J. Math. Anal. Appl. ••• (••••) ••••••

Contents lists available at ScienceDirect

Journal of Mathematical Analysis and Applications

www.elsevier.com/locate/jmaa

Learning delay dynamics for multivariate stochastic processes, with application to the prediction of the growth rate of COVID-19 cases in the United States ☆

Paromita Dubey [a], Yaqing Chen [b,1], Álvaro Gajardo [b,1], Satarupa Bhattacharjee [b,1], Cody Carroll [b,1], Yidong Zhou [b,1], Han Chen [b,1], Hans-Georg Müller [b,*,2]

[a] *Department of Statistics, Stanford University, United States of America*
[b] *Department of Statistics, University of California, Davis, United States of America*

ARTICLE INFO

Article history:
Received 11 October 2020
Available online xxxx
Submitted by J.J. Nieto

Keywords:
Random delay differential equation
Functional data analysis
History index model
Time dynamics
Stochastic process
Economic activity

ABSTRACT

Delay differential equations form the underpinning of many complex dynamical systems. The forward problem of solving random differential equations with delay has received increasing attention in recent years. Motivated by the challenge to predict the COVID-19 caseload trajectories for individual states in the U.S., we target here the inverse problem. Given a sample of observed random trajectories obeying an unknown random differential equation model with delay, we use a functional data analysis framework to learn the model parameters that govern the underlying dynamics from the data. We show the existence and uniqueness of the analytical solutions of the population delay random differential equation model when one has discrete time delays in the functional concurrent regression model and also for a second scenario where one has a delay continuum or distributed delay. The latter involves a functional linear regression model with history index. The derivative of the process of interest is modeled using the process itself as predictor and also other functional predictors with predictor-specific delayed impacts. This dynamics learning approach is shown to be well suited to model the growth rate of COVID-19 for the states that are part of the U.S., by pooling information from the individual states, using the case process and concurrently observed economic and mobility data as predictors.

© 2021 Elsevier Inc. All rights reserved.

☆ All authors declare no financial or other competing interests. The data for this study was accessed from https://github.com/OpportunityInsights/EconomicTracker [21] on September 19, 2020. The authors are not responsible for collecting the data, declare no financial interests and maintain ethical neutrality.

* Corresponding author.
E-mail address: hgmueller@ucdavis.edu (H.-G. Müller).

[1] Contributed equally to the paper.

[2] Research supported in part by NSF Grant DMS-2014626.

https://doi.org/10.1016/j.jmaa.2021.125677
0022-247X/© 2021 Elsevier Inc. All rights reserved.

Please cite this article in press as: P. Dubey et al., Learning delay dynamics for multivariate stochastic processes, with application to the prediction of the growth rate of COVID-19 cases in the United States, J. Math. Anal. Appl. (2022), https://doi.org/10.1016/j.jmaa.2021.125677

1. Introduction

The modeling of time dynamical systems is of interest in multiple scientific fields. While ordinary differential equations (ODE) have a long history, interest in delay differential equations (DDE) is more recent, albeit both have been extensively studied [1,3,22,26,61]. A DDE is a natural extension of an ODE when observed processes have an aftereffect. Unlike the situation for ODEs, the solution of a DDE depends not only on the initial state but on the entire history of the process in a time interval of length equal to the delay prior to the initial time point.

Moving beyond the basic notion of a deterministic ODE, random differential equations (RDE) [23,46, 54,55] and stochastic differential equations (SDE) [2] are used to accommodate probabilistic uncertainty in temporal stochastic processes. In this paper we focus on RDEs, where the random effects are manifested in the model parameters. These include coefficients, initial conditions and forcing terms that are typically smooth in time, leading to differentiable sample path solutions of the RDE, in contrast to the situation for SDEs, which have a non-differentiable forcing term, typically a Wiener process, and therefore require stochastic calculus [34].

Extensive developments in the theory of the forward problem of obtaining solutions for RDEs stand in contrast with statistical approaches, which include data-oriented methodology for the inverse problem, that is, learning the nature of the differential equation from data. This approach has been referred to as empirical dynamics [45]. The starting point is a sample of random trajectories that are viewed as independent and identically distributed (i.i.d.) realizations of an underlying smooth (continuously differentiable) stochastic process. For a smooth stochastic process $X(t)$, there always exists a function f such that

$$E(X'(t)|X(t)) = f(t, X(t)), \tag{1}$$

$$X'(t) = f(t, X(t)) + Z(t), \tag{2}$$

with $E(Z(t)|X(t)) = 0$ almost surely. This forms the basis of empirical dynamics and has led to both parametric and nonparametric modeling approaches for f [57,66], which can be characterized as dynamics learning from data. Due to the minimal assumptions, dynamics learning covers a large number of specific ODEs that do not need to be a priori specified, however it requires that a sample of repeated realizations of the underlying stochastic process is available. Such samples form the backbone of functional data analysis [48,58]. A pertinent example is provided by COVID-19 caseload trajectories, where one observes samples of such trajectories across geographic units such as states or countries [18].

Related but quite distinct statistical inference concerns the problem of identifying the parameters of an a priori specified dynamic system [12,20,37], usually just from one or very few realizations. Here the trajectories are considered non-stochastic but are typically assumed to have been measured with noise. This leads to a curve fitting problem that in simple cases can be addressed with nonlinear least squares, but for which also nonparametric and semiparametric statistical methods have been employed [38,47]. A key difference is that in dynamics learning one requires data that can be considered as an independent and identically distributed sample of realizations of an underlying smooth stochastic process $X_i(t), i = 1, \ldots, n$, while the parameter identification problem is often addressed given (noisy) data from one trajectory. The modeling approach that we introduce here follows the paradigm of dynamics learning, albeit in a somewhat more structured framework than empirical dynamics.

The general form of DDEs for vector functions is

$$\frac{d}{dt}\mathbf{x}(t) = f(t, \mathbf{x}(t), \mathbf{x}_\tau(t)),$$

where $\mathbf{x}(t) = (x_0(t), \ldots, x_m(t))$ is the function value or state at time t and $\mathbf{x}_\tau(t)$ can be either a vector of the states evaluated at discrete time delays $\tau_i \geq 0$, $i = 1, ..., m$, so that $\mathbf{x}_\tau(t) = (x_0(t - \tau_0), x_1(t - \tau_1), \ldots,$

Please cite this article in press as: P. Dubey et al., Learning delay dynamics for multivariate stochastic processes, with application to the prediction of the growth rate of COVID-19 cases in the United States, J. Math. Anal. Appl. (2022), https://doi.org/10.1016/j.jmaa.2021.125677

$x_m(t-\tau_m)$), or alternatively an integral of the trajectory $\mathbf{x}(\cdot)$ over a past period, representing a continuum of delays often called distributed delays [4–6,8,15,27,31,35,41]. Results on the existence and uniqueness of random differential equations with delay (RDED) are relatively recent, where [14] studied conditions for $\mathcal{L}_p$ existence and uniqueness of the solutions of a RDED and [24] the specific case of a linear RDED with a forcing term.

In this paper we propose dynamics learning from a sample of multivariate functional data, where the model generating the observed derivative trajectories is assumed to be a RDE with a delay component. The randomness in the RDE is included in the forcing term, part of which is explained by additional covariates. These are stochastic processes with individual delay components, which in the motivating COVID-19 application correspond to trajectories of mobility and economic activity. The unexplained remainder is a drift process, which appears as $Z(t)$ in equation (2).

We utilize tools from the theory of RDE to provide a foundation for the proposed population models and establish that all the models described in Section 2 correspond to RDED with unique solutions, which depend on the model parameters. The goal is then to learn these unknown parameters, which govern the dynamics encapsulated in the population RDED, by pooling information across the sample of observed trajectories. The presence of a large number of covariates and delay components to choose from gives rise to the challenge of model selection. To address this we propose an initial pruning step as described in Section 4.2 followed by a backfitting step to optimize over a set of delay components as outlined in Section 4.3. The proposed approach differs from existing optimization strategies [41,59,60,65] and statistical methodology [36] for parameter estimation that has been previously deployed to identify a DDE based on noisy data observed for a single non-stochastic trajectory. To our knowledge, dynamics learning where one has samples of stochastic processes has so far not been explored for the case of an underlying RDED, even for the case of a one-dimensional stochastic process. We furthermore demonstrate in this paper that such an approach is well suited for modeling the growth rates of COVID-19.

We consider both discrete and distributed delay models and establish existence and uniqueness of the solutions in the $\mathcal{L}_p$ sense [14] in Section 3. For the case of discrete delays we harness functional concurrent regression models [13,33,51,53] and for distributed delays the functional history-index model [40,52], which incorporates a range of recent past values of the process. For dynamics learning of RDEDs, we adopt a two stage procedure. In a first step we utilize functional linear regression [17,42,63] with history index to learn the distributed delay, where the regression parameter function then corresponds to a history index function for the process of interest. In a second step the resulting linear predictor, which is the inner product of the history index function and the predictor process of interest, is used as a predictor along with additional covariate processes with covariate-specific delays in a concurrent model to fit the derivative process.

We apply this model to predict the time-dynamic growth rate of COVID-19 for individual states in the United States, where the sample of case trajectories of all states provides the sample of trajectories that is the starting point for the proposed dynamics learning. Predictor processes that we consider include the cumulative case (caseload) process, daily economic indicators and changes in mobility patterns. Modeling the time evolution dynamics of the COVID-19 pandemic is of great importance to understand and interpret the underlying associations as well as for deploying resources and formulating policies in the face of great uncertainty [7,11,32]. The proposed methodology is also of interest to assess the dynamics of many other empirically observed complex multivariate stochastic processes for which samples of observed trajectories are available. We show by means of leave-one-out predictions that by employing dynamics learning for the proposed RDED model with delay components one obtains considerably more accurate time-dynamic growth rate predictions for COVID-19 caseload curves compared to models without delays, demonstrating the importance of considering the inclusion of lags when modeling the growth rate of COVID-19.

Please cite this article in press as: P. Dubey et al., Learning delay dynamics for multivariate stochastic processes, with application to the prediction of the growth rate of COVID-19 cases in the United States, J. Math. Anal. Appl. (2022), https://doi.org/10.1016/j.jmaa.2021.125677

2. Proposed models

Let $(X(\cdot), \mathbf{U}(\cdot))$ denote a multivariate stochastic process where $X(\cdot)$ is a continuously differentiable process of interest, $\mathbf{U}(\cdot) = (U_1(\cdot), \ldots, U_J(\cdot))^\intercal$ is a vector function of additional covariates, and $[t_0, T]$ is a time window of interest. Consider a RDED with discrete delays,

$$\begin{aligned} \frac{dX(t)}{dt} &= \alpha(t) + \beta_0(t)X(t-\tau_0) + \sum_{j=1}^{J} \beta_j(t)U_j(t-\tau_j) + Z(t) \\ &:= f(X(t-\tau_0), t), \quad t \in [t_0, T], \\ X(t) &= g(t), \ t \in [t_0 - \tau_0, t_0], \end{aligned} \tag{3}$$

where g corresponds to the initial condition stochastic process, the τ_j, $j = 0, \ldots, J$, are discrete delays, and $\alpha(t), \beta_0(t), \beta_j(t)$ are smooth functions whose regularity will be specified below in Section 3. In the above, $Z(\cdot)$ is a random drift process that is independent of $(X(\cdot), \mathbf{U}(\cdot))$.

While the inclusion of discrete delays is an extension of the classical functional concurrent regression model [48], the functional linear regression model with history index [40,52] might be a better choice when the derivative trajectories depend not only on the predictor process value at a single past instant but on the entire continuum in the recent past. We model these distributed delays as

$$\begin{aligned} \frac{dX(t)}{dt} &= \alpha(t) + \int_0^{\tau_0} \gamma(s,t)X(t-s)\ ds \\ &\quad + \int_0^{\tau_1} \gamma_1(s,t)U(t-s)\ ds + Z(t), \quad t \in [t_0, T], \\ X(t) &= g(t), \ t \in [t_0 - \tau_0, t_0], \end{aligned} \tag{4}$$

where g again is an initial condition process. For the purpose of illustration and technical derivations, we assume that $U(\cdot)$ is a univariate process in (4); the corresponding multivariate generalization is straightforward.

For COVID-19 modeling a hybrid model that combines distributed delay in $X(\cdot)$ with discrete delays for all other predictors $U_j(\cdot)$ turned out to be particularly suitable. In this practically relevant variant of a RDED one postulates that for all $t \in [t_0, T]$,

$$X'(t) = \alpha(t) + \int_0^{\tau_0} \gamma(t,s)X(t-s)ds + \sum_{j=1}^{J} \beta_j(t)U_j(t-\tau_j) + Z(t), \tag{5}$$

with an initial condition $X(t) = g(t), \ t \in [t_0 - \tau_0, t_0]$.

We remark here that the RDED models proposed above are linear in the model parameters, although general RDED models determined by nonlinear functions are possible and might provide greater modeling flexibility, at the cost of additional modeling complexity. For example, it is not difficult to develop population models for quadratic and polynomial versions, in analogy to functional polynomial models [62]. Modeling with linear RDEDs has the advantages that such models are easy to apply and implement, come with excellent interpretability of the model parameters, and exhibit good empirical performance in our application. For the rest of the manuscript we therefore only consider linear RDEDs and leave the development of nonlinear RDEDs for future research.

Please cite this article in press as: P. Dubey et al., Learning delay dynamics for multivariate stochastic processes, with application to the prediction of the growth rate of COVID-19 cases in the United States, J. Math. Anal. Appl. (2022), https://doi.org/10.1016/j.jmaa.2021.125677

3. Existence and uniqueness of solutions

3.1. Functional concurrent model with discrete delays

Here we discuss the existence and uniqueness of the solutions of the RDEDs corresponding to the concurrent model (3) and the history-index functional linear model in (4), as well as the hybrid model in (5) used in the data application.

