Statistical Estimation of Local Lyapunov Exponents: Toward Characterizing Predictability in Nonlinear Systems

Zhan-Qian Lu

Abstract

While Lyapunov exponents are among the global dynamical invariants studied for detecting chaos and nonlinear structure in time series analysis, *local* Lyapunov exponents, defined as the local divergence within a finite-time horizon, are a more useful measure of predictability of nonlinear systems and a more powerful tool for testing nonlinearity in time series.

This paper is concerned with estimating local Lyapunov exponents in time series using the kernel regression method of local polynomial fitting. The estimators of local Lyapunov exponents are shown to approach joint asymptotic normality. Approximate formulae for computing the asymptotic bias and the asymptotic variance are derived. A method for constructing confidence intervals which adaptively correct for bias in the estimators is prescribed. This methodology is illustrated with applications to the Nicholson blowfly population data and the daily maximum temperature series at Charleston, SC.

KEY WORDS: Lyapunov exponents; Chaos; Nonlinear dynamics; Testing nonlinearity; Multi-step prediction errors; Charleston temperature series.

1 Introduction

Lyapunov exponents and fractal dimension are among the global dynamical invariants studied for detecting chaos and nonlinear structure in nonlinear systems. However, practical concerns such as prediction require quantifying the finite-time behavior of a nonlinear system. The *local* Lyapunov exponents, defined as the divergence rates of two nearby trajectories within a finite-time horizon, is one way of characterizing predictability in a nonlinear system, and is a powerful tool for quantifying and testing nonlinearity in time series.

Introductions to the increasely popular area of chaos and nonlinear time series from the statisticians' viewpoint include Tong (1990), Tong and Smith (1992), which contains several papers cited here, Berliner (1992), and Tong (1995). The diversity of applications of chaos theory to substantial fields can be seen in the collections of papers edited by Berry, Percival and Weiss (1987), Drazin and King (1992), and Grenfell, May, and Tong (1994).

While much progress has been made on the statistical properties of dimension estimation methods in recent years, see for example Cutler (1993) and Smith (1992), fundamental statistical issues remain about estimating Lyapunov exponents, as shown in the discussions by McCaffrey, Nychka, Ellner, and Gallant (1992), and Nychka, Ellner, McCaffrey, and Gallant (1992). While these global dynamical methods have been successfully applied to many noise-free and high quality data, there are both conceptual and practical difficulties in applications to time series data which are subject to significant noises and often they fail to find nonlinearity (cf. Example 3 of Section 3). In this paper, we develop a finite-time and local state space approach for studying nonlinear time series via the local Lyapunov exponents. This approach appears to be able to detect nonlinearity in time series when global nonlinear dynamical techniques fail.

In deterministic chaotic dynamical systems, small errors in initial conditions can amplify rapidly in the future and errors in the initial conditions are the main source of forecast errors. For example, error in determination of present weather state is among the main sources of errors in current numerical weather prediction. Local Lyapunov exponents provide a way of quantifying the influence of small error in the specification of initial condition and thus of predictability of a nonlinear system. The concept of local Lyapunov exponents has a long history in meteorology. Right after his discovery of chaos in a physical context, Lorenz (1965) developed a similar concept in the context of weather forecasting. This pioneering work has played a key role in many meteorological predictability studies and has been a key method for the recent ensemble weather forecasting. The local Lyapunov exponents defined in this paper are a direct generalization of this classical local Lyapunov exponents for a deterministic system to a noisy system.

In a scalar time series model, the local Lyapunov exponents (or LLEs) are defined through the state space representation and the Jacobian matrix of the state space dynamics. One way to define nonlinearity of a time series model is through the property that first-order partial derivatives are *state dependent*. For example Priestley (1988, Ch. 5) developed a model identification procedure based on this property. LLEs for a scalar time series are defined from the Jacobian matrix of the first-order partial derivatives in a state space representation. Thus, variability of LLEs over states is a test of nonlinearity. Further, LLEs characterize sensitivity to initial conditions in the state space representation. Local Lyapunov exponents for one-dimensional systems, in somewhat different forms from our approach, have also been proposed and studied by Wolff (1992) and Yao and Tong (1994). The statistical theory for estimating local Lyapunov exponents from time series is developed in this paper. This appears to be the first systematic study of LLEs for multidimensional systems. This study is related to estimating global Lyapunov exponents in the sense that the estimators for global Lyapunov exponents are given by those of local Lyapunov exponents for very large time horizon. This methodology is illustrated with applications to the blowfly population data due to A. J. Nicholson and the daily maximum temperature data at Charleston, SC.

The main statistical issue in this paper is estimation of partial derivatives of an autoregression function, for which nonparametric regression method of local polynomial fitting is employed. An advantage of local polynomial derivative estimators is that they are easy to analyze, and in particular the asymptotic bias and asymptotic variance can be characterized. This paper develops a method for transforming the results on derivative estimators to those for the LLE estimators using the theory of eigenvalues from a random matrix. An alternative approach, based on neural net and spline approaches to nonlinear regression, has been given by Bailey, Ellner and Nychka (1997). However, they do not obtain explicit expressions for the asymptotic bias and variance.

The rest of this paper is organized in the following manner. A precise definition of LLEs is given in Subsection 1.1. Issues of defining LLEs for time series with unknown state space are briefly discussed in Subsection 1.2. In Section 2 the statistical theory of estimating local Lyapunov exponents is developed. The theory of local polynomial fitting for derivative estimation is reviewed in Subsection 2.1. The general theory for singular values of a random matrix is reviewed in Subsection 2.3. In Subsection 2.4, joint asymptotic normality of the LLE estimators is established, and approximate formulae for computing the asymptotic bias and asymptotic variance are derived. In Subsection 2.5, this theory is applied to develop a method for constructing pointwise confidence intervals which adaptively correct for bias in the estimators. Applications are given in Section 3. Concluding remarks are given in Section 4.

1.1 Local Lyapunov exponents

Observations of time series are subject to various errors. We consider the following model for noisy time series

$$x_{i+1} = m(x_i, \ldots, x_{i-p+1}) + \nu^{1/2}(x_i, \ldots, x_{i-p+1})\varepsilon_{i+1}, \tag{1.1}$$

where $m : \Re^p \to \Re$ is some nonlinear function, $\nu : \Re^p \to \Re$ is a nonnegative function, and $\{\varepsilon_i, i \geq 1\}$ is a sequence of i.i.d. random variables with mean 0 and variance 1; moreover, ε_{i+1} is assumed to be independent of $\{x_j, j \leq i\}$. While an estimation theory will be discussed for the general form (1.1), the definition of LLEs for this setup is complicated. To simplify our discussion, we will assume that ν is a constant function and $\nu = \sigma^2$.

Model (1.1) can be written in *state space form*: define X_i to be the *embedded state* vector $(x_i, \ldots, x_{i-p+1})^T$ and $e_i = (\sigma\varepsilon_i, 0, \ldots, 0)$. Define

$$M(X_i) = (m(x_i, \ldots, x_{i-p+1}), x_i, \ldots, x_{i-p+2})^T, \tag{1.2}$$

Then

$$X_{i+1} = M(X_i) + e_{i+1}. \tag{1.3}$$

Our purpose is to quantify and to derive a diagnostic test for nonlinearity of model (1.3). This is equivalent to characterizing nonlinearity of M or the state dependency of the Jacobian matrix of M. For multi-step prediction, both uncertainty in the present state and dynamical noise contribute to the forecast uncertainty. The forecast uncertainty is controlled by the sensitivity on initial values of the system, which is, for small errors, quantified by the local Lyapunov exponents.

Now we discuss the motivations that lead to a definition of local Lyapunov exponents for the state space model (1.3). The specific form (1.2) of M is not important here, though this type of state space representation does have an effect on defining LLEs for a scalar time series when p is unknown (cf. Sec. 1.2). To derive the definition, we consider the problem of *Monte Carlo prediction.* Suppose that at a given time instant i, there exists a reconstructed state vector $\tilde{X}_i$ which is close to X_i. Assume further that the noise shocks $e_{i+1}, \ldots, e_{i+L}$ are *known*, then one can consider the multi-step prediction $\tilde{X}_{i+1}, \ldots, \tilde{X}_{i+L}$ defined by

$$\tilde{X}_{i+j} = M(\tilde{X}_{i+j-1}) + e_{i+j}, 1 \leq j \leq L. \tag{1.4}$$

That is, $\tilde{X}_{i+1}, \ldots, \tilde{X}_{i+L}$ differ from $X_{i+1}, \ldots, X_{i+L}$ only in the specification of the initial condition at time i.

The one-step prediction error is approximated by

$$\tilde{X}_{i+1} - X_{i+1} = M(\tilde{X}_i) - M(X_i) \approx D_M(X_i)(\tilde{X}_i - X_i), \tag{1.5}$$

where D_M denotes the Jacobian matrix of M. We will write $J(\mathbf{x}) = D_M(\mathbf{x})$ as a shorthand notation. The *initial error* $\delta X_i = \tilde{X}_i - X_i$ is assumed to be *small* enough in (1.5).

Similarly, the L-step prediction error is given by

$$\begin{aligned} \tilde{X}_{i+L} - X_{i+L} &\approx D_M(X_{i+L-1})(\tilde{X}_{i+L-1} - X_{i+L-1}) \\ &\approx \ldots \\ &\approx D_M(X_{i+L-1}) \cdots D_M(X_i)(\tilde{X}_i - X_i). \end{aligned} \tag{1.6}$$

provided that all intermediate errors $\{\tilde{X}_{i+\ell} - X_{i+\ell}, (0 \leq \ell \leq L-1)\}$ are *small* enough in (1.6).

