Jianqing Fan, Qiwei Yao and Howell Tong

Estimation of conditional densities and sensitivity measures in nonlinear dynamical systems

Article (Accepted version) (Refereed)

Original citation:
Fan, Jianqing and Yao, Qiwei and Tong, Howell (1996) Estimation of conditional densities and sensitivity measures in nonlinear dynamical systems. Biometrika, 83 (1). pp. 189-206.
DOI: 10.1093/biomet/83.1.189

Estimation of Conditional Densities and Sensitivity Measures in Nonlinear Dynamical Systems

Jianqing Fan

Department of Statistics, University of North Carolina

Chapel Hill, NC 27599-3260, U.S.A.

Qiwei Yao and Howell Tong

Institute of Mathematics and Statistics, University of Kent

Canterbury, Kent CT2 7NF, U.K.

Summary

Using locally polynomial regression, we develop nonparametric estimators for the conditional density function and its square root, and their partial derivatives. Two measures of sensitivity to initial conditions in nonlinear stochastic dynamic systems are proposed, one of which relates Fisher information with initial-value sensitivity in dynamical systems. We propose estimators for these, and show asymptotic normality for one of them. We further propose a simple method for choosing the bandwidth. The methods are illustrated by simulation of two well-known models in dynamical systems.

Some key words: Conditional density function; Kullback-Leibler information; Locally polynomial regression; Nonlinear time series; Sensitivity to initial values.

1 Introduction

Nonlinear dynamical systems which exhibit chaos are characterized by the phenomenon that a small perturbation in the initial condition can lead to a considerable divergence of the states of the system in the short or medium term. In a deterministic dynamical system, this phenomenon has been very well documented and is usually analysed by the well-known Lyapunov exponents (cf. Eckmann and Ruelle 1985). However, for a stochastic, i.e. noisy system, further understanding is required. The issue of initial-value sensitivity in a stochastic dynamical system is at the heart of

a proper understanding of chaos in a random environment (see, e.g., Yao and Tong, 1994b) and has in addition important implications for the theory and practice of nonlinear prediction (see, e.g., Yao and Tong 1994a). Tong (1995) and the discussion therein has summarized the various recent approaches to date, including those proposed by Crutchfield *et al.* (1982), Kifer (1986), Wolff (1992) and Yao and Tong (1994a, b).

The goal of this paper is two-fold. First, we note the increasing recent use of nonparametric density estimation to provide diagnostic tools for nonlinear time series modelling. Thus Robinson (1991) used the Kullback-Leibler information criterion for testing nested hypotheses. Skaug and Tjøstheim (1993, unpublished) applied several different distance measures for density functions in testing serial independence. See also Tjøstheim (1994) and the references therein. In all the above work, the standard kernel estimator of a density function was used as the basic building block, and the conditional density function was typically estimated indirectly. We aim to develop a direct estimation method, with good sampling properties. In this paper, the conditional density functions and their square roots, and their partial derivatives, are estimated directly using locally polynomial regression. For more details of the latter method, see, e.g., Fan (1992), Fan *et al.* (1993, unpublished), and Ruppert and Wand (1994). The proposed estimators have been applied to construct predictive distributions of nonlinear time series (cf. Yao and Tong 1995), and to test for independence, as will be reported in a forthcoming paper.

Secondly, we set out to develop some suitable statistical tools to aid understanding of initial-value sensitivity in a stochastic dynamical system. Following Yao and Tong (1994b), we adopt the Kullback-Leibler mutual information and a simple L^2-distance to measure the initial-value sensitivity of the conditional distribution of the state variables in a nonlinear dynamical system. Since both measures are the functionals of the conditional density function, we estimate them using our proposed estimators of the conditional density and its partial derivatives.

The plan of the paper is as follows. In Section 2, we concentrate on the estimators of conditional density functions and the derivatives using, respectively, locally linear and locally quadratic regression. In Section 3, we discuss two sensitivity measures for a stochastic dynamical system and their estimators. In both sections, the asymptotic normality of the estimators are stated, some methods for bandwidth selection are also suggested, and two simulated examples are used as illustration. All technical proofs are briefly outlined in the appendix.

2 Estimation of conditional density and its derivative

2.1 Estimators

We assume that $\{(Y_t, X_t)\}$ is a strictly stationary process having the same marginal distribution as $(Y,\ X)$, where Y is a scalar and X is a d-dimensional vector. Let $g(y|x)$ be the conditional density of Y given X, assumed smooth in both x and y. We use $\dot{g}(y|x)$ to denote the partial derivative of $g(y|x)$ with respect to x. Of interest is the estimation of the functions $g(y|x)$ and $\dot{g}(y|x)$ based on a sequence of observations $(Y_1, X_1), \cdots, (Y_n, X_n)$.

Estimating the conditional density and its derivatives can be regarded as a nonparametric regression problem. To make this connection, note that

$$E\{K_{h_2}(Y-y)|X=x\} \approx g(y|x), \quad \text{as } h_2 \to 0, \tag{2.1}$$

where K is a nonnegative density function and $K_h(z)$ denotes $K(z/h)/h$. The left hand side of (2.1) can be regarded as the regression of $K_{h_2}(Y_i - y)$ on $\{X_i\}$. Recent nonparametric regression theory (see Fan 1992, and Ruppert and Wand 1994) suggests that we may use a locally polynomial regression to estimate $g(y|x)$ and $\dot{g}(y|x)$. For the conditional density, a locally linear fit should be employed, while for its first derivative, locally quadratic fitting is preferable (see Fan and Gijbels, 1995). We treat here the locally quadratic fit more thoroughly, since it is more involved. By Taylor's expansion about $x = (x_1, \ldots, x_d)^T \in R^d$, we have

$$\begin{aligned} E\{K_{h_2}(Y-y)|X=z\} &\approx g(y|z) \\ &\approx g(y|x) + \dot{g}(y|x)^T(z-x) + \frac{1}{2}(z-x)^T\ddot{g}(y|x)(z-x) \\ &\equiv \beta_0 + \beta_1^T(z-x) + \beta_2^T \text{vec}\{(z-x)(z-x)^T\}, \end{aligned}$$

where $\ddot{g}(y|x)$ is the Hessian matrix of $g(y|x)$ with respect to x, $\text{vec}(A) := (a_{11}, a_{22}, \ldots, a_{d,d}, a_{12}, \ldots, a_{1,d}, a_{23}, \ldots, a_{d-1,d})^T \in R^{d(d+1)/2}$ for any $d \times d$ symmetric matrix $A = (a_{ij})$, and

$$\beta_2 := \left(\frac{\partial^2 g(y|x)}{2\partial x_1^2}, \frac{\partial^2 g(y|x)}{2\partial x_2^2}, \ldots, \frac{\partial^2 g(y|x)}{2\partial x_d^2}, \frac{\partial^2 g(y|x)}{\partial x_1 \partial x_2}, \ldots, \frac{\partial^2 g(y|x)}{\partial x_1 \partial x_d}, \frac{\partial^2 g(y|x)}{\partial x_2 \partial x_3}, \ldots, \frac{\partial^2 g(y|x)}{\partial x_{d-1} \partial x_d}\right)^T.$$

