Stochastic Dynamic Models for Functional Data

by

Bin Zhu

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Biostatistics)
in The University of Michigan
2010

Doctoral Committee:

Professor Peter X. Song, Co-Chair
Professor Jeremy M.G. Taylor, Co-Chair
Professor Naisyin Wang
Assistant Professor Brisa N. Sanchez

To my parents and zhenzhen

ACKNOWLEDGEMENTS

I would like to express my deepest gratitude to my advisors Jeremy Taylor and Peter Song for supporting me to explore this research topic and for advising my dissertation work along the way. Their broad knowledge and attitude towards research will serve as academic role models for me. I also appreciate Morton Brown and Jeremy Taylor for providing my financial support and letting me be involved several interesting collaborative research projects. I am grateful to Naisyin Wang and Brisa Sanchez for serving as my dissertation committee. Their advice and comments are so valuable.

I would like to thank all faculty members of the Department of Biostatistics for teaching me biostatistics. I have taken many great classes, especially 699, and have had a lot of valuable discussions with my teachers and my classmates.

I have spent my most wonderful time at the Department of Biostatistics at the University of Michigan.

TABLE OF CONTENTS

LIST OF FIGURES

Figure

LIST OF TABLES

Table

LIST OF APPENDICES

Appendix

CHAPTER I

Introduction

1.1 Scope and Aims

Functional data arise frequently in many fields of biomedical research as sequential observations over time. The observations are generated by an unknown dynamic mechanism. This dynamic process has unspecified mean function, and the observations can be considered as arising from this mean function plus noise. The goal is typically estimation and inference for the unknown mean function and its derivatives as well as other parameters involved in the underlying dynamic system. Methods from nonparametric and semiparametric regression are typically used to achieve this goal.

1.1.1 Smoothing methods

Two widely used methods to estimate the unknown mean function are kernel smoothing (Wand and Jones, 1995; Fan and Gijbels, 1996) and spline smoothing (Wahba, 1990; Green and Silverman, 1994; Ramsay and Silverman, 2005). Suppose we have observations $Y(t_j)$ at design points t_j, $j = 1, 2, \ldots, J$. The observations are assumed to follow the model

$$Y(t_j) = U(t_j) + \varepsilon(t_j), \qquad \varepsilon(t_j) \overset{i.i.d}{\sim} \mathcal{N}(0, \sigma_\varepsilon^2). \tag{1.1}$$

The kernel smoothing estimator at a target point t is defined as a weighted average:

$$\hat{U}(t) = \sum_{j=1}^{J} W_{j,h}(t) Y(t_j),$$

where $W_{j,h}(t)$ is the weight function centered at t with bandwidth h. A popular choice of weight function is to use a kernel $K_h(\cdot)$, and the $W_{j,h}(t)$ is given by

$$W_{j,h}(t) = \frac{K_h(t_j - t)}{\sum_{j=1}^{J} K_h(t_j - t)}.$$

This is called the Nadaraya−Watson local average kernel smoothing. Very often the kernel function $K_h(\cdot)$ is chosen to be a density function with mean zero.

Spline smoothing is achieved by minimizing an objective function, namely the penalized sum-of-squares

$$S_\lambda(U) = \sum_{j=1}^{J} [Y(t_j) - U(t_j)]^2 + \lambda \int L(U)^2(t) dt,$$

where L is linear differential operator and λ is the smoothing parameter. The first term of $S_\lambda(U)$ measures the goodness-of-fit of $U(\cdot)$ to the data, while the second term penalizes the roughness of $U(\cdot)$. The smoothing spline estimator of $U(\cdot)$ is obtained by minimizing $S_\lambda(U)$, where λ is determined typically through cross-validation. When $L(U) = D^2(U) = U''$, the minimizer corresponds to a natural cubic smoothing spline, which is in fact a piecewise cubic polynomial.

The above two smoothing methods essentially regard the mean function $U(t)$ as a deterministic function and mainly focus on estimation of $U(t)$. In this dissertation, we take a different perspective , and treat $U(t)$ as the realization or sample path of a stochastic process. This will enable us to study dynamics of the underlying process, including how the stochastic process and its derivatives evolve over time, both within the observation time(through estimation and inference) and also afterwards(through

forecasting).

1.1.2 Stochastic dynamic model

A stochastic dynamic model(SDM) refers to a class of hierarchical models, where observations are sampled from distributions, which depend on an function determined by some diffusion processes that evolve continuously and stochastically over time. The SDM consists of two hierarchies: A d-dimensional discrete-time observation process $\{\boldsymbol{Y}_t\}_{t\in\mathcal{T}_o}$ with $\mathcal{T}_o := \{t_j : t_1 < t_2 < \cdots < t_J\}$ and m-dimensional continuous-time latent state process $\{\boldsymbol{\theta}_t\}_{t\in\mathcal{T}_s}$ with $\mathcal{T}_s := \{t : t_0 \leq t \leq t_J\}$, detailed as follows:

$\boldsymbol{P_o}$: The observation process $\{\boldsymbol{Y}_t\}_{t\in\mathcal{T}_o}$ defined on state space $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ gives a collection of observations; given latent process $\boldsymbol{\theta}_t$, $\boldsymbol{Y}_t$ is conditionally independent of $\{\boldsymbol{Y}_s\}_{s\neq t}$ and $\{\boldsymbol{\theta}_s\}_{s\neq t}$ with the conditional distribution

$$\boldsymbol{Y}_t \mid \boldsymbol{\theta}_t \sim f_t(\boldsymbol{y} \mid \boldsymbol{\theta}_t, \boldsymbol{X}_o, \boldsymbol{\phi}_o), \boldsymbol{Y}_t \in \mathbb{R}^d \tag{1.2}$$

where $\boldsymbol{X}_o$ and $\boldsymbol{\phi}_o$ are respectively the covariates and the parameter vector involved.

$\boldsymbol{P_s}$: The state process $\{\boldsymbol{\theta}_t\}_{t\in\mathcal{T}_s}$ defined on state space $(\mathbb{R}^m, \mathcal{B}(\mathbb{R}^m))$ is a latent diffusion process, which is a Markov process with almost surely continuous sample paths, governed by a stochastic differential equation(SDE),

$$d\boldsymbol{\theta}(t) = \boldsymbol{a}\{\boldsymbol{\theta}(t); \boldsymbol{X}_s, \boldsymbol{\phi}_s, t\}dt + \boldsymbol{b}\{\boldsymbol{\theta}(t); \boldsymbol{X}_s, \boldsymbol{\phi}_s, t\}d\boldsymbol{W}(t), \tag{1.3}$$

where

$$\boldsymbol{\theta}(t) = \left[\begin{array}{c} \theta_1(t), \cdots, \theta_m(t) \end{array} \right]^{\mathrm{T}},$$

$$\boldsymbol{a}\{\boldsymbol{\theta}(t); \boldsymbol{X}_s, \boldsymbol{\phi}_s, t\} = \left[\begin{array}{c} a_1\{\boldsymbol{\theta}(t); \boldsymbol{X}_s, \boldsymbol{\phi}_s, t\}, \cdots, a_m\{\boldsymbol{\theta}(t); \boldsymbol{X}_s, \boldsymbol{\phi}_s, t\} \end{array} \right]^{\mathrm{T}},$$

$$\boldsymbol{b}\{\boldsymbol{\theta}(t); \boldsymbol{X}_s, \boldsymbol{\phi}_s, t\} = \left[\begin{array}{ccc} b_{11}\{\boldsymbol{\theta}(t); \boldsymbol{X}_s, \boldsymbol{\phi}_s, t\} & \ldots & b_{1n}\{\boldsymbol{\theta}(t); \boldsymbol{X}_s, \boldsymbol{\phi}_s, t\} \\ \vdots & & \vdots \\ b_{m1}\{\boldsymbol{\theta}(t); \boldsymbol{X}_s, \boldsymbol{\phi}_s, t\} & \ldots & b_{mn}\{\boldsymbol{\theta}(t); \boldsymbol{X}_s, \boldsymbol{\phi}_s, t\} \end{array} \right],$$

is a n-dimensional independent Wiener process $\boldsymbol{W}(t) = [W_1(t), \cdots, W_n(t)]^{\mathrm{T}}$, $\boldsymbol{X}_s$ and $\boldsymbol{\phi}_s$ are the covariates and parameter vector involved. The elements of $\boldsymbol{\theta}(t)$ usually include a process $U(t)$ and its first to qth order derivatives besides other covariate processes. The vector of drift terms $\boldsymbol{a}\{\boldsymbol{\theta}(t); \boldsymbol{X}_s, \boldsymbol{\phi}_s, t\}$ and the matrix of diffusion terms $\boldsymbol{b}\{\boldsymbol{\theta}(t); \boldsymbol{X}_s, \boldsymbol{\phi}_s, t\}$ are assumed to satisfy regularity conditions; see Grimmett and Stirzaker (2001, Chap. 13) and Feller (1970, Chap. 10).

The SDM can be regarded as the generalization of the continuous-discrete state space model (Jazwinski, 1970, Chap. 6) by allowing continuous and discrete observations as well as by including possible covariates. SDM has some similarity to other modeling approaches, such as state space models (Jones, 1993; Durbin and Koopman, 2001), linear dynamic models (West and Harrison, 1997) and many other diffusion type models in finance, e.g. the stochastic volatility model (Hull and White, 1987), but are not identical to them. In the linear dynamic models, both the observations and states exist at discrete time and the states are specified by conditional densities. The diffusion type models in finance have both continuous time observation and states, and the states are usually directly or partially observed.

Estimation and inference for $\{\boldsymbol{\theta}_t\}_{t \in \mathcal{T}_s}$, $\boldsymbol{\phi}_s$ and $\boldsymbol{\phi}_o$ in SDM is very challenging, due to the lack of closed form expression for the likelihood function, which involves high-dimensional integrals, and to the requirement of inferring continuous unknown func-

tions. To overcome those challenges, we develop Markov Chain Monte Carlo(MCMC) based Bayesian inference utilizing the Euler approximation and data augmentation. The estimation methods are described in chapter II.

The advantages of SDM as a modeling approach for functional data analysis include: i) SDM treats the unknown function as a realization of a stochastic diffusion processes, governed by stochastic differential equations. In this way, we can investigate not only the mean function nonparametrically but also its dynamics(or derivatives), that may be of great interest in many applications; ii) Because it is a fully specified probability model, inference and forecasting can be achieved using an MCMC approach. In contrast, smoothing techniques are essentially point estimation methods; iii) SDM is able to model not only smooth functions, but also non-smooth functions, such as those with structural changes or breakpoints; iv), Covariates can be incorporated into both the distribution of the observations, and the drift and diffusion terms of the SDE in equation (1.3). Thus, covariate effects can be assessed in the observation and/or state processes simultaneously.

In this dissertation, we mainly focus on the situation when the observations $\{\boldsymbol{Y}_t\}_{t\in\mathcal{T}_o}$ are continuous and explore various specific forms of SDEs in (1.3) for different purposes.

1.2 Organization of Dissertation

1.2.1 Organization

Chapter II introduces a new modeling strategy for time series functional data. We consider the problem of estimating an unknown smooth function. The model consists of an observation equation for the observation process, an ordinary differential equation(ODE) and a stochastic differential equation(SDE) for the state process. The method of smoothing spline, introduced in section 1.1.1, is connected to a spe-

cial case of this approach. The resulting models offer great flexibility to capture the dynamic features of functional data, and allows straightforward and meaningful interpretation. The likelihood of the models is derived with Euler approximation and data augmentation. A unified Bayesian inference method for those models is carried out via a MCMC algorithm with simulation smoother. The proposed models and methods are illustrated on some prostate specific antigen(PSA) data, where we use an Ornstein-Uhlenbeck process to model the rate of changes of PSA and achieve more precise forecasting.

Chapter III presents a novel semiparametric stochastic approach to model the rate functions for functional data in a multi-subject setting. The rate functions, one for each subject, characterize the rate of change of individual profiles, and are assumed to follow covariate-dependent Gaussian processes, which obey stochastic differential equations. Consequently, each rate function is expected to be centered on a parametric function, which may be dependent on the covariates, and nonparametric departure is allowed from the parametric functional expectation. The profiles for different subjects are linked to each other, because some parameters in the SDE are common whereas others are unique and assumed to arise from a mixing distribution. The proposed approach is flexible to capture the patterns of subject-specific rate of change, convenient to study the covariate effects on the rate function, and easy for interpretation of the results. The proposed methods are illustrated by analyzing a prostate-specific antigen dataset with many subjects and through simulation studies.

Chapter IV proposes a time-varying stochastic position model with an observation equation and an SDE, where the diffusion term is time-varying. In this way, the model can approximate the breakpoints in a function. The discretized version of this model assumes a t-distribution for both the measurement errors and the signal differences. The discretized model is applied to array comparative genomic hybridization(CGH) data analysis. The CGH profiles are regarded as functional unknown signals estimated

by a MCMC algorithm. Both breakpoints and outliers are identified by a backward selection procedure. Glioblastoma Multiforme(GBM) data are used to demonstrate the characteristics of the proposed method. Compared to three other popular methods, our approach has superior detection ability, which is illustrated on a simulated dataset and a breast tumor dataset.

Chapter V summarize the results in this dissertation and discusses some future work.

CHAPTER II

Stochastic Functional Data Analysis: A Diffusion Model-based Approach

2.1 Introduction

Conventional nonparametric regression theory concerns primarily the estimation of a population mean function through kernel smoothing (Fan and Gijbels, 1996) and spline smoothing techniques (Green and Silverman, 1994). 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. With the advent of many high-throughput technologies, functional data analysis has received much attention in the recent statistical literature(Ramsay and Silverman, 2005), where tremendous challenges have emerged in statistical methodology development.

This paper presents a new modeling strategy in functional data analysis. Our primary focus in modeling lies on a system of stochastic differential equations connected in a hierarchical fashion, in the hope that it not only models the mechanism of the population mean function but also captures various features of the dynamics of the mean function. Consequently, we can make statistical inference on both means and dynamics of the underlying process.

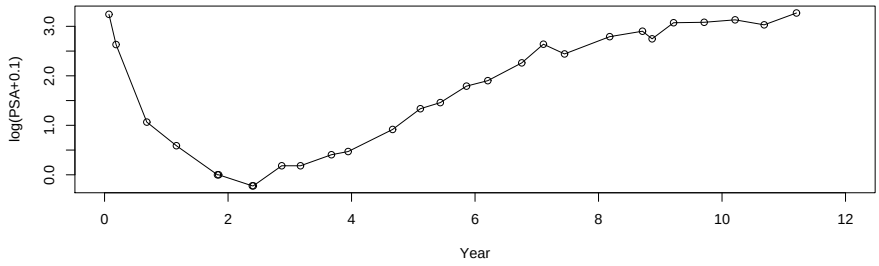
For example, in the study of prostate specific antigen (PSA), an important biomarker of prostate cancer, we are not only interested in the PSA level but also the dynamics of PSA. Figure 2.1 displays raw data of one patient's PSA level (panel (a)) and the scaled difference (panel (b)) over time (Proust-Lima et al., 2008b), 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. Hence, modeling the process of the scaled difference properly 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 first order derivative of the PSA level. While panel (a) and panel (b) show the data, by using the model(to be described in section 4.1), we can make probability statements about features of the curves. For example, in Figure 2.1 panel (c) we show the posterior probability of the time when PSA level reaches its minimum.

Consider a regression model for functional data of the form:

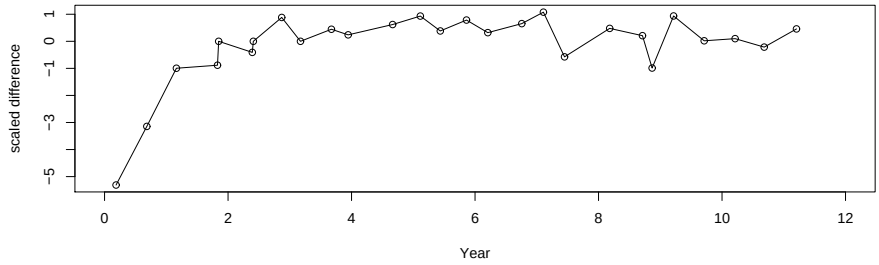
$$Y(t) = U(t, \omega) + \varepsilon(t),\ \omega \in \Omega,\ t \in \mathcal{T}_o \tag{2.1}$$

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 $U(\cdot\,, \omega)$ given time series observations, $\boldsymbol{Y}_o = [Y(t_1), Y(t_2), \ldots, Y(t_J)]^{\mathrm{T}}$. In this paper, 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 sample path realization of an underlying stochastic process $U := U(\cdot\,, \cdot\,)$ and the observed data is thus process plus measurement error.

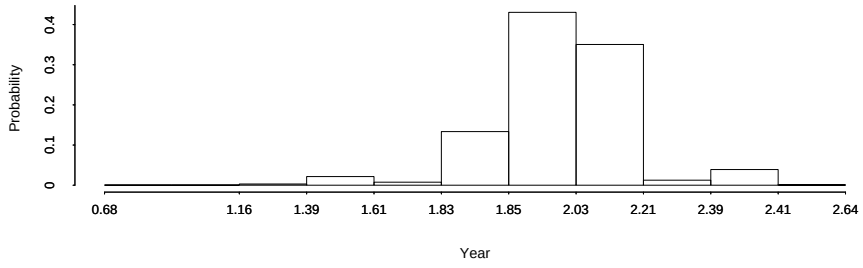
Model (2.1) is useful to model the PSA level of prostate cancer nonparametrically,



(a)



(b)



(c)

Figure 2.1: PSA plots: (a) the raw data; (b) the scaled difference; (c) posterior probability of minimum of PSA level at interval $[t_j, t_{j+1}]$ in SVM-OU

where $U(t,\omega)$ describes the population mean PSA process. In addition, to understand the dynamics profiles concerning the evolution of the biomarker process, we incorporate models of rate(or velocity) and/or higher order derivatives into model (2.1). To proceed, we begin by treating $U(t,\omega)$ in model (2.1) 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),

$$\frac{d^{m-1}U(t)}{dt^{m-1}} = V(t), \tag{2.2}$$

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

$$dV(t) = a\{V(t), \boldsymbol{\phi}_s\}dt + b\{V(t), \boldsymbol{\phi}_s\}dW(t), \quad t \in \mathcal{T}_s, \tag{2.3}$$

where $W(t)$ is the standard Wiener process, $\boldsymbol{\phi}_s$ is the parameter vector and $\mathcal{T}_s := \{t : t_0 \leq t \leq t_J\}$ 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 paper, we use continuous time stochastic processes $\boldsymbol{U}$ and $\boldsymbol{V}$ to model the underlying dynamics of interest. 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 (2.3) 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 (2.2) and (2.3), along with the observation equation (2.1), 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 (2.3) 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. Two special cases are considered in this paper. They are, (i) SVM and SAM with Wiener process $V(t)$, denoted SVM-W and SAM-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 and SAM with Ornstein-Uhlenbeck(OU) process $V(t)$, denoted SVM-OU and SAM-OU, where $a\{V(t), \boldsymbol{\phi}_s\} = -\rho\{V(t) - \bar{\nu}\}$, $b\{V(t), \boldsymbol{\phi}_s\} = \sigma_\xi$ and $\boldsymbol{\phi}_s = [\rho, \bar{\nu}, \sigma_\xi^2]^{\mathrm{T}}$. These two processes will be discussed in detail in Section 2.

When $V(t)$ follows a Wiener process, 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 and Sidhu, 1980). By equivalence, we mean that the two methods give the same estimate of $U(t)$. To elaborate, let $\hat{U}(t; \sigma_0^2) := \mathbb{E}\{U(t) \mid \boldsymbol{Y}_o; \sigma_\varepsilon, \sigma_\xi, \sigma_0^2\}$ be the posterior mean of $U(t)$ in SVM-W . Wahba (1978) showed that $\hat{U}(t) := \lim_{\sigma_0^2 \to \infty} \hat{U}(t; \sigma_0^2)$ exists and is the same as the estimate obtained by the smoothing spline with degree $2m - 1$ and $2m - 2$ continuous derivatives. Wahba's estimation method minimizes the penalized sum-of-squares,

$$\sum_{j=1}^{J} [y(t_j) - U(t_j)]^2 + \lambda P_m(U), \tag{2.4}$$

where $\lambda = \frac{\sigma_\varepsilon^2}{\sigma_\xi^2}$ and the roughness penalty $P_m(U)$ is given by

$$P_m(U) = \int_{\mathcal{T}_s} [U^{(m)}(t,\omega)]^2 dt, \quad m = 2, 3, \ldots, \tag{2.5}$$

Kimeldorf and Wahba (1970) and Wahba (1978) have shown the "equivalence" by treating penalized sum-of-squares (2.4) as the minimal norm optimization problem in a Reproducing Kernel Hilbert Space, where the kernel is regarded as the variance covariance function of the stochastic process U in SVM-W ; see also Ansley and Kohn (1986) for a detailed discussion. In short, the posterior mean of $U(t)$ in SVM-W can be equivalently obtained by the smoothing spline. Diggle and Hutchinson (1989) and Kohn and Ansley (1988) found that the equivalence results can hold for more general covariance matrices than the diagonal matrix of independent measurement errors $\varepsilon(t)$.

For the PSA data example in figure 2.1 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. In other applications, the data may demonstrate different patterns, for example, periodic patterns, which require more flexible models to address those complicated dynamic features. The diffusion model and the consideration of higher derivative in (2.2) 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, which is very useful in biomedical and other research.

Note that in this paper we treat the function $U(t,\omega)$ in model (2.1) as a sample path of the stochastic process U. This treatment is different from kernel smoothing and spline smoothing, where $U(t,\omega)$ is regarded as a deterministic unknown function. We also note that our treatment of $U(t,\omega)$ is similar to that considered in

the Gaussian process models for nonparametric Bayesian data analysis, where the nonparametric function $U(t, \omega)$ 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). In term of statistical estimation and inference, our method offers the estimation and inference for parameters of the stochastic differential equation from noisy data. We note that this differs from the approaches to parameters estimation for models based on ordinary differential equations as recently developed by, for example, Ramsay et al. (2007) and Liang and Wu (2008).

The rest of the paper is organized as follows. Section 2.2 concerns two special cases of the proposed stochastic velocity model, one with the Wiener process $V(t)$ and the other with the OU process $V(t)$. For each cases, we give model interpretations and discuss several interesting relationships. Section 2.3 develops Bayesian inference for stochastic functional data analysis models, illustrated by the SVM, where the likelihood is derived using Euler approximation and data augmentation. In Section 2.4, the proposed models and methods are applied to estimate the PSA profile from prostate cancer data. Concluding remarks are given in Section 2.5. Technical details are included in the Appendix.

2.2 Examples of stochastic velocity model

We focus on two SVMs, where velocities are modelled by specific forms for the SDE. The first is the SVM-W, for which we will point out its connections to the problem of model adequacy. The second is the SVM-OU, in which we examine its relationship to the Wiener process and AR(1) model. For both special cases, we also give the interpretation of the parameters $\boldsymbol{\phi}_s$. Similar interpretations can be applied to the SAM, substituting the acceleration for the velocity.

2.2.1 Wiener process for velocity

In SVM-W, $V(t)$ follows the 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 (2.2) and (2.3) for $m = 2$, $a\{V(t), \boldsymbol{\phi}_s\} = 0$ and $b\{V(t), \boldsymbol{\phi}_s\} = \sigma_\xi$, we have

$$U(t) = U(t_0) + \int_{t_0}^{t} V(s)\, ds = U(t_0) + V(t_0)(t - t_0) + \sigma_\xi \int_{t_0}^{t} W(s)\, ds, \tag{2.6}$$

$$V(t) = V(t_0) + \sigma_\xi W(t). \tag{2.7}$$

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, $\sigma_\xi \int_0^t W(s)\, ds$.

2.2.2 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:

$$dV(t) = -\rho\{V(t) - \bar{\nu}\}dt + \sigma_\xi dW(t), \quad t \in \mathcal{T}_s, \tag{2.8}$$

where $\rho \in \mathbb{R}^+, \bar{\nu} \in \mathbb{R}$, and $\sigma_\xi \in \mathbb{R}^+$. In contrast to the Wiener process, 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 towards $\bar{\nu}$.

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

2.3 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 closed form. Since an analytical solution of the SDE is rarely available, the resulting 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 preciser 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 the point is the SVM-OU.

The resulting likelihood with this approximate method involves high-dimensional integrals, and we adopt a Bayesian approach in which we use Markov Chain Monte Carlo (MCMC, Geman and Geman 1987, Gelfand and Smith 1990, Gilks et al. 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, 2002b).

Different approaches for inference for discretely observed diffusions are reviewed

by Beskos et al. (2006). These includes numerical approximations to obtain likelihood functions (Aït-Sahalia, 2002) and methods based on iterated filtering (Ionides et al., 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 et al. (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 paper are that we specify an observation equation to address the measurement errors. Most methods of inference for diffusion process do not extend easily when there is measurement errors (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 volatility or variance of the process of interest, for example, derivative securities.