The covariance of prediction error has the form

$$(J^L)^T \Omega J^L, \tag{1.7}$$

where $J^L = J(X_{i+L-1}) \cdots J(X_i)$ and $\mathrm{Cov}(\tilde{X} - X) = \Omega$, and in particular this is of the form $\rho^2 (J^L)^T J^L$ if $\Omega = \rho^2 I$.

In addition, provided that the noise shocks $e_{i+1}, \ldots, e_{i+L-1}$ are small enough, we have from (1.3),

$$\begin{aligned} X_{i+2} &= M(M(X_i) + e_{i+1}) + e_{i+2} \\ &\approx M(M(X_i)) + J(M(X_i))e_{i+1} + e_{i+2} \end{aligned}$$

and in general

$$\begin{aligned} X_{i+L} \approx\ & M^L(X_i) + J(M^{L-1}(X_i)) \ldots J(M(X_i))e_{i+1} \\ & + J(M^{L-1}(X_i)) \ldots J(M^2(X_i))e_{i+2} \\ & + \cdots + J(M^{L-1}(X_i))e_{i+L-1} + e_{i+L} \end{aligned}$$

where M^ℓ is the $\ell-$fold composition of M with itself. If X_i is completely *known*, an $L-$step predictor at time i is $M^L(X_i)$ and the predictive covariance is of the form

$$\tilde{J}_2^L \Sigma (\tilde{J}_2^L)^T + \tilde{J}_3^L \Sigma (\tilde{J}_3^L)^T + \cdots + \tilde{J}_L^L \Sigma (\tilde{J}_L^L)^T + \Sigma \tag{1.8}$$

where $\mathrm{Cov}(e_1) = \Sigma$ and $\tilde{J}_j^L = J(M^{L-1}(X_i)) \ldots J(M^{j-1}(X_i))$ for $j \leq L$.

In general, let $\{\mathbf{x}_0, \mathbf{x}_1, \ldots, \mathbf{x}_{L-1}\}$ denote generically any consecutive observed states or their approximations. From now on we will fix the notation

$$J^L = J(\mathbf{x}_{L-1}) \ldots J(\mathbf{x}_0).$$

From previous discussions it is seen that the size of J^L controls the prediction error in both deterministic and probabilistic senses. Define $\delta_1^2(L) \geq \cdots \geq$

$\delta_p^2(L)$ to be the ordered the eigenvalues of of $(J^L)^T J^L$. Equivalently, $\delta_1(L) \geq \cdots \geq \delta_p(L)$ are the singular values of J^L. Then, the normalized growth rates

$$\lambda_i(L) = \frac{1}{L} \log \delta_i(L), \ i = 1, \ldots, p, \tag{1.9}$$

are called the $L-$step *local Lyapunov exponents* (LLEs) (wrt $\mathbf{x}_0, \mathbf{x}_1, \ldots, \mathbf{x}_{L-1}$). The corresponding eigenvectors $V_1, \ldots, V_p(L)$ are called the $L-$step *local Lyapunov vectors.*

The relation to prediction error is given by, for a given initial error $a \in \Re^p$,

$$\|J^L a\|^2 =< a, V_1(L) >^2 \exp(2L\lambda_1(L)) + \ldots + < a, V_p(L) >^2 \exp(2L\lambda_p(L)), \tag{1.10}$$

where $< \cdot, \cdot >$ denotes the inner product. Thus, if $\lambda_i(L) > 0$, an initial error which has significant projection onto the subspace spanned by $V_1(L), \ldots, V_i(L)$ is amplified at the $L-$th step, while if $\lambda_i(L) < 0$, an initial error lying in the subspace spanned by $V_i(L), \ldots, V_p(L)$ is reduced at the $L-$th step and thus has less influence on prediction. If the process $\{X_i\}$ is ergodic, a general theorem states that $\lambda_i = \lim_{L \to \infty} \lambda_i(L)$ exist almost surely. The quantities $\lambda_1, \ldots, \lambda_p$ characterize the property of the whole process $\{X_i\}$ and are called the (global) Lyapunov exponents (cf. Kifer, 1986; McCaffrey et al, 1992). A bounded process with $\lambda_1 > 0$ is called chaotic (Nychka et al 1992, McCaffrey et al 1992). This appears to be a viable and simple way of defining noisy chaos, though there are possible alternative approaches, see Tong (1995).

For the one-dimensional case, e.g. $p = 1$ in (1.1), the only $L-$step local Lyapunov exponent is given by

$$\lambda(L) = \frac{1}{L} \left\{ \log |m'(x_{L-1})| + \ldots \log |m'(x_0)| \right\}. \tag{1.11}$$

This can be compared with those of Wolff (1992) and Yao and Tong (1994) where alternative definitions of local Lyapunov type quantities are given.

Nevertheless, these alternative definitions can be treated as in some sense "averaged" versions of LLE (cf. Discussions to Tong 1995).

The LLEs so defined controls the multi-step prediction error of a nonlinear system in various ways. For example, the magnitude of $L-$step prediction variance in (1.8) is controlled by the LLEs at time steps $L-1, L-2, \ldots, 1$. An additional interpretation may be given as follows. Consider the one-dimensional system ($p = 1$) for simplicity. Let $\tilde{x}_i$ be an estimate of true state x_i, and let $m_L = H(x_i, \varepsilon_{i+1}, \ldots, \varepsilon_{i+L})$ denote the $L-$step evolution of (1.3). Assume that the distribution of ε_i is *known*, the $L-$step predictive distribution given by the random variable $\tilde{m}_L = H(\tilde{x}_i, \varepsilon_{i+1}, \ldots, \varepsilon_{i+L})$ is available. One can evaluate the mean squared prediction error given by

$$\begin{aligned} \mathrm{E}(\tilde{m}_L - m_L)^2 &= \mathrm{E}\{\mathrm{E}[(\tilde{m}_L - m_L)^2 | \varepsilon_{i+1}, \ldots, \varepsilon_{i+L}]\} \\ &= \mathrm{E}_{\varepsilon_{i+1},\ldots,\varepsilon_{i+L}} \mathrm{E}_{\tilde{x}_i} (\tilde{m}_L - m_L)^2 \\ &\approx \mathrm{E}_{\varepsilon_{i+1},\ldots,\varepsilon_{i+L-1}} [m'(x_{t+L-1}) \cdots m'(x_i)]^2 \rho^2 \\ &= \rho^2 \mathrm{E}_{\varepsilon_{i+1},\ldots,\varepsilon_{i+L-1}} \exp\{2L\lambda(L)\}, \end{aligned} \tag{1.12}$$

where $x_{i+\ell+1} = m(x_{i+\ell}) + \varepsilon_{i+\ell}$ for $0 \le \ell \le L-2$. The 2nd line assumes the independence of $\tilde{x}_i$ and $\varepsilon_{i+1}, \ldots, \varepsilon_{i+L}$ and $\mathrm{E}(\tilde{x}_i - x_i)^2 = \rho^2$, while the 3rd line uses the local linear approximation. This shows that the prediction error is controlled by the averaged growth rate given by the local Lyapunov exponent.

1.2 Effect of embedding

Given a univariate time series $\{x_i, i = 1, 2, \ldots, N\}$, if its phase space is unknown, one can consider the reconstructed state vectors

$$X_i = (x_i, x_{i-\tau}, \ldots, x_{i-(p-1)\tau})^T, i = (p-1)\tau + 1, \ldots, N,$$

for some selection of the embedding dimension p and time delay τ. $\tau > 1$ is useful if the time series is finely sampled. If we only consider LLEs for integer time steps of τ, the definition of LLEs is defined through $X_{i+\tau} = M_\tau(X_i) + e_{i+\tau}$ where M_τ is the $\tau-$step evolution map and is similar to the case of $\tau = 1$. For simplicity this is the only situation considered in this paper. For theoretical discussions we can assume $\tau = 1$ without loss of generality. The selection of p has been considered by a number of authors (e.g. Cheng and Tong 1992, Tjøstheim and Auestad 1994) and in this paper we do not consider this.

For estimating global Lyapunov exponents or fractal dimension, if p is large enough, one hopes that estimates of these global quantities stabilize and converge to the true ones. However, it is not known how the embedding dimension will affect the definition of local Lyapunov exponents. As far as we are aware of, we have not seen any statement to this effect. Thus, in this subsection, we will derive some results on this.

Suppose the time series comes from an autoregressive model (1.1) of order p_0. Obviously if $p < p_0$, the LLEs defined based on embedding p will be biased. Thus, we will just consider the case $p > p_0$.

Note that, for the reconstructed state space map (1.2), its Jacobian matrix at a fixed point $\mathbf{u} = (u_p, \cdots, u_2, u_1)^T$ has the simple form

$$D_M(\mathbf{u}) = \begin{pmatrix} \multicolumn{2}{c}{D_m^T(\mathbf{u})} \\ I_{p-1} & 0_{p-1,1} \end{pmatrix}, \tag{1.13}$$

where $D_m^T(\mathbf{u}) = (\partial m(\mathbf{u})/\partial u_p, \cdots, \partial m(\mathbf{u})/\partial u_1)$, and we use I_k to denote the identity matrix of dimension k and $0_{k,q}$ as the $k \times q$ zero matrix.

Let $\tilde{\lambda}_1(L) \geq \ldots \geq \tilde{\lambda}_p(L)$ denote the LLEs corresponding to embedding dimension p. Let $J^i = J(\mathbf{x}_{i-1}) \cdots J(\mathbf{x}_0)$ where $J(\mathbf{x}) = D_M(\mathbf{x})$ denotes the Jacobian matrix corresponding to the true dimension p_0. Denote the ℓ-th row of a matrix A by $A[\ell,]$. Let $E(i,j)$ denotes the $p_0 \times p_0$ matrix which consists of one at the (i,j)th position and of zeros elsewhere. Given a positive definite matrix A which has the spectral decomposition $A = U\Lambda U^T$ where U is orthogonal and $\Lambda = \mathrm{diag}\{d_1, \ldots, d_p\}$, we define $\log A = U(\log \Lambda)U^T$ where $\log \Lambda = \mathrm{diag}\{\log d_1, \ldots, \log d_p\}$.

