Supplementary Material for "Additive Function-on-Function Regression"

This document contains six sections of supplementary material. Section A discusses extension of our method to various realistic scenarios. Section B illustrates the transformations of the functional covariates required by our estimation procedure. Section C details the settings of our simulation experiment, describes the evaluation criteria, and discusses the fitting of these methods in practice, which includes the selection of the tuning parameters for our method, as well as for the competitive approaches. Additional simulation results are presented in Section D. Section E includes additional investigation for the capital bike share study. Section F discusses another data application, the yield curves study.

A Methodology Extensions

The methodology presented in Section 2 of the main paper is based on the assumption that both the response and the covariate are observed on a fine and regular grid of points and that covariates are observed without noise. However, extensions to diverse realistic scenarios are possible as we describe now.

(i) We first consider the case of functional covariates observed on a fine and regular grids of points but with error; i.e., $W_{ik} = X_i(s_k) + \delta_{ik}$, where the random deviations δ_{ik} are independent with a common variance $\tau^2 > 0$. In functional data analysis (FDA), various smoothing techniques have been applied to remove the observational errors δ_{ik}; see for example Ramsay and Silverman (2005) and the references therein. Zhang and Chen (2007) proposed approximating the true latent process $X_i(\cdot)$ by smoothing each noisy trajectory using local polynomial kernel smoothing, and they proved that the estimated curves, say $\widehat{X}_i(\cdot)$, are asymptotically identical to the true latent process. The mean and the covariance function of the covariate $X_i(s_k)$ are then estimated by their sample estimators.

(ii) If the covariates are observed on a sparse and irregular grid of points with errors, i.e., $W_{ik} = X_i(s_{ik}) + \delta_{ik}$ and $\tau^2 > 0$, we estimate the underlying smooth curves, say $\widehat{X}_i(\cdot)$, by employing functional principal component analysis (FPCA) techniques for sparsely sampled functional data (Yao et al. 2005). Their method uses local linear smoothers to estimate the mean and covariance functions of the covariate $X_i(s_k)$, and then estimates the pairs of eigenvalues/eigenfunctions by representing the random curves in a mixed model framework. They predict FPC scores using conditional expectations and predict the latent trajectories using finite Karhunen-Loève (KL) expansions. If the new covariate $X_0(s)$ is observed on a sparse design, we first estimate eigenfunctions from the original observed data assuming that $X_0(s)$ follows the same process as $X_i(s)$, and then predict the FPC scores for the new covariates via conditional expectation formula in Yao et al. (2005).

(iii) Finally, when the sampling design of the response is sparse, we follow the same approach used for the sparsely observed covariates, and then estimate the smooth underlying curves for the response.

When $Y_i(\cdot)$ are observed on a sparse and irregular grid of points, $\{t_{i1}, \ldots, t_{im_{Y,i}}\}$, numerical integration may not provide an accurate approximate of the ξ_{ik}'s. In this case, we use best linear unbiased predictors (BLUPs) proposed by Yao et al. (2005). For completeness, we review some of the results presented in Yao et al. (2005). Let $\mathbb{Y}_i = [Y_i(t_{i1}), \ldots, Y_i(t_{im_{Y,i}})]^T$, and let $\Sigma_{Y,i} = [\text{cov}\{Y_i(t_{ij}), Y_i(t_{ij'})\}]_{1 \le j,j' \le m_{Y,i}}$ be the covariance matrix evaluated corresponding to the observed time points t_{ij} and $t_{ij'}$. For $i = 1, \ldots, n$, define the $m_{Y,i}$-dimensional vector $\widetilde{\phi}_{ik} = [\widetilde{\phi}_k(t_{i1}), \ldots, \widetilde{\phi}_k(t_{im_{Y,i}})]^T$, where $\widetilde{\phi}_k(\cdot)$ is the estimated eigenfunctions obtained from the spectral decomposition of the estimated covariance matrix of response. The BLUPs of ξ_{ik} can be found as $\widetilde{\xi}_{ik} = \widetilde{\sigma}_k^2 \widetilde{\phi}_{ik}^T \widetilde{\Sigma}_{Y,i}^{-1} \mathbb{Y}_i$, where $\widetilde{\sigma}_k^2$ and $\widetilde{\Sigma}_{Y,i}$ are the eigenvalues and the covariance matrix of marginal response estimated using the entire data set. Then, it follows that $\widetilde{\nu}_{kk} = \widehat{\text{var}}(\widetilde{\xi}_{ik}) = \widetilde{\sigma}_k^2 \widetilde{\phi}_{ik}^T \widetilde{\Sigma}_{Y,i}^{-1} \widetilde{\phi}_{ik} \widetilde{\sigma}_k^2$ and $\widetilde{\nu}_{kk'} = \widehat{\text{cov}}(\widetilde{\xi}_{ik}, \widetilde{\xi}_{ik'}) = \widetilde{\sigma}_k^2 \widetilde{\phi}_{ik}^T \widetilde{\Sigma}_{Y,i}^{-1} \widetilde{\phi}_{ik'} \widetilde{\sigma}_{k'}^2$ $(k \neq k')$. The inference procedure described in Section 3 of the main paper is now available with $\widetilde{\nu}_{kk}$ and $\widetilde{\nu}_{kk'}$.