2.3.1 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 $[\cdot \mid \cdot]$ denote conditional density. The conditional density of the observations is given by

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

since the observations are mutually independent given the latent states and follow the normal distribution according to model (2.1), where $\phi(\cdot \mid U_G, \sigma_G^2)$ is the normal density function 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 complex form. For SVM-OU,

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

with

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

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

the proof of which is in Appendix C. When using data augmentation, we may take the component wise first-order Taylor approximation of $\boldsymbol{m}_{OU}, \boldsymbol{V}_{OU}$ with respected to

δ_j and get

$$\tilde{\boldsymbol{m}}_{OU} = [U(t_j) + V(t_j)\delta_j, V(t_j) - \rho\{V(t_j) - \bar{\nu}\}\delta_j]^{\mathrm{T}}, \tag{2.9}$$

$$\tilde{\boldsymbol{V}}_{OU} = \sigma_\xi^2 \begin{bmatrix} 0 & 0 \\ 0 & \delta_j \end{bmatrix}. \tag{2.10}$$

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 (2.9) and (2.10) 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 be updated and converges much faster in the following MCMC algorithm.

For a general SDE, e.g. (2.3), 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 (2.2) and (2.3) are given by, respectively,

$$U^{(J)}(t_j) = U^{(J)}(t_{j-1}) + V^{(J)}(t_{j-1})\delta_j, \tag{2.11}$$

$$V^{(J)}(t_j) = V^{(J)}(t_{j-1}) + a\{V^{(J)}(t_{j-1}), \boldsymbol{\phi}_s\}\delta_j + b\{V^{(J)}(t_{j-1}), \boldsymbol{\phi}_s\}\eta_j, \quad t_j \in \mathcal{T}_o \tag{2.12}$$

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

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

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:

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

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 paper, 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 those 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 (2.11) and (2.12) 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 (2.11), 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

$$[\boldsymbol{y}_o \mid \boldsymbol{V}_o, \boldsymbol{\phi}_o] = \prod_{j=1}^{J} \phi(y(t_j) \mid U(t_j), \sigma_\varepsilon^2),$$
$$\prec \boldsymbol{V}_o \mid V(t_0), \boldsymbol{\phi}_s \succ = \prod_{j=1}^{J} \prec V(t_j) \mid V(t_{j-1}), \boldsymbol{\phi}_s \succ,$$

and

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

In a real data analysis, the Euler approximation may not always achieve the desired precision, simply because $\max(\delta_j)$ for the observations is not sufficiently small. A solution to this is data augmentation as described in the next subsection.

2.3.2 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

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

where

$$\begin{aligned}
[\boldsymbol{y}_o, \boldsymbol{y}_a \mid \boldsymbol{U}, \boldsymbol{V}, \boldsymbol{\phi}_o] &= \prod_{j=0}^{J} \prod_{m=0}^{M_j} \phi(y_{j,m} \mid U_{j,m}, \sigma_\varepsilon^2), \\
[\boldsymbol{U}, \boldsymbol{V} \mid \boldsymbol{\theta}_0, \boldsymbol{\phi}_s] &= \prod_{j=1}^{J} \prod_{m=1}^{M_j+1} [\, U_{j-1,m}, V_{j-1,m} \mid U_{j-1,m-1}, V_{j-1,m-1}, \boldsymbol{\phi}_s].
\end{aligned}$$

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 $t \in \mathcal{T}_{ao}$ given as follows:

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

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 2.3.3.

2.3.3 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 in the case of data augmentation, $[\boldsymbol{\theta}_0, \boldsymbol{V}, \boldsymbol{\phi}_o, \boldsymbol{\phi}_s, \boldsymbol{y}_a \mid \boldsymbol{y}_o]$. We use the approximate transition density to obtain the likelihood and apply the data augmentation technique whenever necessary.

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:

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

where

$$\begin{aligned}[\boldsymbol{y}_o \mid \boldsymbol{\theta}_0, \boldsymbol{V}, \boldsymbol{\phi}_o, \boldsymbol{\phi}_s] &= \prod_{j=1}^{J} \phi(y_{j,0} \mid U_{j,0}(\boldsymbol{\theta}_0, \boldsymbol{V}), \boldsymbol{\phi}_o), \\ [\boldsymbol{y}_a \mid \boldsymbol{\theta}_0, \boldsymbol{V}, \boldsymbol{\phi}_o, \boldsymbol{\phi}_s] &= \prod_{j=0}^{J-1} \prod_{m=1}^{M_j} \phi(y_{j,m} \mid U_{j,m}(\boldsymbol{\theta}_0, \boldsymbol{V}), \boldsymbol{\phi}_o), \\ \prec \boldsymbol{V} \mid \boldsymbol{\theta}_0, \boldsymbol{\phi}_s \succ &= \prod_{j=1}^{J} \prod_{m=1}^{M_j+1} \prec V_{j-1,m} \mid V_{j-1,m-1}, \boldsymbol{\phi}_s \succ,\end{aligned}$$

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,

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

and $[\boldsymbol{\theta}_0]$, $[\boldsymbol{\phi}_s]$, $[\boldsymbol{\phi}_o]$ are non-informative prior densities. See the Appendix A for specification of the prior distributions and details of MCMC algorithm. Note that we use the simulation smoother(Durbin and Koopman, 2002b) 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.

A desirable property of this approach is the ease of deriving forecasts of states at

future times, which is of great interest in some studies.

2.3.4 Posterior forecasting with SVM

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,

$$[\boldsymbol{\theta}^f_{J+k} \mid \boldsymbol{y}_o] = \int\int\int [\boldsymbol{\theta}^f_{J+k} \mid \boldsymbol{y}_a, \boldsymbol{y}_o, \boldsymbol{\phi}_s, \boldsymbol{\phi}_o][\boldsymbol{y}_a, \boldsymbol{\phi}_s, \boldsymbol{\phi}_o \mid \boldsymbol{y}_o]\, d\boldsymbol{y}_a\, d\boldsymbol{\phi}_s\, d\boldsymbol{\phi}_o,$$

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 (A.1) and (A.2) in Appendix A. 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 (A.2) 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 Appendix A for the SVM-OU and SVM-Ws, 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 of states at future times take the variation of parameter draws into consideration.

2.4 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 Akaikes information criterion. 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.

2.4.1 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 varying 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 re-growing and that the treatment did not cure the patient. Therefore, PSA has strong prognostic significant 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. Yu et al. (2008) applied a joint longitudinal survival-cure model to make individual prediction of cancer recurrence, where the "true" PSA trajectory is specified by a nonlinear exponential decay and growth model(Zagars and Pollack, 1993),

$$PSA(t) = r_1 \exp(-r_2 t) + r_3 \exp(r_4 t),$$

where r_1,r_2,r_3 and r_4 are parameters. 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 2.1 and 2.2 shows the means and quantiles of the SVM and SAM parameters from the Wiener and the OU process $V(t)$, respectively. Figure 2.2 and 2.3 shows 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 fluctuates with higher volatility, compare to the SAMs. In addition, there are the non-zero 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 2.4 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 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, Wiener process corresponds to a nonstationary process for the rate 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.

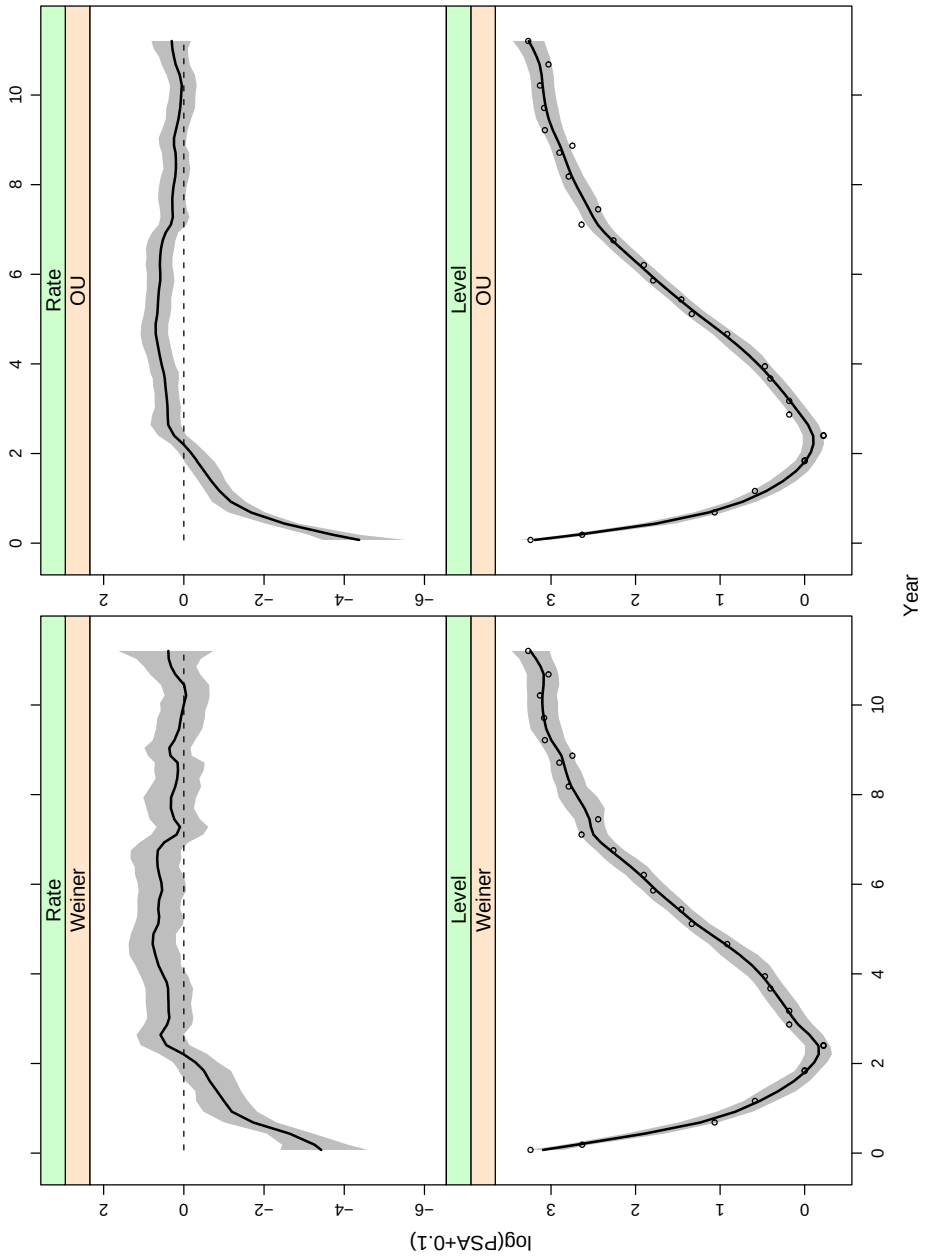


Figure 2.2: PSA: Plots of data points(∘), posterior means(—) and 95% credible intervals(gray shades) for the SVM with the Wiener process and OU process, respectively. In the graph the upper panels show the rate, $V(t)$, and the lower panels show the level, $U(t)$.

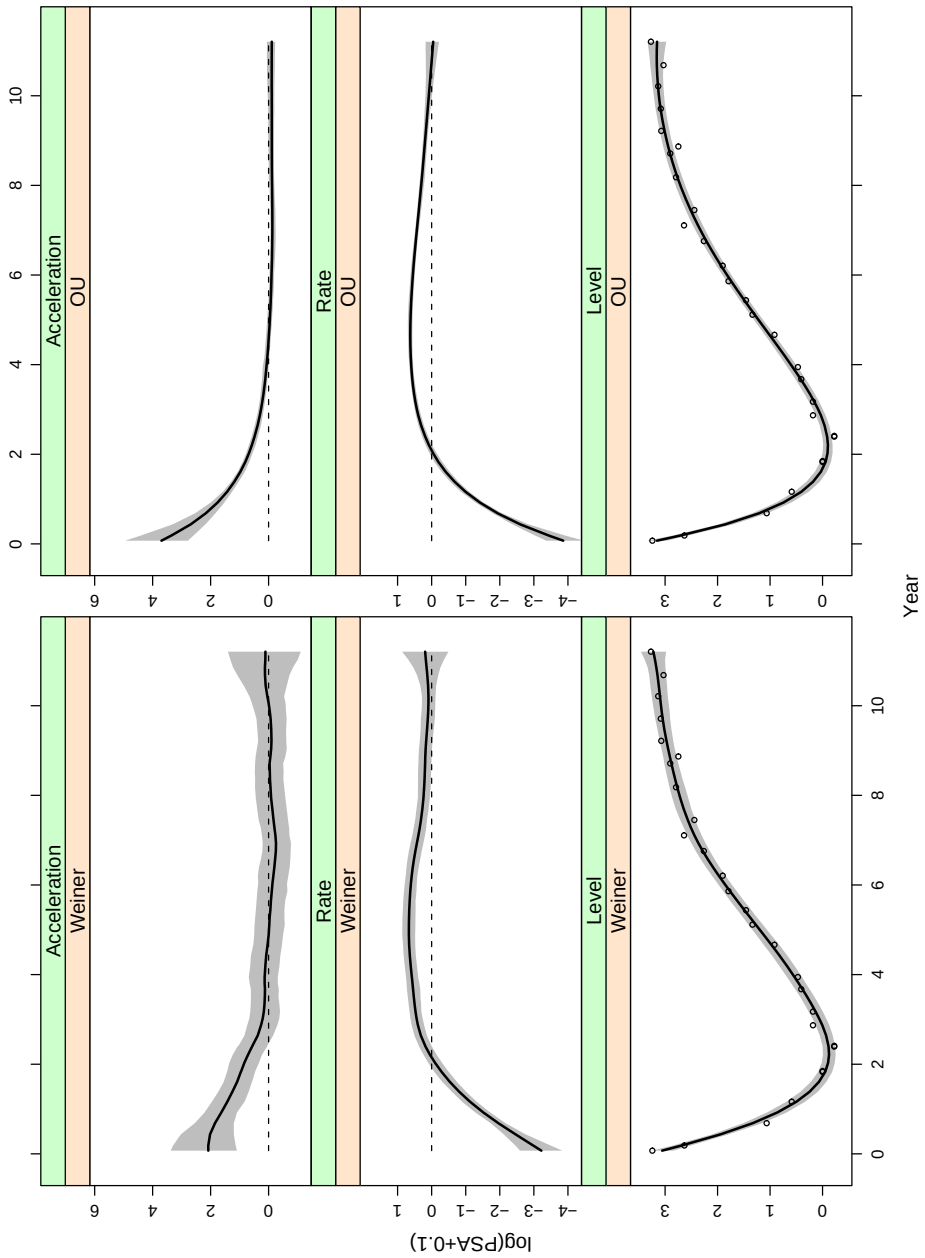


Figure 2.3: PSA: Plots of data points(∘), posterior means(—) and 95% credible intervals(gray shades) for the SAM with the Wiener process and OU process, respectively. In the graph, the upper panels show the acceleration $V(t)$, the middle panels show the rate $\dot{U}(t)$, and the lower panels show the level, $U(t)$.

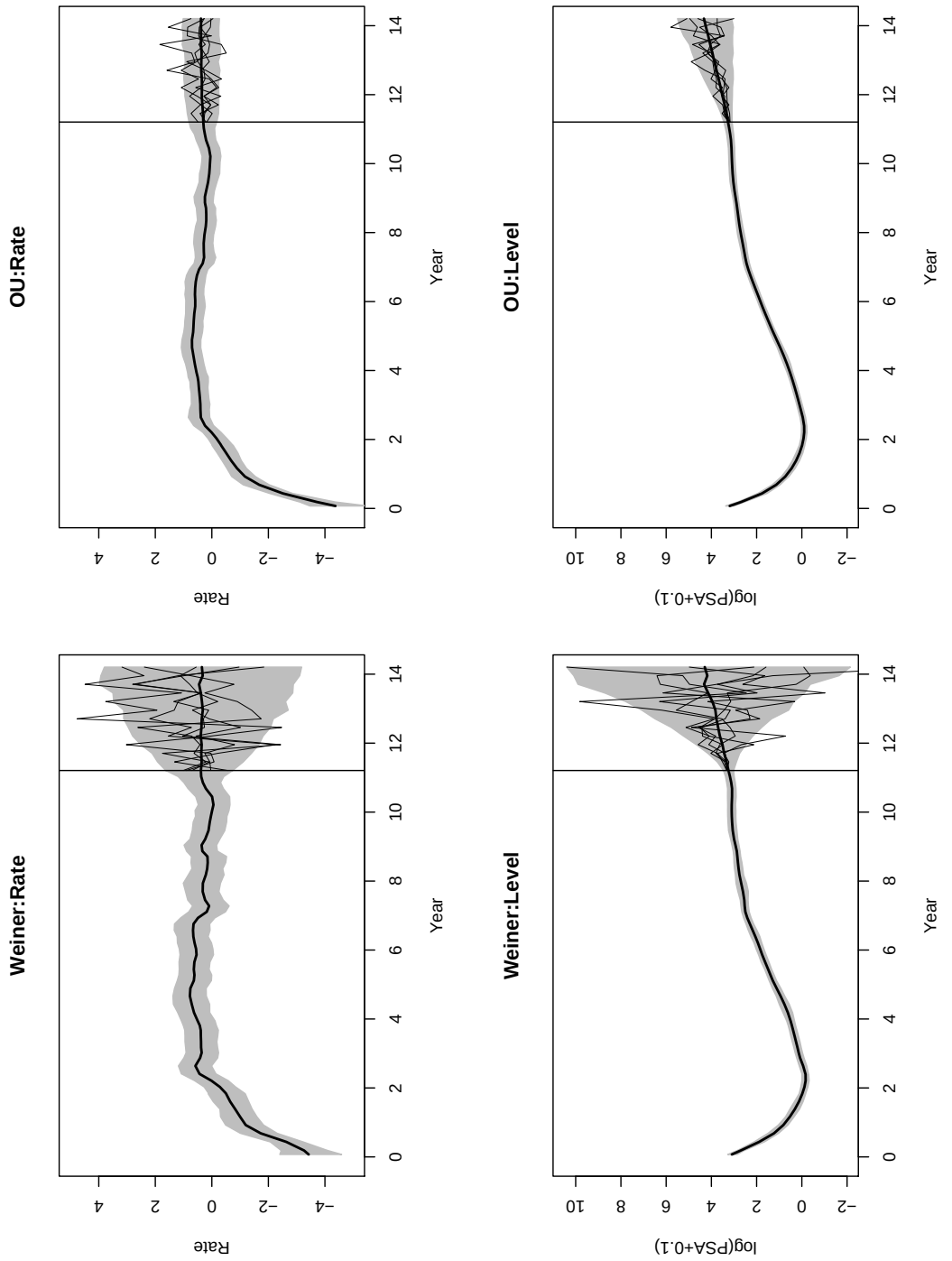


Figure 2.4: PSA Forecasting: Plots of posterior means(—) and 95% credible intervals(gray shades) for the SVM with the Wiener process and OU process, respectively, till year 11.2. The future rates and levels are forecasted for the next 3 years, illustrated by the forecasting means (—) and 95% forecasting credible intervals(gray shades). The 5 randomly picked realizations for each plot are also illustrated.

Table 2.1: 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

Table 2.2: PSA data:Posterior mean and quantiles for the SAMs.

	Wiener Process $\bar{D} = -34.812$, $P_D = 8.985$, $DIC = -25.827$					OU Process $\bar{D} = -38.867$, $P_D = 6.213$, $DIC = -32.654$				
	Mean	SD	2.5%	50%	97.5%	Mean	SD	2.5%	50%	97.5%
σ^2_ε	0.018	0.007	0.009	0.017	0.036	0.015	0.005	0.008	0.015	0.028
σ^2_ξ	0.386	0.408	0.074	0.275	1.327	0.011	0.095	0.000	0.002	0.043
$\bar{\nu}$						-0.119	0.048	-0.193	-0.122	-0.004
ρ						0.741	0.170	0.573	0.723	0.990

2.5 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 paper 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, a 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 application. 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.

As noted above, the SVM with the Wiener process corresponds to the smoothing spline with $m = 2$. If no augmented data are involved, the model can be rewritten as a linear mixed model for both situations of exact and approximate transition densities.

As shown in the Appendix B, when data are equally spaced, the latter case is identical to a linear spline model with truncated line function basis (Ruppert et al., 2003). In this sense, the linear spline model can be regarded as a numerical approximation of the smoothing spline. In addition, with no data augmentation, one can easily fit the SVM with the Wiener process using existing software for the linear mixed model.

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,

$$d\,\Phi\{Y(t) \mid \boldsymbol{F}^{\mathbf{T}}\boldsymbol{\theta}(t), \sigma_\varepsilon\}, \quad t \in \mathcal{T}_o, \tag{2.13}$$

where $\Phi(\cdot \mid U_G, \sigma_G)$ is the normal CDF with mean U_G and standard deviation σ_G. Then, (2.13) can be extended to,

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

where one specifies the corresponding observation distribution $\mathbb{F}$ in the exponential family, with state equations (2.2) and (2.3) unchanged.

CHAPTER III

Semiparametric Stochastic Modeling of the Rate Function in Longitudinal Studies

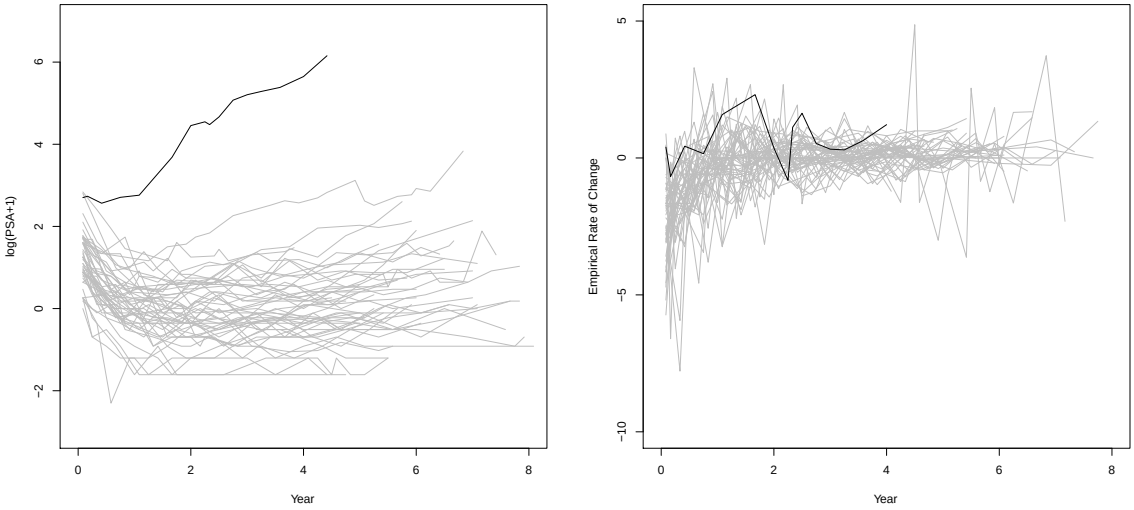
3.1 Introduction

This chapter 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 added measurement errors. The rate function describes the functional rate of change or slope with respect to time, a quantity which has been recently of great interest in longitudinal biomedical studies (Mungas et al., 2005; Lloyd-Jones et al., 2007; Strasak et al., 2008; Kariyanna et al., 2009). 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. A challenge in this research is to model the rate function without making a strong parametric assumption. Further challenges include modeling the rate functions across subjects as well as a function of the covariates of interest.

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

PSA is roughly 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 3.1(a) shows the log-transformed PSA level over time after EBRT treatment for 50 selected patients, and Figure 3.1(b) illustrates individual empirical rates of change, one for each subject. Figure 3.1(b) suggests that the individual rate of change in PSA follows roughly a common pattern. That is, it begins with a negative value caused by the EBRT, decreases over time in magnitude when 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, 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 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 be modeled nonparametrically (Rice and Silverman, 1991; Wang and Taylor, 1995; Zeger and Diggle, 1994; Zhang et al., 1998;



(a) (b)

Figure 3.1: 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.

