Journal of Econometrics 239 (2024) 105573

Contents lists available at ScienceDirect

Journal of Econometrics

journal homepage: www.elsevier.com/locate/jeconom

Dynamic modeling for multivariate functional and longitudinal data

Siteng Hao [a], Shu-Chin Lin [a], Jane-Ling Wang [*,a], Qixian Zhong [b]

[a] Department of Statistics, University of California, Davis, CA 95616, USA
[b] MOE Key Lab of Econometrics, WISE, Department of Statistics & Data Science, School of Economics, Xiamen University, Xiamen, 361000, China

ARTICLE INFO

Keywords:
Concurrent regression
Derivatives
Functional data
Local polynomial regression
Varying-coefficient model

ABSTRACT

Dynamic interactions among several stochastic processes are common in many scientific fields. It is crucial to model these interactions to understand the dynamic relationship of the corresponding multivariate processes with their derivatives and to improve predictions. In reality, full observations of the multivariate processes are not feasible as measurements can only be taken at discrete locations or time points, and often only sparingly and intermittently in longitudinal studies. This results in multivariate longitudinal data that are measured at different times for different subjects. We propose a time-dynamic model to handle multivariate longitudinal data by modeling the derivatives of multivariate processes using the values of these processes. Starting with a linear concurrent model, we develop methods to estimate the regression coefficient functions, which can accommodate irregularly measured longitudinal data that are possibly contaminated with noise. Our approach can also be applied to settings when the observational times are the same for all subjects. We establish the convergence rates of our estimators with phase transitions and further illustrate our model through a simulation study and a real data application.

1. Introduction

Since the seminal work of Fan (1992, 1993), local polynomial regression has received ever increasing attention in Statistics, Econometrics and many other scientific disciplines for both theoretical research and applications. In particular, it provides a valuable tool to capture the shape or derivatives of a curve and surface without suffering from the boundary effects encountered by kernel regression, such as the Nadaraya-Watson estimator. The monograph Fan and Gijbels (1996) is an excellent source to get a comprehensive overview of the subject of local polynomial smoothing. A second area that Professor Fan made landmark contributions is the area of varying-coefficient models (Cai et al., 2000; Fan et al., 2003; Fan and Zhang, 2000; 1999), which extends classical linear models for vector responses and covariates to flexible linear models that can accommodate functional responses and functional covariates by allowing time-dependent covariate effects that are reflected in the corresponding coefficient functions of the covariates. In this paper, we extend the reach of both the local polynomial regression method and varying-coefficient approach to tackle a new problem that aims at dynamic modeling of multivariate functional and longitudinal data.

The study of the dynamic relationship between a random function X and its derivative X' has gained increasing attention recently. Such random functions are termed "functional data" and the analysis of functional data are termed "functional data analysis (FDA)".

* Corresponding author.
E-mail address: janelwang@ucdavis.edu (J.-L. Wang).

https://doi.org/10.1016/j.jeconom.2023.105573
Received 7 December 2021; Received in revised form 16 September 2023; Accepted 27 September 2023
Available online 4 November 2023

Functional data occur frequently in economics and finance studies, e.g. daily stock prices (measured every second), daily interest rates of different countries over a year, and GDP or unemployment rates of countries over long periods of time, to name just a few. The first derivative of a random function is closely related to the dynamic structure of the function and provides information on changing trends. Hence, it is of interest to study the dynamic relationship of a random function with its derivatives. For example, Bapna et al. (2008) use the first derivative of the bidding process in auctions to model the bid velocity, and this can be extended to make online bidding predictions.

Similar to a first-order system in engineering, there are several ways to study the relation of a function with its derivatives. Existing work by Ramsay et al. (2007) and Liang and Wu (2008) use a pre-selected differential equation model, which is a parametric method that relies heavily on specific knowledge of the data. A more flexible dynamic model is the linear or non-linear concurrent model (Müller and Yao, 2010; Verzelen et al., 2012), which is a nonparametric approach. Such an approach often leads to a simple and interpretable model. For instance, when the goal is to explore the relationship between a function X and its derivative at a time t, the linear concurrent model assumes a time-varying linear relation between $X(t)$ and its derivative $X^{'}(t)$. This model bears similarities to the varying-coefficient model that has been widely used to model functional response with functional covariates (Chiang et al., 2001; Fan and Zhang, 2000; Hoover et al., 1998; Huang et al., 2002; 2004; Liu and Müller, 2008; Wang et al., 2008). However, there is a fundamental difference here as the response, which is $X^{'}(t)$, is not observed at all.

We exploit the concurrent approach to jointly model the derivatives of multivariate functional data using the information from the multivariate functional data themselves. This allows all components of a multivariate functional process to contribute to the dynamics feature of any specific component, leading to an efficient modeling scheme. The dynamics of multivariate functional data have been explored in Müller and Yang (2010) and Pigoli and Sangalli (2012). However, our goal and approach are different. Müller and Yang (2010) proposed methods to estimate the regression functions between $Y^{(\nu_1)}$ and $X^{(\nu_2)}$, where ν_1 and ν_2 can be any integers that denote the order of derivative and $(X(t), Y(t))$ is a bivariate stochastic process for which we have data available. Pigoli and Sangalli (2012) focused on estimating the derivatives of multivariate functional data but not their relation to the multivariate functional data. To the best of our knowledge, a rigorous study that establishes a dynamic relationship between the derivatives of multiple processes and these multiple processes does not exist and is the goal of this paper.

Different linear models using derivatives as predictors or responses have been explored in Reddy and Dass (2006), Mas and Pumo (2009), Bhat and Madushani (2016), and Rawat (2018). Verzelen et al. (2012) and Bone (2016) further provide a tool that estimates a data-driven non-linear dynamic relationship. However, the estimation of their method relies on a precise estimate of $X^{'}$, which like most other existing works assume that the functional data are fully or densely observed (Gervini and Gasser, 2005; Ramsay and Silverman, 2005; Rice, 2004; Zhang and Wang, 2016). This is rarely the case for longitudinal studies involving human subjects, where data are often recorded on a subject-specific irregular time grid that covers just a few time points for each subject. The measurements may be further contaminated with noise or have random fluctuations (Müller and Yao, 2010; Yao et al., 2015; Zhu and Dunson, 2012). In this paper, we therefore propose a method that can handle both intensely and irregularly sampled functional data. Here "irregular" means that the sampling schedule varies from subject to subject as opposed to a regular schedule, in which all subjects are sampled at the same times. In addition to irregularity, most longitudinal studies can only afford to collect a few measurements per subject, resulting in sparsely observed data. Such irregular and sparse designs are ubiquitous in longitudinal studies, where parametric approaches, such as the popular linear mixed-effects model of Laird and Ware (1982), are routinely applied. To substitute these parametric models by nonparametric models has been the prime interest of the nonparametric community of FDA. Rice (2004) and the (Davidian et al., 2004) succinctly characterize the distinctions in statistical modeling approaches between the longitudinal and functional data communities.

Our approach benefits from previous work on estimating the derivatives of longitudinal data. For instance, Liu and Müller (2009) provided a method to estimate the derivatives of any order for sparsely observed data and Dai et al. (2018) used functional principal component analysis (Yao et al., 2005) to estimate derivatives. Müller and Yao (2010) studied a concurrent linear model between a univariate functional data X and its derivative $X^{'}$. Under Gaussian assumptions, X and $X^{'}$ satisfy a first order stochastic differential equation with a drift process and they estimated the components of this differential equation and studied the rates of convergence of these estimates. We broaden their scope to the multivariate setting in this paper and remove the Gaussian assumption.

More specifically, let $\boldsymbol{X}(t) = (X^1(t), X^2(t), ..., X^p(t))^\top$ denote a multivariate stochastic processes, where all X^l share the same domain $\mathscr{T}$ and $\boldsymbol{\mu}(t) = \mathrm{E}(\boldsymbol{X}(t))$ denote the mean function of $\boldsymbol{X}(t)$. Our goal is to simultaneously model $(X^1(t), \ldots, X^p(t))^\top$ and the derivatives $\boldsymbol{X}^{'}(t) = ((X^1)^{'}(t), (X^2)^{'}(t), ..., (X^p)^{'}(t))^\top$ with the following concurrent dynamic model:

$$\boldsymbol{X}^{'}(t) - \boldsymbol{\mu}^{'}(t) = B(t)\{\boldsymbol{X}(t) - \boldsymbol{\mu}(t)\} + \boldsymbol{Z}(t), \tag{1}$$

where $\boldsymbol{\mu}^{'}(t) = ((\mu^1)^{'}(t), \ldots, (\mu^p)^{'}(t))^\top$ is the derivative of $\boldsymbol{\mu}(t) = (\mu^1(t), \ldots, \mu^p(t))^\top$, $\boldsymbol{Z}(t)$ is a p-dimensional vector independent of $\boldsymbol{X}(t)$, and $B(t) = (B^{dl}(t))_{1\le d,l\le p}$ is a $p \times p$ matrix with (d,l)th element $B^{dl}(t)$, a function of t.