B Transformation of Covariates

As discussed in Section 4 of the main paper, transformation of the covariates is particularly important in our procedure. For completeness, we summarize the pointwise center/scaling transformation of the functional covariate proposed by Kim et al. (2016). The pointwise center/scaling transformation of $X(t)$ is defined by $X_i^*(s) = \{X_i(s) - \mu_X(s)\}/\sigma_X(s)$, where $\mu_X(s)$ and $\sigma_X(s)$ are mean and standard deviation of the covariate $X_i(s)$. In practice, for

a fixed point s_k, we estimate the mean and the standard deviation by the sample mean $\widehat{\mu}_X(s_k)$ and the sample standard deviation $\widehat{\sigma}_X(s_k)$, respectively, based on the covariates $\{X_i(s_k)\}_{i=1}^n$. The transformed version of the new covariates, $X_0^*(s)$, is also needed in the prediction of a new response curve; for a fixed point s_k, we obtain a realization of the transformed covariate $X_0^*(s_k)$ based on the sample mean $\widehat{\mu}_X(s_k)$ and the sample standard deviation $\widehat{\sigma}_X(s_k)$ obtained from the original observed data $\{X_i(s_k)\}_{i=1}^n$. By applying the transformation technique, we can avoid numerical stability issues, while still preserving prediction accuracy.

C Details of the Simulation Setup

In this section, we describe the design of the numerical study and the evaluation criteria used in Sections 4.2.1 and 4.2.2.

C.1 Simulation Design

We construct simulation data sets using the model

$$Y(t) = \int_{\mathcal{T}_X} F\{X(s), s, t\}ds + \epsilon_i(t), \tag{1}$$

with the true covariate $X(s)$ given by $X(s) = a_1 + a_2\sqrt{2}\sin(\pi s) + a_3\sqrt{2}\cos(\pi s)$, where $a_p \sim N(0, 2^{(1-p)2})$ for $p = 1, 2, 3$. Throughout the study, it is assumed that each covariate $X_i(t)$ is not observed directly due to noise. We construct noisy trajectories W_{ik} by $W_{ik} = X_i(s_{ik})+\text{Normal}(0,\tau^2)$ with a noise variance equal to $\tau = 0.5$. Define $\phi_1(t) = 1$, $\phi_2(t) = \sqrt{2}\sin(2\pi t)$, $\phi_3(t) = \sqrt{2}\cos(2\pi t)$, $\phi_4(t) = \sqrt{2}\sin(4\pi t)$, and $\phi_5(t) = \sqrt{2}\cos(4\pi t)$. Then, the response $Y_i(t)$ is generated based on model (1) for all of the following factors:

(1) True function $F(x, s, t)$:

- **(i)** linear case: $F_1(x, s, t) = [\sqrt{2}\sin(2\pi t)s + \sqrt{2}\cos(2\pi t)2\cos(\pi s)]x$
- **(ii)** simple nonlinear case: $F_2(x, s, t) = \phi_1(t)xs + \phi_2(t)(4 - 2(\frac{x}{5})^2 - 10s)$
- **(iii)** complex nonlinear case: $F_3(x, s, t) = \phi_1(t)xs+\phi_2(t)(4-2(\frac{x}{5})^2-10s)+\phi_3(t)\sin(2-x-2s) + \phi_4(t)2x\cos(\pi s) + \phi_5(t)[\exp\{-(\frac{x}{5})^2 - (\frac{s-0.5}{0.3})^2\} - 0.5]$

(2) Error process $\mathbb{E}_i = [\epsilon_i(t_{i1}), \ldots, \epsilon_i(t_{im_{Y,i}})]^T$:

- **(i)** $\mathbb{E}_i^1 \sim\text{Normal}(0,\sigma_\epsilon^2 I_{m_{Y,i}})$

(ii) $\mathbb{E}_i^2 \sim$ Normal(0,$\sigma_\epsilon^2\Sigma$)+Normal(0, $\sigma_\epsilon^2 I_{m_{Y,i}}$) where Σ has $AR_\rho(1)$ structure

(iii) $\mathbb{E}_i^3 \sim \omega_{i1} + \omega_{i2}\sqrt{2}\sin(2\pi t)$+ Normal(0, $\sigma_\epsilon^2 I_{m_{Y,i}}$)

(iv) $\mathbb{E}_i^4 \sim \omega_{i1}\sqrt{2}\cos(2\pi t) + \omega_{i2}\sqrt{2}\sin(4\pi t)$+ Normal(0, $\sigma_\epsilon^2 I_{m_{Y,i}}$)

We set $\sigma_\epsilon = 0.3$ and $\rho = 0.1$. The random variables ω_{i1} and ω_{i2} are uncorrelated and are generated from N(0,0.15^2) and N(0, 0.1^2), respectively.

(3) Sampling design:

(i) dense: $m_Y = 101$ and $m_W = 81$ equidistant time points in [0,1]

(ii) sparse:
$m_{Y,i} \overset{iid}{\sim}$ Uniform($\{35, \ldots, 44\}$) (preserve 34.7%∼43.6% of the data per curve)
$m_{W,i} \overset{iid}{\sim}$ Uniform($\{45, \ldots, 54\}$) (preserve 55.6%∼66.7% of the data per curve)

We do not consider the cases where the sampling design of the covariate is sparse but the response curves are sampled densely, and vice versa.

(4) Number of subjects: (i) n = 50, (ii) n = 100, and (iii) n = 300

Throughout this study, the covariate in the test data is observed with noise. We generated the covariates $X_{0,i'}(s)$ $(i' = 1, 2, \ldots, n')$ of the test data set assuming that the test data follow the same distribution as the training data. We assessed the predictive performance using the sample size setting $n' = 50$, $m_Y = 101$, and $m_X = 81$. When the coverage performance was assessed, we increased the sample size in the test data to $n' = 100$.

C.2 Evaluation Criteria

For each scenario, we used Monte-Carlo simulations with $N = 1000$ replications. Our performance measures are the in-sample and out-of-sample root mean squared prediction error (RMSPE) and the average coverage probability (ACP) of the pointwise prediction intervals. We define the in-sample RMSPE by

$$\text{RMSPE}^{\text{in}} = N^{-1}\textstyle\sum_{r=1}^{N}\left[n^{-1}\sum_{i=1}^{n} m_{Y,i}{}^{-1}\sum_{j=1}^{m_{Y,i}}\{Y_i^{(r)}(t_{ij}) - \widehat{Y}_i^{(r)}(t_{ij})\}^2\right]^{\frac{1}{2}},$$

where $Y_i^{(r)}(t_{ij})$ and its estimate $\widehat{Y}_i^{(r)}(t_{ij})$ are from the r^{th} Monte Carlo simulation. We define the out-of-sample RMSPE by

$$\text{RMSPE}^{\text{out}} = N^{-1}\textstyle\sum_{r=1}^{N}\left[(50\cdot 101)^{-1}\sum_{i'=1}^{50}\sum_{j=1}^{101}\{\widetilde{Y}_{0,i'}^{(r)}(t_j) - \widehat{Y}_{0,i'}^{(r)}(t_j)\}^2\right]^{\frac{1}{2}},$$

where $\widetilde{Y}^{(r)}_{0,i'}(t) = \int_{\mathcal{T}_X} F\{X^{(r)}_{0,i'}(s), s, t\}ds$ and $\widehat{Y}^{(r)}_{0,i'}(t_j)$ are from the r^{th} Monte Carlo simulation. The RMSPE$^{\text{out}}$ captures prediction errors, and one would expect that the values of RMSPE$^{\text{out}}$ converge to zero as the sample size increases.