This suggests the following least squares problem: let $\hat{\beta}_0$ and $\hat{\beta}_1$ and $\hat{\beta}_2$ minimize

$$\sum_{i=1}^{n} \left[K_{h_2}(Y_i - y) - \beta_0 - \beta_1^T(X_i - x) - \beta_2^T \text{vec}\{(X_i - x)(X_i - x)^T\}\right]^2 W_{h_1}(X_i - x), \tag{2.2}$$

where W is a nonnegative kernel function, and h_1 is the bandwidth. We can estimate

$$\hat{g}(y|x) = \hat{\beta}_0 \text{ and } \dot{g}(y|x) = \hat{\beta}_1.$$

Here

$$\hat{\beta} := (\hat{\beta}, \hat{\beta}_1^T, \hat{\beta}_2^T)^T = (\mathcal{X}^T \mathcal{W} \mathcal{X})^{-1} \mathcal{X}^T \mathcal{W} \mathcal{Y}, \tag{2.3}$$

where $\mathcal{X}$ is the design-matrix of the least-squares problem (2.2), $\mathcal{W} = \text{diag}(W_{h_1}(X_1 - x), \cdots, W_{h_1}(X_n - x))$, and $\mathcal{Y} = (K_{h_2}(Y_1 - y), \cdots, K_{h_2}(Y_n - y))^T$.

If we use locally constant fitting, setting β_1 and β_2 to 0 in (2.2), the least-squares approach will lead to the conventional kernel estimator for the conditional density function (cf. Rosenblatt 1969). For a locally linear fit, we set $\beta_2 = 0$ in (2.2).

For simplicity of presentation, in the rest of this section we treat only univariate x, i.e. $d = 1$. We have

$$\hat{\beta}_j(x, y) = h_1^{-1} \sum_{i=1}^{n} W_j^n \left(\frac{X_i - x}{h_1} \right) K_{h_2}(Y_i - y), \quad j = 0, 1,$$

where

$$W_j^n(t) := \tau_j^T S_n^{-1} (1, h_1 t, h_1^2 t^2)^T \times W(t),$$

with τ_j the unit vector with $(j+1)^{th}$ element 1, and

$$S_n = \begin{pmatrix} s_{n,0} & s_{n,1} & s_{n,2} \\ s_{n,1} & s_{n,2} & s_{n,3} \\ s_{n,2} & s_{n,3} & s_{n,4} \end{pmatrix}, \quad s_{n,j} = \sum_{i=1}^{n} (X_i - x)^j W_{h_1}(X_i - x). \tag{2.4}$$

2.2 Selection of bandwidths

In this section, we propose a simple and intuitively appealling method for choosing the smoothing parameters. For given bandwidth h_2, (2.2) is a standard nonparametric problem of regressing $Z_i(y) = K_{h_2}(Y_i - y)$ on X_i. A simple and appealling bandwidth selection rule is the Residual Squares Criterion proposed in Fan and Gijbels (1995), which translates into our specific case as follows. Let $\hat{Z}_i(y)$ be the fitted value for the regression problem (2.2), and define the normalized weighted residual sum of squares by

$$\hat{\sigma}^2(x, y; h_1) = \frac{1}{\text{tr}(W - S_n^{-1} T_n)} \sum_{i=1}^{n} \Big\{ Z_i(y) - \hat{Z}_i(y) \Big\}^2 W_{h_1}(X_i - x),$$

where $S_n = \mathcal{X}^T\mathcal{W}\mathcal{X}$ and $T_n = \mathcal{X}^T\mathcal{W}^2\mathcal{X}$. Let

$$RSC(x, y; h_1) = \sigma^2(x, y; h_1)\{1 + 3V_n(x; h_1)\}, \tag{2.5}$$

where $V_n(x; h_1)$ is the first diagonal element of the matrix $S_n^{-1}T_nS_n^{-1}$. This estimates the mean squared error at the point x.

For given h_2 and y, the proposed bandwidth h_1 for estimating $\dot{g}(y|x)$ using (2.2) is

$$\hat{h}_1(y) = adj \ \times \ \text{argmin}_h \int RSC(x, y; h)dx, \tag{2.6}$$

where the integration is over the region for x where the curve has to be estimated. Here, the constant adj, depending on the kernel function W, is used to adjust the selected bandwidth so that it converges to the theoretically optimal one. From Table 1 of Fan and Gijbels (1995), $adj = 0.7643$ for the Epanechnikov kernel $W(x) = 0.75(1-x^2)_+$ and $adj = 0.8403$ for the Gaussian kernel $W(x) = (2\pi)^{-1/2}\exp(-x^2/2)$. A similar discussion can be made for the locally linear fit.

The proposed bandwidth (2.6) depends on y. If a constant suffices, we could select

$$\check{h}_1 = adj \ \times \ \text{argmin}_h \int\int RSC(x, y; h)dxdy, \tag{2.7}$$

where the integration is over the region of x and y of interest.

Now consider h_2. For simplicity, we use the normal referencing rule (p.45 of Silverman, 1986), yielding

$$\hat{h}_2 = \left[\frac{8\pi^{1/2}\int K^2(x)dx}{3\{\int x^2K(x)dx\}^2}\right]^{1/5} s_y n^{-1/5}, \tag{2.8}$$

where s_y is the sample standard deviation of $\mathcal{Y}$. When K is the Gaussian kernel, $\hat{h}_2 = 1.06s_yn^{-1/5}$; for the Epanechnikov kernel, $\hat{h}_2 = 2.34s_yn^{-1/5}$.

2.3 Examples

We illustrate the methods via two simulated models. We choose both kernels K and W to be Gaussian.