Consider a complete probability space $(\Omega, \mathcal{F}, P)$, where $\mathcal{F} \subset 2^{\Omega}$ is the σ-algebra on Ω and P is a probability measure. Denote the space of random variables with finite p^{th} moment by $\mathcal{L}_p$ $(p \geq 1)$, that is, for $Y : \Omega \to \mathbb{R}$, $Y \in \mathcal{L}_p$ if $E|Y|^p < \infty$. A sequence of real-valued random variables Y_m is said to converge to a random variable Y in the p^{th} moment if $||Y_m - Y||_p := (E|Y_m - Y|^p)^{1/p} \to 0$ as $m \to \infty$. We define a process $Y(\cdot)$ to be p^{th} moment continuous if for any $t_m \to t \in \mathcal{T}$, $||Y(t_m) - Y(t)||_p \to 0$ as $m \to \infty$. The notions of p^{th}-moment differentiability and p^{th}-moment Riemann integrability are defined similarly (see [54] Chapter 4, pages 92 and 100).

We say that a stochastic process $X : \mathcal{T} \to \mathcal{L}_p$ is a solution of an RDED in the $\mathcal{L}_p$ sense on the interval $\mathcal{T}$ if X is p^{th}-moment differentiable on I, p^{th}-moment continuous on $\mathcal{T}' := [t_0 - \tau_0, t_0] \cup \mathcal{T}$, and X satisfies the RDED including the corresponding initial condition [14]. We require the following assumptions on the coefficient functions and predictor processes in model (3).

(A0) $\alpha(\cdot)$ and $\beta_0(\cdot)$ are continuous on $\mathcal{T} = [t_0, T]$.
(A1) $Z(\cdot)$ is p^{th}-moment continuous on $\mathcal{T}$.
(A2) The predictor processes U_j satisfy $U_j(\cdot) \in \mathcal{L}_p$ and are p^{th}-moment continuous on $D_j := [t_0 - \tau_j, T - \tau_j]$, for $j = 1, \ldots, J$.

Calatayud et al. [14] showed the existence and uniqueness of the solution of the following general form of a random delay differential equation with discrete delay $\tau > 0$, given by

$$\begin{aligned} x'(t,\omega) &= r(x(t,\omega), x(t-\tau,\omega), t, \omega), \quad t \in [t_0, T], \\ x(t,\omega) &= r_0(t,\omega), \quad t_0 - \tau \leq t \leq t_0 \end{aligned} \tag{6}$$

and established the following result.

Theorem 2.2 of Calatayud et al. (2019) [14]. *If r satisfies the Lipschitz condition $\|r(x,y,t) - r(u,v,t)\|_p \leq k(t)\max\{\|x-u\|_p, \|y-v\|_p\}$, for $x, y \in \mathcal{L}_p$, $t \in [t_0, T]$, and a k with $\int_{t_0}^{T} |k(t)|dt < \infty$, then the random delay differential equation (6) has a unique solution in the $\mathcal{L}_p$ sense.*

The concurrent model in (3) is a special case of the random delay differential equation (6) with a specific choice of r. A key step is to show that the function f defined in (3) is p^{th}-moment continuous. The existence and uniqueness of the solutions of (3) can then be obtained following similar arguments to those in the proof of Theorem 2.2 in [14]. We summarize this result in Proposition 1. All proofs are in the Appendix.

Proposition 1. *When $\tau_0 > 0$, under the regularity conditions (A0)–(A2) and assuming that the coefficients $\beta_j(\cdot)$ are continuous on $\mathcal{T} = [t_0, T]$, $j = 1, \ldots, J$, the functional concurrent DDE model given by (3) has a unique solution in the $\mathcal{L}_p$ sense provided that the initial condition $g(\cdot)$ is p^{th}-moment continuous.*

Remark 1. Proposition 1 requires the p^{th}-moment continuity of $g(\cdot)$ on $[t_0 - \tau_0, t_0]$ for the uniqueness of the solution in the $\mathcal{L}_p$ sense. If $g(\cdot) \in C^1([t_0 - \tau_0, t_0], \mathbb{R})$ almost surely, then there exists a version of the $\mathcal{L}_p$-solution which solves the RDED in the sample path sense, see Theorem 2.3 in [14] and [16]. Note that in

Please cite this article in press as: P. Dubey et al., Learning delay dynamics for multivariate stochastic processes, with application to the prediction of the growth rate of COVID-19 cases in the United States, J. Math. Anal. Appl. (2022), https://doi.org/10.1016/j.jmaa.2021.125677

(3), we do not include the process $X(t)$ itself as one of the predictors; included is only the delayed version $X(t-\tau_0)$, where $\tau_0 > 0$. Hence solutions in the sample path sense do not require g to be differentiable almost surely. For $\tau_0 = 0$, see Remark 3.

Remark 2. The explicit sample path solution in the special case of a linear RDED has been derived in [14,16,24]. The linear RDED often includes random coefficients and a random time-varying forcing term. In (3), the terms involving the predictor processes $U_j(\cdot)$, the drift process $Z(\cdot)$, and the intercept coefficient function $\alpha(\cdot)$ form the constituents of the random forcing term; the varying coefficient functions α, β_0, β_j are non-random. Using the method of steps as outlined in the proof of Theorem 2.3 in [14], we can provide a solution to (3) in the sample path sense as follows. For $t \in [t_0, t_0+\tau_0]$, (3) is equivalent to an ordinary RDE given by

$$\begin{aligned} X'(t) &= f(g(t-\tau_0), t) \\ X(t_0) &= g(t_0), \end{aligned} \tag{7}$$

where f is defined as in (3). Then, the solution of (7) is given by

$$X_1(t) := X(t) = \int_{t_0}^{t} f(g(s-\tau_0), s)ds + g(t_0),\ t \in [t_0, t_0+\tau_0].$$

Next, for $t \in [t_0+\tau_0, t_0+2\tau_0]$,

$$\begin{aligned} X'(t) &= f(X_1(t-\tau_0), t) \\ X(t_0+\tau_0) &= X_1(t_0+\tau_0). \end{aligned} \tag{8}$$

Solving (8) leads to

$$X_2(t) := X(t) = \int_{t_0}^{t} f(X_1(s-\tau_0), s)ds + X_1(t_0+\tau_0),\ t \in [t_0+\tau_0, t_0+2\tau_0].$$

Repeating this argument until the entire domain $\mathcal{T} = [t_0, T]$ is covered one obtains the sample path solution. Due to the existence and uniqueness of the solution of (3) in the $\mathcal{L}_p$ sense and by Remark 1, the sample path solution obtained as above is unique.

Remark 3. In applications, the delay parameter can take the value zero and when $\tau_0 = 0$ in (3) one arrives at an ordinary RDE without delay. For this case, existence and uniqueness of solutions has been well studied (see, e.g., Theorem 5.1.1 in [54] and [55]). The sample path solution in this case can be obtained via the integration factor method and is given by

$$X(t) = e^{B(t)} \left(g(t_0) + \int_{t_0}^{t} q(s) e^{-B(s)} ds \right),$$

$$B(t) := \int_{t_0}^{t} \beta_0(u) du,$$

Please cite this article in press as: P. Dubey et al., Learning delay dynamics for multivariate stochastic processes, with application to the prediction of the growth rate of COVID-19 cases in the United States, J. Math. Anal. Appl. (2022), https://doi.org/10.1016/j.jmaa.2021.125677

where $q(t) := f(0,t)$ with f defined as in (3). Here $q(\cdot)$ consists of the intercept term $\alpha(t)$, the predictor processes $U_j(t)$ along with the corresponding coefficient functions $\beta_j(t),\ j = 1,\ldots,J$, and the drift process $Z(t)$.

3.2. Functional linear model with history index for distributed delay

We start with the simplest history index model with only one predictor process $U(\cdot)$, characterized as a first order RDED described in (4). We prove the existence of a unique solution of (4) by adopting similar arguments as those behind Proposition 2.1 and Theorem 2.2 of [14]. The extension to multivariate predictor processes $\mathbf{U}(\cdot)$ is straightforward. Observe that one can rewrite model (4) as

$$X'(t) = \alpha(t) + \int_{t-\tau_0}^{t} \gamma(t-s,t)X(s)ds + \int_{t-\tau_1}^{t} \gamma_1(t-s,t)U(s)ds + Z(t), \tag{9}$$

where $t \in \mathcal{T} = [t_0, T]$ with the initial condition $X(t) = g(t),\ t \in [t_0 - \tau_0, t_0]$.

Assume that $g(\cdot)$ is p^{th}-moment continuous on $[t_0 - \tau_0, t_0]$ and define $\tilde{f} : \mathcal{L}_p \times I_b \to \mathbb{R}$ such that for a p^{th}-moment continuous process $Y(\cdot) : [t_0 - \tau_0, b] \to \mathbb{R}$ and $t \in I_b := [t_0, b]$, $t_0 < b$,

$$\tilde{f}(Y,t) = \alpha(t) + \int_{t-\tau_0}^{t} \gamma(t-s,t)Y(s)ds + \int_{t-\tau_1}^{t} \gamma_1(t-s,t)U(s)ds + Z(t). \tag{10}$$

Here $\tilde{f}$ is a random functional whose first argument is the trajectory $\{Y(s) : s \in [t_0 - \tau_0, b]\}$ and whose second argument is $t \in [t_0, b]$. The following proposition with proof in Appendix B provides the continuity of $\tilde{f}$ in t in the $\mathcal{L}_p$ sense.

Proposition 2. *Suppose $Y : [t_0 - \tau_0, b] \to \mathcal{L}_p$ is p^{th}-moment continuous for $\tau_0 > 0$. Then $\tilde{f}$ as defined in (10) is continuous in t in the $\mathcal{L}_p$ sense on the domain $[t_0, b]$, $t_0 < b$.*

For the random process X on $[t_0 - \tau_0, T]$ and the time variable $t \in [t_0, T]$, the history index model (9) can be expressed as

$$X'(t) = \tilde{f}(X,t). \tag{11}$$

For the existence of a unique solution of (11) we need the following assumptions, in addition to (A0)–(A2):

(B0) γ is continuous on $\mathcal{T}' \times \mathcal{T} = [t_0 - \tau_0, T] \times [t_0, T]$ and γ_1 is continuous on $\mathcal{T}'' \times \mathcal{T} = [t_0 - \tau_1, T] \times [t_0, T]$.
(B1) $X(\cdot) \in \mathcal{L}_p$ on $\mathcal{T}' = [t_0 - \tau_0, T]$ and is p^{th}-moment continuous on I'.

The following result characterizes the solution of (11), where the integral in condition (b) of the following Proposition is defined in the p^{th}-moment Riemann integral sense [for definition of p^{th}-moment Riemann integrability see [54], Chapter 4 page 100].

Proposition 3. *The process $X : \mathcal{T} \to \mathcal{L}_p$ is a $\mathcal{L}_p$ solution of (11) if and only if for $\tau_0 > 0$,*

(a) *X is p^{th}-moment continuous on $[t_0 - \tau_0, T]$*
(b) *$X(t) = g(t_0) + (\mathcal{L}_p)\int_{t_0}^{t} \tilde{f}(X,s)ds$ for each $t \in \mathcal{T}$*
(c) *$X(t) = g(t)$ for each $t \in [t_0 - \tau_0, t_0]$.*

Please cite this article in press as: P. Dubey et al., Learning delay dynamics for multivariate stochastic processes, with application to the prediction of the growth rate of COVID-19 cases in the United States, J. Math. Anal. Appl. (2022), https://doi.org/10.1016/j.jmaa.2021.125677

The following result establishes the existence and uniqueness of a solution for (11). The proof extends the classical Picard theorem for deterministic ODE to RDED in the $\mathcal{L}_p$ sense, via the Banach fixed-point theorem, following a similar line of arguments as in [14] (see Appendix B).

Theorem 1. *Assume the regularity conditions (A0)–(A2) and (B0)–(B1). Then for $\tau_0 > 0$ we have that*

$$k(t) = \int_{t-\tau_0}^{t} |\gamma(t-s,t)|ds \in L^1(\mathcal{T}),\ t \in \mathcal{T} = [t_0, T],$$

and a solution of (4) exists and is unique in the $\mathcal{L}_p$ sense.

For the existence and uniqueness of the solution of the hybrid model in (5), which is the most promising of the models considered for modeling the growth rate of COVID-19, the following corollary applies.

Corollary 1. *Suppose the regularity conditions (A1), (A2) and (B1) hold. If $\alpha(\cdot)$ and the coefficients $\beta_j(\cdot)$ are continuous on $\mathcal{T} = [t_0, T]$, $j = 1, \ldots, J$, and γ is continuous on $\mathcal{T}' \times \mathcal{T} = [t_0 - \tau_0, T] \times [t_0, T]$, then the model in (5) has a unique $\mathcal{L}_p$ solution.*

4. Learning dynamics from samples of multivariate stochastic processes and practical considerations for COVID-19 caseload modeling

4.1. Obtaining derivatives

In data-driven differential equation modeling, given a sample of observed processes, a necessary first step is the estimation of derivatives. In applications such as the COVID-19 case trajectories the available data are usually noise-contaminated, and therefore some care is needed to obtain viable derivative estimates. The situation for the observed data for one of the trajectories is reflected by the nonparametric regression model

$$Y_k = X(t_k) + \epsilon_k, \quad k = 1, ..., K, \tag{12}$$

where $\{t_k\}_{k=1}^K$ are the grid points where the process X is measured and ϵ_k are the measurement errors, which typically are considered to be independent mean zero random variables with finite variance.

In our data application, the available data Y_k correspond to the cumulative case counts of COVID-19, which we consider to be noisy realizations of a smooth underlying function $X(\cdot)$ that is observed on a daily grid. For the practical estimation of the derivatives $X'(t)$ one has various choices that all require a tuning parameter. Since in preliminary studies local polynomial fitting yielded the smallest leave-one-out prediction error relative to the difference quotients, which serve as nearly unbiased targets, we opted to obtain derivatives by applying local polynomial regression [28,44] as per Appendix C, where further details are provided.