The following theorem is obtained. The proof, which involves algebraic calculations and exploiting the special structure of (1.13), is fairly straightforward and is omitted.

Theorem 1 *For $L \le p - p_0$, $\tilde{\lambda}_1(L), \tilde{\lambda}_2(L), \ldots, \tilde{\lambda}_{p_0}(L)$ are the eigenvalues of*

$$\frac{1}{2L} \log(\{J^L\}^T J^L + \sum_{j=1}^{L-1} \{J^{L-j}[p_0,]\}^T J^{L-j}[p_0,] + E(p_0, p_0)),$$

and $\tilde{\lambda}_{p_0+1}(L) = \cdots = \tilde{\lambda}_{p-L}(L) = 0, \tilde{\lambda}_{p-L+1}(L) = \cdots = \tilde{\lambda}_p(L) = -\infty$.

For $L > p - p_0$, $\tilde{\lambda}_1(L), \tilde{\lambda}_2(L), \ldots, \tilde{\lambda}_{p_0}(L)$ are the eigenvalues of

$$\frac{1}{2L} \log(\{J^L\}^T J^L + \sum_{j=1}^{p-p_0} \{J^{L-j}[p_0,]\}^T J^{L-j}[p_0,]),$$

and $\tilde{\lambda}_{p_0+1}(L) = \cdots = \tilde{\lambda}_p(L) = -\infty$.

Theorem 1 implies that the time series LLEs are positively biased due to the effect of embedding. However, if p is large enough, the biases die out quickly as L is increased. This also implies that the definition of global Lyapunov exponents is consistent, as stated in the next corollary.

Corollary 1 *If the embedding dimension $p \ge p_0$, where p_0 is the true dimension, the first p_0 Lyapunov exponents are preserved, i.e.*

$$\tilde{\lambda}_i = \lambda_i, \ 1 \le i \le p_0; \tilde{\lambda}_i = -\infty, \ i > p_0.$$

While a positive global Lyapunov exponent λ_1 implies sensitive dependence on initial conditions or chaos (e.g. McCaffrey et al 1992), one should be cautious about interpreting the sign of local Lyapunov exponents due to the embedding effect. The following example further illustrates this point.

Example 1. Consider the NAR(1) model

$$x_{i+1} = m(x_i) + \sigma\varepsilon_{i+1}.$$

Consider the two-dimensional embedded space or $p = 2$. The two-step Jacobian product has the form

$$J^L = \begin{pmatrix} a & 0 \\ b & 0 \end{pmatrix}, \text{ where } b = \textstyle\prod_{j=i}^{i+L-2} m'(x_j), a = m'(x_{i+L-1})b,$$

for $L \geq 2$. The two LLE are given by

$$\begin{aligned} \lambda_1(2) &= \frac{1}{2L}[\log b^2 + \log(1 + \{m'(x_{i+L-1})\}^2)] \\ &= \frac{L-1}{L} \cdot \frac{1}{L-1} \sum_{j=i}^{i+L-2} \log|m'(x_j)| + \frac{1}{2L}\log(1 + \{m'(x_{i+L-1})\}^2), \\ \lambda_2(2) &= -\infty. \end{aligned}$$

It is seen that $\lambda_1 = \lim_{L\to\infty} \lambda_1(L) = \mathrm{E}\log|m'(x)|$, the global Lyapunov exponent. As an example, consider the AR(1) model $m(x) = \rho x$ embedded in the two-dimensional state space, the largest 2-step LLE is given by $\lambda_1(2) = 0.25\log\rho^2(1+\rho^2)$ which is positive if and only if $\rho^2(1+\rho^2) > 1$, i.e. $\rho^2 > (\sqrt{(5)} - 1)/2$.

2 Estimation of local Lyapunov exponents

Given an embedding dimension p, estimators of LLEs are given by plugging in the estimates of partial derivatives. Various nonlinear regression methods have been proposed for estimating Lyapunov exponents, including the neural net and spline regression methods (McCaffrey et al, 1992; Nychka et al, 1992). In this paper we will focus on the local polynomial regression method for which explicit statistical results can be derived. The next subsection reviews recent results for derivative estimation using the local polynomial fitting.

2.1 Derivative estimation

Multivariate local polynomial fitting is used for estimation of m and its partial derivative vector

$$D_m(\mathbf{x}) = (\partial m(\mathbf{x})/\partial x_1, \cdots, \partial m(\mathbf{x})/\partial x_p)^T$$

in model (1.1). The main statistical theory for partial derivative estimation in the time series context will be reviewed in this subsection. The details are given in Lu (1996a, b). First, we embed the time series model (1.1) in the regression framework

$$Y_i = m(X_i) + \nu(X_i)\varepsilon_i \tag{2.1}$$

by setting

$$\{X_i = (x_{i-1}, \ldots, x_{i-p}), Y_i = x_i\}. \tag{2.2}$$

Given time series data $x_1, x_2, \ldots, x_N$, there are $n = N - p$ observed vectors of (2.2) corresponding to $i = p+1, p+1, \ldots, N$. If $\tau > 1$, the embedded data are of the form

$$\{X_i = (x_{i-\tau}, \ldots, x_{i-p\tau}), Y_i = x_i; i = p\tau + 1, N\}$$

and $n = N - p\tau$.

In this paper we consider the local quadratic partial derivative estimator. That is, at any given point $\mathbf{x} = (x_1, \ldots, x_p)^T$, the estimator is derived from minimizing the weighted sum

$$\sum_{i=p+1}^{N} \{Y_i - a - b^T(X_i - \mathbf{x}) - (X_i - \mathbf{x})^T L(X_i - \mathbf{x})\}^2 \frac{1}{h^p} K(\frac{X_i - \mathbf{x}}{h}), \quad (2.3)$$

where a is a real number, b is a $p-$dimensional vector, and L is a $p \times p$ matrix which is restricted to be a lower triangular matrix for identifiability. The solution corresponding to minimizing (2.3) consists of $\widehat{a} = \widehat{m}(\mathbf{x})$, of $\hat{b} = \hat{D}_m(\mathbf{x})$ which corresponds to an estimate of $D_m(\mathbf{x})$, and of $\hat{L}$ which corresponds to estimates of elements in the Hessian matrix $H_m(\mathbf{x}) = (\partial^2 m(\mathbf{x})/\partial x_i \partial x_j)$ at $\mathbf{x}$. That is, $L(x) = (l_{ij})$ satisfies $l_{ij} = h_{ij}$ if $i > j$ and $= h_{ii}/2$ if $i = j$, where $H_m(\mathbf{x}) = (h_{ij})$ is the Hessian. Let $\widehat{\beta} = (\hat{a}, \hat{b}^T, \mathrm{vech}^T\{\hat{L}\})^T$ and we have

$$\hat{\beta} = (\mathbf{X}^T W \mathbf{X})^{-1} \mathbf{X}^T W Y, \quad (2.4)$$

where $Y = (Y_1, \cdots, Y_n)^T, W = \mathrm{diag}\{K(\frac{X_1 - \mathbf{x}}{h}), \cdots, K(\frac{X_n - \mathbf{x}}{h})\}$ and

$$\mathbf{X} = \begin{pmatrix} 1 & (X_{p+1} - \mathbf{x})^T & \mathrm{vech}^T\{(X_1 - \mathbf{x})(X_{p+1} - \mathbf{x})^T\} \\ \vdots & \vdots & \vdots \\ 1 & (X_N - \mathbf{x})^T & \mathrm{vech}^T\{(X_n - \mathbf{x})(X_N - \mathbf{x})^T\} \end{pmatrix}. \quad (2.5)$$

Here vech^T denotes the row vector consisting of the columns on and below the diagonal of a symmetric matrix. Other local polynomial estimators can be similarly considered (cf. Lu 1996a, b), though the local quadratic estimator appears to be simplest and is thus mostly used in practice.

The following assumptions are made for the regression model (2.1):

(A) The sequence $\{\varepsilon_i, \mathcal{F}_i\}$ is a martingale difference with $\mathrm{E}\{\varepsilon_i | \mathcal{F}_{i-1}\} = 0, \mathrm{E}\{\varepsilon_i^2 | \mathcal{F}_{i-1}\} = 1$ and $X_i \in \mathcal{F}_{i-1}$ for all $i \geq 1$.

(B) The noise sequence further satisfies $\sup_{i \geq 1} \mathrm{E}\{|\varepsilon_i|^{2+\delta} | \mathcal{F}_{i-1}\} < \infty$ for some $\delta > 0$.

(C) The vector sequence $\{X_i\}$ is strictly stationary and satisfies the short-range dependence condition: let $f_j(\cdot,\cdot)$ denote the joint density of X_1, X_{j+1} and $f(\cdot)$ denote the marginal density, then

$$\sup_{\mathbf{u},\mathbf{v}\in\Re^p} \sum_{j=1}^{\infty} |f_j(\mathbf{u},\mathbf{v}) - f(\mathbf{u})f(\mathbf{v})| < \infty. \tag{2.6}$$

(D) K is assumed to be spherically symmetric and to satisfy

$$\int u_1^{12} K(u_1, \cdots, u_p) du_1 \cdots du_p < \infty.$$

Let U denote an open neighborhood of $\mathbf{x} = (x_1, \cdots, x_p)^T$ in $\Re^p$, and let $C^d(U)$ be the class of functions which have up to order d continuous partial derivatives in U. The following theorem is from Lu (1996b, Theorem2).