Example 1. We begin with a simple quadratic model

$$X_t = 0.23X_{t-1}(16 - X_{t-1}) + 0.4\epsilon_t \qquad (t \geq 1), \tag{2.9}$$

where ϵ_t, $t \geq 1$, are independent random variables with the same distribution as the sum of 48 independent random variables each uniformly distributed on [-0.25, 0.25]. According to the central

limit theorem, ϵ_t can be treated as nearly a standard normal variable. However, it has bounded support [-12, 12]. Bounded support is necessary for stationarity (Chan and Tong, 1994). A sample of 1000 was generated from (2.9). We consider three cases: $Y_t = X_{t+m}$ for $m = 1, 2, 3$. We obtained $\hat{h}_2 = 0.98$ from (2.8). Using (2.7), the selected values for $\check{h}_1$ are 0.62 for $m = 1$, 0.70 for $m = 2$, and 0.71 for $m = 3$. The estimated conditional density functions $\hat{g}_m(y|x) \equiv \hat{\beta}_0(x, y)$ are displayed in Figure 1, which shows that, given $X_t = x$, the density of X_{t+m} is around $f^{(m)}(x)$, where $f(x) = 0.23x(16 - x)$, and $f^{(m)}$ denotes the m-th fold composition of f $(m = 1, 2, 3)$.

(Figures 1 is about here.)

Example 2. Consider the cosine model

$$X_t = 20 \cos\left(\frac{\pi X_{t-1}}{10}\right) + \epsilon_t, \tag{2.10}$$

where ϵ_t, $t \geq 1$, are independent standard normal random variables. A sample of 1000 was generated from the above model. ¿From (2.8), we obtain $\hat{h}_2 = 3.65$. Using (2.7) again, the selected values for $\check{h}_1$ are 1.12 for $m = 1$, 1.32 for $m = 2$, and 1.51 for $m = 3$. The estimated conditional density functions $\hat{g}_m(y|x) \equiv \hat{\beta}_0(x, y)$ are displayed in Figure 2.

(Figure 2 is about here.)

2.4 Sampling properties

Let $\mathcal{F}_i^k$ be the σ-algebra of events generated by the random variables $\{X_j, Y_j, i \leq j \leq k\}$ and $L_2(\mathcal{F}_i^k)$ the collection of all $\mathcal{F}_i^k$-measurable square integrable random variables. Let

$$\rho(k) := \sup_{U \in L_2(\mathcal{F}_{-\infty}^0), V \in L_2(\mathcal{F}_k^\infty)} \frac{|\mathrm{cov}(U, V)|}{\mathrm{var}^{1/2}(U)\mathrm{var}^{1/2}(V)} \tag{2.11}$$

denote the ρ-mixing coefficient (Kolmogorov and Rozanov, 1960). We first impose some regularity conditions:

(C1) The kernel functions W and K are symmetric and bounded with bounded supports.

(C2) The process $\{X_j, Y_j\}$ is ρ-mixing with $\sum \rho(\ell) < \infty$. Further, there exists a sequence of positive integers $s_n \to \infty$ such that $s_n = o\{(nh_1h_2)^{1/2}\}$ and $\{n/(h_1h_2)\}^{1/2}\rho(s_n) \to 0$.

(C3) The function $g(y|x)$ has bounded continuous third order derivatives with respect to x at (x, y), and $p(x)$ is continuous at x.

(C4) The joint density of the distinct elements of $(X_0, Y_0, X_\ell, Y_\ell)$ $(\ell > 0)$ is bounded by a constant independent of ℓ.

(C5) The bandwidths h_1 and h_2 converge to zero in such a way that $nh_1^3 h_2 \to \infty$.

Condition (C1) is imposed for brevity of proofs, and could be removed. In particular, the Gaussian kernel is allowed. The assumption on the convergence rate of $\rho(\ell)$ in (C2) is also for technical convenience, and not the weakest possible.

Theorem 1. *Under Conditions (C1) — (C5), for $x \in \{x : p(x) > 0\}$, the two random variables $(nh_1h_2)^{1/2}\{\hat{g}(y|x) - g(y|x) - \vartheta_{n,1}\}$ and $(nh_1^3h_2)^{1/2}\{\hat{\dot{g}}(y|x) - \dot{g}(y|x) - \vartheta_{n,2}\}$ are jointly asymptotically normal with means values 0, variance $\sigma_1^2(x,y)$ and $\sigma_2^2(x,y)$, and covariance 0, where*

$$\vartheta_{n,1} = \frac{1}{2}\mu_K \frac{\partial^2 g(y|x)}{\partial y^2} h_2^2 + o(h_1^3 + h_2^2), \quad \sigma_1^2(x,y) = \frac{g(y|x)\nu_0\nu_K}{p(x)} \frac{\mu_4^2\nu_0 - 2\mu_2\mu_4\nu_2 + \frac{1}{2}\mu_2^2\nu_4}{(\mu_4 - \mu_2^2)^2},$$

$$\vartheta_{n,2} = \frac{\mu_4}{6\mu_2}\frac{\partial^3 g(y|x)}{\partial x^3} h_1^2 + \frac{1}{2}\mu_K \frac{\partial^3 g(y|x)}{\partial x \partial y^2} h_2^2 + o(h_1^2 + h_2^2), \quad \sigma_2^2(x,y) = \frac{g(y|x)\nu_K}{p(x)} \frac{\nu_0\nu_2}{\mu_2^2},$$

and $\mu_K = \int t^2 K(t)\mathrm{d}t$, $\nu_K = \int \{K(t)\}^2 \mathrm{d}t$, $\mu_j = \int t^j W(t)\mathrm{d}t$, $\nu_j = \int t^j \{W(t)\}^2 \mathrm{d}t$ $(j \geq 0)$.

Remark. If our interest is to estimate the conditional density, then locally linear, rather than locally quadratic, regression suffices. In that case, the asymptotic normality admits a more symmetric form:

$$(nh_1h_2)^{1/2}\left\{\hat{g}(y|x) - g(y|x) - \frac{h_1^2\mu_2}{2}\frac{\partial^2 g(y|x)}{\partial x^2} - \frac{h_2^2\mu_K}{2}\frac{\partial^2 g(y|x)}{\partial y^2}\right\} \xrightarrow{\mathcal{L}} N\left\{0,\ \nu_K\nu_0\frac{g(y|x)}{p(x)}\right\},$$

under the assumptions (C1) – (C4) and $nh_1h_2 \to \infty$. Our results and proofs can be readily extended to higher order polynomial regression.