Finally, we approximate $(1-\alpha)$ level pointwise prediction intervals to observe coverage probabilities at the nominal level. We define the ACP of the prediction intervals at the $(1-\alpha)$ level by

$$\text{ACP}_p(1-\alpha) = (100 \cdot 101 \cdot N)^{-1} \sum_{r=1}^{N} \sum_{i'=1}^{100} \sum_{j=1}^{101} I\{Y_{0,i'}(t_j) \in P^{(r)}_{1-\alpha,i'}(t_j)\},$$

where $P^{(r)}_{1-\alpha,i'}(t_j)$ are the pointwise prediction intervals from the r^{th} Monte Carlo simulation, and $I(\cdot)$ is the indicator function. In this calculation the prediction intervals are constructed using the same fixed test data set over the Monte Carlo replications.

C.3 Implementation

In this section we detail the implementation of our additive function-on-function based on principal components regression model (AFF-PC) as well as comment on the implementation for the competing approaches. The description includes the selection of the tuning parameters that was used throughout the simulation study.

The additive function-on-function principal component-based regression model (AFF-PC) was fit using $K_x = K_s = 7$ dimension cubic B-splines for x and s. The number of eigenbasis functions K was selected to have 95% of the variance explained, and so K can vary between data sets. The smoothing parameters λ_x and λ_s were selected by restricted maximum likelihood (REML) implemented with the `gam()` function in the `mgcv` package (Wood 2006), and the integration was approximated by Simpson's rule. Also, the FPC decompositions were implemented using `fpca.sc` function in the `refund` (Huang et al. 2015) `R` package. The number of basis functions in x and s directions were such that the product K_xK_s is smaller than the smallest sample size considered in the simulation experiment (which is $n = 50$); this constraint is needed for computational feasibility.

The functional linear model (FLM) was also fit using the `sff()` function n the `refund` package; the bivariate coefficient function, $\beta(s,t)$, was modeled using the tensor product of two univariate spline bases one for each direction. We used seven cubic B-splines for each direction and the two corresponding smoothness parameters were selected using REML.

The B-spline based estimation, labeled AFF-S, (Scheipl et al. 2015) was fit using the `sff()` function in the `refund` package with six cubic B-splines for direction t and five cubic B-splines for directions x and s. The three corresponding smoothness parameters were selected using REML.

Table 1: Summary of (1) RMSPE^{in} and (2) $\text{RMSPE}^{\text{out}}$ based on 1000 simulated data sets.

		$\mathbb{E}_i = \mathbb{E}_i^1$		$\mathbb{E}_i = \mathbb{E}_i^2$		$\mathbb{E}_i = \mathbb{E}_i^3$		$\mathbb{E}_i = \mathbb{E}_i^4$		$\mathbb{E}_i = \mathbb{E}_i^1$		$\mathbb{E}_i = \mathbb{E}_i^2$		$\mathbb{E}_i = \mathbb{E}_i^3$		$\mathbb{E}_i = \mathbb{E}_i^4$	
		(1)	(2)	(1)	(2)	(1)	(2)	(1)	(2)	(1)	(2)	(1)	(2)	(1)	(2)	(1)	(2)
n	method	$F(x,s,t) = F_2(x,s,t)$, **moderately sparse design**								$F(x,s,t) = F_3(x,s,t)$, **moderately sparse design**							
50	`FLM`	0.328	0.149	0.445	0.150	0.372	0.156	0.372	0.155	0.514	0.461	0.595	0.463	0.543	0.465	0.544	0.464
	`AFF-PC`	0.310	0.100	0.429	0.099	0.348	0.095	0.352	0.110	0.345	0.207	0.456	0.213	0.384	0.220	0.385	0.224
100	`FLM`	0.330	0.143	0.446	0.144	0.375	0.146	0.375	0.146	0.522	0.442	0.602	0.443	0.550	0.442	0.551	0.442
	`AFF-PC`	0.309	0.081	0.429	0.077	0.350	0.071	0.353	0.083	0.339	0.167	0.453	0.172	0.380	0.173	0.380	0.176
300	`FLM`	0.331	0.138	0.447	0.138	0.377	0.140	0.377	0.139	0.522	0.427	0.602	0.428	0.552	0.428	0.552	0.428
	`AFF-PC`	0.306	0.058	0.428	0.057	0.352	0.051	0.353	0.056	0.332	0.133	0.448	0.136	0.377	0.138	0.377	0.139

