Biometrics 67, 1295–1304
December 2011
DOI: 10.1111/j.1541-0420.2011.01591.x

Stochastic Functional Data Analysis: A Diffusion Model-Based Approach

Bin Zhu,[1,2,*] Peter X.-K. Song,[3,] and Jeremy M.G. Taylor[3,***]**

[1]Department of Statistical Science, Duke University, Durham, North Carolina 27708, U.S.A.
[2]Center for Human Genetics, Duke University Medical Center, Durham, North Carolina 27710, U.S.A.
[3]Department of Biostatistics, University of Michigan, Ann Arbor, Michigan 48109, U.S.A.
[*]*email:* bin.zhu@duke.edu
[**]*email:* pxsong@umich.edu
[***]*email:* jmgt@umich.edu

Summary. This article presents a new modeling strategy in functional data analysis. We consider the problem of estimating an unknown smooth function given functional data with noise. The unknown function is treated as the realization of a stochastic process, which is incorporated into a diffusion model. The method of smoothing spline estimation is connected to a special case of this approach. The resulting models offer great flexibility to capture the dynamic features of functional data, and allow straightforward and meaningful interpretation. The likelihood of the models is derived with Euler approximation and data augmentation. A unified Bayesian inference method is carried out via a Markov chain Monte Carlo algorithm including a simulation smoother. The proposed models and methods are illustrated on some prostate-specific antigen data, where we also show how the models can be used for forecasting.

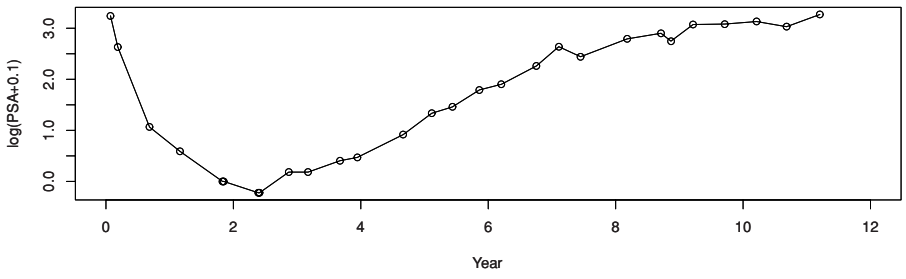
Key words: Diffusion model; Euler approximation; Nonparametric regression; Simulation smoother; Stochastic differential equation; Stochastic velocity model.

1. Introduction

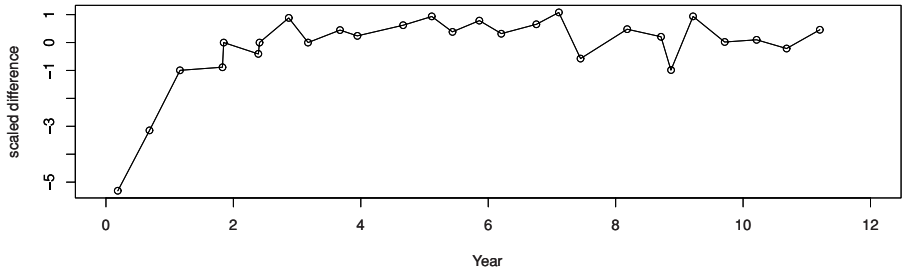
With the advent of many high-throughput technologies, functional data are routinely collected. To analyze those data, we usually assume that the observations are generated from an unknown mean function with additive errors. This article aims to use a diffusion model to estimate the mean function and its derivatives under this assumption.

There is a rich literature on penalized methods to regulate the mean function and to incorporate smoothness assumptions by using the penalized functions, with the focus on estimation of the unknown mean function (Wahba, 1990; Green and Silverman, 1994; Ramsay and Silverman, 2005). In many practical settings, not only the mean function but also its derivatives (in general referred to as dynamics) offer useful insights regarding the underlying mechanism of a physical or biological process. For example, in the study of prostate-specific antigen (PSA), an important biomarker of prostate cancer, we are not only interested in the PSA level but also the dynamics of PSA. Figure 1 displays raw data of one patient's PSA level (panel (a)) and the scaled difference (panel (b)) over time (Proust-Lima et al., 2008), where $Y(t) = \log(PSA(t) + 0.1)$ and scaled difference is $\frac{\Delta Y}{\Delta t}$. It is easy to observe that the PSA level is largely driven by the behavior of the scaled difference that itself provides meaningful clinical interpretation. Modeling the process of the scaled difference will facilitate the modeling of the PSA level. However, the connection between the PSA level and the scale difference cannot be established simply by association, but instead by hierarchical models of dynamics, as the scaled difference may be regarded as the derivative of the PSA level.

This article presents a new modeling strategy in functional data analysis, where the a priori smoothness assumption is specified by stochastic diffusion processes, using a set of ordinary and stochastic differential equations connected in a hierarchical fashion, in the hope that it not only models the mean function but also captures its various dynamic features. Note that this approach treats the unknown mean function and its dynamics as a sample path of stochastic processes. This treatment is different from kernel smoothing and spline smoothing, where the mean function is regarded as a deterministic unknown function. Our treatment of the mean function is similar to that considered in the Gaussian process models for nonparametric Bayesian data analysis, where the mean function is governed by a prior Gaussian process with a mean function $M(t;\boldsymbol{\phi})$ and a covariance function $C(t,t';\boldsymbol{\phi})$ with hyperparameters $\boldsymbol{\phi}$ (Muller and Quintana, 2004; Rasmussen and Williams, 2006). However, the hierarchical structure of the proposed model enables us to make inference of the mean function and its dynamics simultaneously. In particular, the method provides the estimation and inference for parameters of the stochastic differential equation from noisy data. We note that this differs from the approaches to parameter estimation for models based on ordinary differential equations as recently developed by, for example, Ramsay et al. (2007) and Liang and Wu (2008).



(a)



(b)

Figure 1. PSA plots: (a) the raw data; (b) the scaled difference.

The rest of the article is organized as follows. Section 2 introduces the proposed model and considers two special cases. For each case, we give model interpretation and discuss several interesting relationships. Section 3 develops Bayesian inference for stochastic functional data analysis models, where the likelihood is derived using Euler approximation and data augmentation. Section 4 presents a simulation study. In Section 5, the proposed models and methods are applied to estimate the PSA profile from prostate cancer data. Concluding remarks are given in Section 6. Technical details are included in the Web Appendix.

2. Stochastic Velocity Model

2.1 *Model Specification*

Consider a regression model for functional data of the form:

$$Y(t) = U(t,\omega) + \varepsilon(t),\ \omega \in \Omega,\ t \in \mathcal{T}_o, \tag{1}$$

where Ω is the sample space, $\mathcal{T}_o$ is the index set of observation times, defined as $\mathcal{T}_o := \{t_j : t_1 < t_2 < \cdots < t_J\}$, and $U(\cdot,\, \omega)$ is an unknown function of interest to be estimated and $\varepsilon(t) \sim \mathcal{N}(0, \sigma_\varepsilon^2)$ at each time t. The goal is to estimate the function $U(\cdot,\, \omega)$ and its derivatives given time series observations, $\boldsymbol{Y}_o = [Y(t_1), Y(t_2), \ldots, Y(t_J)]^{\mathrm{T}}$. In this article, we develop methods based on diffusion type models for estimation of $U(\cdot,\, \omega)$ and its derivatives $U^{(p)}(\cdot,\, \omega)$, $p = 1, \ldots,\, m - 1$. Here, $U(\cdot,\, \omega)$ is regarded as a realization of an underlying stochastic process $U := U(\cdot,\, \cdot)$ and thus the observed data are the sample path of the process plus measurement error.