3 Initial-value sensitivity of a stochastic dynamical system

3.1 Sensitivity measures

A discrete-time stochastic dynamical system can be described by the equation

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

for $t \geq 1$, where X_t denotes a state vector in R^d, $F(.)$ is a real vector-valued function, and $\{e_t\}$ is a noise process satisfying $\mathrm{E}(\, e_t \,|\, X_{t-s}$ for $s \geq 1\,) = 0$. The additive dynamic noise model, $X_t = F(X_{t-1}) + e_t$, is a special case. The nonlinear autoregressive model can also be regarded as a special case of model (3.1). Suppose that $\{Y_t,\ -\infty < t < \infty\}$ is a one-dimensional strictly stationary time series, which is d-dependent ($d \geq 1$) in the sense that, given $\{Y_i,\ i \leq t\}$, the conditional distribution of Y_{t+1} depends on $\{Y_i,\ i \leq t\}$ only through $X_t := (Y_t, Y_{t-1}, \ldots, Y_{t-d+1})^T$. Let $f(x) = E(Y_1|X_0 = x)$. Then Y_t can be expressed as

$$Y_t \ = \ f(X_{t-1}) \ + \ \epsilon_t, \tag{3.2}$$

where $\epsilon_t = Y_t - f(X_{t-1})$. Define $F(X_{t-1}) = (f(X_{t-1}), Y_{t-1}, \ldots, Y_{t-d+1})^T$, $e_t = (\epsilon_t, 0, \ldots, 0)^T$. Then equation (3.1) holds with additive noise.

For a stochastic system with additive noise, several recent attempts have been made to extend the notion of a Lyapunov exponent from a deterministic system to a stochastic system. However, the problem of measuring the sensitivity of a stochastic dynamical system is still open. For example, Crutchfield *et al.*(1982) and Kifer (1986) suggested the use of a probability average in the conventional definition of the Lyapunov exponent, initially designed for a deterministic system. However, this seems to lose its intuitive appeal. Wolff (1992) proposed a local Lyapunov exponent, which replaces the above probability average by a local average. Yao and Tong (1994a) considered the divergence of the conditional expectation with respect to a small disturbance in initial values. Wolff's and Yao and Tong's approaches appear to be closely related to each other, in that both concentrate on the divergence of the average orbit and both are designed to capture only the short- to medium-term divergence. The measures proposed by Yao and Tong (1994a) were directly motivated by the pointwise prediction problem.

An alternative and more informative way is to consider the global divergence of the conditional distribution of X_m given X_0. Similar to Wolff (1992) and Yao and Tong (1994a), we only consider the case that m is finite, because due to the accumulation of noise through the time evolution, the system seems unlikely to have a strong memory of its initial value after a long time. Let $g_m(y|x)$ denote the conditional density of X_m given $X_0 = x$. Several measures for the discrepancy of two densities are available. See, for example, Blyth (1994). In this paper, we adopt the following two

indices. Let x and $x+\delta \in R^d$ be two nearby initial values. The L_2-distance is simply defined as

$$D_m(x;\delta) = \int \{g_m(y|x+\delta) - g_m(y|x)\}^2 \mathrm{d}y.$$

We also consider the mutual information based on the Kullback-Leibler information, which may be expressed as follows

$$K_m(x;\delta) = \int \{g_m(y|x+\delta) - g_m(y|x)\} \log\{g_m(y|x+\delta)/g_m(y|x)\} \mathrm{d}y.$$

We assume that $g_m(y|x)$ is smooth in both x and y, and partial differentiation with respect to x and integration with respect to y of the function $g_m(y|x)$ are interchangeable where required. We also assume that integrations in (3.3) and (3.5) below exist and are finite. It follows from the Taylor expansion that

$$D_m(x;\delta) = \delta^T I_{1,m}(x)\delta + o(||\delta||^2),$$

where

$$I_{1,m}(x) = \int \dot{g}_m(y|x)\dot{g}_m^T(y|x)\mathrm{d}y. \tag{3.3}$$

Also for small δ, $K_m(x;\delta)$ has the approximation

$$K_m(x;\delta) = \delta^T I_{2,m}(x)\delta + o(||\delta||^2), \tag{3.4}$$

where

$$I_{2,m}(x) = \int \dot{g}_m(y|x)\dot{g}_m^T(y|x)/g_m(y|x)\mathrm{d}y, \tag{3.5}$$

(cf. §2.6 of Kullback 1967). If we treat the initial value x as a parameter vector of the distribution, $I_{2,m}(x)$ is the Fisher information matrix, which represents the information on the initial value $X_0 = x$ contained in X_m. Roughly speaking, (3.4) may be interpreted as saying that the more information X_m contains about the parameter, the more sensitively the distribution depends on the initial condition.

The measures defined above are more informative than those which only focus on the divergence of some characteristics, e.g. the mean, of the conditional distribution (cf. §2.2 of Yao and Tong 1995). For example, the measure $I_{2,m}(x)$ is directly useful in assessing the initial-value sensitivity of predictive intervals (cf. Proposition 2 of Yao and Tong 1995). Further, by Theorem 4.1 of Blyth (1994), we have the following inequality when the system is one-dimensional:

$$I_{2,m}(x) \geq \frac{\lambda_m^2(x)}{\sigma_m^2(x)}, \tag{3.6}$$

where $\sigma_m^2(x) := \mathrm{var}(X_m|X_0 = x)$, and $\lambda_m(x) := \mathrm{d}\{E(X_m|X_0 = x)\}/\mathrm{d}x$ measures the sensitivity of the conditional expectation (cf. Yao and Tong 1994a). Relation (3.6) suggests that, when the conditional variance is large, the sensitivity measure would be small, in agreement with intuition; also see Figures 3 and 4 below. Inequality (3.6) generalises to multivariate cases. As for other measures for the divergence in short or medium term, the actual numerical values of $I_{1,m}(x)$ and $I_{2,m}(x)$ are not informative, so much as their relative magnitudes. For example, the maximizer of $I_{2,m}(x)$ is the location in the the state space from which the system diverges the most after m steps of time evolution.

3.2 Estimating sensitivity measures

It is of both practical and technical interest to consider the divergence in one particular component, e.g. the first component, of system (3.1); see also the time series model (3.2). Thus, given data $(Y_1, X_1), \cdots, (Y_n, X_n)$ as in §2.1, of interest is estimation of the functionals

$$I_1(x) = \int \dot{g}(y|x)\dot{g}^T(y|x)\mathrm{d}y,$$

and

$$I_2(x) = \int \dot{g}(y|x)\dot{g}^T(y|x)/g(y|x)\mathrm{d}y.$$

For clarity, we assume henceforth that $d = 1$.