Different from Müller and Yao (2010), where a single process is considered, we are interested in the hidden relationships among the different processes, quantified by the off-diagonal elements of $B(t)$. For instance, at time t, a negative value of $B^{12}(t)$ implies that a trajectory X^2 with positive value at t is associated with a downward trend of the trajectory X^1 shortly after time t, while an upward trend is predicted for X^1 if $X^2(t)$ is negative. Müller and Yao (2010) used $(1/2)d\log(\widehat{V}(t))/dt$, where $\widehat{V}(t)$ is an estimate of $\mathrm{var}\{X^1(t)\}$, to estimate the univariate coefficient function $B(t)$ in (1). However such an approach does not work for multivariate processes. Instead, we propose a plug-in type estimator (see (6)) based on the normal equation in (2). This leads to substantial difference in theoretical tools and results. In particular, we establish the uniform convergence rate of the estimates while Müller and Yao (2010) focused on

L^2-convergence rate. Moreover, Müller and Yao (2010) developed the convergence rate of the coefficient estimate for the case of sparse functional data, where each process $X^j(t)$ can only be observed at a finite number of time points. In contrast, we provide the convergence rate for sparse functional data as well as the case when the number of observation grows with sample size, leading to a phase transition in the convergence rates.

We now compare model (1) with the conventional varying-coefficient model for functional data. If $p = 1$ and $X^1(t)$ and its derivative $(X^1)'(t)$ can be observed at some time points t, then model (1) collapses to the conventional varying-coefficient model (Cai and Li, 2008; Cai et al., 2009; Şentürk and Nguyen, 2011; Fan and Zhang, 2000; Hoover et al., 1998; Huang et al., 2002; 2004; Li et al., 2002; Morris and Carroll, 2006), which has also been called a functional linear model (Shen and Faraway, 2004). If $p > 1$ and all the processes $X^j(t), j = 1, \ldots, p$, and their derivatives $(X^j)'(t), j = 1, \ldots, p$, can be observed at some time points, we can extend the conventional varying-coefficient model to a multivariate setting, although this seems to have not been explored yet in the literature. In reality, the derivative processes $(X^j)'(t)$ cannot be observed at any time point t. This is a key distinction between the estimation approach of a varying-coefficient model and the proposed model (1).

In summary, this paper proposes an interpretable dynamic model for multivariate functional data by leveraging the univariate coefficient function B^{dl} and the additive structure for varying-coefficient models. Specifically, the sign and value of the univariate coefficient functions reveals the relationship among the different functional covariates. We establish convergence rates of the estimated coefficient functions under different design settings that include both densely and sparsely observed functional data. A byproduct of the proposed dynamic model is that it is well suited to make short-term predictions of a process by leveraging not only the current value of other covariates but also the current value of the process itself. This offers a major advantage for prediction as traditional varying-coefficient models only use the current values of other covariates as predictors. We demonstrate this advantage in an applied setting for cancer and GDP data in Section 4.

The rest of paper is organized as follows. We propose an estimate of the coefficient matrix B in Section 2. Assumptions and theoretical results are included in Section 3. The numerical performance of the proposed method is illustrated in Section 4 through Monte Carlo simulations and a real data application that links Cancer Death Rates and GDP (per capita). Section 5 contains a closing discussion. Detailed proofs and additional numerical results are provided in the Supplementary Material.

2. Methodology

Let $X^l, l = 1, \ldots, p$, be differentiable stochastic processes on a compact interval $\mathcal{T} \subset \mathbb{R}$, which is assumed to be [0,1], without loss of generality.

Proposition 2.1. *Assume* $\mathrm{E}\{X^l(t)\}^2 < \infty, l = 1, \cdots, p$, *for all* $t \in [0,1]$. *Denote* $\Sigma_{00}(t) = \mathrm{E}\{\boldsymbol{X}(t) - \boldsymbol{\mu}(t)\}\{\boldsymbol{X}(t) - \boldsymbol{\mu}(t)\}^\top$ *and* $\Sigma_{01}(t) = \mathrm{E}\{\boldsymbol{X}(t) - \boldsymbol{\mu}(t)\}\{\boldsymbol{X}'(t) - \boldsymbol{\mu}'(t)\}^\top$. *If the inverse* $\Sigma_{00}^{-1}(t)$ *of the matrix* $\Sigma_{00}(t)$ *exist for all* $t \in [0,1]$, *then model (1) implies that*

$$B(t) = \Sigma_{01}(t)\Sigma_{00}^{-1}(t), \tag{2}$$

and $\boldsymbol{Z}(t) = \boldsymbol{X}'(t) - \boldsymbol{\mu}'(t) - \Sigma_{01}(t)\Sigma_{00}^{-1}(t)\{\boldsymbol{X}(t) - \boldsymbol{\mu}(t)\}$

Proposition 2.1 suggests that we can estimate $B(t) = \Sigma_{01}(t)\Sigma_{00}^{-1}(t)$ from the observed data, even if they are contaminated with errors. To see this, assume that for individuals $i = 1, \ldots, n$, we observe the repeated measurements $(T_{ij}, \boldsymbol{Y}_{ij}), j = 1, \ldots, N_i$, with $\boldsymbol{Y}_{ij} = (Y_{ij}^1, \ldots, Y_{ij}^p)^\top$ and

$$Y_{ij}^l = X^l(T_{ij}) + \epsilon_{ij}^l, l = 1, \ldots, p,$$

where the measurement errors $\boldsymbol{\epsilon}_{ij} = (\epsilon_{ij}^1, \ldots, \epsilon_{ij}^p)^\top, j = 1, \ldots, N_i$ are i.i.d. with $\mathrm{E}(\epsilon_{ij}^l) = 0$ and $\mathrm{E}(\epsilon_{ij}^l)^2 < \infty, l = 1, \ldots, p$. For simplicity, here we assume that each coordinate $X^l, l = 1, \ldots, p$ is observed at the same time points $T_{ij}, j = 1, \ldots, N_i$. Our method can be adapted to cases where each coordinate is observed at irregular and different time grids with noise corruption, and the theoretical results in this paper also hold for such a general setting.

Next we introduce an approach to estimate $\Sigma_{01}(t)$ and $\Sigma_{00}(t)$, which then leads to a plug-in estimator of $\Sigma_{01}(t)\Sigma_{00}^{-1}(t)$. Because the mean function can be estimated at a faster rate than the covariance, for technical brevity and following the practice in the FDA literature, we assume that the mean function $\boldsymbol{\mu}(t) = (\mu^1(t), \ldots, \mu^p(t))^\top$ is known.

For $1 \le d, l \le p$, denote the (d, l) component of the covariance function $\Sigma_{00}(t)$ by

$$G^{dl}(s,t) = \mathrm{E}\{X^d(s) - \mu^d(s)\}\{X^l(t) - \mu^l(t)\}. \tag{3}$$

Note that

$$\begin{aligned} \frac{\partial}{\partial t} G^{dl}(s,t) &= \frac{\partial}{\partial t} \mathrm{E}\{X^d(s) - \mu^d(s)\}\{X^l(t) - \mu^l(t)\} \\ &= \mathrm{E}\{X^d(s) - \mu^d(s)\}\{(X^l)'(t) - (\mu^l)'(t)\}. \end{aligned}$$

This shows that $\Sigma_{01}(t) = \left(\frac{\partial}{\partial t} G^{dl}(s,t)|_{s=t}\right)_{1\le d,l\le p}$. Along with $\Sigma_{00}(t) = (G^{dl}(s,t)|_{s=t})_{1\le d,l\le p}$, we only need to estimate $G^{dl}(s,t), d,l = 1,\ldots,p$, and their partial derivatives $\frac{\partial}{\partial t} G^{dl}(s,t)$. These can be achieved by the local polynomial regression method (Fan and Gijbels, 1996) as follows.

