Biometrika, *pp.* 1–20

Printed in Great Britain

Inferring Stochastic Dynamics from Functional Data

By Nicolas Verzelen,

Institut National de Recherche Agronomique, 2, place Pierre Viala
F-34060 Montpellier, FRANCE.
nicolas.verzelen@supagro.inra.fr

Wenwen Tao and Hans-Georg Müller

Department of Statistics, University of California, Davis, One Shields Avenue, Davis, California, 95616, U.S.A.
wtao@ucdavis.edu mueller@wald.ucdavis.edu

Summary

In most current data modelling for time-dynamic systems, one works with a pre-specified differential equation and attempts to estimate its parameters. In contrast, we demonstrate that in the case of functional data, the equation itself can be inferred. Assuming only that the dynamics are described by a first order nonlinear differential equation with a random component, we obtain data-adaptive dynamic equations from the observed data via a simple smoothing-based procedure. We prove consistency and introduce diagnostics to ascertain the fraction of variance that is explained by the deterministic part of the equation. This approach is shown to yield useful insights into the time-dynamic nature of human growth.

Some key words: Empirical Dynamics, Functional Data Analysis, Goodness of Fit, Growth Curves, Smoothing

1. Introduction

In recent years, there has been increasing interest in fitting nonlinear differential equations to data arising in engineering, economics or biology. A major motivation is to understand the dynamics underlying physical or biological processes (Holte et al., 2006; Perelson et al., 1997) or to predict the future behavior of such systems from current observations. These challenges arise in growth studies (Gasser et al., 1984), where, in addition to scientific interest in understanding the dynamics of human growth by studying how growth velocity relates to current age and current height, differential equation models can also be used to assess clinical aspects of a child's growth patterns. A differential equation model that fits the data can be applied to predict the size of the derivative of growth for a healthy child that is low on height for current age. This predicted derivative can then be checked against the observed derivative for monitoring purposes.

Substantial work has been devoted to parametric estimation procedures for dynamic systems (Bellman & Roth, 1971; Brunel, 2008; Liang & Wu, 2008; Ramsay et al., 2007). These, and also recent semiparametric approaches (Chen & Wu, 2008; Paul et al., 2011) for modelling dynamic systems, rely on the fact that a pre-specified non-random differential equation applies to the data. However, this is often not the case, particularly in the study of dynamics that are repeatedly observed for many subjects or experiments. There are two major reasons for discrepancies between stipulated dynamic models and actual behavior of systems. First, differential equation models have been traditionally accepted and based on their inherent plausibility and concordance with presumed underlying mechanisms. All too often, this leads to models that actually do not fit the data well (Hooker, 2009), because the presumed underlying mechanisms that the model reflects are not well understood or do not provide good approximations to the actual mechanisms. Second, deterministic models rarely provide satisfactory fits to phenomena that are inherently stochastic in nature, because the dynamics vary across subjects or experiments. Dynamics of viral level in HIV studies (Miao et al., 2009) that are subject-specific and the dynamics of auction price trajectories (Reddy & Dass, 2006; Wang et al., 2008) provide examples of this difficulty. In such cases, subject-specific effects come into play that cannot be controlled for, and it is then not reasonable to expect a deterministic dynamic equation to provide a good fit across subjects.

All of this motivates an alternative bottom-up approach, namely to directly obtain information about underlying dynamic systems from repeated observations of the trajectories that result from the dynamics, in contrast to the customary top-down approach of a priori postulating what the dynamic equations should be. Our aim thus is to derive differential equations from functional data, i.e., learning these equations from observing many realizations of the trajectories that they generate. To allow for random variation between subjects, it is necessary to add stochastic elements to a deterministic equation. For this, inclusion of an additional stochastic drift process is expedient.

Nonparametric analysis of stochastic differential equations has been previously studied for diffusion processes (Hoffmann, 1999; Jacod, 2000), with solutions that are versions of Brownian motion and have non-differentiable trajectories. As growth and many other dynamic phenomena are usually considered to be quite smooth, the stochastic differential equation approach is not useful for most non-financial data. Recently, Müller & Yao (2010) have investigated an empirical dynamic approach, where one determines linear dynamics empirically from a sample of trajectories. Specifically, each trajectory of a differentiable Gaussian process is shown to satisfy a first order linear differential equation, which can be determined for various types of longitudinal data by suitable estimation procedures. However, this approach does not extend to nonlinear dynamic systems or non-Gaussian processes.

Here we show that each trajectory of a smooth stochastic process X satisfies a first order nonlinear differential equation with a random component, where the stochastic part is an additive smooth drift process Z. We call this representation of the process the data-driven differential equation. The variance of the process Z determines to what extent the process X is driven by the deterministic part of the differential equation. Whenever the variance of the drift Z is small in comparison to the variance of X, a deterministic version of the differential equation explains most of the observed behavior of the process. Obtaining data-driven dynamics reveals underlying mechanisms generating the observed functional data and provides diagnostic tools for assessing the linearity of the dynamics or the quality of a parametric fit. Implementation proceeds via a two-step kernel estimation procedure, which we show to be consistent. We illustrate the method by constructing the data-driven differential equation governing the growth of children for the Berkeley Growth Study.

We conclude this section by describing the data structure of the available observations from which the dynamics will be learned. Given n realizations X_i of the underlying process X on a domain $\mathcal{T}$, we assume that N_i measurements Y_{ij} ($i = 1, \ldots, n$, $j = 1, \ldots, N_i$), where $N = \inf_{i=1,\ldots,n} N_i$, are obtained at times t_{ij} according to

$$Y_{ij} = Y_i(t_{ij}) = X_i(t_{ij}) + \epsilon_{ij}. \tag{1}$$

Here ϵ_{ij} are zero mean independent identically distributed measurements errors with finite and constant variance $\mathrm{var}(\epsilon_{ij}) = \sigma^2$, independent of all other random components. The design points t_{ij} are considered deterministic and densely spaced. This model reflects typical measurements obtained in growth studies.

2. Data-driven differential equation

In the following we consider a differentiable stochastic process $X(t)$ such that X and its derivative X' are square integrable. A simple representation of the derivative process is to decompose it into a mean function $\mu_{X'}$ and a mean zero stochastic process Z_1,

$$X'(t) = \mu_{X'}(t) + Z_1(t). \tag{2}$$

Nonparametric estimation of individual derivative trajectories and of $\mu_{X'}$ provides data-driven descriptions (Gasser & Müller, 1984; Gasser et al., 1984; Mas & Pumo, 2009).

Considering a dynamic equation that captures the relationship between the process $X(t)$ and its derivative $X'(t)$, the simplest such relation is a linear relationship between X' and X. The corresponding linear empirical dynamics is a natural approach for Gaussian processes, since the joint Gaussianity of X and X' implies that there exists a deterministic function β with

$$X'(t) = \mu_{X'}(t) + \beta(t)\{X(t) - \mu_X(t)\} + Z_2(t). \tag{3}$$

Here Z_2 is a zero mean drift process with $\mathrm{cov}\{Z_2(t), X(t)\} = 0$, implying independence between X and Z_2 in the Gaussian case (Müller & Yao, 2010).

Many complex biological processes, including growth, cannot be expected to be adequately represented by linear dynamics. For more complex dynamics, it is therefore of interest to model the dynamics of X with a nonlinear differential equation. There always exists a function f with

$$E\{X'(t) \mid X(t)\} = f\{t, X(t)\}, \quad X'(t) = f\{t, X(t)\} + Z(t), \tag{4}$$

with $E\{Z(t) \mid X(t)\} = 0$ almost surely. When f is unknown and is determined from the data, (4) is a data-driven nonlinear differential equation. The function f and the properties of the drift process Z determine the underlying non-linear dynamics. In some applications, comparisons with the special case of a simpler autonomous system

$$E\{X'(t) \mid X(t)\} = f_1\{X(t)\}, \tag{5}$$

for a function f_1, which is time-independent, are of interest.

Parametric differential equations with random effects provide alternatives to modelling with Equation (4). Upon integration, these become nonlinear random effects models, which are difficult to fit, especially if they contain many random effects. A typical example is the nonlinear Preece–Baines model (Preece & Baines, 1978) for human growth, which can be derived from a non-autonomous differential equation. Such nonlinear models are nearly always fitted by least squares separately for each child, not taking advantage of the availability of a sample of growth curves and not including any random effects. These model fits are usually not efficient and have been shown to be inferior to nonparametric smoothing and differentiation methods in Gasser et al. (1984). These parametric growth models can be expressed in the form of the proposed general equation $X'(t) = f\{t, X(t)\}$, which thus provides a general and flexible framework that is informed by all data in the sample. As is typical for the life sciences, for growth data the nature of the underlying dynamics is largely unknown. The popular Preece–Baines model and related models have been derived purely based on data fitting considerations, while the model parameters are not interpretable (Hansen et al., 2003).

Models (2), (3) and (4) are characterized by increasing complexity, as

$$\mathrm{var}\{Z(t)\} \le \mathrm{var}\{Z_2(t)\} \le \mathrm{var}\{Z_1(t)\} = \mathrm{var}\{X'(t)\},$$

by definition of these drift processes. This means that the dynamic behavior of the process X is better predictable by the data-driven nonlinear differential equation (4) than by the empirical linear differential equation (3). If $\mathrm{var}\{Z(t)\} = \mathrm{var}\{Z_2(t)\}$, there is no gain in adopting a non-linear as compared to a simpler linear differential equation, but there can be substantial gains when the variance of $Z(t)$ is strictly smaller than the variance of $Z_2(t)$. Thus, the estimation of a data-driven nonlinear differential equation also can be used to assess the linearity of the underlying dynamics.