Model (1) is useful to model the PSA level of prostate cancer nonparametrically, where $U(t,\, \omega)$ describes the population mean PSA process. To understand the dynamics of the process, we incorporate models of rate and/or higher-order derivatives into model (1). To proceed, we begin by treating $U(t,\, \omega)$ in model (1) as a realization of $U(t) := U(t,\, \cdot)$, which enables us to express $U(t)$ in the form of a stochastic diffusion model. That is, the stochastic process U satisfies the following ordinary differential equation (ODE):

$$\frac{d^{m-1}U(t)}{dt^{m-1}} = V(t), \tag{2}$$

and its $(m - 1)$th order derivative $V(t)$ is governed by a stochastic differential equation (SDE), given as follows:

$$dV(t) = a\{V(t), \boldsymbol{\phi}_s\}\, dt + b\{V(t), \boldsymbol{\phi}_s\}\, dW(t), \quad t \in \mathcal{T}_s, \tag{3}$$

where $W(t)$ is the standard Wiener process, $\boldsymbol{\phi}_s$ is the parameter vector, and $\mathcal{T}_s := \{t : t_0 \leqslant t \leqslant t_J\}$ is a continuous index set. In addition, the initial condition at time t_0 is assumed to be $\boldsymbol{\theta}(t_0) := [U(t_0), U^{(1)}(t_0), \ldots, U^{(m-2)}(t_0), V(t_0)]^{\mathrm{T}} \sim \mathcal{N}_m(0, \sigma_0^2 \boldsymbol{I}_m)$. In this article, we use continuous time stochastic processes $\boldsymbol{U}$ and $\boldsymbol{V}$ to model the underlying dynamics. Let $\boldsymbol{V} := \{V(t,\omega) : t \in \mathcal{T}_s, \omega \in \Omega\}$, defined on a probability space $(\Omega, \mathcal{F}, \mathcal{P})$. We limit $\boldsymbol{V}$ to a one-dimensional continuous state space and a continuous index set $\mathcal{T}_s$. Similar definition and limitation hold for $\boldsymbol{U}$. The SDE in (3) defines a stochastic diffusion process $\boldsymbol{V}$, which is a Markov process with almost surely continuous sample paths. The existence and uniqueness of the process can be shown rigorously; see Grimmett and Stirzaker (2001, Chap. 13) and Feller (1970, Chap. 10).

The state equations (2) and (3), along with the observation equation (1), make up a continuous-discrete state space model (CDSSM; Jazwinski, 1970, Chap. 6). Although inference methods will be demonstrated for the stochastic velocity model (SVM), namely the CDSSM with $m = 2$, they are applicable to any higher order of m. For example, $m = 3$ corresponding to a stochastic acceleration model (SAM). For SVM, the latent process $U(t)$ represents position, and its

first derivative $V(t)$ is the velocity of $U(t)$. Similarly, in the SAM, the processes $\boldsymbol{\theta}(t) := [U(t), U^{(1)}(t), V(t)]^{\mathrm{T}}$ represent the position, velocity, and acceleration respectively. Coefficients $a\{V(t), \boldsymbol{\phi}_s\}$ and $b\{V(t), \boldsymbol{\phi}_s\}$ in (3) are typically specified according to the objectives of a given study. The drift term $a\{V(t), \boldsymbol{\phi}_s\}$ can be interpreted as the instantaneous mean of velocity; it represents the expected conditional acceleration when $V(t)$ denotes velocity. Likewise, $b^2\{V(t), \boldsymbol{\phi}_s\}$ measures the instantaneous variance or volatility of velocity. The diffusion model and the consideration of higher derivative in (2) allow considerable flexibility and the incorporation of various dynamic features into the two coefficients $a\{V(t), \boldsymbol{\phi}_s\} \in \mathbb{R}$ and $b\{V(t), \boldsymbol{\phi}_s\} \in \mathbb{R}^+$. By this model-based approach, various stochastic processes can be specified for $V(t)$, the model fitting can be evaluated by likelihood-based model assessment, and forecasting can also be easily carried out.

Two special cases are considered in this article. They are, (i) SVM with Wiener process $V(t)$, denoted SVM-W, where $a\{V(t), \boldsymbol{\phi}_s\} = 0, b\{V(t), \boldsymbol{\phi}_s\} = \sigma_\xi$ and $\boldsymbol{\phi}_s = \sigma_\xi^2$; (ii) SVM with Ornstein-Uhlenbeck(OU) process $V(t)$, denoted SVM-OU, where $a\{V(t), \boldsymbol{\phi}_s\} = -\rho\{V(t) - \bar{\nu}\}, b\{V(t), \boldsymbol{\phi}_s\} = \sigma_\xi$ and $\boldsymbol{\phi}_s = [\rho, \bar{\nu}, \sigma_\xi^2]^{\mathrm{T}}$.

2.2 *Wiener Process for Velocity*

In SVM-W, $V(t)$ follows a Wiener process, the instantaneous variance σ_ξ^2 measures the disturbance of velocity and influences the smoothness of $U(t)$. With the smaller the σ_ξ^2, $V(t)$ will appear less wiggly and hence $U(t)$ will be smoother. If $\sigma_\xi = 0$, the velocity $V(t)$ is constant over time, so $U(t)$ becomes a straight line.

Integrating (2) and (3) for $m = 2, a\{V(t), \boldsymbol{\phi}_s\} = 0$ and $b\{V(t), \boldsymbol{\phi}_s\} = \sigma_\xi$, we have

$$\begin{aligned} U(t) &= U(t_0) + \int_{t_0}^{t} V(s)\, ds \\ &= U(t_0) + V(t_0)(t - t_0) + \sigma_\xi \int_{t_0}^{t} W(s)\, ds, \end{aligned} \tag{4}$$

$$V(t) = V(t_0) + \sigma_\xi W(t). \tag{5}$$

The velocity $V(t)$ follows the Wiener process starting at $V(t_0)$. The position $U(t)$ follows a linear trend with deviation governed by the integrated Wiener process, $\sigma_\xi \int_0^t W(s)\, ds$. As shown in the literature, there exists an interesting "equivalence" between smoothing splines and Bayesian estimation of SVM-W (Kimeldorf and Wahba, 1970; Wahba, 1978; Weinert, Byrd, and Sidhu, 1980). By equivalence, we mean that the two methods give the same estimate of $U(t)$, see Web Appendix A for details.

2.3 *Ornstein–Uhlenbeck Process for Velocity*

The OU process originated as a model for the velocity of a particle suspended in fluid (Uhlenbeck and Ornstein, 1930). The velocity $V(t)$ takes the form

$$dV(t) = -\rho\{V(t) - \bar{\nu}\}\, dt + \sigma_\xi\, dW(t), \quad t \in \mathcal{T}_s, \tag{6}$$

where $\rho \in \mathbb{R}^+, \bar{\nu} \in \mathbb{R}$, and $\sigma_\xi \in \mathbb{R}^+$. In contrast to the Wiener process, the OU process is a stationary Gaussian process with stationary mean $\bar{\nu}$ and variance $\sigma_\xi^2/2\rho$. σ_ξ^2 has the same interpretation as that of the Wiener process. The instantaneous mean or the expected conditional acceleration $-\rho\{V(t) - \bar{\nu}\}$ describes how fast the process moves. The larger the ρ, the more rapidly the process evolves toward $\bar{\nu}$. The farther $V(t)$ departs from $\bar{\nu}$, the faster the process moves back toward $\bar{\nu}$.

If the data are equally spaced, namely $\delta_j := t_j - t_{j-1} = \delta$, $V(t)$ coincides with the first-order autoregression (AR(1)) process with autocorrelation $\exp(-\rho\delta)$. The converse also holds; AR(1) converges weakly to the OU process as $\delta_j \to 0$ (Cumberland and Sykes, 1982).

For the PSA data example in Figure 1, it is obvious that the scaled difference is varying around a certain level after about 3 years, which is more consistent with the behavior of an OU process than a Wiener process, suggesting that the SVM-OU may fit better.

3. Estimation and Inference

Statistical inference for CDSSM is challenging because we consider a vector of stochastic processes $\boldsymbol{\theta}(t) := [U(t), U^{(1)}(t), \ldots, U^{(m-2)}(t), V(t)]^{\mathrm{T}}$ simultaneously. This leads to a complex likelihood function, which may not even exist in a closed form. Since an analytical solution of the SDE is rarely available, the conditional distribution of $\boldsymbol{\theta}(t)$ given $\boldsymbol{\theta}(t')$, for $t' < t$, which we call the exact transition density, does not have a simple closed-form expression. Thus exact inference for the latent processes and its parameters is not generally possible. Hence, a numerical approximation will usually be needed. We will use the Euler approximation of the SDE to approximate the transition density, which enables us to obtain a simple closed form of the likelihood. To alleviate the errors associated with this approximation, it may be helpful to augment the observed data by adding virtual data at extra time points (Tanner and Wong, 1987), so that the interval between adjacent time points is shorter and a more precise approximation is achieved. Even when the exact transition density exists, using the approximated one will significantly simplify the estimation of parameters $\boldsymbol{\phi}_s$. A case in point is the SVM-OU.

