Published in final edited form as:
J Am Stat Assoc. 2011 December 1; 106(496): 1485–1495. doi:10.1198/jasa.2011.tm09294.

Semiparametric Stochastic Modeling of the Rate Function in Longitudinal Studies

Bin Zhu[Postdoctoral Associate]
Department of Statistical Science and Center for Human Genetics, Duke University, Durham, NC 27708, (bin.zhu@duke.edu)

Jeremy M.G. Taylor[Professor] and **Peter X.-K. Song[Professor]**
Department of Biostatistics, University of Michigan, Ann Arbor, MI 48109, (jmgt@umich.edu and pxsong@umich.edu)

Abstract

In longitudinal biomedical studies, there is often interest in the rate functions, which describe the functional rates of change of biomarker profiles. This paper proposes a semiparametric approach to model these functions as the realizations of stochastic processes defined by stochastic differential equations. These processes are dependent on the covariates of interest and vary around a specified parametric function. An efficient Markov chain Monte Carlo algorithm is developed for inference. The proposed method is compared with several existing methods in terms of goodness-of-fit and more importantly the ability to forecast future functional data in a simulation study. The proposed methodology is applied to prostate-specific antigen profiles for illustration. Supplementary materials for this paper are available online.

Keywords

Euler approximation; Functional data analysis; Gaussian process; Rate function; Stochastic differential equation; Semiparametric stochastic velocity model

1 Introduction

This paper focuses on semiparametric stochastic modeling of rate functions for functional data in a multi-subject setting, where the data consists of a set of subjects, and for each subject, the observations are discrete samples from a curve with additive measurement errors. The rate function describes the functional rate of change or slope with respect to time, and has been of recent interest in longitudinal biomedical studies (Mungas et al., 2005; Lloyd-Jones et al., 2007; Strasak et al., 2008; Kariyanna et al., 2010). For example, from subject-matter knowledge it may be the rate of change, rather than the level of some biomarker, that can explain and predict the disease outcomes. One challenge in this research is to model the rate function without making a strong parametric assumption. Further challenges include modeling the rate functions across the subjects and allowing it to depend on the covariates of interest.

Our development has been largely motivated by a longitudinal study in prostate cancer patients (Proust-Lima et al., 2008), where prostate-specific antigen (PSA) profiles were collected for patients who received external beam radiation therapy (EBRT). PSA is roughly

Supplementary Materials Supplementary Materials are available at the *JASA* website http://pubs.amstat.org/loi/jasa.

proportional to the prostate tumor size, and its rate of change has been shown to be associated with the recurrence of prostate cancer (Sartor et al., 1997). Figure 1(a) shows the log-transformed PSA level over time after EBRT treatment for 50 selected patients, and Figure 1(b) illustrates individual empirical rates of change, one for each subject. Figure 1(b) suggests that the individual rate of change of PSA roughly follows a common pattern. That is, it begins with a negative value caused by the EBRT, decreases over time in magnitude as the rate of tumor shrinkage gets lower, and eventually reaches a certain stable level. It is also apparent that rates of change vary considerably from this common pattern. For example, for the subject highlighted in black in Figure 1(b), his empirical rate of change fluctuates around zero and his PSA level appears very different from the others. Hence it is desirable to model the rate of change semiparametrically by incorporating empirical evidence or prior knowledge through a parametric function of time while accounting for deviation from the common pattern nonparametrically. Additionally, it is clear that for some subjects the long term stable rates of change are near zero, while for others they are positive. It is thus appealing not only to model a common stable rate of change across the subjects but also to let it follow a distribution, say a normal distribution with its mean depending on some baseline covariates. This flexibility will benefit the forecasts of future observations.

A number of methods have been used to study the rate of change in longitudinal studies. A popular approach is through a parametric linear mixed model (Laird and Ware, 1982; Diggle et al., 2002; Verbeke and Molenberghs, 2009), for example the random intercept and slope mixed model for disease progression (Zhang et al., 2008). This model assumes the subject's mean function follows a straight line with constant rate of change, which in turn is dependent on the covariates. In contrast to parametric models, the mean function have been modeled nonparametrically (Rice and Silverman, 1991; Wang and Taylor, 1995; Zeger and Diggle, 1994; Zhang et al., 1998; Verbyla et al., 1999). For these approaches, the rate function, as the first order derivative of the mean function, does not have any parametric form, and usually is not dependent on covariates. For other relevant literature that considers population dynamic models with multiple subjects see Wang et al. (2008), Paul et al. (2009) and Müller and Yao (2010). Additionally, in a time-varying coefficient model (Hastie and Tibshirani, 1993; Hoover et al., 1998) or functional mixed model (Guo, 2002; Morris and Carroll, 2006), the mean function $U_i(t)$ of the ith subject is specified as $U_i(t) = \Sigma_{k=0}^{K} X_{ik}\beta_k(t)$. Hence, $U_i(t)$ is a linear combination of several arbitrary smooth functions $\beta_k(t)$ with covariates X_{ik} as the weights and depends on covariates linearly. Thus there seems to be a need for a model that allows flexible relationships between the rate function and covariates. Moreover, note that except for few approaches (Qin and Guo, 2006; Welham et al., 2006), nonparametric approaches seldom incorporate any prior knowledge from the subject-matter science, if available, in the modeling of the shape of the rate function.

Our goal is to develop a semiparametric stochastic model for the analysis of the rate function, which is called in this paper a semiparametric stochastic velocity model (SSVM). A key feature of SSVM is to utilize a stochastic process as a prior for the rate function, in a similar spirit to the work of Wahba (1978) and Zhu et al. (2011) for functional data in a single-subject setting. Formally, for each rate function $V_{x_i}(t) \in \mathbb{R}$ for subject $i \in N = \{1, 2, \ldots, n\}$ and time $t \in \mathcal{T}_s = [0, \infty)$, its prior is assumed to be a Gaussian process, conditional on $\boldsymbol{x}_i = (x_{i0}, x_{i1}, \ldots, x_{ip})'$, the vector of covariates for the ith subject. As an important special case of the proposed SSVM, we consider $V_{x_i}(t) = f_{x_i}(t) + \sigma_\xi W_i(t)$, where $f_{x_i}(t)$ has a pre-specified parametric functional form dependent on covariates $\boldsymbol{x}_i$ and $\sigma_\xi W_i(t)$ is a scaled standard Wiener process. Hence, $E\{V_{x_i}(t)\} = f_{x_i}(t)$ implies that $V_{x_i}(t)$, the rate function of the ith subject, is expected to be centered about $f_{x_i}(t)$, while the second term $\sigma_\xi W_i(t)$ allows deviations from the parametric functional expectation $f_{xi}(t)$.

The remainder of the paper is organized as follows. Section 2 first presents the model and then is devoted to an important special case with the Ornstein-Uhlenbeck process as the prior for the rate function. Section 3 develops MCMC based methods for posterior inference and forecasting. Section 4 applies the methods to analyze the data of PSA profiles. Section 5 presents simulation results to evaluate and compare the performance of the proposed method with other existing methods. The paper concludes with a discussion in Section 6. Some supplementary materials related to the technical details of the proof of Theorem 1 are available online.

2 Semiparametric Stochastic Velocity Model

2.1 Model Specification

Suppose that $Y_i(t_{ij})$, $j = 1, 2, \ldots, m_i$, $i = 1, 2, \ldots, n$, is the response of the ith subject at time t_{ij} and satisfies the following hierarchical model, SSVM:

$$Y_i(t) = U_{\boldsymbol{x}_i}(t) + \varepsilon_i(t), \quad t \in \mathscr{T}_{io} = \left\{t_{ij} : t_{i1} < t_{i2} < \cdots < t_{im_i}\right\}, \tag{1}$$

$$dU_{\boldsymbol{x}_i}(t) = V_{\boldsymbol{x}_i}(t)\, dt, \qquad t \in \mathscr{T}_s = [t_0, \infty), \tag{2}$$

$$dV_{\boldsymbol{x}_i}(t) = a\{V_{\boldsymbol{x}_i}(t); \boldsymbol{x}_i, \phi_i\}\, dt + b\{V_{\boldsymbol{x}_i}(t); \boldsymbol{x}_i, \phi_i\}\, dW_i(t), t \in \mathscr{T}_s, \tag{3}$$

where $U_{x_i}(t)$ is the mean function for the ith subject's outcome curve, $V_{x_i}(t)$ is the corresponding rate function and $W_i(t)$ denotes the standard Wiener process. Note that in this specification, although the mean function is defined at continuous time $\mathscr{T}_s$, it is observed at discrete times $\mathscr{T}_{io}$ only and is subject to measurement error. Equation (3) may be regarded as a prior for the rate function $V_{x_i}(t)$, in which the behavior of $V_{x_i}(t)$ is governed by a stochastic differential equation (SDE), with drift term $a\{V_{x_i}(t); \boldsymbol{x}_i, \boldsymbol{\phi}_i\}$ and diffusion term $b\{V_{x_i}(t); \boldsymbol{x}_i, \boldsymbol{\phi}_i\}$, where $\boldsymbol{x}_i$ and $\boldsymbol{\phi}_i$ are the covariate vector and subject-specific parameter vector. We assume that the initial values $[U_{x_i}(t_0), V_{x_i}(t_0)]' \overset{\text{iid}}{\sim} \mathscr{N}_2\left(0, \sigma_0^2 \boldsymbol{I}_2\right)$ with large value of variance σ_0^2 to make it non-informative, and that the measurement error $\varepsilon_i(t) \overset{\text{iid}}{\sim} \mathscr{N}\left(0, \sigma_\varepsilon^2\right)$. Here $\boldsymbol{I}_k$ is the $k \times k$ identity matrix and $\mathscr{N}_k(m, \Sigma)$ denotes the k-dimensional normal distribution with mean vector $\boldsymbol{m}$ and covariance matrix Σ. Furthermore, $[U_{x_i}(t_0), V_{x_i}(t_0)]'$, $\varepsilon_i(t)$ and $W_i(t)$ are assumed mutually independent.