3. Estimating the components of data-driven non-linear dynamics

3·1. *Estimation of the deterministic component*

We adopt a two-step kernel smoothing approach to obtain an estimator $\hat{f}$ of the deterministic part of the nonlinear differential equation (4), corresponding to the function f, which from now on we assume to be a smooth function. This two-step procedure proceeds from the same ideas as the method of Ellner et al. (2002) for autonomous dynamics.

Step 1: Obtaining the trajectories of $X(t)$ and $X'(t)$. For any $i = 1, \dots, n$, we estimate the trajectory $X_i(t)$ and its derivative $X_i'(t)$ by a convolution kernel smoothing method (Gasser et al., 1984). Using a nonnegative symmetric kernel function K and an anti-symmetric kernel function with one sign change K_2 for derivative estimation, such that $\int K(u)du = 1$, $\int K_2(u)du = 0$ and $\int K_2(u)udu = 1$, we obtain the estimates

$$\widehat{X}_i(t) = \frac{1}{h_X} \sum_{j=1}^{N_i} \int_{s_{j-1}}^{s_j} Y_{ij} K\left(\frac{u-t}{h_X}\right) du, \tag{6}$$

$$\widehat{X'}_i(t) = \frac{1}{h_{X'}^2} \sum_{j=1}^{N_i} \int_{s_{j-1}}^{s_j} Y_{ij} K_2\left(\frac{u-t}{h_{X'}}\right) du, \tag{7}$$

where $s_j = (t_{ij} + t_{i,j+1})/2$ and $h_X > 0$ and $h_{X'} > 0$ are smoothing bandwidths.

Step 2: Estimation of f. Trajectory estimates $\widehat{X}(t)$ and $\widehat{X'}(t)$ from Step 1 are combined to obtain a Nadaraya–Watson kernel estimator for f,

$$\widehat{f}(t, x) = \frac{\sum_{i=1}^n K\{\frac{\widehat{X}_i(t)-x}{b_X}\}\widehat{X'}_i(t)}{\sum_{i=1}^n K\{\frac{\widehat{X}_i(t)-x}{b_X}\}}. \tag{8}$$

utilizing bandwidths $b_X > 0$.

When estimators (6), (7) are supplemented with suitably chosen boundary kernels for estimating the regression function near endpoints of the domain of X (Jones & Foster, 1996; Müller, 1991), these convolution kernel estimates are equivalent to fitting local linear estimates for $\widehat{X}_i(t)$, taking the intercept as estimator, and to fitting local quadratic estimates for $\widehat{X'}(t)$, taking the linear term as estimator (Fan & Gijbels, 1996; Müller, 1987). Thus, one can conveniently implement these estimators by local polynomial fitting.

3·2. *Decomposition of variance*

By definition (4) of the differential equation, we have the following decomposition of variance,

$$\mathrm{var}\{X'(t)\} = \mathrm{var}[f\{t, X(t)\}] + \mathrm{var}\{Z(t)\}. \tag{9}$$

Therefore, on subdomains where the variance of the drift process $\mathrm{var}\{Z(t)\}$ is small, the solution of (4) will not deviate much from the solution that is obtained with the deterministic approximation

$$X'(t) = f\{t, X(t)\} \quad (t \in \mathcal{T}), \tag{10}$$

that corresponds to the population equation. In this situation, the future changes of individual trajectories are easily predictable. This motivates to consider the fraction of the variance of $X'(t)$ that is explained by the deterministic part of the data-driven differential equation itself as a key quantity for assessing the predictability of the process, leading to a coefficient of determination

$$R^2(t) = \frac{\mathrm{var}[f\{t, X(t)\}]}{\mathrm{var}\{X'(t)\}} = 1 - \frac{\mathrm{var}\{Z(t)\}}{\mathrm{var}\{X'(t)\}}. \tag{11}$$

It is of interest to locate subdomains of $\mathcal{T}$ where $R^2(t)$ is large. On such subdomains, the drift process is small compared to $X'(t)$. An obvious estimate for the coefficient of determination $R^2(t)$ is obtained by plugging in estimates of the unknown quantities, yielding

$$\widehat{R}^2(t) = 1 - \frac{\sum_{i=1}^n \left[\widehat{X}_i'(t) - \widehat{f}\{t, \widehat{X}_i(t)\}\right]^2}{\sum_{i=1}^n \left\{\widehat{X}_i'(t) - \overline{\widehat{X}'}(t)\right\}^2}. \tag{12}$$

The coefficient of determination $R^2(t)$ assesses the fraction of $X'(t)$ explained by the deterministic differential at a given time t. However, for some processes the predictability of the process may depend on the time t and on the position x of the process. Considering the nonlinear regression model (4), we define the dynamic signal over noise ratio $S(t,x)$ by

$$S(t,x) = \frac{f^2(t,x)}{E\{X'^2(t) \mid X(t) = x\}} = \frac{f^2(t,x)}{f^2(t,x) + \mathrm{var}\{Z(t) \mid X(t) = x\}}. \tag{13}$$

Obviously, $S(t,x)$ lies between 0 and 1. When $S(t,x)$ is close to one, then $f^2(t,x)$ is large compared to $\mathrm{var}\{Z(t) \mid X(t) = x\}$ and the process is well predictable when $X(t) = x$. In contrast, small values of $S(t,x)$ indicate that the variability of $Z(t)$ given $X(t) = x$ is large. The functions S quantify the predictability of X as a function of the level of the process at time t.

By plugging in the estimate $\widehat{f}(t,x)$ for $f(t,x)$, one obtains the estimator

$$\widehat{S}(t,x) = \frac{\widehat{f}^2(t,x)}{\widehat{E}\{X'^2(t) \mid X(t)=x\}}, \quad \widehat{E}\{X'^2(t) \mid X(t)=x\} = \frac{\sum_{i=1}^n K\{\frac{\widehat{X}_i(t)-x}{b_X}\}\widehat{X'}_i^2(t)}{\sum_{i=1}^n K\{\frac{\widehat{X}_i(t)-x}{b_X}\}}. \quad (14)$$

3·3. *Applying data-driven nonlinear dynamics for goodness-of-fit*

It is of interest to determine whether linear dynamics, implied by Gaussianity of the underlying processes, suffices to describe the dynamics, or whether a more complex non-linear model is needed, reflecting increased complexity. A simple diagnostic of this can be obtained by comparing the variance of the drift process $Z(t)$ of the nonlinear dynamic model (4) with that of the drift process $Z_2(t)$ of the linear dynamic model (3), as follows.

For the coefficient of determination for the linear empirical dynamic model (3),

$$R_L^2(t) = \frac{\text{var}\,\{\beta(t)X(t)\}}{\text{var}\{X'(t)\}} = 1 - \frac{\text{var}\{Z_2(t)\}}{\text{var}\{X'(t)\}}, \quad (15)$$

one expects that $R^2(t) \geq R_L^2(t)$. Similar to equation (12),

$$\widehat{R}_L^2(t) = 1 - \frac{\sum_{i=1}^n \left\{\widehat{X}_i'(t) - \widehat{\beta}(t)\widehat{X}_i(t)\right\}^2}{\sum_{i=1}^n \left\{\widehat{X}_i'(t) - \overline{\widehat{X}'}(t)\right\}^2}, \quad (16)$$

where we note that both $\widehat{R}^2(t)$ in (12) and $\widehat{R}_L^2(t)$ in (16) might be negative when the fits are bad. On subdomains of $\mathcal{T}$ where $R(t)$ is close to $R_L(t)$, var$\{Z(t)\}$ is close to var$\{Z_2(t)\}$ and one may infer that the data-driven differential equation is almost linear, so Equation (3) provides a simpler description.

On subdomains where the diagnostic function $R(t) - R_L(t)$ is large, the linear differential equation (3) is probably insufficient to provide a good description of the underlying dynamics, and then one would then choose the data-driven non-linear dynamic model (4).

4. Asymptotic properties

4·1. *Assumptions*

In the following, we describe consistency results for the estimation of the smooth bivariate function f that determines the deterministic part of the proposed data-driven dynamic model (4) and for the estimate (12) of the fraction of variance explained at time t. In the sequel, $g(t,x)$ denotes the density of the random variable $X(t)$ at x. The assumptions C.1–C.7 are listed below.

C.1 The kernels K and K_2 have a compact support $[-1,1]$ and are Lipschitz continuous with respective constants μ_K and $\mu_{K'}$. Moreover, K is positive and satisfies $\int_{-1}^1 K(u)du = 1$, $\int_{-1}^1 K(u)udu = 0$ and $\int_{-1}^1 K(u)u^2du \neq 0$. The kernel K_2 satisfies $\int_{-1}^1 K_2(u)du = 0$, $\int_{-1}^1 K_2(u)udu = 1$, $\int_{-1}^1 K_2(u)u^2du = 0$ and $\int_{-1}^1 K_2(u)u^3du \neq 0$.
C.2 The random function X is almost surely three times continuously differentiable and for all $t \in \mathcal{T}$, $|X(t)| \leq C_0$, $|X'(t)| \leq C_1$, $|X^{(2)}(t)| \leq C_2$ and $|X^{(3)}(t)| \leq C_3$ almost surely.

C.3 The random variables ϵ_{ij} ($i = 1, \ldots, n;\ j = 1, \ldots, N$) are centered and have a finite moment of order 8.
C.4 The functions $f(t, \cdot)$ and $g(t, \cdot)$ are Lipschitz with constants μ_f and μ_g, twice continuously differentiable and have compact support.
C.5 The conditional variance $s(t, u) = \mathrm{var}\{X'(t) \mid X(t) = u\}$ is continuous and is nonzero.
C.6 We have $(N, n) \to \infty$ and $(b_X, h_X, h_{X'}) \to 0$ such that $nb_X \geq \log^2 n \to \infty$, $Nh_X b_X^4 \geq 1$, $Nh_{X'}^3 \to \infty$ and $h_X \leq b_X$.
C.7 There exists a constant $C > 0$ such that $g(t, x) > C$ for any $x \in [x_1; x_2]$.