Let $\Gamma_{ijk}^{dl} = \{Y_{ij}^{d} - \mu^{d}(T_{ij})\}\{Y_{ik}^{l} - \mu^{l}(T_{ik})\}$ be the "raw covariance" and $K(\cdot)$ be a univariate density function on $[-1,1]$. We estimate $G^{dl}(s,t)$ by $\widehat{\alpha}_{00}(= \widehat{\alpha}_{00}^{dl}(s,t))$, which is the arg min with respect to the argument α_{00} for

$$M_{0n}\left(\alpha_{00}^{dl}, \alpha_{10}^{dl}, \alpha_{01}^{dl}\right) := \sum_{i=1}^{n} w_i \sum_{1\le j\ne k\le N_i} \left[K_{h_{G,0}}(T_{ij} - s) K_{h_{G,0}}(T_{ik} - t) \left\{ \Gamma_{ijk}^{dl} - \alpha_{00}^{dl} - \alpha_{10}^{dl}(T_{ij} - s) - \alpha_{01}^{dl}(T_{ik} - t) \right\}^2 \right], \tag{4}$$

where $K_h(s) = (1/h)K(s/h)$ with bandwidth h, and the weight w_i for the ith subject satisfies $\sum_i^n w_i N_i(N_i - 1) = 1$. The assignment of the weights w_i, including the per-observation scheme $w_i = 1/\sum_i^n N_i(N_i - 1)$ and the per-subject scheme $w_i = 1/\{nN_i(N_i - 1)\}$, is discussed in Zhang and Wang (2016). Moreover, the partial derivatives $\partial G^{dl}(s,t)/\partial t$ can be estimated by $\widehat{\alpha}_{01}(= \widehat{\alpha}_{01}^{dl}(s,t))$ which is the arg min with respect to the argument α_{01} for

$$\begin{aligned} M_{1n}\left(\alpha_{00}^{dl}, \alpha_{10}^{dl}, \alpha_{01}^{dl}, \alpha_{11}^{dl}, \alpha_{20}^{dl}, \alpha_{02}^{dl}\right) \;=\; & \sum_{i=1}^{n} w_i \sum_{1\le j\ne k\le N_i} \Big[K_{h_{G,1}}(T_{ij} - s) K_{h_{G,1}}(T_{ik} - t) \times \Big\{ \Gamma_{ijk}^{dl} - \alpha_{00}^{dl} \\ & - \alpha_{10}^{dl}(T_{ij} - s) - \alpha_{01}^{dl}(T_{ik} - t) - \alpha_{11}^{dl}(T_{ij} - s)(T_{ik} - t) - \alpha_{20}^{dl}(T_{ij} - s)^2 - \alpha_{02}^{dl}(T_{ik} - t)^2 \Big\}^2 \Big] \end{aligned} \tag{5}$$

Note that we estimate $G^{dl}(s,t)$ through Eq. (4) rather than Eq. (5) since (4) involves fewer parameters and thus produces an estimator with lower variance.

Based on (4) and (5), we use

$$\widehat{\Sigma}_{00}(t) = \left(\widehat{\alpha}_{00}^{dl}(s,t)|_{s=t}\right)_{1\le d,l\le p} \text{ and } \widehat{\Sigma}_{01}(t) = \left(\widehat{\alpha}_{01}^{dl}(s,t)|_{s=t}\right)_{1\le d,l\le p}$$

as the estimators of $\Sigma_{00}(t)$ and $\Sigma_{01}(t)$, respectively. Therefore, the estimator of $B(t)$ is given by

$$\widehat{B}(t) = \{\widehat{\Sigma}_{01}(t)\}\{\widehat{\Sigma}_{00}(t)\}^{-1}. \tag{6}$$

3. Theoretical properties

In this section we provide theoretical analysis of the methodology in Section 2. The following assumptions are needed to establish the results.

A. Sampling design and covariance function.

(A1) The observed times $\{T_{ij} : i = 1,\ldots,n, j = 1,\ldots,N_i\}$ are i.i.d. copies of a random variable T on [0,1] whose density function $f(\cdot)$ satisfies, for some constant $0 < c_1 < \infty$,

$$c_1 \le \min_{t\in[0,1]} f(t) \le \max_{t\in[0,1]} f(t) \le 1 \Big/ c_1.$$

Moreover, the second derivative of f is bounded on [0,1].

(A2) The time variable T, stochastic vector process $\boldsymbol{X}$ and the error $\boldsymbol{\epsilon}$ are jointly independent.

(A3) For all non-negative integers ν_1 and ν_2 with $\nu_1 + \nu_2 = 3$, the partial derivatives

$$\frac{\partial^{\nu_1+\nu_2}}{\partial s^{\nu_1}\partial t^{\nu_2}} G^{dl}(s,t), 1 \le d,l \le p,$$

exist and are continuous on $[0,1]\times[0,1]$.

(A4) The matrix Σ_{00} satisfies

$$0 < \inf_{t\in[0,1]} \lambda_{\min}(\Sigma_{00}(t)) \le \sup_{t\in[0,1]} \lambda_{\max}(\Sigma_{00}(t)) < \infty,$$

where $\lambda_{\min}(A)$ and $\lambda_{\max}(A)$ are the smallest and largest eigenvalues of a matrix A, respectively.

The boundedness assumption in (A1) and independence of T, $\boldsymbol{X}$ and $\boldsymbol{\epsilon}$ in (A2) are common assumptions in functional data analysis (Li and Hsing, 2010; Yao et al., 2005; Zhang and Wang, 2016). Assumption (A3) requires that the partial derivatives $\partial^{\nu_1+\nu_2} G^{dl}(s,t)/\partial s^{\nu_1}\partial t^{\nu_2}, (\nu_1 + \nu_2 = 1)$ be twice differentiable, which is a standard assumption. Assumption (A4) is another standard assumption in varying-coefficient models to guarantee the uniqueness of $B(t)$ (Gu and Volgushev, 2019; Huang et al., 2002; Wang et al., 2008).

B. Kernel function and bandwidth. Let $\beta_{n,\nu} = [\log(n)\sum_{i=1}^n w_i^2\{N_i(N_i-1)/h_{G,\nu}^2 + N_i(N_i-1)(N_i-2)/h_{G,\nu} + N_i(N_i-1)(N_i-2)(N_i-4)\}]^{1/2}$.

(B1) The kernel function $K(\cdot)$ is a symmetric probability density on $[-1,1]$ and

$$\max\left\{\int_0^1 [K(s)]^2 ds, \int_0^1 s^4 K(s) ds\right\} < \infty.$$

(B2) There exists $0 < L_1 < \infty$, such that $|K(s) - K(t)| \le L_1 |s - t|$, for any $s, t \in [0,1]$.

(B3) For some $\delta > 0$, $\max\{\mathrm{E}|\epsilon^l|^{2+\delta}, \mathrm{E}[\sup_{t\in[0,1]}|X^l(t) - \mu^l(t)|^{2+\delta}]\} < \infty, l = 1, \ldots, p$, and

$$\beta_{n,\nu} h_{G,\nu}^2 \left\{\frac{n}{\log(n)}\right\}^{\delta/(2+\delta)} \to \infty, \text{ as } n \to \infty.$$

(B4) $\max\{h_{G,\nu}, \beta_{n,\nu}/h_{G,\nu}^{\nu}\} \to 0$, as $n \to \infty$ and $\sup_n\{n \max_i w_i N_i (N_i - 1)\} < \infty$.

Assumptions (B1) and (B2) are classical assumptions for smoothing methods. Here the Lipschitz continuity condition simplifies the proof and is satisfied by a rich class of kernel functions, such as the Epanechnikov kernel function $K(s) = 3/4(1 - x^2)1(|s| \le 1)$. Assumptions (B3) and (B4) are standard and entail some restrictions on the bandwidth and moments of X and ϵ to ensure the consistency of the estimators. They are fairly standard assumptions.

Our first theorem establishes the convergence rate of $\widehat{\alpha}_{0\nu}^{dl}(s,t), \nu = 0, 1$ in (4) and (5), respectively. Denote $\partial^0 G^{dl}(s,t) / \partial t^0 = G^{dl}(s,t)$.

Theorem 3.1. *Under (A1) – (A3) and (B1) – (B4), for* $\nu = 0, 1$,

$$\max_{1\le d,l\le p} \sup_{s,t\in[h_{G,\nu}, 1-h_{G,\nu}]} \left| \widehat{\alpha}_{0\nu}^{dl}(s,t) - \frac{\partial^\nu}{\partial t^\nu} G^{dl}(s,t) \right| = O_p\left(\frac{\beta_{n,\nu}}{h_{G,\nu}^{\nu}} + h_{G,\nu}^2\right) \tag{7}$$

Theorem 3.1 provides unified convergence rates for both the estimated covariance ($\nu = 0$) and partial derivatives ($\nu = 1$) with different kinds of weights w_i and generalized time design N_i. To better understand the meaning of (7), we consider a special case when the number N_i of repeated measurements are the same for all subjects, i.e., $N_i = N$. Under this setting the weights, for both per-observation and per-subject scheme, are equal (i.e., $w_i = 1/\{nN(N-1)\}$) and we have a more transparent result.