The SDE in equation (3) gives rise to a general class of Markovian Gaussian processes (Feller, 1970; Grimmett and Stirzaker, 2001). In our model, this stochastic process is considered as the prior for the rate function $V_{x_i}(t)$. According to the specific research interest or context of a given study, we can choose different forms for $a\{V_{x_i}(t); \boldsymbol{x}_i, \boldsymbol{\phi}_i\}$, which measures the instantaneous mean or the expected conditional acceleration, and for $b^2\{V_{x_i}(t); \boldsymbol{x}_i, \boldsymbol{\phi}_i\}$, which reflects the instantaneous variance of the rate process. In particular, we have the SSVM-W, when $a\{V_{x_i}(t); \boldsymbol{x}_i, \boldsymbol{\phi}_i\} = 0$ and $b\{V_{x_i}(t); \boldsymbol{x}_i, \boldsymbol{\phi}_i\} = \sigma_\xi$, and the prior for $V_{x_i}(t)$ is the Wiener process. The resulting mean function takes the form $U_{x_i}(t) = U_{x_i}(t_0) + V_{x_i}(t_0)(t - t_0) + \sigma_\xi \int_{t_0}^{t} W(s)\, ds$, which is the partially integrated Wiener process leading to a smoothing spline (Wahba, 1978; Wecker and Ansley, 1983; Ansley and Kohn, 1986) for a given subject. Note that this prior is independent of covariates.

For the PSA data analysis given in Section 4, we specify $a\{V_{x_i}(t); x_i, \phi_i\} = -\rho\left\{V_{x_i}(t) - \bar{\nu}_i(x_i, \beta)\right\}$ and $b\{V_{xi}(t); \boldsymbol{x}_i, \boldsymbol{\phi}_i\} = \sigma_\xi$. This specification

corresponds to an Ornstein-Uhlenbeck (OU) process for $V_{xi}(t)$, and the resulting rate function is given by $V_{x_i}(t) = f_{x_i}(t) + \sigma_\xi W_i(t) = V_{x_i}(t_0) - \int_{t_0}^{t} \rho \left\{ V_{x_i}(s) - \bar{\nu}_i(x_i, \beta) \right\} ds + \sigma_\xi W_i(t)$. More details and properties of the OU process can be found in Section 2.2 below. We refer to this specification as SSVM-OU. For the PSA data analysis, it is of interest to estimate the stable rate $\bar{\nu}_i(x_i, \beta)$, since $V_{xi}(t)$ will eventually stabilize and fluctuate around the level given by $\bar{\nu}_i(x_i, \beta)$, which describes the long term rate of tumor growth after radiation treatment. In addition, to address the relationship between the long term tumor growth rate $\bar{\nu}_i(x_i, \beta)$ and the patients' baseline characteristics, we propose a linear model $\bar{\nu}_i(x_i, \beta) = \nu_i + x_i'\beta$, where $\beta = (\beta_0, \beta_1, \ldots, \beta_p)'$ is the vector of fixed effect parameters and $\nu_i \overset{\text{iid}}{\sim} \mathcal{N}\left(0, \sigma_\nu^2\right)$ are random effects. This subject-specific SSVM-OU is very useful to understand the dynamics of tumor growth, to assess the effect of covariates, and to predict a patient's future PSA values using the baseline covariate information.

2.2 The OU and IOU Processes

The OU process was first proposed as a physical model for the velocity of a particle suspended in a fluid (Uhlenbeck and Ornstein, 1930). It describes a homeostasis system that fluctuates around some stable level and has been applied in biology (Trost et al., 2010), finance (Nicolato and Venardos, 2003) and engineering (Kulkarni and Rolski, 2009), among many others. In the statistics literature, Aalen and Gjessing (2004) studied the first-passage time of an OU process, and Taylor and Law (1998) modeled the serial correlation in a linear mixed model by an integrated OU (IOU) process with mean zero. The OU process is particularly suitable for the PSA profiles considered in this paper, where the rate function of tumor growth reaches a stable level that potentially depends on baseline covariates.

Now we present some properties for both the OU and IOU processes. For ease of exposition, we suppress the subject index i in the discussion. Let $U_j := U(t_j)$ and $V_j := V(t_j)$. The IOU and OU processes are given by, respectively,

$$dU(t) = V(t)\, dt, \tag{4}$$

$$dV(t) = -\rho \left\{ V(t) - \bar{\nu} \right\} dt + \sigma_\xi dW(t). \tag{5}$$

Theorem 1 *For IOU and OU processes at time t_j, conditional on the values at time t_{j-1} and parameters $\bar{\nu}$, ρ, σ_ξ, the transition distribution is given by*

$$U_j, V_j | U_{j-1}, V_{j-1}, \bar{\nu}, \rho, \sigma_\xi \sim \mathcal{N}_2\left(\boldsymbol{m}_j, \Sigma_j\right),$$

$\delta_j = t_j - t_{j-1}$, with conditional mean and covariance matrix given, respectively, by,

$$\boldsymbol{m}_j = \left[U_{j-1} + \bar{\nu}\,\delta_j + \left\{ V_{j-1} - \bar{\nu} \right\} \left\{ \frac{1 - \exp\left(-\rho\delta_j\right)}{\rho} \right\}, \bar{\nu} + \left\{ V_{j-1} - \bar{\nu} \right\} \exp\left(-\rho\delta_j\right) \right]',$$

$$\Sigma_j=\sigma_\xi^2\begin{bmatrix} \frac{\delta_j}{\rho^2}+\frac{1}{2\rho^3}\left\{-3+4\exp\left(-\rho\delta_j\right)-\exp\left(-2\rho\delta_j\right)\right\} & \frac{1}{2\rho^2}\left\{1-2\exp\left(-\rho\delta_j\right)+\exp\left(-2\rho\delta_j\right)\right\} \\ \frac{1}{2\rho^2}\left\{1-2\exp\left(-\rho\delta_j\right)+\exp\left(-2\rho\delta_j\right)\right\} & \frac{1}{2\rho}\left\{1-\exp\left(-2\rho\delta_j\right)\right\} \end{bmatrix}.$$

The proof is included in the supplementary materials.

Corollary 1 *For* $\delta_j \to \infty$ *and fixed* $\rho > 0$, *such that* $\exp(-\rho\delta_j) = o(1)$, *then the conditional mean and variance in Theorem 1 can be approximated by,*

$$\boldsymbol{m}_j=\left[U_{j-1}+\bar{\nu}\,\delta_j,\bar{\nu}\right]'+\boldsymbol{R}_{m_j}(1),$$

$$\Sigma_j=\sigma_2^\xi\begin{bmatrix} \frac{\delta_j}{\rho^2}-\frac{3}{2\rho^3} & \frac{1}{2\rho^2} \\ \frac{1}{2\rho^2} & \frac{1}{2\rho} \end{bmatrix}+\boldsymbol{R}_{\Sigma_j}(1),$$

where the errors in the approximation are $\boldsymbol{R}_{m_j}(1)=[o(1),o(1)]'$ *and* $\boldsymbol{R}_{\Sigma_j}(1)=\begin{bmatrix} o(1) & o(1) \\ o(1) & o(1) \end{bmatrix}$. The proof is straightforward by noting that $\rho\delta_j \to \infty$ as δ_j satisfies $\exp(-\rho\delta_j) = o(1)$.

Corollary 2 *For OU and IOU processes with* $\rho > 0$ *and* $\delta_j = o(1)$, *the approximate transition density denoted by* $\prec U_j, V_j|U_{j-1}, V_{j-1}, \bar{\nu}, \rho, \sigma_\xi \succ$ *is given by,*

$$\begin{aligned}\prec U_j, V_j,|U_{j-1}, V_{j-1}, \bar{\nu}, \rho, \sigma_\xi \succ &= \prec V_j|V_{j-1}, \bar{\nu}, \rho, \sigma_\xi \succ \delta\left(U_j-U_{j-1}-V_{j-1}\delta_j\right)\\ &=\phi\left(\tilde{m}_j,\tilde{\Sigma}_j\right)\delta\left(U_j-U_{j-1}-V_{j-1}\delta_j\right)\end{aligned}$$

where $\phi\left(\tilde{m}_j,\tilde{\Sigma}_j\right)$ *is the normal density with mean* $\tilde{m}_j=V_{j-1}-\rho\left\{V_{j-1}-\bar{\nu}\right\}\delta_j$ *and variance* $\tilde{\Sigma}_j=\sigma_\xi^2\delta_j$, *and* $\delta(\cdot)$ *is the Dirac Delta function.*

This corollary can be proved by taking the component-wise first-order Taylor approximation of $\boldsymbol{m}_j$ and Σ_j in Theorem 1 with respect to δ_j.