Conditions on the kernels K and K_2 are given by C.1, while C.2–C.5 are essentially regularity assumptions on the process X and on the deterministic part f. Finally, C.6 provides conditions on the bandwidth of the kernels. Interestingly, the estimated functions $\widehat{X}(t)$ is less regularized than $\widehat{f}(t, x)$ ($h_X \leq b_X$).

4·2. *Results*

Theorem 1. *Under assumptions C.1–C.6, for any $t \in \mathcal{T}$ and x such that $g(t, x) \neq 0$,*

$$E\left[\left\{\widehat{f}(t, x) - f(t, x)\right\}^2\right] = O\left(b_X^4 + \frac{h_X^4}{b_X^2} + h_{X'}^4 + \frac{\sigma^2}{Nh_X b_X^2} + \frac{1}{nb_X} + \frac{\sigma^2}{Nh_{X'}^3}\right). \quad (17)$$

With suitable choices of the bandwidths b_X, h_X, and $h_{X'}$, one obtains

$$E\{\widehat{f}(t, x) - f(t, x)\}^2 = O\left\{\max\left(N^{-8/15}, n^{-4/5}\right)\right\} \ . \quad (18)$$

If $n \leq N^{2/3}$, the classical convergence rate $n^{-4/5}$ for nonparametric regression is obtained. Conversely, when $n \geq N^{2/3}$, the estimation error in $\widehat{X}_i$ is non-negligible and the lower bound N on the number of measurements per curve becomes the limiting quantity for the convergence rate.

Regarding $R^2(t)$, the rate of convergence of $\widehat{R}^2(t)$ depends on that of $f(t, \cdot)$ near the boundary of the support of $g(t, \cdot)$, where there are few observations. Therefore, we consider bounded domains for asymptotic study. For positive numbers x_1 and x_2 in the support of $g(t, \cdot)$, define

$$R^2_{x_1,x_2}(t) = \frac{\mathrm{var}\left[f\left\{t, X(t)\right\} \mid x_1 \leq X(t) \leq x_2\right]}{\mathrm{var}\{X'(t) \mid x_1 \leq X(t) \leq x_2\}} = 1 - \frac{\mathrm{var}\{Z(t) \mid x_1 \leq X(t) \leq x_2\}}{\mathrm{var}\{X'(t) \mid x_1 \leq X(t) \leq x_2\}}, \quad (19)$$

so that $R^2_{x_1,x_2}(t)$ quantifies the ratio of these variances when $X(t)$ is conditioned to lie between x_1 and x_2. With $\hat{n}_{x_1,x_2} = \#\{i : \ x_1 \leq \widehat{X}_i(t) \leq x_2\}$, we estimate $R^2_{x_1,x_2}(t)$ by

$$\widehat{R}^2_{x_1,x_2}(t) = 1 - \frac{\sum_{i=1}^n \left[\widehat{f}\{t, \widehat{X}_i(t)\} - \widehat{X}'_i(t)\right]^2 1_{x_1 \leq \widehat{X}_i(t) \leq x_2}}{\sum_{i=1}^n \widehat{X}'^2_i(t) 1_{x_1 \leq \widehat{X}_i(t) \leq x_2} - \{\sum_{i=1}^n \widehat{X}'_i(t) 1_{x_1 \leq \widehat{X}_i(t) \leq x_2} / \hat{n}_{x_1,x_2}\}^2}. \quad (20)$$

Theorem 2. *Under assumptions C.1–C.7,*

$$\widehat{R}^2_{x_1,x_2}(t) - R^2_{x_1,x_2}(t) = O_p\left\{b_X^2 + \frac{h_X^2}{b_X} + h_{X'}^2 + (nb_X)^{-1/2} + \frac{1}{(Nh_X)^{1/2} b_X} + \frac{1}{N^{1/2} h_{X'}^{3/2}}\right\}.$$

Corollary 1. *Under assumptions C.1–C.6, for the dynamic signal over noise ratio (13),*

$$\widehat{S}(t,x) - S(t,x) = O_p\left\{ b_X^2 + \frac{h_X^2}{b_X} + h_{X'}^2 + (nb_X)^{-1/2} + \frac{1}{(Nh_X)^{1/2} b_X} + \frac{1}{N^{1/2} h_{X'}^{3/2}} \right\}.$$

5. Nonlinear concurrent model

Our methodology provides an estimation procedure for a nonlinear version of the concurrent model, also known as varying-coefficient model (Chiang et al., 2001). We aim at investigating the relationship between two stochastic processes $X(t)$ and $U(t)$ at each time $t \in \mathcal{T}$. The linear concurrent model captures a linear relationship between X and U through a deterministic function $\beta(t)$,

$$U(t) = \mu_U(t) + \beta(t)\{X(t) - \mu_X(t)\} + Z_2(t), \tag{21}$$

where $Z_2(t)$ is a zero mean drift process with $\mathrm{cov}\{Z_2(t), X(t)\} = 0$. Versions of this functional linear varying coefficient linear model were mentioned in Ramsay & Silverman (2005) and estimators and its asymptotics were studied in Sentürk & Müller (2010).

Our methodology covers the more general situation where the link between $U(t)$ and $X(t)$ is nonlinear, i.e., where one has a smooth function $f(\cdot,\cdot)$ and a drift process $Z(t)$ such that

$$U(t) = f\{t, X(t)\} + Z(t), \tag{22}$$

with $E\{Z(t) \mid X(t)\} = 0$ almost surely and $f\{t, X(t)\} = E\{U(t) \mid X(t)\}$. This nonlinear varying coefficient model can be studied with the methods that we have developed for the nonlinear dynamic model (4).

Given n realizations X_i and U_i of the underlying processes X and U on a domain $\mathcal{T}$, we assume that N noisy measurements Y_{ij} and V_{ij} ($i = 1,\ldots,n$, $j = 1,\ldots,N$) have been obtained at times t_{ij} analogously to (1). Following the arguments of Section 3·1, we propose a two-step estimator. For any $i = 1,\ldots,n$, we first estimate the trajectory $X_i(t)$ and $U_i(t)$ with a convolution kernel K with bandwidths h_X and h_U. Then, using another bandwidth b_X, these trajectory estimates $\widehat{X}_i(t)$ and $\widehat{U}_i(t)$ are combined to obtain

$$\widehat{f}(t,x) = \frac{\sum_{i=1}^n K\{\frac{\widehat{X}_i(t)-x}{b_X}\}\widehat{U}_i(t)}{\sum_{i=1}^n K\{\frac{\widehat{X}_i(t)-x}{b_X}\}}.$$

Arguing as for the estimation of the non linear dynamic, we obtain the rate of convergence for $\widehat{f}$.

Corollary 2. *Suppose that assumptions D.1–D.6 in the Appendix hold. For any $t \in \mathcal{T}$ and any x such that $g(t,x) \neq 0$*

$$E\left[\left\{\widehat{f}(t,x) - f(t,x)\right\}^2\right] = O\left(b_X^4 + \frac{h_X^4}{b_X^2} + h_U^4 + \frac{\sigma^2}{N h_X b_X^2} + \frac{1}{nb_X} + \frac{\sigma^2}{N h_U^3}\right).$$

With suitable choices of the bandwidths b_X, h_X, and h_U, one obtains

$$E\{\widehat{f}(t,x) - f(t,x)\}^2 = O\left\{\max\left(N^{-8/15}, n^{-4/5}\right)\right\}. \tag{23}$$

As before, one can compute a coefficient of determination

$$R^2(t) = \frac{\mathrm{var}[f\{t, X(t)\}]}{\mathrm{var}\{U(t)\}} = 1 - \frac{\mathrm{var}\{Z(t)\}}{\mathrm{var}\{U(t)\}},$$

to decompose the variance of $U(t)$ into a part explained by the model and a part left unexplained.

6. Nonlinear dynamics of human growth data

The proposed model and estimation procedures can be used to illuminate the dynamics of human growth. We illustrate the nonlinear differential equation in (4) using the Berkeley Growth Study (Jones & Bayley, 1941), in which, the heights of 54 girls and 39 boys aged from 1 to 18 years were recorded. Since male and female growth patterns differ substantially, with girls entering puberty much earlier than boys (Tanner et al., 1966), we focus on girls only. For each of the 54 girls in the study, 31 measurements are available, which were recorded at different time intervals ranging from three months to one year. The purpose of characterizing the dynamics of human growth and especially the time domains where the dynamics is nonlinear is twofold . First, it allows us to gain a better understanding of the growth process. Second, it of clinical interest to distinguish between normal and pathological patterns of development.

In order to estimate the data-driven differential equation, we apply the two-step procedure described in Section 3.1, which is implemented through local weighted least-squares methods (Fan & Gijbels, 1996) with a Gaussian kernel K. For $t \in [0, 18]$, we obtain estimates $\widehat{X_i}(t) = \widehat{a}_{i0}(t)$, where

$$(\widehat{a}_{i0}, \widehat{a}_{i1})(t) = \underset{a' \in \mathbb{R}^2}{\arg\min} \frac{1}{h_X} \sum_{j=1}^{N} K\left(\frac{t_{ij} - t}{h_X}\right) \{X_{ij} - a_0 - a_1(t_{ij} - t)\}^2, \tag{24}$$

with $N = 31$. The growth velocities $X_i'(t)$ are estimated analogously by taking the slope of weighted local quadratic fits, $\widehat{X_i'}(t) = \hat{b}_{i1}(t)$, where

$$(\widehat{b}_{i0}, \widehat{b}_{i1}, \widehat{b}_{i2})(t) = \underset{b \in \mathbb{R}^3}{\arg\min} \frac{1}{h_{X'}} \sum_{j=1}^{N_i} K\left(\frac{t_{ij} - t}{h'_X}\right) \left\{X_{ij} - b_0 - b_1(t_{ij} - t) - b_2(t_{ij} - t)^2\right\}^2. \tag{25}$$

In a second step, $f(t, x)$ is obtained by another local linear estimator based on $\widehat{X}_i(t)$ and $\widehat{X}_i'(t)$, setting $\widehat{f}(t, x) = \widehat{d}_0(t, x)$, where

$$\{\widehat{d}_0(t, x), \widehat{d}_1(t, x)\} = \underset{d \in \mathbb{R}^2}{\arg\min} \sum_{i=1}^{n} \frac{1}{b_X} K\left\{\frac{\widehat{X}_i(t) - x}{b_X}\right\} \left[\widehat{X}_i'(t) - d_0 - d_1\{\widehat{X}_i(t) - x\}\right]^2. \tag{26}$$

A practically relevant feature is that for given t the function $\widehat{f}(t, \cdot)$ is only defined on the interval $(\min_i \widehat{X}_i(t), \max_i \widehat{X}_i(t))$. A second implementation issue is the choice of the smoothing bandwidths h_X, $h_{X'}$, and b_X that are needed for local polynomial estimators (24), (25) and (26). We select these tuning parameters by generalized cross-validation (Golub et al., 1979).