Using the estimators of $g(y|x)$ and $\dot{g}(y|x)$ derived in §2.1 leads to the following estimator for $I_1(x)$:

$$\begin{aligned}\hat{I}_1(x) &:= \int \hat{\beta}_1^2(x,y)\mathrm{d}y \\ &= \frac{1}{h_1^2}\sum_{i=1}^n\sum_{j=1}^n W_1^n\left(\frac{X_i - x}{h_1}\right) W_1^n\left(\frac{X_j - x}{h_1}\right)\int K_{h_2}(Y_i - y)K_{h_2}(Y_j - y)\mathrm{d}y.\end{aligned}$$

Assume that the kernel $K(\cdot)$ is symmetric. Then,

$$\int K_{h_2}(Y_i - y)K_{h_2}(Y_j - y)\mathrm{d}y = K_{h_2}^*(Y_i - Y_j),$$

where $K^* = K * K$ is the convolution of the kernel function K with itself. Thus, the proposed estimator can be expressed as

$$\hat{I}_1(x) = \frac{1}{h_1^2}\sum_{i=1}^n\sum_{j=1}^n W_1^n\left(\frac{X_i - x}{h_1}\right) W_1^n\left(\frac{X_j - x}{h_1}\right) K_{h_2}^*(Y_i - Y_j). \tag{3.7}$$

Estimation of a quadratic functional of the form $\theta_k = \int f^{(k)}(y)^2 dy$, with $f(y)$ a density, has been extensively studied in the literature. See for example, Hall and Marron (1987), Fan (1991) and Hall and Wolff (1995b) and the references therein. Hall and Marron (1987) propose to reduce the bias by leaving diagonal terms out, similar to the terms with $i = j$ in $\hat{I}_1(x)$; while Jones and Sheather (1991) argue in favour of leaving them. When estimating θ_0, under mild conditions both versions have bias of order $o(n^{-1/2})$ for a large range of bandwidths. For estimating $\hat{I}_1(x)$, since the derivative is taken with respect to x, its behaviour is analogous, and hence the difference between the "diagonal-in" and "diagonal-out" estimators is negligible under mild conditions.

For $I_2(x)$, an intuitive estimator is

$$\hat{I}_2(x) = \int \hat{\beta}_1^2(x,y)/\hat{\beta}(x,y)\mathrm{d}y, \tag{3.8}$$

with the usual convention $0/0 = 0$. The integral is typically finite. However, (3.8) cannot easily be simplified. We thus propose an alternative. Let $q(x,y)$ denote $\{g(y|x)\}^{1/2}$. Then

$$I_2(x) = 4\int \{\dot{q}(x,y)\}^2 \mathrm{d}y.$$

For given bandwidths h_1 and h_2, define, for $1 \le i \le n$,

$$C(X_i, Y_i) = \#\{(X_t, Y_t), 1 \le t \le n \ : \ |X_t - X_i| \le h_1 \text{ and } |Y_t - Y_i| \le h_2\},$$

$$C(X_i) = \#\{X_t, 1 \le t \le n, \ : \ |X_t - X_i| \le h_1\}.$$

Then

$$Z_t := [C(X_t, Y_t)/\{C(X_t)\, h_2\}]^{1/2}$$

is a natural estimate of $q(x,y)$ at $(x,y) = (X_t, Y_t)$. Using locally quadratic regression, we may estimate $q(x,y)$ and its first and second order partial derivatives with respect to x, $\dot{q}(x,y)$ and $\ddot{q}(x,y)$, by $\hat{q}(x,y) = \hat{a}$, $\hat{\dot{q}}(x,y) = \hat{b}$, and $\hat{\ddot{q}}(x,y) = \hat{c}$, where $(\hat{a}, \hat{b}, \hat{c})$ minimises of the function

$$\sum_{t=1}^{n} \{Z_t - a - b(X_t - x) - c(X_t - x)^2/2\}^2 H\left(\frac{X_t - x}{h_1}, \frac{Y_t - y}{h_2}\right),$$

H being a probability density function on R^2. Then, we estimate $I_2(x)$ by

$$\check{I}_2(x) = 4\int \{\hat{\dot{q}}(x,y)\}^2 \, \mathrm{d}y. \tag{3.9}$$

3.3 Selecting bandwidths for $I_1(x)$ and $I_2(x)$

We propose a simple and intuitively appealling method for choosing the smoothing parameters h_1 and h_2 for estimating $I_1(x)$ and $I_2(x)$ (using (3.7) and (3.8)). For estimator (3.9), we have not found a systematic way to choose h_1 and h_2.

For estimating the first derivative, the optimal bandwidth of h_1 is of the order $O(n^{1/7})$ under the assumption that the third derivative with respect to x exists. For that choice of h_1, there are about $N = O(n^{6/7})$ data points in the neighborhood of $x \pm h_1$. The choice of bandwidth h_2 is not very crucial to $\hat{I}_1(x)$ and $\hat{I}_2(x)$, owing to the integration over y (Fan 1991, and Hall and Marron 1991). The choice of order $O(N^{-7/30}) = O(n^{-1/5})$ would be sufficient. To make this order of magnitude meaningful in terms of the scale of y and that of K, we suggest using

$$\check{h}_2 := \alpha \left[\frac{8\pi^{1/2} \int K^2(x)dx}{3\{\int x^2 K(x)dx\}^2} \right]^{1/5} s_y n^{-1/5}, \tag{3.10}$$

where $\alpha \in [0.5, 1)$ is a specified constant, which makes $\check{h}_2$ smaller than $\hat{h}_2$ in (2.8). This is natural, since the integration over y in the definitions of $\hat{I}_1(x)$ and $\hat{I}_2(x)$ reduces the noise level of the estimators, and allows us to use a smaller bandwidth to reduce bias. The above choice of $\check{h}_2$ is also supported by Theorem 2 below: see Remark in §3.5.

Once h_2 is selected, the choice of the bandwidth $\check{h}_1$ is determined by (2.7), which minimizes the average mean squared errors for the derivative curve estimation, as explained above.

We do not claim that any one of the bandwidths (2.6) – (2.8) and (3.10) would be the best choice for all statistical problems. They are quick and simple selection procedures which take the structure and the scale of the data into account, and give us an initial idea as to how much smoothing should be done.