Theorem 2 *Under model (2.1) and Assumptions (A)-(D), for l distinct points $\mathbf{x}_1, \ldots, \mathbf{x}_\ell$ such that $f(\mathbf{x}_j) > 0, \nu(\mathbf{x}_j) > 0$ for all j, if there exist open neighborhoods U_i of $\mathbf{x}_i$ such that $m \in C^4(U_j), f \in C^0(U_j), \nu \in C^0(U), j = 1, 2, \ldots, \ell$, then for $h \to 0, nh^p \to \infty$ as $n \to \infty$, the local quadratic estimators $\widehat{\beta}(\mathbf{x}_1), \ldots, \widehat{\beta}(\mathbf{x}_\ell)$ are asymptotically independent and jointly normal. In particular, at each point $\mathbf{x} = (x_1, \cdots, x_p)^T$, we have that*

$$(nh^{p+2})^{1/2}\{\hat{D}_m(\mathbf{x}) - D_m(\mathbf{x}) - \beta(\mathbf{x}, h)\} \xrightarrow{d} N(0, \frac{\gamma_2 \nu(\mathbf{x})}{\mu_2^2 f(\mathbf{x})} I_p). \tag{2.7}$$

Here $\beta(\mathbf{x}, h) = (h^2/3!\mu_2)b(\mathbf{x})$,

$$b(\mathbf{x}) = \begin{pmatrix} \mu_4 \frac{\partial^3 m(\mathbf{x})}{\partial x_1^3} + 3\mu_2^2 \sum_{i=2}^p \frac{\partial^3 m(\mathbf{x})}{\partial x_i^2 \partial x_1} \\ \mu_4 \frac{\partial^3 m(\mathbf{x})}{\partial x_2^3} + 3\mu_2^2 \sum_{i\neq 2} \frac{\partial^3 m(\mathbf{x})}{\partial x_i^2 \partial x_2} \\ \vdots \\ \mu_4 \frac{\partial^3 m(\mathbf{x})}{\partial x_p^3} + 3\mu_2^2 \sum_{i=1}^{p-1} \frac{\partial^3 m(\mathbf{x})}{\partial x_i^2 \partial x_p} \end{pmatrix}, \tag{2.8}$$

and $\mu_\ell = \int u_1^\ell K(\mathbf{u}) d\mathbf{u}, \gamma_\ell = \int u_1^\ell K^2(\mathbf{u}) d\mathbf{u}$ *for any nonnegative integers* ℓ.

2.2 LLE estimators

Consider in the embedded state space any L distinct fixed points $\{\mathbf{x}_i = (x_{ip}, \ldots, x_{i2}, x_{i1})^T\}, 0 \le i \le L-1$. Let $J(\mathbf{x}_i)$ denote the Jacobian matrix of M at $\mathbf{x}_i$ and $J^L = J(\mathbf{x}_{L-1}) \ldots J(\mathbf{x}_0)$ denote the $L-$step Jacobian product. An estimator of J^L is given by $\hat{J}^L = \hat{J}(\mathbf{x}_{L-1}) \cdots \hat{J}(\mathbf{x}_0)$ by plugging in the respective partial derivative estimators. We denote the corresponding singular values of $\hat{J}^L$ by $\hat{\delta}_i(L), 1 \le i \le p$ and the $L-$step LLE estimators are given by

$$\hat{\lambda}_i(L) = \frac{1}{L} \log\{\hat{\delta}_i(L)\}, \text{ for } 1 \le i \le p.$$

Define a sequence of matrices by

$$B(\mathbf{x}, h) = \begin{pmatrix} h^2/(3!\mu_2)\{b^T(\mathbf{x}_i) + o(1)\} \\ 0_{p-1,p} \end{pmatrix}, \tag{2.9}$$

where $b(\mathbf{x}_i) = (b_{ip}, \ldots, b_{i2}, b_{i1})^T$ is as given in (2.8). A sequence of random matrices is defined by

$$W_i = \frac{\sigma\sqrt{\gamma_2}}{\mu_2\sqrt{f(\mathbf{x}_i)}} \begin{pmatrix} Z_i^T \\ 0_{p-1,p} \end{pmatrix}, \; i = 0, 1, \ldots, L-1 \tag{2.10}$$

where $Z_0, \ldots, Z_{L-1}$ are iid $N(0, I_p)$.

We introduce notations for the intermediate Jacobian products:

$$J_k^j = J(\mathbf{x}_{j-1}) \cdots J(\mathbf{x}_{k-1}), \text{ for } 0 \le k \le j \le L, J^k = J_1^k, \text{ for } k \ge 0, \tag{2.11}$$

and $J^0 = I_p$. The following corollary follows from Theorem 2.

Corollary 2 *Assume that model (1.1) satisfies the conditions of Theorem 2, we have as $h \to 0, nh^p \to \infty$,*

$$(nh^{p+2})^{1/2}\{\hat{J}^L - J^L - \sum_{k=0}^{L-1} J_{k+2}^L B(\mathbf{x}_k, h) J^k - O(h^4)\}$$

$$\xrightarrow{d} \tilde{W} = \sum_{k=0}^{L-1} J_{k+2}^{L} W_k J^k, \tag{2.12}$$

with the convention $J^0 = J_{L+1}^L = I_p$.

To derive asymptotic results for the LLE estimators based on Corollary 2, we need the theory of singular values from a random matrix, which is considered in the next subsection.

2.3 Singular values from a random matrix

Consider the following setup: assume that an asymmetric matrix T_n satisfies

$$c_n^{1/2}(T_n - A_n) \xrightarrow{d} W, \tag{2.13}$$

where $c_n \to \infty$, and A_n is a sequence of matrices tending to a nonrandom matrix A, i.e.

$$A_n = A + B_n, B_n \xrightarrow{p} 0, \tag{2.14}$$

and W is a random matrix. We will denote the vector of singular values in decreasing order of a matrix G by $\delta(G) = (\delta_1(G), \ldots, \delta_p(G))^T$. Our purpose is to study the asymptotic behavior of

$$Z_n = \delta(T_n) - \delta(A). \tag{2.15}$$

By the singular value decomposition,

$$A = U\Delta V^T, \text{ where } \Delta = \mathrm{diag}\{\delta_1(A), \ldots, \delta_r(A), 0, \ldots, 0\},$$

where $\delta_r > 0, r$ is the rank of A, and $U = (U_1, \cdots, U_p), V = (V_1, \cdots, V_p)$ are orthogonal matrices. Let $\mathrm{dgv}(G)$ denote the column vector consisting of the diagonal elements of a square matrix G, and so for any G,

$$\mathrm{dgv}(U^T G V) = (U_1^T G V_1, \cdots, U_p^T G V_p)^T. \tag{2.16}$$

We have the following theorem.

Theorem 3 *Under (2.13) and (2.14), if A is full rank and all the singular values of A have multiplicity one (otherwise consider the subvector consisting of those simple positive singular values), we have*

$$c_n^{1/2}\left\{\delta(T_n)-\delta(A)-\mathrm{dgv}(U^TB_nV)-o(B_n)\right\}\xrightarrow{d}\mathrm{dgv}(U^TWV). \tag{2.17}$$

If A is degenerate and $\delta_p=0$ has multiplicity one and $B_n=o_p(c_n^{-1/2})$, the following result holds

$$c_n^{1/2}|\delta_p(T_n)|\xrightarrow{d}\|WV_p\|.$$

PROOF. The first part of the theorem follows from the fact that a singular value from a matrix is differentiable if of multiplicity one. The second part of the theorem concerning the zero singular value follows directly from Theorem 4.2 of Eaton and Tyler (1994). □

The special case $B_n=0$ is also given in Eaton and Tyler (1994). One consequence of Theorem 3 is that, if W is jointly normal, the limiting distribution of the singular values which have positive and simple limits is jointly normal.

2.4 Asymptotic theory for estimating local Lyapunov exponents

We denote the singular value decomposition of T^L by

$$J^L=U(L)\mathrm{diag}\{\delta_1(L),\cdots,\delta_p(L)\}V^T(L),$$

where $U(L),V(L)$ are orthogonal, and whose columns are denoted by

$$U(L)=(U_1(L),\cdots,U_p(L)),V(L)=(V_1(L),\cdots,V_p(L)).$$

Recall that $J_{k+2}^L=J(\mathbf{x}_{L-1})\ldots J(\mathbf{x}_{k+1})$ for $k=0,\ldots,L-1$. We denote its columns by

$$J_{k+2}^L=(J_{k+2}^L(1),\cdots,J_{k+2}^L(p)). \tag{2.18}$$

The following lemma is obtained.

Lemma 1 *Assuming that all the singular values of J^L have multiplicity one and positive, under conditions of Corollary 2, we have*

$$(nh^{p+2})^{1/2}\left\{\begin{pmatrix}\hat{\delta}_1(L)\\ \vdots\\ \hat{\delta}_p(L)\end{pmatrix}-\begin{pmatrix}\delta_1(L)\\ \vdots\\ \delta_p(L)\end{pmatrix}\right.$$

$$\left.-\frac{h^2}{6\mu_2}\sum_{k=0}^{L-1}\begin{pmatrix}\{U_1^T(L)J_{k+2}^L(1)\}\{b^T(\mathbf{x}_k)J^kV_1(L)\}\\ \vdots\\ \{U_p^T(L)J_{k+2}^L(1)\}\{b^T(\mathbf{x}_k)J^kV_p(L)\}\end{pmatrix}-o(h^2)\right\}\xrightarrow{d}N(0,\Sigma^s),$$

where $\Sigma^s=(\sigma_{ij}^s)$,

$$\begin{aligned}\sigma_{ii}^s &= \frac{\sigma^2\gamma_2}{\mu_2^2}\sum_{k=0}^{L-1}\frac{1}{f(\mathbf{x}_k)}\{U_i^T(L)J_{k+2}^L(1)\}^2\{V_i^T(L)(J^k)^TJ^kV_i(L)\},\\ \sigma_{ij}^s &= \frac{\sigma^2\gamma_2}{\mu_2^2}\sum_{k=0}^{L-1}\frac{1}{f(\mathbf{x}_k)}\{U_i^T(L)J_{k+2}^L(1)\}\{U_j^T(L)J_{k+2}^L(1)\}\\ &\quad\{V_i^T(L)(J^k)^TJ^kV_j(L)\},\end{aligned}$$

for $1\le i\le p, 1\le j\le p$.