Verbyla et al., 1999). Consequently, the resulting rate function, as the first order derivative of the mean function, does not have any parametric form, and usually it is not dependent on covariates. 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 $U_i(t) = \sum_{k=0}^{K} X_{ik}\beta_k(t)$ and $U_i(t) = X_{i0}\beta_0(t) + X_{i1}\beta_{i1}(t)$, respectively. Hence, $U_i(t)$ is a linear combination of several arbitrary smooth functions $\beta_k(t)$ or the summation of smooth functions $\beta_0(t)$ and $\beta_{i1}(t)$, with covariates X_{ik} as the weights. As the result, the rate function depends on covariates through linear combinations. 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 a 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). Formally, for each rate function $V_{\boldsymbol{x}_i}(t) \in \mathbb{R}$ with subject $i \in \mathcal{N} = \{1, 2, \ldots, n\}$ and time $t \in \mathcal{T}_s = [0, \infty)$, its prior is assumed to be a Gaussian process, conditional on $\boldsymbol{x}_i = (x_{i0}, x_{i1}, \ldots, x_{ip})^{\mathrm{T}}$, the vector of covariates for the ith subject. As an important special case of the proposed SSVM, we consider $V_{\boldsymbol{x}_i}(t) = f_{\boldsymbol{x}_i}(t) + \sigma_\xi W_i(t)$, with $f_{\boldsymbol{x}_i}(t)$ having a pre-specified parametric functional form dependent on covariates $\boldsymbol{x}_i$, and $\sigma_\xi W_i(t)$ a scaled standard Wiener process. Hence, $E\{V_{\boldsymbol{x}_i}(t)\} = f_{\boldsymbol{x}_i}(t)$ implies that $V_{\boldsymbol{x}_i}(t)$, the rate function of the ith subject, is expected to be centered about $f_{\boldsymbol{x}_i}(t)$, while the second term $\sigma_\xi W_i(t)$ allows deviations from the parametric functional expectation $f_{\boldsymbol{x}_i}(t)$.

The remainder of the chapter is organized as follows. Section 3.2 presents first the model and then is devoted to an important special case with the Ornstein-Uhlenbeck process as the prior for the rate function. Section 3.3 develops MCMC based methods for posterior inference and forecasting. Section 3.4 applies the methods to analyze the data of PSA profiles. Section 3.5 presents simulation results to evaluate and compare the performance of the proposed method with other existing methods. The chapter concludes with a discussion in Section 3.6.

3.2 Semiparametric Stochastic Velocity Model

3.2.1 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:

$$Y_i(t) = U_{\boldsymbol{x}_i}(t) + \varepsilon_i(t), \qquad t \in \mathcal{T}_{io} = \{t_{ij} : t_{i1} < t_{i2} < \cdots < t_{im_i}\}, \tag{3.1}$$

$$dU_{\boldsymbol{x}_i}(t) = V_{\boldsymbol{x}_i}(t)dt, \qquad t \in \mathcal{T}_s = [t_0, \infty), \tag{3.2}$$

$$dV_{\boldsymbol{x}_i}(t) = a\{V_{\boldsymbol{x}_i}(t); \boldsymbol{x}_i, \boldsymbol{\phi}_i\}dt + b\{V_{\boldsymbol{x}_i}(t); \boldsymbol{x}_i, \boldsymbol{\phi}_i\}dW_i(t), \quad t \in \mathcal{T}_s, \tag{3.3}$$

where $U_{\boldsymbol{x}_i}(t)$ is the mean function for the ith subject's outcome curve , $V_{\boldsymbol{x}_i}(t)$ 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 times $\mathcal{T}_s$, it is observed at discrete times $\mathcal{T}_{io}$ only and is subject to measurement error. Equation (3.3) may be regarded as a prior for the rate function $V_{\boldsymbol{x}_i}(t)$, in which the behavior of $V_{\boldsymbol{x}_i}(t)$ is governed by the stochastic differential equation(SDE), with drift term $a\{V_{\boldsymbol{x}_i}(t); \boldsymbol{x}_i, \boldsymbol{\phi}_i\}$ and diffusion term $b\{V_{\boldsymbol{x}_i}(t); \boldsymbol{x}_i, \boldsymbol{\phi}_i\}$, where $\boldsymbol{x}_i$ and $\boldsymbol{\phi}_i$ are the covariate vector and subject-specific parameter vector. We assume that the initial values $[U_{\boldsymbol{x}_i}(t_0), V_{\boldsymbol{x}_i}(t_0)]^{\mathrm{T}} \overset{\text{iid}}{\sim} \mathcal{N}_2(0, \sigma_0^2 \boldsymbol{I}_2)$ with large value of variance σ_0^2 to make it non-informative, and that the measurement error $\varepsilon_i(t) \overset{\text{iid}}{\sim} \mathcal{N}(0, \sigma_\varepsilon^2)$. Here $\boldsymbol{I}_k$ is the $k \times k$ identity matrix and $\mathcal{N}_k(\boldsymbol{m}, \boldsymbol{\Sigma})$ denotes the k-dimensional normal distribution with mean vector $\boldsymbol{m}$ and covariance matrix $\boldsymbol{\Sigma}$. Furthermore, $[U_{\boldsymbol{x}_i}(t_0), V_{\boldsymbol{x}_i}(t_0)]^{\mathrm{T}}$, $\varepsilon_i(t)$ and $W_i(t)$ are assumed mutually independent.

The SDE in equation (3.3) 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 $V_{\boldsymbol{x}_i}(t)$. According to the specific research interests or contexts of a given study, we can choose different forms for

$a\{V_{\boldsymbol{x}_i}(t); \boldsymbol{x}_i, \boldsymbol{\phi}_i\}$, which measures the instantaneous mean or the expected conditional acceleration, and for $b^2\{V_{\boldsymbol{x}_i}(t); \boldsymbol{x}_i, \boldsymbol{\phi}_i\}$, which reflects the instantaneous variance of the rate process. In particular, we have the SSVM-W, with $a\{V_{\boldsymbol{x}_i}(t); \boldsymbol{x}_i, \boldsymbol{\phi}_i\} = 0$ and $b\{V_{\boldsymbol{x}_i}(t); \boldsymbol{x}_i, \boldsymbol{\phi}_i\} = \sigma_\xi$, and the prior for $V_{\boldsymbol{x}_i}(t)$ is the Wiener process. Thus the resulting mean function takes the form $U_{\boldsymbol{x}_i}(t) = U_{\boldsymbol{x}_i}(t_0) + V_{\boldsymbol{x}_i}(t_0)(t - t_0) + \sigma_\xi \int_{t_0}^{t} W(s)ds$, 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 3.4, we specify $a\{V_{\boldsymbol{x}_i}(t); \boldsymbol{x}_i, \boldsymbol{\phi}_i\} = -\rho\{V_{\boldsymbol{x}_i}(t) - \bar{\nu}_i(\boldsymbol{x}_i, \boldsymbol{\beta})\}$ and $b\{V_{\boldsymbol{x}_i}(t); \boldsymbol{x}_i, \boldsymbol{\phi}_i\} = \sigma_\xi$. This specification corresponds to an Ornstein-Uhlenbeck(OU) process for $V_{\boldsymbol{x}_i}(t)$, and the resulting rate function is given by $V_{\boldsymbol{x}_i}(t) = f_{\boldsymbol{x}_i}(t) + \sigma_\xi W_i(t) = V_{\boldsymbol{x}_i}(t_0) - \int_{t_0}^{t} \rho\{V_{\boldsymbol{x}_i}(s) - \bar{\nu}_i(\boldsymbol{x}_i, \boldsymbol{\beta})\}\, ds + \sigma_\xi W_i(t)$. More details and properties of the OU process can be found in Section 3.2.2 below. We refer to this specification as SSVM-OU. For the PSA data analysis, it is of interest to estimate the stable rate $\bar{\nu}_i(\boldsymbol{x}_i, \boldsymbol{\beta})$, since $V_{\boldsymbol{x}_i}(t)$ will eventually stabilize and fluctuate around the level given by $\bar{\nu}_i(\boldsymbol{x}_i, \boldsymbol{\beta})$, 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 $\bar{\nu}_i(\boldsymbol{x}_i, \boldsymbol{\beta})$ and the patients' baseline characteristics, we propose a linear model $\bar{\nu}_i(\boldsymbol{x}_i, \boldsymbol{\beta}) = \nu_i + \boldsymbol{x}_i^{\mathrm{T}}\boldsymbol{\beta}$, where $\boldsymbol{\beta} = (\beta_0, \beta_1, \ldots, \beta_p)^{\mathrm{T}}$ is the vector of fixed effects parameters and $\nu_i \overset{\text{iid}}{\sim} \mathcal{N}(0, \sigma_\nu^2)$ are random effects. This subject-specific SSVM-OU is very useful to understand the dynamics of tumor growth, to assess the effect of coefficients, and to predict a patient's future PSA values using the baseline covariate information.

3.2.2 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 sys-

tem that fluctuates around some stable level and has been applied in biology (Trostab et al., 2009), 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. This process is particularly suitable for the PSA profiles considered in this paper, where the rate function of tumor growth reaches the stable level that may depend 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 OU and IOU processes are given by, respectively,

$$dU(t) = V(t)dt, \tag{3.4}$$

$$dV(t) = -\rho\{V(t) - \bar{\nu}\}dt + \sigma_\xi dW(t). \tag{3.5}$$

Theorem 1. For IOU and OU processes at time t_j, conditioned on the values at time t_{j-1} and parameters $\bar{\nu}, \rho, \sigma_\xi$. We have

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

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

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

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

The proof is in Appendix C.

Corollary 1. For $\delta_j \to \infty$ and fixed $\rho > 0$, such that $\exp(-\rho\delta_j) = \mathrm{o}(1)$, then the

conditional mean and variance from in Theorem 1 can be approximated by,

$$\boldsymbol{m}_j = [U_{j-1} + \bar{\nu}\delta_j, \bar{\nu}]^{\mathrm{T}} + \boldsymbol{R}_{\boldsymbol{m}_j}(1),$$
$$\boldsymbol{\Sigma}_j = \sigma_\xi^2 \begin{bmatrix} \frac{\delta_j}{\rho^2} - \frac{3}{2\rho^3} & \frac{1}{2\rho^2} \\ \frac{1}{2\rho^2} & \frac{1}{2\rho} \end{bmatrix} + \boldsymbol{R}_{\boldsymbol{\Sigma}_j}(1),$$

where the errors in the approximation $\boldsymbol{R}_{\boldsymbol{m}_j}(1) = [\mathrm{o}\big(1\big), \mathrm{o}\big(1\big)]^{\mathrm{T}}$ and $\boldsymbol{R}_{\boldsymbol{\Sigma}_j}(1) = \begin{bmatrix} \mathrm{o}\big(1\big) & \mathrm{o}\big(1\big) \\ \mathrm{o}\big(1\big) & \mathrm{o}\big(1\big) \end{bmatrix}$.

The proof is straightforward by noting that $\rho\delta_j \to \infty$ as δ_j satisfies $\exp(-\rho\delta_j) = \mathrm{o}\big(1\big)$.

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

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

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

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

3.3 Inference and Forecasting

3.3.1 Inference

In this section, we present Bayesian estimation for the mean function $U_{\boldsymbol{x}_i}(t)$, the rate function $V_{\boldsymbol{x}_i}(t)$ and parameters $\boldsymbol{\phi}_i$ and σ_ε with $i = 1, 2, \ldots, n$ and $t \in \mathcal{T}_{s_1} = [t : t_0 \le t \le t_m]$. Let $[\cdot \mid \cdot]$ denote the exact conditional density, $\prec \cdot \mid \cdot \succ$ the approximate

conditional density and $\boldsymbol{U} = [\boldsymbol{U}_1^{\mathrm{T}}, \boldsymbol{U}_2^{\mathrm{T}}, \ldots, \boldsymbol{U}_n^{\mathrm{T}}]^{\mathrm{T}}$ with $\boldsymbol{U}_i = [U_{i1}, U_{i2}, \ldots, U_{im_i}]^{\mathrm{T}}$. Similar notation is used for $\boldsymbol{V}$,$\boldsymbol{Y}$ and $\boldsymbol{x}$. For the model specified by equations (3.1), (3.2) and (3.3), we first consider the posterior density $[\boldsymbol{\phi} \mid \boldsymbol{U}, \boldsymbol{V}, \boldsymbol{Y}, \boldsymbol{x}]$ for $\boldsymbol{\phi}$, where $\boldsymbol{\phi} = [\boldsymbol{\phi}_1^{\mathrm{T}}, \boldsymbol{\phi}_2^{\mathrm{T}}, \ldots, \boldsymbol{\phi}_n^{\mathrm{T}}]^{\mathrm{T}}$. The posterior distribution is given by

$$[\boldsymbol{\phi} \mid \boldsymbol{U}, \boldsymbol{V}, \boldsymbol{Y}, \boldsymbol{x}] \propto \prod_{i=1}^{n} \prod_{j=1}^{m_i} [U_{ij}, V_{ij} \mid U_{i,j-1}, V_{i,j-1}, \boldsymbol{\phi}_i, \boldsymbol{x}_i][U_0, V_0][\boldsymbol{\phi}_i], \quad t \in \mathcal{T}_{io} \tag{3.6}$$

where $[U_{ij}, V_{ij} \mid U_{i,j-1}, V_{i,j-1}, \boldsymbol{\phi}_i, \boldsymbol{x}_i]$ is the exact transition density derived from the SDE in equation (3.3) and $[U_0, V_0]$ and $[\boldsymbol{\phi}_i]$ are non-informative prior densities. Unfortunately, except for a very few specific forms for the drift and diffusion terms in equation (3.3), $[U_{ij}, V_{ij} \mid U_{i,j-1}, V_{i,j-1}, \boldsymbol{\phi}_i, \boldsymbol{x}_i]$ 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, where the exact transition density is given in Theorem 1, the posterior density for $\boldsymbol{\phi}$ 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 with no measurement errors. Our approach is related to theirs, but with an important distinction that instead of being partially observed, both processes $V_{\boldsymbol{x}_i}(t)$ and $U_{\boldsymbol{x}_i}(t)$ are completely unobserved, and will be sampled as part of an MCMC algorithm. In this manner, we will be estimating the processes $V_{\boldsymbol{x}_i}(t)$ and $U_{\boldsymbol{x}_i}(t)$, as well as estimating the parameters $\boldsymbol{\phi}$.

To carry out the data augmentation and the Euler approximation, we first specify

time points at which data would be augmented. Let $\mathcal{T}_{ia} = [t : t = t_{ij} + k\tau_{ij}, \tau_{ij} = \frac{t_{i,j+1}-t_{ij}}{M_{ij}} < \tau_c, t \in (t_{ij}, t_{i,j+1}), k = 1, 2, \ldots, M_{ij}, j = 1, 2, \ldots, m_i - 1]$ denote 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 $\mathcal{T} = \cup_{i=1}^{n} (\mathcal{T}_{io} \cup \mathcal{T}_{ia}) = [t : t_j, j = 1, 2, \ldots, m]$ denote the set of all possible time points of the observed and augmented data across all subjects. With further data augmentation at times $t \in \mathcal{T}_{im} = [t : t \in \mathcal{T}, t \notin \mathcal{T}_{io}, t \notin \mathcal{T}_{ia}]$, each subject would have either observed or augmented data $\tilde{\boldsymbol{Y}}_i = [Y_{i1}, Y_{i2}, \ldots, Y_{im}]^{\mathrm{T}}$ at the common time set $\mathcal{T}$. The Euler approximation to equations (3.2) and (3.3) for $t \in \mathcal{T}$ leads to the following difference equations:

$$U_{ij} = U_{i,j-1} - V_{i,j-1}\delta_j, \tag{3.7}$$

$$V_{ij} = V_{i,j-1} + a\{V_{i,j-1}; \boldsymbol{x}_i, \boldsymbol{\phi}_i\}\delta_j + b\{V_{i,j-1}; \boldsymbol{x}_i, \boldsymbol{\phi}_i\}(W_j - W_{j-1}), \tag{3.8}$$

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

$$\prec \boldsymbol{\phi} \mid \tilde{\boldsymbol{U}}, \tilde{\boldsymbol{V}}, \tilde{\boldsymbol{Y}}, \boldsymbol{x} \succ \propto \prod_{i=1}^{n} \prod_{j=1}^{m} \prec U_{ij}, V_{ij} \mid U_{i,j-1}, V_{i,j-1}, \boldsymbol{\phi}_i, \boldsymbol{x}_i \succ [U_0, V_0][\boldsymbol{\phi}_i], \tag{3.9}$$

where $\tilde{\boldsymbol{U}} = [\tilde{\boldsymbol{U}}_1^{\mathrm{T}}, \tilde{\boldsymbol{U}}_2^{\mathrm{T}}, \ldots, \tilde{\boldsymbol{U}}_n^{\mathrm{T}}]^{\mathrm{T}}$ with $\tilde{\boldsymbol{U}}_i = [U_{i1}, U_{i2}, \ldots, U_{im}]^{\mathrm{T}}$ and similarly for $\tilde{\boldsymbol{V}}$ 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 (3.9) is given by,

$$\begin{aligned} \prec U_{ij}, V_{ij} \mid U_{i,j-1}, V_{i,j-1}, \boldsymbol{\phi}_i, \boldsymbol{x}_i \succ = \mathcal{N}_1 \left(V_{i,j-1} + a\{V_{i,j-1}; \boldsymbol{x}_i, \boldsymbol{\phi}_i\}\delta_j, b^2\{V_{i,j-1}; \boldsymbol{x}_i, \boldsymbol{\phi}_i\}\delta_j \right) \\ \times \boldsymbol{\delta}(U_{ij} - U_{i,j-1} - V_{i,j-1}\delta_j), \end{aligned} \tag{3.10}$$

which is derived from equations (3.7) and (3.8). 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 $t \in \mathcal{T}_{s_1}$, we follow equations (3.7) and (3.8) to come up with their approximations, denoted by $U^{(m)}_{\boldsymbol{x}_i}(t)$ and $V^{(m)}_{\boldsymbol{x}_i}(t)$, 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,

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

This indicates that if m is sufficiently large, then $V^{(m)}_{\boldsymbol{x}_i}(t)$ will approach to its continuous counterpart $V_{\boldsymbol{x}_i}(t)$ with arbitrary precision. Similar arguments hold for $U^{(m)}_{\boldsymbol{x}_i}(t)$. Note that we will sample m instead of m_i data points for $U^{(m)}_{\boldsymbol{x}_i}(t)$ and $V^{(m)}_{\boldsymbol{x}_i}(t)$ with possibly $m \gg m_i$. Hence, the benefit of introducing augmented data is two fold: (i) it reduces the error of approximation, when $U^{(m)}_{\boldsymbol{x}_i}(t)$ or $V^{(m)}_{\boldsymbol{x}_i}(t)$, instead of $U^{(m_i)}_{\boldsymbol{x}_i}(t)$ or $V^{(m_i)}_{\boldsymbol{x}_i}(t)$, is used to replace $U_{\boldsymbol{x}_i}(t)$ or $V_{\boldsymbol{x}_i}(t)$; (ii) it gives a more accurate approximation for the exact transition density, as suggested 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 $V^{(m)}_{\boldsymbol{x}_i}(t)$ throughout the rest of the chapter. $U_{\boldsymbol{x}_i}(t)$ is treated similarly.

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

$$\prec U_0, V_0, \tilde{\boldsymbol{U}}, \tilde{\boldsymbol{V}} \mid \tilde{\boldsymbol{Y}}, \boldsymbol{x}, \boldsymbol{\phi}, \sigma^2_\varepsilon \succ \propto \prod_{i=1}^{n} \prod_{j=1}^{m} [\tilde{Y}_{ij} \mid U_{ij}, \sigma^2_\varepsilon] \prec U_{ij}, V_{ij} \mid U_{i,j-1}, V_{i,j-1}, \boldsymbol{\phi}_i, \boldsymbol{x}_i \succ \times [U_0, V_0], \tag{3.11}$$

where $[\tilde{Y}_{ij} \mid U_{ij}, \sigma^2_\varepsilon] = \phi(U_{ij}, \sigma^2_\varepsilon)$, $\prec U_{ij}, V_{ij} \mid U_{i,j-1}, V_{i,j-1}, \boldsymbol{\phi}_i, \boldsymbol{x}_i \succ$ is given in equation

(3.10) and $[U_0, V_0]$ is non-informative prior. Equivalently, the posterior density (3.11) may be derived from a state space model representation(Durbin and Koopman, 2001), which is a useful reformulation of the SSVM in equations (3.1), (3.2) and (3.3) 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 $\prec U_{ij}, V_{ij} \mid U_{i,j-1}, V_{i,j-1}, \boldsymbol{\phi}_i, \boldsymbol{x}_i \succ$ is given in Corollary 2. Let $\tilde{\boldsymbol{Y}}_j = [\tilde{Y}_{1j}, \tilde{Y}_{2j}, \ldots, \tilde{Y}_{nj}]^{\mathrm{T}}$ denote the observed or augmented data for n subjects at time t_j, and let $\boldsymbol{\theta}_j = [\boldsymbol{\theta}_{1j}^{\mathrm{T}}, \boldsymbol{\theta}_{2j}^{\mathrm{T}}, \ldots, \boldsymbol{\theta}_{nj}^{\mathrm{T}}]^{\mathrm{T}}$ be the latent states with $\boldsymbol{\theta}_{ij} = [U_{\boldsymbol{x}_i}(t_j), V_{\boldsymbol{x}_i}(t_j), \bar{\nu}_i(\boldsymbol{x}_i, \boldsymbol{\beta})]^{\mathrm{T}}$. The corresponding SSVM can be expressed as a state space model, given as follows:

$$\begin{aligned} \tilde{\boldsymbol{Y}}_j &= \boldsymbol{F}_j^{\mathrm{T}} \boldsymbol{\theta}_j + \boldsymbol{\varepsilon}_j, \qquad & \boldsymbol{\varepsilon}_j &\sim \mathcal{N}_n(0, \sigma_\varepsilon^2 \boldsymbol{I}_n) \\ \boldsymbol{\theta}_j &= \boldsymbol{G}_j \boldsymbol{\theta}_{j-1} + \boldsymbol{\xi}_j, \qquad & \boldsymbol{\xi}_j &\sim \mathcal{N}_{3n}(0, \sigma_\xi^2 \boldsymbol{I}_n \otimes \boldsymbol{\Sigma}_j) \end{aligned}$$

where $\boldsymbol{F}_j = \boldsymbol{I}_n \otimes \boldsymbol{F}_{ij}$, $\boldsymbol{G}_j = \boldsymbol{I}_n \otimes \boldsymbol{G}_{ij}$, $\boldsymbol{F}_{ij} = [1, 0, 0]^{\mathrm{T}}$ with

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

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.

3.3.2 MCMC Algorithm

Under this state space model reformulation, both latent states $U_{\boldsymbol{x}_i}(t)$ and $V_{\boldsymbol{x}_i}(t)$ at times $t \in t_o \bigcup \mathcal{T}$ could be sampled by using the simulation smoother(de Jong and Shephard, 1995; Durbin and Koopman, 2002b), an efficient MCMC algorithm to sample the latent states simultaneously in blocks instead of one variable at a time,

and thus the convergence of the algorithm is fast.

The proposed MCMC algorithm iterates through the following steps.

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

2. Update latent states $U_{\boldsymbol{x}_i}(t)$ and $V_{\boldsymbol{x}_i}(t)$ for $t \in t_0 \bigcup \mathcal{T}$ from the posterior density (3.11) by using the simulation smoother.

3. Update $\boldsymbol{\phi}$ by sampling from the posterior density (3.9). In particular, when $V_{\boldsymbol{x}_i}(t)$ follows an OU process and is discretized through the Euler approximation, the collection of equations (3.8) could be equivalently reformulated as a linear mixed model,

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

where $\boldsymbol{Y}_j^* = \frac{\boldsymbol{V}_j - \boldsymbol{V}_{j-1}}{\sqrt{\delta_j}}$, $\boldsymbol{X}_j^* = [\boldsymbol{X}\sqrt{\delta_j}, \boldsymbol{V}_{j-1}\sqrt{\delta_j}]$, $\boldsymbol{Z}_j^* = -\sqrt{\delta_j}\boldsymbol{I}_n$ with $\boldsymbol{V}_j = [V_{1j}, V_{2j}, \ldots, V_{nj}]^{\mathrm{T}}$ and $\boldsymbol{X} = [\boldsymbol{x}_1^{\mathrm{T}}, \boldsymbol{x}_2^{\mathrm{T}}, \ldots, \boldsymbol{x}_n^{\mathrm{T}}]^{\mathrm{T}}$. Further, $\boldsymbol{\beta}^* = [\rho\boldsymbol{\beta}^{\mathrm{T}}, -\rho]^{\mathrm{T}}$, $\boldsymbol{b}^* = \rho\boldsymbol{\nu}$, $\boldsymbol{\nu} = [\nu_1, \nu_2, \ldots, \nu_n]^{\mathrm{T}}$, $\boldsymbol{\xi}_j^* \sim \mathcal{N}_n(0, \sigma_\xi^2 \boldsymbol{I}_n)$, $\boldsymbol{b}^* \sim \mathcal{N}_n(0, \rho^2\sigma_\nu^2 \boldsymbol{I}_n)$. As a result, the set of model parameters is $\boldsymbol{\phi}^* = [\boldsymbol{\beta}^*, \boldsymbol{b}^*, \sigma_\xi^2, \rho^2\sigma_\nu^2]^{\mathrm{T}}$, which 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, $\boldsymbol{\beta}^* \sim \mathcal{N}_{p+2}(0, \sigma_{\beta^*}^2 \boldsymbol{I}_{p+2})$, $\sigma_\xi^2 \sim \mathcal{IG}(A_\xi, B_\xi)$, and $\rho^2\sigma_\nu^2 \sim \mathcal{IG}(A_{b^*}, B_{b^*})$. Here $\mathcal{IG}(A, B)$ denotes the inverse gamma distribution with shape parameter A and scale parameter B.