4.2. Target model and initial lag and covariate process selection

For modeling the COVID-19 caseload process X for the states of the U.S. we adopt the RDED model (5),

$$X'(t) = \alpha(t) + \int_0^{\tau_0} \gamma(t,s)X(t-s)ds + \sum_{j=1}^{J} \beta_j(t)U_j(t-\tau_j) + Z(t), \quad t \in \mathcal{T}$$

Please cite this article in press as: P. Dubey et al., Learning delay dynamics for multivariate stochastic processes, with application to the prediction of the growth rate of COVID-19 cases in the United States, J. Math. Anal. Appl. (2022), https://doi.org/10.1016/j.jmaa.2021.125677

where the time delay of the effect of the process X at time $t \in \mathcal{T}$ is distributed on the interval $[t-\tau_0, t]$, and $\{\tau_j\}_{j=1}^J$ are discrete time delays or time lags for the covariate or predictor processes U_j. Our starting point is that the data $\{(X_i, U_{i1}, \ldots, U_{iJ})\}_{i=1}^n$ represent n independent realizations of the stochastic process $(X, U_1, \ldots, U_J)$, where each realization (corresponding to a state) is indexed by i and n is the number of states included in the analysis.

Given τ_0, to estimate the history index weight function $\gamma(s,t)$ in the historical model without covariates, we preform a function-to-scalar linear regression for each grid point t_k, i.e.,

$$X'(t_k) = \alpha(t_k) + \int_0^{\tau_0} \gamma(s, t_k) X(t_k - s)\, ds + Z(t_k), \;\; 1 \le k \le K. \tag{13}$$

This can be implemented using function `FLM` in the R package `fdapace` [19]. With estimates of $\gamma(s, t_k)$ for each k in hand, the estimate of the history index weight function, $\widehat{\gamma}(s,t)$ for all $s, t \in \mathcal{T}$, is obtained using a two dimensional kernel smoother.

For each of the covariates U_j, we conduct an initial lag selection based on functional varying-coefficient regression including the corresponding variable as a single predictor, considering the models

$$X'(t_k) = \alpha_j(t_k) + \beta_j^*(t_k) U_j(t_k - \tau) + \epsilon_{jk}, \;\; 0 \le j \le J, \; 1 \le k \le K, \tag{14}$$

where $U_0 := X$. Given a lag τ, for each subject i and time point t_k, we obtain the leave-one-out prediction

$$\widetilde{X}'_{i,-i}(t_k; \tau) = \widetilde{\alpha}_{j,-i}(t_k) + \widetilde{\beta}^*_{j,-i}(t_k) U_{ij}(t_k - \tau),$$

where $\widetilde{\alpha}_{j,-i}(t_k)$ and $\widetilde{\beta}^*_{j,-i}(t_k)$ are estimated by fitting model (14) by ordinary least squares, excluding the data from subject i. The lags are chosen to minimize the leave-one-out prediction error across subjects,

$$\widetilde{\tau}_j = \underset{\tau \in \{0,1,\ldots,21\}}{\operatorname{argmin}} \frac{1}{nK} \sum_{i=1}^n \sum_{k=1}^K \left(X'_i(t_k) - \widetilde{X}'_{i,-i}(t_k; \tau) \right)^2. \tag{15}$$

With initial choices of the lags $\{\widetilde{\tau}_j\}_{j=1}^J$ and the estimate $\widehat{\gamma}$ obtained from model (13), for each t_k, variable selection is then performed using LASSO [56],

$$\left(\check{\alpha}(t_k), (\check{\beta}_j(t_k))_{j=0}^J \right) = \underset{\alpha_k, \boldsymbol{\beta}_k}{\operatorname{argmin}} Q(\alpha_k, \boldsymbol{\beta}_k),$$

$$Q(\alpha_k, \boldsymbol{\beta}_k) = \sum_{i=1}^n \left(X'_i(t_k) - \alpha_k - \beta_{0k} \int_0^{\tau_0} \widehat{\gamma}(t_k, s) X(t_k - s) ds - \sum_{j=1}^J \beta_{jk} U_{ij}(t_k - \widetilde{\tau}_j) \right)^2 + \lambda \sum_{j=0}^J |\beta_{jk}|,$$

where $\boldsymbol{\beta}_k = (\beta_{0k}, \ldots, \beta_{Jk})$. This can be implemented using the R package `ncvreg` [9,10], where the tuning parameter λ is chosen by leave-one-out cross-validation. Considering all K time points simultaneously, we obtain proportions of the number of time points t_k where each variable U_j is selected, i.e.,

$$p_j = \#\{k : \check{\beta}_j(t_k) \neq 0\}/K, \quad \text{for } j = 1, \ldots, J.$$

Subsequently, selection among the time-varying predictors $\{U_j\}_{j=1}^J$ other than the process X is performed by applying a threshold on these proportions $\{p_j\}_{j=1}^J$. This threshold p^* is chosen by leave-one-out cross-validation,

$$\{U_j : p_j \ge p^*\}. \tag{16}$$

Please cite this article in press as: P. Dubey et al., Learning delay dynamics for multivariate stochastic processes, with application to the prediction of the growth rate of COVID-19 cases in the United States, J. Math. Anal. Appl. (2022), https://doi.org/10.1016/j.jmaa.2021.125677

4.3. Lag selection by backfitting

Let $\mathcal{V}$ denote the set of selected predictor processes obtained using the method described in Section 4.2. Updated lag parameters are then chosen for the variables in $\mathcal{V}$ through an iterative backfitting algorithm such that the total leave-one-out prediction error is minimized. Given the b^{th} iteration of the lag vector, one obtains the $(b+1)^{st}$ update for the k^{th} lag by minimizing the leave-one-out integrated mean squared prediction error, i.e.,

$$\tau_k^{(b+1)} = \underset{\tau_k}{\operatorname{argmin}} \frac{1}{n} \sum_{i=1}^{n} \int_{\mathcal{T}} (X_i'(t) - \hat{X}_{i,-i}'^{(b)}(t;\tau_k))^2 dt,$$

where for $k \in \mathcal{V}$

$$\hat{X}_{i,-i}'^{(b)}(t;\tau_k) = \hat{\alpha}_{-i}^{(b)}(t) + \hat{\beta}_{0,-i}^{(b)}(t) \int_0^{\tau_0} \widehat{\gamma}(t,s) X_i(t-s+\tau_0) ds + \sum_{j \neq k, j \in \mathcal{V}} \hat{\beta}_{j,-i}^{(b)}(t) U_{j,i}(t-\tau_j^{(b)}) + \hat{\beta}_{k,-i}^{(b)}(t) U_{k,i}(t-\tau_k).$$

Here the functions $\hat{\beta}_{k,-i}^{(b)}(t)$ represent the fitted coefficients corresponding to the b^{th} iteration of the j^{th} lag, $j \in \mathcal{V}$ with the i^{th} observation held out for prediction. This process is iterated over $k \in \mathcal{V}$ until the lag vector converges. Lags are initialized at the vector $\boldsymbol{\tau}^{(0)} = (\widetilde{\tau}_1, \ldots, \widetilde{\tau}_j, \ldots, \widetilde{\tau}_{|\mathcal{V}|})^T$, where $\widetilde{\tau}_j$ are as described in Section 4.2 and the stopping rule is such that the algorithm terminates if no lags have changed after a complete cycle of updates for all predictors. After the final lags are estimated, we lightly smooth the estimated coefficients with local linear smoothing, which is a standard procedure in concurrent functional regression modeling [25,29]. Specifically, the initial estimates $\hat{\beta}_j(t_k)$ are smoothed to obtain the effect functions presented in Figs. 3 and 4, defined as $\hat{\beta}_j^{smooth}(t) = \hat{b}_j^{(0)}$ where

$$(\hat{b}_j^{(0)}, \hat{b}_j^{(1)}) = \underset{b_j^{(0)}, b_j^{(1)}}{\operatorname{argmin}} \sum_{k=1}^{K} K\left(\frac{t_k - t}{h}\right) (\hat{\beta}_j(t_k) - b_j^{(0)} - b_j^{(1)}(t - t_k))^2, \tag{17}$$

with $K(\cdot)$ representing the Epanechnikov kernel with a bandwidth of $h = 20$ days.

5. Random differential equation with delay fitting and prediction for COVID 19 case trajectories in the United States

5.1. Data description

We obtained daily confirmed cases across states in the United States from the COVID-19 Data Repository by the Center for Systems Science and Engineering (CSSE) at Johns Hopkins University, which is publicly available at https://github.com/CSSEGISandData/COVID-19 and was accessed on September 19, 2020. The positive and negative COVID-19 test results per day and state were obtained from the COVID Tracking Project and are publicly available at https://COVIDtracking.com/ (accessed on September 19, 2020). The former variable refers to the number of people with confirmed or probable case of COVID-19 (see https://COVIDtracking.com/about-data/data-definitions for more details) and the latter refers to the total number of people with a completed negative PCR test. We used effective positivity rates (EPR), defined as the ratio of effective positive results (positive test or probable case) to the effective total tests (positive or probable case plus negative test results).

Please cite this article in press as: P. Dubey et al., Learning delay dynamics for multivariate stochastic processes, with application to the prediction of the growth rate of COVID-19 cases in the United States, J. Math. Anal. Appl. (2022), https://doi.org/10.1016/j.jmaa.2021.125677

Table 1
List of all variables considered (see Data Description Section for further details).

Item	Feature description
1	cases per million
2	ratio of effective positive results to effective total tests
3	ratio of effective positive results to state population in 2019
4	individuals who are currently hospitalized with COVID-19
5	spending in all categories
6	spending in arts, entertainment, and recreation categories
7	spending in accommodation and food service categories
8	spending in general merchandise stores and apparel and accessories categories
9	spending in grocery and food store categories
10	spending in health care and social assistance categories
11	spending in transportation and warehousing categories
12	spending among high (top quartile) income ZIP codes in all categories
13	spending among low (bottom quartiles) income ZIP codes in all categories
14	spending among middle (middle two quartiles) income ZIP codes in all categories
15	time spent at retail and recreation locations
16	time spent at grocery and pharmacy locations
17	time spent at parks
18	time spent at inside transit stations
19	time spent at work places
20	time spent at residential locations
21	time spent outside of residential locations
22	number of small businesses open
23	number of small businesses open among high(top quartile) income ZIP codes
24	number of small businesses open among low(bottom quartile) income ZIP codes
25	number of small businesses open among middle(middle two quartiles) income ZIP codes
26	number of small businesses open in transportation
27	number of small businesses open in education and health services
28	number of small businesses open in leisure and hospitality
29	net revenue for small businesses
30	net revenue for small businesses among high(top quartile) income ZIP codes
31	net revenue for small businesses among low(bottom quartile) income ZIP codes
32	net revenue for small businesses among middle(middle two quartiles) income ZIP codes
33	net revenue for small businesses in transportation
34	net revenue for small businesses in education and health services
35	net revenue for small businesses in leisure and hospitality

We also obtained features from the *Opportunity Insights Economic Tracker Data* [21], which is publicly available at https://github.com/OpportunityInsights/EconomicTracker (accessed on September 19, 2020). This database contains daily and state level information for several economic activity indicators such as credit and debit card expenditure across different types of activities, some of them stratified by zip codes with low/medium/high income level. We refer to [21] for more details. Additionally, we downloaded Google mobility data, publicly available at https://www.google.com/COVID19/mobility/ (accessed on September 19, 2020). These data include indicators of mobility pattern changes in different areas such as parks, residential locations, retail locations, among others; and information such as the percent change in revenue as well as total number of small businesses for high/low income consumers and across different economic activities; see Table 1 for a complete listing.

We briefly describe a few features below and refer to the Economic Tracker source [21] for further details:

- 'spend all': Seasonally adjusted credit and debit card spending relative to the period Jan 4 to Jan 31, 2020, in all merchant categories, reported as a seven day moving average formed by using the values for the current and the previous six days.
- 'spend tws': Seasonally adjusted credit and debit card spending relative to the period Jan 4 to Jan 31, 2020, in the category transportation and warehousing, and also reported as a seven day moving average.
- 'spend all inchigh': Seasonally adjusted credit and debit card spending by consumers living in high median income (above 75% quantile) zip codes, and relative to the period Jan 4 to Jan 31, 2020, in all merchant categories, again reported as a seven day moving average.
- 'revenue all': Seasonally adjusted percent change in net revenue for small businesses and indexed to the period Jan 4 to Jan 31, 2020, reported as a seven day moving average.

Please cite this article in press as: P. Dubey et al., Learning delay dynamics for multivariate stochastic processes, with application to the prediction of the growth rate of COVID-19 cases in the United States, J. Math. Anal. Appl. (2022), https://doi.org/10.1016/j.jmaa.2021.125677

Table 2
Predictors and their corresponding lags (see Data Description Section for further details).

Predictor	variable	lag
Park mobility activity	$U_1(t)$	0
Number of people currently hospitalized	$U_2(t)$	14
Effective test positivity rate	$U_3(t)$	7
Credit/debit card spending on arts, entertainment, and recreation	$U_4(t)$	3

- 'revenue ss70': Seasonally adjusted percent change in net revenue for small businesses in leisure and hospitality and indexed to the period Jan 4 to Jan 31, 2020, reported as a seven day moving average.
- 'merchant inchigh': Seasonally adjusted percent change in the number of small businesses open in high median income (above 75% quantile) zip codes and indexed to the period Jan 4 to Jan 31, 2020, reported as a seven day moving average.
- 'gps parks': Corresponds to the time spent at parks relative to a baseline which is defined as the median value, for the corresponding day of the week, during the period Jan 3–Feb 6, 2020.