PROOF: From Corollary 2 and Theorem 3, we have

$$(nh^{p+2})^{1/2}\left\{\begin{pmatrix}\hat{\delta}_1(L)\\ \vdots\\ \hat{\delta}_p(L)\end{pmatrix}-\begin{pmatrix}\delta_1(L)\\ \vdots\\ \delta_p(L)\end{pmatrix}\right.$$

$$\left.-\begin{pmatrix}\sum_{k=0}^{L-1}U_1^T(L)J_{k+2}^LB(\mathbf{x}_k,h)J^kV_1(L)\\ \vdots\\ \sum_{k=0}^{L-1}U_p^T(L)T_{k+2}^LB(\mathbf{x}_k,h)J^kV_p(L)\end{pmatrix}-o(h^2)\right\}$$

$$\xrightarrow{d}\begin{pmatrix}\sum_{k=0}^{L-1}U_1^T(L)J_{k+2}^LW_kJ^kV_1(L)\\ \vdots\\ \sum_{k=0}^{L-1}U_p^T(L)J_{k+2}^LW_kJ^kV_p(L)\end{pmatrix}.$$

From the definition of $B(\mathbf{x}, h)$, we have that

$$J_{k+2}^L B(\mathbf{x}_k, h) = \frac{h^2}{3!\mu_2} J_{k+2}^L(1) b(\mathbf{x}_k) + o(h^2),$$

and

$$\begin{aligned} &U_k^T(L) J_{k+2}^L B(\mathbf{x}_k, h) J^k V_i(L) = \\ &\quad \frac{h^2}{3!\mu_2} \{U_i^T(L) J_{k+2}^L(1)\}\{b^T(\mathbf{x}_k) J^k V_i(L)\}, \text{ for } 1 \le i \le p. \end{aligned}$$

Similarly, from the definition of W_k in (2.10), we have

$$J_{k+2}^L W_k = \frac{\sigma\sqrt{\gamma_2}}{\mu_2\sqrt{f(\mathbf{x}_k)}} J_{k+2}^L(1) Z_j^T,$$

and

$$U_i^T(L) J_{k+2}^L W_k J^k V_i(L) = \frac{\sigma\sqrt{\gamma_2}}{\mu_2\sqrt{f(\mathbf{x}_k)}} \{U_i^T(L) J_{k+2}^L(1)\}\{Z_j^T J^k V_i(L)\},$$

for $1 \le i \le p, 1 \le j \le L$.

So we have proved that

$$\begin{aligned} &(nh^{p+2})^{1/2} \left\{ \begin{pmatrix} \hat{\delta}_1(L) \\ \vdots \\ \hat{\delta}_p(L) \end{pmatrix} - \begin{pmatrix} \delta_1(L) \\ \vdots \\ \delta_p(L) \end{pmatrix} \right. \\ &\quad \left. - \frac{h^2}{6\mu_2} \sum_{k=0}^{L-1} \begin{pmatrix} \{U_1^T(L) J_{k+2}^L(1)\}\{b^T(\mathbf{x}_k) J^k V_1(L)\} \\ \vdots \\ \{U_p^T(L) J_{k+2}^L(1)\}\{b^T(\mathbf{x}_k) J^k V_p(L)\} \end{pmatrix} - o(h^2) \right\} \\ &\quad \xrightarrow{d} \frac{\sigma\sqrt{\gamma_2}}{\mu_2} \sum_{k=0}^{L-1} \frac{1}{\sqrt{f(\mathbf{x}_k)}} \begin{pmatrix} \{U_1^T(L) J_{k+2}^L(1)\}\{Z_j^T J^k V_1(L)\} \\ \vdots \\ \{U_p^T(L) J_{k+2}^L(1)\}\{Z_j^T J^k V_p(L)\} \end{pmatrix}, \quad (2.19) \end{aligned}$$

from which the theorem follows easily. $\square$

Applying the "delta method" to $\log x$ and Lemma 1, we obtain our main theorem.

Theorem 4 *Under conditions of Lemma 1, we have*

$$
\begin{aligned}
&(nh^{p+2})^{1/2}\left\{\begin{pmatrix}\hat\lambda_1(L)\\ \vdots \\ \hat\lambda_p(L)\end{pmatrix}-\begin{pmatrix}\lambda_1(L)\\ \vdots \\ \lambda_p(L)\end{pmatrix}\right.\\
&\qquad\left.-\frac{h^2}{6\mu_2 L}\begin{pmatrix}\delta_1^{-1}(L)\sum_{k=0}^{L-1}\{U_1^T(L)J_{k+2}^L(1)\}\{b^T(\mathbf{x}_k)J^kV_1(L)\}\\ \vdots \\ \delta_p^{-1}(L)\sum_{k=0}^{L-1}\{U_p^T(L)J_{k+2}^L(1)\}\{b^T(\mathbf{x}_k)J^kV_p(L)\}\end{pmatrix}-o(h^2)\right\}\\
&\qquad\stackrel{d}{\to} N(0,\Sigma^L),
\end{aligned}
$$

where $\Sigma^L=(\sigma_{ij}^L)$,

$$
\begin{aligned}
\sigma_{ii}^L &= \frac{\sigma^2\gamma_2}{L^2\mu_2^2\delta_i^2(L)}\sum_{k=0}^{L-1}\frac{1}{f(\mathbf{x}_k)}\{U_i^T(L)J_{k+2}^L(1)\}^2\|J^kV_i(L)\|^2,\\
\sigma_{ij}^L &= \frac{\sigma^2\gamma_2}{L^2\mu_2^2\delta_i(L)\delta_j(L)}\\
&\quad \sum_{k=0}^{L-1}\frac{1}{f(\mathbf{x}_k)}\{U_i^T(L)J_{k+2}^L(1)\}\{U_j^T(L)J_{k+2}^L(1)\}\{V_i^T(L)(J^k)^TJ^kV_j(L)\},
\end{aligned}
$$

for $1\le i\le p, 1\le j\le p$.

2.5 Confidence intervals

Theorem 4 serves as a basis for constructing confidence intervals for local Lyapunov exponents. One immediate difficulty is to deal with bias in the estimators. Theoretically, bias may be avoided by using a smaller bandwidth, though in practice it is hard to decide when a bandwidth is small enough. Furthermore, for the asymptotically optimal bandwidth, the bias term is not negligible. To deal with this issue, the following *bias-correction* method is proposed. The following corollary follows directly from Theorem 4.

Corollary 3 *Under conditions of Theorem 4, and given consistent estimators $\hat{\beta}_i$ and $\hat{\sigma}_{ii}^L$, we have as $h \to 0, nh^p \to \infty$,*

$$(nh^{p+2})^{1/2}(\hat{\sigma}_{ii}^L)^{-1}(\hat{\lambda}_i - \lambda_i - \frac{h^2}{6\mu_2 L}\hat{\beta}_i - o(h^2)) \xrightarrow{d} N(0,1).$$

Based on Corollary 3, the pointwise "plug-in" confidence intervals (CIs) can be constructed for $\lambda_i(L)$. Let

$$\begin{aligned} I_n &= (\hat{\lambda}_i - (nh^{p+2})^{-1/2}\hat{\sigma}_{ii}^L Z_{\alpha/2} + \frac{h^2}{6\mu_2 L}\hat{\beta}_i, \\ &\quad \hat{\lambda}_i + (nh^{p+2})^{-1/2}\hat{\sigma}_{ii}^L Z_{\alpha/2} + \frac{h^2}{6\mu_2 L}\hat{\beta}_i), \end{aligned} \tag{2.20}$$

where $Z_{\alpha/2}$ is the $(1-\alpha/2)$th percentile of the standard normal. If the remainder term of order $o(h^2)$ in the asymptotic bias is negligible, in the sense that $(nh^{p+2})^{1/2}h^2$ is bounded, i.e. $h \le c_1 n^{-1/(p+6)}$ for some constant c_1, Corollary 3 implies that

$$P(I_n) \to 1-\alpha, \text{as } h \to 0, nh^p \to \infty, nh^{p+6} = O(1). \tag{2.21}$$

Simultaneous confidence intervals for several $\lambda_i(L)$'s can also be constructed using e.g. the Bonferroni method.

In the examples in Section 3, the local cubic fit with bandwidth h_4 is used for estimating the bias term $b(\mathbf{x}_k), k = 1, 2, \ldots, L$. Consistent estimators of $U(L)$ and $V(L)$ are given by the orthogonal matrices $\hat{U}(L)$ and $\hat{V}(L)$ derived from the singular value decomposition of $\hat{J}^L$. Thus, a consistent estimator for the asymptotic bias of $\hat{\lambda}_i$ is given by $\frac{h^2}{6\mu_2 L}\hat{\beta}_i$, where

$$\hat{\beta}_i \triangleq \hat{\delta}_i^{-1}(L)\sum_{k=0}^{L-1}\{\hat{U}_i^T(L)\hat{J}_{k+2}^L(1)\}\{\hat{b}^T(\mathbf{x}_k)\hat{J}^k\hat{V}_i(L)\}, \tag{2.22}$$

for any $1 \le i \le p$.