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

$$[\sigma_\varepsilon^2 \mid \tilde{\boldsymbol{U}}, \tilde{\boldsymbol{V}}, \tilde{\boldsymbol{Y}}, \boldsymbol{x}] \sim \mathcal{IG}(A_\varepsilon + \frac{1}{2}mn, B_\varepsilon + \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{m}(Y_i(t_j) - U_{\boldsymbol{x}_i}(t_j))^2),$$

where the prior distribution for σ_ε^2 is $\mathcal{IG}(A_\varepsilon, B_\varepsilon)$.

3.3.3 Bayesian Posterior Forecasting

The proposed model is useful to forecast processes of interest, including $U_{\boldsymbol{x}_i}(t)$, $V_{\boldsymbol{x}_i}(t)$ and $Y_i(t)$, for $t \in \mathcal{T}_{s_2} = [t : t > t_m]$. With the availability of posterior samples for $U_{\boldsymbol{x}_i}(t)$, $V_{\boldsymbol{x}_i}(t)$, $\boldsymbol{\phi}_i$ and σ_ε with $i = 1, 2, \ldots, n$ and $t \in \mathcal{T}_{s_1} = [t : t_1 \leq t \leq t_m]$, it is straightforward to derive Bayesian posterior forecasting. Note that the posterior forecasting distributions are,

$$[U_{\boldsymbol{x}_i}(t), V_{\boldsymbol{x}_i}(t) \mid \boldsymbol{Y}, \boldsymbol{x}] = \int\int\int [U_{\boldsymbol{x}_i}(t), V_{\boldsymbol{x}_i}(t) \mid U_{\boldsymbol{x}_i}(t_m), V_{\boldsymbol{x}_i}(t_m), \boldsymbol{\phi}_i, \boldsymbol{x}] \times$$
$$[U_{\boldsymbol{x}_i}(t_m), V_{\boldsymbol{x}_i}(t_m), \boldsymbol{\phi}_i \mid \boldsymbol{Y}, \boldsymbol{x}] dU_{\boldsymbol{x}_i}(t_m) dV_{\boldsymbol{x}_i}(t_m) d\boldsymbol{\phi}_i,$$

and

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

Thus, we draw $U^r_{\boldsymbol{x}_i}(t)$, $V^r_{\boldsymbol{x}_i}(t)$ and $Y^r_i(t)$ from $[U^r_{\boldsymbol{x}_i}(t), V^r_{\boldsymbol{x}_i}(t) \mid U^r_{\boldsymbol{x}_i}(t_m), V^r_{\boldsymbol{x}_i}(t_m), \boldsymbol{\phi}^r_{is}, \boldsymbol{x}]$ and $[Y^r_i(t) \mid U^r_{\boldsymbol{x}_i}(t), \sigma^{2\,r}_\varepsilon]$ for $r = 1, 2, \ldots, R$, where $[U^r_{\boldsymbol{x}_i}(t_m), V^r_{\boldsymbol{x}_i}(t_m)]^{\mathrm{T}}$,$\boldsymbol{\phi}^r_{is}$ and $\sigma^{2\,r}_\varepsilon$ are the rth posterior samples from the MCMC algorithm. When $[U_{\boldsymbol{x}_i}(t), V_{\boldsymbol{x}_i}(t) \mid U_{\boldsymbol{x}_i}(t_m), V_{\boldsymbol{x}_i}(t_m), \boldsymbol{\phi}_i, \boldsymbol{x}]$ does not have a closed form, the approximate transition density $\prec U_{\boldsymbol{x}_i}(t), V_{\boldsymbol{x}_i}(t) \mid U_{\boldsymbol{x}_i}(t_m), V_{\boldsymbol{x}_i}(t_m), \boldsymbol{\phi}_i, \boldsymbol{x} \succ$ could be used instead along with data augmentation.

3.4 Application to the PSA Data

We apply the proposed SSVM-OU to analyze the PSA data discussed in Section 3.1. The prior of the rate function $V_{\boldsymbol{x}_i}(t)$ is assumed to be the OU process with $a\{V_{\boldsymbol{x}_i}(t); \boldsymbol{x}_i, \boldsymbol{\phi}_i\} = -\rho\{V_{\boldsymbol{x}_i}(t) - \bar{\nu}_i(\boldsymbol{x}_i, \boldsymbol{\beta})\}$ and $b\{V_{\boldsymbol{x}_i}(t); \boldsymbol{x}_i, \boldsymbol{\phi}_i\} = \sigma_\xi$ in equation (3.3).

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 after EBRT treatment is obtained at the first month or 0.083 year, 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 3.3.1, 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 3.5. We investigate the association of the pretreatment covariates baseline PSA, Gleason score and T stage with the stable PSA rate via the model $\bar{\nu}_i(\boldsymbol{x}_i, \boldsymbol{\beta}) = \nu_i + \beta_0 + \beta_1 X_{Pi} + \beta_2 X_{Ti} + \beta_3 X_{Gi}$, where the random effect $\nu_i \sim \mathcal{N}(0, \sigma_\nu^2)$; 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, 0 otherwise ; X_{Ti} takes the value of 1 if T stage is at level 2 or higher, 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 3.1 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 3.2 displays $E[V_{\boldsymbol{x}_i}(t) \mid \boldsymbol{Y}]$, the posterior means of the rate function for each subject (shown as dashed lines), and $E[V(t) \mid \boldsymbol{Y}] = E[E[V_{\boldsymbol{x}_i}(t) \mid \boldsymbol{Y}]]$, the posterior mean of the rate function in the population(shown as solid line). It is clear that although the rate function in the population level 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 3.3 shows the posterior means and credible intervals of $U_{\boldsymbol{x}_i}(t)$ for six randomly selected subjects, including the forecasted $U_{\boldsymbol{x}_i}(t)$ after year 5. Note that the width of the forecasted credible intervals is comparable to the theoretical results given in Corollary 1.

Table 3.1: PSA data: Posterior mean 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

For comparison, we also analyze the PSA data using smoothing splines and a parametric linear mixed model(LMM). The model fits are evaluated by the Deviance Information Criterion(DIC, Speigelhalter et al., 2003). 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 $V_{\boldsymbol{x}_i}(t)$ from the SSVM-W with Wiener process as the prior for $V_{\boldsymbol{x}_i}(t)$, where $a\{V_{\boldsymbol{x}_i}(t); \boldsymbol{x}_i, \boldsymbol{\phi}_i\} = 0$ and $b\{V_{\boldsymbol{x}_i}(t); \boldsymbol{x}_i, \boldsymbol{\phi}_i\} = \sigma_\xi$ in equation (3.3). As mentioned in Section 3.2.1, the estimation of $V_{\boldsymbol{x}_i}(t)$ from this model, is equivalent to the estimation by a smoothing spline with a common smoothing parameter $\lambda = \frac{\sigma_\xi^2}{\sigma_\varepsilon^2}$. The exact transition density in this SSVM-W, is given by

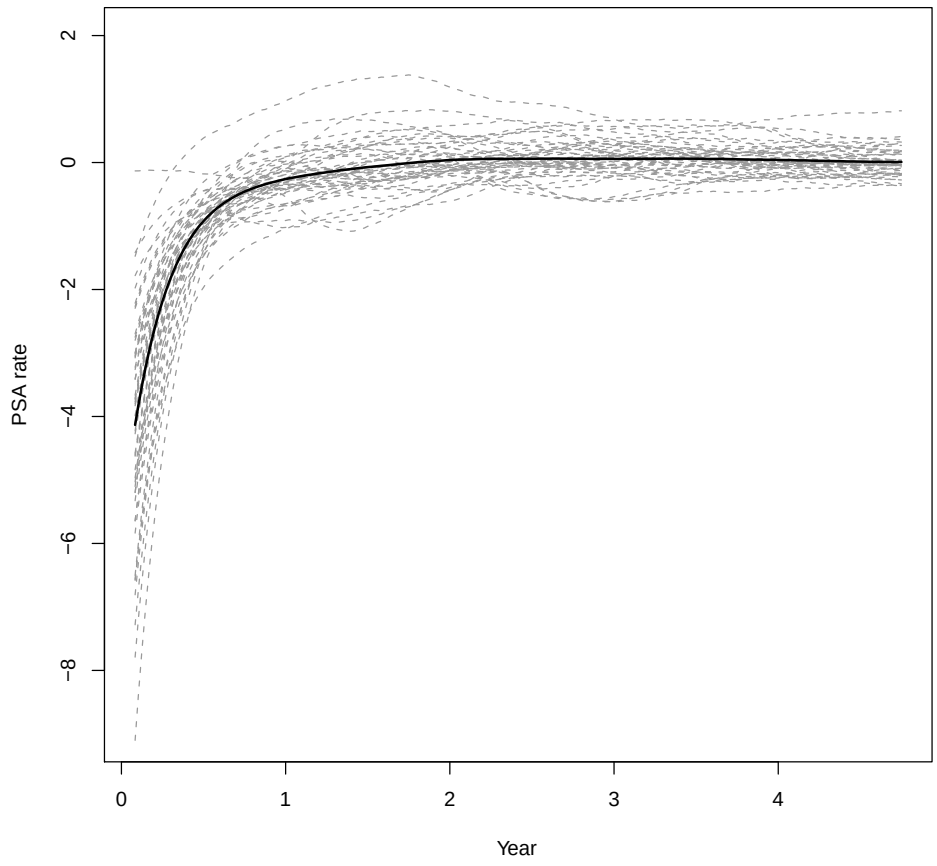


Figure 3.2: Posterior means of $V_{\boldsymbol{x}_i}(t)$ for each subject as gray dash lines and the population-level rate function $V(t)$ as black solid line

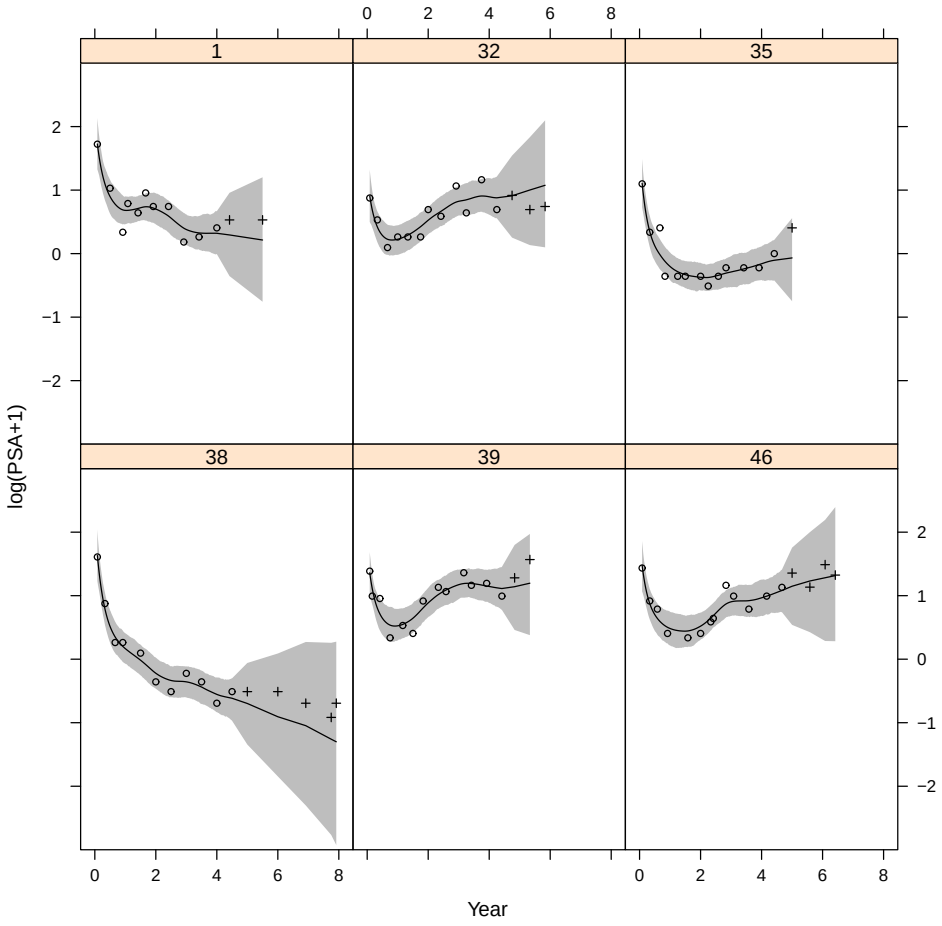


Figure 3.3: Plots of training data points(∘), validation data points(+), posterior means(—) and 95% credible intervals(gray shades) of $U_{\boldsymbol{x}_i}(t)$ for six randomly selected subjects.

Wecker and Ansley (1983) as

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

with

$$\boldsymbol{m}_j = [U_{j-1} + V_{j-1}\delta_j, V_{j-1}]^{\mathrm{T}},$$
$$\boldsymbol{V}_j = \sigma_\xi^2 \begin{bmatrix} \frac{\delta_j^3}{3} & \frac{\delta_j^2}{2} \\ \frac{\delta_j^2}{2} & \delta_j \end{bmatrix},$$

and will be used in the proposed MCMC algorithm. The forecasting of future observations is outlined in Section 3.3.3 for the SSVM-OU and SSVM-Ws. The parametric linear mixed model is specified similarly to the one given by Proust-Lima et al. (2008a),

$$\begin{aligned} Y_i(t_{ij}) &= U_{\boldsymbol{x}_i}(t_{ij}) + \varepsilon_i(t_{ij}) \\ &= U^0_{\boldsymbol{x}_i}(t_{ij}) + U^1_{\boldsymbol{x}_i}(t_{ij}) + U^2_{\boldsymbol{x}_i}(t_{ij}) + \varepsilon_i(t_{ij}) \\ &= (\beta_{00} + \nu_{0i} + \beta_{01}X_{Pi}) + (\beta_{10} + \nu_{1i} + \beta_{11}X_{Pi} + \beta_{12}X_{Ti})f_1(t_{ij}) \\ &+ (\beta_{20} + \nu_{2i} + \beta_{21}X_{Pi} + \beta_{22}X_{Ti} + \beta_{23}X_{Gi})f_2(t_{ij}) + \varepsilon_i(t_{ij}), \end{aligned} \tag{3.12}$$

where the mean function $U_{\boldsymbol{x}_i}(t)$ consists of three parts: (i) post-therapy level $U^0_{\boldsymbol{x}_i}(t)$, (ii) short-term evolution $U^1_{\boldsymbol{x}_i}(t)$, and (iii) long-term evolution $U^2_{\boldsymbol{x}_i}(t)$. In addition, $f_1(t) = (1+t)^{-1.5} - 1$ and $f_2(t) = t$; the fixed effects $\boldsymbol{\beta}_{lmm} = [\beta_{00}, \beta_{01}, \beta_{10}, \beta_{11}, \beta_{12}, \beta_{20}, \beta_{21}, \beta_{22}, \beta_{23}]^{\mathrm{T}} \sim \mathcal{N}_9(\boldsymbol{0}, \sigma^2_{\beta,lmm}\boldsymbol{I}_9)$ a non-informative prior with large value of $\sigma^2_{\beta,lmm}$; the random effects $[\nu_{0i}, \nu_{1i}, \nu_{2i}]^{\mathrm{T}} \sim \mathcal{N}_3(\boldsymbol{0}, \Sigma_{\nu,lmm})$ with $\Sigma_{\nu,lmm}$ a diagonal matrix with its main diagonal entries $\boldsymbol{\nu}_{lmm} = [\sigma^2_{0\nu,lmm}, \sigma^2_{1\nu,lmm}, \sigma^2_{2\nu,lmm}]^{\mathrm{T}}$; measurement error $\varepsilon_i(t_{ij}) \sim \mathcal{N}_1(0, \sigma^2_{\varepsilon,lmm})$. We further assume noninformative prior distributions $\mathcal{IG}(A, B)$ 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 $\sigma^2_{\varepsilon,lmm}$,

respectively. The MCMC algorithm for the linear mixed model (Ruppert et al., 2003, Chap. 16) is conducted to draw the posterior samples with the same burn-in stage and thinning scheme as for the MCMC algorithm for the SVC-OU model. Table 3.1 presents the posterior summary of the parameters β_{20}, β_{21}, β_{22}, and β_{23}, which are involved in the long-term evolution $U^2_{\boldsymbol{x}_i}(t)$ in equation (3.12). Note that for those parameters in LMM, they are aimed 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 $\boldsymbol{\beta}^r_{lmm}$, $\boldsymbol{\nu}^r_{lmm}$ and $\sigma^{2\,r}_{\varepsilon,lmm}$, the forecasts of PSA at time t for the ith subject can be drawn from $Y^r_i(t) \sim \mathcal{N}(U^r_{\boldsymbol{x}_i}(t), \sigma^{2\,r}_{\varepsilon,lmm})$, where $U^r_{\boldsymbol{x}_i}(t) = (\beta^r_{00} + \nu^r_{0i} + \beta^r_{01}X_{Pi}) + (\beta^r_{10} + \nu^r_{1i} + \beta^r_{11}X_{Pi} + \beta^r_{12}X_{Ti})f_1(t) + (\beta^r_{20} + \nu^r_{2i} + \beta^r_{21}X_{Pi} + \beta^r_{22}X_{Ti} + \beta^r_{23}X_{Gi})f_2(t)$.

Among these three models, the SSVM-OU fits the data best, with DIC equal to 71.809, which is less than 119.400 and 151.048, the DICs of the SSVM-W and the linear mixed model, respectively. Next, we predict the 164 validation data points and evaluate the posterior predictive ability of the three models. Table 3.2 presents the magnitude of biases and mean squared errors(MSE) of the point forecasts, for which we use the posterior means, and the coverage rate and averaged length of credible intervals. For the 69 validation data points within 1 year distance from the last training data points, the SSVM-OU performs best, in term of smaller Bias and MSE, higher coverage rate and narrower interval length. For the later time validation data points, the SSVM-W outperforms the other two in terms of bias and MSE. Overall the coverage rates of the SSVM-OU intervals are closest to the nominal 95% level. Note that the interval lengths from the SSVM-W are significantly wider than those of SSVM-OU and LMM.

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

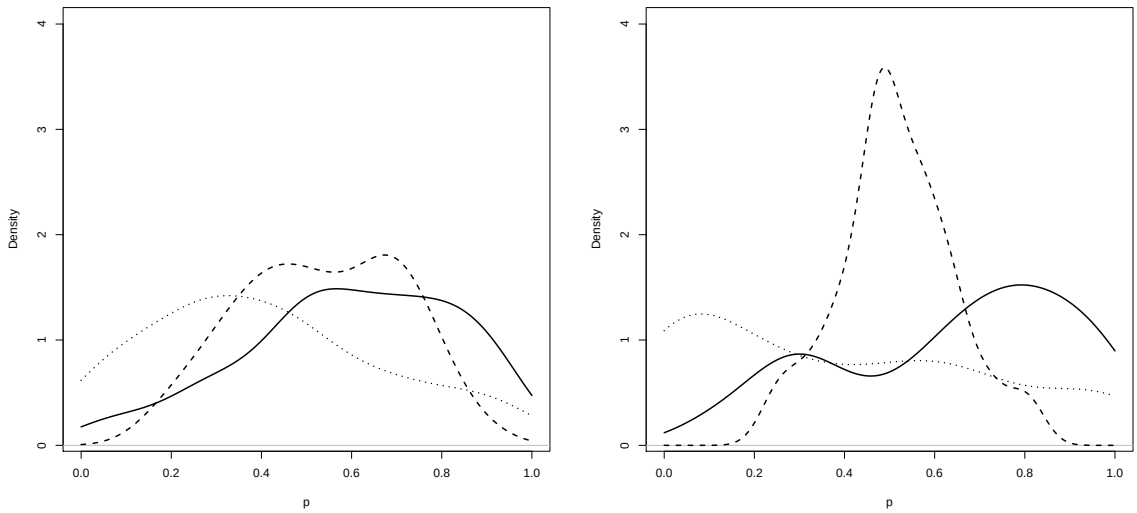
forecast can be expressed as the posterior predictive cumulative distribution functions (CDFs) $F_{ij}(Y)$, where Y is the forecasted validation data point at time t_{ij} for the ith subject and is assumed to be generated from the true unknown CDF $G_{ij}(Y)$. For the observed validation data point Y_{ij}, the PIT value $p_{ij} = F_{ij}(Y_{ij})$ should has 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 of the three models. The corresponding smoothed density plots of $\tilde{p}_{ij}$ are displayed in Figure 3.4. 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.

Table 3.2: PSA data: Posterior forecasting of the validation data points.

Method	Type	Bias	MSE	Coverage Rate	Interval Length
SSVM-OU	≤ 1 year	0.225	0.085	1	1.404
	> 1 year	0.537	0.489	0.905	2.378
	All	0.406	0.319	0.945	1.968
SSVM-W	≤ 1 year	0.250	0.098	1	2.383
	> 1 year	0.479	0.403	1	8.426
	All	0.383	0.275	1	5.883
LMM	≤ 1 year	0.246	0.110	0.899	1.240
	> 1 year	0.520	0.476	0.726	1.626
	All	0.405	0.322	0.799	1.464

3.5 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 $\bar{\nu}_i(\boldsymbol{x}_i, \boldsymbol{\beta})$; (ii)



(a) (b)

Figure 3.4: PIT density plots for (a) $t_{ij} \leq 1$ year, (b) $t_{ij} > 1$ year of SSVM-OU (—), SSVM-W $(- - -)$, LMM $(\cdots)$

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 include 20 subjects with 14 equidistant observations and three validation data points per subject. The three validation data points are 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 and stable rates, we analyze the same dataset using the proposed MCMC algorithm without data augmentation, and with one and three augmented data points between the consecutive observed data points. The corresponding time interval between the adjacent data points, either observed or augmented, decreases from 0.08 in the original datasets to 0.04 and 0.02 for the MCMC algorithm with one and three augmented data points between neighboring observations.

Table 3.3 presents simulation results of model parameters and stable rates, assessed by the magnitude of biases and mean squared errors of the posterior means

and the coverage rate and average length of credible intervals. Those results indicate that the data augmentation is critical to estimation of the parameter ρ. With data augmentation, the bias of ρ reduces from 0.394 to 0.014 and the coverage rate increases from 0.12 to 0.90. The data augmentation also moderately improves the estimation of σ_ξ^2 and σ_ν^2 but has little effect on the estimation of other parameters. Note that when we augment three data points between the adjacent observations, the estimates of the parameters of interest, β_1,β_2 and β_3 are virtually unbiased with satisfactory coverage rates.

For the data simulated from the SSVM-OU, we further forecast the validation data points by the SSVM-OU. Those data points are also predicted by the SSVM-W and LMM (3.12). Table 3.4 compares the forecasting ability of posterior mean and credible intervals for those three models, evaluated by Bias, MSE, coverage rate and interval length. As we expected, the 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 robustness or sensitivity of the forecasting ability of the SSVM-OU. We simulate another 100 datasets from the LMM specified as equation (3.12) in which the parameters are the same as those estimated from the analysis of real PSA data. The simulation design, including the number of subjects and observation and the position of validation data points, are identical to those of the dataset generated from the SSVM-OU. The forecasting results are given in the second part of Table 3.4. We find that SSVM-OU forecasts relatively well. Especially, when the time distance is 0.08, the forecasts of SSVM-OU shows the same bias, smaller MSE, slightly lower coverage rate and obviously shorter interval length, compared to the forecasts from the true LMM. The forecasts from the SSVM-W are inferior to those by the SSVM-OU and the LMM.

3.6 Discussion

This chapter considers modeling and inference for the rate function 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 stochastic process is specified by an SDE and is expected to be centered on a pre-specified parametric function, while allowing significant deviation 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.

We propose an MCMC algorithm to estimate the posterior distribution of the model parameters and rate functions. We apply the Euler approximation to facilitate the sampling of the model parameters and use data augmentation of reduce the approximation error. The accuracy of this approximation is supported by the results from the simulation studies. We demonstrate that our proposed model has superior forecasting ability, as least for the case when the rate function will evolve to some stable level in the long term.

One can extend our model to discrete outcomes and to include the covariates in equation (3.1). Moreover, a similar modeling and inference approach can be applied to analyze the acceleration function, which is the third-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. In future research, it will also be interesting to consider alternative efficient algorithms to sample the parameters and rate functions without relying on the Euler approximation.

Table 3.3: Simulation results on the estimation of SSVM-OU parameters and stable rates.

