Semiparametric Functional Factor Models with Bayesian Rank Selection[*]

Daniel R. Kowal[†] and Antonio Canale[‡]

Abstract. Functional data are frequently accompanied by a parametric template that describes the typical shapes of the functions. However, these parametric templates can incur significant bias, which undermines both utility and interpretability. To correct for model misspecification, we augment the parametric template with an infinite-dimensional nonparametric functional basis. The nonparametric basis functions are learned from the data and constrained to be orthogonal to the parametric template, which preserves distinctness between the parametric and nonparametric terms. This distinctness is essential to prevent functional confounding, which otherwise induces severe bias for the parametric terms. The nonparametric factors are regularized with an ordered spike-and-slab prior that provides consistent rank selection and satisfies several appealing theoretical properties. The versatility of the proposed approach is illustrated through applications to synthetic data, human motor control data, and dynamic yield curve data. Relative to parametric and semiparametric alternatives, the proposed semiparametric functional factor model eliminates bias, reduces excessive posterior and predictive uncertainty, and provides reliable inference on the effective number of nonparametric terms—all with minimal additional computational costs.

Keywords: factor analysis, nonparametric regression, shrinkage prior, spike-and-slab prior, yield curve.

MSC2020 subject classifications: Primary 62R10, 62F15; secondary 62P20.

1 Introduction

1.1 Setting and goals

As high-resolution monitoring and measurement systems generate vast quantities of complex and highly correlated data, *functional data analysis* has become increasingly vital for many scientific, medical, business, and industrial applications. Functional data are (noisy) realizations of random functions $\{Y_i\}_{i=1}^n$ observed over a continuous domain, such as time, space, or wavelength, and exhibit a broad variety of shapes. The concurrence of *complex* and *voluminous* data prompts the common use of nonparametric

[†]Dobelman Family Assistant Professor, Department of Statistics, Rice University, daniel.kowal@rice.edu

[‡]Associate Professor, Department of Statistics, University of Padova, antonio.canale@unipd.it

models for functional data. Yet in many cases, there is valuable information regarding the functional form of Y_i. Template curves that describe the shape of Y_i are derived from fundamental scientific laws or motivated by extensive empirical studies, and often are the focal point of the analysis. Prominent examples include human motor control (Ramsay, 2000; Goldsmith and Kitago, 2016), interest rates (Nelson and Siegel, 1987; Diebold and Li, 2006), and basal body temperature (Scarpa and Dunson, 2009, 2014; Canale et al., 2017).

Our goal is to construct a functional data modeling framework that simultaneously (i) incorporates parametric templates in a coherent and interpretable manner, (ii) maintains the modeling flexibility of nonparametric methods, and (iii) provides computationally scalable inference with reliable uncertainty quantification. The approach is fully Bayesian, accompanied by an efficient Markov chain Monte Carlo (MCMC) algorithm for posterior and predictive inference, and equally applicable to both densely-observed and sparsely- or irregularly-sampled functional data.

The parametric templates are represented as a spanning set

$$\mathcal{H}_0 = \text{span}\{g_1(\cdot;\gamma),\ldots,g_L(\cdot;\gamma)\}$$

of functions $\{g_\ell(\cdot;\gamma)\}_{\ell=1}^L$ known up to γ. Any function belonging to $\mathcal{H}_0$ is a linear combination of $\{g_\ell\}$; the corresponding coefficients and the parameters γ must be learned. Important examples are presented in Table 1. The linear basis is routinely used for longitudinal data analysis and here is equivalent to a random slope model (Molenberghs and Verbeke, 2000). A change in slope, $(\tau - \gamma)_+$ with $(x)_+ = \max\{0, x\}$, is useful for modeling structural changes, such as a change in disease transmissions due to policy interventions (Wagner et al., 2020). Cosinor functions model circadian rhythms (Mikulich et al., 2003) and other periodic behaviors (Welham et al., 2006). Biphasic curves offer utility in modeling basal body temperature of women during the menstrual cycle (Scarpa and Dunson, 2009, 2014; Canale et al., 2017). In general, interest centers on learning the linear coefficients associated with each g_ℓ, the nonlinear parameters γ, and an adequate yet interpretable model for the functions Y_i.

Linear	Linear change	Cosinor	Biphasic
$\{1, \tau\}$	$\{1, \tau, (\tau - \gamma)_+\}$	$\{1, \sin(2\pi\tau/\gamma), \cos(2\pi\tau/\gamma)\}$	$\{1, \exp(\gamma\tau)/\{1 + \exp(\gamma\tau)\}\}$

Table 1: Examples of parametric templates, or spanning sets, for $\mathcal{H}_0$.

The advantages of the parametric templates are clear: they incorporate domain knowledge, lend interpretability to the model, and often produce low-variance estimators relative to nonparametric alternatives. These templates are restrictive by design, and therefore can incur significant bias and other model misspecifications. Such effects erode model interpretability and can induce variance inflation. For illustration, we present model fits for two datasets in Figure 1 using a parametric model and our proposed semiparametric alternative; details and analyses of these data are in Section 6. In both instances, the templates capture the general shape of the data. However, regions of substantial bias are present, which produce uniformly inflated prediction bands over the domain. By comparison, the proposed approach preserves the essential shape of the curves, yet corrects the bias and shrinks the prediction bands appropriately—and crucially does so without overfitting.
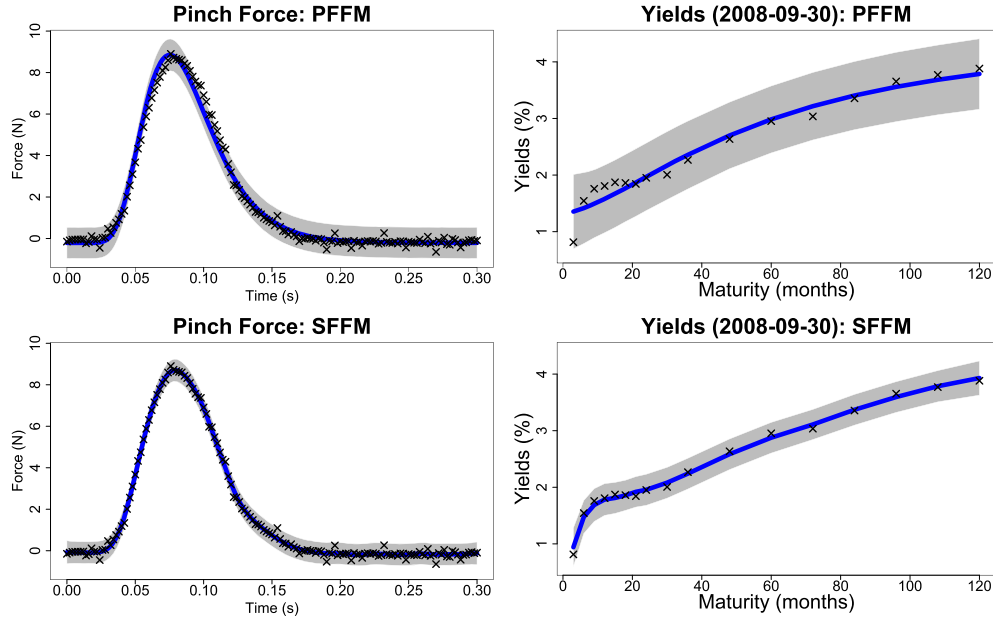
Figure 1: Posterior expectations and 95% simultaneous prediction bands for the parametric (PFFM, top) and semiparametric functional factor model (SFFM, bottom) for one replicate of the pinch force data (left; see Section 6.1) and one yield curve (right; see Section 6.2). The proposed SFFM corrects the bias of the PFFM and offers narrower prediction bands with correct coverage.

1.2 Overview of the proposed approach

The *semiparametric functional factor model* (SFFM) bridges the gap between parametric and nonparametric functional data models. The SFFM augments a parametric template $\mathcal{H}_0$ with a nonparametric and infinite-dimensional basis expansion for the functions $Y_i \in L^2(\mathcal{T})$:

$$Y_i(\tau) = \sum_{\ell=1}^{L} \alpha_{\ell,i} g_\ell(\tau; \gamma) + \sum_{k=1}^{\infty} \beta_{k,i} f_k(\tau), \quad \tau \in \mathcal{T}, \tag{1.1}$$

where $\{f_k\}$ are unknown functions, $\{\alpha_{\ell,i}\}$ and $\{\beta_{k,i}\}$ are unknown factors for the parametric and nonparametric components, respectively, and $\mathcal{T} \subseteq \mathbb{R}^d$ is the domain, usually with $d = 1$ for curves or $d = 2$ for images. Without the nonparametric terms, (1.1) is a *parametric functional factor model* (PFFM), which serves as our parametric baseline.

Our implementation of (1.1) is fully Bayesian and is based on three essential aspects.

First, the nonparametric basis $\{f_k\}$ is treated as unknown. The SFFM pools information across all functions $\{Y_i\}$ to learn the key functional features in the data—in particular, the systemic biases unresolved by $\mathcal{H}_0$. The functions $\{f_k\}$ are learned jointly

with the parametric terms γ and $\{\alpha_{\ell,i}\}$ and the nonparametric factors $\{\beta_{k,i}\}$, so uncertainty about $\{f_k\}$ is automatically absorbed into the joint posterior distribution.

Second, the nonparametric basis $\{f_k\}$ is constrained to be orthogonal to the parametric templates $\{g_\ell\}$. While orthogonalization has demonstrated its advantages in similar contexts (e.g., regression; Rossell and Rubio, 2023), our specific scenario introduces additional complexities due to the presence of unknown basis functions $\{f_k\}$. The constraint enforces distinctness between the parametric and nonparametric components of the SFFM. We show that without such a constraint, inference on the parametric terms becomes severely biased and loses interpretability due to *functional confounding*. The constraint is also essential for valid inference on the effective number of nonparametric terms, applies even in the challenging setting in which there is an unknown nonlinear parameter γ, and produces computational simplifications that improve both the algorithmic efficiency and the ease of implementation of the MCMC algorithm.

Third, the nonparametric factors $\{\beta_{k,i}\}$ are endowed with an ordered spike-and-slab prior distribution similar to successful recent approaches for modeling multivariate data via infinite factorizations (Legramanti et al., 2020; Schiavon et al., 2022; Frühwirth-Schnatter, 2023). The proposed prior is critical for coherence of the infinite-dimensional basis expansion in (1.1): it provides much-needed regularization, encourages selection of a finite number of factors, and provides posterior inference for the effective number of nonparametric terms—including an assessment of whether any nonparametric component is needed at all. The prior admits a parameter expansion that offers substantial improvements in MCMC efficiency and satisfies several key properties that broaden applicability beyond the SFFM.

The SFFM in (1.1) is accompanied by an observation error equation to accommodate noisy and sparsely- or irregularly-sampled functional data. The observed data $\boldsymbol{y}_i = (y_{i,1}, \ldots, y_{i,m_i})'$ are modeled as noisy realizations of Y_i on a discrete set of points $\{\tau_{i,j}\}_{j=1}^{m_i} \subset \mathcal{T}$ for $i = 1, \ldots, n$:

$$y_{i,j} = Y_i(\tau_{i,j}) + \epsilon_{i,j}, \quad \epsilon_{i,j} \overset{indep}{\sim} N(0, \sigma_\epsilon^2), \tag{1.2}$$

although non-Gaussian versions are available. We proceed using common observation points $\tau_{i,j} = \tau_j$ and $m_i = m$ for notational simplicity.

Although we focus on the parametric and semiparametric versions of (1.1), the proposed modeling framework remains useful without any parametric template ($L = 0$). In this case, (1.1) resembles a Karhunen-Loève decomposition and $\{f_k\}$ correspond to the eigenfunctions of the covariance function of $\{Y_i\}$ (assuming the Y_i have been centered). The Karhunen-Loève decomposition provides the theoretical foundation for functional principal components analysis (FPCA; e.g., Shang, 2014). With this perspective, our model, priors, and algorithms for (1.1) also constitute a new Bayesian view of FPCA.

1.3 Review of related approaches

Semiparametric models for functional data are predominantly non-Bayesian. To model a single function, L-splines combine a goodness-of-fit criterion with a penalty on deviations

from $\mathcal{H}_0$ (Ramsay and Dalzell, 1991; Heckman and Ramsay, 2000). Although L-splines guarantee the existence of basis functions $\{f_k\}_{k=1}^m$ that describe the deviations from $\mathcal{H}_0$, these functions are analytically intractable or highly challenging to derive for all but the simplest choices of $\mathcal{H}_0$, which inhibits widespread practical use. In addition, L-splines are not well-equipped to handle unknown γ and do not offer finite-sample inference on the adequacy of the parametric templates $\mathcal{H}_0$. By comparison, the SFFM learns $\{f_k\}$ directly from the data—which sidesteps the challenging derivations—and provides direct posterior inference on the effective number of nonparametric terms. Other non-Bayesian approaches for semiparametric functional data analysis seek to replace nonparametric functions with parametric alternatives. Sang et al. (2017) attempt to simplify FPCA by using polynomials for each FPC instead of splines or Fourier functions. In functional regression analysis, Chen et al. (2019) develop hypothesis tests to determine whether an unknown regression function deviates from a parametric template.