We employ local linear regression to slightly smooth the feature trajectories; this also allows to impute missing observations for some states and days. For this step, a Gaussian kernel with bandwidth 1.5 days was employed and we utilized the R package `fdapace` version 0.5.5 for the computational implementation [19]. The effective test positivity rate was smoothed analogously. We use local quadratic regression [28] to obtain the derivatives $X'(t)$, implemented by using the R package `locfit` version 1.5-9.4 [39]. For data analysis we consider states with population at least 1 million as the states with smaller populations had less COVID-19 cases, and their inclusion would therefore increase the overall noise level in the data and negatively impact the analysis. This is a pragmatic choice and is not a limitation of the proposed method. In the end, we included the data from 44 states (excluding Alaska, Delaware, North and South Dakota, Vermont and Wyoming).

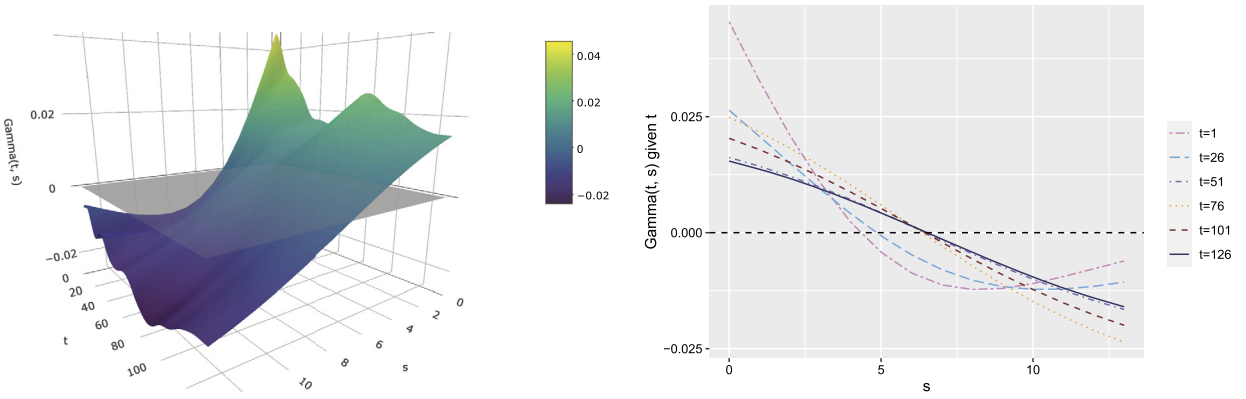
As a result of the LASSO variable selection described in Section 4.3, which was implemented with the R package `ncvreg` version 3.12.0 [9] with threshold $p^* = 0.3$ in (16), the features utilized in the final model are as follows, where the selected lags are listed in Table 2.

1. $X_i(t)$, the total cases per million at time t for state i, $i = 1, \ldots, 44$.
2. $U_{1i}(t)$, the relative time spent at park areas at time t in state i.
3. $U_{2i}(t)$, the number of hospitalized people at time t in state i.
4. $U_{3i}(t)$, the "positive testing rate" defined as the ratio of effective positive results (positive test or probable case) to the effective total tests (positive or probable case plus negative test results) at time t for state i.
5. $U_{4i}(t)$, the credit/debit card relative spending in the arts, entertainment, and recreation category (see 'spend aer' predictor in Table 1) at time t in state i.

5.2. Model fitting and results

In order to study the delay time dynamics of COVID-19 in the United States, we focus on the state-specific time-varying cumulative confirmed cases per million people, $X(t)$, for $t \in \mathcal{T}$, which is referred to as the *case process* in the following. Here, the time domain $\mathcal{T}$ that we consider starts on April 5 and ends on August 12 in 2020, where data are recorded once per day, on an equidistant grid for 130 days. We estimate the history coefficient function γ in the RDED model (5) as described in Section 4.2 with $\tau_0 = 14$, and adopt the functional varying-coefficient regression in model (5) with $J = 34$ other variables as listed in Table 1. The estimated coefficient surface $\hat{\gamma}(s,t)$ and some of its cross-sections are displayed in Fig. 1.

Please cite this article in press as: P. Dubey et al., Learning delay dynamics for multivariate stochastic processes, with application to the prediction of the growth rate of COVID-19 cases in the United States, J. Math. Anal. Appl. (2022), https://doi.org/10.1016/j.jmaa.2021.125677



(a) (b)

Fig. 1. (a) Estimated coefficient surface $\hat{\gamma}(s,t)$ in the RDED model (5), where t corresponds to number of days since April 5, 2020 and s takes values from 0 to τ_0. (b) t-sliced cross sections for $t = 1, 26, 51, 76, 101, 126$ of $\hat{\gamma}(s,t)$ as functions of s.

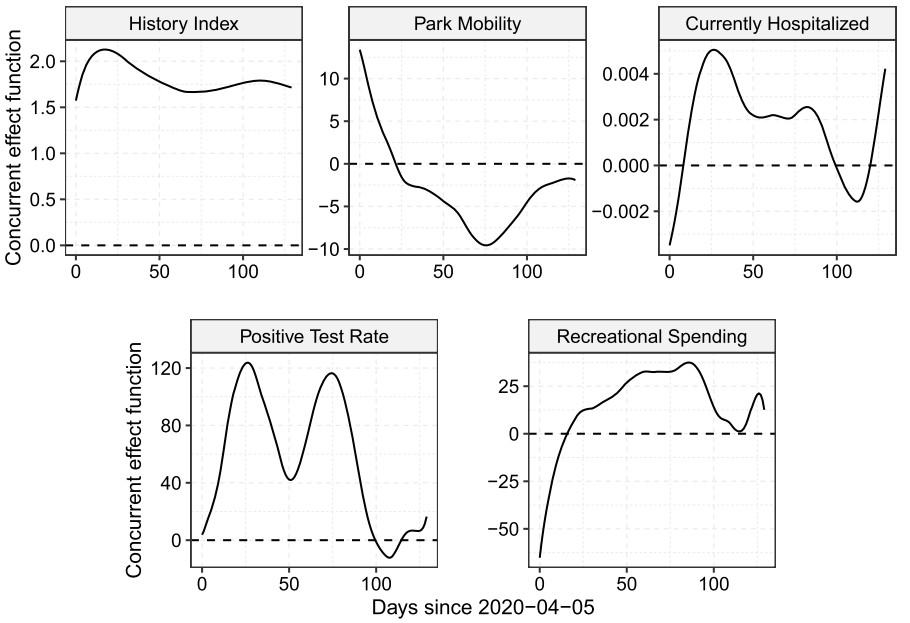


Fig. 2. Concurrent effect functions $\hat{\beta}_j$ for predictor processes U_j in the RDED (5).

For each fixed t, $\hat{\gamma}(\cdot, t)$, the estimated history index in model (5) is found to change from positive near current time to negative with increasing lags. This means that the recent past of the process has a positive association with the current derivative while days further into the past are negatively associated. The right panel shows that the shape of $\hat{\gamma}(\cdot, t)$ changes from more or less quadratic to linear from the early days of COVID-19 to recent times, indicating that in April 2020 the contrast between the effects of near-current and past caseload on the current derivative was more pronounced than it is more recently.

5.3. Predictor process effects on growth rate

The results are visualized in Fig. 2. States which are suffering from faster spread of the virus are characterized by higher growth rates. The historically weighted integrated case count per million has a positive effect on the growth rate. This reflects dynamic explosive behavior [45], so that states will tend to have

Please cite this article in press as: P. Dubey et al., Learning delay dynamics for multivariate stochastic processes, with application to the prediction of the growth rate of COVID-19 cases in the United States, J. Math. Anal. Appl. (2022), https://doi.org/10.1016/j.jmaa.2021.125677

caseloads that move even further away from the mean caseload per million across states as time moves forward, which could be either moving further above the average across states or further below it. Thus the positive effect of the history index reflects increasing variability in the caseloads across states. Potential reasons for this are the distinct policies that states implement in terms of mobility, business and social restrictions, as well as cultural acceptance of such restrictions. Spatial infectivity dynamics with waves of infections moving spatially across the U.S. also might play a major role.

In terms of the effects of covariate processes, our results indicate that higher park mobility during late spring and summer (May onwards) is associated with a decrease in growth rates. For each percent increase in park mobility, the growth rate decreased by as much as 10 cases per million per day, with the effect reaching its maximum in mid-summer. This indicates negligible risk and potential benefits from outdoors recreation activities.

The other covariate processes that were selected by LASSO as predictors in the RDED model (5) are patient counts at hospitals, the testing positivity rate, and credit card spending on recreational activities, including dining at restaurants. All of these were found to have positive effects on the growth rate throughout nearly the entire time domain. These effects were included in model (5) with covariate process-specific lags, as per Table 2. The number currently hospitalized was found to be most predictive for the case growth rate when including a two-week lag, whereas for test positivity rate a one-week lag was found to be most predictive. This may suggest that individuals who are infected by individuals with less severe symptoms (i.e. a non-hospitalized COVID-positive person) tend to learn of their disease one week after their infector did.

The two-week difference for hospitalized individuals is likely a reflection of the time needed to develop serious symptoms which warrant admission for medical care. The effect of recreational spending on viral spread was much more immediate; the optimal lag in this case is only 3 days. This suggests that states with higher recreational economies during the pandemic see an almost immediate uptick in growth rate, with cases increasing by as much as 25 infections per million per day for each percent of additional recreational spending.

5.4. Predictive performance of the random differential equation with delay model

Figs. 3 and 4 display the fitted growth rates provided by the RDED model (5), with covariate specific lags, and also the results of fitting a differential equation model where the covariates do not have individual lags, i.e. a concurrent RDE model given by (5) in which $\tau_j = 0, j = 0, 1, 2, 3, 4$. Fitted curves are the result of leave-one-out predictions for each state, where the data from the left-out state were not used for fitting the model and then the state's data are plugged into the fitted model to obtain the predicted case growth rate. These predictions are compared against the actual observed growth rates as obtained by local polynomial regression (see Section 4.1 and Appendix C). Visual inspection (as well as measures of goodness-of-fit, such as integrated mean squared prediction error) indicates that the incorporation of delays with predictor specific lags in the model substantially improves the fits. This provides compelling evidence for modeling COVID-19 growth rates with RDEDs that include covariate-specific lags. A vanilla RDE without predictor delays does not incorporate historical information beyond the previous day and struggles to explain deviations in phase (as in, e.g. Illinois). The problem of modeling phase variation is instead foisted upon the coefficient functions, which are ill-equipped to reflect it, and thus at times the straight RDE model without predictor delays falters in terms of amplitude modeling as well (e.g. for Colorado). In contrast, the proposed RDED incorporates past information through the included delays and thus reflects phase variability in a natural way, especially through the inclusion of distributed delays via the history index. Including a two-week history of case counts through the history index function γ in model (5) substantially enhances the flexibility of the model. The estimates of coefficient functions are no longer uncontrollably altered by the effects of unaccounted lags, and as a result the fitted model performs much better.

Please cite this article in press as: P. Dubey et al., Learning delay dynamics for multivariate stochastic processes, with application to the prediction of the growth rate of COVID-19 cases in the United States, J. Math. Anal. Appl. (2022), https://doi.org/10.1016/j.jmaa.2021.125677

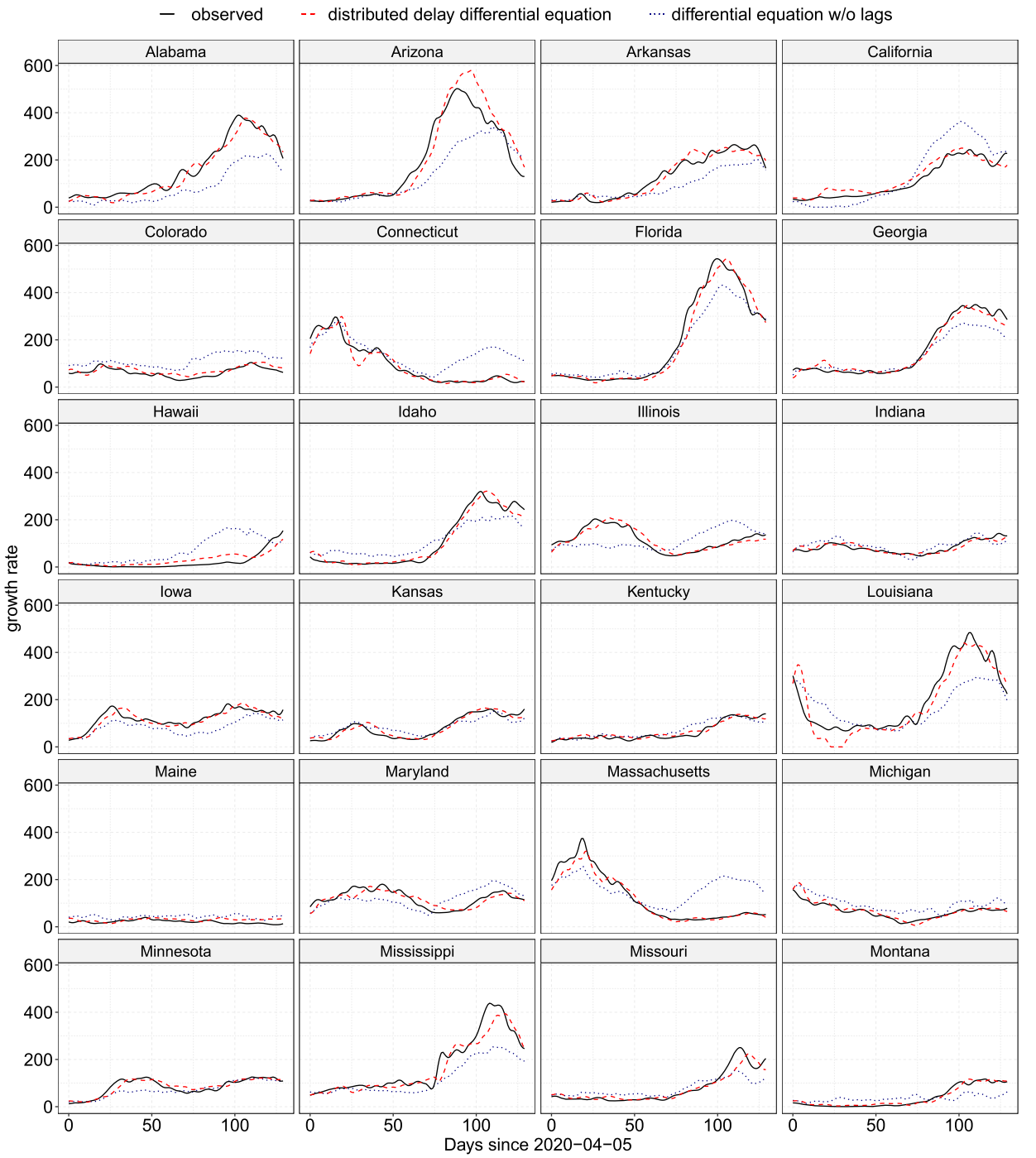


Fig. 3. Predicted growth rates by the proposed RDED (5) (red, dashed) and by a RDE without lags (blue, dotted) for US States, Alabama–Montana. Observed rates are the black solid curves. (For interpretation of the colors in the figure(s), the reader is referred to the web version of this article.)