Corollary 3.1. *Under (A1) – (A3) and (B1) – (B4), if* $N_i = N$ *and* $w_i = 1/\{nN(N-1)\}$, *then, for* $\nu = 0, 1$,

$$\max_{1\le d,l\le p} \sup_{s,t\in[h_{G,\nu}, 1-h_{G,\nu}]} \left| \widehat{\alpha}_{0\nu}^{dl}(s,t) - \frac{\partial^\nu}{\partial t^\nu} G^{dl}(s,t) \right| = O\left(\frac{1}{h_{G,\nu}^{\nu}} \sqrt{\left(\frac{1}{N^2 h_{G,\nu}^2} + 1\right)\frac{\log(n)}{n}} + h_{G,\nu}^2\right)$$

Furthermore,

(i) if $N < \infty$ *and* $h_{G,\nu} = O(\{n/\log(n)\}^{-1/(2\nu+6)})$ *hold, then, for* $\nu = 0, 1$,

$$\max_{1\le d,l\le p} \sup_{s,t\in[h_{G,\nu}, 1-h_{G,\nu}]} \left| \widehat{\alpha}_{0\nu}^{dl}(s,t) - \frac{\partial^\nu}{\partial t^\nu} G^{dl}(s,t) \right| = O\left(\left\{\frac{n}{\log(n)}\right\}^{-\frac{1}{\nu+3}}\right); \tag{8}$$

(ii) if $N/\{n/\log(n)\}^{(1-\gamma)/(2\nu+2\gamma+4)} \to C_\nu$ *for* $0 \le \gamma < 1, 0 < C_\nu < \infty$, *and* $h_{G,\nu} = O(\{n/\log(n)\}^{-1/(2\nu+2\gamma+4)})$ *hold, then, for* $\nu = 0, 1$,

$$\max_{1\le d,l\le p} \sup_{s,t\in[h_{G,\nu}, 1-h_{G,\nu}]} \left| \widehat{\alpha}_{0\nu}^{dl}(s,t) - \frac{\partial^\nu}{\partial t^\nu} G^{dl}(s,t) \right| = O\left(\left\{\frac{n}{\log(n)}\right\}^{-\frac{1}{\nu+\gamma+2}}\right);$$

(iii) if $N/\{n/\log(n)\}^{1/(2\nu+4)} \to \infty$ *and* $h_{G,\nu} = O(\{n/\log(n)\}^{-1/(2\nu+4)})$ *hold, then, for* $\nu = 0, 1$,

$$\max_{1\le d,l\le p} \sup_{s,t\in[h_{G,\nu}, 1-h_{G,\nu}]} \left| \widehat{\alpha}_{0\nu}^{dl}(s,t) - \frac{\partial^\nu}{\partial t^\nu} G^{dl}(s,t) \right| = O\left(\left\{\frac{n}{\log(n)}\right\}^{-\frac{1}{\nu+2}}\right)$$

The rates (7) in Theorem 3.1 are combinations of estimation variance $\beta_{n,\nu}/h_{G,\nu}^{\nu}$ and bias $h_{G,\nu}^2$. We observe that if the same bandwidth is used to estimate the covariance and derivatives, i.e., $h_{G,0} = h_{G,1}$, it will trigger larger variance for the derivative estimates. This suggests to estimate the covariance and derivatives separately and to employ larger bandwidths for the derivative estimates. For example, as shown in (8) of Corollary 3.1, for sparse functional data with a finite number of observations per subject, the bandwidths $h_{G,0} = O(\{n/\log(n)\}^{-1/6})$ and $h_{G,1} = O(\{n/\log(n)\}^{-1/8})$ lead to the bias-variance trade-off convergence rates $O(\{n/\log(n)\}^{-1/3})$ and $O(\{n/\log(n)\}^{-1/4})$ for the covariances and derivatives estimates, respectively. These two rates are similar to the ones established in Müller and Yao (2010) for covariance function and its derivatives but here we went further to pursue cross-covariance functions as well. We also provide the convergence rates of the estimates when the number of time designs N_i increases with the sample size n, leading to a phase transition in the convergence rates. In setting (ii) of Corollary 3.1, the rate of derivative estimates can be of any order between $\{n/\log(n)\}^{1/4}$ and $\{n/\log(n)\}^{1/3}$ as N grows with the sample size up to the order $\{n/\log(n)\}^{1/6}$. Furthermore, when N is faster than the order $\{n/\log(n)\}^{1/6}$, it attains the fastest convergence rate, i.e., $\{n/\log(n)\}^{1/3}$. This is substantially different from the result that the covariate estimates is able to get a root-n rate when the data is very dense (Zhang and Wang, 2016). Whether root-n

convergence rate is attainable for derivative estimates is still unknown in the existing literature and we believe this discovery is of independent interest.

Next, we discuss the convergence rates of the coefficient function estimates $\widehat{B}$ in (6).

Theorem 3.2. *Under (A1) – (A4) and (B1) – (B4),*

$$\max_{1\le d,l\le p}\sup_{s,t\in[h_m,1-h_m]}\left|\widehat{B}^{dl}(t)-B^{dl}(t)\right| = O_p\left(\beta_{n,0}+\frac{\beta_{n,1}}{h_{G,1}}+h_{G,0}^2+h_{G,1}^2\right), \tag{9}$$

where $\widehat{B}(t)=(\widehat{B}^{dl}(t))_{1\le d,l\le p}$, $B(t)=(B^{dl}(t))_{1\le d,l\le p}$ *and* $h_m=\max\{h_{G,0},h_{G,1}\}$

(i) If $N_i=N<\infty, w_i=1/\{nN(N-1)\}$, *and* $h_{G,\nu}=O(\{n/\log(n)\}^{-1/(2\nu+6)})$ *hold, then*

$$\max_{1\le d,l\le p}\sup_{s,t\in[h_m,1-h_m]}\left|\widehat{B}^{dl}(t)-B^{dl}(t)\right| = O\left(\left\{\frac{n}{\log(n)}\right\}^{-\frac{1}{4}}\right); \tag{10}$$

(ii) If $N_i=N, w_i=1/\{nN(N-1)\}$, $N/\{n/\log(n)\}^{(1-\gamma)/(2\nu+2\gamma+4)}\to C_\nu$ *for* $0\le\gamma<1, 0<C_\nu<\infty$, *and* $h_{G,\nu}=O(\{n/\log(n)\}^{-1/(2\nu+2\gamma+4)})$ *hold, then*

$$\max_{1\le d,l\le p}\sup_{s,t\in[h_m,1-h_m]}\left|\widehat{B}^{dl}(t)-B^{dl}(t)\right| = O\left(\left\{\frac{n}{\log(n)}\right\}^{-\frac{1}{\gamma+3}}\right); \tag{11}$$

(iii) If $N_i=N, w_i=1/\{nN(N-1)\}$, $N/\{n/\log(n)\}^{-1/(2\nu+4)}\to\infty$ *and* $h_{G,\nu}=O(\{n/\log(n)\}^{-1/(2\nu+4)})$, $\nu=1,2$ *hold, then*

$$\max_{1\le d,l\le p}\sup_{s,t\in[h_m,1-h_m]}\left|\widehat{B}^{dl}(t)-B^{dl}(t)\right| = O\left(\left\{\frac{n}{\log(n)}\right\}^{-\frac{1}{3}}\right) \tag{12}$$

Since we plug the estimators $\widehat{\Sigma}_{00}(t)$ and $\widehat{\Sigma}_{01}(t)$ into $B(t)=\Sigma_{01}(t)\Sigma_{00}^{-1}(t)$ to get the estimator $\widehat{B}(t)=\{\widehat{\Sigma}_{01}(t)\}\{\widehat{\Sigma}_{00}(t)\}^{-1}$, the convergence rate of $\widehat{B}(t)$ is naturally determined by the rates of $\widehat{\Sigma}_{00}(t)$ and $\widehat{\Sigma}_{01}(t)$, or the maximum convergence rates of $\widehat{\Sigma}_{00}(t)$ and $\widehat{\Sigma}_{01}(t)$. As discussed after Theorem 3.1 and Corollary 3.1, the convergence rates of the estimated derivatives are generally slower than those of the corresponding covariance estimators. Actually the convergence rates of $\widehat{B}$ hinges on the convergence rates of $\widehat{\Sigma}_{01}(t)$. For instance, when optimal bandwidths $h_{G,0}=O((\log(n)/n)^{1/6})$ and $h_{G,1}=O((\log(n)/n)^{1/8})$ are used, the resulting rate of $\widehat{B}$ is $O((\log(n)/n)^{1/4})$ in (10). This is the expected order of convergence for a derivative estimate which is consistent with the rate in (8) when $\nu=1$. The convergence rates for the other settings (11) and (12) matches with the results in Corollary 3.1.

4. Numerical studies

We now carry out numerical studies to illustrate the finite sample performance of the proposed methods. All simulations were implemented via the software R (R Core Team, 2021) and the R package *fdapace* version 0.5.5 (Carroll et al., 2020) is used for the estimation of mean and covariance functions.