Data Augmented	Parameter	Truth	Bias	MSE $\times 10^{-2}$	Coverage Rate	Interval Length
0	σ_ε^2	0.05	0.001	0.003	0.93	0.019
	σ_ξ^2	1.5	0.394	17.845	0.95	1.545
	ρ	4	0.469	24.971	0.12	0.574
	σ_ν^2	0.05	0.042	0.192	0.99	0.172
	β_0	-0.2	0.128	14.615	0.98	1.481
	β_1	0.15	0.021	5.747	0.95	0.914
	β_2	0.25	0.065	11.267	0.96	1.277
	β_3	0.05	0.006	9.934	0.97	1.325
	$\bar{\nu}_i$	$\boldsymbol{X}_i^{\mathrm{T}}\boldsymbol{\beta}$[a]$+\nu_i$	0.197	18.618	0.96	1.747
1	σ_ε^2	0.05	0.001	0.003	0.94	0.019
	σ_ξ^2	1.5	0.357	17.930	0.96	1.585
	ρ	4	0.173	7.302	0.80	0.687
	σ_ν^2	0.05	0.030	0.111	1.00	0.142
	β_0	-0.2	0.127	15.221	0.94	1.366
	β_1	0.15	0.025	5.699	0.95	0.859
	β_2	0.25	0.065	11.671	0.94	1.168
	β_3	0.05	0.008	9.832	0.98	1.259
	$\bar{\nu}_i$	$\boldsymbol{X}_i^{\mathrm{T}}\boldsymbol{\beta}$[a]$+\nu_i$	0.202	19.237	0.95	1.675
3	σ_ε^2	0.05	0.001	0.003	0.93	0.019
	σ_ξ^2	1.5	0.294	14.862	0.93	1.585
	ρ	4	0.014	4.294	0.90	0.687
	σ_ν^2	0.05	0.025	0.079	1.00	0.142
	β_0	-0.2	0.217	16.706	0.88	1.366
	β_1	0.15	0.022	4.808	0.96	0.859
	β_2	0.25	0.005	9.336	0.93	1.168
	β_3	0.05	0.003	8.735	0.98	1.259
	$\bar{\nu}_i$	$\boldsymbol{X}_i^{\mathrm{T}}\boldsymbol{\beta}$[a]$+\nu_i$	0.215	19.803	0.93	1.643

[a] $\boldsymbol{X}_i^{\mathrm{T}}\boldsymbol{\beta} = \beta_0 + \beta_1 X_{Pi} + \beta_2 X_{Ti} + \beta_3 X_{Gi}$

Table 3.4: Simulation results on forecasting of three models

Simulation Model	Fitted Model	Year Distance	Bias	MSE	Coverage Rate	Interval Length
SSVM-OU	SSVM-OU	0.08	0.028	0.068	0.947	1.026
		0.5	0.091	0.138	0.948	1.427
		1	0.202	0.335	0.920	2.081
	SSVM-W	0.08	0.040	0.142	0.993	1.986
		0.5	0.240	1.491	1	14.954
		1	0.595	5.533	1	39.040
	LMM	0.08	0.283	0.342	0.920	2.028
		0.5	1.456	2.444	0.235	2.174
		1	3.494	12.679	0.008	2.500
LMM	SSVM-OU	0.08	0.015	0.016	0.919	0.438
		0.5	0.049	0.023	0.995	0.826
		1	0.105	0.040	0.998	1.360
	SSVM-W	0.08	0.062	0.038	0.988	0.952
		0.5	0.386	0.699	1	7.073
		1	0.759	2.654	1	18.446
	LMM	0.08	0.014	0.055	0.957	0.948
		0.5	0.012	0.055	0.957	0.959
		1	0.014	0.057	0.956	0.975

CHAPTER IV

Signal extraction and breakpoint identification for array CGH data using robust state space model

4.1 Introduction

Almost all types of cancer share one common characteristic, genetic instability, including DNA copy number variation(CNV). During cancer progression some genes will lose one of the two copies or are completely deleted, while others may gain one copy, or become amplified up to hundreds of copies. These chromosomal alterations can lead to abnormal cell proliferation, DNA repair, senescence and apoptotic mechanisms and can provide a selective advantage for cells and result in cancer. Identification of CNV not only enhances the understanding of oncogenesis but also facilitates the treatment of cancer. For example, Trastuzumab is a monoclonal antibody interfering with ERBB2 receptor and is used for the treatment of breast cancers with amplified, and multiple copies of the ERBB2 gene (Vogel et al., 2002).

Array comparative genomic hybridization (CGH) is a technique that is used to detect differences in DNA copy number (Solinas-Toldo et al., 1997; Pinkel et al., 1998). The isolated DNA from tumor and the normal tissue from each patient are labeled with different fluorescent dyes and then cohybridized to the microarray. The $\log_2$ fluorescent intensity ratios are measured at different chromosomal positions to

define each CGH profile. This CGH profile is supposed to be proportional to the copy number ratio for tumor and normal cells across the chromosome. See Pinkel and Albertson (2005) and Davies et al. (2005) for detail reviews. Array CGH data exhibit three challenging characteristics. First, the data displays abrupt changes at the positions where DNA copy number is possibly altered. Second, the data usually contain biological variations and experimental errors, which hinder the accurate identification of breakpoints where copy number changes. Biological variations refer to heterogeneity of copy number within tumor cells and experimental errors include contamination of the tumor cells with normal cells, measurement errors and errors caused by processing tissue samples. Third, the data are spatially dependent. That is, neighboring genes are more likely to share the same copy number than remote ones. The primary aim of array CGH data analysis is to estimate the CGH profiles and to identify breakpoints from available noisy observations.

A number of statistical methods have been proposed for array CGH data analysis. Most of the methods postulate that the observed $\log_2$ intensity ratio $Y(t_j)$ is governed by the following model,

$$Y(t_j) = U(t_j) + \varepsilon(t_j), \quad j = 1, 2, \ldots, J \tag{4.1}$$

where signal $U(t_j)$ is the unobserved $\log_2$ intensity ratio at jth probe, $\varepsilon(t_j)$ is the noise or measurement error and t_j denotes the physical position of jth probe on a chromosome. Different assumptions and interpretations of $U(t_j)$ and $\varepsilon(t_j)$ lead to various estimation approaches, which may be categorized into three types. The first type is based on the segmentation method. It assumes that the CGH profile $U(t)$, is piecewise constant, i.e. $U(t_j) = \sum_{m=1}^{M} U_m I[t_j \in \mathcal{T}_m]$, where $\mathcal{T}_m$ is segment m with level U_m and $I(\cdot)$ is the indicator function. Also $\varepsilon(t_j)$ follows independent and identically distributed (i.i.d.) $\mathcal{N}(0, \sigma_\varepsilon^2)$. To detect breakpoints that enable us to

classify chromosome into blocks, Olshen et al. (2004) and Venkatraman and Olshen (2007) proposed the method of circular binary segmentation(CBS); Hupe et al. (2004) developed the adaptive weighted smoothing procedure; and Erdman and Emerson (2008) implemented a Bayesian change point model.The second type is the method of hidden Markov models (HMM), which restricts $U(t)$ to take a finite number of values and uses a Markov chain to model probabilities: $\Pr(U(t_{j+1}) = U_{m'} \mid U(t_j) = U_m)$, $U_m, U_{m'} \in \mathcal{U}$, with state space $\mathcal{U} = \{U_m; m = 1, 2, \ldots, M\}$. Note that M is a prespecified number of states. The HMM method was first applied to array CGH data analysis by Fridlyand et al. (2004). Shah et al. (2006) modified the HMM method to achieve robustness against outliers. A continuous-index HMM was developed by Stjernqvist et al. (2007). Guha et al. (2008) derived a Bayesian approach to the HMM with objective decision rules. A segmental maximum posteriori approach(SMAP) by Andersson et al. (2008) has incorporated both genomic distance and overlap between clones into the HMM. Finally, the third type is built upon penalization methods, which essentially relax the piecewise constant assumption by imposing a roughness penalty on CGH profile U(t). In a penalization method, we consider minimizing an objective function of the form $Q = Q_{gf} + Q_{sp}$, where the first term Q_{gf} measures the goodness of fit for profile $U(t)$ to the observed process $Y(t)$ at observed probes t_j's, and the second term Q_{sp} regularizes the smoothness of $U(t)$. Various forms of Q_{gf} and Q_{sp} have been proposed in the literature, including quantile smoothing (Eilers and De Menezes, 2005), LASSO (Huang et al., 2005), fused quantile regression(Li and Zhu, 2007) and fused LASSO (FLASSO) by Tibshirani and Wang (2008).

Besides the three types of methods, there are other approaches; for example, clustering algorithm (Wang et al., 2005; Liu et al., 2006), wavelet transform (Hsu et al., 2005) and ridge regression (van de Wiel et al., 2009), among many others. Comprehensive comparisons among some of aforementioned methods were given by Lai et al. (2005) and Willenbrock and Fridlyand (2005). Some of the methods only estimate

the profiles but do not directly call the breakpoints. Further, most methods do not control the false positive rate for breakpoint identification, and their performances are significantly effected by the experimental errors, such as outliers.

In this paper, we propose a new method based on robust state space models for array CGH data to estimate the CGH profile and to identify breakpoints under controlled false positive rates. In addition, this new method has a number of desirable properties: (1) it is robust against outliers; (2) it incorporates physical distance between probes into CNV identification; (3) it enables us to quantify estimation uncertainties of signals via posterior credible intervals; (4) all the parameters are estimated as part of the MCMC algorithm and thus are highly data-adaptive; (5) the computational efficiency of the MCMC algorithm for profile estimation is proportional to the number of probes, which helps the computation speed for high-throughput array CGH data analysis.

The rest of the paper is organized as follows. In Section 2, we first present the robust state space model, then describe an MCMC algorithm to draw samples of both profiles and parameters, and also outline a novel procedure of calling the breakpoints and outliers using MCMC samples. In Section 3, the proposed model and method are applied to both simulated and real datasets for illustration, where we compare our new method to three popular existing methods. We finally give conclusions and discussion in Section 4.

4.2 Methods

4.2.1 Model

For the ease of exposition, we denote $Y_j = Y(t_j)$, $U_j = U(t_j)$ and $\varepsilon_j = \varepsilon(t_j)$. The proposed robust state space model(RSSM) comprises two equations: an observation equation and a state equation. The observation equation is given in equation (4.1),

where measurement error ε_j is assumed to be i.i.d and follow t-distribution, $\mathcal{T}_{\upsilon_\varepsilon}$, with degree of freedom(d.f.) υ_ε. Note that t-distribution is a scale-mixture of normal distribution and gamma distribution. Thus, we rewrite $\varepsilon_j \sim \mathcal{N}(0, \sigma^2_{\varepsilon,j})$ a normal distribution with mean 0 and variance $\sigma^2_{\varepsilon,j}$, and let $\sigma^{-2}_{\varepsilon,j} = \lambda_{\varepsilon,j}\tau_\varepsilon$ and $\lambda_{\varepsilon,j} \sim \mathcal{G}(\upsilon_\varepsilon/2, \upsilon_\varepsilon/2)$ a gamma distribution with shape parameter $\upsilon_\varepsilon/2$ and rate parameter $\upsilon_\varepsilon/2$. The non-informative priors are specified as $\upsilon_\varepsilon \sim \mathcal{G}(10^{-3}, 10^{-3})I(2, 10)$ and $\tau_\varepsilon \sim \mathcal{G}(10^{-3}, 10^{-3})$ throughout the paper.

We regard the signal $U(t_j)$ as a continuous quantity which measures the $\log_2$ of average copy number of heterogeneous tumor cells versus homogeneous normal cells. The state equation is:

$$U_{j+1} = U_j + \xi_j, \tag{4.2}$$

where the evolution error or signal difference ξ_j follows an i.i.d t-distribution with d.f υ_ξ. Similar to the specification of ε_j, we let $\xi_j \sim \mathcal{N}(0, \sigma^2_{\xi,j}\delta_j)$, $\delta_j = t_{j+1} - t_j$, $\sigma^{-2}_{\xi,j} = \lambda_{\xi,j}\tau_\xi$ and $\lambda_{\xi,j} \sim \mathcal{G}(\upsilon_\xi/2, \upsilon_\xi/2)$, with the priors $\upsilon_\xi \sim \mathcal{G}(10^{-3}, 10^{-3})I(0.01, 2)$ and $\tau_\xi \sim \mathcal{G}(10^{-3}, 10^{-3})$. As a result, $\varepsilon_j \sim \mathcal{T}_{\upsilon_\varepsilon}(0, \tau_\varepsilon^{-1})$ and $\xi_j \sim \mathcal{T}_{\upsilon_\xi}(0, \delta_j\tau_\xi^{-1})$ marginally. Unlike other robust state space models (West, 1984; Fahrmeir and Künstler, 1999), our model incorporates the physical distance δ_j between two probes to address the feature that the farther two probes are apart, the larger the signal difference ξ_j is likely to be. Note that degree of freedom υ_ξ is limited below 2. In this way, we hope that the distribution $\mathcal{T}_{\upsilon_\xi}$ can accommodate extremely large values of signal difference probably caused by breakpoints. A similar strategy was suggest by Kitagawa (1987), where differences of signals are modeled by a distribution in the Pearson system with no finite second moments. As shown in his paper, the Pearson system distribution facilitates the detection of mean structure changes.

4.2.2 Signal extraction by MCMC

With the model formulation given in Section 2.1, we now outline an MCMC algorithm to sample from the posterior distribution for signals $\boldsymbol{U} = [U_1, U_2, \ldots, U_J]^\top$, parameters $\boldsymbol{\phi_o} = [\lambda_{\varepsilon,j}, \upsilon_\varepsilon, \tau_\varepsilon]$ and $\boldsymbol{\phi_s} = [\lambda_{\xi,j}, \upsilon_\xi, \tau_\xi]$ for $j = 1, 2, \ldots, J$, given the data $\boldsymbol{Y} = [Y_1, Y_2, \ldots, Y_J]$.

- Given $\boldsymbol{Y}$, $\boldsymbol{\phi_o}$ and $\boldsymbol{\phi_s}$, update the $\boldsymbol{U}$ by the simulation smoother (Durbin and Koopman, 2002a), a multi-state Gibbs sampler which very efficiently draws samples from the posterior distribution of signals $\boldsymbol{U}$.

- Given $\boldsymbol{Y}$ and $\boldsymbol{U}$, update $\boldsymbol{\phi_o}$ according to the following steps:

  $[\lambda_{\varepsilon,j} \mid \cdot] \sim \mathcal{G}(\frac{\upsilon_\varepsilon}{2} + \frac{1}{2}, \frac{\upsilon_\varepsilon}{2} + \frac{(Y_j - U_j)^2 \tau_\varepsilon}{2})$;

  $[\upsilon_\varepsilon \mid \cdot] = \prod_{j=1}^{J} \mathcal{G}(\lambda_{\varepsilon j} \mid \frac{\upsilon_\varepsilon}{2}, \frac{\upsilon_\varepsilon}{2}) \mathcal{G}(\upsilon_\varepsilon \mid 10^{-3}, 10^{-3}) I(2, 10)$, by Adaptive Metropolis Rejection Sampling (ARMS; Gilks et al., 1995);

  $\tau_\varepsilon \sim \mathcal{G}(\frac{J}{2} + 10^{-3}, \sum_{j=1}^{J} \frac{(Y_j - U_j)^2 \lambda_{\varepsilon j}}{2} + 10^{-3})$.

- Given $\boldsymbol{U}$, update $\boldsymbol{\phi_s}$ through the following steps:

  $[\lambda_{\xi,j} \mid \cdot] \sim \mathcal{G}(\frac{\upsilon_\xi}{2} + \frac{1}{2}, \frac{\upsilon_\xi}{2} + \frac{(U_{j+1} - U_j)^2 \tau_\xi}{2\delta_j})$;

  $[\upsilon_\xi \mid \cdot] = \prod_{j=1}^{J} \mathcal{G}(\lambda_{\xi j} \mid \frac{\upsilon_\xi}{2}, \frac{\upsilon_\xi}{2}) \mathcal{G}(\upsilon_\xi \mid 10^{-3}, 10^{-3}) I(0.01, 2)$, by the ARMS;

  $\tau_\xi \sim \mathcal{G}(\frac{J}{2} + 10^{-3}, \sum_{j=1}^{J} \frac{(U_{j+1} - U_j)^2 \lambda_{\xi j}}{2\delta_j} + 10^{-3})$.

According to the definition of errors $\varepsilon_j = Y_j - U_j$ and $\xi_j = U_{j+1} - U_j$, we obtain the posterior draws of the errors $\boldsymbol{\varepsilon} = [\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_J]^\top$ and the signal differences $\boldsymbol{\xi} = [\xi_2, \xi_3, \ldots, \xi_J]^\top$. Samples of $\boldsymbol{\varepsilon}$ and $\boldsymbol{\xi}$ are essential to identify outliers and breakpoints through a novel backward selection procedure detailed in Section 2.3 below.

4.2.3 Breakpoints and outliers calling

Breakpoints are called by our backward selection procedure outlined in Algorithm 1 given in the Appendix D. The input to the algorithm is the posterior draws of

signal differences $\boldsymbol{\xi}$, in an $m \times n$ matrix, with m denoting the number of draws and n equal to the number of probes minus one, as well as an input of a threshold q_α. The specification of q_α is discussed in detail below. At line 6 in Algorithm 1, we calculate $\tilde{P}_j$, which is an estimate of the posterior probability $P[|\xi_j| > |\xi_{-j}| \mid \boldsymbol{Y}]$. This is the probability of the absolute value of signal difference at position j is larger than those at any other positions, given the data. The quantity $P[|\xi_j| > |\xi_{-j}| \mid \boldsymbol{Y}]$ represents the area under the ROC curve or AUC (Pepe, 2004, Ch.4). It is known that AUC measures the separation between the posterior distribution of $|\xi_j|$ and that of the remaining $|\xi_{-j}|$, namely all $|\xi_i|$ with $i \neq j$. Under the null hypothesis that probe j is not a breakpoint, we expect $\tilde{P}_j$ to be near 0.5. The decision of rejection of the null hypothesis will be based on the comparison of $\tilde{P}_j$ with the threshold q_α. In the first iteration of procedure, several $\tilde{P}_j'$s may be larger than q_α; we take the largest one and call it a breakpoint. This called position will be excluded from the subsequent iterations. We repeat his calling procedure for the remaining ξ_j until none of the remaining $\tilde{P}_j$ is above the threshold q_α or all $n-1$ breakpoints are selected. The output of the algorithm is a list of identified breakpoints. Likewise, we utilize this backward selection procedure to call outliers, based on the posterior draws of errors $\boldsymbol{\varepsilon}$.

In the above backward selection procedure for the calling of breakpoints, the q_α is determined such that under the null hypothesis that probe i is not a breakpoint, it will be chosen with probability α(i.e. α is false positive rate). When a normal reference array is available, we can measure $\log_2$ intensity ratio of normal versus normal tissue. Fitting the proposed state space model to the normal reference array, we obtain the posterior draws of signal differences $\boldsymbol{\xi}^o$, where we can obtain $\tilde{P}_j^{o\prime}$s according to Algorithm 1. These $\tilde{P}_j^o$, $j = 1, 2, \ldots, J^o$, can be regarded as a random sample from a distribution under the null hypothesis. Then, the q_α is obtained as the $(1-\alpha)$ quantile of all $\tilde{P}_j^{o\prime}$s. This quantile q_α implies that under the null hypothesis, the rate

of false positive is α. In some real experiments, normal reference arrays however may not be available. In this case, we can generate a pseudo normal reference array $\boldsymbol{Y}^o = [Y_1^o, Y_2^o, \ldots, Y_{J^o}^o]$ by sampling with replacement from the data $\boldsymbol{Y}$. In this case, if some Y_j's in the aberration region are sampled, they will be dispersed and scattered randomly within the set $\boldsymbol{Y}^o$. Thus, they appear most likely as outliers rather than a contiguous pattern of changes. Since the proposed state space model is robust against outliers, the q_α under the null hypothesis can be reasonably determined. Given the pseudo normal array, the steps to obtain the q_α are the same as those given in the scenario of the normal reference array being available. Note that for calling the outliers, the q_α can only been obtained when normal reference array is available. Sampling with replacement from the data $\boldsymbol{Y}$ will resulting in a similar set of outliers probes.

4.3 Applications

4.3.1 Simulation study

We first evaluate our proposed method and compare it with three other popular methods, FLASSO, CBS and SMAP, using well known artificial chromosomes simulated by Lai et al. (2005)(multi-subject functional data for download at `http://www.chip.org/~ppark/ Supplements/Bioinformatics05b.html`. Lai *et al.*'s data consist of 100 chromosomes, each with length 100. In the center of each chromosome is added an aberration of copy number gain, which has one of the four different width (5,10,20 and 40). The signal-to-noise ratio(SNR) is 1, and the noise follows a normal distribution with standard deviation 0.25.

We use the Receiver Operating Characteristic(ROC) curve to compare the performance of the four methods in each width case. To obtain ROC curves, we compare the estimated signal $\hat{U}_j$ at each location with a cutoff varying from the minimum to

the maximum of $\boldsymbol{Y}$, and regard the location i where $\hat{U}_i$ is above the cutoff as the detected aberration region. The true positive rate(TPR) is defined as the proportion of the true aberration region detected as an aberration region, while the false positive rate(FPR) is defined as the proportion of the normal region declared as an aberration region. The TPRs and FPRs are plotted as ROC curves in Figure 4.1. For the Lai *et al.*'s data, the plots at the first row in Figure 4.1 indicate that our approach performs clearly better than CBS and SMAP methods, in terms of higher TPR and lower FPR, not as well as FLASSO for the narrow regions but comparably to FLASSO for the wide aberrations(20 and 40).

The simulated data in Lai et al. (2005) is idealized, and does not contain any of the complex features that occur in real data. Outliers are commonly seen in real datasets for various reasons, including single probe amplification/deletion or experimental errors. To investigate the effect of outliers, in Lai *et al.*'s simulated dataset, we add five percent of outliers in each chromosome at randomly selected positions with magnitudes uniformly distributed over interval $(3, 6)$. The ROC curves given at the second row in Figure 4.1 clearly show the advantage of the proposed method. Comparing to the corresponding cases in the first row, the ROC curves of FLASSO, CBS and SMAP are considerably closer to the diagonal line, demonstrating a significant loss of prediction power for the detection of CNVs. In contrast, the ROC curves of the proposed approach are affected very little, indicating clearly that our method is robust to outliers.

Another feature of the real data is the possibility of more than one region of aberration with different magnitudes. To evaluate the performance of the methods, we explore cases when two aberration regions are present in the simulated chromosome simultaneously. For each Lai *et al.*'s simulated chromosome, a randomly selected normal region of width five is replaced by an aberration block with SNR 4. Based on the ROC curves plotted in the third row of Figure 4.1, the proposed approach

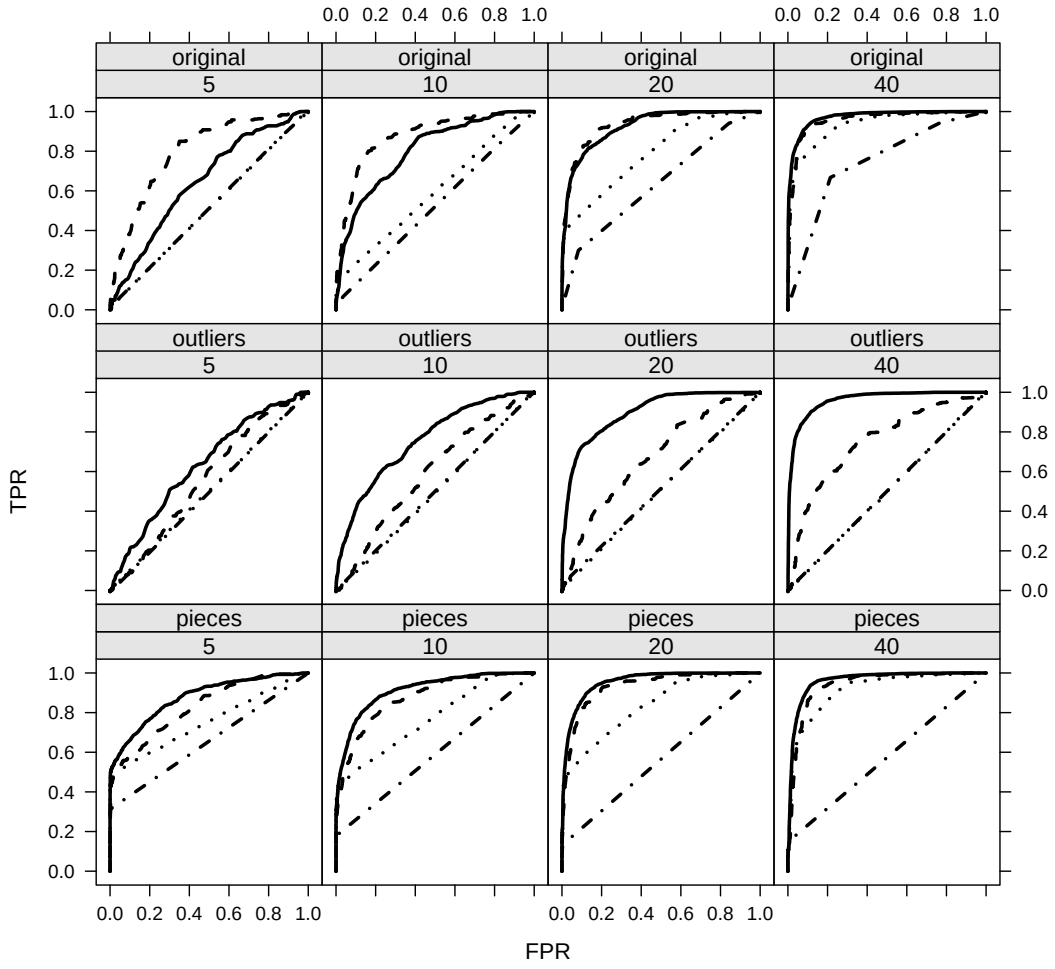


Figure 4.1: ROC curves of four methods at SNR 1. — Our model, – – – FLASSO, – · – CBS, ⋯ SMAP.

outperforms the three other approaches.