3 Inference and Forecasting

3.1 Posterior Distribution Approximation

In this section, we present Bayesian estimation for the mean function $U_{x_i}(t)$, the rate function $V_{x_i}(t)$ and parameters ϕ_i and σ_ε for $i = 1,2, –, n$. Let $[\cdot \mid \cdot]$ denote the exact conditional density, $\prec \cdot|\cdot \succ$ the approximate conditional density and $\boldsymbol{U}=\left[\boldsymbol{U}_1',\boldsymbol{U}_2',\cdots,\boldsymbol{U}_n'\right]'$ where $\boldsymbol{U}_i=[\boldsymbol{U}_{i1},\boldsymbol{U}_{i2},\cdots,\boldsymbol{U}_{\mathrm{im}_i}]'$. Similar notation is used for $\boldsymbol{V}$, $\boldsymbol{Y}$ and $\boldsymbol{x}$. For the model specified by equations (1), (2) and (3), we first consider the posterior density $[\phi \mid \boldsymbol{U}, \boldsymbol{V}, \boldsymbol{Y}, \boldsymbol{x}]$ for ϕ, where $\phi=\left[\phi_1',\phi_2',\cdots,\phi_n'\right]'$. The posterior distribution is given by

$$[\phi|\boldsymbol{U},\boldsymbol{V},\boldsymbol{Y},\boldsymbol{x}]\propto\prod_{i=1}^{n}\prod_{j=1}^{m_i}\left[U_{ij},V_{ij}|U_{i,j-1},V_{i,j-1},\phi_i,\boldsymbol{x}_i\right][U_0,V_0][\phi_i],t\in\mathscr{T}_{io} \tag{6}$$

where $[U_{ij}, V_{ij} \mid U_{i;j-1}M, V_{i;j-1}, \phi_i, x_i]$ is the exact transition density derived from the SDE in equation (3) and $[U_0, V_0]$ and $[\phi_i]$ are non-informative prior densities. Unfortunately, except for a very few specific forms for the drift and diffusion terms in equation (3), $[U_{ij}, V_{ij} \mid U_{i;j-1}, V_{i;j-1}, \phi_i, x_i]$ is usually analytically intractable. Even when the exact transition density does have a closed form, as is the case for the OU and IOU processes, for which the exact transition density is given in Theorem 1, the posterior density for ϕ still does not have an explicit form. Hence, we will use the Euler approximation to approximate the exact transition density, while applying the method of data augmentation (Tanner and Wong, 1987) to minimize the error in this approximation.

The strategy of combining data augmentation and Euler approximation to approximate the exact transition density has been discussed by Elerian et al. (2001), Eraker (2001), Roberts and Stramer (2001) and Durham and Gallant (2002), in the context of estimating parameters in the SDE for a single diffusion process observed at discrete times without measurement errors. Our approach is related to theirs, but with an important distinction that instead of being partially observed, both processes $V_{x_i}(t)$ and $U_{x_i}(t)$ are completely unobserved, and will be sampled as part of an MCMC algorithm. In this manner, we will estimate the processes $V_{x_i}(t)$, $U_{x_i}(t)$ and the parameters ϕ. It is worth pointing out that although augmentation only needs to take place for the latent process, augmenting the data themselves will facilitate the operation of the simulation smoother, as this algorithm requires observations (either observed or augmented) available at each corresponding time. In addition, the data augmentation allows us to create augmented longitudinal data with a common set of time points, and consequently this method enables us to handle longitudinal data with irregularly spaced times which may vary across the subjects.

To carry out the data augmentation and the Euler approximation, we first specify time points at which data would be augmented. Let

$$\mathscr{T}_{ia}=\left\{t:t=t_{ij}+k\tau_{ij},\tau_{ij}=\frac{t_{i,j+1}-t_{ij}}{M_{ij}}<\tau_c,\ t\in\left(t_{ij},t_{i,j+1}\right),\ k=1,2,\ldots,M_{ij,j}=1,2,\ldots,m_i-1\right\}$$

denotes the set of augmentation times for the ith subject. Consequently, the time interval τ_{ij} between adjacent data points, either observed or augmented, is less than τ_c. In addition, let $\mathscr{T}=\cup_{i=1}^{n}(\mathscr{T}_{io}\cup\mathscr{T}_{ia})=t:t_j,\ j=1,2,\ldots,m$ denote the set of all possible time points of the observed and augmented data across subjects. With further data augmentation at times $t\in\mathscr{T}_{im}=[t:t\in\mathscr{T},t\notin\mathscr{T}_{io},t\notin\mathscr{T}_{ia}]$, each subject would have either observed or augmented data $\tilde{Y}_i=[Y_{i1},Y_{i2},\cdots,Y_{im}]'$ at the common time set $\mathscr{T}$. The Euler approximation to equations (2) and (3) for $t\in\mathscr{T}$ leads to the following difference equations:

$$U_{ij}=U_{i,j-1}-V_{i,j-1}\delta_j, \tag{7}$$

$$V_{ij}=V_{i,j-1}+a\left\{V_{i,j-1};\boldsymbol{x}_i,\phi_i\right\}\delta_j+b\left\{V_{i,j-1};\boldsymbol{x}_i,\phi_i\right\}\left(W_j-W_{j-1}\right), \tag{8}$$

where $W_j-W_{j-1}\sim\mathscr{N}_1\left(0,\delta_j\right)$ and $j = 1, 2, \ldots, m$. Thus, the conditional posterior density for ϕ is approximated by,

$$\prec \phi \mid \tilde{\boldsymbol{U}}, \tilde{\boldsymbol{V}}, \tilde{\boldsymbol{Y}}, \boldsymbol{x} \succ \propto \prod_{i=1}^{n} \prod_{j=1}^{m} \prec U_{ij}, V_{ij}, \mid U_{i,j-1}, V_{i,j-1}, \phi_i, \boldsymbol{x}_i \succ [U_0, V_0][\phi_i], \tag{9}$$

where $\tilde{\boldsymbol{U}} = \left[\tilde{\boldsymbol{U}}_1', \tilde{\boldsymbol{U}}_2', \ldots, \tilde{\boldsymbol{U}}_n'\right]'$ with $\tilde{\boldsymbol{U}}_i = [U_{i1}, U_{i2}, \ldots, U_{im}]'$ and similarly for $\tilde{V}$ and $\tilde{\boldsymbol{Y}}$. Note that the approximate transition density $\prec U_{ij}, V_{ij} \mid U_{i,j-1}, V_{i,j-1}, \boldsymbol{\phi}_i, \boldsymbol{x}_i \succ$ in equation (9) is given by,

$$\prec U_{ij}, V_{ij} \mid U_{i,j-1}, V_{i,j-1}, \phi_i, \boldsymbol{x}_i \succ = \mathcal{N}_1\left(V_{i,j-1} + a\left\{V_{i,j-1}; \boldsymbol{x}_i, \phi_i\right\}\delta_j, b^2\left\{V_{i,j-1}; \boldsymbol{x}_i, \phi_i\right\}\delta_j\right) \times \delta\left(U_{ij} - U_{i,j-1} - V_{i,j-1}\delta_j\right), \tag{10}$$

which is derived from equations (7) and (8). This implies that it is feasible to directly sample from the posterior distribution of $\boldsymbol{\phi}$, if the conjugate priors for $\boldsymbol{\phi}$ are chosen.

With regard to the posterior samples of $U_{\boldsymbol{x}_i}(t)$ and $V_{\boldsymbol{x}_i}(t)$ for $t \in \mathscr{T}_{s_1}$, we follow equations (7) and (8) to come up with their approximations, denoted by $U_{x_i}^{(m)}(t)$ and $V_{x_i}^{(m)}(t)$, with linear interpolation for t between t_{j-1} and t_j for $j = 1, 2, \ldots, m$. Bouleau and Lepingle (1992) showed that under some regularity conditions, with constant C_i, the L_p-norm of the approximation error for $V_{\boldsymbol{x}_i}(t)$ is bounded at the rate of $\sqrt{\frac{\log m}{m}}$; that is,

$$\left\| \sup_{t \in \mathscr{T}_{s_1}} \mid V_{\boldsymbol{x}_i}(t) - V_{\boldsymbol{x}_i}^{(m)}(t) \mid \right\|_p \leq C_i \left(\frac{1+\log m}{m}\right)^{1/2}, 1 \leq p < \infty,$$

where $\|f\|_p = \{\int_\Omega |f(z)|^p d\mu(z)\}^{1/p}$ for a real function f on the space $(\Omega, \mathscr{A})$ with measure μ on random variable z. This indicates that if m is sufficiently large, then $V_{x_i}^{(m)}(t)$ will approach to its continuous counterpart $V_{\boldsymbol{x}_i}(t)$ with arbitrary precision. Similar arguments hold for $U_{x_i}^{(m)}(t)$. Note that we will sample m instead of m_i data points for $U_{x_i}^{(m)}(t)$ and $V_{x_i}^{(m)}(t)$ with possibly $m \gg m_i$. Hence, the benefit of introducing augmented data is two fold: (i) it reduces the error of approximation, when $U_{x_i}^{(m)}(t)$ or $V_{x_i}^{(m)}(t)$, instead of $U_{x_i}^{(m)}(t)$ or $V_{x_i}^{(m)}(t)$, is used to replace $U_{\boldsymbol{x}_i}(t)$ or $V_{\boldsymbol{x}_i}(t)$; (ii) it gives a more accurate approximation to the exact transition density, as shown by Pedersen (1995), which benefits estimation of model parameters $\boldsymbol{\phi}$. Under the assumption that m is large enough such that the approximation error is small, for the ease of exposition, we still use $V_{\boldsymbol{x}_i}(t)$ instead of $V_{x_i}^{(m)}(t)$ throughout the rest of the paper. $U_{\boldsymbol{x}_i}(t)$ is treated similarly.

In the MCMC algorithm to update the values of $U_{\boldsymbol{x}_i}(t)$ and $V_{\boldsymbol{x}_i}(t)$ for $t \in t_0 \bigcup \mathscr{T}$, we draw samples from

$$\prec U_0, V_0, \tilde{\boldsymbol{U}}, \tilde{\boldsymbol{V}} \mid \tilde{\boldsymbol{Y}}, \boldsymbol{x}, \phi, \sigma_\varepsilon^2 \succ \propto \prod_{i=1}^{n} \prod_{j=1}^{m} \left[\tilde{Y}_{ij} \mid U_{ij}, \sigma_\varepsilon^2\right] \prec U_{ij}, V_{ij} \mid U_{i,j-1}, V_{i,j-1}, \phi_i, \boldsymbol{x}_i \succ \times [U_0, V_0], \tag{11}$$

where $\left[\tilde{Y}_{ij} | U_{ij}, \sigma_\varepsilon^2\right] = \phi\left(U_{ij}, \sigma_\varepsilon^2\right)$, $\prec U_{ij}, V_{ij} \mid U_{i,j-1}, \boldsymbol{\phi}_i, \boldsymbol{x}_i \succ$ is given in equation (10) and $[U_0, V_0]$ is a non-informative prior. Equivalently, the posterior density (11) may be derived from a state space model representation (Durbin and Koopman, 2001), which is a useful

reformulation of the SSVM in equations (1), (2) and (3) when it is discretized using the Euler approximation and data augmentation.