4.1. Simulations

We consider the random process $X=(X^1,X^2)$ with three cases below.

Case 1: The two random processes are

$$X^1(t)=\mu^1(t)+A_1\phi_1(t)+A_2\phi_2(t) \text{ and } X^2(t)=\mu^2(t)+B_1\phi_1(t)+B_2\phi_2(t),$$

where

$$\mu^1(t)=4t+e^{-(t-0.5)^2/0.02}\Big/\left(0.1\sqrt{2\pi}\right),\ \mu^2(t)=3t+e^{-(t-0.7)^2/0.32}\Big/\left(0.4\sqrt{2\pi}\right), \tag{13}$$

$\phi_1(t)=\sqrt{2}\sin(\pi t), \phi_2(t)=\sqrt{2}\cos(\pi t)$, and $(A_1,A_2,B_1,B_2)^\top$ is a Gaussian vector with zero mean and covariance matrix

$$\Sigma=\begin{pmatrix}10 & 0 & 1 & 0.1\\ 0 & 3 & 2 & 0.3\\ 1 & 2 & 5 & 0\\ 0.1 & 0.3 & 0 & 1\end{pmatrix}.$$

Case 2: The two random processes are

Table 1
Median and median absolute deviation (in parentheses) of relative integrated squared errors (RISE) of $\widehat{B}^{dl}(t), 1 \leq d, l \leq 2$ under *Case 1 – Case 3*.

		Case 1				Case 2				Case 3			
	n	$\overline{N}=8$	$\overline{N}=20$	$\overline{N}=30$	$\overline{N}=51$	$\overline{N}=8$	$\overline{N}=20$	$\overline{N}=30$	$\overline{N}=51$	$\overline{N}=8$	$\overline{N}=20$	$\overline{N}=30$	$\overline{N}=51$
$B^{11}(t)$	100	0.287	0.139	0.083	0.027	0.391	0.258	0.152	0.081	0.498	0.296	0.153	0.083
		(0.242)	(0.103)	(0.052)	(0.019)	(0.306)	(0.211)	(0.113)	(0.057)	(0.341)	(0.231)	(0.110)	(0.062)
	200	0.190	0.086	0.038	0.016	0.281	0.132	0.059	0.041	0.348	0.135	0.075	0.036
		(0.165)	(0.060)	(0.027)	(0.010)	(0.193)	(0.122)	(0.058)	(0.028)	(0.265)	(0.131)	(0.060)	(0.033)
$B^{12}(t)$	100	0.414	0.243	0.189	0.089	0.468	0.334	0.237	0.160	0.536	0.343	0.249	0.157
		(0.345)	(0.155)	(0.138)	(0.075)	(0.373)	(0.248)	(0.175)	(0.140)	(0.377)	(0.290)	(0.176)	(0.143)
	200	0.227	0.148	0.074	0.053	0.313	0.196	0.112	0.071	0.355	0.184	0.094	0.082
		(0.197)	(0.126)	(0.075)	(0.041)	(0.251)	(0.133)	(0.096)	(0.072)	(0.259)	(0.131)	(0.092)	(0.075)
$B^{21}(t)$	100	0.395	0.231	0.173	0.106	0.688	0.581	0.425	0.219	0.695	0.596	0.433	0.230
		(0.378)	(0.176)	(0.141)	(0.096)	(0.513)	(0.381)	(0.252)	(0.184)	(0.550)	(0.377)	(0.274)	(0.189)
	200	0.198	0.145	0.077	0.058	0.592	0.413	0.276	0.161	0.616	0.423	0.303	0.173
		(0.147)	(0.119)	(0.073)	(0.050)	(0.365)	(0.221)	(0.135)	(0.109)	(0.350)	(0.225)	(0.140)	(0.111)
$B^{22}(t)$	100	0.276	0.110	0.066	0.029	0.197	0.105	0.062	0.024	0.218	0.102	0.059	0.029
		(0.260)	(0.088)	(0.055)	(0.022)	(0.165)	(0.080)	(0.049)	(0.017)	(0.176)	(0.085)	(0.048)	(0.017)
	200	0.137	0.075	0.033	0.019	0.153	0.061	0.029	0.016	0.127	0.082	0.027	0.018
		(0.123)	(0.054)	(0.030)	(0.013)	(0.103)	(0.045)	(0.020)	(0.010)	(0.105)	(0.056)	(0.021)	(0.012)

$$X^1(t) = \mu^1(t) + \sum_{m=1}^{M} A_m \phi_m(t) \text{ and } X^2(t) = \mu^2(t) + \sum_{m=1}^{M} B_m \phi_m(t),$$

where the mean functions μ^1 and μ^2 are the same as (13), $M = 20$, $\phi_{2m-1}(t) = \sqrt{2}\sin(m\pi t), \phi_{2m}(t) = \sqrt{2}\cos(m\pi t)$ and $(A_1, \ldots, A_M, B_1, \ldots, B_M)^\top$ is a centered Gaussian vector with $\text{cov}(A_k, A_l) = 5\delta_{kl}/k^2$, $\text{cov}(B_k, B_l) = 8\delta_{kl}/k^3, \text{cov}(A_k, B_l) = e^{-(|k-l|+1)}\sqrt{\text{var}(A_k)\text{var}(B_l)}$ with $\delta_{kl} = 1$ if $k = l$, otherwise 0.

Case 3: The two random processes are similar to *Case 2* except that the number of eigenvalues is $M = 50$ for each process.

These three cases were designed to study the performance of the proposed method under different numbers of eigenvalues. For each case, we generated the data $\{(T_{ij}, Y_{ij}^1, Y_{ij}^2) : i = 1, \ldots, n; j = 1, \ldots, N_i\}$ with

$$Y_{ij}^l = X_i^l(T_{ij}) + \epsilon_{ij}^l, l = 1, 2.$$

The sample size n is set to be 100 and 200. The measurement errors ϵ_{ij}^l are independently sampled from a centered Gaussian distribution with variance $\sigma^2 = 0.25$. We consider the following four settings for the numbers of observations N_i and the time points T_{ij} : (i) N_i are independently sampled from $2 + \text{Po}(6)$ ($\text{Po}(\xi)$ represents the Poisson distribution with mean ξ) and T_{ij} are independently sampled from a uniform distribution on [0,1]. This leads to $\overline{N} = 8$. (ii) N_i are independently sampled from $10 + \text{Po}(10)$ and T_{ij} are independently sampled from a uniform distribution on [0,1], resulted in $\overline{N} = 20$. (iii) N_i are independently sampled from $10 + \text{Po}(20)$ and $T_{ij}, j = 1, \ldots, N_i$ are randomly selected from the grids $\{(j-1)/50 : j = 1, \ldots, 51\}$ without replacement, hence $\overline{N} = 30$. (iv) $N_i = 51$ and $T_{ij} = (j-1)/50$ for all $i = 1, \ldots, n$ and $j = 1, \ldots, 51$.

Settings (i) is for sparse functional data and (ii) pertains to the setting with moderate number of measurements per subject. Setting (iii) mimics the design of the data application in Section 4.2, and (iv) pertains to the setting of dense and regular functional data.

In the implementation, we used a Gaussian kernel to get the local polynomial estimates. For the settings (i) and (ii) with irregular time designs ($\overline{N} = 8, 20$), the bandwidths used to estimate the covariance G^{12} between X^1 and X^2 are the same as those used to estimate the covariance functions, $G^{dd}, d = 1, 2$, for X^1 and X^2, which are selected by the generalized cross-validation (GCV) method. For setting (iii) and (iv) with relatively dense and regular design ($\overline{N} = 30, 51$), $\widehat{G}^{dd}, d = 1, 2$, are calculated by the sample covariance while we used