Estimated growth curves and estimated growth velocities for the sample of girls are depicted in Figure 1. The estimated function $\widehat{f}(t, x)$, corresponding to the deterministic

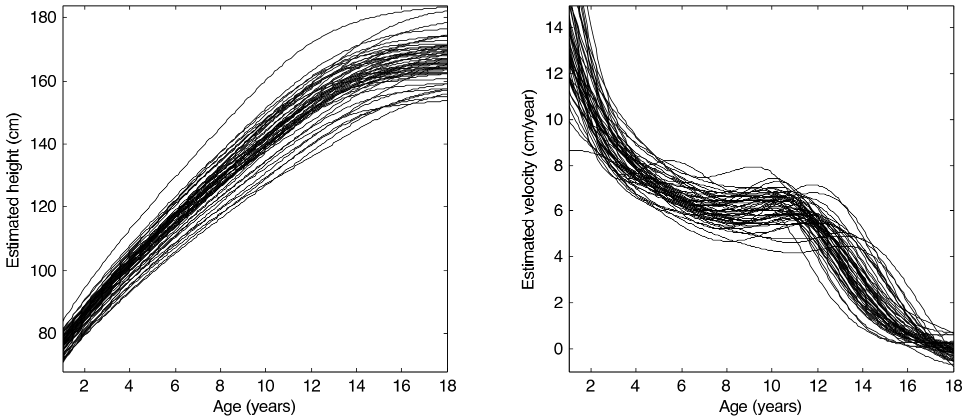

Figure 1. Estimated curves. Estimated growth curves and estimated growth velocities for 54 girls.

part of the data-driven nonlinear differential equation, is displayed as a contour plot in Figure 2. Growth velocity has a tendency to decrease with age, with the exception of the pubertal growth spurt at age between 10 and 13.

A more detailed study of the function f, considering $\widehat{f}(t, \cdot)$ as a function of current height x for ages $t = 2, 4, 6, 8, 12$ or 16, as shown in Figure 3, reveals that at earlier ages, there is a sizeable difference between the fits of the linear and the nonlinear differential equation and furthermore that an autonomous differential equation is inadequate. The clearly more appropriate proposed nonlinear non-autonomous model shows that there is only a weak relationship between growth velocity and height, while between ages 4 and 8, taller girls also tend to have a higher growth velocity, which can be interpreted as manifestation of an inherent growth momentum in this age range. In contrast, for ages between 12 and 16, $\widehat{f}(t, \cdot)$ is no longer monotone. At age 12, the relationship is weak, probably because the taller girls had their puberty growth peak prior to this age and their growth velocity then is decreasing during the post-pubertal growth deceleration, while the smaller girls did yet not enter the pubertal spurt with its growth acceleration. At age 16, all girls are growing much more slowly, though both shorter and taller girls grow relatively faster than medium sized girls, indicating a strongly nonlinear relationship.

The nonlinear dynamic coefficient of determination $R^2(t)$ defined in Equation (11) quantifies to which extent the deterministic part of the nonlinear differential equation (4) explains the variance of $X'(t)$. When estimating this coefficient with $\widehat{R}^2_{x_1(t),x_2(t)}(t)$ defined in Equation (19), we chose $x_1(t)$, respectively $x_2(t)$, as the third smallest, respectively largest, value among $\widehat{X}_i(t)$, $(i = 1, \ldots, n)$. We also estimated the linear dynamic coefficient of determination $R^2_L(t)$ defined in Equation (15) for the linear dynamic model

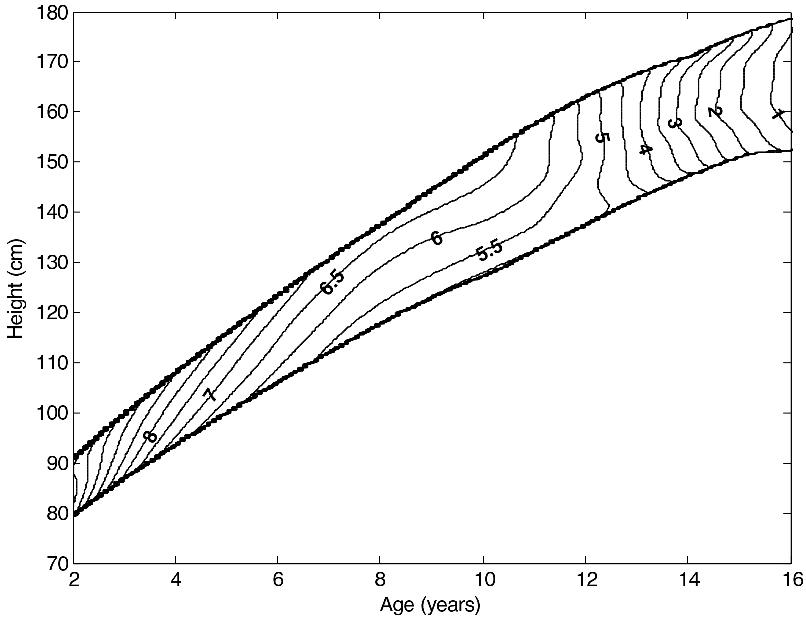

Figure 2. Contour plot of the nonparametric estimate of the surface $\hat{f}(t, x)$

(3). A comparison of the two coefficients of determination $\widehat{R}^2(t)$ and $\widehat{R}^2_L(t)$ is shown in Figure 4, and bootstrap confidence bands for the nonlinear version $R^2(t)$ are shown in the right panel.

For the proposed nonlinear dynamic model, $\widehat{R}^2(t)$ is seen to be close to 0.5 from approximately age 4 to 8. This implies that the deterministic part of the data-driven differential equation captures the behavior of the growth curves during these periods quite well. In contrast, $\widehat{R}^2(t)$ decays sharply from around age 11, as growth velocities are difficult to predict during this period, likely due to time variation in the occurrence of menarche and pubertal growth spurts. For the simpler linear dynamic model, the corresponding $\widehat{R}^2_L(t)$ is always smaller than the corresponding $\widehat{R}^2(t)$ for the proposed model, but comes closest during ages 8 to 10, where the discrepancy between the fits from the linear and the nonlinear systems is relatively small. In conclusion, growth dynamics around the pubertal growth spurt are highly nonlinear.

Acknowledgements

We wish to thank the Editor and several reviewers for helpful comments. The third author acknowledges support from the U.S. National Science Foundation.

Appendix 1

Assumptions for Corollary 2

In these assumptions, $g(t, \cdot)$ stands for the density of $X(t)$.

D.1 The kernel K has compact support $[-1, 1]$ and is Lipschitz continuous with constant μ_K. Moreover, K is positive and satisfies $\int_{-1}^{1} K(u)du = 1$, $\int_{-1}^{1} K(u)udu = 0$ and $\int_{-1}^{1} K(u)u^2du \neq 0$.

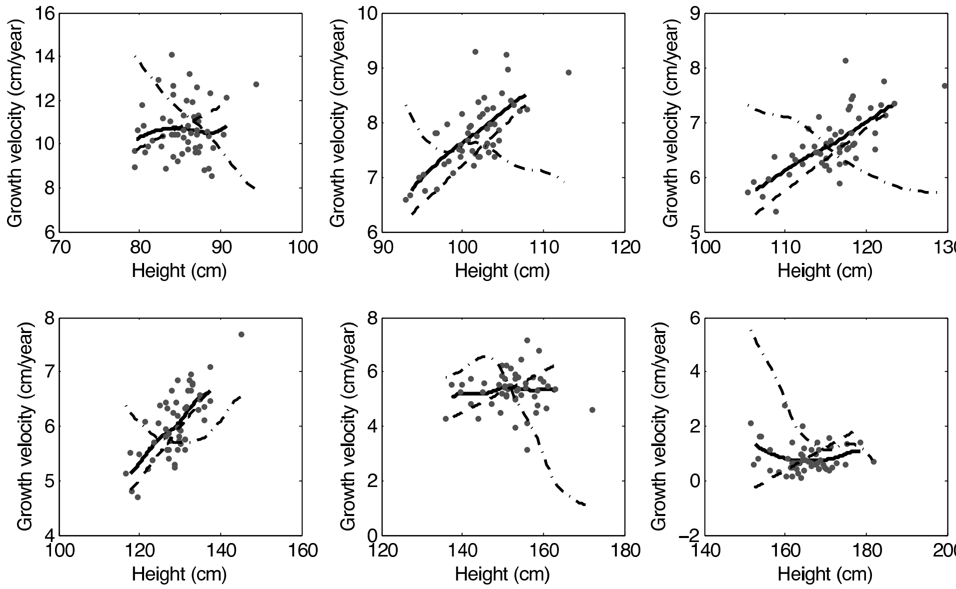

Figure 3. Comparison between nonlinear and linear dynamic estimation. Each of the panels, arranged for ages $t = 2, 4, 6, 8, 12$ years from left to right and top to bottom, respectively, illustrates the estimates $\hat{f}(t, \cdot)$ of the deterministic part of the proposed nonlinear dynamic model (4) (solid), estimates for the alternative linear dynamic model with time-varying coefficients (3) (dash-dash), and estimates for an autonomous differential equation (5) (dash-dot). Overlaid is the scatterplot of observed data pairs $\{x(t), x^{(1)}(t)\}$.