5.5. Model-benchmarked evaluation of performance

A comparison of observed vs. RDED model-predicted growth rates allows one to assess the performance of states in terms of controlling spread. A state whose observed growth rate falls below the predicted curve corresponds to a negative residual and suggests better-than-expected viral control, and vice versa for

Please cite this article in press as: P. Dubey et al., Learning delay dynamics for multivariate stochastic processes, with application to the prediction of the growth rate of COVID-19 cases in the United States, J. Math. Anal. Appl. (2022), https://doi.org/10.1016/j.jmaa.2021.125677

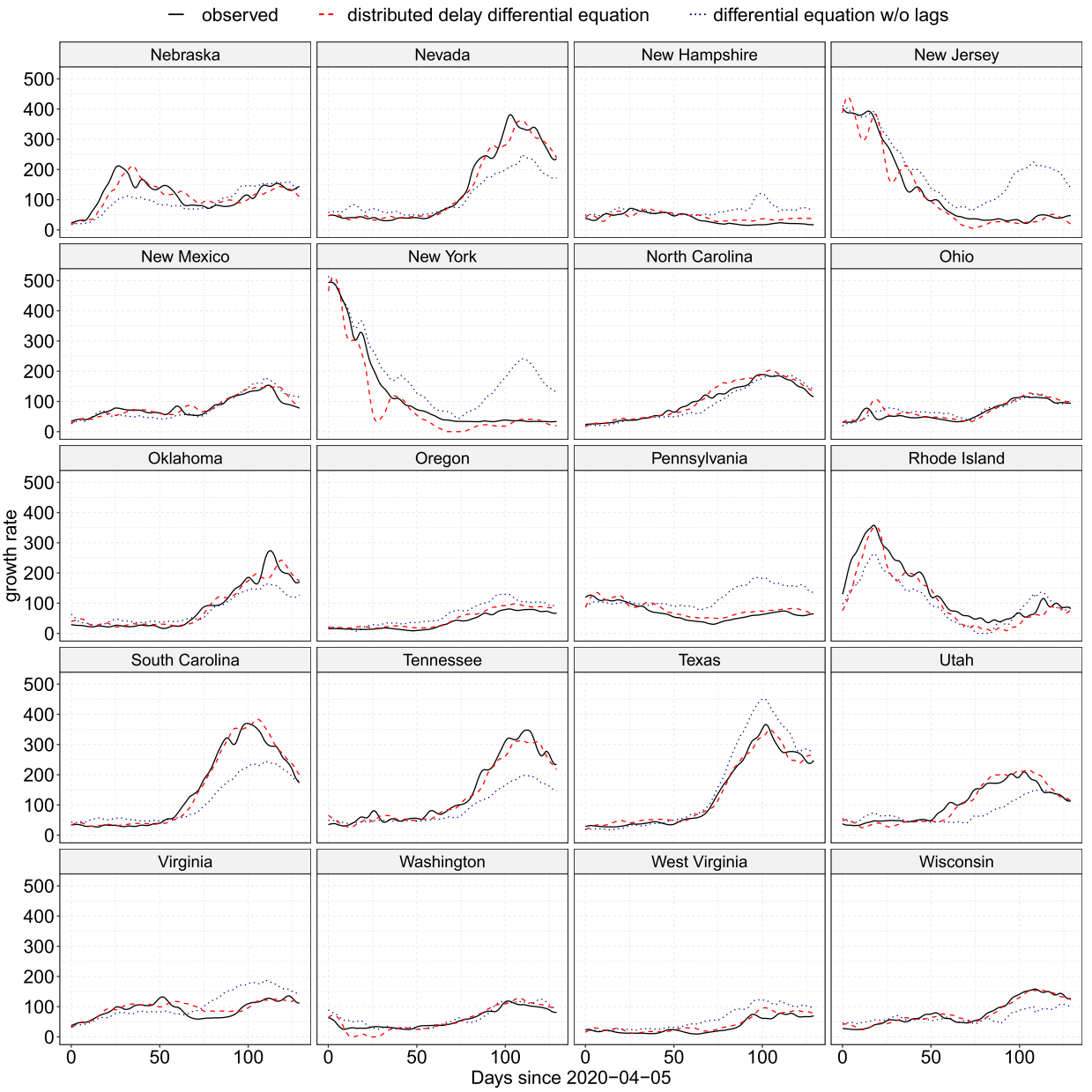


Fig. 4. Predicted growth rates by the proposed RDED (5) (red, dashed) and by a RDE without lags (blue, dotted) for US States, Nebraska–Wisconsin. Observed rates are the black solid curves.

observed rates which fall above the predicted curve. Figs. 5 and 6 display the residual curves for the fitted RDED model (5), where positive residuals correspond to the number of excess new cases per day relative to the predictions obtained from fitting model (5), as described above, including the two-week history index and other covariate processes with their respective lags. These residuals reflect the unexplained residual stochastic process Z that is an integral and principally unpredictable part of the RDED model (5) on which our predictions are based.

Major outbreaks, which can be interpreted as an unforeseen amount of excess cases, are revealed through positive spikes in the residual plots. For example, New York's residuals leap up to as high as 170 excess cases per million per day during a spike in Spring. Arizona's summer outbreak is captured in the residuals as well, subsequently course-correcting with a streak of fewer-than-expected new cases around day 100. These

Please cite this article in press as: P. Dubey et al., Learning delay dynamics for multivariate stochastic processes, with application to the prediction of the growth rate of COVID-19 cases in the United States, J. Math. Anal. Appl. (2022), https://doi.org/10.1016/j.jmaa.2021.125677

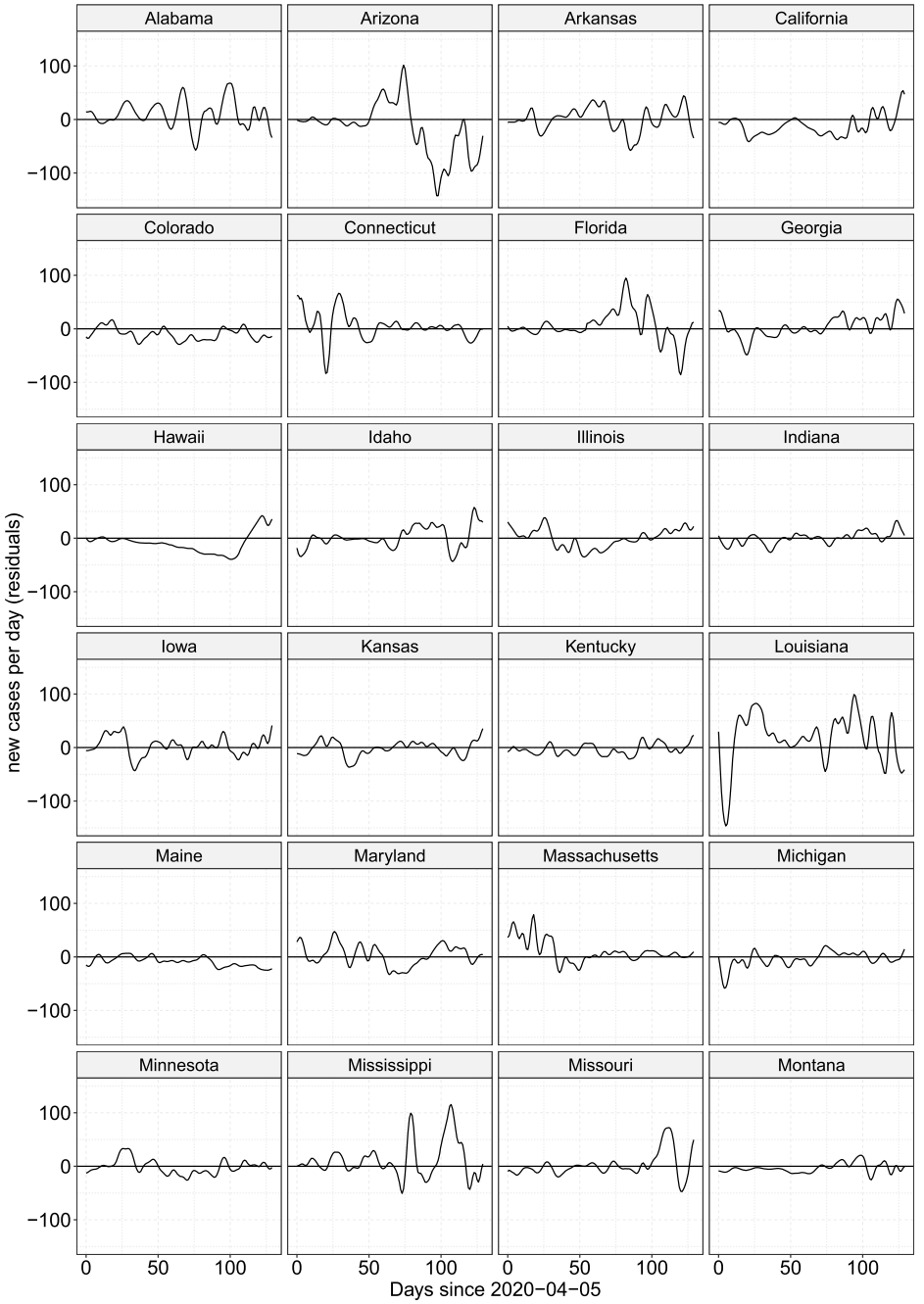


Fig. 5. Residuals for the predictions from the RDED (5) for US States, Alabama–Montana.

spikes also suggest that in the two weeks prior to the uptick, there were substantially more cases in the state population than the numbers reported in the JHU dataset.

The consistency of a state's handling of the virus can be seen through the volatility of the residual curve. States like Colorado and Oregon exhibit only minor deviations from zero, which suggests their new cases are quite predictable given a two-week history of the state case counts and covariate processes. Some states, particularly in the Northeast, including New York, New Jersey, and Rhode Island, only exhibit high volatility during the early days of the pandemic before achieving a period of stability. On the other hand, states such as Alabama and Louisiana exhibit less-predictable case counts throughout.

Please cite this article in press as: P. Dubey et al., Learning delay dynamics for multivariate stochastic processes, with application to the prediction of the growth rate of COVID-19 cases in the United States, J. Math. Anal. Appl. (2022), https://doi.org/10.1016/j.jmaa.2021.125677

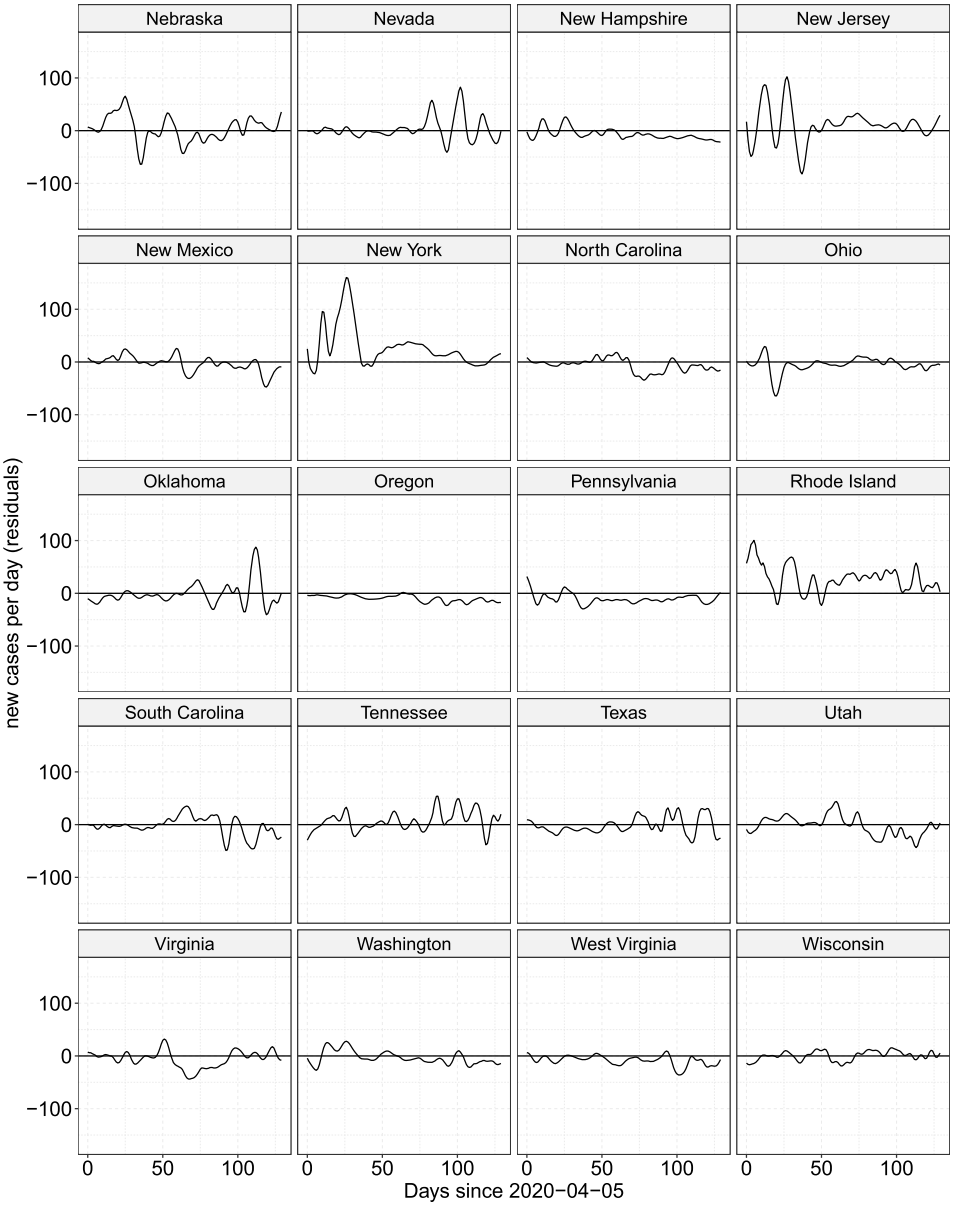


Fig. 6. Residuals for US States, Nebraska–Wisconsin.