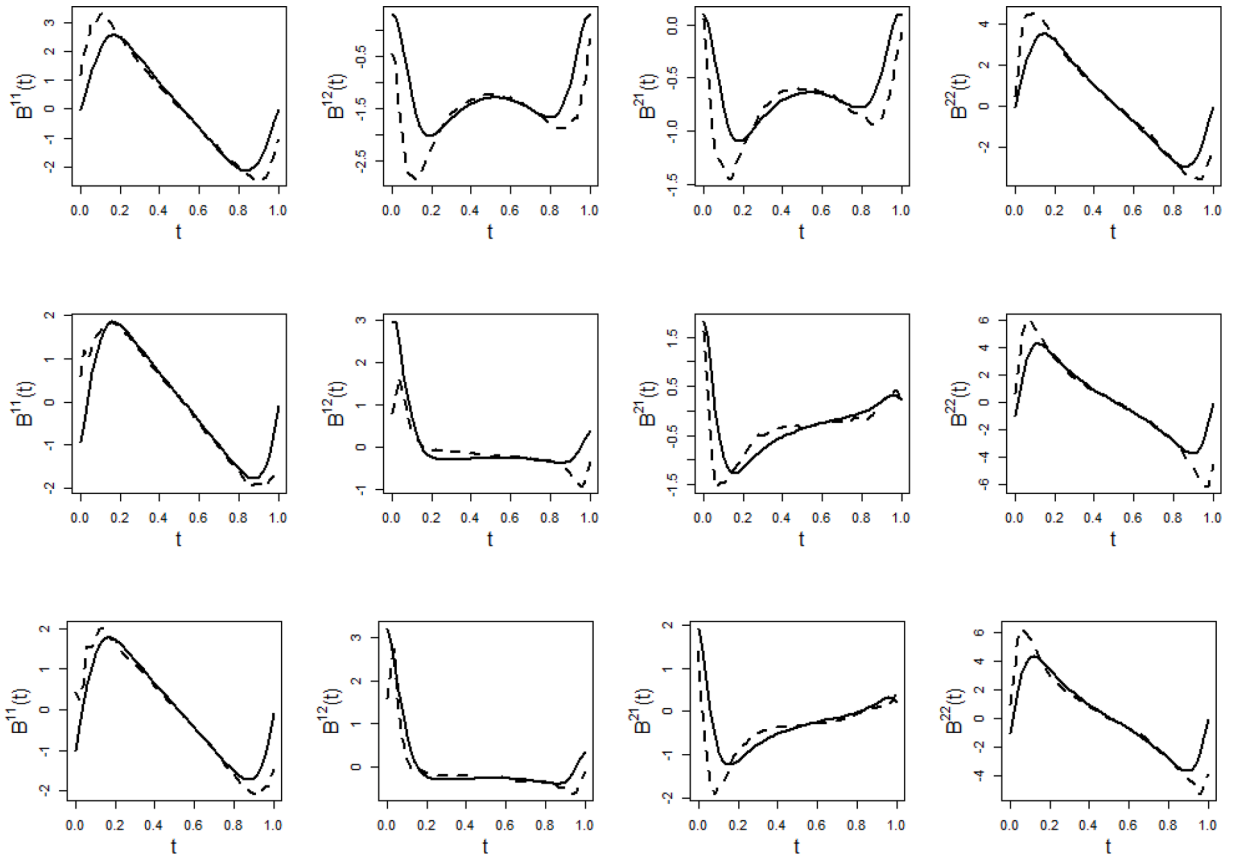

Fig. 1. Median of 200 coefficient function estimates under *Case 1* (top row), *Case 2* (middle row) and *Case 3* (bottom row) for the setting $\overline{N} = 8, n = 100$. Solid line: true coefficient function; dashed line: median of function estimates.

10% of the observed time domain as the bandwidths to get $\widehat{G}^{12}$. The bandwidths to estimate the partial derivatives are a constant factor c times the bandwidths to estimate the corresponding covariance functions, i.e., $h_{G,1} = ch_{G,0}$, where the factor c is selected from [1.5,2.5] empirically.

We performed $Q = 200$ simulation runs to evaluate the estimation results and evaluate the performance of the qth run by the relative integrated squared errors (RISE):

$$\mathrm{RISE}\left(\widehat{B}_q^{dl}\right) = \frac{\int_0^1 \left|\widehat{B}_q^{dl}(t) - B^{dl}(t)\right|^2 dt}{\int_0^1 \left|B^{dl}(t)\right|^2 dt},$$

where $B^{dl}(d, l = 1, 2)$ are the true coefficient functions. Table 1 reports the median and mean absolute deviations (MAD) of $\{\mathrm{RISE}(\widehat{B}_q^{dl}) : 1 = 1, \ldots, Q\}$

As expected, the performance of the proposed method, in terms of median and median absolute deviation (MAD), improves as the sample size or average number of observations per subject increases; and the performance deteriorates as the number of eigenfunctions increases from 2 in *Case 1* to 50 in *Case 3*.

Fig. 1 presents the median of the 200 estimated coefficient function against the true coefficient functions for *Case 1, Case 2* and *Case 3* under the sparse setting $\overline{N} = 8$ when $n = 100$, the most challenging setting. Overall, the proposed method is able to capture the shape of the target functions although the estimated functions are more extreme at the peak and valley values due to the fewer observations available near the boundary. Given that this is the worst scenario (largest bias) among the four settings, the result is encouraging.

4.2. Case study: cancer death rate and GDP per capita

We apply the proposed method to study the relations between the cancer death rate (number of death per 100,000 individuals) of a nation/region and its gross domestic product per capita (GDP per capita). The primary interest is to investigate whether and how GDP per capita affects the cancer death rate over time around the world.

The data, collected annually from 1990 to 2016, is available from https://ourworldindata.org/cancer. We only include observations for which both GDP per capita and the cancer death rates are recorded. This results in a total 4,724 observations from $n = 182$ countries or regions whose number of observations ranges from 4 to 27. Fig. 2 shows the logarithm of GDP per capita (left panel) and the logarithm of cancer death rates (right panel) of five randomly selected regions and the estimated mean functions (solid lines) among 182 countries/regions. Among these five selected countries/regions and the average estimates, the cancer death rates were stationary in the 1990s except for Russia, then steadily dropped untill 2016. In contrast, the GDP per capita went up slowly except for Russia, where the economic shock therapy, a sudden release of price and currency control, was implemented during the 1990s, after the dissolution of Soviet Union in 1991.

Denoting LGDP(t) and LCDR(t) as the logarithm of GDP per capita and the logarithm of cancer death rates at time t, repectively, we consider the following dynamic model:

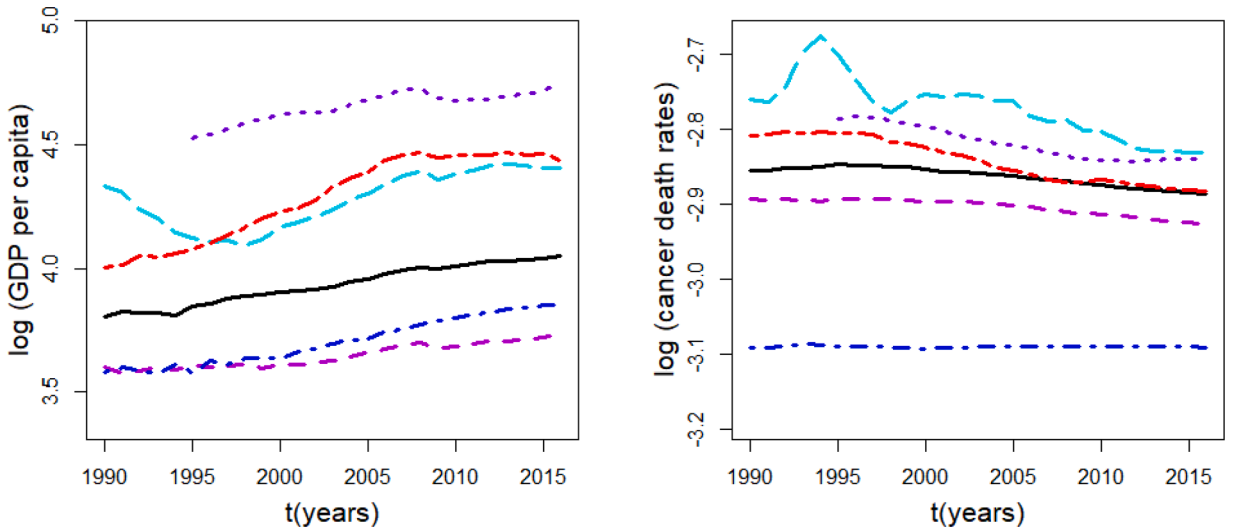

Fig. 2. The logarithm of GDP per capita (left panel) and the logarithm of cancer death rates (right panel) of five randomly selected countries/regions and the estimated mean functions (solid lines) among 182 countries/regions. Zooming in at year 2015 of the left panel, the six curves from the top to bottom are Iceland (dotted), Trinidad and Tobago (twodash), Russia (longdash), estimated mean of log GDP (solid), Morocco (dotdash), and Honduras (dashed) from top to bottom, respectively.

$$\begin{aligned} \mathrm{LGDP}'(t) &\approx B_1(t)\mathrm{LGDP}(t) + B_2(t)\mathrm{LCDR}(t) + \mathrm{INT}_1(t), \\ \mathrm{LCDR}'(t) &\approx B_3(t)\mathrm{LGDP}(t) + B_4(t)\mathrm{LCDR}(t) + \mathrm{INT}_2(t), \end{aligned} \tag{14}$$

where $\mathrm{INT}_1(t)$ and $\mathrm{INT}_2(t)$ are intercept functions. We applied the proposed methods in Section 2 to estimate $B_k, k = 1, 2, 3, 4$. The covariance functions of LGDP and LCDR are estimated by their sample covariance. The bandwidths are chosen by visual inspection to be $h_{G,0} = 0.75$ for the estimated covariances between LGDP and LCDR and $h_{G,1} = 1.6$ for the estimated derivatives.