The resulting likelihood with this approximate method involves high-dimensional integrals, and we adopt a Bayesian approach using Markov chain Monte Carlo (MCMC; Gelfand and Smith, 1990; Geman and Geman, 1993; Gilks, Richardson, and Spiegelhalter, 1996) to estimate $U(t)$, $V(t)$, and the parameters $(\sigma_\varepsilon^2, \boldsymbol{\phi}_s)$, with the assistance of the simulation smoother (Durbin and Koopman, 2002).

Different approaches for inference for discretely observed diffusions are reviewed by Beskos et al. (2006). These include numerical approximations to obtain likelihood functions (Aït-Sahalia, 2002) and methods based on iterated filtering (Ionides, Breto, and King 2006). The idea of Euler approximation has been applied to the stochastic volatility model in the finance literature. Pedersen (1995) applied the approximation and data augmentation to facilitate Monte Carlo integration and it was further developed by Durham and Gallant (2002). Bayesian analysis of the diffusion model, especially the stochastic volatility model, has been developed by many authors, including Elerian, Chib, and Shephard (2001), Eraker (2001), and Roberts and Stramer (2001). Sorensen (2004) gave a survey on inference methods for stochastic diffusion models in finance. Distinctions between the models considered in financial statistics and the models considered in this article are that we specify an observation equation to allow for the measurement errors. Most methods of inference for diffusion

process do not extend easily when there is measurement error (Beskos et al., 2006). However, MCMC methods can be extended. A further distinction is that we consider the case $m > 1$ for the ODE, and that we apply the ODE and SDE to model various biomedical phenomena via $U(t)$ and $V(t)$. Thus, the SVM is focused on estimating the unknown sample paths of the latent stochastic process $U(t)$ and $V(t)$, whereas the diffusion models commonly used in the finance literature do not include an observation equation for measurement errors and typically focus on estimating the volatility or variance of the process of interest, for example, derivative securities.

3.1 *Likelihood and Euler Approximation*

To develop Bayesian inference with an MCMC algorithm, we begin with the likelihood of the SVM

$$[\boldsymbol{y}_o \mid \boldsymbol{\phi}_o, \boldsymbol{\phi}_s, \boldsymbol{\theta}_0] = \int\int [\boldsymbol{y}_o \mid \boldsymbol{U}_o, \boldsymbol{V}_o, \boldsymbol{\phi}_o][\boldsymbol{U}_o, \boldsymbol{V}_o \mid \boldsymbol{\theta}_0, \boldsymbol{\phi}_s]\, d\boldsymbol{U}_o\, d\boldsymbol{V}_o,$$

where $\boldsymbol{U}_o := [U(t_1), U(t_2), \ldots, U(t_J)]^{\mathrm{T}}$, $\boldsymbol{V}_o := [V(t_1), V(t_2), \ldots, V(t_J)]^{\mathrm{T}}$, and $\boldsymbol{y}_o := [y(t_1), y(t_2), \ldots, y(t_J)]^{\mathrm{T}}$ are vectors of the latent states and observations at $t \in \mathcal{T}_o$. $\boldsymbol{\theta}_0 = [U(t_0), V(t_0)]^{\mathrm{T}}$ is the unknown initial vector of the latent states, and $[.|.]$ denote conditional density. The conditional density of the observations is given by

$$[\boldsymbol{y}_o \mid \boldsymbol{U}_o, \boldsymbol{V}_o, \boldsymbol{\phi}_o] = \prod_{j=1}^{J} \phi\left[y(t_j) \mid U(t_j), \sigma_\varepsilon^2\right],$$

since the observations are mutually independent given the latent states and follow a normal distribution according to model (1), where $\phi(\cdot \mid U_G, \sigma_G^2)$ is the normal density with mean U_G and variance σ_G^2. In principle, the density of latent states $\boldsymbol{U}_o$ and $\boldsymbol{V}_o$ can be written as

$$[\boldsymbol{U}_o, \boldsymbol{V}_o \mid \boldsymbol{\theta}_0, \boldsymbol{\phi}_s] = \prod_{j=1}^{J} [\, U(t_j), V(t_j) \mid U(t_{j-1}), V(t_{j-1}), \boldsymbol{\phi}_s],$$

due to the Markov property. The exact transition density $[U(t_j), V(t_j) \mid U(t_{j-1}), V(t_{j-1}), \boldsymbol{\phi}_s)$ exists in a closed form only for few models with simple SDEs. Even in those cases, the exact transition density may have a complex form. For SVM-OU,

$$[U(t_j), V(t_j) \mid U(t_{j-1}), V(t_{j-1}), \boldsymbol{\phi}_s] = \mathcal{N}_2(\boldsymbol{m}_{OU}, \boldsymbol{V}_{OU}) \quad (7)$$

with

$$\boldsymbol{m}_{OU} = \left[U(t_{j-1}) + \bar{\nu}\delta_j + \{V(t_{j-1}) - \bar{\nu}\}\left\{\frac{1 - \exp(-\rho\delta_j)}{\rho}\right\},\quad \bar{\nu} + \{V(t_{j-1}) - \bar{\nu}\}\exp(-\rho\delta_j)\right]^{\mathrm{T}},$$

$$\boldsymbol{V}_{OU} = \sigma_\xi^2 \begin{bmatrix} \frac{\delta_j}{\rho^2} + \frac{1}{2\rho^3}\{-3 + 4\exp(-\rho\delta_j) - \exp(-2\rho\delta_j)\} & \frac{1}{2\rho^2}\{1 - 2\exp(-\rho\delta_j) + \exp(-2\rho\delta_j)\} \\ \frac{1}{2\rho^2}\{1 - 2\exp(-\rho\delta_j) + \exp(-2\rho\delta_j)\} & \frac{1}{2\rho}\{1 - \exp(-2\rho\delta_j)\} \end{bmatrix},$$

the proof of which is given by Zhu (2010). When using data augmentation, we may take the component-wise first-order Taylor approximation of $\boldsymbol{m}_{OU}, \boldsymbol{V}_{OU}$ with respect to δ_j and get

$$\tilde{\boldsymbol{m}}_{OU} = [U(t_j) + V(t_j)\delta_j, V(t_j) - \rho\{V(t_j) - \bar{\nu}\}\delta_j]^{\mathrm{T}}, \quad (8)$$

$$\tilde{\boldsymbol{V}}_{OU} = \sigma_\xi^2 \begin{bmatrix} 0 & 0 \\ 0 & \delta_j \end{bmatrix}. \quad (9)$$

We note that these are the same expressions as those obtained by applying Euler approximation to SVM-OU. Thus, although $\tilde{\boldsymbol{m}}_{OU}$ and $\tilde{\boldsymbol{V}}_{OU}$ as given in (8) and (9) are not strictly necessary for calculating $[\boldsymbol{U}_o, \boldsymbol{V}_o \mid \boldsymbol{\theta}_0, \boldsymbol{\phi}_s]$ when $\boldsymbol{m}_{OU}$ and $\boldsymbol{V}_{OU}$ are available, they however lead to a simpler form for parameter ρ, which is much easier to update and converges much faster in the following MCMC algorithm.

For a general SDE, e.g., (3), the forms for $U(t)$ and $V(t)$ are

$$U(t) = U(t_0) + \int_{t_0}^{t} V(s)\, ds,$$

$$V(t) = V(t_0) + \int_{t_0}^{t} a\{V(s), \boldsymbol{\phi}_s\}\, ds + \int_{t_0}^{t} b\{V(s), \boldsymbol{\phi}_s\} dW(s), \quad t \in \mathcal{T}, s,$$

where $[U(t_j), V(t_j) \mid U(t_{j-1}), V(t_{j-1}), \boldsymbol{\phi}_s]$ is implicitly defined but in general is not available analytically. To deal with this difficulty, we use the Euler approximation to obtain a numerical approximation of the transition density in the general SDE case.