D.2 The random functions X and U are almost surely two times continuously differentiable. For $t \in \mathcal{T}$, $|X(t)| \leq C_0$, $|X'(t)| \leq C_1$, $|X^{(2)}(t)| \leq C_2$, $|U(t)| \leq C_3$, $|U'(t)| \leq C_4$, $|U^{(2)}(t)| \leq C_5$.
D.3 The random variables ϵ_{ij} $(i = 1, \ldots, n;\ j = 1, \ldots, N)$ and ζ_{ij} $(i = 1, \ldots, n;\ j = 1, \ldots, N)$ are centered and have a finite moment of order 8.
D.4 The functions $f(t, \cdot)$ and $g(t, \cdot)$ are Lipschitz with constants μ_f and μ_g, twice continuously differentiable and have a compact support.
D.5 The conditional variance $s(t, x) = \text{var}\{U(t) \mid X(t) = x\}$ is continuous and is nonzero.
D.6 We have $(N, n) \to \infty$ and $(b_X, h_X, h_U) \to 0$, such that $nb_X \geq \log^2 n \to \infty$, $Nh_X b_X^4 \geq 1$, $Nh_U \to \infty$ and $h_X \leq b_X$.

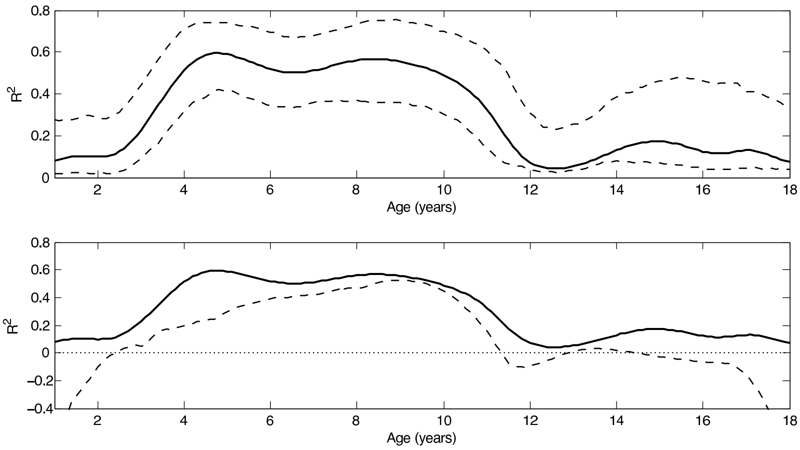

Figure 4. Coefficients of determination. Upper panel: 95% bootstrap confidence intervals for $R^2(t)$. Lower panel: Estimated coefficients of determination $\widehat{R}^2(t)$ (12), corresponding to the fraction of variance explained by the deterministic part of the nonlinear dynamic model (4) (solid), in comparison with the corresponding fractions of variance $\widehat{R}^2_L(t)$ (15) explained by linear dynamics (3) (dashed).

Appendix 2

Proofs

Proof of Theorem 1. We decompose the difference $\widehat{f}(t,x) - f(t,x)$ into the sum of two terms,

$$A = \frac{\sum_{i=1}^n K\{\frac{X_i(t)-x}{b_X}\} X_i'(t)}{\sum_{i=1}^n K\{\frac{X_i(t)-x}{b_X}\}} - f(t,x),$$

$$B = \frac{\sum_{i=1}^n K\{\frac{\widehat{X}_i(t)-x}{b_X}\} \widehat{X'}_i(t)}{\sum_{i=1}^n K\{\frac{\widehat{X}_i(t)-x}{b_X}\}} - \frac{\sum_{i=1}^n K\{\frac{X_i(t)-x}{b_X}\} X_i'(t)}{\sum_{i=1}^n K\{\frac{X_i(t)-x}{b_X}\}}.$$

The term A is simply the difference between a Nadaraya–Watson estimator and its target. Under Assumptions C.1–C.2,C.4–C.6, the pointwise risk of this estimator is known (Schimek, 2000, pages 43–70) to be equivalent to

$$\left\{ b_X^2 \int_{-1}^1 u^2 K(u)du \right\}^2 \left\{ \frac{1}{2} \frac{d^2 f(t,x)}{dx^2} + \frac{\frac{df(t,x)}{dx}\frac{dg(t,x)}{dx}}{g(t,x)} \right\}^2 + \frac{s(t,x)}{g(t,x)nb_X} \int_{-1}^1 K^2(u)du, \qquad \text{(A1)}$$

if the quantities involved in the last expression are nonzero. Hence, we have

$$E(A) = O\left\{ b_X^4 + (nb_X)^{-1} \right\}.$$

By Assumption C.2, we have $|X^{'}(t)| \leq C_1$, $|X^{(2)}(t)| \leq C_2$, and $|X^{(3)}(t)| \leq C_3$ almost surely. Applying classical results in kernel estimation (Gasser et al., 1984), one finds

$$E\left[\left\{\widehat{X}(t) - X(t)\right\}^2 \mid X\right] = O_p\left[\left\{h_X^2 C_2 \int_{-1}^{1} K(u)u^2 du\right\}^2 + \frac{\sigma^2}{Nh_X}\int_{-1}^{1} K^2(u)du\right], \quad \text{(A2)}$$

$$E\left[\left\{\widehat{X'}(t) - X'(t)\right\}^2 \mid X\right] = O_p\left[\left\{h_{X'}^2 C_3 \int_{-1}^{1} K_2(u)u^3 du\right\}^2 + \frac{\sigma^2}{Nh_{X'}^3}\int_{-1}^{1} K_2^2(u)du\right]. \quad \text{(A3)}$$

For the sake of simplicity, we respectively denote the rates in (A2) and (A3) by r_1^2 and r_2^2. Moreover, we have $E\left[\{\widehat{X}(t) - X(t)\}^4 \mid X\right] = O_p(r_1^4)$. To prove that

$$E\left(B^2\right) = O\left(\frac{r_1^2}{b_X^2} + \frac{r_1^4}{b_X^6} + \frac{r_1^8}{b_X^{14}} + r_2^2 + \frac{1}{n}\right),$$

we decompose B into the sum of two terms,

$$B_1 = \frac{\sum_{i=1}^n K\left\{\frac{\widehat{X}_i(t)-x}{b_X}\right\}\widehat{X'}_i(t)}{\sum_{i=1}^n K\left\{\frac{\widehat{X}_i(t)-x}{b_X}\right\}} - \frac{\sum_{i=1}^n K\left\{\frac{X_i(t)-x}{b_X}\right\}\widehat{X'}_i(t)}{\sum_{i=1}^n K\left\{\frac{X_i(t)-x}{b_X}\right\}},$$

$$B_2 = \frac{\sum_{i=1}^n K\left\{\frac{X_i(t)-x}{b_X}\right\}\left\{\widehat{X'}_i(t) - X_i'(t)\right\}}{\sum_{i=1}^n K\left\{\frac{X_i(t)-x}{b_X}\right\}}.$$

Let us first control the term B_1. We write

$$\alpha_i = \frac{K\left\{\frac{X_i(t)-x}{b_X}\right\}}{\sum_{j=1}^n K\left\{\frac{X_j(t)-x}{b_X}\right\}}, \qquad \widehat{\alpha}_i = \frac{K\left\{\frac{\widehat{X}_i(t)-x}{b_X}\right\}}{\sum_{j=1}^n K\left\{\frac{\widehat{X}_j(t)-x}{b_X}\right\}}. \quad \text{(A4)}$$

Applying Equation (A3), we get the following upper bound

$$\begin{aligned} B_1^2 &= \sum_{1\leq i_1,i_2\leq n} (\widehat{\alpha}_{i_1} - \alpha_{i_1})(\widehat{\alpha}_{i_2} - \alpha_{i_2})\widehat{X'}_{i_1}(t)\widehat{X'}_{i_2}(t) \\ &\leq O(1)\sum_{1\leq i_1,i_2\leq n} |(\widehat{\alpha}_{i_1} - \alpha_{i_1})(\widehat{\alpha}_{i_2} - \alpha_{i_2})| + \sum_{1\leq i_1,i_2\leq n} n^2(\widehat{\alpha}_{i_1} - \alpha_{i_1})^2(\widehat{\alpha}_{i_2} - \alpha_{i_2})^2 \\ &\quad + \frac{1}{n^2}\sum_{1\leq i_1,i_2\leq n}\left\{\widehat{X'}_{i_1}(t)\widehat{X'}_{i_2}(t) - X_{i_1}'(t)X_{i_2}'(t)\right\}^2, \end{aligned}$$

since the random variables $X_i'(t)$ are uniformly bounded above. As explained after (A3), we have

$$E\left[\frac{1}{n^2}\sum_{1\leq i_1,i_2\leq n}\left\{\widehat{X'}_{i_1}(t)\widehat{X'}_{i_2}(t) - X_{i_1}'(t)X_{i_2}'(t)\right\}^2\right] = O(r_2^2).$$

Define the event Ω by

$$\Omega = \left\{\sum_{j=1}^n K\left\{\frac{X_j(t)-x}{b_X}\right\} \geq nb_X g(t,x),\ \sum_{j=1}^n K\left\{\frac{\widehat{X}_j(t)-x}{b_X}\right\} \geq nb_X g(t,x)\right\}. \quad \text{(A5)}$$