Consider an example where $V_{\boldsymbol{x}_i}(t)$ follows the OU process and $\prec U_{ij}, V_{ij} \mid U_{i,j-1}, \boldsymbol{\phi}_i, \boldsymbol{x}_i \succ$ is given in Corollary 2. Let $\tilde{Y}_j = \left[\tilde{Y}_{1j}, \tilde{Y}_{2j}, \ldots, \tilde{Y}_{nj}\right]'$ denote the observed or augmented data for n subjects at time t_j, and let $\theta = \left[\theta'_{1j}, \theta'_{2j}, \ldots, \theta'_{nj}\right]'$ be the latent states with $\theta_{ij} = \left[U_{x_i}(t_j), V_{x_i}, \bar{\nu}(\boldsymbol{x}_i, \beta)\right]'$. The corresponding SSVM can be expressed as a state space model, given as follows:

$$\tilde{\boldsymbol{Y}}_j = \boldsymbol{F}'_j \theta_j + \varepsilon_j, \quad \varepsilon_j \sim \mathcal{N}_n\left(0, \sigma_\varepsilon^2 \boldsymbol{I}_n\right)$$

$$\theta_j = \boldsymbol{G}_j \theta_{j-1} + \xi_j, \quad \xi_j \sim \mathcal{N}_{3n}\left(0, \sigma_\xi^2 \boldsymbol{I}_n \otimes \Sigma_j\right)$$

where $\boldsymbol{F}_j = \boldsymbol{I}_n \mathrm{I} \otimes \boldsymbol{F}_{ij}$, $\boldsymbol{G}_j = \boldsymbol{I}_n \otimes \boldsymbol{G}_{ij}$, $\boldsymbol{F}_{ij} = [1, 0, 0]'$ with $\otimes$ denoting Kronecker product,

$$\boldsymbol{G}_{ij} = \begin{bmatrix} 1 & \delta_j & 0 \\ 0 & 1 - \rho\delta_j & \rho\delta_j \\ 0 & 0 & 1 \end{bmatrix}, \quad \Sigma_j = \begin{bmatrix} 0 & 0 & 0 \\ 0 & \delta_j & 0 \\ 0 & 0 & 0 \end{bmatrix}.$$

Likewise, when $V_{\boldsymbol{x}_i}(t)$ follows a Wiener process, the corresponding reformulation as a state space model can be obtained in a similar manner.

In this paper we have adopted the MCMC method for Bayesian inference. In the literature other likelihood-based or sampling-based methods have also been developed for nonlinear and/or non-Gaussian state space models, including Kitagawa's (1987) numeric algorithm using piecewise linear approximation, Durbin and Koopman's (1997) simulated maximum likelihood estimation, Jørgensen et al.'s (1999) Kalman estimating equations and some recent work on sequential Monte Carlo methods using particulate filtering (Gordon et al., 1993; Pitt and Shephard, 1999; Liu, 2008; Andrieu et al., 2010), among others.

3.2 MCMC Algorithm

Under the state space model formulation, Gibbs sampler was first developed to sample one latent state θ_j at a time, this was later improved by various algorithms that use simultaneous block-based sampling schemes (e.g. Frühwirth-Schnatter 1994; Carter and Kohn 1994). The simulation smoother proposed first by de Jong and Shephard (1995) and later improved by Durbin and Koopman (1997) provides a remarkably efficient sampling tool. It draws samples of θ_j through sampling independent innovations ξ_j, rather than realizations of a Markov process, so the entire sampling is based on very low dimensional distributions and free of autocorrelation. Thus, the rate of mixing and moreover burn-in can be achieved quickly. We will use the simulation smoother in our implementation.

The proposed MCMC algorithm iterates through the following steps.

1. Draw augmented data according to $Y_i(t) \sim \mathcal{N}\left(U_{x_i}(t), \sigma_\varepsilon^2\right)$ at times $t \in \mathcal{T}_{ia} \bigcup \mathcal{T}_{im}$ for the ith subject, $i = 1, 2, \ldots, n$.

2. Update latent states $U_{xi}(t)$ and $V_{xi}(t)$ for $t \in t_0 \bigcup \mathscr{T}$ from the posterior density (11) by using the simulation smoother.

3. Update ϕ by sampling from the posterior density (9). In particular, when $V_{xi}(t)$ follows an OU process and is discretized through the Euler approximation, the collection of equations (8) can be equivalently reformulated as a linear mixed model,

$$\boldsymbol{Y}_j^* = \boldsymbol{X}_j^* \beta^* + \boldsymbol{Z}_j^* \boldsymbol{b}^* + \xi_j^*,$$

where $\boldsymbol{Y}_j^* = \frac{V_j - V_j - 1}{\sqrt{\delta_j}}$, $\boldsymbol{X}_j^* = \left[\boldsymbol{X}\ \sqrt{\delta_j},\ \boldsymbol{V}_{j-1}\sqrt{\delta_j}\right]$, $\boldsymbol{Z}_j^* = -\sqrt{\delta_j}\boldsymbol{I}_n$ with $\boldsymbol{V}_j = [V_{1j}, V_{2j}, \ldots, V_{nj}$ and $\boldsymbol{X} = \left[x_1', x_2', \ldots, x_n'\right]'$. Further, $\boldsymbol{\beta^*} = [\rho\beta', -\rho]'$, $\boldsymbol{b^*} = \rho\boldsymbol{\nu}$, $\boldsymbol{\nu} = [\nu_1, \nu_2, \ldots, \nu_n]'$, $\xi_j^* \sim \mathscr{N}_n\left(0,\ \sigma_\xi^2 \boldsymbol{I}_n\right)$, $\boldsymbol{b}* \sim \mathscr{N}_n\left(0,\ \rho^2\sigma_\nu^2 \boldsymbol{I}_n\right)$. As a result, the set of model parameters is $\phi* = \left[\beta*,\ \boldsymbol{b}*,\ \sigma_\xi^2,\ \rho^2\sigma_\nu^2\right]'$, can be sampled straightforwardly by using the standard Gibbs sampler in the linear mixed model (Ruppert et al., 2003, Chap. 16) with non-informative conjugate priors, $\beta* \sim \mathscr{N}_{p+2}\left(0,\ \sigma_{\beta*}^2 \boldsymbol{I}_{p+2}\right)$, $\sigma_\xi^2 \sim \mathscr{IG}\left(a_\xi,\ b_\xi\right)$, and $\rho^2\sigma_\nu^2 \sim \mathscr{IG}\left(a_\sigma,\ b_\sigma\right)$. Here $\mathscr{IG}(a,\ b)$ denotes the inverse gamma distribution with shape parameter a and scale parameter b.

4. Update σ_ε^2 by sampling from the following posterior distribution

$$\left[\sigma_\varepsilon^2 \mid \tilde{\boldsymbol{U}}, \tilde{\boldsymbol{V}}, \tilde{\boldsymbol{Y}}, \boldsymbol{x}\right] \sim \mathscr{IG}\left(a_\varepsilon + \frac{1}{2}mn,\ b_\varepsilon + \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{m}\left(Y_i\left(t_j\right) - U_{xi}\left(t_j\right)\right)^2\right),$$

where the prior distribution for σ_ε^2 is $\mathscr{IG}\left(a_\varepsilon,\ b_\varepsilon\right)$.

3.3 Bayesian Posterior Forecasting

The proposed model is useful to forecast processes of interest, including $U_{xi}(t)$, $V_{xi}(t)$ and $Y_i(t)$, for $t \in \mathscr{T}_{s2} = \{t : t > t_m\}$. With the availability of posterior samples for $U_{xi}(t)$, $V_{xi}(t)$, ϕ_i and σ_ε with $i = 1, 2, \ldots, n$ and $t \in \mathscr{T}$, it is straightforward to derive Bayesian posterior forecasting. Note that the posterior forecasting distributions are,

$$\left[U_{xi}(t), V_{xi}(t) \mid \boldsymbol{Y}, \boldsymbol{x}\right] = \int\int\int \left[U_{xi}(t), \left[V_{xi}(t) \mid U_{xi}(t_m), V_{xi}(t_m), \phi_i, \boldsymbol{x}\right] \times \left[U_{xi}(t_m), V_{xi}(t_m), \phi_i \mid \boldsymbol{Y}, \boldsymbol{x}\right] dU_{xi}(t_m)\, dV_{xi}(t_m)\, d\phi_i,$$

and

$$\left[Y_i(t) \mid \boldsymbol{Y}, x\right] = \int\int\int \left[Y_i(t) \mid U_{xi}(t), \sigma_\varepsilon^2\right] \left[U_{xi}(t), V_{xi}(t) \mid \boldsymbol{Y}, \boldsymbol{x}\right] \times \left[\sigma_\varepsilon^2 \mid \boldsymbol{Y}, \boldsymbol{x}\right] dU_{xi}(t)\, dV_{xi}(t)\, d\sigma_\varepsilon^2,$$

Thus, we draw $U_{x_i}^r(t)$, $V_{x_i}^r(t)$ and $Y_i^r(t)$ from $\left[U_{x_i}^r(t),\ V_{x_i}^r(t) \mid U_{x_i}^r(t_m),\ V_{x_i}^r(t_m),\ \phi_{is}^r,\ x\right]$ and $\left[Y_i^r(t) \mid U_{x_i}^r(t),\ \sigma_\varepsilon^{2r}\right]$ for $r = 1,2, \ldots$, where $U_{x_i}^r(t_m)$, $V_{x_i}^r(t_m)$, ϕ_{is}^r and σ_ε^{2r} are the rth posterior samples from the MCMC algorithm. If $[U_{xi}(t), V_{xi}(t) \mid U_{xi}(t_m), V_{xi}(t_m), \phi_i, \boldsymbol{x}]$ does not have a

closed form, the approximate transition density $\prec U_{x_i}(t), V_{x_i}(t) \mid U_{x_i}(t_m), V_{x_i}(t_m), \phi_i, \boldsymbol{x} \succ$ could be used instead along with data augmentation.