The Euler approximation is a discretization method for the SDE through the first-order strong Taylor approximation (Kloeden and Platen, 1992). The resulting discretized versions of the ODE and the SDE in (2) and (3) are given by, respectively,

$$U^{(J)}(t_j) = U^{(J)}(t_{j-1}) + V^{(J)}(t_{j-1})\delta_j, \quad (10)$$

$$V^{(J)}(t_j) = V^{(J)}(t_{j-1}) + a\{V^{(J)}(t_{j-1}), \boldsymbol{\phi}_s\}\delta_j + b\{V^{(J)}(t_{j-1}), \boldsymbol{\phi}_s\}\eta_j, \quad t_j \in \mathcal{T}_o, \quad (11)$$

where $\delta_j = t_j - t_{j-1}$ and $\eta_j := W(t_j) - W(t_{j-1}) \sim \mathcal{N}(0, \delta_j)$. For $t \in [t_{j-1}, t_j\,]$, a linear interpolation takes the form

$$\tilde{V}^{(J)}(t) = V^{(J)}(t_{j-1}) + \frac{t - t_{j-1}}{t_j - t_{j-1}} \times \left(V^{(J)}(t_j) - V^{(J)}(t_{j-1})\right), \quad t \in \mathcal{T}_s.$$

A similar linear interpolation is applied to $\tilde{U}^{(J)}(t)$. Bouleau and Lepingle (1992) showed that under some regularity conditions, with constant C, the L^p-norm of the discretization error is bounded and given by

$$|| \sup_{t \in \mathcal{T}_s} | V(t) - \tilde{V}^{(J)}(t) | \,||_p \leqslant C \left(\frac{1 + \log J}{J}\right)^{1/2}.$$

This indicates that if J is sufficiently large, which can be achieved when the maximum of δ_j is sufficiently small for

fixed interval $[t_1, t_J]$, then $\tilde{V}^{(J)}(t)$ will be close to its continuous counterpart $V(t)$ with arbitrary precision.

In the rest of this article, we assume the δ_j is sufficiently small and the approximation is well achieved. To simplify notation, we replace $\tilde{V}^{(J)}(t)$ with $V(t)$ and $\tilde{U}^{(J)}(t)$ with $U(t)$, for $t \in \mathcal{T}_s$. Under these assumptions, the exact transition density, if it exists, is well approximated by the approximate transition density, as shown in the SVM-OU. Note that equations (10) and (11) imply the approximate transition densities of $U(t)$ and $V(t)$ are Gaussian for $t \in \mathcal{T}_o$, because they are linear combinations of η_j, $U(t_0)$, and $V(t_0)$, which are all Gaussian random variables. Under the Euler approximation, $[\boldsymbol{U}_o, \boldsymbol{V}_o \mid \boldsymbol{\theta}_0, \boldsymbol{\phi}_s]$ degenerates to $\prec \boldsymbol{V}_o \mid V(t_0), \boldsymbol{\phi}_s \succ$ because of equation (10), where $\prec \cdot \mid \cdot \succ$ denotes the approximate conditional density. Consequently, the likelihood based on the approximated processes $U(t)$ and $V(t)$ for $t \in \mathcal{T}_o$ is given by

$$\prec \boldsymbol{y}_o \mid \boldsymbol{\phi}_o, \boldsymbol{\phi}_s, V(t_0) \succ = \int [\boldsymbol{y}_o \mid \boldsymbol{V}_o, \boldsymbol{\phi}_o] \prec \boldsymbol{V}_o \mid V(t_0), \boldsymbol{\phi}_s \succ d\boldsymbol{V}_o,$$

where

$$[\boldsymbol{y}_o \mid \boldsymbol{V}_o, \boldsymbol{\phi}_o] = \prod_{j=1}^{J} \phi\big(y(t_j) \mid U(t_j), \sigma_\varepsilon^2\big),$$

$$\prec \boldsymbol{V}_o \mid V(t_0), \boldsymbol{\phi}_s \succ = \prod_{j=1}^{J} \prec V(t_j) \mid V(t_{j-1}), \boldsymbol{\phi}_s \succ,$$

and

$$V(t_j) \mid V(t_{j-1}), \boldsymbol{\phi}_s \sim \mathcal{N}\big(V(t_{j-1}) + a\{V(t_{j-1}), \boldsymbol{\phi}_s\}\delta_j, b^2\{V(t_{j-1}), \boldsymbol{\phi}_s\}\delta_j\big).$$

3.2 *Data Augmentation*

If observational time intervals are not short enough, the Euler approximation will not work well, because linear interpolation of $V(t)$ and $U(t)$ for $t \in \mathcal{T}_o$ is not accurate enough. A solution to reduce the approximation error is simply to add sufficiently dense virtual data in each time interval and consider the latent states at these times in addition to those at $t \in \mathcal{T}_o$. The corresponding values of $Y(\cdot)$ at added times can be regarded as missing data. They will be sampled as part of the MCMC scheme in the Bayesian analysis.

To carry out data augmentation, we add M_j equally spaced data at times $t_{j-1,1}, \ldots, t_{j-1,M_j}$ over a time interval $(t_{j-1}, t_j]$. Denote $\delta_{M_j} := \frac{\delta_j}{M_j+1}$. The resulting augmented index set is $\mathcal{T}_{ao} = \{t_{j,m} : j = 0, 1, \ldots, J, m = 0, 1, 2, \ldots, M_j, M_J = 0\}$. Note that $\mathcal{T}_{ao} = \mathcal{T}_o$, if $M_j = 0$ for all j. The observed data and the augmented data are denoted by $\boldsymbol{y}_o := [y(t_{1,0}), y(t_{2,0}), \ldots, y(t_{J,0})]^{\mathrm{T}}$ and $\boldsymbol{y}_a := [\boldsymbol{y}_{a,0}^{\mathrm{T}}, \boldsymbol{y}_{a,1}^{\mathrm{T}}, \ldots, \boldsymbol{y}_{a,J-1}^{\mathrm{T}}]^{\mathrm{T}}$, respectively, where $\boldsymbol{y}_{a,j} := [y(t_{j,1}), y(t_{j,2}), \ldots, y(t_{j,M_j})]^{\mathrm{T}}$. We also denote $\boldsymbol{V} := [\boldsymbol{V}_1^{\mathrm{T}}, \boldsymbol{V}_2^{\mathrm{T}}, \ldots, \boldsymbol{V}_J^{\mathrm{T}}]^{\mathrm{T}}$ where $\boldsymbol{V}_j := [V(t_{j,0}), V(t_{j,1}), \ldots, V(t_{j,M_j})]^{\mathrm{T}}$. Similar notation is applied to $\boldsymbol{U}$ and $\boldsymbol{U}_j$. For ease of exposition, we let $y_{j,m} := y(t_{j,m})$, and similarly for other variables.

If the exact transition densities exist, the augmented likelihood is

$$[\boldsymbol{y}_o \mid \boldsymbol{\phi}_o, \boldsymbol{\phi}_s, \boldsymbol{\theta}_0] = \int\int\int [\boldsymbol{y}_o, \boldsymbol{y}_a \mid \boldsymbol{U}, \boldsymbol{V}, \boldsymbol{\phi}_o] \times [\boldsymbol{U}, \boldsymbol{V} \mid \boldsymbol{\theta}_0, \boldsymbol{\phi}_s]\, d\boldsymbol{y}_a\, d\boldsymbol{U}\, d\boldsymbol{V},$$

where

$$[\boldsymbol{y}_o, \boldsymbol{y}_a \mid \boldsymbol{U}, \boldsymbol{V}, \boldsymbol{\phi}_o] = \prod_{j=0}^{J}\prod_{m=0}^{M_j} \phi\big(y_{j,m} \mid U_{j,m}, \sigma_\varepsilon^2\big),$$

$$[\boldsymbol{U}, \boldsymbol{V} \mid \boldsymbol{\theta}_0, \boldsymbol{\phi}_s] = \prod_{j=1}^{J}\prod_{m=1}^{M_j+1} [\, U_{j-1,m}, V_{j-1,m} \mid U_{j-1,m-1}, V_{j-1,m-1}, \boldsymbol{\phi}_s].$$

If the exact transition densities do not exist, the discretized versions of the ODE and the SDE are modified from $t \in \mathcal{T}_o$ to $t \in \mathcal{T}_{ao}$ and given as follows:

$$\begin{aligned} U_{j-1,m} &= U_{j-1,m-1} + V_{j-1,m-1}\delta_{M_j}, \\ V_{j-1,m} &= V_{j-1,m-1} + a\{V_{j-1,m-1}, \boldsymbol{\phi}_s\}\delta_{M_j} \\ &\quad + b\{V_{j-1,m-1}, \boldsymbol{\phi}_s\}\eta_{j-1,m}, \end{aligned}$$

where $t_{0,0} := t_0, t_{j-1,M_j+1} := t_{j,0}$, and $\eta_{j-1,m} := W(t_{j-1,m}) - W(t_{j-1,m-1}) \sim \mathcal{N}(0, \delta_{M_j})$. The approximate transition density and the corresponding likelihood are given in Section 3.3.