The functional additive models, labeled FAM, (Müller and Yao 2008) was implemented using the `FPCfam()` function in `PACE` package written in `Matlab`. For the estimation of the additive model components, a Gaussian kernel was used and the bandwidth was selected via generalized cross-validation (GCV); these choices are the default for fitting FAM.

D Additional Simulation Results

D.1 Additional Simulations for Irregular and Sparse Design

We further investigated predictive performance for sparse design with a different level of sparseness. In this experiment, we generated the data using increased number of time points for s and t; for convenience, we call this setting moderately sparse design. Although the simulations in the main paper used smaller numbers of time points per curve, we continue to call this setting sparse design. For the response trajectories, $m_{Y,i} = 55$–66 points were randomly selected from [0, 1] interval for each curve, while the simulations in the main paper used only $m_{Y,i} = 35$–44 points for t. For covariate trajectories, we randomly selected $m_{W,i} = 50$–62 points from [0, 1] interval for each curve, while the settings in the main paper used $m_{W,i} = 45$–54 points for s. Results are shown in Table 1 and show improved prediction performance with AFF-PC over FLM for both in-sample and out-of-sample prediction.

Table 2: Summary of (1) RMSPE^{in} and (2) $\text{RMSPE}^{\text{out}}$ based on 1000 simulated data sets. Results are obtained by applying the AFF-PC model.

	dense design				**sparse design**			
	PVE=95%		PVE=99%		PVE=95%		PVE=99%	
n	(1)	(2)	(1)	(2)	(1)	(2)	(1)	(2)
50	0.374	0.201	0.369	0.194	0.387	0.229	0.382	0.221
100	0.372	0.161	0.366	0.150	0.386	0.186	0.380	0.176
300	0.370	0.131	0.364	0.116	0.380	0.143	0.375	0.129

Table 3: Summary of ACP for the new response $Y_0(t)|X_0(\cdot)$, i.e., ACPp, at nominal significance levels $1-\alpha$ =0.85, 0.90, and 0.95. Results are based on 1000 simulated data sets with 100 bootstrap replications per data.

	dense design						**sparse design**					
	PVE=95%			PVE=99%			PVE=95%			PVE=99%		
n	0.85	0.90	0.95	0.85	0.90	0.95	0.85	0.90	0.95	0.85	0.90	0.95
50	0.912	0.947	0.978	0.910	0.946	0.977	0.931	0.958	0.982	0.937	0.962	0.984
100	0.895	0.935	0.972	0.893	0.934	0.971	0.903	0.941	0.975	0.908	0.944	0.976
300	0.870	0.916	0.962	0.868	0.914	0.960	0.879	0.923	0.966	0.877	0.922	0.965

D.2 Sensitivity Analysis

We also investigated the finite sample performance of our method using a different choice of K. Both prediction and inference performances were assessed for the cases where $F(x,s,t) = F_3(x,s,t)$ and $\mathbb{E}_i = \mathbb{E}^3$. We compared selection of K so that 95% or 99% of the variance is explained. Simulation results are presented in Table 2 and Table 3. Table 2 summarizes the in-sample and out-of-sample RMSPEs obtained by fixing the percent of variance explained (PVE) to 95% and 99%. The choice of K slightly affects the values of RMSPE, but the overall predictive performance does not change significantly. Table 3 summarizes the ACP of the pointwise prediction intervals obtained by fixing the PVE to 95% and 99%. The coverage performance is not affected by the choice of K.