Bayesian semiparametric functional data models are less common. The "semiparametric" model of Lee et al. (2018) refers to additive rather than linear effects, but does not include a parametric template like $\mathcal{H}_0$. Scarpa and Dunson (2009) construct a Dirichlet process mixture of a parametric function and a Gaussian process contamination, which is generalized by Scarpa and Dunson (2014) to include prior information on the frequencies of certain functional features. These methods are designed primarily for clustering: they identify individual curves Y_i that deviate substantially from the parametric model, while the remaining curves are presumably well modeled parametrically. The SFFM is capable of modeling total deviations from $\mathcal{H}_0$ for a particular Y_i, but also captures—and corrects—partial deviations from $\mathcal{H}_0$ that persist for some or many Y_i. Unlike the mixture models, the SFFM is well-suited for including additional layers of dependence, such as hierarchical (Section 6.1) or dynamic (Section 6.2) models, while maintaining efficient posterior computing. Notably, these existing methods do not address functional confounding.

The remainder of the paper is organized as follows. The model for the parametric and nonparametric functions is in Section 2. The ordered spike-and-slab prior is introduced and studied in Section 3. The MCMC algorithm is discussed in Section 4. A simulation study is in Section 5. The model is applied to real datasets in Section 6. We conclude in Section 7. Online supplementary material includes proofs of all results, the full MCMC algorithm, additional simulations, and R code for reproducibility (Kowal and Canale, 2023a,b).

2 Modeling the nonparametric functions

2.1 The need for constraints

The error-free latent functions $\{Y_i\}$ belong to the space spanned by the template parametric curves $\{g_\ell\}$ and the nonparametric curves $\{f_k\}$. Any systemic bias resulting from the inadequacies of $\{g_\ell\}$ must be corrected by $\{f_k\}$, which demands substantial flexibility of the nonparametric basis $\{f_k\}$. However, interpretability of the parametric terms $\{g_\ell\}$ and $\{\alpha_{\ell,i}\}$ requires a strict distinction between the parametric and nonparametric components. In conjunction, these points demand both *flexibility* and *restraint* from the

nonparametric term. The challenges for modeling and computing are exacerbated when the parametric nonlinear term γ is unknown.

To emphasize the importance of the distinction between $\{g_\ell\}$ and $\{f_k\}$, consider a seemingly reasonable alternative to (1.1) that replaces the nonparametric term with a Gaussian process (PFFM+GP):

$$Y_i(\tau) = \sum_{\ell=1}^{L} \alpha_{\ell,i} g_\ell(\tau; \gamma) + h_i(\tau), \quad h_i \sim \mathcal{GP}(0, \mathcal{K}_h). \tag{2.1}$$

The PFFM+GP is a special case of the SFFM (1.1), and implies that each latent curve Y_i is a Gaussian process centered at the parametric term. Naturally, it may be expected that each h_i corrects for the biases of the parametric component. However, there is a significant cost: despite being centered at zero, the Gaussian processes $\{h_i\}$ are not constrained to be sufficiently distinct from the parametric term. This introduces a *functional confounding* that biases the parametric factors $\{\alpha_{\ell,i}\}$.

To illustrate this point, we fit the PFFM, PFFM+GP, and SFFM to the pinch force data (see Section 6.1) with identical parametric models (see (6.1)). The resulting posterior expectations of each $\{\alpha_{\ell,i}\}$ are presented in Figure 2. Most striking, the point estimates of the parametric factors are nearly identical between the PFFM and SFFM, while the PFFM+GP estimates are substantially different. The discrepancy of the PFFM+GP is concerning: under mild conditions, the PFFM will produce (approximately) unbiased estimates of the linear coefficients $\{\alpha_{\ell,i}\}$ even in the presence of dependent errors, so the point estimates are a reliable benchmark. This is not an issue of model inadequacy for PFFM+GP: its fitted values are nearly identical to those from the SFFM. These results are also confirmed in the simulation study using the ground truth values of $\{\alpha_{\ell,i}\}$ (Section 5). We emphasize that despite this agreement between the PFFM and the SFFM for point estimation of $\{\alpha_{\ell,i}\}$, the SFFM *does* identify inadequacies of the PFFM and provides key improvements in bias removal for $\{Y_i\}$ and uncertainty quantification for both $\{Y_i\}$ and $\{\alpha_{\ell,i}\}$ (see Sections 5–6).

The simple augmentation of a Gaussian process in (2.1) offers flexibility but lacks restraint: $\{h_i\}$ absorbs variability otherwise explained by $\{g_\ell\}$, which ultimately corrupts inference and interpretation of the parametric term. This functional confounding is not simply an artifact of the PFFM+GP sampling algorithm, which uses a *joint* sampler for $\{\alpha_{\ell,i}, h_i\}$ (see the supplement), and is robust to the choice of the covariance function $\mathcal{K}_h$. This issue is related to spatial confounding (e.g., Reich et al., 2006) but has not been thoroughly explored or resolved for functional data.

2.2 Models and constraints for the nonparametric functions

The proposed SFFM achieves both flexibility and distinctness by (i) modeling each f_k in (1.1) as a smooth unknown function and (ii) constraining each f_k to be orthogonal to $\{g_\ell\}$. The model for f_k can be any Bayesian curve-fitting model, such as splines, wavelets, or Gaussian processes, usually with a prior that encourages smoothness; our specifications are discussed subsequently. Crucially, our choice of constraints not only
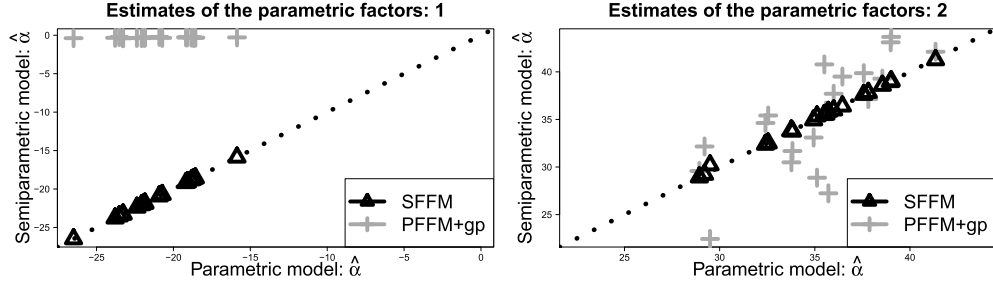
Figure 2: Posterior expectations of $\{\alpha_{\ell,i}\}_{i=1}^{n}$ for $\ell = 1$ (left) and $\ell = 2$ (right) for the pinch force data (see Section 6.1) comparing the PFFM estimates (x-axis) against both the SFFM and PFFM+GP (y-axis). The point estimates of the parametric factors are nearly identical between the PFFM and SFFM, while the PFFM+GP estimates are substantially different, and near zero for α_1. The dotted line denotes $y = x$.

ensure distinctness of $\{f_k\}$ from $\{g_\ell\}$, but also offer key computational simplifications that improve scalability and increase MCMC efficiency. These results maintain regardless of the specification of $\{g_\ell\}$ or $\{f_k\}$, including when γ is unknown.

First, each unknown function f_k expressed using known basis functions $\boldsymbol{b}$, and a prior is placed on the unknown coefficients to encourage smoothness:

$$f_k(\tau) = \boldsymbol{b}'(\tau)\boldsymbol{\psi}_k, \quad [\boldsymbol{\psi}_k \mid \lambda_{f_k}] \sim N(\boldsymbol{0}, \lambda_{f_k}^{-1}\boldsymbol{\Omega}^-), \quad \lambda_{f_k}^{-1/2} \sim \text{Uniform}(0, 10^4), \qquad (2.2)$$

where $\boldsymbol{\Omega}$ is the (known) matrix of integrated squared second derivatives, $[\boldsymbol{\Omega}]_{j,j'} = \int \ddot{b}_j(\tau)\ddot{b}_{j'}(\tau)d\tau$. This prior is a Bayesian analog of the classical roughness penalty

$$-2 \log p(\boldsymbol{\psi}_k \mid \lambda_{f_k}) \overset{c}{=} \lambda_{f_k} \boldsymbol{\psi}_k' \boldsymbol{\Omega} \boldsymbol{\psi}_k = \lambda_{f_k} \int_{\mathcal{T}} \{\ddot{f}_k(\tau)\}^2 d\tau$$

with a smoothing parameter λ_{f_k} that is assigned a prior in (2.2) and learned jointly with the model parameters. Although we proceed under general conditions, our default specification uses low-rank thin plate splines for $\boldsymbol{b}$, which are flexible, computationally efficient, and defined on $\mathcal{T} \subseteq \mathbb{R}^d$.

Consider model (1.1)–(1.2) evaluated at the observation points $\{\tau_j\}_{j=1}^{m}$:

$$\boldsymbol{y}_i = \boldsymbol{G}_\gamma \boldsymbol{\alpha}_i + \boldsymbol{F}\boldsymbol{\beta}_i + \boldsymbol{\epsilon}_i, \quad \boldsymbol{\epsilon}_i \overset{iid}{\sim} N(\boldsymbol{0}, \sigma_\epsilon^2 \boldsymbol{I}_m), \quad i = 1, \dots, n, \qquad (2.3)$$

where

- $\boldsymbol{G}_\gamma = (\boldsymbol{g}_{1;\gamma}, \dots, \boldsymbol{g}_{L;\gamma})$ is the $m \times L$ parametric basis matrix, $\boldsymbol{g}_{\ell;\gamma} = (g_\ell(\tau_1; \gamma), \dots, g_\ell(\tau_m; \gamma))'$;

- $\boldsymbol{\alpha}_i = (\alpha_{1,i}, \dots, \alpha_{L,i})'$ is the vector of parametric factors;

- $\boldsymbol{F} = (\boldsymbol{f}_1, \boldsymbol{f}_2, \ldots)$ is the $m \times \infty$ nonparametric basis matrix with $\boldsymbol{f}_k = (f_k(\tau_1), \ldots, f_k(\tau_m))'$;

- $\boldsymbol{\beta}_i = (\beta_{1,i}, \beta_{2,i}, \ldots)'$ is the vector of nonparametric factors.

By constraining each $\boldsymbol{f}_k$ to be orthogonal to $\boldsymbol{G}_\gamma$, the likelihood (2.3) factors into parametric and nonparametric terms:

Lemma 1. *When* $\boldsymbol{G}'_\gamma \boldsymbol{f}_k = \boldsymbol{0}_L$ *for all* $k = 1, 2, \ldots$, *the likelihood* (2.3) *factorizes:*

$$p(\boldsymbol{y} \mid \gamma, \{\boldsymbol{\alpha}_i\}, \boldsymbol{F}, \{\boldsymbol{\beta}_i\}, \sigma_\epsilon^2) = p_0(\boldsymbol{y} \mid \gamma, \{\boldsymbol{\alpha}_i\}, \sigma_\epsilon^2) \, p_1(\boldsymbol{y} \mid \boldsymbol{F}, \{\boldsymbol{\beta}_i\}, \sigma_\epsilon^2),$$

where p_0 *is invariant of the nonparametric terms* ($\boldsymbol{F}$ *and* $\{\boldsymbol{\beta}_i\}$) *and* p_1 *is invariant of the parametric terms* (γ *and* $\{\boldsymbol{\alpha}_i\}$).