3.3 *Bayesian Inference*

MCMC enables us to draw samples from the joint posterior $[\boldsymbol{\theta}_0, \boldsymbol{V}, \boldsymbol{\phi}_o, \boldsymbol{\phi}_s \mid \boldsymbol{y}_o]$ or $[\boldsymbol{\theta}_0, \boldsymbol{V}, \boldsymbol{\phi}_o, \boldsymbol{\phi}_s, \boldsymbol{y}_a \mid \boldsymbol{y}_o]$. For the latter case, we will augment M_j equally spaced data points between time interval $(t_{j-1}, t_j]$. To assess whether M_j is sufficiently large we suggest a sensitivity analysis in which M_j is increased until the parameter estimates are stable. An illustration of this is given in Section 4. We may also compare the trace plots of parameter estimates and deviance information criteria (DIC) values for different M_j values to evaluate the numerical performance of MCMC and goodness of fit respectively.

MCMC draws samples from $[\boldsymbol{\theta}_0, \boldsymbol{V}, \boldsymbol{\phi}_o, \boldsymbol{\phi}_s, \boldsymbol{y}_a \mid \boldsymbol{y}_o]$ by iteratively simulating from each full conditional density of $\boldsymbol{\theta}_0, \boldsymbol{V}, \boldsymbol{\phi}_o, \boldsymbol{\phi}_s$, and $\boldsymbol{y}_a$. The joint posterior density is proportional to the product of the likelihood and prior densities:

$$[\boldsymbol{\theta}_0, \boldsymbol{V}, \boldsymbol{\phi}_o, \boldsymbol{\phi}_s, \boldsymbol{y}_a \mid \boldsymbol{y}_o] \propto [\boldsymbol{y}_o \mid \boldsymbol{\theta}_0, \boldsymbol{V}, \boldsymbol{\phi}_o, \boldsymbol{\phi}_s] \times [\boldsymbol{y}_a \mid \boldsymbol{\theta}_0, \boldsymbol{V}, \boldsymbol{\phi}_o, \boldsymbol{\phi}_s] \prec \boldsymbol{V} \mid \boldsymbol{\theta}_0, \boldsymbol{\phi}_s \succ \times [\boldsymbol{\theta}_0][\boldsymbol{\phi}_s][\boldsymbol{\phi}_o],$$

where

$$[\boldsymbol{y}_o \mid \boldsymbol{\theta}_0, \boldsymbol{V}, \boldsymbol{\phi}_o, \boldsymbol{\phi}_s] = \prod_{j=1}^{J} \phi(y_{j,0} \mid U_{j,0}(\boldsymbol{\theta}_0, \boldsymbol{V}), \boldsymbol{\phi}_o),$$

$$[\boldsymbol{y}_a \mid \boldsymbol{\theta}_0, \boldsymbol{V}, \boldsymbol{\phi}_o, \boldsymbol{\phi}_s] = \prod_{j=0}^{J-1}\prod_{m=1}^{M_j} \phi(y_{j,m} \mid U_{j,m}(\boldsymbol{\theta}_0, \boldsymbol{V}), \boldsymbol{\phi}_o),$$

$$\prec \boldsymbol{V} \mid \boldsymbol{\theta}_0, \boldsymbol{\phi}_s \succ = \prod_{j=1}^{J}\prod_{m=1}^{M_j+1} \prec V_{j-1,m} \mid V_{j-1,m-1}, \boldsymbol{\phi}_s \succ,$$

and $V_{j-1,M_j+1} = V_{j,0}$. The approximate transition density $\prec V_{j-1,m} \mid V_{j-1,m-1}, \boldsymbol{\phi}_s \succ$ with augmented data is given by,

$$\begin{aligned}&\prec V_{j-1,m} \mid V_{j-1,m-1}, \boldsymbol{\phi}_s \succ :=\\ &\quad \phi\big(V_{j-1,m} \mid V_{j-1,m-1} + a\{V_{j-1,m-1}, \boldsymbol{\phi}_s\}\delta_{M_j},\\ &\qquad b^2\{V_{j-1,m-1}, \boldsymbol{\phi}_s\}\delta_{M_j}\big),\end{aligned}$$

and $[\boldsymbol{\theta}_0], [\boldsymbol{\phi}_s], [\boldsymbol{\phi}_o]$ are noninformative prior densities. See Web Appendix B for specification of the prior distributions and details of the MCMC algorithm. We use the simulation smoother (Durbin and Koopman, 2002) to achieve an efficient MCMC algorithm. In the simulation smoother, the latent states are recursively backward sampled in blocks instead of one state at a time. This leads to low autocorrelation between successive draws, and hence faster convergence.

3.4 *Posterior Forecasting with SVM*

A desirable property of this approach is the ease of deriving forecasts of states at future times. To forecast the k-step future latent state $\boldsymbol{\theta}^f_{J+k}$ given the observations $\boldsymbol{y}_o$, we simulate $\boldsymbol{\theta}^f_{J+k}$ from the following posterior forecasting distribution:

$$\begin{aligned}\left[\boldsymbol{\theta}^f_{J+k} \mid \boldsymbol{y}_o\right] = \int\int\int &\left[\boldsymbol{\theta}^f_{J+k} \mid \boldsymbol{y}_a, \boldsymbol{y}_o, \boldsymbol{\phi}_s, \boldsymbol{\phi}_o\right]\\ &\times [\boldsymbol{y}_a, \boldsymbol{\phi}_s, \boldsymbol{\phi}_o \mid \boldsymbol{y}_o]\, d\boldsymbol{y}_a\, d\boldsymbol{\phi}_s\, d\boldsymbol{\phi}_o,\end{aligned}$$

where $\boldsymbol{y}_a, \boldsymbol{\phi}_s$, and $\boldsymbol{\phi}_o$ are drawn from $[\boldsymbol{y}_a, \boldsymbol{\phi}_s, \boldsymbol{\phi}_o \mid \boldsymbol{y}_o]$ by the MCMC algorithm. Given $\boldsymbol{y}_a, \boldsymbol{\phi}_s$, and $\boldsymbol{\phi}_o$, we first discretize the SVM. For SVM-OU, this will lead to equations (B.3) and (B.4) in Web Appendix B. Let $\boldsymbol{\theta}_J$ denote the latent state of the last observation. Then, $\mathbb{E}(\boldsymbol{\theta}_J) = \boldsymbol{a}_J$ and $\mathbb{V}ar(\boldsymbol{\theta}_J) = \boldsymbol{R}_J$ are obtained via the Kalman filter. Moreover, it follows from (B.4) that the mean and variance of $\boldsymbol{\theta}^f_{J+k}$ can be recursively obtained as follows:

$$\begin{aligned}\boldsymbol{a}_{J+k} &= \boldsymbol{G}_{J+k-1}\boldsymbol{a}_{J+k-1}\\ \boldsymbol{R}_{J+k} &= \boldsymbol{G}_{J+k-1}\boldsymbol{R}_{J+k-1}\boldsymbol{G}^{\mathrm{T}}_{J+k-1} + \Sigma_{\omega_{J+k-1}},\quad k = 1, 2, \ldots,\end{aligned}$$

where $\boldsymbol{G}_{J+k-1}$ and $\Sigma_{\omega_{J+k-1}}$ are specified in Web Appendix B for the SVM-OU and SVM-W, respectively. Finally, we draw $\boldsymbol{\theta}^f_{J+k}$ from $\boldsymbol{\theta}^f_{J+k} \mid \boldsymbol{y}_a, \boldsymbol{y}_o, \boldsymbol{\phi}_s, \boldsymbol{\phi}_o \sim \mathcal{N}(\boldsymbol{a}_{J+k}, \boldsymbol{R}_{J+k})$. By this way, the forecasts at future times take the variation of parameter draws into consideration.