We bound B_1^2 under the event Ω^c,

$$E\left(B_1^2 1_{\Omega^c}\right) \leq \left\{E\left(B_1^4\right)\mathrm{pr}(\Omega^c)\right\}^{1/2} \leq n^2\left[E\left\{\widehat{X}'^4(t)\right\}\mathrm{pr}(\Omega^c)\right]^{1/2}.$$

To obtain an upper bound for $\mathrm{pr}(\Omega^c)$, we bound the first two moments of $K[\{X_j(t)-x\}/b_X]$:

$$E\left[K\left\{\frac{X_j(t)-x}{b_X}\right\}\right] \geq 2b_X\left\{g(t,x)-\mu_g b_X\right\} = 2b_X g(t,x)\{1+o(1)\},$$

$$E\left[K^2\left\{\frac{X_j(t)-x}{b_X}\right\}\right] \leq 2b_X g(t,x)\{1+o(1)\}\|K\|_\infty^2,$$

$$E\left[K\left\{\frac{\widehat{X}_j(t)-x}{b_X}\right\}\right] \geq 2b_X g(t,x)\{1+o(1)\},$$

since b_X goes to 0, h_X^2/b_X goes to 0 and Nh_Xb_X goes to infinity. Since the kernel K is bounded, we can apply Bernstein's inequality

$$\mathrm{pr}\left[\sum_{j=1}^n K\left\{\frac{X_j(t)-x}{b_X}\right\} \leq nb_X g(t,x)\right] \leq \exp\left[-\frac{nb_X g(t,x)}{5\|K\|_\infty^2}\{1+o(1)\}\right].$$

Since $nb_X \geq \log^2 n$, it follows that

$$E\left(B_1^2 1_{\Omega^c}\right) = o(n^{-1}). \tag{A6}$$

Considering $E(B_1^2 1_\Omega)$, we aim to find bounds for terms of the form $E\left\{|(\widehat{\alpha}_{i_1}-\alpha_{i_1})(\widehat{\alpha}_{i_2}-\alpha_{i_2})|1_\Omega\right\}$. We note that $\widehat{\alpha}_i-\alpha_i$ decomposes as

$$\frac{K\left\{\frac{\widehat{X}_i(t)-x}{b_X}\right\}-K\left\{\frac{X_i(t)-x}{b_X}\right\}}{\sum_{j=1}^n K\left\{\frac{\widehat{X}_j(t)-x}{b_X}\right\}} + K\left\{\frac{X_i(t)-x}{b_X}\right\}\frac{\sum_{j=1}^n\left[K\left\{\frac{X_j(t)-x}{b_X}\right\}-K\left\{\frac{\widehat{X}_j(t)-x}{b_X}\right\}\right]}{\sum_{j=1}^n K\left\{\frac{\widehat{X}_j(t)-x}{b_X}\right\}\sum_{j=1}^n K\left\{\frac{X_j(t)-x}{b_X}\right\}}.$$

Applying Assumption C.1, under the event Ω,

$$|\widehat{\alpha}_i-\alpha_i|1_\Omega = O\left\{\frac{1}{nb_X^2 g(t,x)}\right\}\left[\left|X_i(t)-\widehat{X}_i(t)\right|1_{\{|X_i(t)-x|\leq 2b_X\}\cup\{|\widehat{X}_i(t)-X_i(t)|\geq b_X\}}\right.$$
$$\left.+\frac{1_{|X_i(t)-x|\leq b_X}}{nb_X g(t,x)}\sum_{j=1,\ j\neq i}^n\left|X_j(t)-\widehat{X}_j(t)\right|1_{\{|X_j(t)-x|\leq 2b_X\}\cup\{|\widehat{X}_j(t)-X_j(t)|\geq b_X\}}\right].$$

Applying (A2), the Cauchy–Schwarz inequality and Tchebychev's inequality, for $i_1 \neq i_2$,

$$E\left\{|(\widehat{\alpha}_{i_1}-\alpha_{i_1})(\widehat{\alpha}_{i_2}-\alpha_{i_2})|1_\Omega\right\} = \frac{1}{n^2}O\left\{\frac{r_1^2}{b_X^2}+\frac{r_1^4}{b_X^6 g^2(t,x)}+\frac{1}{n}\right\}. \tag{A7}$$

Similarly, bounding the second moment for $i_1 \neq i_2$,

$$E\left\{(\widehat{\alpha}_{i_1}-\alpha_{i_1})^2(\widehat{\alpha}_{i_2}-\alpha_{i_2})^2 1_\Omega\right\} = \frac{1}{n^2}O\left\{\frac{r_1^4}{b_X^6 g^2(t,x)}+\frac{r_1^8}{b_X^{14} g^6(t,x)}+\frac{1}{n}\right\}. \tag{A8}$$

The terms corresponding to $i_1 = i_2$ are negligible. Combining the upper bounds (A7) and (A8) with (A6), we conclude that

$$E\left(B_1^2\right) = O\left\{r_2^2+\frac{r_1^2}{b_X^2}+\frac{r_1^4}{b_X^6 g^2(t,x)}+\frac{r_1^8}{b_X^{14} g^6(t,x)}+\frac{1}{n}\right\}.$$

The term B_2 is simply a weighted sum of the differences $\widehat{X}_i'(t)-X_i'(t)$. Recall the weights α_i $(i=1,\ldots,n)$ defined in (A4). Conditioning on $X_i(t)$ $(i=1,\ldots,n;\ t\in\mathcal{T})$, we get

$$E\left(B_2^2\right) = E\left[\sum_{1\leq i,j\leq n}\alpha_i\alpha_j\left\{\widehat{X'}_i(t)-X_i'(t)\right\}\left\{\widehat{X'}_j(t)-X_j'(t)\right\}\right] = O(r_2^2).$$

All in all, we conclude that

$$E\left(B^2\right) = O\left\{r_2^2 + \frac{r_1^2}{b_X^2} + \frac{r_1^4}{b_X^6 g^2(t,x)} + \frac{r_1^8}{b_X^{14} g^6(t,x)} + \frac{1}{n}\right\} .$$

It then follows from Assumption C.6, (A2) and (A3) that

$$E\left(B^2\right) = O\left(h_{X'}^4 + \frac{h_X^4}{b_X^2} + \frac{\sigma^2}{N h_X b_X^2} + \frac{\sigma^2}{N h_{X'}^3} + \frac{1}{n}\right). \tag{A9}$$

Combining this last bound with (A1) allows us to prove the first part of the theorem. Setting $h_X = N^{-1/5}$, $h_{X'} = N^{-1/7}$ and $b_X = N^{-2/15}$ if $n \geq N^{2/3}$, while $b_X = n^{-1/5}$ if $n \leq N^{2/3}$, assumption C.6 is satisfied and one obtains

$$E\left\{\widehat{f}(t,x) - f(t,x)\right\}^2 = O\left\{\max\left(N^{-8/15}, n^{-4/5}\right)\right\}.$$

□

Proof of Theorem 2. We first consider the denominator of (20) divided by $\hat{n}_{x_1,x_2}$ and then the numerator of (20) divided by $\hat{n}_{x_1,x_2}$. We note that

$$\widehat{\mathrm{var}}_{x_1,x_2}\{X'(t)\} = \frac{\sum_{i=1}^n \widehat{X}_i'^2(t) 1_{x_1 \leq \widehat{X}_i(t) \leq x_2} - \{\sum_{i=1}^n \widehat{X}_i'(t) 1_{x_1 \leq \widehat{X}_i(t) \leq x_2} / \hat{n}_{x_1,x_2}\}^2}{\hat{n}_{x_1,x_2}}.$$

In the sequel, $\tilde{n}_{x_1,x_2}$ stands for $\#\{i: \ x_1 \leq X_i(t) \leq x_2\}$. The difference $\widehat{\mathrm{var}}_{x_1,x_2}\{X'(t)\} - \mathrm{var}_{x_1,x_2}\{X'(t)\}$ behaves like

$$O_p\left(n^{-1/2}\right) + \frac{\sum_{i=1}^n \widehat{X'}_i^2(t) 1_{x_1 \leq \widehat{X}_i(t) \leq x_2}}{\hat{n}_{x_1,x_2}} - \frac{\sum_{i=1}^n X_i'^2(t) 1_{x_1 \leq X_i(t) \leq x_2}}{\tilde{n}_{x_1,x_2}}$$
$$+ \left\{\frac{\sum_{i=1}^n X_i'(t) 1_{x_1 \leq X_i(t) \leq x_2}}{\tilde{n}_{x_1,x_2}}\right\}^2 - \left\{\frac{\sum_{i=1}^n \widehat{X}_i'(t) 1_{x_1 \leq \widehat{X}_i(t) \leq x_2}}{\hat{n}_{x_1,x_2}}\right\}^2 . \tag{A10}$$

Consider the following upper bound of $|\widehat{X'}_i^2(t) 1_{x_1 \leq \widehat{X}_i(t) \leq x_2} - X_i'^2(t) 1_{x_1 \leq X_i(t) \leq x_2}|$

$$|\widehat{X'}_i^2(t) - X_i'^2(t)| + X_i'^2(t) |1_{x_1 \leq \widehat{X}_i(t) \leq x_2} - 1_{x_1 \leq X_i(t) \leq x_2}|.$$

Since $\widehat{X'}$ is a kernel estimator of $X'(t)$, we have

$$E\left[\left\{\widehat{X'}^2(t) - X'^2(t)\right\}^2 \mid X\right] = O_p\left(h_{X'}^4 + \frac{1}{N h_{X'}^3}\right).$$