The primary implication of Lemma 1 is that, under prior independence between the parametric and nonparametric factors, $p(\{\boldsymbol{\alpha}_i\}, \{\boldsymbol{\beta}_i\}) = p(\{\boldsymbol{\alpha}_i\}) \, p(\{\boldsymbol{\beta}_i\})$, the parametric factors $\{\boldsymbol{\alpha}_i\}$ and nonparametric factors $\{\boldsymbol{\beta}_i\}$ are (conditionally) independent in the posterior. Thus, the sampling steps for $\{\boldsymbol{\alpha}_i\}$ and $\{\boldsymbol{\beta}_i\}$—under orthogonality $\boldsymbol{G}'_\gamma \boldsymbol{f}_k = \boldsymbol{0}_L$ for all k and conditional on γ, $\boldsymbol{F}$, and σ_ϵ^2—are completely decoupled: the sampling steps for $\{\boldsymbol{\alpha}_i\}$ are identical to the fully parametric case, while the sampling steps for $\{\boldsymbol{\beta}_i\}$ proceed exactly as in a fully nonparametric setting (see Section 3). Further, these separate sampling steps for $\{\boldsymbol{\alpha}_i\}$ and $\{\boldsymbol{\beta}_i\}$ are equivalent to a *joint* sampler for $\{\boldsymbol{\alpha}_i, \boldsymbol{\beta}_i\}$. Joint sampling steps are typically preferred for greater MCMC efficiency, but often increase the computational burden. For instance, a joint sampler for $\{\boldsymbol{\alpha}_i, h_i\}$ under the PFFM+GP requires either a large block sampler or marginalization over either $\boldsymbol{\alpha}_i$ or h_i, both of which become more challenging when the model for $\boldsymbol{\alpha}_i$ is complex. The SFFM—with the proposed orthogonality constraint—completely avoids this tradeoff and guarantees joint sampling even for complex models. Such a reduction is not available in general, including for the L^2-orthogonality constraint $\int_{\mathcal{T}} g_\ell(\tau; \gamma) f_k(\tau) \, d\tau = 0$ that appears often in functional data analysis.

We enforce the orthogonality constraint $\boldsymbol{G}'_\gamma \boldsymbol{f}_k = \boldsymbol{0}_L$ by *conditioning* on it in the model (2.2) for $\boldsymbol{f}_k$. Conditioning is a natural Bayesian mechanism for incorporating such information into estimation and inference. This is especially useful when γ is unknown, and thus the orthogonality constraint updates at every MCMC iteration. Samples from the full conditional distribution of $\boldsymbol{\psi}_k$ conditional on $\boldsymbol{G}'_\gamma \boldsymbol{f}_k = \boldsymbol{0}_L$ can be obtained by first sampling from the unconstrained distribution $[\boldsymbol{\psi}_k \mid -]$ and then applying a simple shift. Specifically, $[\boldsymbol{\psi}_k \mid -] \sim N(\boldsymbol{Q}_{\psi_k}^{-1} \boldsymbol{\ell}_{\psi_k}, \boldsymbol{Q}_{\psi_k}^{-1})$, where

$$\boldsymbol{Q}_{\psi_k} = \sigma_\epsilon^{-2} (\boldsymbol{B}'\boldsymbol{B}) \sum_{i=1}^n \beta_{k,i}^2 + \lambda_{f_k} \boldsymbol{\Omega}, \quad \boldsymbol{\ell}_{\psi_k} = \sigma_\epsilon^{-2} \boldsymbol{B}' \sum_{i=1}^n \big\{ \beta_{k,i} \big(\boldsymbol{y}_i - \boldsymbol{G}_\gamma \boldsymbol{\alpha}_i - \sum_{\ell \neq k} \boldsymbol{f}_\ell \beta_{\ell,i} \big) \big\},$$

and $\boldsymbol{B} = (\boldsymbol{b}'(\tau_1), \ldots, \boldsymbol{b}'(\tau_m))'$, and the shift is applied as follows:

Lemma 2. *Suppose* $\boldsymbol{G}_\gamma$ *is* $L \times m$ *with rank* L. *Let* $\boldsymbol{f}_k^0 = \boldsymbol{B}\boldsymbol{\psi}_k^0$ *where* $\boldsymbol{\psi}_k^0 \sim N(\boldsymbol{Q}_{\psi_k}^{-1} \boldsymbol{\ell}_{\psi_k}, \boldsymbol{Q}_{\psi_k}^{-1})$. *The shifted term*

$$\boldsymbol{f}_k = \boldsymbol{B}\boldsymbol{\psi}_k, \quad \boldsymbol{\psi}_k = \boldsymbol{\psi}_k^0 - \boldsymbol{Q}_{\psi_k}^{-1} \boldsymbol{B}' \boldsymbol{G}_\gamma \big(\boldsymbol{G}'_\gamma \boldsymbol{B} \boldsymbol{Q}_{\psi_k}^{-1} \boldsymbol{B}' \boldsymbol{G}_\gamma \big)^{-1} \boldsymbol{G}'_\gamma \boldsymbol{B} \boldsymbol{\psi}_k^0$$

satisfies $\boldsymbol{f}_k \stackrel{d}{=} [\boldsymbol{f}_k^0 \mid \boldsymbol{G}'_\gamma \boldsymbol{f}_k^0 = \boldsymbol{0}]$ *and* $\mathbb{P}(\boldsymbol{G}'_\gamma \boldsymbol{f}_k = \boldsymbol{0}) = 1$.

Lemma 2 provides a simple and effective mechanism to adapt any standard (basis) regression sampling step for $\boldsymbol{\psi}_k$ to accommodate this crucial orthogonality constraint, including when the nonlinear term γ is unknown. The computational cost to sample the unconstrained coefficients is cubic in the length of $\boldsymbol{\psi}_k$ (i.e., the number of columns of $\boldsymbol{B}$); enforcing the constraints is cubic in the number of constraints. Although it is possible to adapt Lemma 2 for the PFFM+GP, the resulting model still does not provide direct inference on the necessity of the nonparametric factors (i.e., rank selection) or describe the systemic biases of $\mathcal{H}_0$, which are captured by $\{f_k\}$ in the SFFM.

For further computational simplifications along with parameter identifiability, we additionally constrain $\boldsymbol{G}_\gamma$ and $\boldsymbol{F}$ to be *orthonormal*, which implies that the joint basis $\{g_1, \ldots, g_L, f_1, f_2, \ldots\}$ in (1.1) is orthonormal. First, for each nonparametric term $\boldsymbol{f}_k$, we incorporate the *orthogonality* constraints via conditioning on $\boldsymbol{f}_{k'}'\boldsymbol{f}_k = 0$ for all $k' \neq k$ and the *unit-norm* constraint by rescaling $\boldsymbol{f}_k$ appropriately. These steps follow Kowal (2021) and are described briefly. The linear orthogonality constraint is enforced using a straightforward modification of Lemma 2, which augments the parametric orthogonality of $\boldsymbol{G}_\gamma$ with the nonparametric terms $\boldsymbol{f}_{k'}$ for $k' \neq k$. The unit-norm constraint is enforced by suitably rescaling $\boldsymbol{\psi}_k$ after sampling; we equivalently rescale $\beta_{k,i}$ to preserve the product $\boldsymbol{f}_k \beta_{k,i}$ in the likelihood (2.3). This operation does not change the shape of the curve f_k nor the likelihood (2.3).

When paired with the orthogonality $\boldsymbol{G}_\gamma'\boldsymbol{f}_k = \boldsymbol{0}_L$ for all k, the orthonormality of $\boldsymbol{F}$ offers convenient simplifications for posterior sampling:

Corollary 1. *Under model* (2.3) *with prior independence between* $\{\boldsymbol{\alpha}_i\}$ *and* $\{\boldsymbol{\beta}_i\}$ *and subject to* $\boldsymbol{G}_\gamma'\boldsymbol{f}_k = \boldsymbol{0}_L$ *for all* k *and* $\boldsymbol{F}'\boldsymbol{F} = \boldsymbol{I}_\infty$, *the full conditional posterior for the nonparametric factors is*

$$p(\{\boldsymbol{\beta}_i\} \mid \boldsymbol{y}, \gamma, \boldsymbol{F}, \{\boldsymbol{\alpha}_i\}) \propto p_{1,F}(\boldsymbol{y} \mid \boldsymbol{F}, \{\boldsymbol{\beta}_i\})\ p(\{\boldsymbol{\beta}_i\}),$$

where $p_{1,F}$ *is the likelihood defined by* $[\boldsymbol{f}_k'\boldsymbol{y}_i \mid -] \overset{indep}{\sim} N(\beta_{k,i}, \sigma_\epsilon^2)$ *for all* k.

The likelihood from (2.3) reduces to a low-dimensional, independent Gaussian model, which makes for convenient and efficient posterior sampling steps for $\{\beta_{k,i}\}$. Thus, the SFFM is capable of incorporating complex prior specifications for $\{\beta_{k,i}\}$—which are essential for inference in the SFFM with the *infinite* expansion in (1.1).

Next, we construct an orthonormal parametric basis $\boldsymbol{G}_\gamma$ using a QR decomposition. For any value of γ, let $\boldsymbol{G}_\gamma^0 = \boldsymbol{Q}_\gamma \boldsymbol{R}_\gamma$ be the QR decomposition of the initial basis matrix $\boldsymbol{G}_\gamma^0 = (\boldsymbol{g}_{1;\gamma}, \ldots, \boldsymbol{g}_{L;\gamma})$, where $\boldsymbol{Q}_\gamma$ is orthogonal and $\boldsymbol{R}_\gamma$ is upper triangular. By setting $\boldsymbol{G}_\gamma = \boldsymbol{Q}_\gamma$, we ensure that $\boldsymbol{G}_\gamma'\boldsymbol{G}_\gamma = \boldsymbol{I}_L$ and the columns of $\boldsymbol{G}_\gamma$ span the same space as the columns of $\boldsymbol{G}_\gamma^0$. When γ is unknown and endowed with a prior distribution, the QR decomposition is incorporated into the likelihood evaluations of (2.3) for posterior sampling of γ, with computational complexity $\mathcal{O}(mL^2 - L^3/3)$ (Boyd and Vandenberghe, 2004). Although this QR step is not strictly necessary for the SFFM, we include it because (i) it places the parametric and nonparametric factors on the same scale; (ii) it simplifies the likelihood for $\{\boldsymbol{\alpha}_i\}$ analogous to Corollary 1, thus providing modularity to incorporate more complex models for $\{\boldsymbol{\alpha}_i\}$ (see Section 6); and (iii) orthonormal design

matrices ($\boldsymbol{G}_\gamma$) often improve MCMC efficiency for the corresponding coefficients ($\{\boldsymbol{\alpha}_i\}$). Since $\boldsymbol{G}_\gamma^0 \boldsymbol{\alpha}_i^0 = \boldsymbol{G}_\gamma \boldsymbol{\alpha}_i$ for $\boldsymbol{\alpha}_i = \boldsymbol{R}_\gamma \boldsymbol{\alpha}_i^0$ under the QR decomposition, we can recover the parametric factors on the original scale by setting $\boldsymbol{\alpha}_i^0 = \boldsymbol{R}_\gamma^{-1} \boldsymbol{\alpha}_i$, which can be computed draw by draw within the MCMC sampler.

3   Models and rank selection for the nonparametric factors

3.1   Ordered spike-and-slab priors

Regularization of the SFFM is critical: the infinite-dimensional nonparametric term in (1.1) is clearly overparametrized for finite m. The nonparametric components increase model complexity and should be removed whenever the added complexity is not supported by the data. This *rank selection* is complicated by the presence of the parametric template with unknown $\{\alpha_{\ell,i}\}$ and γ.

We focus on *ordered spike-and-slab* priors that apply joint shrinkage and selection to the nonparametric factors $\{\beta_{k,i}\}_{i=1}^n$ to remove unnecessary nonparametric terms from the SFFM. Because of the proposed orthogonality constraints on $\{g_\ell\}$ and $\{f_k\}$ and Corollary 1, we are able to introduce a sophisticated prior for $\{\beta_{k,i}\}$ within the broader SFFM framework of (1.1)—and do so with minimal impact on computational cost and MCMC efficiency relative to the PFFM (see Section 6).

We leverage the *cumulative shrinkage process* (CUSP; Legramanti et al., 2020), which uses an ordered spike-and-slab prior for the factors $[\beta_{k,i} \mid \theta_k] \overset{indep}{\sim} N(0, \theta_k)$:

$$[\theta_k \mid \pi_k] \sim P_k, \quad P_k = (1 - \pi_k)P_{slab} + \pi_k P_{spike}, \tag{3.1}$$

$$\pi_k = \sum_{h=1}^k \omega_h, \quad \omega_h = \nu_h \prod_{\ell=1}^{h-1}(1 - \nu_\ell), \quad \nu_\ell \overset{iid}{\sim} \text{Beta}(\iota, \iota\kappa), \tag{3.2}$$

where P_{slab} is the distribution of the active component and P_{spike} is the distribution of the inactive component. The cumulative summation for π_k ensures an increasing sequence of (spike) probabilities, $\pi_k < \pi_{k+1}$, that converges: $\lim_{k\to\infty} \pi_k = 1$. The hyperparameters are determined by the distributions P_{slab} and P_{spike} and the scalars $\iota, \kappa > 0$. Legramanti et al. (2020) used $P_{spike} = \delta_{\theta_\infty}$ with $\theta_\infty = 0.05$, $P_{slab} = \text{Inv-Gamma}(a_\theta, b_\theta)$ with $a_\theta = b_\theta = 2$, and fixed $\iota = 1$ and $\kappa = 5$.