An important task in array CGH analysis is to correctly identify breakpoints. We investigate the number of breakpoints identified by the four methods for each chromosome in the above simulated data. In addition, we simulate normal chromosomes without any aberration regions. For these we generate 100 normal chromosomes, each with 100 probes simulated from $\mathcal{N}(0, 0.25^2)$. In addition, another 100 chromosomes are generated by adding outliers to the 100 normal chromosomes, in the same way described above. For FLASSO, CBS and SMAP, a breakpoint is defined as a position

j, if the difference is non-zero, that is, $\Delta\hat{U}_j = \hat{U}_{j+1} - \hat{U}_j \neq 0$. For the proposed method, a breakpoint is called by the backward selection procedure as described in Algorithm 1. To determine the q_α, we simulate a normal reference array with each probe as $\mathcal{N}(0, 0.25^2)$ with length $J^o = 1000$ and generate the pseudo normal reference arrays with length $J^o = 1000$ through sampling with replacement from the artificial chromosomes. The false positive rate α is set at 0.001, which means that for every 1000 probes in the normal reference array, one probe is expected to be falsely called as a breakpoint. Figure 4.2 shows the side-by-side boxplots of the number of breakpoints identified by each of four method respectively, where the q_α is determined with simulated normal reference arrays and pseudo normal reference arrays, respectively, for RSSM0 and RSSM1 corresponding to the first two boxplots in each panel. From a comparison of these boxplots, it is clear that the number of breakpoints is over-estimated substantially by FLASSO in all the three scenarios although the magnitude of the signal difference at some of these breakpoints may by quite small. The true number of breakpoints, on average, is more likely to be correctly achieved by the proposed method, in scenario of two pieces of aberration regions or in the cases where the aberration widths are as wide as 20 and 40. For the normal chromosomes with or without outliers, both CBS and our method correctly conclude that there are no breakpoints, while FLASSO identifies a few number of false breakpoints. Note that our method identifies a total of 6 and 13 breakpoints for 10,000 probes in 100 normal chromosomes by using, respectively, simulated and pseudo normal reference arrays. These number of false discoveries numbers are close to the expected number 10, given the false positive rate 0.001. We also notice that the numbers of breakpoints identified in the simulated and pseudo normal reference arrays are very close to each other, which validates the utility of pseudo normal reference arrays when the normal reference arrays are not available.

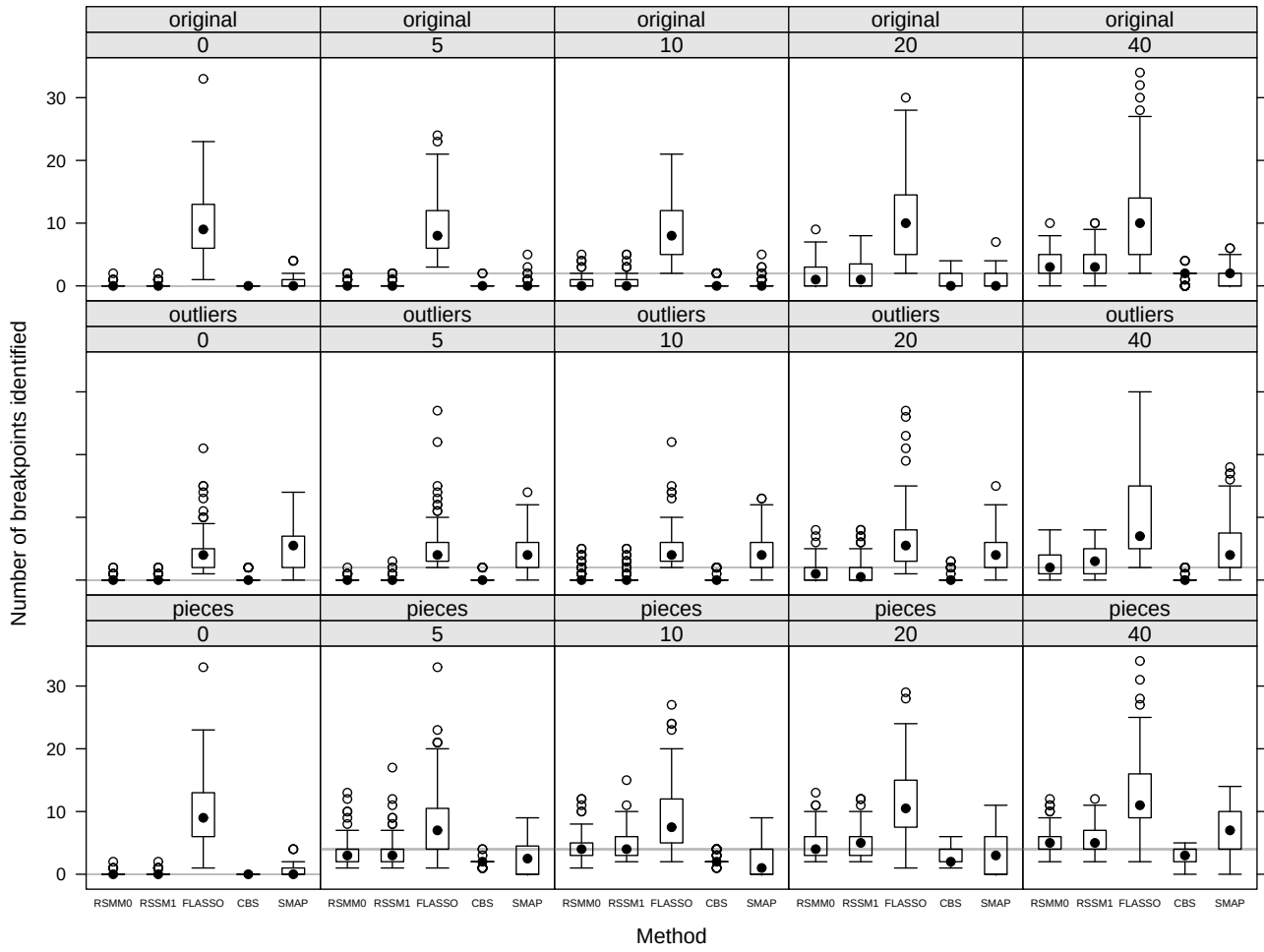


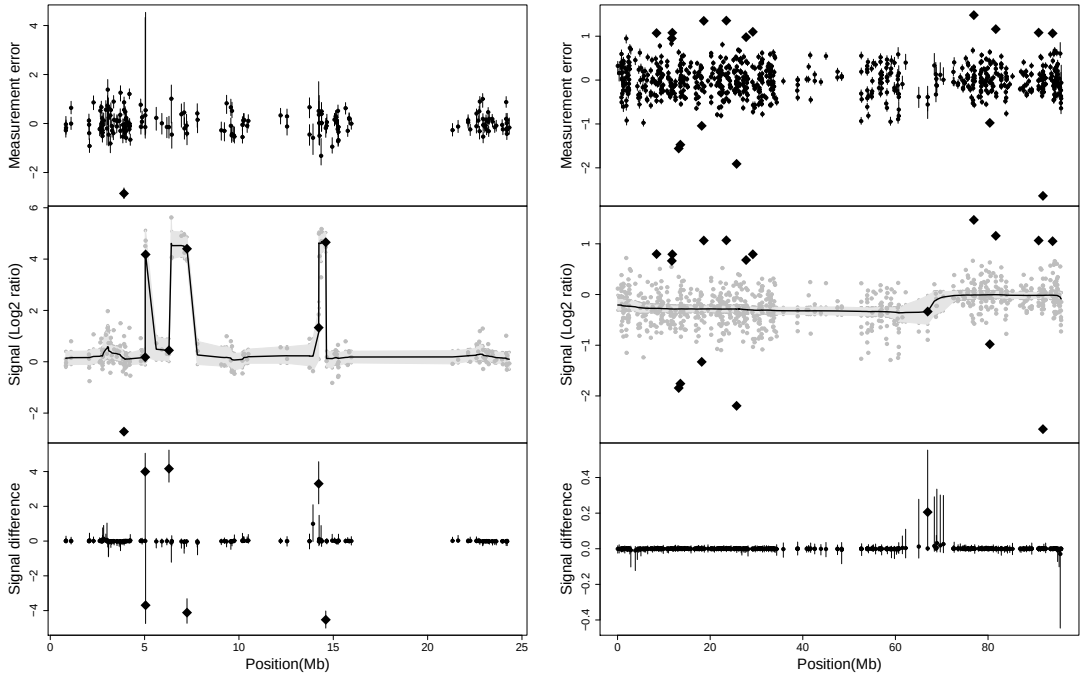
Figure 4.2: Breakpoints identification using simulated and pseudo normal reference arrays. The horizontal reference lines indicate the true number of breakpoints. The simulated normal reference array is labeled by RSSM0, while RSSM1 utilizes pseudo normal reference arrays based on resampling the observed data. The panel on bottom left is the replicate of the one on top left.

4.3.2 Glioblastoma Multiforme(GBM) data

GBM data by Bredel et al. (2005) include 26 samples representing primary GBMs, the most malignant type of brain tumor. In sample GBM31, a large region of loss is demonstrated on chromosome 13, which is also observed by Koschny et al. (2002) in a meta-analysis of 509 cases. Besides losses, the GBM data also contain a number of amplifications, one of which is shown on chromosome 7 in sample GBM29. Lai et al. (2005) compared the performance of various methods based on these two chromosomes 13 and 7 with challenging features. They represent wider, low level region of loss, and narrower, high level region of amplification, respectively. To assess our proposed method, we re-analyze these two chromosomes using our method. The analysis is based on 1000 MCMC draws from a single chain of $75,000$ iterations with $25,000$ burn-in period and every 50th being recorded. As shown in Figure 4.3 , our method successfully detects both the loss region and amplification region as well as some outliers. Both breakpoints and outliers are called using the proposed backward selection procedure. The threshold for breakpoints is obtained through the pseudo normal reference arrays with $q^{\xi}_{0.001} = 0.911$ for chomosome 7 and $q^{\xi}_{0.001} = 0.882$ for chomosome 13. The threshold for outliers is chosen as $q^{\varepsilon} = 0.98$. The panels in Figure 4.3 also illustrate posterior means and 95% credible intervals for signal U_j, error ε_j and signal differences ξ_j across the chromosomes. At a given position, the wider interval indicates higher uncertainty. Note that 95% credible intervals of signal difference illustrate the corresponding posterior distributions. The further the credible interval departs from the others along with the narrower width, the stronger it indicates the corresponding position is a breakpoint. We also analyze the GBM data using the methods of FLASSO, CBS and SMAP. As Figure 4.4 shown, all three methods can identify the two aberration regions, except SMAP method that fails to detect any aberration region for chromosome 13.

Table 4.1 lists the number of breakpoints identified by each of the four methods.

Chromosome 7 in GBM 29



(a) Chromosome 7 in GBM 29 (b) Chromosome 13 in GBM31

Figure 4.3: GBM panel plots for the posterior distributions of measurement error, signal, and signal difference by state space model. In the top and bottom panels, the • denotes the posterior mean and | stands for the 95% credible intervals. In the middle panel, gray • is the data point and —is posterior mean and 95% credible intervals are the shaded areas. ◆ denotes the selected outliers and breakpoints.

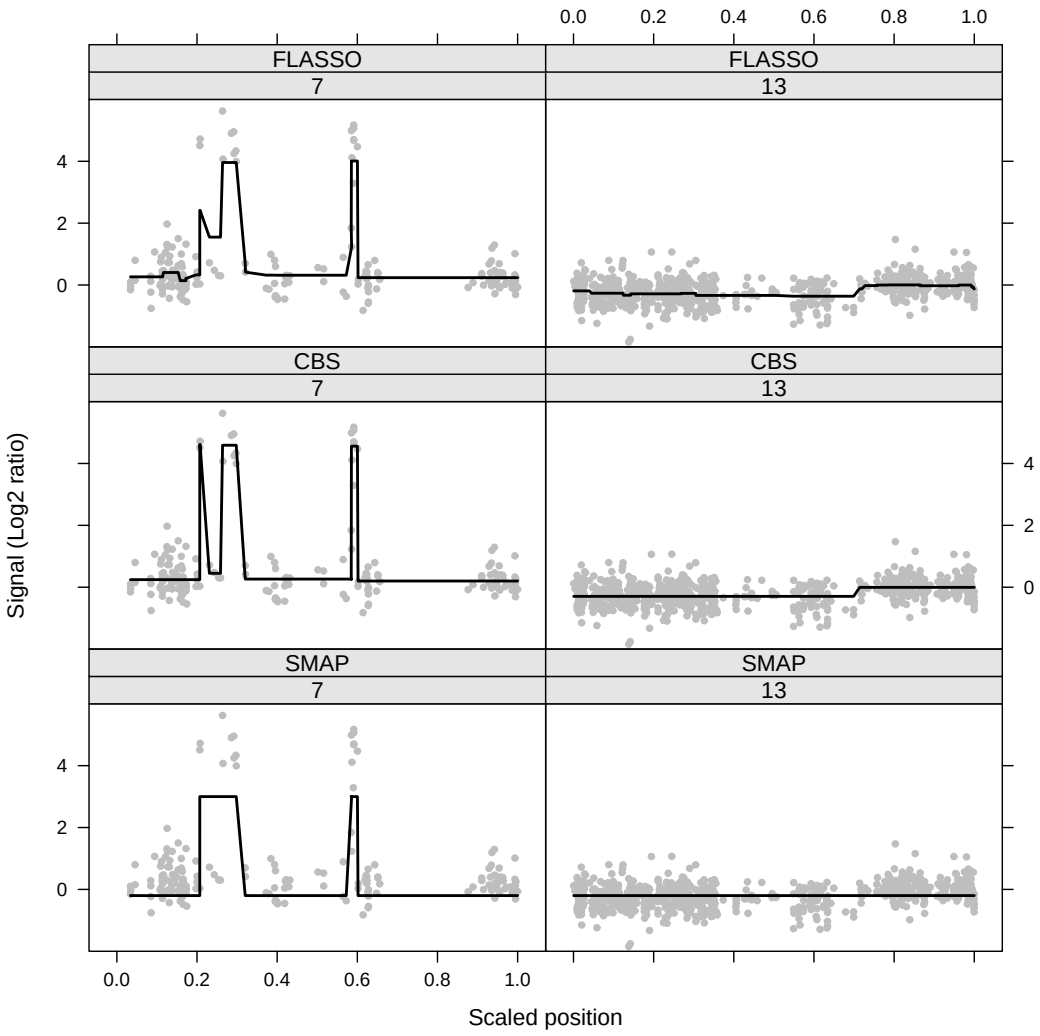


Figure 4.4: Panel plots of signal(—) estimated for GBM data by FLASSO, CBS and SMAP, where gray • denotes the data point.

Our method and CBS reach the same numbers on both chromosomes, which are much less than the those found by FLASSO.

4.3.3 Breast tumor data

Fridlyand et al. (2006) considered array CGH data from across 2464 genomic clones in 62 sporadic ductal invasive breast tumors and 5 BRCA1 mutant tumors. We apply our method as well as other three methods to analyze four chromosomes(8,11,17 and 20) of tumor "S1539", in which there are a number of low level gains and losses as well as high level amplifications. The results of our method are based on 1000 MCMC draws from a single chain of $75,000$ iterations with $25,000$ burn-in period and every 50th being recorded. The backward selection procedure has been applied to identify a number of breakpoints and outliers/amplifications. The q^{ε} is specified as 0.98, and $q^{\xi}_{0.001}$ is determined using the pseudo normal reference arrays, resulting in values of $0.795, 0.789, 0.808$ and 0.807 for chromosome 8,11,17 and 20 respectively. Figure 4.5 displays the posterior means and 95% credible intervals of signal U_j, error ε_j and signal differences ξ_j across the chromosomes, as well as a number of called outliers and breakpoints. These breakpoints define the edges of aberration regions which include several well-known oncogenes, that play key roles in the pathogenesis of breast tumor. The detected regions cover gene FGFR I between 36.4Mb and 39.7Mb on chromosome 8, gene CCND I between 68.5Mb and 77.0Mb on chromosome 11, and gene ZNF217 between 44.4Mb and 62.7Mb on chromosome 20. Gene ERBB2 between 34.1Mb and 38.7Mb on chromosome 17 is a well known gene that can be amplified in

Table 4.1: The number of breakpoints identified in GBM and Breast Tumor data

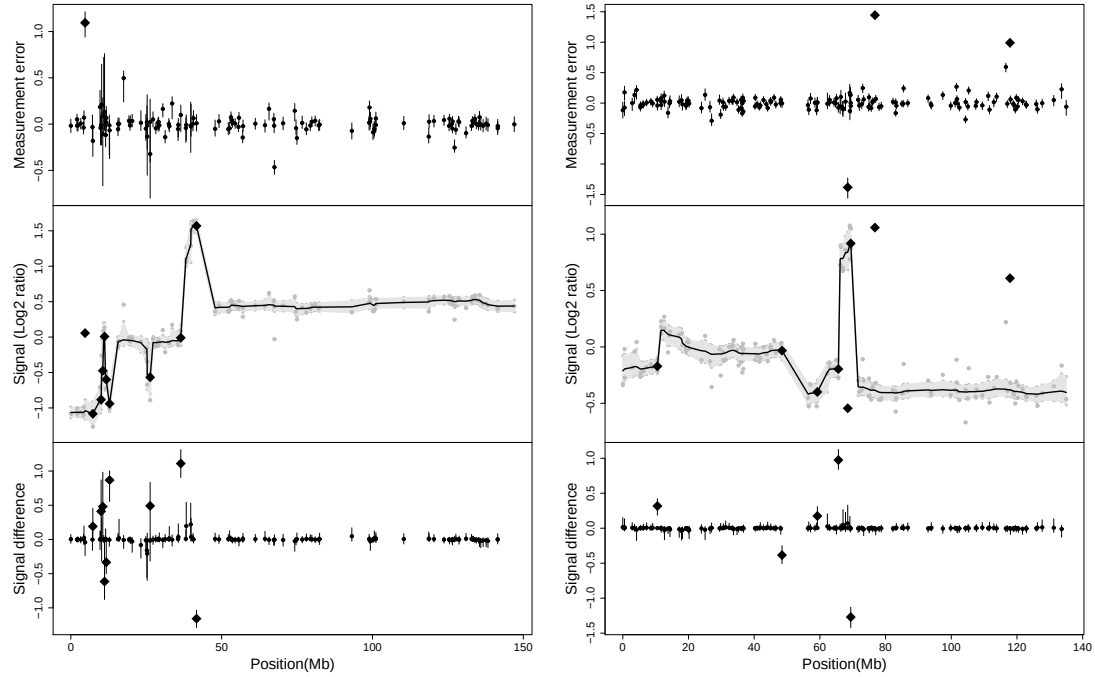
	GBM data		Breast Tumor data			
	CH7	CH13	CH8	CH11	CH17	CH20
RSSM(ours)	6	1	9	5	3	4
FLASSO	15	15	30	19	29	12
CBS	6	1	3	6	0	2
SMAP	4	0	6	11	8	9

breast cancer. There are very few probes close to ERBB2, and the method detected a probe as an outlier in this region.

We also analyze the same breast tumor data by FLASSO, CBS and SMAP methods. The results are shown in Figure 4.6. We can see that the SMAP method appears to be very sensitive to outliers(e.g. in chromosome 11) and local features(e.g. in chromosome 20), which has obscured the estimate of the global trend. The CBS method failed to capture the single probe amplification in the chromosome 17 and the weak gain in chromosome 20. The FLASSO method is also sensitive to outliers, e.g. at the beginning of chromosome 8 and in the middle of chromosome 11. The number of breakpoints identified by each of the methods is summarized in Table 4.1. FLASSO identifies a large number of breakpoints, our method identifies slightly more breakpoints than CBS and slightly fewer than SMAP.

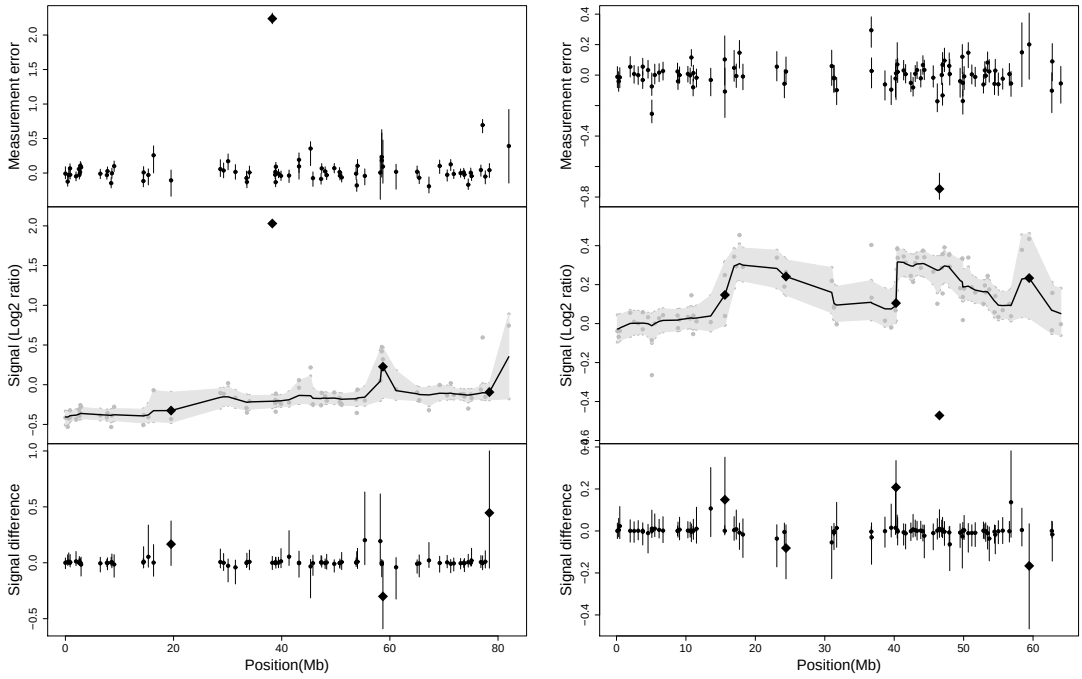
4.4 Discussion

In this paper, we have proposed a powerful new method based on a robust state space model to detect CNVs from array CGH data. A key feature of the proposed method is the use of heavy tail t-distributions, which facilitates the robustness in the calling of breakpoints and outliers. Through an MCMC algorithm, our approach presents an appealing method for CGH profile estimation and detection of breakpoints. Our method is based on a probability model that gives not only point estimation, but also uncertainty intervals for the signal, signal difference and measurement error magnitudes, as illustrated in Figure 4.3 and 4.5. Such displays are very useful for visualizing the data and the degree of confidence in any conclusion. We developed a novel backward selection procedure to effectively utilize the MCMC samples in the identification of breakpoints and outliers/amplifications. Importantly, we control the false positive rate of feature detection at a prespecified level by using real or pseudo normal reference arrays. As illustrated by both simulated and real datasets,



(a) Chromosome 8

(b) Chromosome 11



(c) Chromosome 17

(d) Chromosome 20

Figure 4.5: Breast tumor panel plots for the posterior distributions of measurement error, signal, and signal difference by state space model. In the top and bottom panels, the • denotes the posterior mean and | stands for the 95% credible intervals. In the middle panel, gray • is the data point and —is posterior mean and 95% credible intervals are the shaded areas. ◆ denotes the selected outliers and breakpoints.