To bound the expectation of the term $|1_{x_1 \leq \widehat{X}_i(t) \leq x_2} - 1_{x_1 \leq X_i(t) \leq x_2}|$, we use the rate of convergence (A2) of $\widehat{X}_i(t)$. Since $X'(t)$ is uniformly bounded, we get

$$E\left\{X_i'^2(t) |1_{x_1 \leq \widehat{X}_i(t) \leq x_2} - 1_{x_1 \leq X_i(t) \leq x_2}|\right\} = O\left\{h_X^2 + (N h_X)^{-1/2}\right\},$$

$$E\left\{\left|\widehat{X'}_i^2(t) 1_{x_1 \leq \widehat{X}_i(t) \leq x_2} - X_i'^2(t) 1_{x_1 \leq X_i(t) \leq x_2}\right|\right\} = O\left\{h_{X'}^2 + h_X^2 + \frac{1}{N^{1/2} h_{X'}^{3/2}} + (N h_X)^{-1/2}\right\}.$$

From the rate of convergence (A2) of $\widehat{X}(t)$, we derive that

$$\frac{\hat{n}_{x_1,x_2} - \tilde{n}_{x_1,x_2}}{n} = O_p\left\{h_X^2 + (N h_X)^{-1/2}\right\}. \tag{A11}$$

It follows that $\sum_{i=1}^n \widehat{X'}_i^2(t)1_{x_1\le \widehat{X}_i(t)\le x_2}/\hat{n}_{x_1,x_2} - \sum_{i=1}^n X_i'^2(t)1_{x_1\le X_i(t)\le x_2}/\tilde{n}_{x_1,x_2}$ is

$$O_p\left\{h_{X'}^2 + h_X^2 + \frac{1}{N^{1/2}h_{X'}^{3/2}} + (Nh_X)^{-1/2}\right\}.$$

Arguing similarly for the last terms in (A10), we conclude that

$$\widehat{\mathrm{var}}_{x_1,x_2}\{X'(t)\} - \mathrm{var}_{x_1,x_2}\{X'(t)\} = O_p\left\{h_{X'}^2 + h_X^2 + \frac{1}{N^{1/2}h_{X'}^{3/2}} + (Nh_X)^{-1/2} + n^{-1/2}\right\}.$$

Let us now study the convergence of the numerator of (20). The difference

$$\frac{\sum_{i=1}^n \left[\widehat{f}\{t,\widehat{X}_i(t)\} - \widehat{X}_i'(t)\right]^2 1_{x_1\le \widehat{X}_i(t)\le x_2}}{\hat{n}_{x_1,x_2}} - \mathrm{var}\{Z(t)\mid x_1\le X(t)\le x_2\}$$

behaves like

$$\begin{aligned} O_p\left(n^{-1/2}\right) &+ \frac{\sum_{i=1}^n \widehat{f}^2\{t,\widehat{X}_i(t)\}1_{x_1\le \widehat{X}_i(t)\le x_2}}{\hat{n}_{x_1,x_2}} - \frac{\sum_{i=1}^n f^2\{t,X_i(t)\}1_{x_1\le X_i(t)\le x_2}}{\tilde{n}_{x_1,x_2}} \\ &+ \frac{\sum_{i=1}^n\{\widehat{X}_i'(t)\}^2 1_{x_1\le \widehat{X}_i(t)\le x_2}}{\hat{n}_{x_1,x_2}} - \frac{\sum_{i=1}^n\{X_i'(t)\}^2 1_{x_1\le X_i(t)\le x_2}}{\tilde{n}_{x_1,x_2}} \\ &+ 2\frac{\sum_{i=1}^n f\{t,X_i(t)\}X_i'(t)1_{x_1\le X_i(t)\le x_2}}{\tilde{n}_{x_1,x_2}} - 2\frac{\sum_{i=1}^n \widehat{f}\{t,\widehat{X}_i(t)\}\widehat{X}_i'(t)1_{x_1\le \widehat{X}_i(t)\le x_2}}{\hat{n}_{x_1,x_2}}. \end{aligned} \tag{A12}$$

We only bound the first difference in (A12), the two other differences being handled similarly. For any $1\le i\le n$, we consider the random variable $\widehat{f}^{(-i)}$ which is computed analogously to $\widehat{f}$ with the data $Y_{k,j}$ $(j=1,\ldots,N_k;\ k=1,\ldots,i-1,i+1,\ldots,n)$. Consequently, $\widehat{f}^{(-i)}(t,\cdot)$ is independent of $\widehat{X}_i(t)$. Denoting E_{-i} the expectation with respect to $Y_{k,j}$ $(j=1,\ldots,N_k;$ $k=1,\ldots,i-1,i+1,\ldots,n)$ and E_i the expectation with respect to $(Y_{i,j})$ $(j=1,\ldots,N_k)$, the difference $|\widehat{f}^2\{t,\widehat{X}_i(t)\}1_{x_1\le\widehat{X}_i(t)\le x_2} - f^2\{t,X_i(t)\}1_{x_1\le X_i(t)\le x_2}|$ decomposes into a sum of three terms

$$\begin{aligned} &|\widehat{f}^2\{t,\widehat{X}_i(t)\} - \{\widehat{f}^{(-i)}\}^2\{t,\widehat{X}_i(t)\}|1_{x_1\le\widehat{X}_i(t)\le x_2} \\ &+ |\{\widehat{f}^{(-i)}\}^2\{t,\widehat{X}_i(t)\} - f^2\{t,\widehat{X}_i(t)\}|1_{x_1\le\widehat{X}_i(t)\le x_2} \\ &+ |f^2\{t,\widehat{X}_i(t)\}1_{x_1\le\widehat{X}_i(t)\le x_2} - f^2\{t,X_i(t)\}1_{x_1\le X_i(t)\le x_2}|. \end{aligned} \tag{A13}$$

Let us bound the expected value of the second difference

$$\begin{aligned} &E\left[|\{\widehat{f}^{(-i)}\}^2\{t,\widehat{X}_i(t)\} - f^2\{t,\widehat{X}_i(t)\}|1_{x_1\le\widehat{X}_i(t)\le x_2}\right] \\ &\le E_i\left\{\left(\|f\|_\infty + E_{(-i)}\left[\{\widehat{f}^{(-i)}\}^2(t,\widehat{X}_i(t))\right]^{1/2}\right)\right. \\ &\left.\times E_{(-i)}\left(\left[\widehat{f}^{(-i)}\{t,\widehat{X}_i(t)\} - f\{t,\widehat{X}_i(t)\}\right]^2\right)^{1/2} 1_{x_1\le\widehat{X}_i(t)\le x_2}\right\}. \end{aligned}$$

Arguing as in the proof of Proposition 1, we know that the rate of convergence of $\widehat{f}^{(-i)}$ satisfies

$$E\left[\left\{\widehat{f}^{(-i)}(t,x) - f(t,x)\right\}^2\right] = \frac{O\left(b_X^4 + \frac{h_X^4}{b_X^2} + h_{X'}^4 + \frac{1}{nb_X} + \frac{1}{Nh_Xb_X^2} + \frac{1}{Nh_{X'}^3}\right)}{\min\{g^6(t,x),1\}}.$$

Thus, the expectation of the second difference in (A13) behaves like

$$O\left\{b_X^2 + (nb_X)^{-1/2} + \frac{h_X^2}{b_X} + \frac{\sigma}{(Nh_X)^{1/2}b_X} + h_{X'}^2 + \frac{\sigma}{N^{1/2}h_{X'}^{3/2}}\right\}. \tag{A14}$$

In order to control the difference $\{\widehat{f}^{(-i)}\}^2\{t, \widehat{X}_i(t)\} - \widehat{f}^2\{t, \widehat{X}_i(t)\}$ in (A13), we observe that $\widehat{f}\{t, \widehat{X}_i(t)\}$ decomposes as

$$\frac{K(0)\widehat{X'}_i(t)}{K(0) + \sum_{j\neq i} K\left\{\frac{\widehat{X}_j(t)-\widehat{X}_i(t)}{b_X}\right\}} + \widehat{f}^{(-i)}\{t, \widehat{X}_i(t)\}\left[1 - \frac{K(0)}{K(0) + \sum_{j\neq i} K\left\{\frac{\widehat{X}_j(t)-\widehat{X}_i(t)}{b_X}\right\}}\right].$$

We note $\beta = K(0)/(K(0) + \sum_{j\neq i} K[\{\widehat{X}_j(t) - \widehat{X}_i(t)\}/b_X])$. Thus, the difference $E\left[|\{\widehat{f}^{(-i)}\}^2\{t, \widehat{X}_i(t)\} - \widehat{f}^2\{t, \widehat{X}_i(t)\}|1_{x_1\leq \widehat{X}_i(t)\leq x_2}\right]$ is of the form

$$O(1)E\left\{\beta|\widehat{X'}_i^2(t)|1_{x_1\leq \widehat{X}_i(t)\leq x_2}\right\} + O(1)E\left(\beta[\widehat{f}^{(-i)}\{t, \widehat{X}_i(t)\}]^2 1_{x_1\leq \widehat{X}_i(t)\leq x_2}\right).$$

Applying Bernstein inequality as in the proof of Theorem 1, we bound β above by $O[g\{t, \widehat{X}_i(t)\}/(nb_X)]$ with large probability. We control the random variable on the complementary event applying the Cauchy–Schwarz inequality. All in all, we get

$$E_{-i}\left[|\{\widehat{f}^{(-i)}\}^2\{t, \widehat{X}_i(t)\} - \widehat{f}^2\{t, \widehat{X}_i(t)\} \mid 1_{x_1\leq \widehat{X}_i(t)\leq x_2}\right] \leq O_p\left[\frac{1_{x_1\leq \widehat{X}_i(t)\leq x_2} \max\left\{1, \widehat{X}_i'^2(t)\right\}}{nb_X g\{t, \widehat{X}_i(t)\}}\right].$$