By ordering the spike probabilities $\{\pi_k\}_{k=1}^\infty$ in (3.2), CUSP guarantees an ordering for the parameters $\{\theta_k\}_{k=1}^\infty$ in (3.1).

Proposition 1. *For $\varepsilon > 0$ and fixed θ_0, let $\mathbb{B}_\varepsilon(\theta_0) = \{\theta_k : |\theta_k - \theta_0| < \varepsilon\}$. Prior (3.1)–(3.2) implies that $\mathbb{P}(|\theta_k - \theta_0| \le \varepsilon) < \mathbb{P}(|\theta_{k+1} - \theta_0| \le \varepsilon)$ whenever $P_{slab}\{\mathbb{B}_\varepsilon(\theta_0)\} < P_{spike}\{\mathbb{B}_\varepsilon(\theta_0)\}$.*

The special case of Proposition 1 with $P_{spike} = \delta_{\theta_\infty}$, $\iota = 1$, and $\theta_0 = 0$ is proved by Legramanti et al. (2020). Intuitively, for any θ_0 that is favored under the spike

distribution P_{spike} relative to the slab distribution P_{slab}, CUSP places greater mass around θ_0 as k increases.

For interpretability and efficient MCMC sampling, Legramanti et al. (2020) introduced a convenient data augmentation. Let $z_k \in \{1, \ldots, \infty\}$ denote a categorical variable with $\mathbb{P}(z_k = h \mid \omega_h) = \omega_h$. Then

$$[\theta_k \mid z_k] \sim (1 - \mathbb{I}\{z_k \leq k\})P_{slab} + \mathbb{I}\{z_k \leq k\}P_{spike}$$

induces (3.1) via marginalization over z_k. The number of active (slab) terms is therefore

$$K^* = \sum_{k=1}^{\infty} \mathbb{I}\{z_k > k\}, \tag{3.3}$$

with posterior inference available through the proposed MCMC sampling algorithm (see Algorithm 1). For model (1.1), K^* is the effective number of nonparametric terms, and therefore is a key inferential target to assess the adequacy of the parametric model.

A subtle yet important limitation of CUSP is that spike-and-slab selection is applied jointly to all n factors $\{\beta_{k,i}\}_{i=1}^n$ via a single scalar θ_k. In a related setting, Scheipl et al. (2012) showed that as the size of this collection increases (here, n), it becomes more difficult for the MCMC algorithm to switch between the slab and spike components. Thus, the uncertainty quantification for each factor's activeness is invalidated, and inference for the effective number of nonparametric factors K^* is not reliable. To illustrate this, consider the left panel of Figure 3 which displays the MCMC samples of K^* across five different MCMC runs with random initializations obtained with the standard CUSP prior and sampling algorithm for model (1.1). In this case, the original sampler reported in Legramanti et al. (2020) produces MCMC draws that are essentially stuck at a single value: K^* changes values in only 0.004% of the MCMC iterations.
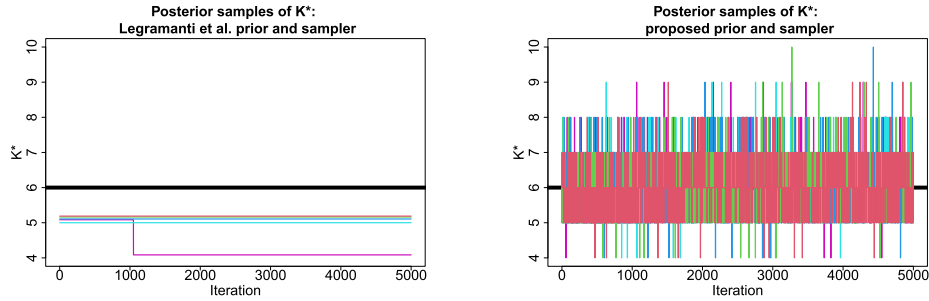

Figure 3: MCMC samples of K^* across 5 chains for Legramanti et al. (2020) (left) compared to the proposed prior and algorithm (right). The horizontal black line denotes the true value $K_{true} = 6$. The proposed approach provides substantial improvements in mixing and produces more accurate estimates.

To address these limitations, we propose the following hierarchical representation

$$\beta_{k,i} = \eta_k \xi_{k,i}, \tag{3.4}$$

$$[\eta_k \mid \pi_k] \sim (1 - \pi_k)t_{2a_1}(0, \sqrt{a_2/a_1}) + \pi_k t_{2a_1}(0, \sqrt{v_0 a_2/a_1}), \tag{3.5}$$

$$[\xi_{k,i} \mid m_{\xi_{k,i}}] \sim N(m_{\xi_{k,i}}, 1), \quad m_{\xi_{k,i}} \sim \frac{1}{2}\delta_1 + \frac{1}{2}\delta_{-1} \tag{3.6}$$

where $\{\pi_k\}$ follows (3.2), $t_d(m, s)$ denotes a t-distribution with mean m, standard deviation s, and degrees of freedom d, and v_0, a_1, and a_2 are hyperparameters.

The prior (3.4)–(3.6) has several key properties. First, ordered shrinkage is applied via $\{\eta_k\}$, which propagates to the nonparametric factors $\{\beta_{k,i}\}$ as desired.

Proposition 2. *For $\varepsilon > 0$, $\mathbb{P}(|\beta_{k,i}| \leq \varepsilon) < \mathbb{P}(|\beta_{k+1,i}| \leq \varepsilon)$ whenever $v_0 < 1$.*

The restriction $v_0 < 1$ is self-evident, since it ensures that the spike distribution is indeed more concentrated around zero. Proposition 2 requires only that $\{\xi_{k,i}\}$ are marginally independent and identically distributed. By a similar argument, the ordering also is preserved across observations: $\mathbb{P}(|\beta_{k,i}| \leq \varepsilon) < \mathbb{P}(|\beta_{k+1,i'}| \leq \varepsilon)$.

Second, the prior on η_k can be equivalently represented as

$$[\eta_k \mid \theta_k, \sigma_k^2] \sim N(0, \theta_k \sigma_k^2), \quad [\theta_k \mid \pi_k] \sim (1 - \pi_k)\delta_1 + \pi_k \delta_{v_0}, \quad [\sigma_k^{-2}] \sim \text{Gamma}(a_1, a_2) \tag{3.7}$$

and referred to as the normal mixture of inverse-gamma (NMIG) prior. The NMIG prior incorporates variable selection by assigning the variance scale parameter θ_k to the slab component, $\theta_k = 1$, or the spike component, $\theta_k = v_0 \ll 1$. Although other spike-and-slab priors are available (George and McCulloch, 1993; O'Hara and Sillanpää, 2009), the prior in (3.7) produces a continuous distribution for the conditional variance of η_k, which is preferable for variable selection and risk properties (Ishwaran and Rao, 2005). The representation in (3.7) is also convenient for computing (Algorithm 1).

Third, (3.6) introduces a redundant parameter expansion to improve MCMC mixing. Similar strategies have been applied for hierarchical models (Gelman, 2006), additive models (Scheipl et al., 2012), and functional autoregression models (Kowal et al., 2019). The Gaussian parameter expansion of $\xi_{k,i}$ is centered around 1 or -1 with variance 1, so η_k assigns the magnitude and selection to $\{\beta_{k,i}\}_{i=1}^n$. The impact is striking: Figure 3 (right) shows the MCMC samples of K^* (again across five different MCMC runs with random initializations). The proposed approach produces draws that are centered at the true value with some variability: K^* changes values in 32% of the MCMC iterations. Consistently, the MCMC-based uncertainty quantification from the proposed approach is more plausible—which is essential for inference on the effective number of nonparametric terms in the SFFM (1.1).

We additionally propose to put a hyperprior on κ as this parameter has the important role of centering the prior on the number of slab terms, i.e., $\mathbb{E}(K^*) = \kappa$. Specifically, we let $\kappa \sim \text{Gamma}(a_\kappa, b_\kappa)$ for $a_\kappa, b_\kappa > 0$, which is conditionally conjugate to (3.2). The hyperparameters may be selected to provide weak prior information regarding the number of factors: in practice, we set $a_\kappa = 2$ and $b_\kappa = 1$ so that $\mathbb{E}(\kappa) = \text{Var}(\kappa) = 2$. Similarly, we select the default value $\iota = 1$ for simplicity.

3.2   Connections to Indian Buffet Processes

A key ingredient of CUSP is the stick-breaking construction for ω_h in (3.2), which is also
commonly used for Dirichlet processes (DPs; Ferguson, 1973). A related nonparametric
Bayesian process is the so-called Indian buffet process (IBP; Griffiths and Ghahramani,
2011), which is widely used for (non-functional) Bayesian factor models (Rai and III,
2009; Ročková and George, 2016; Ohn and Kim, 2021). The IBP defines a prior over
binary matrices $B = \{b_{j,k}\}$ with m rows and infinitely many columns, and specifically
by establishing the independent conditional distributions

$$[b_{j,k} \mid \mu_k] \sim \text{Bernoulli}(\mu_k), \quad [\mu_k] \sim \text{Beta}(\iota\kappa/K, \iota), \tag{3.8}$$

usually with $\iota = 1$, and then integrating over $\{\mu_k\}$ with $K \to \infty$ (the standard IBP
notation has been adapted to match (3.2)).

The distinctions between CUSP and IBP are illuminated by the comparative stick-
breaking representations. In contrast to the DP stick-breaking process for $\{\omega_h\}$ in (3.2),
the IBP stick-breaking process is defined by Teh et al. (2007) for the ordered (slab)
probabilities $\mu_{(1)} > \cdots > \mu_{(K)}$:

$$\mu_{(k)} = \prod_{\ell=1}^{k} \nu'_\ell, \quad \nu'_\ell \overset{iid}{\sim} \text{Beta}(\iota\kappa, \iota). \tag{3.9}$$

The DP stick lengths $\{\omega_h\}$ sum to one but are not decreasing, while IBP stick lengths
$\{\mu_{(k)}\}$ do not necessarily sum to one but are decreasing.

By instead considering the *cumulative summation* in (3.2), we uncover a more direct
connection with IBPs. Unlike the DP stick-breaking process, the cumulative summation
for π_k ensures an increasing sequence of (spike) probabilities, $\pi_k < \pi_{k+1}$, that converges:
$\lim_{k\to\infty} \pi_k = 1$. Most interesting, the CUSP probability sequence $\{\pi_k\}_{k=1}^\infty$ has a simple
mapping to the IBP stick-breaking construction:

Proposition 3. *The CUSP* (3.2) *satisfies* $(1 - \pi_k) = \mu_{(k)}$ *where* $\mu_{(1)} > \cdots > \mu_{(K)}$ *are
the ordered probabilities from the stick-breaking construction* (3.9) *of the IBP* (3.8).

Conventionally, IBPs apply the multiplicative beta process to the slab probability
$(1 - \pi_k)$, which explains the complement in Proposition 3.

Despite these fundamental connections between $\{\pi_k\}$ and $\{\mu_{(k)}\}$, CUSPs and IBPs
remain distinct due to (3.1): CUSPs define a spike-and-slab prior for a sequence of pa-
rameters $\{\theta_k\}_{k=1}^\infty$, while IBPs define a prior over binary matrices $B = \{b_{j,k}\}$. In the
context of (1.1), elementwise sparsity of $\boldsymbol{F} = (\boldsymbol{f}_1, \boldsymbol{f}_2, \ldots)$ for $\boldsymbol{f}_k = (f_k(\tau_1), \ldots, f_k(\tau_m))'$
is unwarranted: although we are interested in rank selection, we prefer priors that en-
courage smoothness rather than sparsity in $f_k(\tau)$ over $\tau \in \mathcal{T}$.

3.3   Consistent rank selection

To motivate the use of the NMIG and CUSP priors, we investigate their asymptotic
behavior in a simplified setting. Suppose we directly observe a noisy version of the

parameters η_k in sequence:

$$y_i = \eta_i + \epsilon_i, \quad \epsilon_i \overset{iid}{\sim} N(0,1), \quad i = 1, 2, \ldots. \tag{3.10}$$

While this setting is quite far from the SFFM model in (1.1), we note that the likelihood implied by (3.10) is equivalent to the likelihood defined in Corollary 1, with $\beta_{k,i} = \eta_{k,i}$, $n = 1$, and $\sigma_\epsilon^2 = 1$. We assume that the true data generating process is

$$y_i = \eta_{0i} + \epsilon_i, \quad \epsilon_i \overset{iid}{\sim} N(0,1), \quad i = 1, 2, \ldots, \tag{3.11}$$

where $\eta_{01}, \ldots, \eta_{0n}$ are "true" mean parameters. This setting is related to the problem of estimating a high-dimensional mean vector from a single multivariate observation (Castillo and van der Vaart, 2012; Rockova, 2018), but with ordered sparsity $\eta_{0k} \neq 0$ for $k \leq K_{0n}$ and $\eta_{0k} = 0$ for $k > K_{0n}$. As such, K_{0n} is the true number of nonzero means and corresponds to the true rank in model (1.1). As customary in posterior asymptotics, we allow K_{0n} to grow with n.