4 Application to the PSA Data

We apply the proposed SSVM-OU to analyze the PSA data discussed in Section 1. The prior of the rate function $V_{x_i}(t)$ is assumed to be the OU process with $a\{V_{x_i}(t); x_i, \phi_i\} = -\rho\left\{V_{x_i}(t) - \bar{\nu}_i(\boldsymbol{x}_i, \beta)\right\}$ and $b\{V_{x_i}(t); \boldsymbol{x}_i, \phi_i\} = \sigma_\xi$ in equation (3). A total of 739 observations are obtained for 50 subjects. The number of observations for each subject varies from 13 to 24. The initial observation for all subjects is at one month (0.083 years) after EBRT treatment, and the time for the last observation ranges from 3.833 to 8.083 years, with the average of 6.050 years. To reduce the approximation error discussed in Section 3.1, we further augment the data to let the time interval between adjacent data points, either observed or augmented, be less than 0.0208 years. The appropriateness of this choice of time interval is confirmed using the simulation studies in Section 5. We investigate the association of the pretreatment covariates (i.e. baseline PSA, Gleason score and T-stage) with the stable PSA rate via the model $\bar{\nu}_i(\boldsymbol{x}_i, \beta) = \nu_i + \beta_0 + \beta_1 X_{P_i} + \beta_2 X_{T_i} + \beta_3 X_{G_i}$, where $\nu_i \sim \mathcal{N}_1\left(0, \sigma_\nu^2\right)$ is a random effect; X_{Pi} denotes the log-transformed baseline PSA for the ith subject, centered around the mean of 2.3; X_{Gi} is equal to 1 if Gleason score is above or equal to level 7, and is 0 otherwise; X_{Ti} takes the value of 1 if T-stage is at level 2 or higher, and is 0 otherwise. We leave out the last observation for each subject as well as the observations after year 5 as validation data to assess the forecasting ability of the model.

The posterior draws are obtained from the proposed MCMC algorithm with 20,000 iterations, discarding the first 10,000 as the burn-in stage and subsequently saving every 10th draws. The trace plots suggest the algorithm converges fast and mixes well. Table 1 presents the posterior summary statistics for the parameters. Baseline PSA and T-stage are found to have significant effect on the PSA stable rate. This result suggests that Baseline PSA and T-stage are predictive of the long term rate of change for PSA, which is in agreement with the finding by Lieberfarb et al. (2002). Figure 2 displays $E[V_{x_i}(t) \mid \boldsymbol{Y}]$, the posterior means of the rate function for each subject (shown as dashed lines), and $E[V(t) \mid \boldsymbol{Y}] = E[E[V_{x_i}(t) \mid \boldsymbol{Y}]]$, the posterior mean of the rate function in the population (shown as a solid line). It is clear that although the rate function for the population is smooth and may be specified by a parametric form, the individual rate functions are much more wiggly, vary significantly across subjects and would be difficult to model parametrically. Figure 3 shows the posterior means and credible intervals of $U_{x_i}(t)$ for six randomly selected subjects, including the forecasted $U_{x_i}(t)$ after year 5. Note that the width of the forecasted credible intervals is comparable to the theoretical results given in Corollary 1.

For comparison, we also analyze the PSA data using smoothing splines and a parametric linear mixed-effects model(LMM). The model fits are evaluated by the Deviance Information Criterion (DIC, Speigelhalter et al., 2003). Note that $DIC = \bar{D} + p_D$, where the posterior mean deviance $\bar{D}$ measures the goodness of fit and the "effective number of parameters" p_D measures the model complexity. According to Speigelhalter et al. (2003), DIC may be regarded asymptotically as a generalization of the Akaike information criterion (AIC). Similar to AIC, a smaller value of DIC indicates a better trade-off between the fit to the data and the complexity of model. We further compare the forecasting ability of these three models on the validation data points. For the smoothing spline approach, we obtain the estimates of $V_{x_i}(t)$ from the SSVM-W with a Wiener process as the prior for $V_{x_i}(t)$, where $a\{V_{x_i}(t); \boldsymbol{x}_i, \phi_i\} = 0$ and $b\{V_{x_i}(t); \boldsymbol{x}_i, \phi_i\} = \sigma_\xi$ in equation (3). As mentioned in Section 2.1, the estimation of $V_{x_i}(t)$ from this model, is equivalent to estimation by a smoothing spline with a

common smoothing parameter $\lambda=\frac{\sigma_\xi^2}{\sigma_\xi^2}$. The exact transition density in this SSVM-W, is given by Wecker and Ansley (1983) as

$$\left[U_i, V_j \mid U_{j-1}, V_{j-1}, \sigma_\xi\right] \sim \mathcal{N}_2\left(\boldsymbol{m}_i, \boldsymbol{V}_j\right),$$

with

$$\boldsymbol{m}_j=\left[U_{j-1}+V_{j-1}\delta_j, V_{j-1}\right]',$$

$$\boldsymbol{V}_j=\sigma_\xi^2\begin{bmatrix} \frac{\delta_j^3}{3} & \frac{\delta_j^2}{2} \\ \frac{\delta_j^2}{2} & \delta_j \end{bmatrix},$$

and is used in the proposed MCMC algorithm. The forecasting of future observations is outlined in Section 3.3 for the SSVM-OU and SSVM-Ws. The parametric linear mixed model is similar to the one given by Proust-Lima et al. (2008),

$$\begin{aligned} Y_i\left(t_{ij}\right) &= U_{\boldsymbol{x}_i}\left(t_{ij}\right)+\varepsilon_i\left(t_{ij}\right) \\ &= U^0_{\boldsymbol{x}_i}\left(t_{ij}\right)+U^1_{\boldsymbol{x}_i}\left(t_{ij}\right)+U^2_{\boldsymbol{x}_i}\left(t_{ij}\right)+\varepsilon_i\left(t_{ij}\right) \\ &= (\beta_{00}+\nu_{0i}+\beta_{01}X_{Pi})+(\beta_{10}+\nu_{1i}+\beta_{11}X_{Pi}+\beta_{12}X_{Ti})\,f_1\left(t_{ij}\right)+(\beta_{20}+\nu_{2i}+\beta_{21}X_{Pi}+\beta_{22}X_{Ti}+\beta_{23}X_{Gi})\,f_2\left(t_{ij}\right)+\varepsilon_i\left(t_{ij}\right), \end{aligned} \tag{12}$$

where the mean function $U_{x_i}(t)$ consists of three parts: (i) post-therapy level $U^0_{x_i}(t)$, (ii) short-term evolution $U^1_{x_i}(t)$, and (iii) long-term evolution $U^2_{x_i}(t)$. In addition, $f_1(t) = (1 + t)^{-1.5} - 1$ and $f_2(t) = t$; the fixed effects $\beta_{lmm}=[\beta_{00},\beta_{01},\beta_{10},\beta_{11},\beta_{12},\beta_{20},\beta_{21},\beta_{22},\beta_{23}]' \sim \mathcal{N}_9\left(0,\sigma^2_{\beta,lmm}\boldsymbol{I}_9\right)$ a non-informative prior with large value of $\sigma^2_{\beta,lmm}$; the random effects $[\nu_{0i},\nu_{1i},\nu_{2i}]' \sim \mathcal{N}_3\left(0,\Sigma^2_{2\nu,lmm}\right)$ where $\Sigma_{\nu,lmm}$ is a diagonal matrix with its main diagonal entries $\nu_{lmm}=\left[\sigma^2_{0\nu,lmm},\sigma^2_{1\nu,lmm},\sigma^2_{3\nu,lmm}\right]'$; measurement error $\varepsilon_i\left(t_{ij}\right) \sim \mathcal{N}_1\left(0,\sigma^2_{\varepsilon,lmm}\right)$. We further assume noninformative prior distributions $\mathcal{IG}$ with small values of a and b for $\sigma^2_{\beta,lmm}$, $\sigma^2_{0\nu,lmm}$, $\sigma^2_{1\nu,lmm}$, $\sigma^2_{2\nu,lmm}$, and $\sigma^2_{\varepsilon\nu,lmm}$ respectively. The MCMC algorithm for the linear mixed model (Ruppert et al., 2003, Chap. 16) is applied to draw the posterior samples with the same burn-in stage and thinning scheme as for the MCMC algorithm for the SSVM-OU. Table 1 presents the posterior summary of the parameters β_{20}, β_{21}, β_{22}, and β_{23}, which are involved in the long-term evolution $U^2_{x_i}(t)$ in equation (12). Note that these parameters in the LMM are designed to measure the association between the long term stable level and the covariates of the interest, similar to the parameter β_0, β_1,β_2, and β_3 in the SSVM-OU. Given the rth samples β^r_{lmm}, ν^r_{lmm} and $\sigma^{2\,r}_{\varepsilon,lmm}$, the forecasts of PSA at time t for the ith subject can be drawn from $Y^r_i(t) \sim \mathcal{N}\left(U^r_{x_i}(t),\sigma^{2\,r}_{\varepsilon,lmm}\right)$, where

$$U^r_{x_i}(t)=\left(\beta^r_{00}+\nu^r_{0i}+\beta^r_{01}X_{Pi}\right)+\left(\beta^r_{10}+\nu^r_{1i}+\beta^r_{11}X_{Pi}+\beta^r_{12}X_{Ti}\right)f_1(t)+\left(\beta^r_{20}+\nu^r_{2i}+\beta^r_{21}X_{Pi}+\beta^r_{22}X_{Ti}+\beta^r_{23}X_{Gi}\right)f_2(t)$$

.