4. Simulations

Using Euler approximation enables us to make inference for the SVMs with the analytically intractable exact transition densities. Although such flexibility seems to induce approximation errors, in practice these errors would be alleviated by applying the data augmentation. We are interested in assessing the performance of the estimation of the parameters, $U(t)$ and $V(t)$ as the number of data augmentation changes, so we simulate 100 replicate datasets from the SVM-OU with exact transition density (7). The parameters are chosen to be close to the ones estimated from a real dataset analyzed in the following section. Each dataset includes 40 observations, equally spaced with interval length 0.5. We fit each dataset by the SVM-OU, under the following three data augmentation schemes: (1) No augmentation; (2) one data point added in the middle of every two adjacent observations; and (3) three evenly spaced data points are inserted between every two adjacent observations. In all cases, MCMC is run for 45,000 iterations, in which the first 35,000 runs are discarded as the burn-in and every 10th draw is saved. Table 1 presents the Bias $E(\tilde{\phi}_{0.5} - \phi)$, and mean squared error (MSE) $E(\tilde{\phi}_{0.5} - \phi)^2$ of the posterior median $\tilde{\phi}_{0.5}$ for each parameter ϕ. These results indicate that the strategy of data augmentation, even for the single data point augmentation, would reduce estimation bias rate for variance parameter σ^2_ξ and drift parameter ρ. The estimations of other parameters σ^2_ε and $\bar{\nu}$ are little affected by the data augmentation. As shown in Web Table 2, the data augmentation improves the estimation of $U(t)$ and $V(t)$ as well, indicated by the smaller average absolute Bias $\frac{1}{40}\sum_{j=1}^{40} E(\mid \hat{\theta}(t_j) - \theta(t_j) \mid)$ and average MSE $\frac{1}{40}\sum_{j=1}^{40} E(\hat{\theta}(t_j) - \theta(t_j))^2$ of posterior means. We observed that the computation time of the proposed MCMC algorithm is proportional to the size of observed and augmented data, and thus the data augmentation would significantly increase the computation burden, which discourages us to add a massive number of augmented data. Those simulation results suggest that in practice we may add few augmented data and achieve accurate estimation.

Table 1
Simulation results for the estimation of SVM-OU parameters and stable rates. For data augmentation, 0, 1*, and* 3 *data points are added between every two adjacent observations.*

No. data points augmented	Parameter	Truth	Bias	MSE
0	σ^2_ε	0.01	−6.723E-04	1.590E-05
	σ^2_ξ	0.2	−9.837E-02	1.322E-02
	ρ	1	−1.863E-01	5.239E-02
	$\bar{V}$	0.3	−7.612E-03	1.297E-02
1	σ^2_ε	0.01	1.120E-04	1.521E-05
	σ^2_ξ	0.2	−6.089E-02	1.130E-02
	ρ	1	−6.426E-02	3.793E-02
	$\bar{V}$	0.3	−9.181E-03	1.304E-02
3	σ^2_ε	0.01	4.489E-04	1.611E-05
	σ^2_ξ	0.2	−3.905E-02	1.219E-02
	ρ	1	8.856E-03	4.687E-02
	$\bar{V}$	0.3	−9.955E-03	1.284E-02

To compare, we analyze these simulated datasets by the SVM-W, and an ODE model, respectively. Note that the posterior mean in the SVM-W is equivalent to the smooth natural cubic spline. The ODE model used in the analysis satisfies the observation equation (1) for functional response, and the mean function $U(\cdot)$ is governed by the following two ODEs:

$$\begin{aligned}dU(t) &= V(t)\, dt,\\ dV(t) &= -\beta_3(V(t) - \beta_1)\, dt,\end{aligned} \tag{12}$$

with boundary conditions $U(t_0) = \beta_0 + \beta_2$ and $V(t_0) = \beta_1 - \beta_2\beta_3$ at $t_0 = 0$. This formulation leads to a deterministic mean function $U(t) = \beta_0 + \beta_1 t + \beta_2 \exp(-\beta_3 t)$, which implies that ODE model is essentially a parametric nonlinear regression model. As shown in Table 2, the proposed SVM-OU model is superior to the ones from the other two models

Table 2
Simulation results for the estimation of $U(t)$ and $V(t)$, when data are simulated from SVM-OU and ODE model. For data augmentation, 0, 1, and 3 data points are added between every two adjacent observations.

Simulation model	Fitting model	No. data points augmented	States	Bias	MSE
SVM-OU	SVM-OU	0	$U(t)$	0.008	0.005
			$V(t)$	0.079	0.085
	SVM-OU	1	$U(t)$	0.008	0.005
			$V(t)$	0.046	0.048
	SVM-OU	3	$U(t)$	0.008	0.005
			$V(t)$	0.028	0.038
	SVM-WN	0	$U(t)$	0.014	0.006
			$V(t)$	0.052	0.079
	ODE	0	$U(t)$	0.051	0.099
			$V(t)$	0.054	0.133
ODE	SVM-OU	3	$U(t)$	0.043	0.026
			$V(t)$	0.045	0.090
	SVM-WN	0	$U(t)$	0.066	0.067
			$V(t)$	0.135	0.303
	ODE	0	$U(t)$	0.030	0.018
			$V(t)$	0.020	0.018

in the estimation of $U(t)$ and $V(t)$, when the SVM-OU model generates the data. We are also interested in exploring how the proposed methods perform when data are generated from other models different from the SVM-OU. For this, we further simulate another 100 replicate datasets from above the ODE model with $\beta_0 \sim \mathcal{N}(-2.5, 0.01), \beta_1 \sim \mathcal{N}(0.3, 0.01), \beta_2 \sim \mathcal{N}(9, 0.04), \beta_3 \sim \mathcal{N}(1, 0.01)$ and $\sigma^2 = 0.04$. The processes $U(t)$ and $V(t)$ are then estimated by the SVM-OU, SVM-W, and ODE model, respectively. As shown Table 2 when data are generated from the ODE model, estimated $U(t)$ and $V(t)$ by the SVM-OU are close to those by the ODE model, and better than the SVM-W in terms of smaller average absolute bias and average MSE.

In addition to the above equally spaced data, we are also consider some other scenarios of datasets simulated from the SVM-OU, including: (1), the unequally spaced data resulted from uniform deletion of 15 data points from the original equally spaced data; (2), unbalanced and unequally spaced data resulted from uniform random deletion of 10 data points in the first half period (0,10] and of 5 data points in the second period (10,20] from the original equally spaced data simulated by the SVM-OU; (3), the data similar to the scenario (2) except with the reserved allocation of the unbalanced data points; (4) and (5), The fivefold larger variance $\sigma_\varepsilon^2 = 0.05$ and 10-fold larger variance $\sigma_\varepsilon^2 = 0.1$. Web Table 1 lists the summarized results of these five scenarios, regarding the estimation of $U(t)$ and $V(t)$ by the SVM-OU. The results from the first three scenarios indicate that the unbalance data distribution between the first and second half period times has different impact on the estimation of $U(t)$ and $V(t)$ in the average absolute bias and MSE. The second case with less data available in the first half time period is the worst. In the last two scenarios with increased measurement errors, the simulation results suggest that the estimation for $U(t)$ and $V(t)$ becomes a harder task.