We study the posterior of η_i under the CUSP (3.1)–(3.2) and NMIG (3.7) prior. A key term is the "remainder" $R_n = \sum_{k \geq K_{0n}} \omega_k$, which represents an upper bound for the probability mass assigned to the slab for the parameters that are null. Under the CUSP, we show that R_n is far from zero with vanishing probability *a priori*:

Lemma 3. *Let $\varepsilon_n \to 0$ with $\varepsilon_n^{1/K_{0n}} > \kappa/(\kappa+1)$. For the CUSP prior* (3.1)–(3.2) *and a positive constant $C > 1$, the remainder term $R_n = \sum_{k \geq K_{0n}} \omega_k$ satisfies*

$$\mathbb{P}(R_n > \varepsilon_n) \leq \exp(-CK_{0n}). \tag{3.12}$$

According to Lemma 3, the prior probability that R_n exceeds a small threshold is exponentially small. Such exponential decay in the prior is essential to obtain optimal posterior behavior in sparse settings (Castillo and van der Vaart, 2012). This result does not require the NMIG prior (3.7).

To connect the prior behavior in Lemma 3 to the posterior, first consider the following mild conditions on the hyperparameters in (3.7):

(C1) $a_1/(a_2 v_0) > n^2/K_{0n}$;

(C2) $AK_{0n}/n < 1/2$ with $A > 1/2$;

(C3) $1/2 < a_1 \leq a_2$.

Condition (C1) implies that the precision of the spike component is increasing with n; condition (C2) requires that K_{0n} is strictly less than the total number of elements (or the truncation limit); and condition (C3) ensures that the variances of the slab component are sufficiently large. Similar conditions appear in Castillo and van der Vaart (2012) and in Rockova (2018). Now let $\mathbb{P}_0$ and $\mathbb{E}_0$ denote the probability and the expectation, respectively, under the true distribution of the data (3.11). Under (C1)–(C3), the posterior probability that the remainder exceeds a shrinking ε_n, for $n \to \infty$, goes to zero in expectation $\mathbb{E}_0$. This is formalized in the following result:

Theorem 1. *Let $\varepsilon_n \to 0$ with $\varepsilon_n^{1/K_{0n}} > \kappa/(\kappa+1)$ and assume (C1)–(C3) and $C > 2Ae$. For the CUSP* (3.1)–(3.2) *and NMIG* (3.7) *priors, the posterior distribution satisfies*

$$\lim_{n \to \infty} \mathbb{E}_0 \left\{ \mathbb{P} \left(R_n > \varepsilon_n \mid y_1, \ldots, y_n \right) \right\} = 0.$$

4 MCMC for posterior inference

We design an MCMC algorithm for posterior inference. The proposed Gibbs sampler is both computationally scalable and MCMC efficient (see Tables 3 and 5) due to two key features: the orthogonality constraints on $\{f_k\}$, which produce important model simplifications, and the parameter-expanded sampler for the ordered spike-and-slab prior, which substantially improves MCMC mixing (see Figure 3). We implement the following sampling steps, with details provided in the supplement: (i) impute $y_i(\tau^*)$ for any unobserved τ^*; (ii) sample the nonlinear parameter γ if unknown; (iii) sample the constrained nonparametric factors $\{f_k\}$ (see Section 2) and the smoothing parameters $\{\lambda_{f_k}\}$; (iv) sample the ordered spike-and-slab parameters (Algorithm 1); (v) sample the parametric factors $\{\alpha_{\ell,i}\}$; and (vi) sample any variance components, such as σ_ϵ^2.

The implementation incorporates a truncation of the infinite summation in (1.1), which provides simpler and faster computations. By applying Theorem 1 of Legramanti et al. (2020), we confirm that finite approximations are accurate for sufficiently large truncation K:

Proposition 4. *Let $\theta^{(K)} = \{\theta_k\}_{k=1}^K$ denote the sequence $\{\theta_k\}_{k=1}^\infty$ truncated at K. Under the CUSP* (3.1)–(3.2) *and NMIG* (3.7) *priors with $0 < v_0 < \varepsilon < 1$, we have $\mathbb{P}\left\{ d_\infty(\theta, \theta^{(K)}) > \varepsilon \right\} \leq \kappa \{\kappa/(1+\kappa)\}^K$.*

The approximation error induced by truncating $\{\theta_k\}_{k=1}^\infty$ to K terms decreases rapidly in K, which suggests that the proposed infinite-dimensional ordered spike-and-slab prior is accurately approximated by a conservative truncation.

We emphasize that the sampling steps for the ordered spike-and-slab parameters, detailed in Algorithm 1, consist of simple, fast, and closed form updates. Because of the orthogonality constraints and Corollary 1, these distributions depend on the data only through $\{\boldsymbol{f}_k'\boldsymbol{y}_i\}$. In addition, we retain posterior samples of K^* in (3.3), which is the effective number of nonparametric terms. If this posterior distribution places mass on values near K, then the truncation level K should be increased. Adaptations of this sampling algorithm for use of the ordered spike-and-slab prior with other models, such as (non-functional) factor models, are available by appropriately replacing $y_{k,i}^F$.

5 Simulation study

A simulation study is performed to assess model performance for point and interval estimation of Y_i, $\boldsymbol{y}_i$, and $\{\boldsymbol{\alpha}_i\}$ and uncertainty quantification for the number of nonparametric terms. We present results for the linear template $\mathcal{H}_0 = \mathrm{span}\{1, \tau\}$ here and include the Nelson and Siegel (1987) template with unknown γ (along with a misspecified example) in the supplementary material.

Algorithm 1: MCMC sampling steps for the ordered spike-and-slab prior.

Let $y_{k,i}^F = \boldsymbol{f}_k' \boldsymbol{y}_i$ for $i = 1, \ldots, n$ and $k = 1, \ldots, K$:

1. **Sample** $[m_{\xi_{k,i}} \mid -]$ from $\{-1, 1\}$ with $\mathbb{P}(m_{\xi_{h,i}} = 1 \mid -) = 1/\{1 + \exp(-2\xi_{h,i})\}$;

2. **Sample** $[\xi_{k,i} \mid -] \sim N(Q_{\xi_{k,i}}^{-1} \ell_{\xi_{k,i}}, Q_{\xi_{k,i}}^{-1})$ where $Q_{\xi_{k,i}} = \eta_k^2/\sigma_\epsilon^2 + 1$ and $\ell_{\xi_{k,i}} = \eta_k y_{k,i}^F/\sigma_\epsilon^2 + m_{\xi_{k,i}}$;

3. **Sample** $[\eta_k \mid -] \sim N(Q_{\eta_k}^{-1} \ell_{\eta_k}, Q_{\eta_k}^{-1})$ where $Q_{\eta_k} = \sum_{i=1}^n \xi_{k,i}^2/\sigma_\epsilon^2 + (\theta_k \sigma_k^2)^{-1}$ and $\ell_{\eta_k} = \sum_{i=1}^n \xi_{k,i} y_{k,i}^F/\sigma_\epsilon^2$;

4. **Rescale** $\eta_k \to (\sum_{i=1}^n |\xi_{k,i}|/n)\eta_k$ and $\boldsymbol{\xi}_k \to (n/\sum_{i=1}^n |\xi_{k,i}|)\boldsymbol{\xi}_k$ and **update** $\beta_{k,i} = \xi_{k,i}\eta_k$;

5. **Sample** $[\sigma_k^{-2} \mid -] \sim \text{Gamma}\left\{a_1 + 1/2, a_2 + \eta_k^2/(2\theta_k)\right\}$;

6. **Sample** $[\nu_k \mid -] \sim \text{Beta}(1 + \sum_{h=1}^K \mathbb{I}\{z_h = k\}, \kappa + \sum_{h=1}^K \mathbb{I}\{z_h > k\})$ for $k = 1, \ldots, K - 1$ and **update** π_k and ω_k from (3.2);

7. **Sample** $[\kappa \mid -] \sim \text{Gamma}\{a_\kappa + K - 1, b_\kappa - \sum_{k=1}^{K-1} \log(1 - \nu_k)\}$;

8. **Sample** $[z_k \mid -]$ from

$$\mathbb{P}(z_k = h \mid -) \propto \begin{cases} \omega_h t_{2a_1}(\eta_k; 0, \sqrt{v_0 a_2/a_1}) & h \leq k \\ \omega_h t_{2a_1}(\eta_k; 0, \sqrt{a_2/a_1}) & h > k \end{cases}$$

9. **Update** $\theta_k = 1$ if $z_k > k$ and $\theta_k = v_0$ if $z_k \leq k$.

Synthetic functional data with $n = 100$ curves and $m = 25$ equally-spaced observation points in $[0, 1]$ are generated as follows. The parametric and nonparametric factors are simulated independently as $\alpha_{\ell,i}^* \sim N(0, 1)$ and $\beta_{k,i}^* \sim N(0, 1/(k+1)^2)$, respectively, for $k = 1, \ldots, K_{true}$ and $K_{true} \in \{0, 1, 3, 8\}$; results for $K_{true} = 1$ are nearly identical to the $K_{true} = 3$ case and are omitted, while $K_{true} = 8$ is presented in the supplement. By design, the variability in the parametric factors outweighs the variability in the nonparametric factors, so the parametric model is at least partially adequate. The parametric basis matrix $\boldsymbol{G}_\gamma^*$ is constructed by evaluating $\{g_\ell\}$ at each observation point and QR-decomposing the resulting matrix as in Section 2. For the nonparametric functions f_k^*, we use orthogonal polynomials of degree $k + 1$, which are orthogonal to the linear template; the Nelson-Siegel version applies an additional orthogonalization step. The error-free latent functions are $\boldsymbol{Y}_i^* = \boldsymbol{G}_\gamma^* \boldsymbol{\alpha}_i^* + \boldsymbol{F}^* \boldsymbol{\beta}_i^*$ and the functional observations are generated as $\boldsymbol{y}_i = \boldsymbol{Y}_i^* + \sigma^* \boldsymbol{\epsilon}_i^*$ where $\sigma^* = \text{sd}(\boldsymbol{Y}_i^*)/\text{RSNR}$ for sample standard deviation $\text{sd}(\cdot)$, root signal-to-noise ratio $\text{RSNR} = 3$, and $\boldsymbol{\epsilon}_i^* \sim N(\boldsymbol{0}, \boldsymbol{I}_m)$ independently. This process was repeated to create 100 synthetic datasets.

We focus primarily on the PFFM, PFFM+GP, and SFFM. Only the PFFM with $K_{true} = 0$ is correctly specified; in the remaining cases, the nonparametric term is misspecified for both K_{true} and $\{f_k^*\}$. For an additional competitor, we modify the PFFM+GP (2.1) to use a spline basis expansion $h_i(\tau) = \boldsymbol{b}'(\tau)\boldsymbol{\chi}_i$, specifically with $\boldsymbol{b}$ from Section 2.2 and $[\chi_{i,r} \mid \sigma_\chi] \sim N(0, \sigma_\chi^2), \sigma_\chi \sim C^+(0, 1)$. To further distinguish this model (PFFM+spline) from PFFM+GP, we orthogonalize the spline basis to the parametric template $\boldsymbol{G}_\gamma$. In effect, PFFM+spline replaces $\{f_k\}$ in the SFFM with $\boldsymbol{b}$,

and thus isolates the importance of the low-rank nonparametric functional data model. However, PFFM+spline does not provide rank selection or inference for K^*, and thus cannot directly assess the necessity of the nonparametric factors.

For all models, we assume the conditionally Gaussian likelihood (2.3) with the hierarchical priors

$$[\alpha_{\ell,i} \mid \sigma_{\alpha_\ell}] \overset{indep}{\sim} N(0, \sigma_{\alpha_\ell}^2), \quad \sigma_{\alpha_\ell} \overset{iid}{\sim} C^+(0,1), \quad p(\sigma_\epsilon^2) \propto 1/\sigma_\epsilon^2, \tag{5.1}$$

and the SFFM hyperparameters $a_1 = 5$, $a_2 = 25$, and $v_0 = 0.001$, with an upper bound $K = 10$ on the number of nonparametric factors. Sensitivity analyses were conducted for $(a_1, a_2) \in \{(5, 25), (5, 50), (10, 30)\}$ and $v_0 \in \{0.01, 0.005, 0.00025\}$. The results for point and interval predictions and estimates of $\boldsymbol{Y}_i^*$ and $\{\alpha_{\ell,i}^*\}$ are highly robust to these hyperparameters. Inference on K^* is typically robust with the exception of $v_0 = 0.00025$, for which the posterior of K^* becomes more sensitive to (a_1, a_2). Hence, we select a larger value of v_0.