5.6. Computation times and code availability

The codes for reproducing the above analysis are available at https://osf.io/w48zd/?view_only=7cd3e2c686a74fd086e4eae8534214c0. The run times of the different steps in the analysis using R version 3.6.3 (2020-02-29) running under CentOS Linux 7 (Core) on x86_64-pc-linux-gnu (64-bit) platform are summarized in Table 3.

6. Discussion and concluding remarks

In this paper we propose dynamics learning exemplified by learning a RDED from a sample of observed trajectories, which is motivated by modeling COVID-19 daily new cases from the previous behavior of the case trajectory itself through a distributed delay or history index component and additional covariate

Please cite this article in press as: P. Dubey et al., Learning delay dynamics for multivariate stochastic processes, with application to the prediction of the growth rate of COVID-19 cases in the United States, J. Math. Anal. Appl. (2022), https://doi.org/10.1016/j.jmaa.2021.125677

Table 3
Run time for the different steps in the data analysis.

Step	Description	Number of Cores	Time in Seconds
1	preprocessing	single core	9.475
2	initial lag selection	8 cores in parallel	422.505
3	history index function estimation	single core	3.414
4	lasso	8 cores in parallel	31.316
5	backfitting	single core	2565.715
6	smoothing	single core	0.012
	Total		3032.437

process components. RDEDs form the driving mechanism of real life processes that include a feedback loop into the future. Depending on the application and the exact nature of the feedback, one might incorporate discrete delays or distributed delays. While discrete delays borrow information from some isolated times in the past, a distributed delay takes into account the entire continuum in the recent past and turns out to be closely related to a functional history index model that has been considered in the statistics literature.

In the current framework, we have not included inference for the estimated model coefficient functions and the lags. It is left for future research to investigate the convergence of these estimated parameters to their population targets and to obtain guarantees on the convergence rates. Some words of caution are in order. Any method that includes a time delay component suffers from the limitation that a larger set of possible lags means one has to sacrifice a part of the data that are available for the analysis, due to the initialization. We have assumed that the trajectories corresponding to the different states are i.i.d. realizations, which may not hold if there is significant spatial correlation based on geographic and demographic similarities across the states. It would be interesting to study adaptations of the proposed methodology when the samples are correlated instead of being independent. From an application point of view, a different and perhaps more refined perspective could potentially be gained by applying the proposed method at a finer level to county level data.

Establishing the existence and uniqueness of the solutions of RDEDs has found increasing interest in recent years [14,24]. We contribute here to the forward problem of obtaining solutions of an RDED by targeting the inverse problem analytically and combine this with functional data analysis approaches to estimate the parameter functions from an observed sample of multivariate stochastic processes for both discrete and distributed delay setups. Our approach builds on and contributes to empirical dynamics and the literature on inferring dynamics and differential equations from functional data, and extends it to models that include time delays.

A key feature of the proposed approach is the statistical estimation of the model parameters in the population RDED, which has inbuilt model selection steps that aid in estimating the optimal lags and also to select the important predictors through the lasso pruning steps, with majority voting across all time points. This reduces the number of predictors in the model and thus its complexity prior to the optimization step of the lags that are associated with the predictors. To handle this complex lag optimization step, instead of a brute force search method, we adopted a backfitting algorithm, which significantly simplifies the optimization step and also leads to faster convergence. We provide a complete toolkit that starts with a sample of multivariate trajectories and extracts from the observed data an estimated RDED that best explains the dynamics inherent in the data, along with automatic variable selection and delay estimation.

The application of our method to the modeling of COVID 19 growth rates in the United States demonstrates excellent performance in terms of predictions of the trajectories. The predictions from the random differential equation model without predictor lags grossly overestimates the COVID-19 growth rates for significant periods of time in New York, New Jersey, Pennsylvania, Massachusetts, Connecticut, Virginia, West Virginia, New Hampshire, California, Texas, Colorado, Oregon, Washington, Maine and Hawaii but underestimates the growth rates for Alabama, Arizona, Arkansas, Florida, Louisiana, Nebraska, Nevada, South Carolina, Tennessee and Wisconsin; see Figs. 3 and 4. This highlights the importance of incorporating

Please cite this article in press as: P. Dubey et al., Learning delay dynamics for multivariate stochastic processes, with application to the prediction of the growth rate of COVID-19 cases in the United States, J. Math. Anal. Appl. (2022), https://doi.org/10.1016/j.jmaa.2021.125677

time lags in differential equation modeling of functional data, when one has a sample of realized stochastic processes, specially when the underlying processes tend to have an aftereffect.

Appendix A. Proof of the existence of a solution for the model in (3)

Proof of continuity of f. The unique solution of the first order DDE in (3) can be shown to exist as a direct extension of Theorem 2.2 of [14].

We first show that $f(y,t)$ is p^{th}-moment continuous, using Assumptions (A0) - (A2) in Section 3.1. Observe that

$$\begin{aligned}
&(E|f(y_m,t_m) - f(y,t)|^p)^{1/p} \\
&= (E|\alpha(t_m) - \alpha(t) \ + \beta_0(t_m)y_m - \beta_0(t)y \ + Z(t_m) - Z(t) \\
&\quad + \sum_{j=1}^{J} \beta_j(t_m)U_j(t_m - \tau_j) - \sum_{j=1}^{J} \beta_j(t)U_j(t-\tau_j)|^p)^{1/p} \\
&\leq |\alpha(t_m) - \alpha(t)| \ + ||\beta_0(t_m)y_m - \beta_0(t)y||_p + ||Z(t_m) - Z(t)||_p \\
&\quad + \sum_{j=1}^{J} ||\beta_j(t_m)U_j(t_m - \tau_j) - \beta_j(t)U_j(t-\tau_j)||_p. \qquad \text{(A.1)}
\end{aligned}$$

The first and third terms in (A.1) converge to 0 using Assumptions (A0) and (A1) respectively. For the second and fourth terms of (A.1),

$$\begin{aligned}
&||\beta_0(t_m)y_m - \beta_0(t)y||_p \\
&\leq ||\beta_0(t_m)\ y_m - \beta_0(t_m)\ y||_p \ + ||\beta_0(t_m)\ y - \beta_0(t)\ y||_p \\
&\leq |\beta_0(t_m)|\ ||y_m - y||_p \ + |\beta_0(t_m) - \beta_0(t)|\ ||y||_p. \qquad \text{(A.2)}
\end{aligned}$$

By Assumption (A0), $\beta_0(t_m)$ is continuous on the compact domain $\mathcal{T}$, and hence uniformly bounded. Also, by assumption $y_m \to y$, in the p^{th}-moment. Since $\mathcal{L}_p$, the space of random variables with finite p^{th} moments, is a Banach space, it follows that $||y||_p := (E|y|^p)^{1/p} < \infty$. Thus (A.2) converges to 0. Next

$$\begin{aligned}
&\sum_{j=1}^{J} ||\beta_j(t_m)U_j(t_m - \tau_j) - \beta_j(t)U_j(t-\tau_j)||_p \\
&\leq \sum_{j=1}^{J} ||\beta_j(t_m)\ (U_j(t_m - \tau_j) - U_j(t-\tau_j)) + (\beta_j(t_m) - \beta_j(t))\ U_j(t-\tau_j)||_p \\
&\leq \sum_{j=1}^{J} |\beta_j(t_m)|\ ||U_j(t_m - \tau_j) - U_j(t-\tau_j)||_p + |\beta_j(t_m) - \beta_j(t)|\ ||U_j(t-\tau_j)||_p, \qquad \text{(A.3)}
\end{aligned}$$

and (A.3) converges to 0 using the assumptions on the continuity of the coefficient functions $\beta_j(\cdot)$ on the compact domain $\mathcal{T}$. This implies the uniform boundedness of β_j and the continuity of the predictor processes $U_j(\cdot) \in \mathcal{L}_p$ in the p^{th} moment for $j = 1, \ldots, J$. Hence the proof of continuity of $f(y,t)$ follows immediately by combining (A.1), (A.2), and (A.3). □

Proof of Proposition 1. Let $k(t) = \beta_0(t),\ t \in \mathcal{T}$. Since, by Assumption (A0), $\beta_0(\cdot)$ is continuous on a compact domain, we have $\int_{t_0}^{T} |k(t)|dt < \infty$. For $y, v \in \mathcal{L}_p$ and $t \in \mathcal{T}$, we obtain

Please cite this article in press as: P. Dubey et al., Learning delay dynamics for multivariate stochastic processes, with application to the prediction of the growth rate of COVID-19 cases in the United States, J. Math. Anal. Appl. (2022), https://doi.org/10.1016/j.jmaa.2021.125677

$$||f(y,t) - f(v,t)||_p = ||\beta_0(t)\ (y-v)||_p \leq |\beta_0(t)|||y-v||_p. \tag{A.4}$$

Since f is p^{th}-moment continuous, the proof follows using the same line of arguments as in [14]. □

Appendix B. Proof of the existence of a solution for the model in (4)

For ease of presentation, we will begin by showing the existence of the unique solution of the first order RDED (4) with one predictor process $U(\cdot)$, rewritten as in (11). The arguments can then be easily extended to the case of multiple predictor processes, and we omit the details.

Proof of Proposition 2. Recall that $t_m \in [t_0, b]$ is a real sequence converging to $t \in [t_0, b]$ and Y is an $\mathcal{L}_p$-continuous process, so that

$$\begin{aligned}
&||\tilde{f}(Y,t_m) - \tilde{f}(Y,t)||_p \\
=&\left\|\alpha(t_m) - \alpha(t) + \int_{t_m-\tau_0}^{t_m} \gamma(t_m - s, t_m)Y(s)ds - \int_{t-\tau_0}^{t} \gamma(t-s,t)Y(s)ds \right. \\
&\left. + \int_{t_m-\tau_1}^{t_m} \gamma_1(t_m - s, t_m)U(s)ds - \int_{t-\tau_1}^{t} \gamma_1(t-s)U(s)ds + Z(t_m) - Z(t)\right\|_p \\
\leq&|\alpha(t_m) - \alpha(t)| + ||Z(t_m) - Z(t)||_p \\
&+ \left\| \int_{t_m-\tau_0}^{t_m} \gamma(t_m - s, t_m)Y(s)ds - \int_{t-\tau_0}^{t} \gamma(t-s,t)Y(s)ds \right\|_p \\
&+ \left\| \int_{t_m-\tau_1}^{t_m} \gamma_1(t_m - s, t_m)U(s)ds - \int_{t-\tau_1}^{t} \gamma_1(t-s,t)U(s)ds \right\|_p .
\end{aligned} \tag{B.1}$$

Using Assumptions (A0) and (A1), respectively, the first and second term of (B.1) converge to 0. As before, we consider the third and fourth terms separately,

$$\begin{aligned}
&\left\| \int_{t_m-\tau_0}^{t_m} \gamma(t_m - s, t_m)Y(s)ds - \int_{t-\tau_0}^{t} \gamma(t-s,t)Y(s)ds \right\|_p \\
\leq&\left\| \int_{t_m-\tau_0}^{t_m} \gamma(t_m - s, t_m)Y(s)ds - \int_{t_m-\tau_0}^{t_m} \gamma(t-s,t)Y(s)ds \right\|_p \\
&+ \left\| \int_{t_m-\tau_0}^{t_m} \gamma(t-s,t)Y(s)ds - \int_{t-\tau_0}^{t} \gamma(t-s,t)Y(s)ds \right\|_p \\
\leq&\left\| \int_{t_m-\tau_0}^{t_m} (\gamma(t_m - s, t_m) - \gamma(t-s,t))Y(s)ds \right\|_p \\
&+ \left\| \int_{t_m-\tau_0}^{t_m} \gamma(t-s,t)Y(s)ds - \int_{t-\tau_0}^{t} \gamma(t-s,t)Y(s)ds \right\|_p .
\end{aligned} \tag{B.2}$$

Please cite this article in press as: P. Dubey et al., Learning delay dynamics for multivariate stochastic processes, with application to the prediction of the growth rate of COVID-19 cases in the United States, J. Math. Anal. Appl. (2022), https://doi.org/10.1016/j.jmaa.2021.125677