5. Application

We now demonstrate an application where the diffusion models are used to investigate dynamic features of the PSA profile for a prostate cancer patient. We fit the SVM and SAM with the Wiener process and the OU process $V(t)$, respectively. We also forecast the future profile of PSA for both models. The models are evaluated by the DIC model selection criterion (Speigelhalter et al., 2003). $DIC = \bar{D} + P_D$, where $\bar{D}$ is posterior mean of the deviance and P_D is the effective number of parameters. DIC has been shown asymptotically to be a generalization of Akaike's information criterion (AIC). Similar to AIC, DIC trades off the model fitting by the model complex and can be easily computed from the MCMC output. The smaller the DIC value indicates better model-fitting. For each application, the posterior draws are from a 400,000 iteration chain with 200,000 burn-in, and every 100th draw is selected. Convergence was assessed by examination of trace plots and autocorrelation plots.

5.1 *Prostate-Specific Antigen*

PSA is a biomarker used to monitor recurrence of prostate cancer after treatment with radiation therapy. When PSA remains low and its rate varies around zero with low volatility, the tumor is stable and the patient may be cured. If PSA increases dramatically with high rate, it is a strong sign of the tumor regrowing and that the treatment did not cure the patient. Therefore, PSA has strong prognostic significance and is important for making clinical decisions. We want to estimate dynamics of the PSA marker, including PSA level, rate, and the volatility of rate. We analyzed the PSA profile of one patient using the SVM and SAM model to estimate PSA(t) nonparametrically. For these data illustrated in the introduction, the average time interval between two observations was 0.4 years with minimum 0.016 and maximum 0.731 years. We added 32 virtual data points to reduce the time span between any pairs of consecutive time points to less than 0.25 year.

Table 3 and Web Table 2 show the means and quantiles of the SVM and SAM parameters from the Wiener and the OU process $V(t)$, respectively. Figure 2 and Web Figure 1 show the posterior means and the corresponding 95% credible intervals of the latent states for SVM and SAMs. Here, the four models demonstrate similar trends of the PSA level. However, the rates in the SVMs fluctuate with higher volatility, compared to the SAMs. In addition, there are the nonzero instantaneous mean terms in the SVM-OU and SAM-OUs, whose rates evolve more stably than those in the models with Wiener process. The SVM-OU gives the smallest DIC, which indicates the best model fitting. In this model, the posterior mean of $\bar{\nu}$ is 0.385 with 95% credible interval [0.143, 0.626]. This stable and clearly positive rate after year 2.2 is a strong indicator of prostate cancer recurrence.

Figure 2 illustrates the forecasting of the PSA latent states for the next 3 years, starting from year 11.2, by SVM-W and SVM-OU. The future states are sampled every 0.25 years and then linearly interpolated, from the posterior forecasting distribution given in Section 3.4. The SVM-OU gives a forecast with narrower credible intervals than the SVM-W. This

Table 3
PSA data: Posterior mean and quantiles for the SVMs

	Wiener process $\bar{D} = -45.1757,\ P_D = 12.303,\ DIC = -32.873$					OU process $\bar{D} = -45.935,\ P_D = 10.658,\ DIC = -35.277$				
	Mean	SD	2.5%	50%	97.5%	Mean	SD	2.5%	50%	97.5%
σ^2_ε	0.014	0.009	0.003	0.012	0.036	0.012	0.005	0.005	0.012	0.024
σ^2_ξ	0.961	0.589	0.297	0.809	2.548	0.177	0.181	0.037	0.122	0.682
$\bar{\nu}$						0.385	0.124	0.143	0.382	0.626
ρ						1.150	0.271	0.756	1.106	1.798

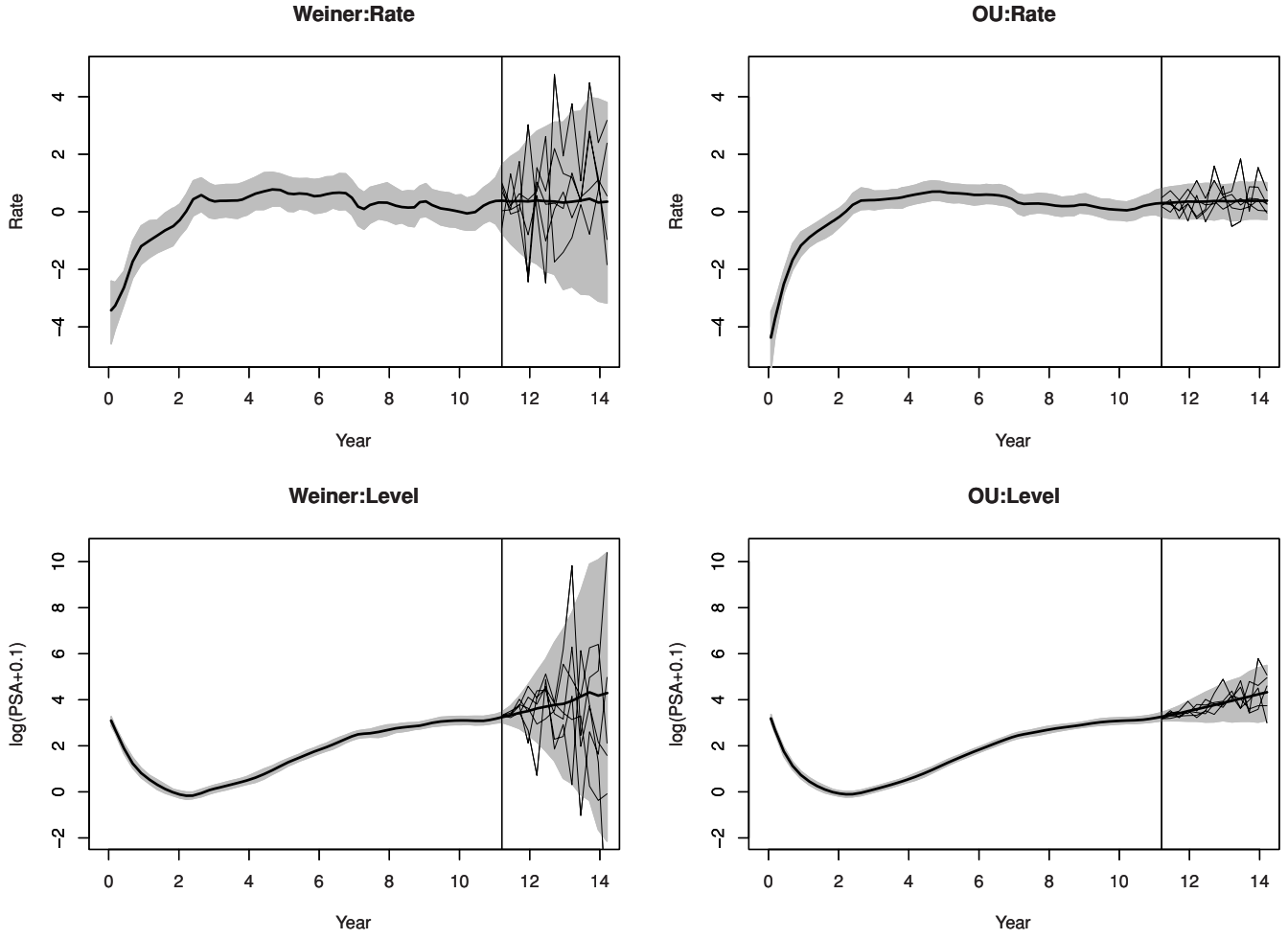


Figure 2. PSA posterior summary and forecasting: Plots of posterior means (|) and 95% credible intervals (gray shades) for the SVM with the Wiener process and OU process, respectively, till year 11.2. The future rates and levels are forecasted for the next 3 years, illustrated by the forecasting means (|) and 95% forecasting credible intervals (gray shades). The five randomly picked realizations for each plot are also illustrated.

result seems clinically more sensible, because several studies, including ours presented in Section 4.2, have found that the rate of PSA follows a stationary process. In contrast, a Wiener process corresponds to a nonstationary process for the rate of change of PSA, resulting in an unbounded variance of the forecast over time. This lacks relevant clinical interpretation. The comparison in the forecasts indicates that specification of the latent process is crucial for adequate forecasting, even though their estimates of the mean function are quite similar. A similar phenomenon has been reported by Taylor and Law (1998) in the linear mixed model of CD4 counts, where the covariance structure matters for individual-level predictions, although it affects little the estimation of fixed effects.