First, we evaluate point prediction of the latent curves $\boldsymbol{Y}_i^*$ and point estimation of the parametric factors $\{\alpha_{\ell,i}^*\}$. Posterior expectations from each model are assessed using root mean squared (prediction) error, and presented in Figure 4. Although the methods perform similarly when the parametric model is true ($K_{true} = 0$), the SFFM offers substantial improvements in the semiparametric setting ($K_{true} > 0$). Notably, when $K_{true} > 0$, the PFFM+GP offers better prediction than the PFFM—as expected—but conversely cannot estimate the parametric factors accurately. Comparatively, the PFFM+spline improves upon PFFM+GP for coefficient estimation—due to the orthogonality constraint—yet underperforms for point prediction.

Next, we assess prediction and credible intervals for $\boldsymbol{y}_i^*$ and $\{\alpha_{\ell,i}^*\}$, respectively, where $\boldsymbol{y}_i^*$ is distributed identically to $\boldsymbol{y}_i$. The 95% interval estimates are evaluated using mean interval width and empirical coverage, which are presented in Figure 5. Most notably, the SFFM provides uniformly narrower interval estimates—and therefore more precise uncertainty quantification—with approximately the correct nominal coverage. The PFFM prediction and credible interval estimates are excessively wide when $K_{true} > 0$, which is reflected in Figure 1 and subsequently in Figures 7–8. The PFFM+GP interval estimates for $\{\alpha_{\ell,i}^*\}$ are quite poor, especially for $K_{true} > 0$, where the interval estimates are excessively wide yet still fail to contain the true parameters $\{\alpha_{\ell,i}^*\}$ at close to the nominal level. In conjunction with Figure 2, these results show that functional confounding adversely impacts both point estimation and uncertainty quantification.

Lastly, we study the posterior distribution of K^* for assessing the necessity of the nonparametric terms. Since the competing methods are not competitors for rank selection, we include a finite mixture model variant of the SFFM (SFFM-fmm) that instead assumes $\omega_{1:K} \sim \text{Dirichlet}(\kappa/K, \ldots, \kappa/K)$. We use the same choice of K as in the SFFM and set the prior expected number of factors to be $\kappa = 1$ to encourage fewer factors. Figure 6 compares the quality of the posterior distribution $\mathbb{P}(K^* \mid \boldsymbol{y})$ for the SFFM and this finite mixture alternative. Notably, the ordered spike-and-slab prior results in much larger probabilities on the true rank K_{true}. Despite fixing $\kappa = 1$, the SFFM-fmm repeatedly overestimates the ranks: the simulation average of $\mathbb{P}(K^* > K_{true} \mid \boldsymbol{y})$ is

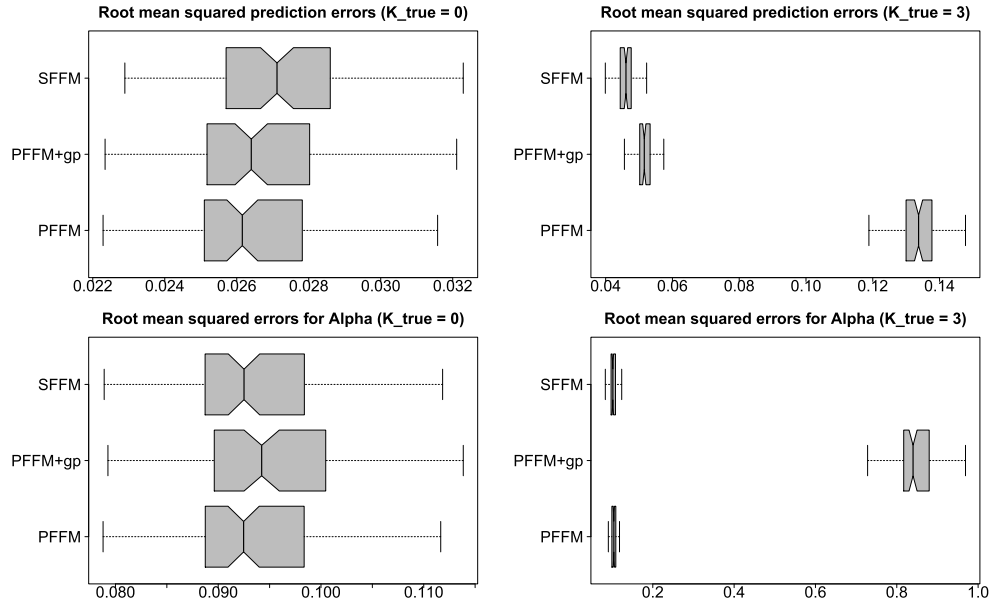Figure 4: Root mean squared (prediction) errors for $\boldsymbol{Y}^*$ (top) and $\{\alpha^*_{\ell,i}\}$ (bottom) for $K_{true} = 0$ (left) and $K_{true} = 3$ (right). All methods perform similarly for $K_{true} = 0$, while the SFFM outperforms all competitors when $K_{true} > 0$. In this setting, the PFFM+GP offers better prediction than the PFFM, but cannot accurately estimate the parametric factors.

0.30 for $K_{true} = 0$, 0.23 for $K_{true} = 3$, and 0.22 for $K_{true} = 8$, while the comparable values for the SFFM are 0.15, 0.16, and 0.15, respectively. Hence, this finite mixture alternative introduces unnecessary parameters and relays deceptively strong evidence for additional nonparametric terms.

Additional simulation results in the supplement include several additional examples, including the Nelson and Siegel (1987) template with unknown γ; cases of misspecification of the parametric template; cases with sparsely-observed functional data; cases with larger number of nonparametric factors ($K_{true} = 8$); and further assessments of $\mathbb{P}(K^* \mid \boldsymbol{y})$, including ranked probability scores and nonparametric detection via $\mathbb{P}(K^* > 0 \mid \boldsymbol{y})$. These results broadly confirm the qualitative results presented here.

6 Applications

6.1 Pinch force data

Human motor control is a critical area of research with implications for human physiology, monitoring and mitigating muscle degeneration, and designing robotic devices. Motor control data are recorded at high resolutions and often modeled as functional
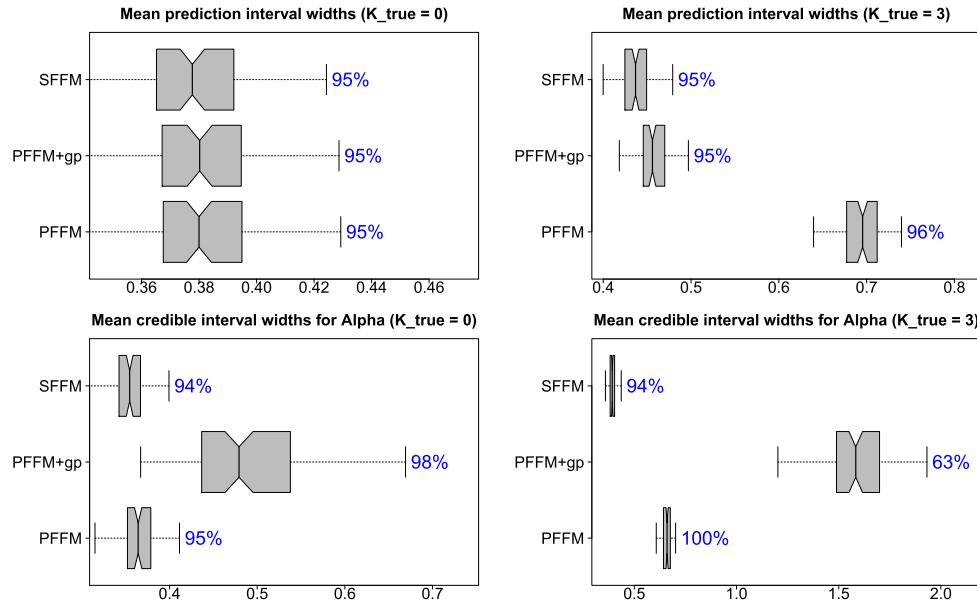
Figure 5: Mean 95% prediction (top) and credible (bottom) interval widths for $\boldsymbol{y}_i^*$ and $\{\alpha_{\ell,i}^*\}$, respectively, with empirical coverage (annotations) for $K_{true} = 0$ (left) and $K_{true} = 3$ (right). The SFFM provides narrow interval estimates with (nearly) the correct nominal coverage in all cases. The PFFM intervals are excessively wide for $K_{true} > 0$, while the PFFM+GP performs poorly for $\{\alpha_{\ell,i}^*\}$.
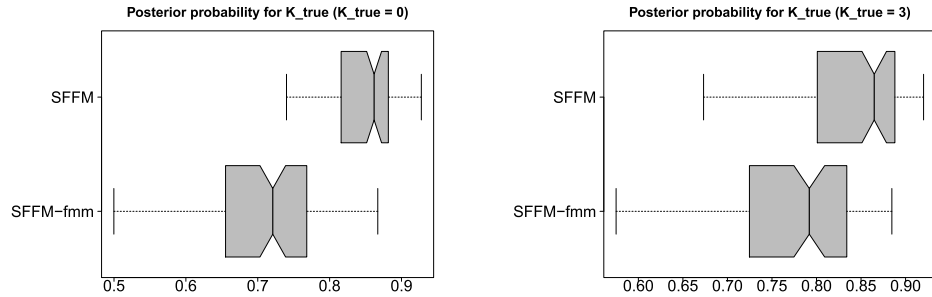

Figure 6: Probability score $\mathbb{P}(K^* = K_{true} \mid \boldsymbol{y})$ for $K_{true} = 0$ (left) and $K_{true} = 3$ (right). The proposed ordered spike-and-slab prior provides better rank selection and inference than the finite mixture alternative.

data (Ramsay, 2000; Goldsmith and Kitago, 2016). Human physiology and the laws of motion dictate a parametric template, which is crucial for understanding the underlying process. However, these parametric models may not fully describe the observed data, which undermines the interpretability of the key parameters.

We analyze human pinching data from Ramsay et al. (1995), which reports the force measured by pinching the thumb and forefinger on opposing sides of a 6cm force meter. The subject was instructed to maintain a background level of constant force, then increase the pinching force to a predetermined maximum level, and finally return to the original background level of constant force. We use data from the `fda` package in R, which consists of $n = 20$ replicate force curves over time each with $m = 151$ observations selected such that the maximum of each curve occurred at 0.076 seconds. An example curve is in Figure 1.

To model the pinch force over time τ, we adapt a parametric model from Ramsay et al. (1995):

$$g_1(\tau; \gamma_i) = 1, \quad g_2(\tau; \gamma_i) = \exp[-(\log \tau - c_i)^2 / \{2 \exp(\gamma_i)\}], \tag{6.1}$$

where $\exp(c_i)$ is the time of the maximum force and $\gamma_i \in \mathbb{R}$ is a shape parameter. Ramsay et al. (1995) argue that the unnormalized log-normal density g_2 matches the shape of the observed data and offers plausible scientific explanations. For computational convenience, we estimate c_i as in Ramsay et al. (1995) by fitting a quadratic regression in $\log(\tau)$ for the response variable $\log(\boldsymbol{y}_i)$ restricted to observations $y_i(\tau) > 0.5$. The estimate of each c_i can be recovered from the estimated regression coefficients and is subsequently treated as fixed.

We fit the PFFM, PFFM+GP, and SFFM to the data using the template (6.1). For partial pooling among subjects, we specify a hierarchical prior on the shape parameters:

$$[\gamma_i \mid \mu_\gamma, \sigma_\gamma] \overset{iid}{\sim} N(\mu_\gamma, \sigma_\gamma^2), \quad \mu_\gamma \sim N(0, 10), \quad \sigma_\gamma \sim C^+(0, 1)$$

with the priors from (5.1) on the remaining parameters. Since the nonlinear parameters γ_i are curve-specific, we construct each $\boldsymbol{G}_{\gamma_i}$ for $i = 1, \ldots, n$ using a QR decomposition (Section 2) and modify the orthogonality constraints for the nonparametric basis $\{f_k\}$:

$$n^{-1} \sum_{i=1}^{n} \boldsymbol{G}_{\gamma_i}' \boldsymbol{f}_k = \boldsymbol{0}_L, \quad k = 1, 2, \ldots,$$

which enforces orthogonality on average across the subject-specific parametric basis matrices. This constraint no longer preserves the posterior factorization (Lemma 1, Corollary 1), but still maintains distinctness between the parametric and nonparametric terms. The SFFM hyperparameters are set to $a_1 = 5$, $a_2 = 25$, $v_0 = 0.001$, and $K = 10$ as in Section 5.