For the first term in (B.2),

$$\begin{aligned}
&\left\| \int_{t_m-\tau_0}^{t_m} (\gamma(t_m-s,t_m)-\gamma(t-s,t))Y(s)\ ds \right\|_p \\
&\leq \int_{t_m-\tau_0}^{t_m} \left\|(\gamma(t_m-s,t_m)-\gamma(t-s,t))Y(s)\right\|_p\ ds \\
&\leq \int_{t_0-\tau_0}^{b} \left\|(\gamma(t_m-s,t_m)-\gamma(t-s,t))Y(s)\right\|_p\ ds \\
&= \int_{t_0-\tau_0}^{b} \left|(\gamma(t_m-s,t_m)-\gamma(t-s,t))\right| \left\|Y(s)\right\|_p\ ds, \qquad \text{(B.3)}
\end{aligned}$$

where the first inequality follows from the fact that the integrand is p^{th} moment continuous, using Assumption (B0) and the assumption that $Y(\cdot) \in \mathcal{L}_p$ (see [54] page 102(3)). The second inequality follows using the properties of $\mathcal{L}_p$ Riemann integration. Now, from Assumption (B0), we have that for any $\epsilon > 0$ there exists $M_0(\epsilon)$ such that $|(\gamma(t_m-s,t_m)-\gamma(t-s,t))| < \varepsilon$ whenever $m \geq M_0(\epsilon)$. Thus, from (B.3) it follows that

$$\int_{t_0-\tau_0}^{b} ||(\gamma(t_m-s,t_m)-\gamma(t-s,t))Y(s)||_p\ ds \leq \int_{t_0-\tau_0}^{b} \epsilon||Y(s)||_p\ ds. \qquad \text{(B.4)}$$

Since $||Y(s)||_p < \infty$ and $[t_0-\tau_0, b]$ is compact, to show that $\int_{t_0-\tau_0}^{b} ||Y(s)||_p ds < \infty$ it is sufficient to prove that the application $s \to ||Y(s)||_p$ is continuous on $[t_0-\tau_0, b]$. For this, let $s_m \to s$ as $m \to \infty$. By Assumption (B1) we have

$$|||Y(s_m)||_p - ||Y(s)||_p| \leq ||Y(s_m)-Y(s)||_p \to 0 \text{ as } m \to \infty.$$

Hence

$$|| \int_{t_m-\tau_0}^{t_m} (\gamma(t_m-s,t_m)-\gamma(t-s,t))Y(s)ds||_p \to 0, \text{ as } m \to 0. \qquad \text{(B.5)}$$

For the function $\Psi(r) := \int_{r-\tau_0}^{r} \gamma(t-s,t)Y(s)ds$, $r \in [t_0, b]$,

$$\begin{aligned}
\Psi(r) &= \int_{r-\tau_0}^{r} \gamma(t-s,t)Y(s)ds \\
&= \int_{t_0-\tau_0}^{r} \gamma(t-s,t)Y(s)ds - \int_{t_0-\tau_0}^{r-\tau_0} \gamma(t-s,t)Y(s)ds \\
&= \int_{t_0-\tau_0}^{r} \gamma(t-s,t)Y(s)ds - \int_{t_0}^{r} \gamma(t-u+\tau_0,t)Y(u-\tau_0)du.
\end{aligned}$$

Please cite this article in press as: P. Dubey et al., Learning delay dynamics for multivariate stochastic processes, with application to the prediction of the growth rate of COVID-19 cases in the United States, J. Math. Anal. Appl. (2022), https://doi.org/10.1016/j.jmaa.2021.125677

From [54] page 103(5), each term in the RHS of the above equation is p^{th}-moment continuous. Hence, by definition, for $t_m \to t$, $||\Psi(t_m) - \Psi(t)||_p = || \int_{t_m-\tau_0}^{t_m} \gamma(t-s,t)Y(s)ds - \int_{t-\tau_0}^{t} \gamma(t-s,t)Y(s)ds||_p \to 0$, and the second term in (B.2) also converges to zero as $m \to \infty$. As for the convergence of the remaining fourth term, involving the predictor process $U(\cdot)$ in (B.1), we can follow a similar line of argument to obtain

$$\begin{aligned}
&\left\| \int_{t_m-\tau_1}^{t_m} \gamma_1(t_m - s, t_m)U(s)ds - \int_{t-\tau_1}^{t} \gamma_1(t-s,t)U(s)ds \right\|_p \\
&\leq \left\| \int_{t_m-\tau_1}^{t_m} (\gamma_1(t_m - s, t_m) - \gamma_1(t-s,t))U(s)ds \right\|_p \\
&\quad + \left\| \int_{t_m-\tau_1}^{t_m} \gamma_1(t-s,t)U(s)ds - \int_{t-\tau_1}^{t} \gamma_1(t-s,t)U(s)ds \right\|_p \\
&\quad \to 0, \text{ using Assumptions (A2) and (B0).}
\end{aligned} \tag{B.6}$$

Thus, $(E|\tilde{f}(Y,t_m) - \tilde{f}(Y,t)|^p)^{1/p} \to 0$, for a sequence $t_m \to t$ as $m \to \infty$, implying the p^{th}-moment continuity of $\tilde{f}(Y,t)$ for all $t \in I$. The p^{th}-moment continuity of the process $\tilde{f}(Y,\cdot)$ follows by definition. □

As a consequence of the above result, we observe that $\tilde{f}(X,t)$ in (11) is continuous in $t \in \mathcal{T} = [t_0, T]$ in the $\mathcal{L}_p$ sense, by simply choosing $b = T$.

Proof of Proposition 3. Suppose X is an $\mathcal{L}_p$ solution of (11). Then by definition, (a) and (c) hold. Since X is p^{th}-moment continuous (see Assumption (B1)) and the process $\tilde{f}(X,\cdot)$ is $\mathcal{L}_p$-continuous on $\mathcal{T} = [t_0, T]$, it follows from [54] page 101(1) that X' is p^{th}-moment continuous and p^{th}-moment Riemann integrable on $\mathcal{T}$. Thus from the Fundamental Theorem of $\mathcal{L}_p$ Calculus ([54] page 104(6))

$$X(t) = g(t_0) + (\mathcal{L}_p) \int_{t_0}^{t} X'(s)ds = g(t_0) + \int_{t_0}^{t} \tilde{f}(X,s)ds,$$

and condition (b) holds. On the other hand, if conditions (a)–(c) are true, X is indeed an $\mathcal{L}_p$ solution of (11), since the process $\tilde{f}(X,\cdot)$ was shown to be continuous. Thus $\tilde{f}(X,\cdot)$ is p^{th}-moment Riemann integrable on $\mathcal{T}$ ([54] page 101 (1)), and $\int_{t_0}^{t} \tilde{f}(X,s)ds$ is well defined in the $\mathcal{L}_p$ sense. From the result on page 103 (5) of [54], it follows that X is p^{th}-moment differentiable on $\mathcal{T}$, with $X'(t) = \tilde{f}(X,t)$. Hence the proposition follows. □

Proof of Theorem 1. From the history index model in (11)

$$\begin{aligned}
X'(t) = \tilde{f}(X,t) = \alpha(t) &+ \int_{t-\tau_0}^{t} \gamma(t-s,t)X(s)ds \\
&+ \int_{t-\tau_1}^{t} \gamma_1(t-s,t)U(s)ds + Z(t).
\end{aligned}$$

For $k(t) = \int_{t-\tau_0}^{t} |\gamma(t-s,t)|ds$, since γ is continuous on a compact domain, hence uniformly bounded, we have $k \in L^1(\mathcal{T})$. Following the arguments outlined in the proof of Theorem 2.2 in [14], this implies

Please cite this article in press as: P. Dubey et al., Learning delay dynamics for multivariate stochastic processes, with application to the prediction of the growth rate of COVID-19 cases in the United States, J. Math. Anal. Appl. (2022), https://doi.org/10.1016/j.jmaa.2021.125677

$\lim_{t\to t_0^+}\int_{t_0}^{t} k(s)ds = 0$. Thus, we can choose $\alpha > t_0$ such that, for all $t \in \mathcal{T}_\alpha = [t_0, \alpha] \subset \mathcal{T}$, $\int_{t_0}^{t} k(s)ds \le 1/2$. Consider the vector space

$$\mathcal{A} = \{X : [t_0 - \tau_0, \alpha] \to \mathcal{L}^p \text{ continuous}, X(t) = g(t) \text{ on } [t_0 - \tau_0, t_0]\},$$

with norm $\|X\|_{\mathcal{A}} = \sup_{t\in[t_0-\tau_0,\alpha]} \|X(t)\|_p$. We note that $\|X\|_{\mathcal{A}}$ is well-defined because, by the p^{th}-moment continuity of X, the real map $t \to \|X(t)\|_p,\ t \in [t_0-\tau_0, \alpha]$ is continuous. Therefore $\sup_{t\in[t_0-\tau_0,\alpha]} \|X(t)\|_p < \infty$.

As shown in the proof of Theorem 2.2 of [14], $\mathcal{A}$ is a Banach space. Consider the map $\Lambda : \mathcal{A} \to \mathcal{A}$ such that

$$\Lambda(X)(t) = \begin{cases} g(t_0) + (\mathcal{L}_p)\int_{t_0}^{t} \tilde{f}(X, s)ds, & t \in \mathcal{T}_\alpha = [t_0, \alpha], \\ g(t), & t \in [t_0 - \tau_0, t_0]. \end{cases}$$

If $X \in \mathcal{A}$, it follows by taking $b = \alpha$ in Proposition 2 that the process $\tilde{f}(X, t)$ is p^{th}-moment continuous over $t \in \mathcal{T}_\alpha$. From arguments on p. 103 (5) of [54], the p^{th}-moment Riemann integral $(\mathcal{L}_p)\int_{t_0}^{t} \tilde{f}(X, s)ds$ is continuous. Also, the initial condition $g(\cdot)$ is continuous on $[t_0 - \tau_0, t_0]$. Thus $\Lambda X : [t_0 - \tau_0, \alpha] \to \mathcal{L}_p$ is continuous on $[t_0 - \tau_0, \alpha]$ satisfying $\Lambda X(t) = g(t)$ on $[t_0 - \tau_0, t_0]$. By definition $\Lambda X \in \mathcal{A}$, so Λ is well defined.

Since from Proposition 3, $X : \mathcal{T}_\alpha \to \mathcal{L}_p$ is a solution of equation (11) if and only if $X \in \mathcal{A}$ and $\Lambda X = X$, by the Banach fixed-point theorem, it suffices to check that Λ is a contraction. Let $X, Y \in \mathcal{A}$ and $t \in \mathcal{T}_\alpha$ and observe that if $t \in [t_0 - \tau_0, t_0]$ then $\Lambda X(t) - \Lambda Y(t) = g(t) - g(t) = 0$). From arguments on p. 102 (3) of [54]

$$\begin{aligned}
\|\Lambda X(t) - \Lambda Y(t)\|_p &= \left\| (\mathcal{L}_p)\int_{t_0}^{t} (\tilde{f}(X, s) - \tilde{f}(Y, s))ds \right\|_p \\
&\le \int_{t_0}^{t} \|\tilde{f}(X, s) - \tilde{f}(Y, s)\|_p ds \\
&= \int_{t_0}^{t} \left\| \int_{s-\tau_0}^{s} \gamma(s - u, s)(X(u) - Y(u))du \right\|_p ds \\
&\le \int_{t_0}^{t} \int_{s-\tau_0}^{s} |\gamma(s - u, s)| \|X(u) - Y(u)\|_p du ds \\
&\le \int_{t_0}^{t} \int_{s-\tau_0}^{s} |\gamma(s - u, s)| du ds \|X - Y\|_{\mathcal{A}} \\
&= \int_{t_0}^{t} k(s)ds \|X - Y\|_{\mathcal{A}} \\
&\le \frac{1}{2}\|X - Y\|_{\mathcal{A}}. \qquad \text{(B.7)}
\end{aligned}$$

Taking the supremum on $t \in [t_0 - \tau_0, \alpha]$, $\|\Lambda X - \Lambda Y\|_{\mathcal{A}} \le \frac{1}{2}\|X - Y\|_{\mathcal{A}}$. Therefore Λ is a contraction. The remainder of the proof follows similar arguments as in the proof of Theorem 2.2 of [14]. □

Please cite this article in press as: P. Dubey et al., Learning delay dynamics for multivariate stochastic processes, with application to the prediction of the growth rate of COVID-19 cases in the United States, J. Math. Anal. Appl. (2022), https://doi.org/10.1016/j.jmaa.2021.125677

Proof of Corollary 1. Similar to (10), define $\tilde{h} : \mathcal{L}_p \times I_b \to \mathbb{R}$ such that for an $\mathcal{L}_p$-continuous function $Y(\cdot) : [t_0 - \tau_0, b] \to \mathbb{R}$ and $t \in I_b := [t_0, b]$, $t_0 < b$,

$$\tilde{h}(Y,t) = \alpha(t) + \int_{t-\tau_0}^{t} \gamma(t-s,t)Y(s)ds + \sum_{j=1}^{J} \beta_j(t)U_j(t-\tau_j) + Z(t).$$

Here $\tilde{h}$ is a random functional with first argument given by trajectories $\{Y(s) : s \in [t_0 - \tau_0, b]\}$ and second argument $t \in [t_0, b]$. Note that $\tilde{h}$ constitutes a distributional delay term on Y and a discrete concurrent delay on the predictors U_j. Thus

$$\begin{aligned}
&||\tilde{h}(Y,t_m) - \tilde{h}(Y,t)||_p \\
\leq &|\alpha(t_m) - \alpha(t)| + ||Z(t_m) - Z(t)||_p \\
&+ \left\| \int_{t_m-\tau_0}^{t_m} \gamma(t_m - s, t_m)Y(s)ds - \int_{t-\tau_0}^{t} \gamma(t-s,t)Y(s)ds \right\|_p \\
&+ \left\| \sum_{j=1}^{J} \beta_j(t_m)U_j(t_m - \tau_j) - \beta_j(t)U_j(t-\tau_j) \right\|_p .
\end{aligned}$$

By similar arguments as in the proof of the continuity of f in Appendix A and the proof of Proposition 2 in Appendix B, the continuity of $\tilde{h}$ in the $\mathcal{L}_p$ sense in t follows. The existence and uniqueness follows by the same arguments as in the proof of Theorem 1. □

Appendix C. Derivative estimation

For derivative estimation, we employ the local polynomial estimator for derivatives, motivated by local polynomial approximation [28,43], which leads to the weighted least squares estimates $\hat{\theta}_l$ that correspond to solving

$$\underset{\theta_l \in \mathbb{R}}{\operatorname{argmin}} \sum_{k=1}^{K} \left\{ Y_k - \sum_{l=0}^{L} \theta_l (t_k - t)^l \right\}^2 K_h(t_k - t), \tag{C.1}$$

where K is a kernel function with $K_h(\cdot) = K(\cdot/h)/h$ and h is a tuning parameter. A well-studied estimator for the ν-th order derivative $x^{(\nu)}(t)$ is given by

$$\hat{x}^{(\nu)}(t) = \nu!\hat{\theta}_\nu, \tag{C.2}$$

for $\nu = 0, 1, \ldots, L$. The whole curve $\hat{x}^{(\nu)}(\cdot)$ is obtained by running the above local polynomial regression with t varying in an appropriate estimation domain.