3.4 Examples (continued)

Example 1. The skeleton of (2.9) is a transformed logistic map with coefficient 3.68 (=16×0.23), which is deterministically chaotic. For further relevant discussion of the logistic map, we refer to Hall and Wolff (1995a). With the same sample as used before ($n = 1000$), we estimate $I_{1,m}(x)$ and $I_{2,m}(x)$ for $m = 1, 2$, and 3. Using $\check{h}_2 = 0.8\hat{h}_2$, we estimate $I_{1,m}(x)$ using (3.7). The estimated curves are plotted in Figure 3(a). The sensitivity does vary with the initial value. For example, for $m = 1$, $\hat{I}_1(x)$ attains its minimal value at $x = 8$, monotonically increasing as x moves away

from 8 in either directions. Similar but more complicated conclusions can be drawn for the cases $m = 2, 3$. See also Section 4.1 of Yao and Tong (1994b), and Example 1 of Yao and Tong (1994a). Figures 3(b)– 3(d) show the estimated curves of $I_{2,m}(x)$ using both (3.8) and (3.9). We expect the curves obtained using (3.8) to be somewhat wiggly, owing to the estimator $\hat{\beta}_0(x, y)$ in the denominator. The curves estimated by (3.9) are smoother. However, it remains open how to choose the smoothing parameters using (3.9). For this example, we manually chose bandwidths $(h_1, h_2) = (0.34, 0.68)$, $(0.41, 0.89)$, and $(0.46, 0.85)$ for $m = 1, 2$, and 3 respectively. Although the magnitudes of the functions $I_{1,m}(x)$ and $I_{2,m}(x)$ are different, their profiles are similar.

(Figure 3 is about here.)

Example 2. For model (2.10), the skeleton of this model has a limit point $x = 20$. With $n = 1000$, we estimate $I_{1,m}(x)$ and $I_{2,m}(x)$ for $m = 1, 2$, and 3 again using $\check{h}_2 = 0.8\hat{h}_2$. The resulting estimates are depicted in Figure 4(a). Figures 4(b)–4(d) report the estimated curves of $I_{2,m}(x)$. As in Example 1, the curves obtained using (3.8) are wiggly, while those estimated by (3.9) are smoother. In applying (3.9), we use bandwidths $(h_1, h_2) = (0.89, 1.88)$, $(0.94, 2.00)$, $(1.48, 2.14)$ for $m = 1, 2$, and 3 respectively. The sensitivity measures drop sharply as m increases. Further, for fixed m, the sensitivity varies with the initial value although, owing to the accumulation of considerable random noise, the variation becomes less pronounced when m increases (cf. (3.6)). This example shows that initial-value sensitivity should be taken into account for a nonlinear stochastic system even when it has a non-chaotic skeleton.

(Figure 4 is about here.)

3.5 An asymptotic result

Theorem 2. *Under Conditions (C1) — (C5) given in §2.2, if* $nh_1^3h_2^2 \to \infty$, *for* $x \in \{x : p(x) > 0\}$, $(nh_1^3)^{1/2}\{\hat{I}_1(x) - I_1(x) - \vartheta_n\}$ *is asymptotically normal with mean value 0 and variance* σ^2, *where*

$$\begin{aligned}
\vartheta_n &= h_1^2 \frac{\mu_4}{3\mu_2} \int \left\{ \frac{\partial g(y|x)}{\partial x} \frac{\partial^3 g(y|x)}{\partial x^3} \right\} \mathrm{d}y + h_2^2 \mu_K \int \left\{ \frac{\partial g(y|x)}{\partial x} \frac{\partial^3 g(y|x)}{\partial x \partial y^2} \right\} \mathrm{d}y + o(h_1^2 + h_2^2), \\
\sigma^2 &= \frac{4\nu_2}{\mu_2^2 p(x)} \left[\int \left\{ \frac{\partial g(y|x)}{\partial x} \right\}^2 g(y|x) \mathrm{d}y - \left\{ \int \frac{\partial g(y|x)}{\partial x} g(y|x) \mathrm{d}y \right\}^2 \right].
\end{aligned}$$

Remark. The choice of h_2 for estimating $I_1(x)$ is not as sensitive as that for estimating the conditional density. In fact, for h_2 in the range $(nh_1^3)^{-1/4} >> h_2 >> (nh_1^3)^{-1/2}$, the asymptotic bias and variance of $I_1(x)$ remain approximately the same; i.e. the term $O(h_2^2)$ in ϑ_n becomes negligible. Thus, the optimal choice of bandwidth is $h_1 = cn^{-1/7}$ and $n^{-1/7} >> h_2 >> n^{-2/7}$, where

$$c = \left\{ \frac{27\mu_2^2\sigma^2}{4\mu_4^2} \left(\int \frac{\partial g}{\partial x} \frac{\partial^3 g(y|x)}{\partial x^3} \mathrm{d}y \right)^{-2} \right\}^{-1/7}.$$

Acknowledgments

JF's work was partially supported by NSF Grants and an NSF Postdoctoral Fellowship. QY and HT acknowledge partial support of the Science and Engineering Research Council (UK). We thank the Editor for concrete suggestions on the restructuring of the paper and for bringing to our attention the reference Blyth (1994). The valuable comments and suggestions from an Associate Editor and a referee are also gratefully acknowledged.

Appendix — Proofs

Proof of Theorem 1. Let $m(x,y) := E\{K_{h_2}(Y_i - y)|X_i = x\}$, $H := \mathrm{diag}(1, h_1, h_1^2)$, and

$$\beta := (m_0(x,y), m_1(x,y), m_2(x,y))^T := \left(m(x,y), \frac{\partial}{\partial x} m(x,y), \frac{1}{2}\frac{\partial^2}{\partial x^2} m(x,y) \right)^T.$$