An example of the fitted values with 95% simultaneous prediction bands for the PFFM and SFFM is in Figure 1. Although the PFFM captures the general shape of the data, it suffers from clear bias around the peak and produces unnecessarily wide prediction bands. The SFFM corrects both issues: the bias is removed and the prediction bands are more precise. Notably, the fitted SFFM curve preserves the same general shape as the PFFM and avoids overfitting despite the increase in modeling complexity.

Posterior uncertainty quantification is also more precise under the SFFM for the parametric factors $\{\alpha_{\ell,i}\}$. Figure 7 presents the posterior standard deviations of $\{\alpha_{\ell,i}\}_{i=1}^{n}$

for each $\ell = 1, \ldots, L$ under the PFFM and SFFM. Although the posterior expectations are similar (see Figure 2) and the prior variances for $\{\alpha_{\ell,i}\}_{i=1}^n$ are the same under the PFFM and SFFM, the posterior standard deviations are substantially smaller under the SFFM. This reduction is even more dramatic compared to the PFFM+GP, which is omitted from Figure 7 due to excessively large posterior standard deviations that range from 1.42 to 4.43. For model (6.1), the linear coefficients $\{\alpha_{\ell,i}\}$ determine the maximum of the force curve, which is the most prominent feature in the data. The SFFM provides more precise posterior inference for these key parametric factors.
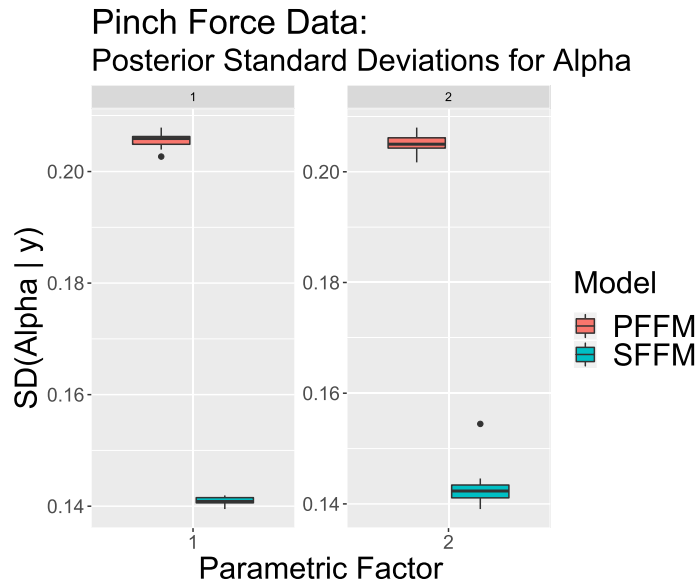

Figure 7: Posterior standard deviations of $\{\alpha_{\ell,i}\}_{i=1}^n$ for $\ell = 1, \ldots, L$ in the parametric (PFFM) and semiparametric functional factor model (SFFM) for the pinch force data. The SFFM consistently reduces posterior standard deviations for the parametric factors by about 30%.

Additional evidence in favor of the SFFM is presented in Table 2, which estimates the posterior distribution of the effective number of nonparametric terms K^*. Clearly, there is substantially posterior probability for at least two nonparametric terms, with limited evidence that more than three terms are needed.

k	0	1	2	3	4	5	≥ 6
$\mathbb{P}(K^* = k \mid \boldsymbol{y})$	0.003	0.147	0.707	0.135	0.007	0.001	0

Table 2: Posterior probabilities $\mathbb{P}(K^* = k \mid \boldsymbol{y})$ for the pinch force data. There is strong evidence for a nonparametric component.

Finally, we highlight both the model performance and the MCMC efficiency in Table 3. Watanabe-Akaike/widely-applicable information criteria (WAIC; Watanabe,

2010) is a comprehensive metric for Bayesian model fitness, and clearly favors the SFFM over competing methods. For all models, we retain 10000 MCMC samples after discarding 5000 iterations as a burn-in. Traceplots of $\{Y_i\}$, $\{\alpha_{\ell,i}\}$, and K^* (not shown) demonstrate excellent mixing and suggest convergence. The SFFM rivals the simpler PFFM in computing time and MCMC efficiency (measured by effective sample size), while the PFFM+GP lags far behind in both.

Metric	PFFM	PFFM+GP	SFFM
WAIC	-922	-2693	-3005
Time (s) / 1000 iterations	1.59	4.35	1.98
Median ESS of $\{\alpha_{\ell,i}\}$	10000	2445	10000

Table 3: WAIC (negatively oriented), computing time (using `R` on a MacBook Pro, 2.8 GHz Intel Core i7), and median effective sample size (ESS) of $\{\alpha_{\ell,i}\}$ for the pinch force data. The SFFM delivers a superior model fit and excellent computational performance.

6.2   Dynamic yield curves

Yield curves are fundamental in economic and financial analyses: they provide essential information about current and future economic conditions, including inflation, business cycles, and monetary policies, and are used to price fixed-income securities and construct forward curves. The yield curve $Y_i(\tau)$ describes how interest rates vary as a function of the length of the borrowing period, or time to maturity τ, at each time i. Naturally, yield curves can be modeled as functional data that evolve dynamically over time.

Yield curve models most commonly employ the Nelson and Siegel (1987) basis:

$$g_1(\tau;\gamma) = 1, \quad g_2(\tau;\gamma) = \{1 - \exp(-\tau\gamma)\}/(\tau\gamma), \quad g_3(\tau;\gamma) = g_2(\tau;\gamma) - \exp(-\tau\gamma) \quad (6.2)$$

where g_1 is the *level*, g_2 is the *slope*, and g_3 is the *curvature*. The nonlinear parameter $\gamma > 0$ is commonly treated as fixed, such as $\gamma = 0.0609$ (Diebold and Li, 2006), but otherwise may be estimated. We use a weakly informative Gamma prior for γ with prior mean 0.0609 and prior variance 0.5. Both the PFFM and the SFFM with this prior on γ are favored by WAIC over their respective counterparts with fixed $\gamma = 0.0609$.

To capture the yield curve dynamics, we model the parametric Nelson-Siegel factors as an AR(1):

$$\alpha_{\ell,i} = \mu_\ell + \phi_\ell(\alpha_{\ell,i-1} - \mu_\ell) + \zeta_{\ell,i}, \quad \zeta_{\ell,i} \overset{indep}{\sim} N(0, \sigma_{\zeta_\ell}^2). \quad (6.3)$$

The PFFM with (6.2) and (6.3) is also known as the *dynamic Nelson-Siegel model* (Diebold and Li, 2006). The dynamic Nelson-Siegel factors $\{\alpha_{\ell,i}\}$ may be viewed as the state variables in a dynamic linear model. Using the orthogonality constraints on $\{g_\ell\}$ and $\{f_k\}$, we derive a convenient and low-dimensional state simulation algorithm for efficient joint sampling of the dynamic factors $\{\alpha_{\ell,i}\}$ (see the supplementary material). Similar simplifications are *not* available for the PFFM+GP, which is omitted.

The dynamic model (6.3) is accompanied by the priors

$$\mu_\ell \overset{iid}{\sim} N(0, 10^6), \quad (\phi_\ell + 1)/2 \overset{iid}{\sim} \mathrm{Beta}(5,2), \quad \sigma_{\zeta_\ell} \overset{iid}{\sim} C^+(0,1), \quad \ell = 1, \ldots, L,$$

where the prior on $\{\phi_\ell\}$ ensures stationarity of the dynamic factors $\{\alpha_{\ell,i}\}$ and therefore $\{Y_i\}$ and $\{\boldsymbol{y}_i\}$. In addition, we generalize (1.2) to accommodate stochastic volatility in the observation error variance,

$$\epsilon_{i,j} \overset{indep}{\sim} N(0, \sigma_{\epsilon_i}^2), \quad \log \sigma_{\epsilon_i}^2 \sim \mathrm{AR}(1),$$

which is an essential component in many economic and financial models (Kim et al., 1998). For these AR(1) parameters, we adopt the priors and sampling algorithm from Kowal (2021).

We evaluate the suitability of the Nelson-Siegel model for monthly unsmoothed Fama-Bliss US government bond yields (Dijk et al., 2014). These data are available from 2000-2009 ($n = 120$) for maturities of 3, 6, 9, 12, 15, 18, 21, 24, 30, 36, 48, 60, 72, 84, 96, 108 and 120 months ($m = 17$). The inferential targets are the latent curves Y_i, the dynamic Nelson-Siegel factors $\{\alpha_{i,\ell}\}$, and the effective number of nonparametric terms K^*. The SFFM hyperparameter and MCMC specifications from Section 6.1 are adopted here, again with excellent mixing and convergence for these key parameters.

Figure 1 shows the PFFM and SFFM fitted values and 95% simultaneous prediction bands for the yield curve in September 2008 during the onset of the Great Recession. As with the pinch force data, the PFFM produces a reasonable shape, yet suffers from clear bias and overconservative prediction bands. The SFFM corrects these deficiencies without distorting the general shape of the curve or overfitting to the data.

Similar results are obtained in Figure 8 for the dynamic Nelson-Siegel factors. The SFFM offers substantial reductions in posterior standard deviation for all three dynamic Nelson-Siegel factors. These factors are fundamental for describing the shape of the yield curve; reducing the posterior uncertainty is a crucial advantage of the SFFM. Importantly, the simulation study confirms that the reduced posterior uncertainty quantification from the SFFM nonetheless retains valid calibration, or more specifically, correct nominal coverage of the posterior credible intervals.

We summarize the posterior distribution $\mathbb{P}(K^* \mid \boldsymbol{y})$ of the effective number of nonparametric terms in Table 4. The evidence for the nonparametric factors is moderate: we estimate $\mathbb{P}(K^* > 0 \mid \boldsymbol{y}) = 0.09$.

k	0	1	2	≥ 3
$\mathbb{P}(K^* = k \mid \boldsymbol{y})$	0.910	0.087	0.003	0

Table 4: Posterior probabilities $\mathbb{P}(K^* = k \mid \boldsymbol{y})$ for the yield curve data. There is moderate evidence for a nonparametric component.

There are compelling reasons to include the nonparametric factors in a dynamic Nelson-Siegel model. First, the uncertainty quantification for $\{Y_i\}$ and $\{\alpha_{\ell,i}\}$ is more precise in the SFFM (see Figures 1 and 8). Second, WAIC decisively prefers the SFFM
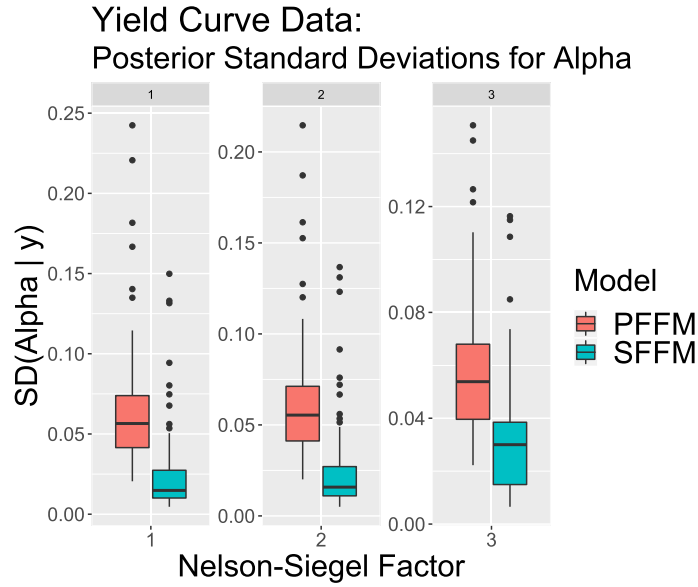
Figure 8: Posterior standard deviations of $\{\alpha_{\ell,i}\}_{i=1}^{n}$ for $\ell = 1, \ldots, L$ in the parametric (PFFM) and semiparametric functional factor model (SFFM) for the yield curve data. Despite only moderate evidence for a nonparametric component, the SFFM provides about a 60% reduction in posterior standard deviation for these key parametric factors.

over the PFFM, which suggests potential improvements in out-of-sample predictive capabilities (see Table 5). Third, the additional computational burdens of the SFFM are minimal. Due to the orthogonality constraints, the sampling steps for $\{\alpha_{\ell,i}\}$ and associated parameters are identical for the PFFM and the SFFM, which permits full and robust model development for the parametric factors. Table 5 showcases the comparable computational performance. Although the SFFM sacrifices some MCMC efficiency, the effective sample sizes nonetheless remain quite substantial for a model of this complexity—including dynamic factors, unknown nonlinear terms, stochastic volatility in the error variance, and an unknown nonparametric functional factor structure.