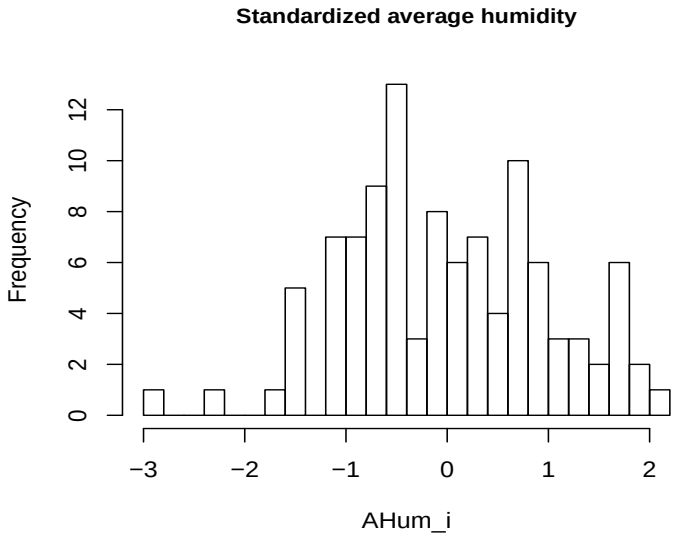

Figure 1: The histogram of the standardized average humidity.

E Further Investigation of Capital Bike Share Study

E.1 Histogram of Standardized Average Humidity

This section provides additional figures for the bike share data application in Section 6 of the main paper. Figure 1 presents a histogram of standardized average humidity.

Next we present another perspective to visualize the estimated surface for the bike data application study. Figure 2 shows the surface plots of the estimated surface, $\widehat{F}(x, s, 0)$ (left), $\widehat{F}(x, s, 12)$ (middle) and $\widehat{F}(x, s, 20)$ (right), while the solid line corresponds to $s = 10$. The nonlinear shape of the curve is indicative of nonlinearity of the surface as function of x.

E.2 Sensitivity to the choice of tuning parameters

As the Associate Editor and one anonymous Reviewer remarked, choosing the number of basis functions to be "sufficiently large" is not trivial in practice. Theoretically, the basis dimension is required to be "large enough" to capture the features of the underlying function; though its exact value is not important and cannot be too large, see Li and Ruppert (2008) and Kauermann and Opsomer (2011). As expected, there are practical limitations: the number of basis functions is generally selected to be smaller than the sample size (or else we could have numerical problems), and the computational burden increases with a larger number of basis functions. Only a few studies have investigated when in practice the basis dimension is "large enough"; Ruppert (2002) and Pya and Wood (2016) shed some

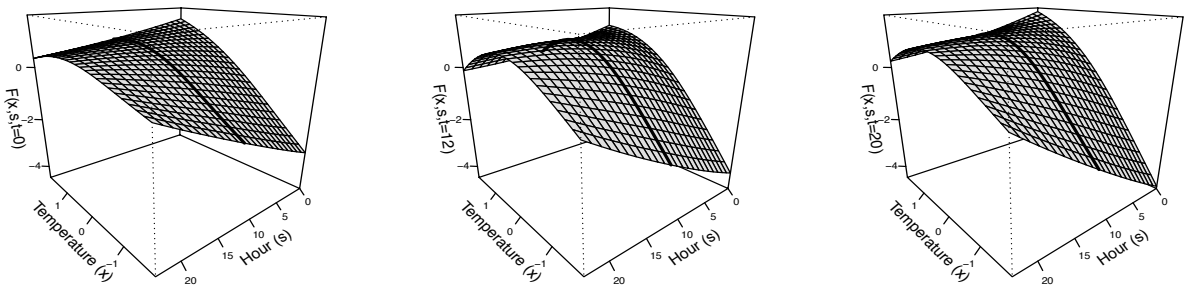

Figure 2: Displayed are surface plots of the estimated surface, $\widehat{F}(x, s, 0)$ (left), $\widehat{F}(x, s, 12)$ (middle) and $\widehat{F}(x, s, 20)$ (right). The solid line corresponds to $s = 10$.

light in this direction. Ruppert (2002) compared using GCV to determine the number of knots of a penalized spline with using a fixed number of knots. Pya and Wood (2016) proposed two approaches, hypothesis testing and residual-based smoothing, to check whether the spline basis dimension is adequate when univariate B-splines functions or thin-plate functions are used. Their ideas are implemented in the `check.gam()` function in the `mgcv` package (Wood 2006).