For local polynomial fitting $L - \nu$ preferably is taken to be odd as shown in Ruppert & Wand [50] and Fan & Gijbels [28]. To obtain an estimate of the first derivative, i.e., for $\nu = 1$, the choice $L = \nu + 1 = 2$ leads to the so-called local quadratic regression and the derivative estimate $\hat{x}'(t)$ is given by the local slope $\hat{\theta}_1$. Other common methods for derivative estimation can be based on smoothing splines or B-splines [49,64]. An alternative method is based on difference quotients, which provides a straightforward approach for pointwise estimation of derivatives. Difference quotient-based estimators have been thoroughly studied in the context of human growth curves in the nonparametric regression literature [30,43,44].

Please cite this article in press as: P. Dubey et al., Learning delay dynamics for multivariate stochastic processes, with application to the prediction of the growth rate of COVID-19 cases in the United States, J. Math. Anal. Appl. (2022), https://doi.org/10.1016/j.jmaa.2021.125677

References

[1] M. Ablowitz, J. Ladik, Nonlinear differential-difference equations, J. Math. Phys. 16 (1975) 598–603.
[2] L. Arnold, Stochastic Differential Equations, Wiley, New York, 1974.
[3] F.M. Asl, A.G. Ulsoy, Analysis of a system of linear delay differential equations, J. Dyn. Syst. Meas. Control 125 (2003) 215–223.
[4] I. Aziz, R. Amin, Numerical solution of a class of delay differential and delay partial differential equations via Haar wavelet, Appl. Math. Model. 40 (2016) 10286–10299.
[5] A. Bellen, M. Zennaro, Numerical Methods for Delay Differential Equations, Oxford University Press, 2013.
[6] E. Beretta, T. Hara, W. Ma, Y. Takeuchi, Global asymptotic stability of an SIR epidemic model with distributed time delay, Nonlinear Anal., Theory Methods Appl. 47 (2001) 4107–4115.
[7] A.L. Bertozzi, E. Franco, G. Mohler, M.B. Short, D. Sledge, The challenges of modeling and forecasting the spread of COVID-19, Proc. Natl. Acad. Sci. 117 (2020) 16732–16738.
[8] S. Bjorklund, L. Ljung, A review of time-delay estimation techniques, in: 42nd IEEE International Conference on Decision and Control (IEEE Cat. No. 03CH37475), vol. 3, IEEE, 2003, pp. 2502–2507.
[9] P. Breheny, ncvreg: regularization paths for SCAD and MCP penalized regression models, R package version 3.12.0, available at, https://CRAN.R-project.org/package=ncvreg, 2020.
[10] P. Breheny, J. Huang, Coordinate descent algorithms for nonconvex penalized regression, with applications to biological feature selection, Ann. Appl. Stat. 5 (2011) 232–253.
[11] T.S. Brett, P. Rohani, Transmission dynamics reveal the impracticality of COVID-19 herd immunity strategies, Proc. Natl. Acad. Sci. 117 (2020) 25897–25903.
[12] N.J.-B. Brunel, Parameter estimation of ODE's via nonparametric estimators, Electron. J. Stat. 2 (2008) 1242–1267.
[13] Z. Cai, J. Fan, R. Li, Efficient estimation and inferences for varying-coefficient models, J. Am. Stat. Assoc. 95 (2000) 888–902.
[14] J. Calatayud, J.-C. Cortés, M. Jornet, Random differential equations with discrete delay, Stoch. Anal. Appl. 37 (2019) 699–707.
[15] T. Caraballo, R. Colucci, L. Guerrini, On a predator prey model with nonlinear harvesting and distributed delay, Commun. Pure Appl. Anal. 17 (2018) 2703.
[16] T. Caraballo, J.-C. Cortés, A. Navarro-Quiles, Applying the random variable transformation method to solve a class of random linear differential equation with discrete delay, Appl. Math. Comput. 356 (2019) 198–218.
[17] H. Cardot, F. Ferraty, P. Sarda, Functional linear model, Stat. Probab. Lett. 45 (1999) 11–22.
[18] C. Carroll, S. Bhattacharjee, Y. Chen, P. Dubey, J. Fan, Á. Gajardo, X. Zhou, H.-G. Müller, J.-L. Wang, Time dynamics of COVID-19, Sci. Rep. 10 (2020) 1–14.
[19] C. Carroll, Á. 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 0.5.5, available at, https://CRAN.R-project.org/package=fdapace, 2020.
[20] J. Chen, H. Wu, Efficient local estimation for time-varying coefficients in deterministic dynamic models with applications to HIV-1 dynamics, J. Am. Stat. Assoc. 103 (2008) 369–384.
[21] R. Chetty, J. Friedman, N. Hendren, M. Stepner, The Opportunity Insights Team, The economic impacts of COVID-19: evidence from a new public database built using private sector data, Nat. Bureau Econ. Res. (2020) (Working Paper No. 27431).
[22] E.A. Coddington, N. Levinson, Theory of Ordinary Differential Equations, Tata McGraw-Hill Education, 1955.
[23] J.C. Cortés, L. Jódar, L. Villafuerte, Numerical solution of random differential equations: a mean square approach, Math. Comput. Model. 45 (2007) 757–765.
[24] J.C. Cortés, M. Jornet, L_p-solution to the random linear delay differential equation with a stochastic forcing term, Mathematics 8 (2020) 1013.
[25] D. Şentürk, H.-G. Müller, Generalized varying coefficient models for longitudinal data, Biometrika 95 (2008) 653–666.
[26] R.D. Driver, Ordinary and Delay Differential Equations, Springer Science & Business Media, 2012.
[27] A. Elnaggar, G.A. Dumont, A.-L. Elshafei, Recursive estimation for system of unknown delay, in: Proceedings of the 28th IEEE Conference on Decision and Control, IEEE, 1989, pp. 1809–1810.
[28] J. Fan, I. Gijbels, Local Polynomial Modelling and Its Applications: Monographs on Statistics and Applied Probability, vol. 66, CRC Press, 1996.
[29] J. Fan, J.-T. Zhang, Two-step estimation of functional linear models with applications to longitudinal data, J. R. Stat. Soc. B 62 (2000) 303–322.
[30] T. Gasser, W. Köhler, H. Müller, A. Kneip, R. Largo, L. Molinari, A. Prader, Velocity and acceleration of height growth using kernel estimation, Ann. Human Biol. 11 (1984) 397–411.
[31] W. Gurney, S. Blythe, R. Nisbet, Nicholson's blowflies revisited, Nature 287 (1980) 17–21.
[32] S.-T. Hao, J.-L. Wang, Dynamic modeling of multivariate longitudinal data, UC Davis, preprint, 2021.
[33] J.Z. Huang, C.O. Wu, L. Zhou, Polynomial spline estimation and inference for varying coefficient models with longitudinal data, Stat. Sin. 14 (2004) 763–788.
[34] K. Itô, On a formula concerning stochastic differentials, Nagoya Math. J. 3 (1951) 55–65.
[35] G. Jacovitti, G. Scarano, Discrete time techniques for time delay estimation, IEEE Trans. Signal Process. 41 (1993) 525–533.
[36] A. Jarne, D. Commenges, L. Villain, M. Prague, Y. Lévy, R. Thiébaut, Modeling CD4+ T cells dynamics in HIV-infected patients receiving repeated cycles of exogenous Interleukin 7, Ann. Appl. Stat. 11 (2017) 1593–1616.
[37] Z. Li, M.R. Osborne, T. Prvan, Parameter estimation of ordinary differential equations, IMA J. Numer. Anal. 25 (2005) 264–285.

Please cite this article in press as: P. Dubey et al., Learning delay dynamics for multivariate stochastic processes, with application to the prediction of the growth rate of COVID-19 cases in the United States, J. Math. Anal. Appl. (2022), https://doi.org/10.1016/j.jmaa.2021.125677

[38] H. Liang, H. Wu, Parameter estimation for differential equation models using a framework of measurement error in regression models, J. Am. Stat. Assoc. 103 (2008) 1570–1583.
[39] C. Loader, locfit: local regression, likelihood and density estimation. R package version 1.5-9.4, available at: https://CRAN.R-project.org/package=locfit, 2020.
[40] N. Malfait, J.O. Ramsay, The historical functional linear model, Can. J. Stat. 31 (2003) 115–128.
[41] S. Mehrkanoon, S. Mehrkanoon, J.A. Suykens, Parameter estimation of delay differential equations: an integration-free LS-SVM approach, Commun. Nonlinear Sci. Numer. Simul. 19 (2014) 830–841.
[42] J.S. Morris, Functional regression, Annu. Rev. Stat. Appl. 2 (2015) 321–359.
[43] H.-G. Müller, Weighted local regression and kernel methods for nonparametric curve fitting, J. Am. Stat. Assoc. 82 (1987) 231–238.
[44] H.-G. Müller, U. Stadtmüller, T. Schmitt, Bandwidth choice and confidence intervals for derivatives of noisy data, Biometrika 74 (1987) 743–749.
[45] H.-G. Müller, F. Yao, Empirical dynamics for longitudinal data, Ann. Stat. 38 (2010) 3458–3486.
[46] T. Neckel, F. Rupp, Random Differential Equations in Scientific Computing, Walter de Gruyter, 2013.
[47] D. Paul, J. Peng, P. Burman, Semiparametric modeling of autonomous nonlinear dynamical systems with application to plant growth, Ann. Appl. Stat. 5 (2011) 2078–2108.
[48] J. Ramsay, B. Silverman, Functional Data Analysis, Springer-Verlag, New York, 2005.
[49] J. Rice, M. Rosenblatt, Smoothing splines: regression, derivatives and deconvolution, Ann. Stat. 11 (1983) 141–156.
[50] D. Ruppert, M.P. Wand, Multivariate locally weighted least squares regression, Ann. Stat. 22 (1994) 1346–1370.
[51] D. Şentürk, H.-G. Müller, Generalized varying coefficient models for longitudinal data, Biometrika 95 (2008) 653–666.
[52] D. Şentürk, H.-G. Müller, Functional varying coefficient models for longitudinal data, J. Am. Stat. Assoc. 105 (2010) 1256–1264.
[53] D. Şentürk, D.V. Nguyen, Varying coefficient models for sparse noise-contaminated longitudinal data, Stat. Sin. 21 (2011) 1831–1856.
[54] T.T. Soong, Random Differential Equations in Science and Engineering, Academic Press, New York, 1973.
[55] J. Strand, Random ordinary differential equations, J. Differ. Equ. 7 (1970) 538–553.
[56] R. Tibshirani, Regression shrinkage and selection via the lasso, J. R. Stat. Soc., Ser. B 58 (1996) 267–288.
[57] N. Verzelen, W. Tao, H.-G. Müller, Inferring stochastic dynamics from functional data, Biometrika 99 (2012) 533–550.
[58] J.-L. Wang, J.-M. Chiou, H.-G. Müller, Functional data analysis, Annu. Rev. Stat. Appl. 3 (2016) 257–295.
[59] L. Wang, J. Cao, Estimating parameters in delay differential equation models, J. Agric. Biol. Environ. Stat. 17 (2012) 68–83.
[60] S. Wang, L. Wang, Adaptive semiparametric Bayesian differential equations via sequential Monte Carlo, arXiv preprint: arXiv:2002.02571, 2020.
[61] L. Weiss, On the controllability of delay-differential systems, SIAM J. Control 5 (1967) 575–587.
[62] F. Yao, H.-G. Müller, Functional quadratic regression, Biometrika 97 (2010) 49–64.
[63] F. Yao, H.-G. Müller, J.-L. Wang, Functional linear regression analysis for longitudinal data, Ann. Stat. 33 (2005) 2873–2903.
[64] C. Zhou, D.A. Wolfe, On derivative estimation in spline regression, Stat. Sin. 10 (2000) 93–108.
[65] Z. Zhou, Statistical Inference of Distributed Delay Differential Equations, Ph.D. thesis, University of Iowa, 2016.
[66] B. Zhu, J.M. Taylor, P.X.-K. Song, Semiparametric stochastic modeling of the rate function in longitudinal studies, J. Am. Stat. Assoc. 106 (2011) 1485–1495.

Please cite this article in press as: P. Dubey et al., Learning delay dynamics for multivariate stochastic processes, with application to the prediction of the growth rate of COVID-19 cases in the United States, J. Math. Anal. Appl. (2022), https://doi.org/10.1016/j.jmaa.2021.125677