The values of DIC for SSVM-OU and SSVM-W are 71.809 and 119.400 respectively, both of which are significantly lower than that of LMM (151.048). Thus, SSVM-OU fits the data best among these three models. This implies that the parametric LMM is less able to capture longitudinal dynamics of subject's trajectories than the other two SSVMs. Next, to compare the prediction capability among these three models, we predict the 164 validation data points and evaluate their posterior predictive ability. Table 2 presents relative bias and mean squared error (MSE) of the point forecast based on the posterior mean, as well as corresponding coverage rate and averaged length of credible interval. For the 69 validation data points within 1 year distance from the last training data points, the SSVM-OU performs best, with the smallest MSE. For the remaining validation data points at later times, the SSVM-W outperforms the other two in terms of relative bias and MSE. However, for the coverage rate, the SSVM-OU intervals are closest to the nominal 95% level, whereas those from the SSVM-W are too wide to be clinically useful. This may be due to the nonstationary variance of the latent process of SSVM-W.

Besides evaluation of the point forecasts and the corresponding credible intervals, we further use the probability integral transform (PIT, Dawid, 1984; Gneiting et al., 2007) value to assess the predictive performance of the probabilistic forecasts. This forecast can be expressed as the posterior predictive cumulative distribution functions (CDFs) $F_{ij}(Y)$, where Y is the forecasted validation data point at time t_{ij} for the ith subject and is assumed to be generated from the true unknown CDF $G_{ij}(Y)$. For the observed validation data point Y_{ij}, the PIT value $p_{ij} = F_{ij}(Y_{ij})$ should have a uniform distribution, if $F_{ij}(Y) = G_{ij}(Y)$ for every i and j. We estimate $F_{ij}(t)$ by the empirical CDF $\tilde{F}_{ij}(Y)$, which is based on the Bayesian posterior forecasting draws for the three models. The corresponding smoothed density plots of $\tilde{p}_{ij}$ are displayed in Figure 4. The density of $\tilde{p}_{ij}$ for the SSVM-OU is left skewed, indicating the forecasts are slightly under predicted, while the density for the linear mixed model is right skewed and the forecasts are slightly over predicted. The density for the SSVM-W is hump-shaped, implying the posterior predictive distribution is over dispersed and the credible intervals are too wide on average. While none of the models gives the ideal PIT plots, the plots of SSVM-OU and the LMM are reasonably close to a uniform density..

5 A Simulation Study

We carry out a simulation study to (i) assess the performance of the proposed MCMC algorithm in estimating the model parameters and stable rates $\bar{\nu}_i(\boldsymbol{x}_i, \beta)$ and (ii) compare the performance of the proposed SSVM-OU with the other two methods for forecasting future observations. We generate 100 replicated datasets from the SSVM-OU with the model parameter set close to those estimated from the analysis of the PSA data. Each dataset includes 20 subjects each with 13 observations and three validation data points per subject. The observations are equidistantly spaced with time interval 0.416, equal to the median of time intervals in the PSA data. The three validation data points are at 0.08, 0.5 and 1 years after the last observation, respectively. To investigate the influence of data augmentation on the estimation of the model parameters, we analyze the same dataset using the proposed MCMC algorithm without data augmentation, and with 9 and 19 augmented data points between the consecutive observed data points. The corresponding time interval between the adjacent data points, either observed or augmented, decreases from 0.416 in the original datasets to 0.0416 and 0.0208 for the MCMC algorithm with 9 and 19 augmented data points between neighboring observations.

Table 3 presents simulation results for the estimation of model parameters, assessed by relative bias, MSE of posterior means, coverage rate and average length of credible interval. All results indicate clearly that the data augmentation is critical to obtain proper estimates of

the second moment parameters, σ_{ε}^2, σ_{ξ}^2, σ_{ν}^2 and ρ. Their relative biases and MSEs decrease significantly even by adding 9 data points between adjacent observations. For example, the relative bias of ρ reduces from 0.47 to 0.052 and the MSE drops from 2.704 to 0.0360. Augmentation with 19 data points can further improve the relative bias in the estimation of parameters σ_{ξ}^2 and ρ, and no additional improvement results from more aggressive augmentation (the results not shown here). The data augmentation, however, has little effect on the relative bias for the estimation of parameters of interest, β_1, β_2 and β_3, implying that the consistent estimation for these parameters may be obtained using observed data. Yet, the data augmentation has noticeable effects on the coverage rates, because it affects the variance of posterior distributions.

For the data simulated from the SSVM-OU, we further forecast the validation data points by the SSVM-OU, SSVM-W and LMM (12). Table 4 compares the forecasting ability of the posterior mean and credible intervals for the three models, evaluated by the relative bias, MSE, coverage rate and interval length. As we expected, the relative biases of posterior means of the forecasting draws from the SSVM-OU are smaller than those from the other models and the corresponding interval lengths are narrower. Furthermore, it is of interest to study the sensitivity of the forecasting ability of SSVM-OU. We simulate another 100 datasets from the LMM specified as equation (12) in which the parameters are the same as those obtained from the PSA data analysis. In addition, the number of subjects, and the number of observations and the validation data points, are set identical to those used to generate datasets from the above SSVM-OU. The forecasting results are given in the second part of Table 4. We find that SSVM-OU has comparable performance to the LMM (for the short-term forecast at time 0.08), with smaller relative bias but slightly larger MSE and wider interval length. For the long-term forecast at time 0.5 or 1, SSVM-OU performs worse than the LMM but is better than SSVM-W.

6 Discussion

This paper considers modeling and inference for the rate functions in longitudinal studies with an application in the analysis of PSA biomarker profiles. For a given subject, the rate of change is described by a rate function whose prior is assumed to follow a Gaussian process conditional on the covariates. A key feature of this approach is that the Gaussian process is specified by an SDE and is expected to be centered on a pre-specified parametric function, while allowing significant deviations from this functional expectation nonparametrically. We have focused on the case where the rate function follows an OU process, motivated by analyzing PSA profiles. The same modeling strategy and inference method should be widely useful in the setting when we aim to model the rate function semiparametrically.

One can extend our model to discrete outcomes and to include the covariates in equation (1). Moreover, a similar modeling and inference approach can be applied to analyze the acceleration function, which is the second-order derivative of the mean function. In addition, for simplicity, we assume the stable rates depend on the covariates through a parametric distribution, which could potentially be replaced by a nonparametric distribution with a stick-breaking process as its prior.

The MCMC algorithm is currently programmed in R (R Development Core Team, 2008). For the PSA application with 50 subjects and 225 observed or augmented data points per subject, it took about 4 hours per 1000 MCMC iterations on a PC with 2.93GHz Intel(R)Core(TM)2Duo CPU. In contrast, it took about 15 minutes per 1000 MCMC iterations if the model was fit with only observed data. One way to speed up computation is to develop a C or C++ program for the proposed method, which is one of our future research

tasks. Our computation-related experiences have suggested that the computation time is approximately linearly proportional to the number of subjects. Hence, we anticipate that with fast computation software this algorithm can be applied to handle studies with relatively large sample sizes.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This research of the second author was partially supported by National Cancer Institute grants CA110518 and CA69568, and the research of the third author was partially supported by National Science Foundation (DMS0904177). The authors are thankful to the first author's dissertation committee members Dr. Naisysin Wang and Dr. Brisa Sanchez for the helpful discussions, and to the editor, associate editor and two anonymous reviewers for valuable comments and suggestions.

References

Aalen OO, Gjessing HK. Survival models based on the Ornstein-Uhlenbeck process. Lifetime Data Analysis. 2004; 10:407–423. [PubMed: 15690993]

Andrieu C, Doucet A, Holenstein R. Particle Markov chain Monte Carlo methods. Journal of the Royal Statistical Society: Series B (Statistical Methodology. 2010; 72:269–342.

Ansley CF, Kohn R. On the Equivalence of Two Stochastic Approaches to Spline Smoothing. Journal of Applied Probability. 1986; 23:391–405.

Bouleau, N.; Lepingle, D. Numerical Methods for Stochastic Process. Wiley; New York: 1992.

Carter CK, Kohn R. On Gibbs sampling for state space models. Biometrika. 1994; 81:541–553.

Dawid AP. Present position and potential developments: Some personal views: Statistical theory: The prequential approach. Journal of the Royal Statistical Society: Series A (Statistics in Society). 1984; 147:278–292.

de Jong P, Shephard N. The simulation smoother for time series models. Biometrika. 1995; 82:339–350.

Diggle, PJ.; Heagerty, P.; Liang, KY.; Zeger, S. Analysis of longitudinal data. Oxford University Press; Oxford: 2002.

Durbin J, Koopman SJ. Monte Carlo maximum likelihood estimation for non-Gaussian state space models. Biometrika. 1997; 84:669–684.

Durbin, J.; Koopman, SJ. Time Series Analysis by State Space Methods. Oxford University Press; Oxford: 2001.

Durham GB, Gallant AR. Numerical techniques for maximum likelihood estimation of continuous-time diffusion processes. Journal of Business & Economic Statistics. 2002; 20:297–338.

Elerian O, Chib S, Shephard N. Likelihood inference for discretely observed nonlinear diffusions. Econometrica. 2001; 69:959–993.

Eraker B. MCMC Analysis of Diffusion Models With Application to Finance. Journal of Business & Economic Statistics. 2001; 19:177–191.

Feller, W. An Introduction to Probability Theory and Its Application. Springer Verlag; New York: 1970.

Frühwirth-Schnatter S. Data augmentation and dynamic linear models. Journal of Time Series Analysis. 1994; 15:183–202.

Gneiting T, Balabdaoui F, Raftery AE. Probabilistic forecasts, calibration and sharpness. Journal of the Royal Statistical Society: Series B (Statistical Methodology). 2007; 69:243–268.

Gordon NJ, Salmond DJ, Smith AFM. Novel approach to nonlinear/non-Gaussian Bayesian state estimation. Radar and Signal Processing, IEE Proceedings F. 1993; 140:107–113.

Grimmett, G.; Stirzaker, D. Probability and Random Processes. Oxford University Press; Oxford: 2001.

