design poles and zeros,” *Canadian Journal of Statistics* **24**, 583-591.
- [27] HERRNDORF, N. (1984): “An Invariance Principle for Weakly Dependent Sequences of Random Variables.” *Annals of Probability* **12**, 141-153.
- [28] LEADBETTER, M.R., AND H. ROOTZÉN (1988): “Extremal theory for stochastic processes,” *Annals of Probability* **16**, 431-478.
- [29] LEBARON, B. (1997): “Has chaos theory found any useful application in the social sciences,” Manuscript, University of Wisconsin, <http://www.ssc.wisc.edu/~blebaron>.
- [30] LOYNES, R.M. (1965): “Extreme values in uniformly mixing stationary stochastic processes,” *Annals of Mathematical Statistics* **66**, 993-999.
- [31] MCCAFFREY, D., S. ELLNER, A.R. GALLANT AND D. NYCHKA, (1992): “Estimating the Lyapunov exponent of a chaotic system with nonparametric regression.” *Journal of the American Statistical Association* **87**, 682-695.
- [32] MASRY, E. (1996). “Multivariate local polynomial regression for time series: Uniform strong consistency and rates,” *Journal of Time Series Analysis* **17**, 571-599.
- [33] NEWAY, W.K., (1994): “The asymptotic variance of semiparametric estimators,” *Econometrica* **62**, 1349-1382.
- [34] NEWAY, W.K., (1995): “Convergence rates for series estimators,” in *Advances in Econometrics and Quantitative Econometrics*, eds. by G.S. Maddala, P.C.B. Phillips, and T.N. Srinivasan. Basil Blackwell.
- [35] NEWAY, W.K., AND K.D. WEST (1987): “A simple positive-definite, heteroskedastic and autocorrelation consistent covariance matrix,” *Econometrica* **55**, 703-708.
- [36] NYCHKA, D., S. ELLNER, A.R. GALLANT AND D. MCCAFFREY (1992): “Finding Chaos in Noisy Systems.” *Journal of the Royal Statistical Society, Series B* **54**, 399-426.
- [37] OULIARIS, S., AND P.C.B. PHILLIPS (1994): “COINT 2.0: Gauss procedures for cointegrated regressions,” *Predicta Software Inc, 133 Concord Drive, Madison CT 06443*.
- [38] PESARAN, M. H., AND S.M. POTTER (1992): “Nonlinear Dynamics and Econometrics: An Introduction,” *Journal of Applied Econometrics* **7**, S1-S7.
- [39] ROBINSON, P.M. (1983): “Nonparametric estimators for time series” *Journal of Time Series Analysis*, **4**, 185-208.

- [40] TONG, H (1990): *Non-linear Time Series: A Dynamical Systems Approach*. Clarendon Press, Oxford.
- [41] WATSON, G.S. (1954): "Extreme values in samples from m-dependent stationary stochastic processes," *The Annals of Mathematical Statistics* **25**, 798-800.
- [42] WHITE, H., (1984): *Asymptotic theory for econometricians*. New York: Academic Press.

TABLE INFORMATION

Tables 1-3 reports the results for the Gaussian autoregression for the full sample estimator, i.e., $n = T$ throughout. We give the mean, median, and standard deviation (sd) of our estimate of λ along with the asymptotic standard deviation (asd), i.e., that value predicted by the theory, for reference. Table 1 gives the case $\gamma = 0.2$, Table 2 gives the case $\gamma = 0.3$, and Table 3 gives the case $\gamma = 0.4$.

Table 4 reports the median and interquartile range (iqr) of our estimates of $\Phi^{1/2}$ in the case $n = T = 500$. The corresponding true values of $\Phi^{1/2}$ for $\rho = 0.5, 0.7, 0.9$, and 0.95 are: 0.077, 0.046, 0.022, and 0.015.

Table 5 reports the 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 $cT^{1/6}$. We give the mean, median, and standard deviation (sd) of our estimate of λ along with the asymptotic standard deviation (asd), i.e., that value predicted by the theory, for reference. We also report the median and interquartile range (iqr) of the standard error estimates.

TABLE 1

		n = 100	n = 200	n = 300	n = 500
$\rho = 0.5$ ($\lambda = -0.693, \Phi = 3.0$)	mean	-0.937	-0.844	-0.793	-0.748
	median	-0.887	-0.801	-0.769	-0.739
	sd	0.292	0.207	0.146	0.094
	asd	0.173	0.122	0.100	0.078
$\rho = 0.7$ ($\lambda = -0.357, \Phi = 1.04$)	mean	-0.552	-0.447	-0.414	-0.388
	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
$\rho = 0.9$ ($\lambda = -0.105, \Phi = 0.24$)	mean	-0.269	-0.179	-0.152	-0.132
	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
$\rho = 0.95$ ($\lambda = -0.051, \Phi = 0.11$)	mean	-0.215	-0.127	-0.099	-0.079
	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
$\rho = 1.0$ ($\lambda = 0, \Phi = 0$)	mean	-0.167	-0.080	-0.053	-0.032
	median	-0.151	-0.073	-0.048	-0.030
	sd	0.091	0.037	0.026	0.015

TABLE 2

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

TABLE 3

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

TABLE 4

		$\rho =$	0.5	0.7	0.9	0.95
$\widehat{\Phi}^{1/2}$	$\gamma = 0.2$	median	0.431	0.079	0.022	0.014
		iqr	4.213	0.328	0.012	0.005
	$\gamma = 0.3$	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
$\widehat{\Phi}_1^{1/2}$	$\gamma = 0.2$	median	0.092	0.044	0.020	0.014
		iqr	0.104	0.018	0.004	0.003
	$\gamma = 0.4$	median	0.057	0.032	0.015	0.011
		iqr	0.027	0.006	0.003	0.002
	$\gamma = 0.6$	median	0.042	0.024	0.012	0.009
		iqr	0.010	0.004	0.002	0.002

TABLE 5

		T = 100	T = 200	T = 300	T = 500
$\widehat{\lambda} \quad (n = T)$	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
$se(\widehat{\lambda})$	median	0.049	0.036	0.030	0.023
	iqr	0.019	0.011	0.011	0.007
$\widehat{\lambda} \quad (n = cT^{1/6})$	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
$se(\widehat{\lambda})$	median	0.205	0.213	0.210	0.191
	iqr	0.141	0.142	0.129	0.124
$\widehat{\lambda} \quad (n = cT^{1/3})$	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
$se(\widehat{\lambda})$	median	0.180	0.158	0.153	0.140
	iqr	0.085	0.070	0.062	0.053
$\widehat{\lambda} \quad (n = cT^{1/2})$	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
$se(\widehat{\lambda})$	median	0.120	0.109	0.100	0.088
	iqr	0.032	0.024	0.022	0.020