The kernel density estimator is used, i.e. at a given point $\mathbf{x}$,

$$\hat{f}(\mathbf{x}) = \frac{1}{nh_2^p}\sum_{i=1}^{n} K(\frac{X_i - \mathbf{x}}{h_2}).$$

For variance estimation, we use the local variance estimator

$$\hat{\nu}(\mathbf{x}) = \sum_{i=1}^{n} Y_i^2 K(\frac{X_i - \mathbf{x}}{h_3}) / \sum_{i=1}^{n} K(\frac{X_i - \mathbf{x}}{h_3}) - \hat{m}_{NW}(\mathbf{x})^2,$$

where $\hat{m}_{NW}$ is the Nadaraya-Watson estimator given by

$$\hat{m}_{NW}(\mathbf{x}) = \sum_{i=1}^{n} Y_i K(\frac{X_i - \mathbf{x}}{h_3}) / \sum_{i=1}^{n} K(\frac{X_i - \mathbf{x}}{h_3}).$$

It is easy to see that the variance estimator is always nonnegative and is consistent under general conditions. Thus, a consistent estimator for the asymptotic variance for $\hat{\lambda}_i$ is given by

$$\hat{\sigma}_{ii}^{L} = \frac{\gamma_2}{L^2 \mu_2^2 \hat{\delta}_i^2(L)} \sum_{k=0}^{L-1} \frac{\hat{\nu}(\mathbf{x}_k)}{\hat{f}(\mathbf{x}_k)} \{\hat{U}_i^T(L) \hat{J}_{k+2}^{L}(1)\}^2 \|\hat{J}^k \hat{V}_i(L)\|^2. \tag{2.23}$$

Though several bandwidths need to be specified in this bias-adjusting procedure, our limited experience suggests that as a rule of thumb, one may choose $h_2 = h_3 = h$, and $h_4 = 1.5h$. Typically a bandwidth can be chosen by the trial and error method. For a given choice of h the asymptotic bias and variance are computed to characterize the estimation error and also to fine tune for a better bandwidth choice.

3 Applications

In this section, we consider applications using the LLE methodology developed in this paper. A large class of multivariate kernels is given by $K_{\alpha\beta}(\mathbf{x}) = c^{-1}(1 - \|\mathbf{x}\|^{\alpha})^{\beta} 1_{\{\|\mathbf{x}\| \le 1\}}$ for different choices of parameters α and β where c is the normalizing constant. In these analyses, the tricube kernel K_{33} is used for local quadratic fitting and biweight kernel K_{23} is used for density and variance estimation. To explore the dynamics change in phase space for a given time series data, we will evaluate $\lambda_1(L)$ along some sequence of embedded state vectors, that is, let $\mathbf{x}_0, \ldots, \mathbf{x}_{L-1}$ taking on $X_i, \ldots, X_{i+L-1}$ where $X_j = (x_{j-(p-1)}, \ldots, x_j)^T$ as i goes through selected time points.

Example 2. The data considered are the bidaily population sizes of blowflies obtained by entomologist A. J. Nicholson in 1957 in his study of population dynamics. This data set has been analyzed by a number of authors using different nonlinear time series models and nonlinearity testing procedures. Tsay (1988) gave an excellent review on analyses of this data set and related references. Due to the concern for stationarity, only the first 206 observations are used in this study. We choose $p = 2$ since nonlinearity in most fitted models involves only the first two lags, though there could be potential bias if additional lags prove to be significant. We also choose $h = 4500, h_2 = h_3 = h$ and $h_4 = 6000$. Two-step LLEs are computed at all possible points in the reconstructed state space. Figure 1a shows the time series of the original data and first LLEs. Figure 1b shows the computed confidence intervals for first LLEs, which are constructed using the method discussed in Section 2.5. Figure 1c shows the distribution of first LLEs in the phase space. Most of the first LLE estimates are positive, indicating that there is sensitive dependence on initial values in most part of the series and the system is likely nonlinear. Furthermore, strongest unstabilities occur at the onsets of these limit cycles when there are big jumps in blowfly population caused by the unusual amount of emergence. This is consistent with the finding of Tsay (1986) that large residuals persistently appear in these time points.

To assess the overall behavior of the limit cycles we have also computed 8-step LLEs based on the same parameters. This result is shown in Figure 2 in similar format as in Figure 1. The implications of Figure 2 are similar to those of Figure 1 in that the rising part of the growth cycle has most unpredictability while the negative first LLEs indicate that there are predictability in other parts of the series during this longer time horizon, consistent with the limit cycle nature of this data set.

This example shows that the LLEs are rather effective in detecting the asymmetry or nonlinearity of limit cycles. Furthermore, it may indicate which part of a time series is predictable and which part is unpredictable.

Example 3. In this example, we apply the methodology to the daily maximum temperature series from 1928 to 1987 at Charleston, SC. To remove seasonal climatology, we consider modelling the temperature series ξ_t by

$$\begin{aligned}
\xi_t &= \alpha(t) + \nu^{1/2}(t)x_t \qquad (3.1)\\
\alpha(t) &= a_0 + \sum_{i=1}^{p_m}(a_{1i}\cos(\frac{2\pi it}{365}) + a_{2i}\sin(\frac{2\pi it}{365})),\\
\nu(t) &= \sum_{j=1}^{p_v}(c_0 + c_{1j}\cos(\frac{2\pi jt}{365}) + c_{2j}\sin(\frac{2\pi jt}{365})),
\end{aligned}$$

where p_m, p_v are the order of harmonics fitted. For this data set, we choose $p_m = 7$ and fit $\alpha(t)$ separately to the two 30-year periods 1928-1957 and 1958-1987. $\nu(t)$ with $p_v = 3$ is fitted to the empirical annual mean of $(\xi_t - \hat{\alpha}(t))^2$. The standardized time series $\{x_t = (\xi_t - \hat{\alpha}(t))/\hat{\nu}^{1/2}(t)\}$ appears to be stationary. The time series plot and autocorrelation plots for both x_t and x_t^2 are shown in Figure 3.

Two-day LLEs are computed based on $\{x_t\}$ for year 1987. We choose $p = 2$ which appears to be adequate and use $h_2 = h_3 = h = 2.5$ and $h_4 = 3$. Figure 4a shows the time series plot of estimates of $\lambda_1(2)$ on different days along with x_t for year 1987. Figure 4b shows the 95% pointwise confidence intervals along with estimated values of the first LLEs. All LLE estimates are negative, indicating that the system is stable and errors in initial value have little effect on the multi-period predictions. On the other hand, the LLEs are able to distinguish some nonlinear structural change in the dynamics. For example, based on Figure 4a and particularly the phase plot in Figure 4c, we define three regions in the phase space

$$\begin{aligned} R^{+} &= \{x_{t-1} \leq 0, x_t \leq -1\}; \\ R^{0} &= \{x_{t-1} \leq 0, x_t > 0\}; \\ R^{\cdot} &= \{x_{t-1} > 0, -\infty < x_t < \infty\}. \end{aligned}$$

It appears that first LLEs are largest on R^+ while being smallest on R^0.

Since the LLEs are negative, larger LLE on R^+ implies more predictability since the local autoregression coefficients are further away from zero while smaller LLE on R^0 implies less predictability since the local AR coefficients are closer to zero. Indeed, this is confirmed by fitting separate autoregressive models for the three regions $R^+, R^0, R^{\cdot}$ to data $\{(x_t, x_{t-1}), x_{t+2}\}$. Including an intercept, the fitted coefficients are given in Table 1. It appears that the model in region R^+ has local AR coefficients furtherest away from zero and hence has the smallest predictive variance while the model in R^0 has local AR coefficients closest to zero and hence the largest predictive variance.

	AR coefficients		intercept	residual variance
R^+:	0.3848,	-0.0369,	0.0531	0.7532
R^0:	0.1038,	0.1248,	0.1070	1.0234
$R^{\cdot}$:	0.3024,	0.0444,	-0.0360	0.9745

Table 1: Fitted parameters for the three autoregressions.

We have also computed 10-day LLEs (not shown) using the same parameters. The estimates of first 10-day LLEs are all less than -0.5. This implies that the 10-step Jacobian products are close to zero at all time points and the 10-step dynamical map is close to a constant. Thus, randomness dominates for the longer time period in this series.

In conclusion, *we find some evidence of nonlinearity in the short-term behavior of this standardized temperature series, while there is little dynamic structure beyond 10 days.* It is not surprising that Sahay and Sreenivasan (1996) failed to find any evidence of nonlinearity in other daily temperature data by using the dimension estimation and other global nonlinear dynamics techniques. This example demonstrates the advantage of the LLE method in studying the short-term behavior of a system in local state space over traditional global dynamical methods which consistently fail to find nonlinearity globally in climate data sets.

With this still limited experience, we think that the local Lyapunov exponents method is a very promising approach for nonlinear time series analysis and is a strong alternative of global nonlinear dynamics techniques for quantifying nonlinearity. The variability of LLEs over the the state space provide a general test for nonlinearity and is a useful diagnostic tool for exploring the change of predictability in empirical time series.

4 Concluding remarks

In this paper, we develop an approach for studying the finite-time behavior of a nonlinear system in local state space via the local Lyapunov exponents. This promising approach appears to be able to detect nonlinearity in time series when global dynamical techniques such as the global Lyapunov exponents and dimension estimation fail. A rigorous statistical theory for estimating local Lyapunov exponents in multidimensional systems is developed. Explicit expressions for the asymptotic bias and variance are also given. These results are useful for constructing confidence intervals and in the choice of bandwidth for LLE estimation.