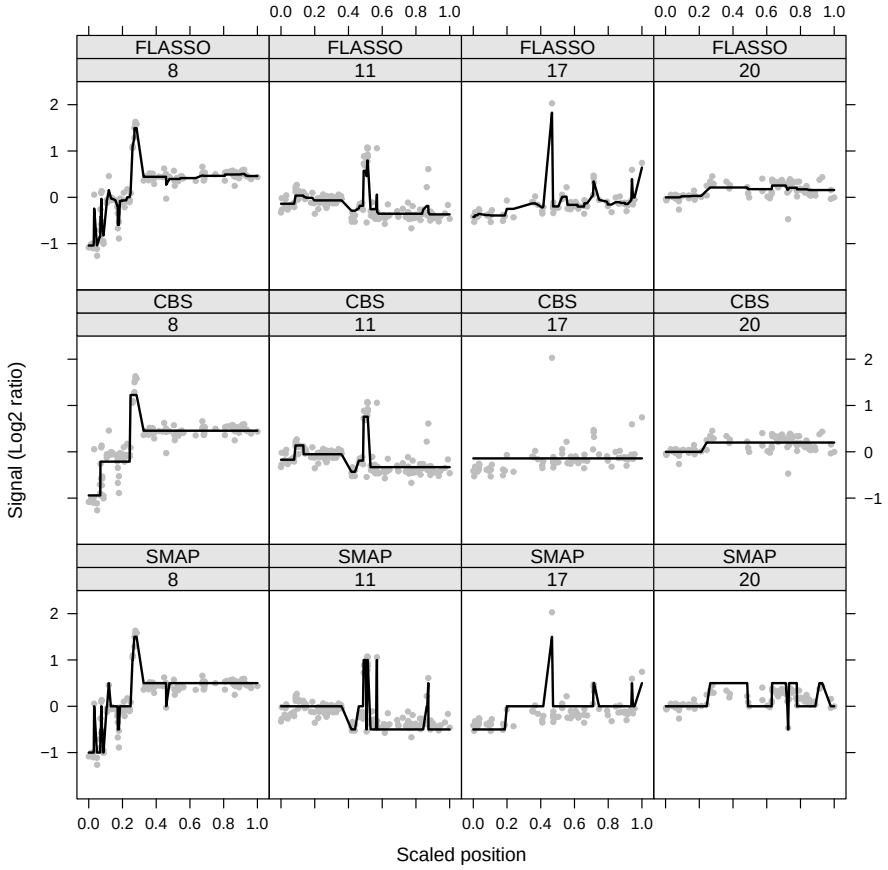


Figure 4.6: Panel plots of signal(—) estimated for breast tumor data by FLASSO, CBS and SMAP, where gray • denotes the data point.

our approach has demonstrated superior detection power for aberration regions and breakpoints, and outperforms other existing methods in most of cases, especially for noisy data with outliers.

CHAPTER V

Conclusions and Future Work

5.1 Conclusions

In this thesis, we propose the stochastic dynamic model(SDM) for functional data analysis. Distinct from the classic smoothing and kernel methods, we treat the unknown function as a realization of a stochastic process, whose distribution is determined by a stochastic different equation(SDE). In this way, the proposed model is very flexible to address various research questions with different applications. Chapter II introduce a special case of SDM, the stochastic velocity and acceleration models to analyze the prostate specific antigen(PSA) profile for time series functional data. Chapter III extend the stochastic velocity model for the multi-subject functional data, where we not only analyze multiple profiles simultaneously, but also consider the effect of covariates on the shape of profiles. Chapter IV considers the time-varying stochastic position model, where the diffusion term in the SDE varies over time. Hence, the model can approximate the breakpoints in the function and is applied to array comparative genomic hybridization(CGH) data.

To estimate the parameters and functions, we apply Euler approximation to discretize the SDE and use data augmentation to reduce the approximation errors. Markov Chain Monte Carlo(MCMC) algorithms are developed for each model, where the simulation smoother can be used to sample the values of the functions simulta-

neously in blocks instead of one value at a time, and thus the convergence of the algorithm is fast. From the MCMC output, the forecasting of future observations can also be easily obtained.

5.2 Future Work

For future work, we will propose another new model, namely a stochastic functional mixed model(SFMM) for multi-subject functional data. We also plan to generalize the Wiener process $W(t)$ to the Lévy Process $Z(t)$.

5.2.1 Model

First, to analyze the data for one subject in a longitudinal study, i.e. time series functional data, we modify the SVM-W proposed in the chapter II by generalizing a zero drift function with a Weiner process. The resulting model, tentatively named as adaptive SVM-W, is more flexible to study the drift function nonparametrically and it is easily extended to the longitudinal study. The model is defined by a hierarchical structure:

$$
\begin{aligned}
Y(t) &= \mu(t) + \varepsilon(t), \quad t \in \mathcal{T}_o = \{t : t_1, t_2, \ldots, t_J\} \\
d\mu(t) &= \nu(t)dt, \quad t \in \mathcal{T}_s = \{t : t_0 \leq t \leq t_J\} \\
d\nu(t) &= \alpha(t)dt + \sigma_\xi dW^\xi(t), \\
d\alpha(t) &= \sigma_\zeta dW^\zeta(t),
\end{aligned}
$$

where $W^\xi(t)$ and $W^\zeta(t)$ are two independent standard Wiener processes and $\varepsilon(t) \sim \mathcal{N}(0, \sigma_\varepsilon^2)$. In this model, the drift term will be data driven, which leads to potentially better model fitting and interpretation.

After some stochastic integrations, we get,

$$\mu(t) = \mu(t_0) + \nu(t_0)t + \frac{1}{2}\alpha(t_0)t^2 + \sigma_\zeta \int_{s=t_0}^{t} \int_{x=t_0}^{s} W^\zeta(x)dxds + \sigma_\xi \int_{s=t_0}^{t} W^\xi(s)ds,$$

where $\mu(t_0)$, $\nu(t_0)$ and $\alpha(t_0)$ are the parameters at initial time t_0, which are assumed $[\mu(t_0), \nu(t_0), \alpha(t_0)]^\top \sim \mathcal{N}_3(0, \sigma_0^2 \boldsymbol{I}_3)$. In addition, we can show that the estimation from this model is equivalent to minimizing a penalized sum-of-squares with nested penalty functions,

$$\sum_{j=1}^{J}[y(t_j) - \mu(t_j)]^2 + \lambda_\xi \int_{t_0}^{t_J} [D^2\mu(s) - \alpha(s)]^2 ds + \lambda_\zeta \int_{t_0}^{t_J} [D\alpha(s)]^2 ds,$$

where $\lambda_\xi = \frac{\sigma_\varepsilon^2}{\sigma_\xi^2}$ and $\lambda_\zeta = \frac{\sigma_\varepsilon^2}{\sigma_\zeta^2}$ are two smoothing parameters, which control the smoothness of $\mu(t)$ and $\alpha(t)$, respectively.

Second, to analyze multi-subject functional data, we extend the above hierarchical model as follows. As an example of including covariates, we assume subjects are divided into k groups. For subject i in group k, we impose a stochastic functional mixed model(SFMM):

$$Y_{i,k}(t) = \mu_{i,k}(t) + \varepsilon_{i,k}(t), \quad t \in \mathcal{T}_o = \{t : t_1, t_2, \ldots, t_J\} \tag{5.1}$$

$$d\mu_{i,k}(t) = \nu_{i,k}(t)dt, \quad t \in \mathcal{T}_s = \{t : t_0 \le t \le t_J\} \tag{5.2}$$

$$d\nu_{i,k}(t) = \alpha_k(t)dt + \sigma_\xi dW_{i,k}^\xi(t), \tag{5.3}$$

$$d\alpha_k(t) = \sigma_\zeta dW_k^\zeta(t), \tag{5.4}$$

Let $f_{i,k}(t) = \mu_{i,k}(t_0) + \nu_{i,k}(t_0)t + \sigma_\xi \int_{s=t_0}^{t} W_{i,k}^\xi(s)ds$, and let $g_k(t) = \mu_k(t_0) + \nu_k(t_0)t + \frac{1}{2}\alpha_k(t_0)t^2 + \sigma_k^\zeta \int_{s=t_0}^{t}\int_{x=t_0}^{s} W_k^\zeta(x)dxds$. Assume $[\mu_{i,k}(t_0), \nu_{i,k}(t_0), \alpha_k(t_0)]^\top \sim \mathcal{N}_3(0, \sigma_0^2 \boldsymbol{I}_3)$.

Then, $\mu_{i,k}(t)$ can be rewritten as

$$\mu_{i,k}(t) = g_k(t) + f_{i,k}(t),$$

where $g_k(t)$ represents the kth group population-average curve and $f_{i,k}(t)$ as the subject-specific curve deviation from the population-average profile. Note that SFMM can be regarded as an extension of the random intercept and random slope mixed effects model, with now additional stochastic terms derived from the two Wiener processes. We will investigate whether the SFMM includes other existing models as special cases.

For estimation and inference, we will first develop Bayesian methods using MCMC. Then, we will establish subject-specific forecasting. We will apply SFMM to analyze simulated and real world data. Results will be compared to those obtained by existing models. We will consider adding covariates into equations (5.1) to (5.4), besides the group stratification.

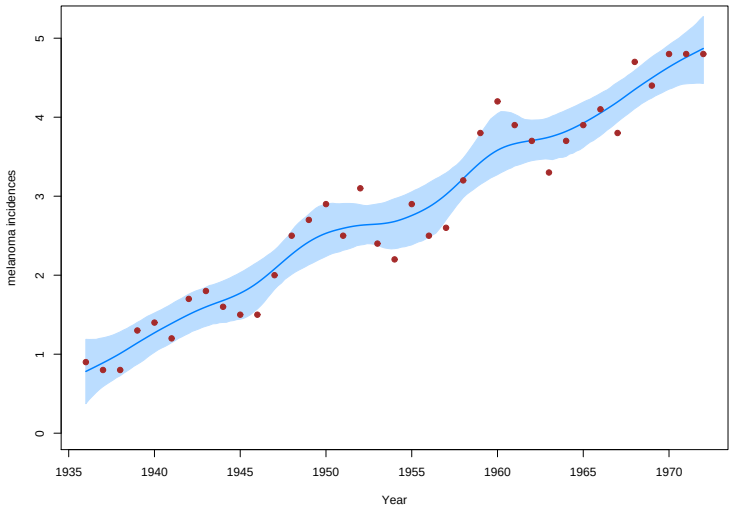
5.3 Some preliminary results

We present two examples to demonstrate the potential of the two proposed models. The adaptive SVM-W is fit to two datasets using MCMC.

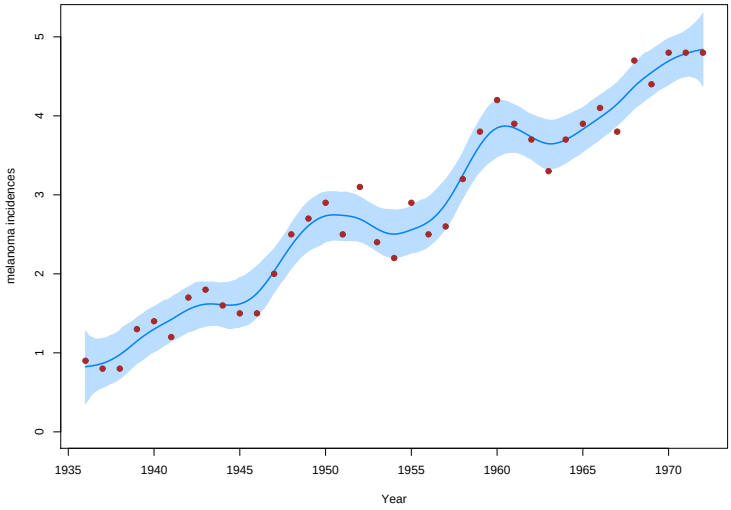
The first example is the analysis of melanoma incidence data reported by Houghton et al. (1980). 37 incidences of malignant melanomas were recorded yearly for males in Connecticut between 1936 and 1972, shown in Figure 5.1. The incidences display an increasing trend plus some periodic pattern, which may be associated with sun radiation fluctuation due to sunspot activity. We fit both the SVM-W described in the chapter II and the adaptive SVM-W in this chapter. The results are shown in Figure 5.1. It can be seen that the adaptive SVM-W fit the data better, which capture both the linear trend and periodic pattern. This conclusion is supported by a

smaller deviance information criteria(DIC). Hence, adaptive SVM-W is more flexible to capture the various pattern of function without subject matter prior knowledge.

The second example is analysis of data concerned with protein contents in milk samples taken from Diggle (1990). Protein contents of milk samples were repeatedly measured for 79 cows weekly up to week 19. The cows were assigned to three treatment groups: diet 1(barley), diet 2(mixed, barley and lupins) and diet 3(lupins). We carry out a two-stage analysis for SFMM. We first fit the adaptive SVM-W for each cow and estimate subject-specific curves. The MCMC samples from the same group are then averaged to estimate the group-average curves. The estimated profiles for each cow in diet 1 group are given in Figure 5.2. It can be seen that the subject-specific curves vary from each other significantly. The group-average profiles are shown in Figure 5.3. Three groups have separable group-average curves. These two figures display subject-specific curve deviations and group-average curves that could be captured by the proposed SFMM.



(a) By SVM-W, DIC=25.95



(b) By adaptive SVM-W, DIC=15.79

Figure 5.1: Melanoma incidences from year 1936 to 1972. Plots of data points(•), posterior means(—) and 95% credible intervals(shades) for the SVM-W and adaptive SVM-W, respectively.

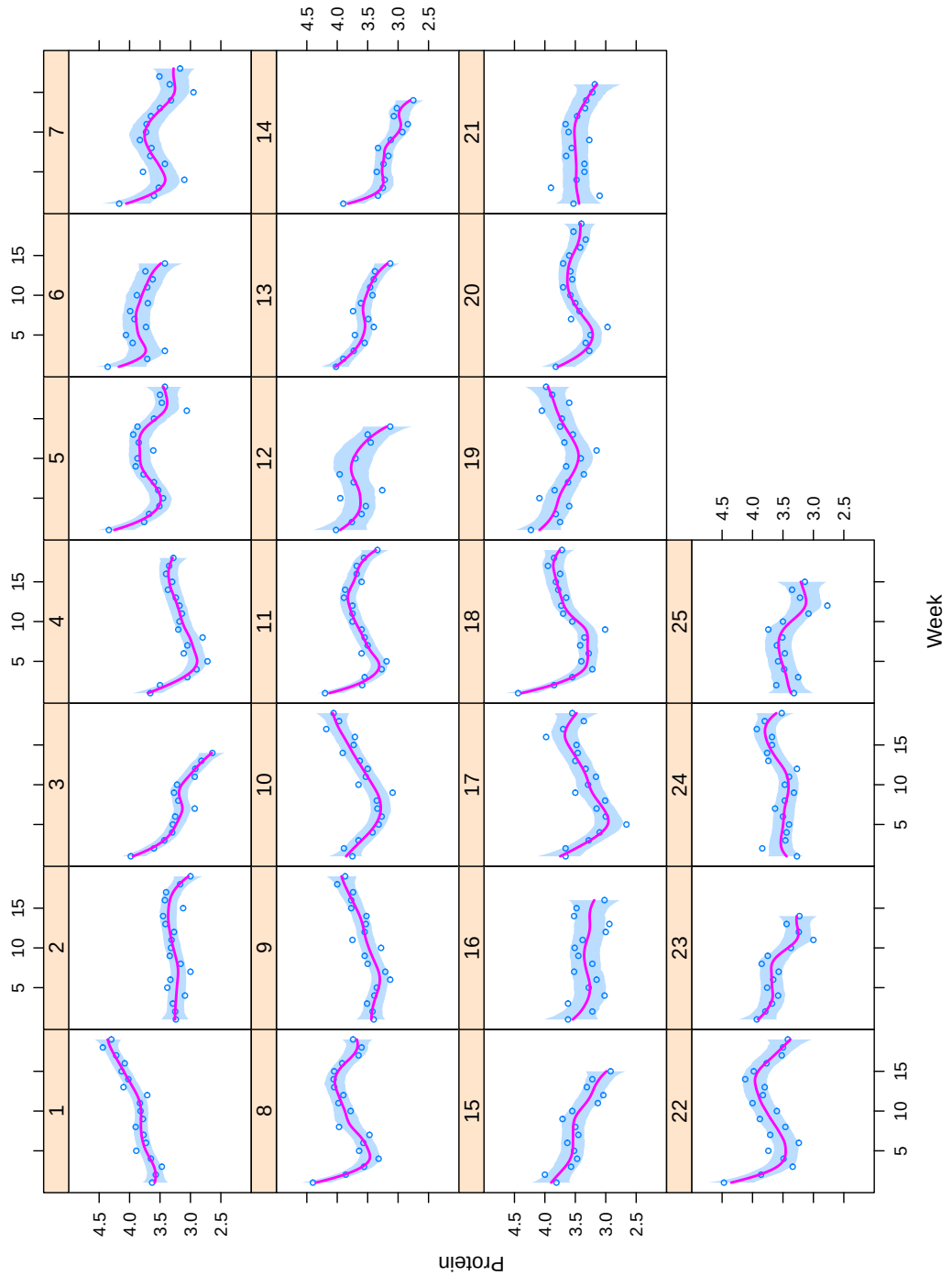


Figure 5.2: Protein contents in milk samples: Plots of posterior means(—) and 95% credible intervals(shades) for each cows in diet group 1.

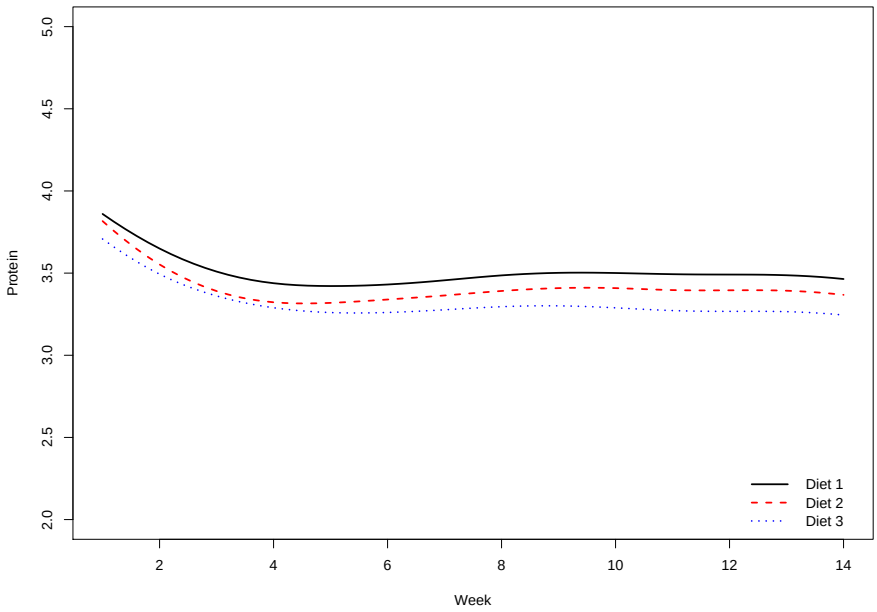


Figure 5.3: Protein contents in milk samples: Plots of posterior means for each diet group.

APPENDICES

APPENDIX A

Efficient MCMC scheme for SVM-OU

Here we outline an efficient MCMC scheme for the SVM-OU. The efficiency takes root in the Markov property of the latent process and is achieved by the simulation smoother.

When $V(t)$ follows an OU process, the Euler approximation gives the following discretized forms:

$$\begin{aligned} U_i &= U_{i-1} + V_{i-1}\delta_i, \\ V_i &= V_{i-1} - \rho V_{i-1}\delta_i + \rho\bar{\nu}\delta_i + \xi_i \\ &= (1-\rho\delta_i)V_{i-1} + \rho\delta_i\bar{\nu} + \xi_i, \quad t_i \in \mathcal{T}_{ao}, \end{aligned}$$

where $\mathcal{T}_{ao} := \{t_i : i = 1, 2, \ldots, J + \sum_{j=0}^{J-1} M_j\}$ and $\xi_i \sim \mathcal{N}(0, \sigma_\xi^2\delta_i)$.

With the observation equation (2.1), we rewrite the above discretized forms as a

standard discrete-discrete state space model:

$$Y_i = U_i + \varepsilon_i, = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}^{\top} \begin{bmatrix} U_i \\ V_i \\ \bar{\nu} \end{bmatrix} = \boldsymbol{F}^T \boldsymbol{\theta}_i + \varepsilon_i \tag{A.1}$$

$$\boldsymbol{\theta}_i = \begin{bmatrix} U_i \\ V_i \\ \bar{\nu} \end{bmatrix} = \begin{bmatrix} 1 & \delta_i & 0 \\ 0 & 1-\rho\delta_i & \rho\delta_i \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} U_{i-1} \\ V_{i-1} \\ \bar{\nu} \end{bmatrix} + \omega_i = \boldsymbol{G}_i \boldsymbol{\theta}_{i-1} + \omega_i, \tag{A.2}$$

where $\varepsilon_i \overset{i.i.d.}{\sim} \mathcal{N}(0, \sigma_\varepsilon^2)$ and $\omega_i \sim \mathcal{N}(0, \Sigma_{\omega_i})$ with $\Sigma_{\omega_i} = \begin{pmatrix} 0 & 0 & 0 \\ 0 & \sigma_\xi^2 \delta_i & 0 \\ 0 & 0 & 0 \end{pmatrix}$. The initial value satisfies

$$\boldsymbol{\theta}_0 \sim \mathcal{N}_3 \left[\begin{pmatrix} 0 \\ 0 \\ \bar{\nu} \end{pmatrix}, \begin{pmatrix} 10^6 & 0 & 0 \\ 0 & 10^6 & 0 \\ 0 & 0 & 0 \end{pmatrix} \right].$$

Given σ_ε^2, $\boldsymbol{\phi_s}$, $\boldsymbol{y}_o$ and $\boldsymbol{y}_a$, we apply the simulation smoother (Durbin and Koopman, 2002b) to update the latent state $\boldsymbol{\theta}_i$.

Given latent state $\boldsymbol{\theta}_i$, $\boldsymbol{y}_o$ and $\boldsymbol{y}_a$, the above state space model can be reformulated as two linear regression models in which parameters σ_ε^2 and $\boldsymbol{\phi_s}$ will be sampled by the standard Gibbs sampling methods.

$$\begin{aligned} Y_i &= U_i + \varepsilon_i, \\ \Delta V_i' &= \frac{V_i - V_{i-1}}{\sqrt{\delta_i}} \\ &= \rho\bar{\nu}\sqrt{\delta_i} - \rho V_{i-1}\sqrt{\delta_i} + \xi_i' \\ &= \beta_0 \sqrt{\delta_i} + \beta_1 V_{i-1}\sqrt{\delta_i} + \xi_i', \end{aligned}$$

where $\xi_i' \overset{i.i.d}{\sim} \mathcal{N}(0, \sigma_\xi^2)$ and prior $[\beta_0, \beta_1]^\top \sim \mathcal{N}_2(0, \sigma_\beta^2 I_2)$ with $\sigma_\beta^2 = 10^6$ and $\beta_1 \in \mathbb{R}^-$; the prior $\sigma_\varepsilon^2 \sim \mathcal{IG}(a, b)$ and $\sigma_\xi^2 \sim \mathcal{IG}(a, b)$ with $a = b = 0.001$. Finally, given both $\boldsymbol{\theta}_i$ and σ_ε^2, the element of $\boldsymbol{y}_a$ are sampled from $\phi(y_i \mid U_i, \sigma_\varepsilon^2)$.

When $V(t)$ follows a Wiener process, the above MCMC scheme can modified to the setting $\boldsymbol{\theta}_i = \begin{bmatrix} U_i \\ V_i \end{bmatrix}$, $\rho = 0$, $\bar{\nu} = 0$, and $\Sigma_{\omega_i} = \begin{pmatrix} 0 & 0 \\ 0 & \sigma_\xi^2 \delta_i \end{pmatrix}$ with little effort. The MCMC scheme of the SAM-W and SAM-OU can be formulated in the same way.

APPENDIX B

Link to linear mixed model for SVM-W

The SVM with the Wiener process $V(t)$ and approximated transition density can be written as a linear mixed model(LMM). It will be identical or similar to the linear spline model with the truncated line function basis, depending on whether or not data are equally spaced.