It follows from (2.3) that

$$H(\hat{\beta} - \beta) = H(\mathbf{X}^T\mathbf{W}\mathbf{X})^{-1}\mathbf{X}^T\mathbf{W}(\mathbf{Y} - \mathbf{X}\beta) = S_n^{*-1}\{(t_{n,0}, t_{n,1}, t_{n,2})^T + (\gamma_{n,0}, \gamma_{n,1}, \gamma_{n,2})^T\}, \quad \text{(A.1)}$$

where

$$t_{n,j} = \frac{1}{n}\sum_{i=1}^n \left(\frac{X_i - x}{h_1} \right)^j W_{h_1}(X_i - x) \left\{ K_{h_2}(Y_i - y) - m(X_i, y) \right\},$$

$$\gamma_{n,j} = \frac{1}{n}\sum_{i=1}^n \left(\frac{X_i - x}{h_1} \right)^j W_{h_1}(X_i - x)\{m(X_i, y) - m(x,y) - m_1(x,y)(X_i - x) - m_2(x,y)\frac{(X_i - x)^2}{2}\},$$

$$s_{n,j}^* = \frac{1}{n}\sum_{i=1}^n \left(\frac{X_i - x}{h_1} \right)^j W_{h_1}(X_i - x),$$

and S_n^* is a 3×3 matrix with the (i,j)-element s_{i+j-2}^*. Let S and Σ be 3×3 matrices with (i,j)-elements μ_{i+j-2}, and ν_{i+j-2} respectively, and $\gamma := (\mu_3, \mu_4, \mu_5)^T$. The basic idea is to establish

(a) S_n^* converges to $p(x)S$ in mean square.

(b) $h_1^{-3}(\gamma_{n,0}, \gamma_{n,1}, \gamma_{n,2})^T$ converges to $6^{-1}p(x)\gamma\partial^3 g(y|x)/\partial x^3$ in mean square.

(c) $(nh_1h_2)^{1/2}(t_{n,0}, t_{n,1}, t_{n,2})$ is asymptotically normal with mean 0 and variance $g(y|x)p(x)\nu_0\nu_K\Sigma$.

Combining these with (A.1), we have

$$pr\left[(nh_1h_2)^{1/2}\left\{\frac{g(y|x)\nu_0\nu_K S^{-1}\Sigma S^{-1}}{p(x)}\right\}^{-1/2}\left\{H(\hat{\beta}-\beta) - \frac{1}{6}h_1^3\frac{\partial^3 g(y|x)}{\partial x^3}S^{-1}\gamma\right\} < x\right] \to \Phi(x), \tag{A.2}$$

where $\Phi(.)$ denotes the standard normal distribution function. It follows from the Taylor expansion that

$$m_j(x,y) = \frac{\partial^j g(y|x)}{\partial x^j} + \frac{1}{2}h_2^2\mu_K\frac{\partial^{j+2}g(y|x)}{\partial x^j \partial y^2} + o(h_2^2).$$

Using this expansion and considering the marginal distribution of (A.2), we obtain the result.

Conclusions (a) and (b) can be proved by computing the means and the variances of $s_{n,j}^*$ and $\gamma_{n,j}$ by using the stationarity and mixing conditions.

To prove (c), we consider arbitrary linear combinations of $t_{n,j}$ with constant coefficients η_j $(j = 0, 1, 2)$. Let

$$\begin{aligned} & Q_n = (nh_1h_2)^{1/2}(\eta_0 t_{n,0} + \eta_1 t_{n,1} + \eta_2 t_{n,2}) \\ = \;& n^{-1/2}\sum_{i=1}^{n}(h_1h_2)^{1/2}D_{h_1}(X_i - x)\{K_{h_2}(Y_i - y) - m(X_i, y)\}, \end{aligned} \tag{A.3}$$

where $D(u) = (\eta_0 + \eta_1 u + \eta_2 u^2)W(u)$. Write $Q_n = n^{-1/2}(Z_{n,0} + \ldots + Z_{n,n-1})$. Note that Q_n is the sum of a stationary mixing sequence. Asymptotic normality follows from standard small-block and large-block arguments. Details can be founded in Fan, Yao and Tong (1993, unpublished).

Proof of Theorem 2. We adopt the notation introduced in the proof of Theorem 1. Let $\xi_{n,j}(x,y) = (t_{n,j} + \gamma_{n,j})/h_1$. To prove Theorem 2, we need the following asymptotic results:

(d) $E\int\{\xi_{n,1}(x,y)\}^2 \mathrm{d}y = O\{h_1^4 + (nh_1^3h_2)^{-1}\} = o\{h_1^2 + (nh_1^3)^{-1/2}\}$,

(e) $(nh_1^3)^{1/2}\{\int \xi_{n,1}(x,y)m_1(x,y)\mathrm{d}y - \frac{\mu_4 p(x)}{6}\int\frac{\partial g(y|x)}{\partial x}\frac{\partial^3 g(y|x)}{\partial x^3}dyh_1^2 + o(h_1^2)\} \xrightarrow{\mathcal{L}} N(0, \sigma_0^2)$,

where $\sigma_0^2 = \nu_2 p(x)\left[\int\{\dot{g}(y|x)\}^2 g(y|x)dy - \{\int \dot{g}(y|x)g(y|x)\mathrm{d}y\}^2\right]$.

By (A.1), we have that $\hat{\beta}_1(x,y) - m_1(x,y) = (0,1,0)S_n^{*-1}(\xi_{n,0},\xi_{n,1},\xi_{n,2})^T$. It follows from (a) that

$$\begin{aligned}
& \hat{I}_1(x) - \int\{m_1(x,y)\}^2 \mathrm{d}y \\
= & \int\{\hat{\beta}_1(x,y) - m_1(x,y)\}^2 \mathrm{d}y + 2\int m_1(x,y)\{\hat{\beta}_1(x,y) - m_1(x,y)\}\mathrm{d}y \\
= & \left\{\frac{1}{p^2(x)\mu_2^2}\int \xi_{n,1}^2(x,y)\mathrm{d}y + \frac{2}{p(x)\mu_2}\int \xi_{n,1}(x,y)m_1(x,y)\mathrm{d}y\right\}\{1+o_p(1)\}. \qquad \text{(A.4)}
\end{aligned}$$

Since $\int\{m_1(x,y)\}^2\mathrm{d}y = I_1(x) + h_2^2\mu_K\int\{\partial g(y|x)/\partial x\}\{\partial^3 g(y|x)/\partial x\partial y^2\}\mathrm{d}y + o(h^2)$, Theorem 5.2 follows immediately from (d), (e), and (A.4).