References

[1] Bailey, B. A., Ellner, S., and Nychka, D. W. (1997). Chaos with confidence: asymptotics and applications of local Lyapunov exponents. In *Nonlinear Dynamics and Time Series, Building a Bridge between the Natural and Statistical Sciences.* (Eds. Colleen D. Cutler and Daniel T. Kaplan.) 115-133. Fields Institute Communications, Vol. 11. Published by American Mathematical Society, Providence, Rhode Island.

[2] Berliner, L. M. (1992). Statistics, probability and chaos. *Statist. Sci.*, **7**, No.1, 69-122.

[3] Berry, M. V., Percival, I. C., and Weiss, N. O. (1987) (eds.). *Dynamical Chaos : Proceedings of a Royal Society Discussion Meeting.* Royal Society, London.

[4] Cheng, B. and Tong, H. (1992). Consistent nonparametric order determination and chaos. *J. R. Statist. Soc. B*, **54**, No. 2, 427-449.

[5] Drazin,P. G. and King, G. P. (1992) (eds.). *Interpretation of Time Series from Nonlinear Systems. Physica* D, **58**, Nos 1-4, 1-506. North Holland, Amsterdam.

[6] Eaton, M. L. and Tyler, D. E. (1994). The asymptotic distribution of singular values with applications to canonical correlations and correspondence analysis. *J. Multi. Ana.*, **50**, 238-264.

[7] Grenfell, B.T., May, R. M. and Tong, H. (1994), Discussion meeting on chaos and forecasting. *Phil. Trans. R. Soc. Lond.*, **A 348**, 325-530.

[8] Kifer, Y. (1986). *Ergodic Theory of Random Transformations.* Bikhäuser, Boston.

[9] Lorenz, E. N. (1965). A study of the predictability of a 28-variable atmospheric model. *Tellus*, **17**, 321-333.

[10] Lu, Z. Q. (1996a). Multivariate locally weighted polynomial fitting and partial derivative estimation. *J. Multivar. Anal.*, **59**, No. 2, 187-205.

[11] Lu, Z. Q. (1996b). *Multivariate local polynomial fitting for nonlinear time series.* To appear in *Annals of Institute of Statistical Mathematics.*

[12] McCaffrey, D., Nychka, D., Ellner, S. and Gallant, A. R. (1992). Estimating Lyapunov exponents with nonparametric regression. *J. Amer. Statist. Soc.* **87**, No.419, 682-695.

[13] Nychka, D., Ellner, S., McCaffrey, D. and Gallant, A. R. (1992). Finding chaos in noisy systems. *J. R. Statist. Soc. B*, **54**, No. 2, 399-426.

[14] Priestley, M. B. (1988). *Non-linear and Non-stationary Time Series Analysis.* Academic Press, London.

[15] Sahay, A. and Sreenivasan, K. R. (1996). The search for a low-dimensional characterization of a local climate system. *Phil. Trans. R. Soc. Lond. A*, **354**, 1715-1750.

[16] Smith, R. L. (1992). Estimating dimension in noisy chaotic time series. *J. R. Statist. Soc. B*, **54**, No. 2, 329-351.

[17] Tjøstheim, D. and Auestad, B. (1994). Nonparametric identification of nonlinear time series: selecting significant lags. *J. Amer. Statist. Soc.* **89**, No. 428, 1410-1419.

[18] Tong, H. (1990). *Nonlinear Time Series: a Dynamical System Approach.* Oxford Univ. Press, Oxford.

[19] Tong, H. and Smith, R. L. (1992). Royal Statistical Society meeting on chaos. *J. R. Statist. Soc. B*, **54**, No. 2, 301-474.

[20] Tong, H. (1995). A personal overview of non-linear time series analysis from a chaos perspective (with discussions). *Scand. J. statist.*, **22**, 399-445.

[21] Tsay, R. S. (1988). Non-linear time series analysis of blowfly population. *J. Time Series Anal.*, **9**, No. 3, 247-263.

[22] Wolff, R. C. L. (1992). Local Lyapunov exponents: looking closely at chaos. *J. R. Statist. Soc. B*, **54**, No. 2, 353-371.

[23] Yao, Q. and Tong, H. (1994). Quantifying the influence of initial values on nonlinear prediction. *J. R. Statist. Soc. B*, **56**, No. 4, p 701-725.

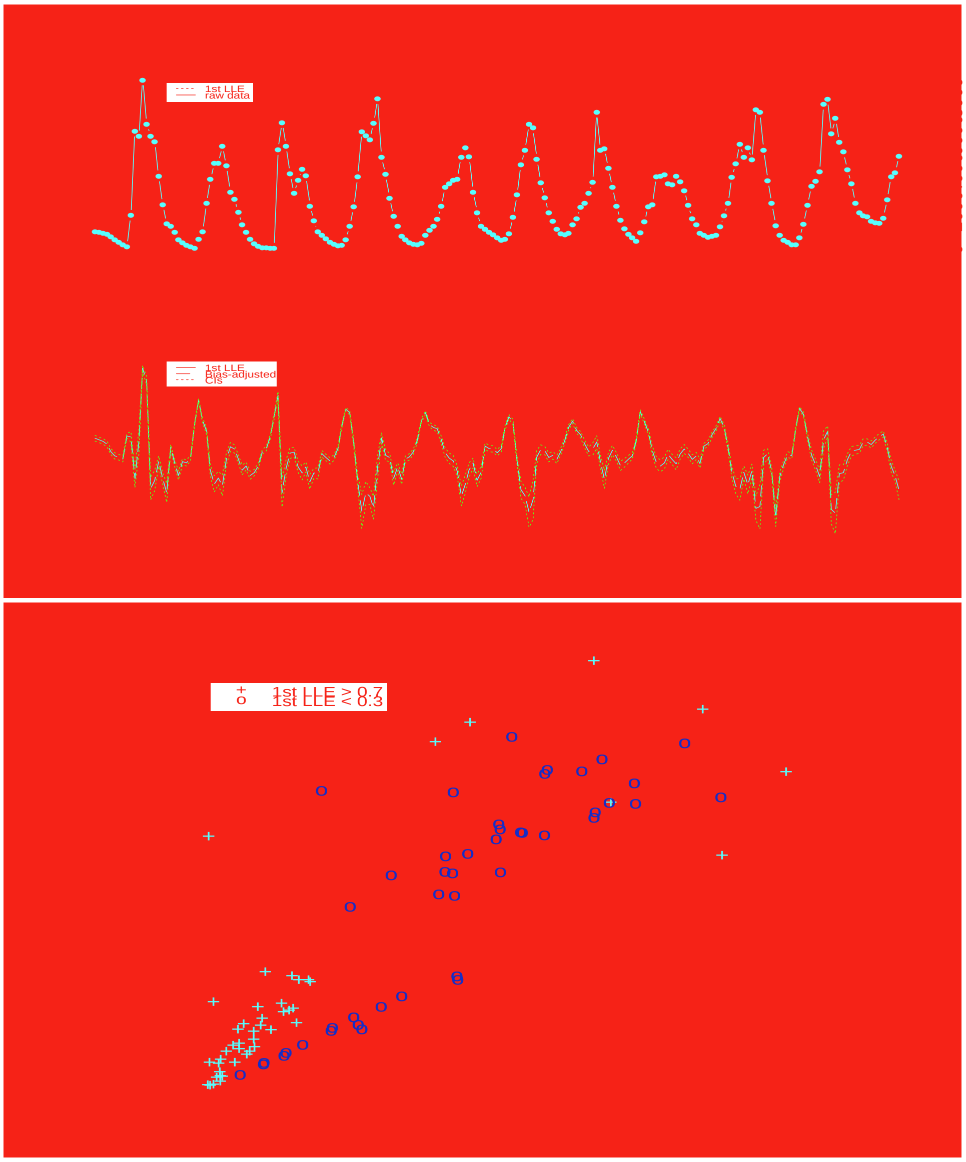

Figure 1: 2-step LLE calculation for the blowfly population data in Example 2. a. Solid line denotes the original data and the dots denote the data points; the dotted line denotes corresponding estimates of first LLE from the same bi-day. The numbers on the right axis denote the scale of the original data. b. 95% pointwise confidence intervals (solid line: raw estimates; long-dashed line: bias-adjusted estimates; dotted lines: confidence bounds). c. Distribution of LLE in the phase space.

