BIOMETRICS 000, 000–000 DOI: 000
000 0000

Stochastic Functional Data Analysis: A Diffusion Model-based Approach

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

[1] Department of Statistical Science, Duke University, Durham, North Carolina 27708, U.S.A.

[2] Center for Human Genetics, Duke University Medical Center, Durham, North Carolina 27708, U.S.A.

[3] Department of Biostatistics,University of Michigan, Ann Arbor, Michigan 48109, U.S.A.

**email:* bin.zhu@duke.edu

***email:* pxsong@umich.edu

****email:* jmgt@umich.edu

A. Equivalence between Smoothing Splines and Bayesian Estimation of SVM-W

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,

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

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

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

Kimeldorf and Wahba (1970) and Wahba (1978) have shown the "equivalence" by treating penalized sum-of-squares (A.1) as a 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. 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 $\varepsilon(t)$.

B. 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 (1), we rewrite the above discretized forms as a standard discrete-discrete state space model:

$$
Y_i = U_i + \varepsilon_i, = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}^\top \begin{bmatrix} U_i \\ V_i \\ \bar{\nu} \end{bmatrix} = \boldsymbol{F}^T\boldsymbol{\theta}_i + \varepsilon_i \tag{B.3}
$$

$$
\boldsymbol{\theta}_i = \begin{bmatrix} U_i \\ V_i \\ \bar{\nu} \end{bmatrix} = \begin{bmatrix} 1 & \delta_i & 0 \\ 0 & 1-\rho\delta_i & \rho\delta_i \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} U_{i-1} \\ V_{i-1} \\ \bar{\nu} \end{bmatrix} + \omega_i = \boldsymbol{G}_i\boldsymbol{\theta}_{i-1} + \omega_i, \tag{B.4}
$$

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, 2002) 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.

C. Web Tables and Figures

[Figure 1 about here.]

[Table 1 about here.]

[Table 2 about here.]

D. Link to linear mixed model for SVM-WN

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) and (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 (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).

References

Ansley, C. F. and Kohn, R. (1986). On the equivalence of two stochastic approaches to spline smoothing. *Journal of Applied Probability* **23,** 391–405.

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

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.

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.

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

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.

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.

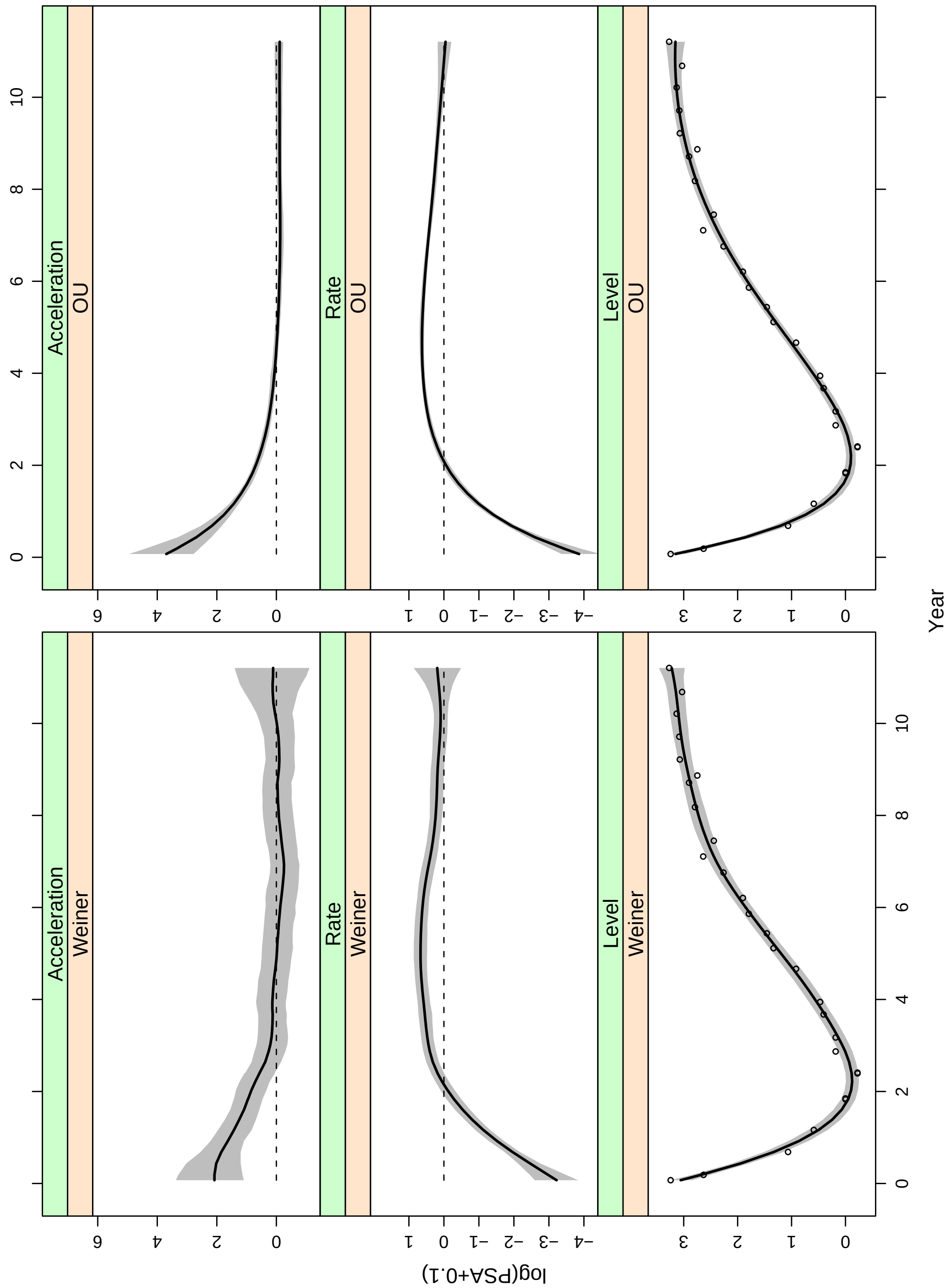

Figure 1. PSA: Plots of data points(∘), posterior means(—) and 95% credible intervals(gray shades) for the SAM with the Wiener process and OU process, respectively. In the graph, the upper panels show the acceleration $V(t)$, the middle panels show the rate $\dot{U}(t)$, and the lower panels show the level, $U(t)$.

Table 1
Simulation results for the estimation of $U(t)$ and $V(t)$ for various observational time interval and measurement errors.

Case	States	Bias	MSE
1. Uniform sparse data	$U(t)$	0.019	0.048
	$V(t)$	0.047	0.375
2. Sparase early data	$U(t)$	0.032	0.163
	$V(t)$	0.106	1.780
3. Sparse late data	$U(t)$	0.016	0.037
	$V(t)$	0.032	0.192
4. $\sigma_\varepsilon^2 = 0.05$	$U(t)$	0.032	0.123
	$V(t)$	0.076	0.717
5. $\sigma_\varepsilon^2 = 0.1$	$U(t)$	0.046	0.165
	$V(t)$	0.090	0.832

Table 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_ε^2	0.018	0.007	0.009	0.017	0.036	0.015	0.005	0.008	0.015	0.028
σ_ξ^2	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