6. Discussion

Diffusion-type models are widely applied in areas such as finance, physics, and ecology. However, other than through the connection with the smoothing spline, they have not played a major role in functional data analysis or nonparametric regression. In this article, we develop a framework that sheds light on more general diffusion models to be used in functional data analysis. Unlike in some applications where the form of the diffusion model is determined by the context, we specify a general form based on an ODE, an SDE, and measurement error. The key advantage of the proposed diffusion model is that is addresses not only the mean function nonparametrically but also its dynamics, which are also of great interest in many applications. Based on this model, we adapt and develop existing ideas for estimation and inference for diffusion models. An additional attractive feature of this stochastic model approach to functional data analysis is that forecasting can be easily implemented.

A number of extensions of the SVM and SAM are possible. Generalizing SVM and SAM to analyze discrete-valued outcomes is of interest. For the SVM, we have an explicit expression for the observation equation given by:

$$\begin{aligned} Y(t_j) &= U(t_j) + \varepsilon(t_j) \\ &= \begin{bmatrix} 1 \\ 0 \end{bmatrix}^{\mathrm{T}} \begin{bmatrix} U(t_j) \\ V(t_j) \end{bmatrix} + \varepsilon(t_j) \\ &= \boldsymbol{F}^{\mathbf{T}}\boldsymbol{\theta}(t_j) + \varepsilon(t_j), \qquad j = 1, 2, \ldots, J. \end{aligned}$$

The observation equation can be expressed as,

$$d\,\Phi\big\{Y(t) \mid \boldsymbol{F}^{\mathbf{T}}\boldsymbol{\theta}(t), \sigma_\varepsilon\big\}, \quad t \in \mathcal{T}_o, \tag{13}$$

where $\Phi(\cdot \mid U_G,\, \sigma_G)$ is the normal CDF with mean U_G and standard deviation σ_G. Then, (13) can be extended to,

$$d\,\mathbb{F}\{Y(t) \mid \boldsymbol{\theta}(\boldsymbol{t}), \boldsymbol{\phi}_o\},$$

where one specifies the corresponding observation distribution $\boldsymbol{F}$ in the exponential family, with state equations (2) and (3) unchanged.

7. Supplementary Materials

Web Appendices, Tables, and Figures referenced in Sections 2, 3, 4, 5, and 6 are available under the Paper Information link at the *Biometrics* website `http://www.biometrics.tibs.org`.

Acknowledgements

This research was partially supported by National Cancer Institute grants CA110518 and CA69568. The authors are thankful to Dr Naisysin Wang and Dr Brisa Sanchez for helpful discussions, and the editor, associate editor, and two anonymous reviewers for valuable comments and suggestions.

References

Aït-Sahalia, Y. (2002). Maximum likelihood estimation of discretely sampled diffusions: A closed-form approximation approach. *Econometrica* **70,** 223–262.

Beskos, A., Papaspiliopoulos, O., Roberts, G., and Fearnhead, P. (2006). Exact and computationally efficient likelihood-based estimation for discretely observed diffusion processes (with discussion). *Journal of the Royal Statistical Society, Series B* **68,** 333–382.

Bouleau, N. and Lepingle, D. (1992). *Numerical Methods for Stochastic Process*. New York: Wiley.

Cumberland, W. and Sykes, Z. (1982). Weak convergence of an autoregressive process used in modeling population growth. *Journal of Applied Probability* **19,** 450–455.

Durbin, J. and Koopman, S. J. (2002). A simple and efficient simulation smoother for state space time series analysis. *Biometrika* **89,** 603–616.

Durham, G. and Gallant, A. (2002). Numerical techniques for maximum likelihood estimation of continuous-time diffusion processes. *Journal of Business and Economic Statistics* **20,** 297–338.

Elerian, O., Chib, S., and Shephard, N. (2001). Likelihood inference for discretely observed non-linear diffusions. *Econometrica* **69,** 959–993.

Eraker, B. (2001). MCMC analysis of diffusion models with application to finance. *Journal of Business and Economic Statistics.* **19,** 177–191.

Feller, W. (1970). *An Introduction to Probability Theory and Its Application*. New York: Springer.

Gelfand, A. E. and Smith, A. F. M. (1990). Sampling-based approaches to calculating marginal densities. *Journal of the American Statistical Association* **85,** 398–409.

Geman, S. and Geman, D. (1993). Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images. *Journal of Applied Statistics* **20**(5), 25–62.

Gilks, W., Richardson, S., and Spiegelhalter, D. (1996). *Markov Chain Monte Carlo in Practice*. New York: Chapman & Hall/CRC.

Green, P. and Silverman, B. (1994). *Nonparametric Regression and Generalized Linear Models*. New York: Chapman & Hall.

Grimmett, G. and Stirzaker, D. (2001). *Probability and Random Processes*. Oxford: Oxford University Press.

Ionides, E., Breto, C., and King, A. (2006). Inference for nonlinear dynamical systems. *Proceedings of the National Academy of Sciences USA* **103,** 18438.

Jazwinski, A. (1970). *Stochastic Processes and Filtering Theory*. New York: Academic Press.

Kimeldorf, G. S. and Wahba, G. (1970). A correspondence between Bayesian estimation on stochastic processes and smoothing by splines. *Annals of Mathematical Statistics* **41,** 495–502.

Kloeden, P. and Platen, E. (1992). *Numerical Solution of Stochastic Differential Equations*. New York: Springer.

Liang, H. and Wu, H. (2008). Parameter estimation for differential equation models using a framework of measurement error in regression models. *Journal of the American Statistical Association* **103,** 1570–1583.

Muller, P. and Quintana, F. (2004). Nonparametric Bayesian data analysis. *Statistical Science* **19,** 95–110.

Pedersen, A. R. (1995). A new approach to maximum likelihood estimation for stochastic differential equations based on discrete observations. *Scandinavian Journal of Statistics* **22,** 55–71.

Proust-Lima, C., Taylor, J. M. G., Williams, S., Ankerst, D., Liu, N., Kestin, L., Bae, K., and Sandler, H. (2008). Determinants of change of prostate-specific antigen over time and its association with recurrence following external beam radiation therapy of prostate cancer in 5 large cohorts. *International Journal of Radiation Oncology, Biology, Physics* **72,** 782.

Ramsay, J. and Silverman, B. (2005). *Functional Data Analysis*. New York: Springer.

Ramsay, J., Hooker, G., Campbell, D., and Cao, J. (2007). Parameter estimation for differential equations: A generalized smoothing approach. *Journal of the Royal Statistical Society, Series B* **69,** 741–796.

Rasmussen, C. and Williams, C. (2006). *Gaussian Processes for Machine Learning*. New York: Springer.

Roberts, G. O. and Stramer, O. (2001). On inference for partially observed nonlinear diffusion models using the Metropolis-Hastings algorithm. *Biometrika* **88,** 603–621.

Sorensen, H. (2004). Parametric inference for diffusion processes observed at discrete points in time: A survey. *International Statistical Review* **72,** 337–354.

Speigelhalter, D., Best, N., Carlin, B., and van der Linde, A. (2003). Bayesian measures of model complexity and fit (with discussion). *Journal of the Royal Statistical Society, Series B* **64,** 583–616.

Tanner, M. and Wong, W. (1987). The calculation of posterior distributions by data augmentation. *Journal of the American Statistical Association* **82,** 528–540.

Taylor, J. M. G. and Law, N. (1998). Does the covariance structure matter in longitudinal modelling for the prediction of future CD4 counts? *Statistics in Medicine* **17,** 2381–2394.

Uhlenbeck, G. and Ornstein, L. (1930). On the theory of the Brownian motion. *Physical Review* **36,** 823–841.

Wahba, G. (1978). Improper priors, spline smoothing and the problem of guarding against model errors in regression. *Journal of the Royal Statistical Society, Series B* **40,** 364–372.

Wahba, G. (1990). *Spline Models for Observational Data*. Philadelphia: Society for Industrial Mathematics.

Weinert, H., Byrd, H., and Sidhu, G. (1980). A stochastic framework for recursive computation of spline functions: Part ii, smoothing splines. *Journal of Optimization Theory and Applications* **01,** 255–268.

Zhu, B. (2010). *Stochastic Dynamic Models for Functional Data*. PhD thesis, University of Michigan, Ann Arbor, MI.

Received January 2010. *Revised January* 2011.
Accepted January 2011.