Fig. 3 displays the estimates of the coefficient functions $B_i, i = 1, 2, 3, 4$. The areas between two dashed lines in each panel are the 95% pointwise confidence intervals (C.I.) obtained from 300 bootstrap samples, where all bandwidths are chosen the same way as those of the correspondent estimates, i.e., $h_{G,0} = 0.75$ and $h_{G,1} = 1.6$. Since we are interested in the effect of GPD per capita on cancer death rates, we focus on the function $B_3(t)$. In the figure for $B_3(t)$ (lower left panel of Fig. 3), the estimated function and the 95% C.I. are consistently negative after 1995. This suggests that LGDP is significantly negatively associated with new cancer death rates after 1995. This may be due to the comprehensive services to get treatment of cancer patient are generally accessible in higher income countries. For instance, it reports that more than 80% of children with cancer were cured in high-income countries, but only 15–45% are cured in many low- and middle-income countries (Lam et al., 2019).

Next, we check the performance of the dynamic model (14) for prediction. We fit the model (14) using the data from year 1990 to 2015, then use the derivatives estimates of $\mathrm{LCDR}'(t)$ in $t = 2015$ to predict the log cancer rate at time $t = 2016$ for each country by $\widehat{\mathrm{LCDR}}_i(t) = \mathrm{LCDR}_i(t-1) + \mathrm{LCDR}_i'(t-1)$. The mean squared prediction error $\mathrm{MSPE} = (1/182)\sum_{i=1}^{182}\{\widehat{\mathrm{LCDR}}_i(t) - \mathrm{LCDR}_i(t)\}^2|_{t=2016}$ is 1.023×10^{-5}. We then compare this with two baseline methods, for which we choose conventional varying-coefficient models without using the derivative information:

$$\begin{aligned} (M1): &\quad \mathrm{LCDR}(t) \approx B_1(t-1)\mathrm{LGDP}(t-1) + B_2(t-1)\mathrm{LCDR}(t-1) + \mathrm{INT}(t-1), \\ (M2): &\quad \mathrm{LCDR}(t) \approx B_1(t-1)\mathrm{LCDR}(t-1) + \mathrm{INT}(t-1). \end{aligned}$$

The resulting MSPE for (M1) and (M2) are 0.928×10^{-5} and 1.093×10^{-2}, respectively. The performance of the proposed dynamic model is comparable to (M1) but substantially better than (M2). This may not come as a surprise as (M1) carries similar information as the dynamic model and is in fact the same model, since we used $\mathrm{LCDR}'(t) \approx \mathrm{LCDR}(t) - \mathrm{LCDR}(t-1)$. This explains why the prediction

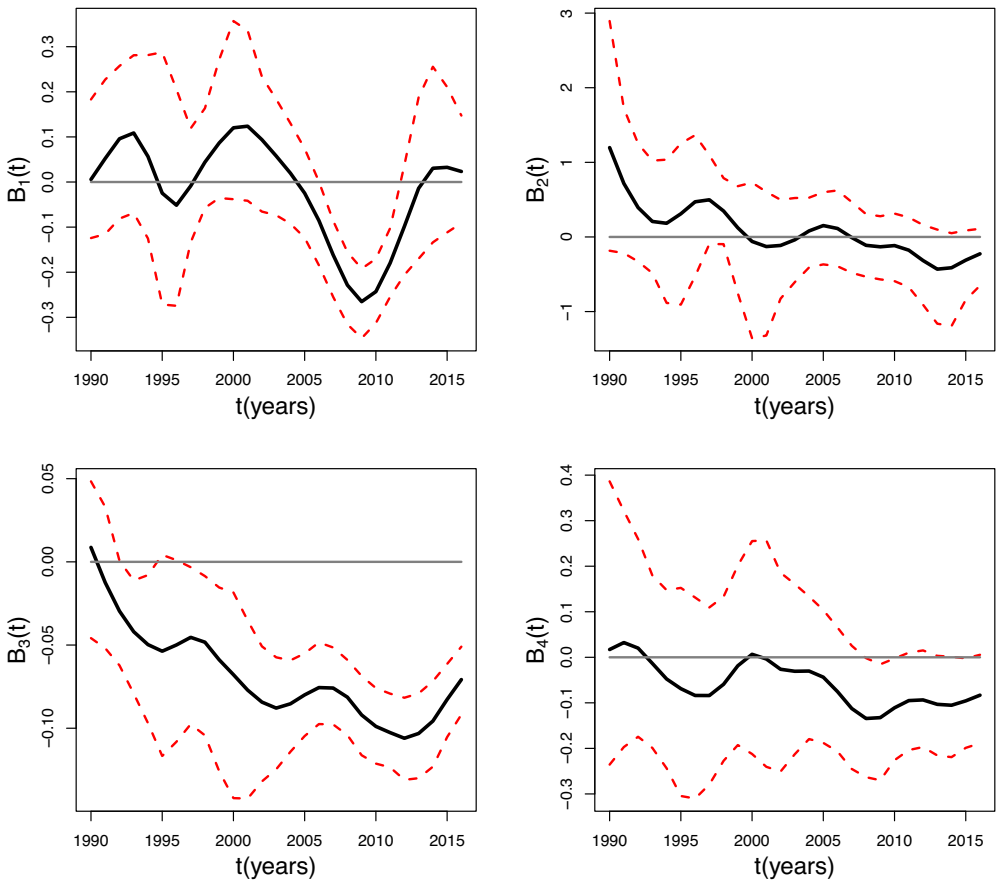

Fig. 3. Estimates of coefficient functions (solid curves) and 95% pointwise bootstrap confidence band for cancer death rates data.

errors for (M1) and the dynamic model (14) are similar.

However, prediction based on (M1) only works for regularly recorded data, such as time-series data, where data for all subjects were collected at regularly spaced time points. In contrast, the dynamic model can be used for irregularly measured functional or longitudinal data and can make predictions at any future time point t even when there were no data collected near time t for a subject.

5. Conclusions and discussion

In this paper, we use both the information of a component process of a multivariate process and those from other components to model the dynamic relation of these processes with the derivatives of the selected component. Such an approach provides an interpretable multivariate dynamic model to reveal the relationship among different functional processes. Once an estimate of $\hat{X}'(t)$ is obtained, we can use it to predict the immediate future values of $X(s)$ for s slightly larger than t by extrapolating X along its tangent at time t. That is, $\widehat{X}(s) \approx X(t) + \widehat{X}'(t)(s - t)$.

The concurrent model in (1) assumes an instantaneous relationship among the multivariate processes $\boldsymbol{X}(t)$ and their derivatives $\boldsymbol{X}'(t)$. While this may be true for a process with its own derivative, there may be a delayed relation in time with other processes. For example, Event A occurs first then leads to event B, hence the derivative for process B at certain time point is more likely to contain the information of process A at a previous time point. These time lags depict the delayed relationship among processes and will be of interest to researchers and practitioners. Similar time lags have been discussed in the study of longitudinal data. Schmid (2001) studied dynamic regression models with lagged covariates. Şentürk and Müller (2008) and Şentürk and Müller (2010) considered the varying coefficient model with the response Y depending on a period of the history of the covariate X. The role of derivatives has not been fully exploited for such purposes, which suggests a direction for future research.

Another limitation of the proposed model is that our focus is primarily on linear modeling of the dynamic mechanism. More complex dynamic mechanism using a nonlinear differential equation may be needed for other applications. Therefore, it is of interest to model the dynamics using a nonlinear differential equation:

$$\left(X^d\right)'(t) = f_d\left(t, X^1(t), \ldots, X^p(t)\right) + Z^d(t), \text{ for } d = 1, \ldots, p, \tag{15}$$

where $f_d(\cdot)$ are unknown functions.

A special case of a nonlinear dynamic model with a univariate process ($p = 1$) was studied in Verzelen et al. (2012). Their approach could be extended to densely observed multivariate functional data but not for sparsely observed functional data. The extension to sparsely observed functional data would be challenging beyond the linear dynamic setting.

Acknowledgments

We are grateful to the editor, associate editor and referees for their constructive comments and suggestions that greatly improve the earlier version of this article. All authors equally contributed to the paper and are listed in the alphabetical order. Wang's research was supported by the US National Science Foundation (DMS19-14917 and 22-10891). Zhong's research was supported by the National Natural Science Foundation of China (12201527 and 12271286) and the Fundamental Research Funds for the Central Universities (20720221032).