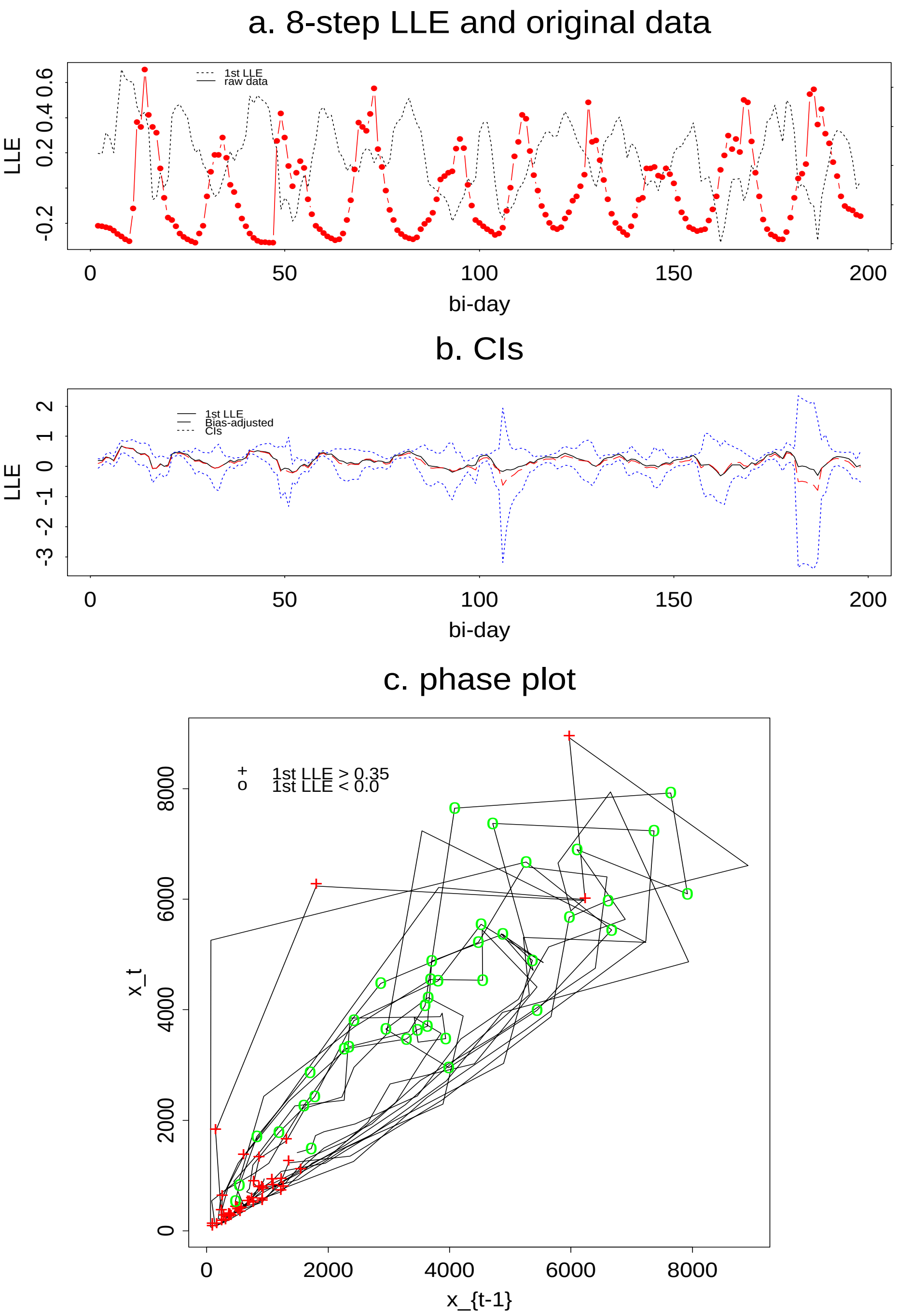

Figure 2: Same as in Figure 1 except for 8-step LLE calculation.

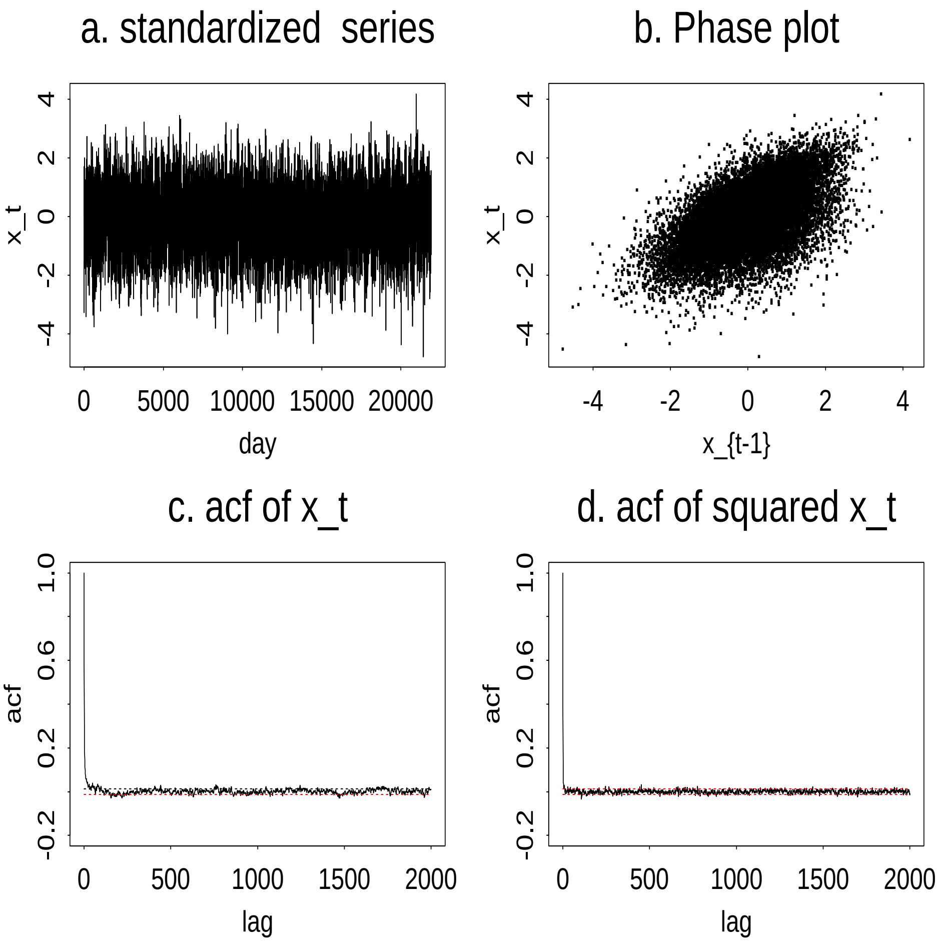

Figure 3: Stationarity test of Charleston temperature anomalies in 1928-1987 in Example 3. a. Standardized series x_t. b. Phase plot. c. Autocorrelation function plot for x_t. d. Autocorrelation function plot for x_t^2. The dotted lines denote 95% confidence intervals under the white noise assumption.

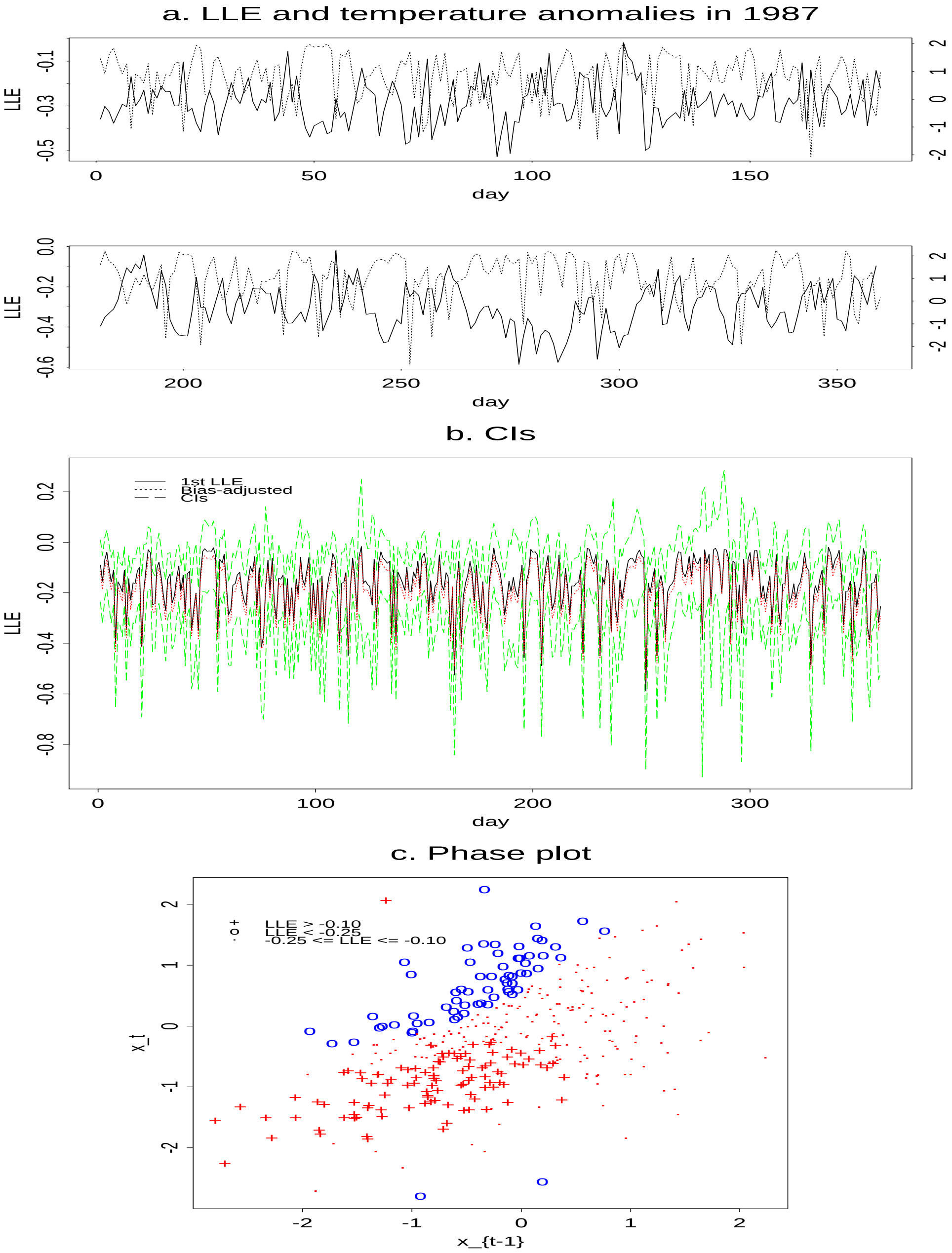

Figure 4: LLE calculations for the standardized temperature series x_t in Figure 3. The numbers on the right axis denote the scale of x_t. a. Solid lines denote x_t while dotted lines denote $\hat{\lambda}_1(2)$. b. 95% confidence intervals (solid line: raw estimates; long-dashed line: bias-adjusted estimates; dotted lines: confidence bounds). c. Distribution of first LLE in the phase space.