Guo W. Functional mixed effects models. Biometrics. 2002; 58:121–128. [PubMed: 11890306]

Hastie T, Tibshirani R. Varying-coefficient models. Journal of the Royal Statistical Society: Series B (Statistical Methodology). 1993; 55:757–796.

Hoover DR, Rice JA, Wu CO, Yang LP. Nonparametric smoothing estimates of time-varying coefficient models with longitudinal data. Biometrika. 1998; 85:809–822.

Jørgensen B, Lundbye-Christensen S, Song PX-K, Sun L. A state space model for multivariate longitudinal count data. Biometrika. 1999; 86:169–181.

Kariyanna SS, Light RP, Agarwal R. A longitudinal study of kidney structure and function in adults. Nephrology Dialysis Transplantation. 2010; 25:1120–1226.

Kitagawa G. Non-Gaussian state-space modeling of nonstationary time series. Journal of the American Statistical Association. 1987; 82:1032–1041.

Kulkarni V, Rolski T. Fluid model driven by an Ornstein-Uhlenbeck process. Probability in the Engineering and Informational Sciences. 2009; 8:403–417.

Laird NM, Ware JH. Random-effects models for longitudinal data. Biometrics. 1982; 38:963–974. [PubMed: 7168798]

Lieberfarb ME, Schultz D, Whittington R, Malkowicz B, Tomaszewski JE, Weinstein M, Wein A, Richie JP, D'Amico AV. Using PSA, biopsy Gleason score, clinical stage, and the percentage of positive biopsies to identify optimal candidates for prostate-only radiation therapy. International Journal of Radiation Oncology Biology Physics. 2002; 53:898–903.

Liu, JS. Monte Carlo strategies in scientific computing. Springer Verlag; New York: 2008.

Lloyd-Jones DM, Liu K, Colangelo LA, Yan LL, Klein L, Loria CM, Lewis CE, Savage P. Consistently stable or decreased body mass index in young adulthood and longitudinal changes in metabolic syndrome components: the Coronary Artery Risk Development in Young Adults Study. Circulation. 2007; 115:1004–1011. [PubMed: 17283263]

Morris JS, Carroll RJ. Wavelet-based functional mixed models. Journal of the Royal Statistical Society: Series B (Statistical Methodology). 2006; 68:179–199.

Müller HG, Yao F. Empirical dynamics for longitudinal data. The Annals of Statistics. 2010; 38:3458–3486.

Mungas D, Harvey D, Reed BR, Jagust WJ, DeCarli C, Beckett L, Mack WJ, Kramer JH, Weiner MW, Schuff N, et al. Longitudinal volumetric MRI change and rate of cognitive decline. Neurology. 2005; 65:565–571. [PubMed: 16116117]

Nicolato E, Venardos E. Option pricing in stochastic volatility models of the Ornstein-Uhlenbeck type. Mathematical Finance. 2003; 13:445–466.

Paul, D.; Peng, J.; Burman, P. Semiparametric modeling of autonomous nonlinear dynamical systems with applications. 2009. submitted

Pedersen AR. A new approach to maximum likelihood estimation for stochastic differential equations based on discrete observations. Scandinavian Journal of Statistics. 1995; 22:55–71.

Pitt MK, Shephard N. Filtering Via Simulation: Auxiliary Particle Filters. Journal of the American Statistical Association. 1999; 94:590–591.

Proust-Lima C, Taylor JMG, Williams S, Ankerst D, Liu N, Kestin L, Bae K, Sandler H. Determinants of change in prostate-specific antigen over time and its association with recurrence after external beam radiation therapy for prostate cancer in five large cohorts. International Journal of Radiation Oncology Biology Physics. 2008; 72:782–791.

Qin L, Guo W. Functional mixed-effects model for periodic data. Biostatistics. 2006; 7:225–234. [PubMed: 16207823]

R Development Core Team. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing; Vienna, Austria: 2008. ISBN 3-900051-07-0

Rice JA, Silverman BW. Estimating the mean and covariance structure nonparametrically when the data are curves. Journal of the Royal Statistical Society: Series B (Statistical Methodology). 1991; 53:233–243.

Roberts GO, Stramer O. On inference for partially observed nonlinear diffusion models using the Metropolis-Hastings algorithm. Biometrika. 2001; 88:603–621.

Ruppert, D.; Wand, MP.; Carroll, RJ. Semiparametric Regression. Cambridge University Press; Cambridge: 2003.

Sartor CI, Strawderman MH, Lin XH, Kish KE, McLaughlin PW, Sandler HM. Rate of PSA rise predicts metastatic versus local recurrence after definitive radiotherapy. International Journal of Radiation Oncology Biology Physics. 1997; 38:941–947.

Speigelhalter DJ, Best NG, Carlin BP, van der Linde A. Bayesian measures of model complexity and fit (with discussion). Journal of the Royal Statistical Society B: (Statistical Methodology). 2003; 64:583–616.

Strasak AM, Kelleher CC, Klenk J, Brant LJ, Ruttmann E, Rapp K, Concin H, Diem G, Pfeiffer KP, Ulmer H. Longitudinal change in serum gamma-glutamyltransferase and cardiovascular disease mortality: a prospective population-based study in 76 113 Austrian adults. Arteriosclerosis, Thrombosis, and Vascular Biology. 2008; 28:1857–1865.

Tanner MA, Wong WH. The calculation of posterior distributions by data augmentation. Journal of the American Statistical Association. 1987; 82:528–540.

Taylor JMG, Law N. Does the covariance structure matter in longitudinal modelling for the prediction of future CD4 counts? Statistics in Medicine. 1998; 17:2381–2394. [PubMed: 9819834]

Trost DC, Overman EA II, Ostroff JH, Xiong W, March P. A model for liver homeostasis using modified mean-reverting Ornstein-Uhlenbeck process. Computational and Mathematical Methods in Medicine. 2010; 11:27–47.

Uhlenbeck GE, Ornstein LS. On the Theory of the Brownian Motion. Physical Review. 1930; 36:823–841.

Verbeke, G.; Molenberghs, G. Linear Mixed Models for Longitudinal Data. Springer Verlag; New York: 2009.

Verbyla AP, Cullis BR, Kenward MG, Welham SJ. The analysis of designed experiments and longitudinal data by using smoothing splines. Journal of the Royal Statistical Society: Series C (Applied Statistics). 1999; 48:269–311.

Wahba G. Improper Priors, Spline Smoothing and the Problem of Guarding Against Model Errors in Regression. Journal of the Royal Statistical Society B: (Statistical Methodology). 1978; 40:364–372.

Wang S, Jank W, Shmueli G, Smith P. Modeling price dynamics in eBay auctions using differential equations. Journal of the American Statistical Association. 2008; 103:1100–1118.

Wang Y, Taylor JMG. Inference for smooth curves in longitudinal data with application to an AIDS clinical trial. Statistics in Medicine. 1995; 14:1205–1205. [PubMed: 7667561]

Wecker WE, Ansley CF. The Signal Extraction Approach to Nonlinear Regression and Spline Smoothing. Journal of the American Statistical Association. 1983; 78:81–89.

Welham SJ, Cullis BR, Kenward MG, Thompson R. The analysis of longitudinal data using mixed model L-splines. Biometrics. 2006; 62:392–401. [PubMed: 16918903]

Zeger SL, Diggle PJ. Semiparametric models for longitudinal data with application to CD4 cell numbers in HIV seroconverters. Biometrics. 1994; 50:689–699. [PubMed: 7981395]

Zhang D, Lin XH, Raz J, Sowers M. Semiparametric Stochastic Mixed Models for Longitudinal Data. Journal of the American Statistical Association. 1998; 93:710–719.

Zhang P, Song PX-K, Qu A, Greene T. Effcient estimation for patient-specific rates of disease progression using nonnormal linear mixed models. Biometrics. 2008; 64:29–38. [PubMed: 17501938]

Zhu, B.; Song, PX-K.; Taylor, JMG. Biometrics. 2011. Stochastic Functional Data Analysis: A Diffusion Model-Based Approach. In Press

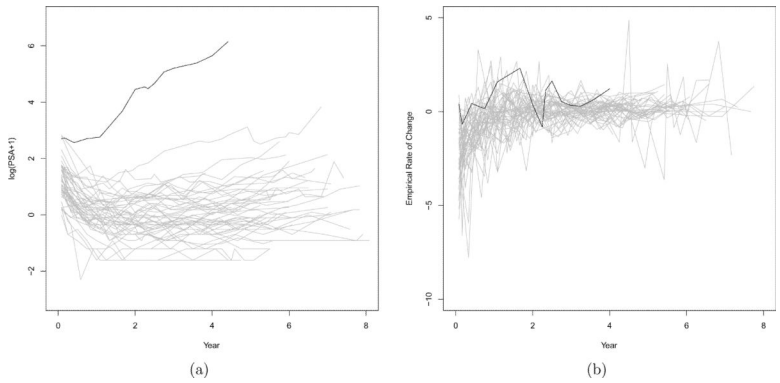

Figure 1.
PSA plots of (a) the raw data, (b) the empirical rate of change, which is defined as $\frac{\Delta Y_{ij}}{\Delta t_{ij}} = \frac{Y_{ij} - Y_{i,j-1}}{t_{ij} - t_{i,j-1}}$, for the give subject i with observation Y_{ij} at time t_{ij}. All profiles are plotted as the gray solid lines, except one profile highlighted in black color.