The proof of (d) is similar to that of (a) and is omitted here. To prove (e), we define

$$U(x_1,y_1;x,y) := h_1^{-2}(x_1-x)W_{h_1}(x_1-x)\{K_{h_2}(y_1-y) - m(x,y) - m_1(x,y)(x_1-x) - m_2(x,y)(x_1-x)^2/2\}$$

and $V(x_1,y_1) := \int U(x_1,y_1;x,y)m_1(x,y)dy$. Then, $\int \xi_{n,1}(x,y)m_1(x,y)\mathrm{d}y = n^{-1}\{V(X_1,Y_1) + \ldots + V(X_n,Y_n)\}$. It can be shown via a Taylor expansion that

$$EV(X_1,Y_1) = \frac{p(x)\mu_4}{6}\int \frac{\partial g(y|x)}{\partial x}\frac{\partial^3 g(y|x)}{\partial x^3}dy\, h_1^2 + o(h_1^2)$$

and that

$$EU^*(X_1,Y_1;x,y)U^*(X_1,Y_1;x,y+h_2z) = h_1^{-3}h_2^{-1}g(y|x)p(x)\nu_2\int K(u)K(u+z)du\{1+o(1)\},$$

where $U^*(x_1,y_1;x,y) = h_1^{-2}(x_1-x)W_{h_1}(x_1-x)K_{h_2}(y_1-y)$. Thus,

$$\begin{aligned}
\mathrm{var}\{V(X_1,Y_1)\} &= EV^2(X_1,Y_1) + O(h_1^4) \\
&= h_2\int\int EU^*(X_1,Y_1;x,y)U^*(X_1,Y_1;x,y+h_2z)m_1(x,y)m_1(x;y+h_2z)dydz \\
&\quad -E\{\int h_1^{-2}(X_1-x)W_{h_1}(X_1-x)m(X_1,y)m_1(x,y)dy\}^2\{1+o(1)\} \\
&= h_1^{-3}p(x)\nu_2[\int m_1^2(x,y)g(y|x)dy - \{\int m_1(x,y)g(y|x)dy\}^2]\{1+o(1)\}.
\end{aligned}$$

Now, using big-small block arguments, we establish (e).

References

Blyth, S. (1994). Local divergence and association. *Biometrika*, **81**, 579-584.

Chan, K.S. and Tong, H. (1994) A note on noisy chaos. *J. Roy. Statist. Soc.* **B**, **56**, 301-311.

Crutchfield, J.P., Farmer, J.D. and Huberman, B.A. (1982). Fluctuations and simple chaotic dynamics. *Phys. Rep.*, **92**, 45-82.

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

Fan, J. (1991). On the estimation of quadratic functionals. *Ann. Statist.*, **19**, 1273-1294.

Fan, J. (1992). Design-adaptive nonparametric regression. *J. Amer. Statist. Assoc.*, **87**, 998-1004.

Fan, J. and Gijbels, I. (1995). Data-driven bandwidth selection in local polynomial fitting: variable bandwidth and spatial adaptation. *J. Roy. Statist. Soc.* **B**, **57**, 371-394.

Kifer, Y. (1986). *Ergodic Theory of Random Transformations.* Birkhäuser, Basel.

Hall, P. and Marron, J. S. (1987). Estimation of integrated squared densit y derivatives. *Statist. and Prob. Letters* **6** 109-115.

Hall, P. and Marron, J.S. (1991). Lower bounds for bandwidth selection in density estimation. *Prob. Th. Rel. Fields.*, **90**, 149-173.

Hall, P. and Wolff, R.C.L. (1995a). Properties of invariant distributions and Lyapunov exponents for chaotic logistic maps. *J. Roy. Statist. Soc.* **B**, **57**, 439-452.

Hall, P. and Wolff, R.C.L. (1995b). Simple, optimal estimators of integrals of powers of density derivatives. *Statist. Prob. Letters*, to appear.

Jones, M.C. and Sheather, J.S. (1991). Using non-stochastic terms to advantage in estimating integrated squared density derivatives. *Statist. Probab. Letters*, **11**, 511 – 514.

Kolmogorov, A.N. and Rozanov, Yu. A. (1960). On strong mixing conditions for stationary Gaussian processes. *Th. Prob. Appl.*, **52**, 204-207.

Kullback, S. (1967). *Information Theory and Statistics.* Dover Publi., New York.

Robinson, P.M. (1991). Consistent nonparametric entropy-based testing. *Rev. Econ. St.*, **58**, 437-453.

Rosenblatt, M. (1969). Conditional probability density and regression estimators. *Multivariate Analysis II* (Edited by P.R. Krishnaiah), 25-31. Academic Press, New York.

Ruppert, D. and Wand, M.P. (1994). Multivariate locally weighted least squares regression. *Ann. Statist.*, **22**, 1346 – 1370.

Silverman, N.B. (1986). *Density Estimation for Statistics and Data Analysis.* Chapman and Hall, London.

Tjøstheim, D. (1994). Non-linear time series: a selective review. *Scand. J. Statist*, **21**, 97-130.

Tong, H. (1995). A personal overview of nonlinear time series from a chaos perspective (with discussion). *Scand. J. Statist.* (in the press.)

Wolff, R.C.L. (1992). Local Lyapunov exponents: looking closely at chaos. *J. Roy. Statist. Soc.* **B**, **54**, 353-371.

Yao, Q. and Tong, H. (1994a). Quantifying the influence of initial values on nonlinear prediction. *J. Roy. Statis. Soc.* **B**, **56**, 701-725.

Yao, Q. and Tong, H. (1994b). On prediction and chaos in stochastic systems. *Phil. Trans. Roy. Soc.* **A**, **348**, 357-369.

Yao, Q. and Tong, H. (1995). On initial-condition sensitivity and prediction in nonlinear stochastic systems. *Bull. Int. Statist. Inst., 50th Session, Beijing, China*, 395-412.

Figure Captions

Figure 1 The estimated $g_m(y|x)$: the conditional density function of Y_{t+m} given $Y_t = x$ for the logistic model (2.9). (a) $m = 1$; (b) $m = 2$; (c) $m = 3$.

Figure 2 The estimated $g_m(y|x)$: the conditional density function of Y_{t+m} given $Y_t = x$ for the cosine model (2.10). (a) $m = 1$; (b) $m = 2$; (c) $m = 3$.

Figure 3 The estimated sensitivity measures for the logistic model (2.9). (a) Estimated curve $I_{1,m}(x)$ with $m = 1$ (solid curve), $m = 2$ (dashed curve) and $m = 3$ (long dashed curve); (b)–(d) Estimated sensitive measure $I_{2,m}(x)$ for $m = 1, 2, 3$. Solid curve: estimated by (3.8); dashed curve: estimated by (3.9).

Figure 4 The estimated sensitivity measures for the cosine model (2.9). See the caption of Figure 3 for details.