When $a\{V(t), \boldsymbol{\phi}_s\} = 0$ and $b\{V(t), \boldsymbol{\phi}_s\} = \sigma_\xi$, we discretize (2.2) and (2.3) for $m = 2$ by Euler approximation without data augmentation, and get,

$$\Delta U(t_j) = U(t_j) - U(t_{j-1}) = V(t_{j-1})\delta_j,$$
$$\Delta V(t_j) = V(t_j) - V(t_{j-1}) = \sigma_\xi \eta_j,$$

where $\delta_j = t_j - t_{j-1}$, $\eta_j = W(t_j) - W(t_{j-1}) \sim \mathcal{N}(0, \delta_j)$, $j = 1, 2, \ldots, J$ with $t_0 = 0$. It is easy to see that

$$U(t_j) = U(t_0) + V(t_0)t_j + \sigma_\xi \sum_{k=1}^{J-1} (t_j - t_k)_+ \eta_k,$$
$$V(t_j) = V(t_0) + \sigma_\xi \sum_{k=1}^{j} \eta_k,$$

where $f(x)_+$ is the positive part of function $f(x)$. Plugging $U(t_j)$ into equation (2.1),

we obtain

$$\begin{aligned} Y_j &= U(t_j) + \varepsilon_j \\ &= U(t_0) + V(t_0)t_j + \sigma_\xi \sum_{k=1}^{J-1} (t_j - t_k)_+ \eta_k + \varepsilon_j \\ &= \boldsymbol{x}_j^\top \boldsymbol{\theta}_0 + \boldsymbol{z}_j^\top \boldsymbol{\gamma} + \varepsilon_j, \end{aligned}$$

where $\boldsymbol{x}_j = [1, t_j]^\top$, $\boldsymbol{\theta}_0 = [U(t_0), V(t_0)]^\top$,
$\boldsymbol{z}_j = [\sqrt{\delta_1}(t_j - t_1), \sqrt{\delta_2}(t_j - t_2), \ldots, \sqrt{\delta_{j-2}}(t_{j-1} - t_{j-2}), 0, \ldots, 0]^\top$, and
$\boldsymbol{\gamma} = \sigma_\xi [\frac{\eta_1}{\sqrt{\delta_1}}, \frac{\eta_2}{\sqrt{\delta_2}}, \ldots, \frac{\eta_{J-1}}{\sqrt{\delta_{J-1}}}]^\top \sim \mathcal{N}_{J-1}(0, \sigma_\xi^2 I_{J-1})$.Thus,

$$\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{\theta}_0 + \boldsymbol{Z}\boldsymbol{\gamma} + \boldsymbol{\varepsilon},$$

where $\boldsymbol{X} = [\boldsymbol{x}_1 \mid \boldsymbol{x}_2 \mid \cdots \mid \boldsymbol{x}_J]^\top$ and $\boldsymbol{Z} = [\boldsymbol{z}_1 \mid \boldsymbol{z}_2 \mid \cdots \mid \boldsymbol{z}_J]^\top$. This is a linear mixed model with J random effects, and parameters $U(t_0)$, $V(t_0)$, σ_ξ^2 and σ_ε^2. If $\delta_j = \delta_{j'}$ for any pair of j and j', then this LMM is sometimes called a linear spline model with truncated line function basis (Ruppert et al., 2003).

APPENDIX C

Proof of theorem 1

C1: The solution for OU SDE:

Let $Z(t) = V(t) - \bar{\nu}$. We have

$$dZ(t) = dV(t) = -\rho Z(t)dt + \sigma_\xi dW(t).$$

Due the fact that for any two continuous functions $g(t)$ and $f(t)$,

$$\int_{t_1}^{t_2} g(s)df(s) = g(t_2)f(t_2) - g(t_1)f(t_1) - \int_{t_1}^{t_2} f(s)dg(s) \ ,$$

we obtain

$$g(t_2)f(t_2) = g(t_1)f(t_1) + \int_{t_1}^{t_2} g(s)df(s) + \int_{t_1}^{t_2} f(s)dg(s) \ .$$

Thus, for $t_1 = t_{j-1}$, $t_2 = t_j$, $g(s) = Z(s)$ and $f(s) = \exp(\rho s)$,

$$\begin{aligned} Z(t_j)\exp(\rho t_j) &= Z(t_{j-1})\exp(\rho t_{j-1}) + \int_{t_{j-1}}^{t_j} Z(s)d\exp(\rho s) + \int_{t_{j-1}}^{t_j} \exp(\rho s)dZ(s) \\ &= Z(t_{j-1})\exp(\rho t_{j-1}) + \int_{t_{j-1}}^{t_j} \exp(\rho s)\left\{\rho Z(s)ds + dZ(s)\right\} \\ &= Z(t_{j-1})\exp(\rho t_{j-1}) + \int_{t_{j-1}}^{t_j} \exp(\rho s)\ \sigma_\xi dW(s). \end{aligned}$$

It follows that with $\delta_j = t_j - t_{j-1}$, then we get

$$Z(t_j) = Z(t_{j-1})\exp(-\rho\delta_j) + \sigma_\xi\zeta_j\ ,$$

where $\zeta_j = \exp(-\rho t_j)\int_{t_{j-1}}^{t_j}\exp(\rho s)dW(s)$. It is easy to show that $E(\zeta_j) = 0$, and

$$\begin{aligned} E(\zeta_j^2) &= \exp(-2\rho t_j)E[\int_{t_{j-1}}^{t_j}\exp(\rho s_1)dW(s_1)\int_{t_{j-1}}^{t_j}\exp(\rho s_2)dW(s_2)] \\ &= \exp(-2\rho t_j)\int_{t_{j-1}}^{t_j}\exp(2\rho s)ds \\ &= \frac{1}{2\rho}\{1-\exp(-2\rho\delta_j)\}, \end{aligned}$$

and

$$Var(\zeta_j) = E(\zeta_j^2) = \frac{1}{2\rho}\left\{1-\exp(-2\rho\delta_j)\right\}.$$

Thus,

$$V(t_j) = \bar{\nu} + \{V(t_{j-1}) - \bar{\nu}\}\exp(-\rho\delta_j) + \sigma_\xi\zeta_j, \tag{C.1}$$

with $E\{V(t_j) \mid U(t_{j-1}), V(t_{j-1}), \bar{\nu}, \rho, \sigma_\xi\}$ and $Var\{V(t_j) \mid U(t_{j-1}), V(t_{j-1}), \bar{\nu}, \rho, \sigma_\xi)\}$ equal to those given in Theorem 1.

C2: The solution for IOU SDE

From equation (C.1), We have

$$\begin{aligned}
U(t_j) &= U(t_{j-1}) + \int_{t_{j-1}}^{t_j} V(t)dt \\
&= U(t_{j-1}) + \bar{\nu}\delta_j + \int_{t_{j-1}}^{t_j} \{V(t_{j-1}) - \bar{\nu}\} \exp\{-\rho(t - t_{j-1})\}dt + \\
&\quad \sigma_\varepsilon \int_{t_{j-1}}^{t_j} \int_{t_{j-1}}^{t} \exp\{-\rho(t-s)\}dW(s)dt \\
&= U(t_{j-1}) + \bar{\nu}\delta_j + \{V(t_{j-1}) - \bar{\nu}\} \left\{ \frac{1 - \exp(-\rho\delta_j)}{\rho} \right\} + \\
&\quad \sigma_\varepsilon \int_{t_{j-1}}^{t_j} \int_{t_{j-1}}^{t} \exp\{-\rho(t-s)\}dW(s)dt.
\end{aligned}$$

This imply that

$$U(t_j) = U(t_{j-1}) + \bar{\nu}\delta_j + \{V(t_{j-1}) - \bar{\nu}\} \left\{ \frac{1 - \exp(-\rho\delta_j)}{\rho} \right\} + \sigma_\varepsilon \xi_j,$$

where

$$\begin{aligned}
\xi_j &= \int_{t_{j-1}}^{t_j} \int_{t_{j-1}}^{t} \exp\{-\rho(t-s)\}dW(s)dt \\
&= \int_{t_{j-1}}^{t_j} \int_{s}^{t_j} \exp\{-\rho(t-s)\}dtdW(s) \\
&= \frac{1}{\rho} \int_{t_{j-1}}^{t_j} \left[1 - \exp\{-\rho(t_j - s)\}\right] dW(s).
\end{aligned}$$

Moreover, we can show that $E[\xi_j] = 0$, and

$$\begin{aligned} Var(\xi_j) = E(\xi_j^2) &= \frac{1}{\rho^2} \int_{t_{j-1}}^{t_j} [\exp\{-2\rho(t_j - s)\} - 2\exp\{-\rho(t_j - s)\} + 1]\, ds \\ &= \frac{\delta_j}{\rho^2} + \frac{1}{2\rho^3} \left\{-3 + 4\exp(-\rho\delta_j) - \exp(-2\rho\delta_j)\right\}, \end{aligned}$$

$$\begin{aligned} E(\xi_j\zeta_j) &= E\left[\frac{1}{\rho} \int_{t_{j-1}}^{t_j} [1 - \exp\{-\rho(t_j - s_1)\} dW(s_1) \int_{t_{j-1}}^{t_j} \exp\{-\rho(t_j - s_2)\} dW(s_2)\right] \\ &= \frac{1}{\rho} \int_{t_{j-1}}^{t_j} [\exp\{-\rho(t_j - s)\} - \exp\{-2\rho(t_j - s)\}]\, ds \\ &= \frac{1}{2\rho^2}\{1 - 2\exp(-\rho\delta_j) + \exp(-2\rho\delta_j)\}. \end{aligned}$$

Thus,

$$Cov(\xi_j, \zeta_j) = \frac{1}{2\rho^2}\{1 - 2\exp(-\rho\delta_j) + \exp(-2\rho\delta_j)\}.$$

Finally, it is trivial to show that $E\{U(t_j) \mid U(t_{j-1}), V(t_{j-1}), \bar{\nu}, \rho, \sigma_\xi\}$, $Var\{U(t_j) \mid U(t_{j-1}), V(t_{j-1}), \bar{\nu}, \rho, \sigma_\xi\}$ and $Cov\{U(t_j), V(t_j) \mid U(t_{j-1}), V(t_{j-1}), \bar{\nu}, \rho, \sigma_\xi\}$ are those given in Theorem 1.

APPENDIX D

Backward selection procedure for the breakpoints

Input: $M_{m\times n}$, q_α

$\mathcal{J} \leftarrow \emptyset$ and flag $\leftarrow$ true
repeat
  for $j = 1$ to n and $j \notin \mathcal{J}$ **do**
    $V_{-j} \leftarrow m$ samples without replacement from columns $\mathcal{I}$ of M, $\mathcal{I} = \{i : i \neq j \text{ and } i \notin \mathcal{J}\}$
    $V_j \leftarrow$ column j of M
    $\tilde{P}_j = \frac{1}{m^2} \sum_{k=1}^{m} \sum_{k'=1}^{m} I(|V_j[k]| > |V_{-j}[k']|)$
  end for
  if $\exists j \in \{1, 2, \ldots, n\}$ and $j \notin \mathcal{J} : \tilde{P}_j > q_\alpha$ **then**
    $j \leftarrow j : \tilde{P}_j > \tilde{P}_{j'}$ all $j' \neq j$
    $\mathcal{J} \leftarrow \mathcal{J} \cup \{j\}$
  else
    flag $\leftarrow$ false
  end if
until flag = false or number of elements in $\mathcal{J} = n - 1$
if number of elements in $\mathcal{J} = n - 1$ **then**
  $\mathcal{J} \leftarrow \{1, 2, \ldots, n\}$
end if

Output: $\mathcal{J}$

BIBLIOGRAPHY

BIBLIOGRAPHY

Aalen, O. and Gjessing, H. (2004), "Survival models based on the Ornstein-Uhlenbeck process," *Lifetime Data Analysis*, 10, 407–423.

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

Andersson, R., Bruder, C., Piotrowski, A., Menzel, U., Nord, H., Sandgren, J., Hvidsten, T., de Stahl, D., Dumanski, J., and Komorowski, J. (2008), "A segmental maximum a posteriori approach to genome-wide copy number profiling," *Bioinformatics*, 24, 751.

Ansley, C. F. and Kohn, R. (1986), "On the Equivalence of Two Stochastic Approaches to Spline Smoothing," *Journal of Applied Probability*, 23, 391–405.

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 B*, 68, 333–382.

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

Bredel, M., Bredel, C., Juric, D., Harsh, G., Vogel, H., Recht, L., and Sikic, B. (2005), "High-resolution genome-wide mapping of genetic alterations in human glial brain tumors," *Cancer Research*, 65, 4088.

Cumberland, W. and Sykes, Z. (1982), "Weak Convergence of an Autoregressive Process Used in Modeling Population Growth." *Journal of Applied Probability*, 19, 450–455.

Davies, J., Wilson, I., and Lam, W. (2005), "Array CGH technologies and their applications to cancer genomes," *Chromosome Research*, 13, 237–248.

Dawid, A. (1984), "Present position and potential developments: Some personal views: Statistical theory: The prequential approach," *Journal of the Royal Statistical Society. Series A (General)*, 147, 278–292.

de Jong, P. and Shephard, N. (1995), "The simulation smoother for time series models," *Biometrika*, 82, 339–350.

Diggle, P. (1990), *Time Series: a Biostatistical Introduction*, Oxford University Press.

Diggle, P., Heagerty, P., Liang, K., and Zeger, S. (2002), *Analysis of longitudinal data*, Oxford University Press, USA.

Diggle, P. and Hutchinson, M. (1989), "On spline smoothing with autocorrelated errors," *Australian & New Zealand Journal of Statistics*, 31, 166–182.

Durbin, J. and Koopman, S. (2001), *Time Series Analysis by State Space Methods*, Oxford University Press.

— (2002a), "A simple and efficient simulation smoother for state space time series analysis," *Biometrika*, 89, 603–616.

Durbin, J. and Koopman, S. J. (2002b), "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 & Economic Statistics.*, 20, 297–338.

Eilers, P. and De Menezes, R. (2005), "Quantile smoothing of array CGH data," *Bioinformatics*, 21, 1146–1153.

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 & Economic Statistics.*, 19, 177–191.

Erdman, C. and Emerson, J. (2008), "A fast Bayesian change point analysis for the segmentation of microarray data," *Bioinformatics*, 24, 2143.

Fahrmeir, L. and Künstler, R. (1999), "Penalized likelihood smoothing in robust state space models," *Metrika*, 49, 173–191.

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

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

Fridlyand, J., Snijders, A., Pinkel, D., Albertson, D., and Jain, A. (2004), "Hidden Markov models approach to the analysis of array CGH data," *Journal of Multivariate Analysis*, 90, 132–153.

Fridlyand, J., Snijders, A., Ylstra, B., Li, H., Olshen, A., Segraves, R., Dairkee, S., Tokuyasu, T., Ljung, B., Jain, A., et al. (2006), "Breast tumor copy number aberration phenotypes and genomic instability," *BMC Cancer*, 6, 96.

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. (1987), "Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images," *Readings in Computer Vision: Issues, Problems, Principles, and Paradigms*, 564–584.

Gilks, W., Best, N., and Tan, K. (1995), "Adaptive rejection Metropolis sampling within Gibbs sampling," *Applied Statistics*, 44, 455–472.

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

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

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.

Guha, S., Li, Y., and Neuberg, D. (2008), "Bayesian hidden Markov modeling of array CGH data," *Journal of the American Statistical Association*, 103, 485–497.

Guo, W. (2002), "Functional mixed effects models," *Biometrics*, 58, 121–128.

Hastie, T. and Tibshirani, R. (1993), "Varying-coefficient models," *Journal of the Royal Statistical Society. Series B (Methodological)*, 55, 757–796.

Hoover, D., Rice, J., Wu, C., and Yang, L. (1998), "Nonparametric smoothing estimates of time-varying coefficient models with longitudinal data," *Biometrika*, 85, 809.

Houghton, A., Flannery, J., and Viola, M. (1980), "Malignant melanoma in Connecticut and Denmark," *International Journal of Cancer*, 25, 95–104.

Hsu, L., Self, S., Grove, D., Randolph, T., Wang, K., Delrow, J., Loo, L., and Porter, P. (2005), "Denoising array-based comparative genomic hybridization data using wavelets," *Biostatistics*, 6, 211–226.

Huang, T., Wu, B., Lizardi, P., and Zhao, H. (2005), "Detection of DNA copy number alterations using penalized least squares regression," *Bioinformatics*, 21, 3811–3817.

Hull, J. and White, A. (1987), "The pricing of options on assets with stochastic volatilities," *The Journal of Finance*, 42, 281–300.

Hupe, P., Stransky, N., Thiery, J., Radvanyi, F., and Barillot, E. (2004), "Analysis of array CGH data: from signal ratio to gain and loss of DNA regions," *Bioinformatics*, 20, 3413.

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.

Jones, R. (1993), *Longitudinal Data with Serial Correlation: a State-space Approach*, Chapman & Hall/CRC.

Kariyanna, S., Light, R., and Agarwal, R. (2009), "A longitudinal study of kidney structure and function in adults," *Nephrology Dialysis Transplantation*.

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.

Kitagawa, G. (1987), "Non-Gaussian state-space modeling of nonstationary time series," *Journal of the American Statistical Association*, 82, 1032–1041.

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

Kohn, R. and Ansley, C. (1988), "Equivalence between Bayesian smoothness priors and optimal smoothing for function estimation," *Bayesian Analysis of Time Series and Dynamic Models*, 1, 393–430.

Koschny, R., Koschny, T., Froster, U., Krupp, W., and Zuber, M. (2002), "Comparative genomic hybridization in glioma a meta-analysis of 509 cases," *Cancer genetics and Cytogenetics*, 135, 147–159.

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

Lai, W., Johnson, M., Kucherlapati, R., and Park, P. (2005), "Comparative analysis of algorithms for identifying amplifications and deletions in array CGH data," *Bioinformatics*, 21, 3763–3770.

Laird, N. and Ware, J. (1982), "Random-effects models for longitudinal data," *Biometrics*, 38, 963–974.

Li, Y. and Zhu, J. (2007), "Analysis of array CGH data for cancer studies using fused quantile regression," *Bioinformatics*, 23, 2470.

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.

Lieberfarb, M., Schultz, D., Whittington, R., Malkowicz, B., Tomaszewski, J., Weinstein, M., Wein, A., Richie, J., and D'Amico, A. (2002), "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*, 53, 898–903.

Liu, J., Mohammed, J., Carter, J., Ranka, S., Kahveci, T., and Baudis, M. (2006), "Distance-based clustering of CGH data," *Bioinformatics*, 22, 1971.

Lloyd-Jones, D., Liu, K., Colangelo, L., Yan, L., Klein, L., Loria, C., Lewis, C., and Savage, P. (2007), "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*, 115, 1004.

Morris, J. and Carroll, R. (2006), "Wavelet-based functional mixed models," *Journal of the Royal Statistical Society. Series B*, 68, 179–199.

Muller, P. and Quintana, F. (2004), "Nonparametric Bayesian Data Analysis," *Statistical Science*, 19, 95–110.

Mungas, D., Harvey, D., Reed, B., Jagust, W., DeCarli, C., Beckett, L., Mack, W., Kramer, J., Weiner, M., Schuff, N., et al. (2005), "Longitudinal volumetric MRI change and rate of cognitive decline," *Neurology*, 65, 565.

Nicolato, E. and Venardos, E. (2003), "Option pricing in stochastic volatility models of the Ornstein-Uhlenbeck type," *Mathematical Finance*, 13, 445–466.

Olshen, A., Venkatraman, E., Lucito, R., and Wigler, M. (2004), "Circular binary segmentation for the analysis of array-based DNA copy number data," *Biostatistics*, 5, 557–572.

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.

Pepe, M. (2004), *The Statistical Evaluation of Medical Tests for Classification and Prediction*, Oxford University Press, USA.

Pinkel, D. and Albertson, D. (2005), "Array comparative genomic hybridization and its applications in cancer," *Nature Genetics*, 37, S11–S17.

Pinkel, D., Segraves, R., Sudar, D., Clark, S., Poole, I., Kowbel, D., Collins, C., Kuo, W., Chen, C., Zhai, Y., et al. (1998), "High resolution analysis of DNA copy number variation using comparative genomic hybridization to microarrays," *Nature Genetics*, 20, 207–211.

Proust-Lima, C., Taylor, J., Williams, S., Ankerst, D., Liu, N., Kestin, L., Bae, K., and Sandler, H. (2008a), "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*, 72, 782–791.

Proust-Lima, C., Taylor, J. M. G., Williams, S., Ankerst, D., Liu, N., Kestin, L., Bae, K., and Sandler, H. (2008b), "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.

Qin, L. and Guo, W. (2006), "Functional mixed-effects model for periodic data," *Biostatistics*, 7, 225.

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 B*, 69, 741–796.

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

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

Rice, J. and Silverman, B. (1991), "Estimating the mean and covariance structure nonparametrically when the data are curves," *Journal of the Royal Statistical Society. Series B*, 53, 233–243.

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

Ruppert, D., Wand, M., and Carroll, R. (2003), *Semiparametric Regression*, Cambridge: Cambridge University Press.

Sartor, C., Strawderman, M., Lin, X., Kish, K., McLaughlin, P., and Sandler, H. (1997), "Rate of PSA rise predicts metastatic versus local recurrence after definitive radiotherapy," *International Journal of Radiation Oncology Biology Physics*, 38, 941–947.

Shah, S., Xuan, X., DeLeeuw, R., Khojasteh, M., Lam, W., Ng, R., and Murphy, K. (2006), "Integrating copy number polymorphisms into array CGH analysis using a robust HMM," *Bioinformatics*, 22.

Solinas-Toldo, S., Lampel, S., Stilgenbauer, S., Nickolenko, J., Benner, A., Dohner, H., Cremer, T., and Lichter, P. (1997), "Matrix-based comparative genomic hybridization: biochips to screen for genomic imbalances," *Genes, Chromosomes and Cancer*, 20.

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 B*, 64, 583–616.

Stjernqvist, S., Ryden, T., Skold, M., and Staaf, J. (2007), "Continuous-index hidden Markov modelling of array CGH copy number data," *Bioinformatics*, 23, 1006.

Strasak, A., Kelleher, C., Klenk, J., Brant, L., Ruttmann, E., Rapp, K., Concin, H., Diem, G., Pfeiffer, K., and Ulmer, H. (2008), "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*, 28, 1857.

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.

Tibshirani, R. and Wang, P. (2008), "Spatial smoothing and hot spot detection for CGH data using the fused lasso," *Biostatistics*, 9, 18–29.

Trostab, D., IIc, E., Ostroff, J., Xiongc, W., and Marchc, P. (2009), "A model for liver homeostasis using modified mean-reverting Ornstein–Uhlenbeck process," *Computational and Mathematical Methods in Medicine*, 99999, 1–1.

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

van de Wiel, M., Brosens, R., Eilers, P., Kumps, C., Meijer, G., Menten, B., Sistermans, E., Speleman, F., Timmerman, M., and Ylstra, B. (2009), "Smoothing waves in array CGH tumor profiles," *Bioinformatics*, 25, 1099–1104.

Venkatraman, E. and Olshen, A. (2007), "A faster circular binary segmentation algorithm for the analysis of array CGH data," *Bioinformatics*, 23, 657.

Verbeke, G. and Molenberghs, G. (2009), *Linear Mixed Models for Longitudinal Data*, Springer Verlag.

Verbyla, A., Cullis, B., Kenward, M., and Welham, S. (1999), "The analysis of designed experiments and longitudinal data by using smoothing splines," *Journal of the Royal Statistical Society Series C*, 48, 269–311.

Vogel, C., Cobleigh, M., Tripathy, D., Gutheil, J., Harris, L., Fehrenbacher, L., Slamon, D., Murphy, M., Novotny, W., Burchmore, M., et al. (2002), "Efficacy and safety of trastuzumab as a single agent in first-line treatment of HER2-overexpressing metastatic breast cancer," *Journal of Clinical Oncology*, 20, 719.

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

— (1990), *Spline Models for Observational Data*, Society for Industrial Mathematics.

Wand, M. and Jones, M. (1995), *Kernel Smoothing*, New York: Chapman & Hall/CRC.

Wang, P., Kim, Y., Pollack, J., Narasimhan, B., and Tibshirani, R. (2005), "A method for calling gains and losses in array CGH data," *Biostatistics*, 6, 45–58.

Wang, Y. and Taylor, J. (1995), "Inference for smooth curves in longitudinal data with application to an AIDS clinical trial," *Statistics in Medicine*, 14, 1205–1205.

Wecker, W. E. and Ansley, C. F. (1983), "The Signal Extraction Approach to Nonlinear Regression and Spline Smoothing," *Journal of the American Statistical Association*, 78, 81–89.

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

Welham, S., Cullis, B., Kenward, M., and Thompson, R. (2006), "The analysis of longitudinal data using mixed model L-splines," *Biometrics*, 62, 392–401.

West, M. (1984), "Outlier models and prior distributions in Bayesian linear regression," *Journal of the Royal Statistical Society. Series B (Methodological)*, 46, 431–439.

West, M. and Harrison, J. (1997), *Bayesian Forecasting and Dynamic Models*, Springer Verlag.

Willenbrock, H. and Fridlyand, J. (2005), "A comparison study: applying segmentation to array CGH data for downstream analyses," *Bioinformatics*, 21, 4084–4091.

Yu, M., Taylor, J. M. G., and Sandler, H. M. (2008), "Individual Prediction in Prostate Cancer Studies Using a Joint Longitudinal Survival Cure Model," *Journal of the American Statistical Association*, 103, 178–187.

Zagars, G. and Pollack, A. (1993), "The fall and rise of prostate-specific antigen: kinetics of serum prostate-specific antigen levels after radiation therapy for prostate cancer," *Cancer*, 72, 832–842.

Zeger, S. and Diggle, P. (1994), "Semiparametric models for longitudinal data with application to CD4 cell numbers in HIV seroconverters," *Biometrics*, 50, 689–699.

Zhang, D., Lin, X., Raz, J., and Sowers, M. (1998), "Semiparametric Stochastic Mixed Models for Longitudinal Data." *Journal of the American Statistical Association*, 93, 710–719.

Zhang, P., Song, P., Qu, A., and Greene, T. (2008), "Efficient estimation for patient-specific rates of disease progression using nonnormal linear mixed models," *Biometrics*, 64, 29–38.