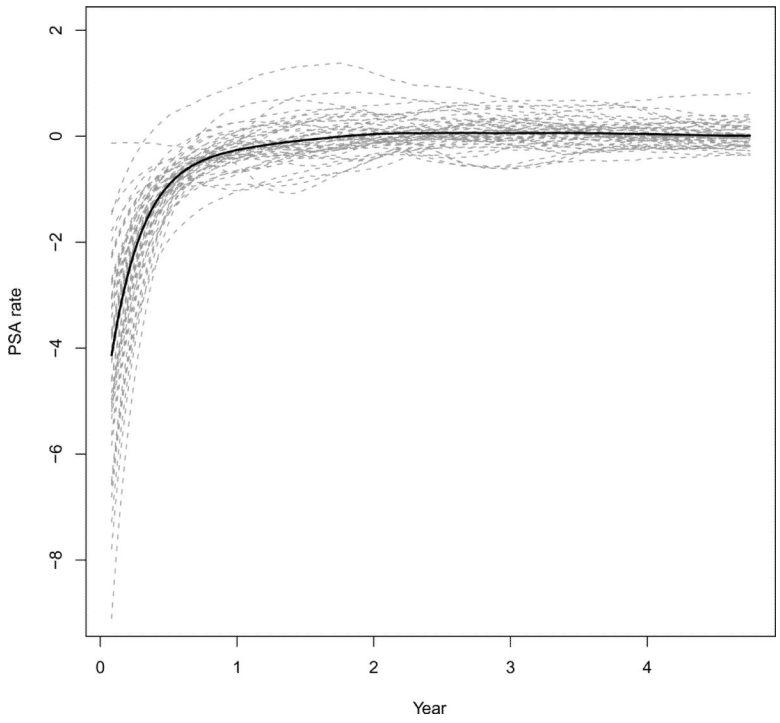

Figure 2.
Posterior means of $V_{x_i}(t)$ for each subject as gray dashed lines and the population-level rate function $V(t)$ as black solid line

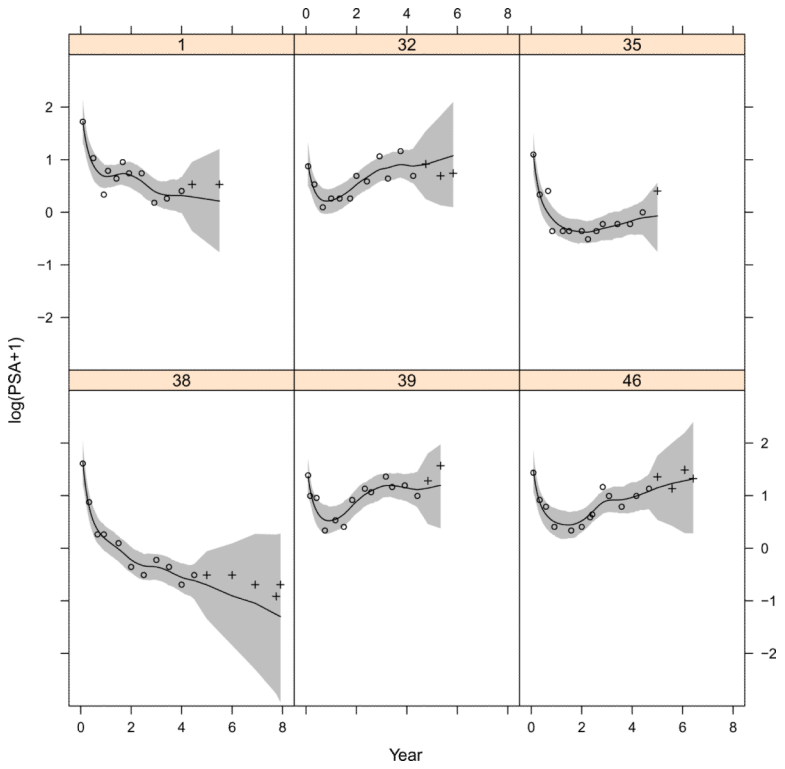

Figure 3.
Plots of training data points (o), validation data points (+), posterior means (−) and 95% credible intervals (gray shades) of $U_{x_i}(t)$ for six randomly selected subjects.

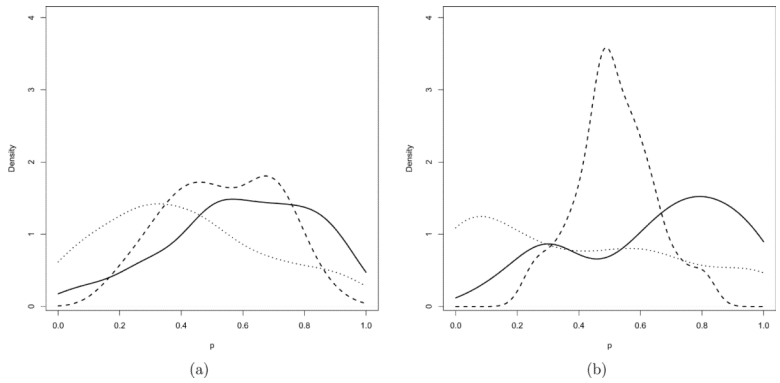

Figure 4.
PIT density plots for (a) $t_{ij} \leq 1$ year, (b) $t_{ij} > 1$ year of SSVM-OU (−), SSVM-W (- - -), LMM (⋯)

Table 1

PSA data: Posterior means and quantiles of parameters for the SSVM-OU and LMM.

Model	Parameter	Mean	SD	2.5%	50%	97.5%
SSVM-OU	σ_{ε}^2	0.044	0.004	0.037	0.044	0.053
	σ_{ξ}^2	1.365	0.297	0.921	1.320	2.108
	ρ	3.721	0.360	3.101	3.690	4.464
	σ_{ν}^2	0.054	0.015	0.031	0.051	0.089
	β_0	−0.171	0.085	−0.335	−0.169	−0.004
	β_1	0.139	0.072	0.001	0.139	0.277
	β_2	0.242	0.095	0.060	0.237	0.438
	β_3	0.061	0.103	−0.157	0.064	0.269
LMM	β_{20}	0.061	0.066	−0.072	0.058	0.200
	β_{21}	0.116	0.056	0.008	0.117	0.225
	β_{22}	0.260	0.076	0.116	0.260	0.411
	β_{23}	0.046	0.078	−0.105	0.046	0.193

Table 2

PSA data: Posterior forecasting of the validation data points. The relative bias is defined as $E\left(\tilde{Y}/Y-1\right)$ for $\tilde{Y}$ the posterior mean of validation data point Y.

Method	Type	Relative Bias	MSE	Coverage Rate	Interval Length
SSVM-OU	≤ 1 year	−0.143	0.076	1	1.403
	> 1 year	−0.913	0.581	0.966	2.356
	All	−0.644	0.404	0.912	2.023
SSVM-W	≤ 1 year	−0.047	0.098	1	2.379
	> 1 year	0.031	0.403	1	8.329
	All	0.040	0.296	1	6.250
LMM	≤ 1 year	0.205	0.108	0.899	1.226
	> 1 year	0.387	0.568	0.672	1.610
	All	0.323	0.407	0.748	1.476

Table 3

Simulation results on the estimation of SSVM-OU parameters. The relative bias is defined as $E\left(\tilde{\phi}/\phi - 1\right)$ for $\tilde{\phi}$ the posterior mean of the parameter ϕ.

Data Augmented	Parameter	Truth	Relative Bias	MSE	Coverage Rate	Interval Length
0	σ_ε^2	0.05	−0.202	1.294e−04	0.678	0.026
	σ_ξ^2	1.00	−0.408	1.689e−01	0.044	0.461
	ρ	3.50	−0.470	2.704e+00	0.000	0.101
	σ_ν^2	0.05	2.101	1.123e−02	0.211	0.299
	β_0	−0.15	0.03	1.992e−02	1.000	1.150
	β_1	0.15	0.139	3.140e−02	1.000	1.307
	β_2	0.25	−0.006	1.693e−02	1.000	1.120
	β_3	0.10	−0.105	1.995e−02	1.000	1.166
9	σ_ε^2	0.05	0.012	2.952e−05	0.967	0.024
	σ_ξ^2	1.00	−0.041	2.253e−02	0.967	0.800
	ρ	3.50	−0.052	3.596e−02	0.256	0.273
	σ_ν^2	0.05	0.548	9.410e−04	0.989	0.133
	β_0	−0.15	0.035	2.033e−02	0.978	0.647
	β_1	0.15	0.135	3.133e−02	0.956	0.739
	β_2	0.25	0.005	1.673e−02	0.989	0.634
	β_3	0.10	−0.099	2.006e−02	0.956	0.660
19	σ_ε^2	0.05	0.013	3.921e−05	0.956	0.024
	σ_ξ^2	1.00	0.007	2.497e−02	0.967	0.802
	ρ	3.50	−0.018	8.104e−03	0.911	0.288
	σ_ν^2	0.05	0.494	7.960e−04	0.989	0.126
	β_0	−0.15	0.022	2.013e−02	0.967	0.625

Data Augmented	Parameter	Truth	Relative Bias	MSE	Coverage Rate	Interval Length
	β_1	0.15	0.138	3.135e−02	0.944	0.713
	β_2	0.25	0.003	1.686e−02	0.978	0.612
	β_3	0.10	−0.113	1.980e−02	0.978	0.635

Table 4

Simulation results on forecasting by three models

Simulation Model	Fitted Model	Year Distance	Relative Bias	MSE	Coverage Rate	Interval Length
SSVM-OU	SSVM-OU	0.08	0.010	0.162	0.949	1.116
		0.5	0.019	0.232	0.951	1.336
		1	0.011	0.343	0.951	1.639
	SSVM-W	0.08	0.008	0.270	0.957	1.455
		0.5	−0.144	6.700	1	9.640
		1	−0.024	39.430	1	23.974
	LMM	0.08	−0.165	1.508	0.994	3.465
		0.5	−0.061	2.206	0.936	3.547
		1	−0.303	3.442	0.742	3.661
LMM	SSVM-OU	0.08	0.001	0.034	0.915	0.483
		0.5	0.077	0.069	0.974	0.781
		1	0.092	0.155	0.996	1.249
	SSVM-W	0.08	−0.021	0.046	0.920	0.565
		0.5	0.075	0.384	1.000	2.152
		1	−0.040	1.905	1.000	5.085
	LMM	0.08	−0.007	0.028	0.943	0.456
		0.5	0.026	0.030	0.943	0.478
		1	−0.006	0.034	0.952	0.510

J Am Stat Assoc. Author manuscript; available in PMC 2012 December 1.