Integrating with respect to X_i, we conclude that

$$E\left[|\{\widehat{f}^{(-i)}\}^2\{t, \widehat{X}_i(t)\} - \widehat{f}^2\{t, \widehat{X}_i(t)\}|1_{x_1\leq \widehat{X}_i(t)\leq x_2}\right] = O\left(\frac{1}{nb_X}\right). \tag{A15}$$

In order to control the third difference in (A13), we bound $|f^2\{t, \widehat{X}_i(t)\}1_{x_1\leq \widehat{X}_i(t)\leq x_2} - f^2\{t, X_i(t))\}1_{x_1\leq X_i(t)\leq x_2}|$ above by $2\mu_f\|f\|_\infty|\widehat{X}_i(t) - X_i(t)|$ if $x_1 \leq X_i(t) \leq x_2$ and $x_1 \leq \widehat{X}_i(t) \leq x_2$, by 0 if $X_i(t) \notin [x_1, x_2]$ and $\widehat{X}_i(t) \notin [x_1, x_2]$, and by $\|f\|_\infty^2$ else. From Equation (A2), we derive

$$E\left[|f^2\{t, \widehat{X}_i(t)\}1_{x_1\leq \widehat{X}_i(t)\leq x_2} - f^2\{t, X_i(t)\}1_{x_1\leq X_i(t)\leq x_2}|\right] = O\left\{h_X^2 + (Nh_X)^{-1/2}\right\}. \tag{A16}$$

Combining (A14), (A15), and (A16) with (A12) and (A13), we obtain

$$\begin{aligned} &E\left[\frac{1}{n}\sum_{i=1}^n \left|\widehat{f}^2\{t, \widehat{X}_i(t)\}1_{x_1\leq \widehat{X}_i(t)\leq x_2} - f^2\{t, X_i(t)\}1_{x_1\leq X_i(t)\leq x_2}\right|\right] \\ &\qquad = O\left\{b_X^2 + \frac{h_X^2}{b_X} + h_{X'}^2 + (nb_X)^{-1/2} + \frac{1}{(Nh_X)^{1/2}b_X} + \frac{1}{n^{1/2}h_{X'}^{3/2}}\right\}. \end{aligned}$$

Combining this bound with (A11), one finds

$$\begin{aligned} &\frac{1}{\widehat{n}_{x_1,x_2}}\sum_{i=1}^n \widehat{f}^2\{t, \widehat{X}_i(t)\}1_{x_1\leq \widehat{X}_i(t)\leq x_2} - \frac{1}{\tilde{n}_{x_1,x_2}}\sum_{i=1}^n f^2\{t, X_i(t)\}1_{x_1\leq X_i(t)\leq x_2} \\ &\qquad = O_p\left\{b_X^2 + \frac{h_X^2}{b_X} + h_{X'}^2 + (nb_X)^{-1/2} + \frac{1}{(Nh_X)^{1/2}b_X} + \frac{1}{n^{1/2}h_{X'}^{3/2}} + h_X\right\}. \end{aligned}$$

Arguing similarly, we obtain the rate of convergence of the two remaining terms in (A12). We conclude that $\widehat{\mathrm{var}}_{x_1,x_2}[\widehat{f}\{t,\widehat{X}(t)\}] - \mathrm{var}_{x_1,x_2}[f\{t,X(t)\}]$ behaves like

$$O_p\left\{b_X^2 + \frac{h_X^2}{b_X} + h_{X'}^2 + (nb_X)^{-1/2} + \frac{1}{(Nh_X)^{1/2}b_X} + \frac{1}{n^{1/2}h_{X'}^{3/2}}\right\}.$$

□

Proof of Corollary 1. We only need to observe that the rate of convergence of $\widehat{E}\{X'^2(t) \mid X(t) = x\}$ towards $E\{X'^2(t) \mid X(t) = x\}$ is the same as that of $\widehat{f}(t,x)$ towards $f(t,x)$. Indeed, $\widehat{E}\{X'^2(t) \mid X(t) = x\}$ is a Nadaraya–Watson estimator based on $\{\widehat{X}_i'(t), X_i'(t)\}$, $(i = 1,\ldots,n)$. Gathering this remark with Theorem 1 allows us to conclude the proof.

□

Proof of Corollary 2. The arguments are the same as in the proof of Theorem 1, the only difference being that the rate of convergence of $\widehat{X}'(t)$ is replaced by the rate of convergence of $\widehat{U}(t)$.

□

Bibliography

Bellman, R. & Roth, R. (1971). The use of splines with unknown end points in the identification of systems. *J. Math. Anal. Appl.* 34 26–33.

Brunel, N. (2008). Parameter estimation of ODE's via nonparametric estimators. *Electron. J. Stat.* 2 1242–1267. URL http://dx.doi.org/10.1214/07-EJS132.

Chen, J. & Wu, H. (2008). Efficient local estimation for time-varying coefficients in deterministic dynamic models with applications to HIV-1 dynamics. *Ann. Stat* 103 369–384.

Chiang, C., Rice, J. & Wu, C. (2001). Smoothing spline estimation for varying coefficient models with repeatedly measured dependent variables. *J. Amer. Statist. Assoc.* 96 605–619.

Ellner, S., Seifu, Y. & Smith, R. (2002). Fitting population dynamic models to time-series data by gradient matching. *Ecology* 83 2256–2270.

Fan, J. & Gijbels, I. (1996). *Local Polynomial Modelling and its Applications*. London: Chapman & Hall.

Gasser, T. & Müller, H.-G. (1984). Estimating regression functions and their derivatives by the kernel method. *Scand. J. Statist.* 11 171–185.

Gasser, T., Müller, H.-G., Köhler, W., Molinari, L. & Prader, A. (1984). Nonparametric regression analysis of growth curves. *Ann. Statist.* 12 210–229.

Golub, G., Heath, M. & Wahba, G. (1979). Generalized cross-validation as a method for choosing a good ridge parameter. *Technometrics* 21 215–223.

Hansen, B., Cortina-Borja, M. & Ratcliffe, S. (2003). Assessing non-linear estimation procedures for human growth models. *Ann Hum Biol* 30 80 –96.

Hoffmann, M. (1999). Adaptive estimation in diffusion processes. *Stochastic Process. Appl.* 79 135–163. URL http://dx.doi.org/10.1016/S0304-4149(98)00074-X.

Holte, S., Melvin, A., Mullins, J., Tobin, N. & Frenkel, L. (2006). Density-dependent decay in HIV-1 dynamics. *J. Acquired Immune Deficiency Syndromes* 41 266–276.

Hooker, G. (2009). Forcing Function Diagnostics for Nonlinear Dynamics. *Biometrics* 65 928–936.

Jacod, J. (2000). Non-parametric kernel estimation of the coefficient of a diffusion. *Scand. J. Statist.* 27 83–96. URL http://dx.doi.org/10.1111/1467-9469.00180.

Jones, H. & Bayley, N. (1941). The Berkeley Growth Study. *Child Development* 12 167–173.

Jones, M. C. & Foster, P. J. (1996). A simple nonnegative boundary correction method for kernel density stimation. *Statistica Sinica* 6 1005–1013.

Liang, H. & Wu, H. (2008). Parameter estimation for differential equation models using a framework of measurement error in regression models. *J. Amer. Statist. Assoc.* 103 1570–1583. URL http://dx.doi.org/10.1198/016214508000000797.

Mas, A. & Pumo, B. (2009). Functional linear regression with derivatives. *J. Nonparametr. Stat.* 21 19–40.

Miao, H., Dykes, C., Demeter, L. M. & Wu, H. (2009). Differential equation modeling of HIV viral fitness experiments: Model identification, model selection, and multimodel inference. *Biometrics* 65 292 – 300.

Müller, H.-G. (1987). Weighted local regression and kernel methods for nonparametric curve fitting. *J. Amer. Statist. Assoc.* 82 231–238.

Müller, H.-G. (1991). Smooth optimum kernel estimators near endpoints. *Biometrika* 78 521–530.

Müller, H.-G. & Yao, F. (2010). Empirical Dynamics for Longitudinal Data. *Ann. Statist.* 38 3458–3486.

Paul, D., Peng, J. & Burman, P. (2011). Semiparametric modeling of autonomous nonlinear dynamical systems with applications. *Ann. Appl. Stat.* 5 2078–2108.

Perelson, A., Essunger, P., Cao, Y., Vesanen, M., Hurley, A., Saksela, K., Markowitz, M. & Ho, D. (1997). Decay characteristics of HIV-l-infected compartments during combination therapy. *Nature* 387 188–191.

Preece, M. & Baines, M. (1978). A new family of mathematical models describing the human growth curve. *Ann. Hum. Biol.* 5 1–24.

Ramsay, J. O., Hooker, G., Campbell, D. & Cao, J. (2007). Parameter estimation for differential equations: a generalized smoothing approach. *J. R. Stat. Soc. Ser. B Stat. Methodol.* 69 741–796. With discussions and a reply by the authors, URL http://dx.doi.org/10.1111/j.1467-9868.2007.00610.x.

Ramsay, J. O. & Silverman, B. W. (2005). *Functional Data Analysis*. New York: Springer, 2nd ed.

Reddy, S. K. & Dass, M. (2006). Modeling on-line art auction dynamics using functional data analysis. *Stat. Sci.* 21 179–193.

Schimek, M., ed. (2000). *Smoothing and Regression*. New York: John Wiley & Sons Inc.

Sentürk, D. & Müller, H.-G. (2010). Functional varying coefficient models for longitudinal data. *J. Amer. Statist. Assoc.* 105 1256–1264.

Tanner, J., Whitehouse, R. & Takaishi, M. (1966). Standards from birth to maturity for height, weight, height velocity, and weight velocity: British children. *Arch. Dis. Child.* 41 613–635.

Wang, S., Jank, W., Shmueli, G. & Smith, P. (2008). Modeling price dynamics in eBay auctions using principal differential analysis. *J. Amer. Statist. Assoc.* 103 1100–1118.