Metric	PFFM	SFFM
WAIC	-5516	-10229
Time (s) / 1000 iterations	21	23
Median ESS of $\{\alpha_{\ell,i}\}$	10000	6788

Table 5: WAIC (negatively oriented), computing time (using R on a MacBook Pro, 2.8 GHz Intel Core i7), and median effective sample size (ESS) of $\{\alpha_{\ell,i}\}$ for the yield curve data. The SFFM delivers a superior model fit and excellent computational performance.

7 Discussion

We proposed a Bayesian semiparametric model for functional data. The semiparametric functional factor model (SFFM) augmented a parametric template with an infinite-dimensional nonparametric functional basis. The nonparametric basis was treated as unknown and learned from the data to correct the biases of the parametric template while appropriately incorporating relevant uncertainties into the posterior distribution. Distinctness between the parametric and nonparametric terms was achieved by conditioning upon an orthogonality constraint, which simultaneously prevented *functional confounding* and admitted highly convenient simplifications for efficient MCMC sampling. The nonparametric component was regularized with an ordered spike-and-slab prior that built upon CUSP (Legramanti et al., 2020), which implicitly provided rank selection for infinite-dimensional models and satisfied appealing theoretical properties. This prior was accompanied by a parameter expansion scheme customized to boost MCMC efficiency, and is broadly applicable for Bayesian factor models. Our analyses of synthetic data, human motor control data, and dynamic interest rates demonstrated clear advantages of the semiparametric modeling framework relative to both parametric and Gaussian process alternatives. The proposed approach eliminated bias, reduced excessive posterior and predictive uncertainty, and provided reliable inference on the effective number of nonparametric terms—all with minimal computational costs.

There are several promising extensions that remain for future work. First, formulation of the parametric template in (1.1) can be generalized, for example to include a functional regression term with scalar or functional covariates. Second, the ordered spike-and-slab prior currently uses independent and identically distributed variables in the parameter expansion (3.6). Adaptations to include dependence among these variables, such as regression models, clustering, and spatio-temporal dependence, would broaden the applicability of the prior. Lastly, we studied only a small subset of many possible parametric templates. For applications that rely on such parametric models, the proposed semiparametric modeling framework can directly assess the adequacy of these models—and perhaps suggest improvements.

Supplementary Material

Proofs, Computing, and Simulations (DOI: 10.1214/23-BA1410SUPPA; .pdf). The supplementary PDF includes proofs of all results, details on the MCMC algorithms, and additional simulation results.

Code (DOI: 10.1214/23-BA1410SUPPB; .zip). R code to reproduce the analysis.

References

Boyd, S. P. and Vandenberghe, L. (2004). *Convex Optimization*. Cambridge University Press. MR2061575. doi: https://doi.org/10.1017/CBO9780511804441. 1169

Canale, A., Lijoi, A., Nipoti, B., and Prünster, I. (2017). "On the Pitman-Yor process with spike and slab base measure." *Biometrika*, 104: 681–697. MR3694590. doi: https://doi.org/10.1093/biomet/asx041. 1162

Castillo, I. and van der Vaart, A. (2012). "Needles and straw in a haystack: Posterior concentration for possibly sparse sequences." *The Annals of Statistics*, 40: 2069–2101. MR3059077. doi: https://doi.org/10.1214/12-AOS1029. 1174

Chen, S. T., Xiao, L., and Staicu, A.-M. (2019). "Model testing for generalized scalar-on-function linear models." *arXiv preprint arXiv:1906.04889*. 1165

Diebold, F. X. and Li, C. (2006). "Forecasting the term structure of government bond yields." *Journal of Econometrics*, 130: 337–364. MR2211798. doi: https://doi.org/10.1016/j.jeconom.2005.03.005. 1162, 1182

Dijk, D. V., Koopman, S. J., der Wel, M., and Wright, J. H. (2014). "Forecasting interest rates with shifting endpoints." *Journal of Applied Econometrics*, 29: 693–712. MR3258059. doi: https://doi.org/10.1002/jae.2358. 1183

Ferguson, T. S. (1973). "A Bayesian analysis of some nonparametric problems." *The Annals of Statistics*, 209–230. MR0350949. 1173

Frühwirth-Schnatter, S. (2023). "Generalized cumulative shrinkage process priors with applications to sparse Bayesian factor analysis." *Philosophical Transactions of the Royal Society A*, 381: 20220148. MR4590506. 1164

Gelman, A. (2006). "Prior distributions for variance parameters in hierarchical models." *Bayesian Analysis*, 1: 515–534. MR2221284. doi: https://doi.org/10.1214/06-BA117A. 1172

George, E. I. and McCulloch, R. E. (1993). "Variable selection via Gibbs sampling." *Journal of the American Statistical Association*, 88: 881–889. 1172

Goldsmith, J. and Kitago, T. (2016). "Assessing systematic effects of stroke on motor control by using hierarchical function-on-scalar regression." *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 65: 215–236. MR3456686. doi: https://doi.org/10.1111/rssc.12115. 1162, 1179

Griffiths, T. L. and Ghahramani, Z. (2011). "The Indian Buffet Process: An introduction and review." *Journal of Machine Learning Research*, 12. MR2804598. 1173

Heckman, N. E. and Ramsay, J. O. (2000). "Penalized regression with model-based penalties." *Canadian Journal of Statistics*, 28: 241–258. MR1792049. doi: https://doi.org/10.2307/3315976. 1165

Ishwaran, H. and Rao, J. S. (2005). "Spike and slab variable selection: Frequentist and Bayesian strategies." *The Annals of Statistics*, 33: 730–773. MR2163158. doi: https://doi.org/10.1214/009053604000001147. 1172

Kim, S., Shephard, N., and Chib, S. (1998). "Stochastic volatility: Likelihood inference and comparison with ARCH models." *Review of Economic Studies*, 65: 361–393. 1183

Kowal, D. R. (2021). "Dynamic regression models for time-ordered functional data." *Bayesian Analysis*, 16(2): 459–487. MR4255337. doi: https://doi.org/10.1214/20-ba1213. 1169, 1183

Kowal, D. R., Matteson, D. S., and Ruppert, D. (2019). "Functional autoregression for sparsely sampled data." *Journal of Business and Economic Statistics*, 37. MR3910228. doi: https://doi.org/10.1080/07350015.2017.1279058. 1172

Kowal, D. R. and Canale, A. (2023a). "Supplementary Material for "Semiparametric functional factor models with Bayesian rank selection"." *Bayesian Analysis*. doi: https://doi.org/10.1214/23-BA1410SUPPA. 1165

Kowal, D. R. and Canale, A. (2023b). "Supplementary Material for "Semiparametric functional factor models with Bayesian rank selection"." *Bayesian Analysis*. doi: https://doi.org/10.1214/23-BA1410SUPPB. 1165

Lee, W., Miranda, M. F., Rausch, P., Baladandayuthapani, V., Fazio, M., Downs, J. C., and Morris, J. S. (2018). "Bayesian semiparametric functional mixed models for serially correlated functional data, with application to glaucoma data." *Journal of the American Statistical Association*. MR3963158. doi: https://doi.org/10.1080/01621459.2018.1476242. 1165

Legramanti, S., Durante, D., and Dunson, D. B. (2020). "Bayesian cumulative shrinkage for infinite factorizations." *Biometrika*, 107: 745–752. MR4138988. doi: https://doi.org/10.1093/biomet/asaa008. 1164, 1170, 1171, 1175, 1185

Mikulich, S. K., Zerbe, G. O., Jones, R. H., and Crowley, T. J. (2003). "Comparing linear and nonlinear mixed model approaches to cosinor analysis." *Statistics in Medicine*, 22: 3195–3211. 1162

Molenberghs, G. and Verbeke, G. (2000). *Linear Mixed Models for Longitudinal Data*. Springer. MR1880596. doi: https://doi.org/10.1007/978-1-4419-0300-6. 1162

Nelson, C. R. and Siegel, A. F. (1987). "Parsimonious modeling of yield curves." *The Journal of Business*, 60: 473. 1162, 1175, 1178, 1182

O'Hara, R. B. and Sillanpää, M. J. (2009). "A review of Bayesian variable selection methods: What, how and which." *Bayesian Analysis*, 4: 85–117. MR2486240. doi: https://doi.org/10.1214/09-BA403. 1172

Ohn, I. and Kim, Y. (2021). "Posterior consistency of factor dimensionality in high-dimensional sparse factor models." *Bayesian Analysis*, 1: 1–24. MR4483228. doi: https://doi.org/10.1214/21-ba1261. 1173

Rai, P. and III, H. D. (2009). "The infinite hierarchical factor regression model." Curran Associates, Inc. 1173

Ramsay, J. O. (2000). "Functional components of variation in handwriting." *Journal of the American Statistical Association*, 95: 9–15. 1162, 1179

Ramsay, J. O. and Dalzell, C. J. (1991). "Some tools for functional data analysis." *Journal of the Royal Statistical Society: Series B (Methodological)*, 53: 539–561. MR1125714. 1165

Ramsay, J. O., Wang, X., and Flanagan, R. (1995). "A functional data analysis of the pinch force of human fingers." *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 44: 17–30. MR1616045. doi: https://doi.org/10.1111/1467-9868.00129. 1180

Reich, B. J., Hodges, J. S., and Zadnik, V. (2006). "Effects of residual smoothing on the posterior of the fixed effects in disease-mapping models." *Biometrics*, 62: 1197–1206. MR2307445. doi: https://doi.org/10.1111/j.1541-0420.2006.00617.x. 1166

Rockova, V. (2018). "Bayesian estimation of sparse signals with a continuous spike and slab prior." *The Annals of Statistics*, 46: 401–437. MR3766957. doi: https://doi.org/10.1214/17-AOS1554. 1174

Rossell, D. and Rubio, F. J. (2023). "Additive Bayesian variable selection under censoring and misspecification." *Statistical Science*, 38: 13–29. MR4534642. doi: https://doi.org/10.1214/21-sts846. 1164

Ročková, V. and George, E. I. (2016). "Fast Bayesian factor analysis via automatic rotations to sparsity." *Journal of the American Statistical Association*, 111: 1608–1622. MR3601721. doi: https://doi.org/10.1080/01621459.2015.1100620. 1173

Sang, P., Wang, L., and Cao, J. (2017). "Parametric functional principal component analysis." *Biometrics*, 73: 802–810. MR3713114. doi: https://doi.org/10.1111/biom.12641. 1165

Scarpa, B. and Dunson, D. B. (2009). "Bayesian hierarchical functional data analysis via contaminated informative priors." *Biometrics*, 65: 772–780. MR2649850. doi: https://doi.org/10.1111/j.1541-0420.2008.01163.x. 1162, 1165

Scarpa, B. and Dunson, D. B. (2014). "Enriched stick-breaking processes for functional data." *Journal of the American Statistical Association*, 109: 647–660. MR3223740. doi: https://doi.org/10.1080/01621459.2013.866564. 1162, 1165

Scheipl, F., Fahrmeir, L., and Kneib, T. (2012). "Spike-and-slab priors for function selection in structured additive regression models." *Journal of the American Statistical Association*, 107: 1518–1532. MR3036413. doi: https://doi.org/10.1080/01621459.2012.737742. 1171, 1172

Schiavon, L., Canale, A., and Dunson, D. B. (2022). "Generalized infinite factorization models." *Biometrika*, 109: 817–835. MR4472850. doi: https://doi.org/10.1093/biomet/asab056. 1164

Shang, H. L. (2014). "A survey of functional principal component analysis." *AStA Advances in Statistical Analysis*, 98: 121–142. MR3254025. doi: https://doi.org/10.1007/s10182-013-0213-1. 1164

Teh, Y. W., Grür, D., and Ghahramani, Z. (2007). "Stick-breaking construction for the Indian buffet process." In *International Conference on Artificial Intelligence and Statistics*, 556–563. Proceedings of Machine Learning Research. 1173

Wagner, A. B., Hill, E. L., Ryan, S. E., Sun, Z., Deng, G., Bhadane, S., Martinez, V. H., Wu, P., Li, D., Anand, A., Acharya, J., and Matteson, D. S. (2020). "Social distancing

has merely stabilized COVID-19 in the US." *Stat*, e302. MR4193409. doi: https://doi.org/10.1002/sta4.302. 1162

Watanabe, S. (2010). "Asymptotic equivalence of Bayes cross validation and widely applicable information criterion in singular learning theory." *Journal of Machine Learning Research*, 11: 3571–3594. MR2756194. 1181

Welham, S. J., Cullis, B. R., Kenward, M. G., and Thompson, R. (2006). "The analysis of longitudinal data using mixed model L-splines." *Biometrics*, 62: 392–401. MR2236846. doi: https://doi.org/10.1111/j.1541-0420.2005.00500.x. 1162

Acknowledgments