Supplementary material

Supplementary material associated with this article can be found, in the online version, at 10.1016/j.fmre.2023.10.003.

References

Bapna, R., Jank, W., Shmueli, G., 2008. Price formation and its dynamics in online auctions. Decis. Support Syst. 44 (3), 641–656.
Bhat, H.S., Madushani, R., 2016. Nonparametric adjoint-based inference for stochastic differential equations. 2016 IEEE International Conference on Data Science and Advanced Analytics (DSAA). IEEE, pp. 798–807.
Bone, J., 2016. Instantaneous dynamics of functional data. University of British Columbia. Ph.D. thesis.
Cai, Z., Fan, J., Li, R., 2000. Efficient estimation and inferences for varying-coefficient models. J. Am. Stat. Assoc. 95 (451), 888–902.
Cai, Z., Li, Q., 2008. Nonparametric estimation of varying coefficient dynamic panel data models. Econom. Theory 24 (5), 1321–1342.
Cai, Z., Li, Q., Park, J.Y., 2009. Functional-coefficient models for nonstationary time series data. J. Econom. 148 (2), 101–113.
C. Carroll, A. Gajardo, Y. Chen, X. Dai, J. Fan, P.Z. Hadjipantelis, K. Han, H. Ji, H.-G. Müller, J.-L. Wang, fdapace: Functional Data Analysis and Empirical Dynamics, R package version 0.5.2, 2020. https://CRAN.R-project.org/package=fdapace.
Chiang, C.-T., Rice, J.A., Wu, C.O., 2001. Smoothing spline estimation for varying coefficient models with repeatedly measured dependent variables. J. Am. Stat. Assoc. 96 (454), 605–619.
Şentürk, D., Nguyen, D.V., 2011. Varying coefficient models for sparse noise-contaminated longitudinal data. Stat. Sin. 21 (4), 1831–1856. https://doi.org/10.5705/ss.2009.328.
Dai, X., Müller, H.-G., Tao, W., 2018. Derivative principal component analysis for representing the time dynamics of longitudinal and functional data. Stat. Sin. 28 (3), 1583–1609.
Davidian, M., Lin, X., Wang, J., 2004. Introduction: emerging issues in longitudinal and functional data analysis. Stat. Sin. 14 (3), 613–614.
Fan, J., 1992. Design-adaptive nonparametric regression. J. Am. Stat. Assoc. 87 (420), 998–1004.

Fan, J., 1993. Local linear regression smoothers and their minimax efficiencies. Ann. Stat. 21 (1), 196–216. https://doi.org/10.1214/aos/1176349022.
Fan, J., Gijbels, I., 1996. Local Polynomial Modelling and Its Applications. In: Monographs on Statistics and Applied Probability, vol. 66. Chapman & Hall, London.
Fan, J., Yao, Q., Cai, Z., 2003. Adaptive varying-coefficient linear models. J. R. Stat. Soc. Ser. B (Statistical Methodology) 65 (1), 57–80.
Fan, J., Zhang, J.-T., 2000. Two-step estimation of functional linear models with applications to longitudinal data. J. R. Stat. Soc. Ser. B (Statistical Methodology) 62 (2), 303–322. https://doi.org/10.1111/1467-9868.00233.
Fan, J., Zhang, W., 1999. Statistical estimation in varying coefficient models. Ann. Stat. 27 (5), 1491–1518.
Gervini, D., Gasser, T., 2005. Nonparametric maximum likelihood estimation of the structural mean of a sample of curves. Biometrika 92 (4), 801–820.
Gu, J., Volgushev, S., 2019. Panel data quantile regression with grouped fixed effects. J. Econom. 213 (1), 68–91. https://doi.org/10.1016/j.jeconom.2019.04.006.
Hoover, D.R., Rice, J.A., Wu, C.O., Yang, L.-P., 1998. Nonparametric smoothing estimates of time-varying coefficient models with longitudinal data. Biometrika 85 (4), 809–822. https://doi.org/10.1093/biomet/85.4.809.
Huang, J.Z., Wu, C.O., Zhou, L., 2002. Varying-coefficient models and basis function approximations for the analysis of repeated measurements. Biometrika 89 (1), 111–128. https://doi.org/10.1093/biomet/89.1.111.
Huang, J.Z., Wu, C.O., Zhou, L., 2004. Polynomial spline estimation and inference for varying coefficient models with longitudinal data. Stat. Sin. 14 (3), 763–788.
Laird, N.M., Ware, J.H., 1982. Random-effects models for longitudinal data. Biometrics 38 (4), 963–974.
Lam, C., Howard, S., Bouffet, E., Pritchard-Jones, K., 2019. Science and health for all children with cancer. Science 363 (6432), 1182–1186.
Li, Q., Huang, C.J., Li, D., Fu, T.-T., 2002. Semiparametric smooth coefficient models. J. Bus. Econ. Stat. 20 (3), 412–422.
Li, Y., Hsing, T., 2010. Uniform convergence rates for nonparametric regression and principal component analysis in functional/longitudinal data. Ann. Stat. 38 (6), 3321–3351. https://doi.org/10.1214/10-AOS813.
Liang, H., Wu, H., 2008. Parameter estimation for differential equation models using a framework of measurement error in regression models. J. Am. Stat. Assoc. 103 (484), 1570–1583.
Liu, B., Müller, H.-G., 2008. Functional data analysis for sparse auction data. Stat. Methods e-Commerce Res. 269–290.
Liu, B., Müller, H.-G., 2009. Estimating derivatives for samples of sparsely observed functions, with application to online auction dynamics. J. Am. Stat. Assoc. 104 (486), 704–717.
Mas, A., Pumo, B., 2009. Functional linear regression with derivatives. J. Nonparam. Stat. 21 (1), 19–40.
Reddy, S.K., Dass, M., 2006. Modeling on-line art auction dynamics using functional data analysis. Stat. Sci. 179–193.
Müller, H.-G., Yang, W., 2010. Dynamic relations for sparsely sampled gaussian processes. Test 19 (1), 1–29.
Müller, H.-G., Yao, F., 2010. Empirical dynamics for longitudinal data. Ann. Stat. 38 (6), 3458–3486.
Pigoli, D., Sangalli, L.M., 2012. Wavelets in functional data analysis: estimation of multidimensional curves and their derivatives. Comput. Stat. Data Anal. 56 (6), 1482–1498.
R Core Team, R: A Language and Environment for Statistical Computing, R Foundation for Statistical Computing, Vienna, Austria, 2021. https://www.R-project.org/.
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 (Statistical Methodology) 69 (5), 741–796.
Ramsay, J.O., Silverman, B.W., 2005. Functional Data Analysis. Springer-Verlag, New York.
Rawat, S., 2018. Learning governing equations for stochastic dynamical systems. UC Merced. Ph.D. thesis.
Morris, J.S., Carroll, R.J., 2006. Wavelet-based functional mixed models. J. R. Stat. Soc. Ser. B (Statistical Methodology) 68 (2), 179–199. https://doi.org/10.1111/j.1467-9868.2006.00539.x.
Rice, J.A., 2004. Functional and longitudinal data analysis: perspectives on smoothing. Stat. Sin. 14 (3), 631–647.
Schmid, C.H., 2001. Marginal and dynamic regression models for longitudinal data. Stat. Med. 20 (21), 3295–3311.
Şentürk, D., Müller, H.-G., 2008. Generalized varying coefficient models for longitudinal data. Biometrika 95 (3), 653–666.
Şentürk, D., Müller, H.-G., 2010. Functional varying coefficient models for longitudinal data. J. Am. Stat. Assoc. 105 (491), 1256–1264.
Shen, Q., Faraway, J., 2004. An *F* test for linear models with functional responses. Stat. Sin. 14 (4), 1239–1257.
Verzelen, N., Tao, W., Müller, H.-G., 2012. Inferring stochastic dynamics from functional data. Biometrika 99 (3), 533–550.
Wang, L., Li, H., Huang, J.Z., 2008. Variable selection in nonparametric varying-coefficient models for analysis of repeated measurements. J. Am. Stat. Assoc. 103 (484), 1556–1569. https://doi.org/10.1198/016214508000000788.
Yao, F., Lei, E., Wu, Y., 2015. Effective dimension reduction for sparse functional data. Biometrika 102 (2), 421–437.
Yao, F., Müller, H.-G., Wang, J.-L., 2005. Functional data analysis for sparse longitudinal data. J. Am. Stat. Assoc. 100 (470), 577–590.
Zhang, X., Wang, J.-L., 2016. From sparse to dense functional data and beyond. Ann. Stat. 44 (5), 2281–2321.
B. Zhu, D.B. Dunson, Stochastic volatility regression for functional data dynamics, arXiv preprint arXiv:1212.0181(2012).