We carried out a sensitivity analysis for the data application. For $K_x = K_s$ we considered values from 5 to 9, and we selected K so that PVE=95% and 99%; the possible choices for $K_x = K_s$ are limited by the sample size of the training set, $n = 89$. Table 4 and Table 5 show the results for PVE=95% and PVE=99%, respectively. The study shows that AFF-PC is not very sensitive to the choices of the tuning parameters; the results presented in the main manuscript are based on $K_x = K_s = 7$ and $\widehat{K} = 3$, corresponding to PVE=95%. The tables also illustrate how the computational cost scales with the choice of the number of basis functions for the AFF-PC, as well as for FLM and AFF-S .

E.3 Bootstrap-based prediction intervals construction

Algorithm 2 below is a modification of Algorithm 1 in Section 3.2 of the main paper. Algorithm 2 allows additional covariates and can be used to construct a $100(1-\alpha)\%$ point-

Table 4: Results from the example in Section 6 of the main paper. Displayed are the summaries of (1) RMSPE^{in}, (2) $\text{RMSPE}^{\text{out}}$ and (3) computation time (in seconds) when regressing $\log(1+\text{count}_{ij})$ on temperature and average humidity with PVE set at 0.95.

		log-transformed data		original data		
method	(K_x, K_s, K_t)	(1)	(2)	(1)	(2)	(3)
FLM	(NA,5,5)	0.746	0.613	64.965	45.787	1.23
AFF-S	(5,5,5)	0.648	0.506	37.689	32.389	3.94
AFF-PC	(5,5, $\widehat{K}=3$)	0.638	0.491	37.846	30.979	0.82
FLM	(NA,7,7)	0.740	0.606	62.079	43.603	2.12
AFF-S	(7,7,7)	0.637	0.494	37.275	28.826	25.36
AFF-PC	(7,7, $\widehat{K}=3$)	0.635	0.493	38.184	31.715	1.97
FLM	(NA,9,9)	0.737	0.603	60.853	42.251	1.53
AFF-S	(9,9,9)	0.625	0.523	38.221	31.587	177.76
AFF-PC	(9,9, $\widehat{K}=3$)	0.634	0.493	38.261	31.242	3.34

Table 5: Results from the example in Section 6 of the main paper. Displayed are the summaries of (1) RMSPE^{in}, (2) $\text{RMSPE}^{\text{out}}$ and (3) computation time (in seconds) obtained by regressing $\log(1+\text{count}_{ij})$ on temperature and average humidity and PVE is set as 0.99.

		log-transformed data		original data		
method	(K_x, K_s, K_t)	(1)	(2)	(1)	(2)	(3)
FLM	(NA,5,5)	0.746	0.613	64.965	45.787	1.40
AFF-S	(5,5,5)	0.648	0.506	37.689	32.389	4.56
AFF-PC	(5,5, $\widehat{K}=5$)	0.637	0.491	37.587	30.970	1.12
FLM	(NA,7,7)	0.740	0.606	62.079	43.603	2.36
AFF-S	(7,7,7)	0.637	0.494	37.275	28.826	25.51
AFF-PC	(7,7, $\widehat{K}=5$)	0.635	0.493	38.174	31.512	4.81
FLM	(NA,9,9)	0.737	0.603	60.853	42.251	1.75
AFF-S	(9,9,9)	0.625	0.523	38.221	31.587	178.57
AFF-PC	(9,9, $\widehat{K}=5$)	0.634	0.493	38.261	31.079	12.38

wise prediction interval for the new response curve $CB_0(t)$ as $\widehat{CB}_0(t) \pm Z_{\alpha/2}\widehat{\text{SE}}\{\widehat{CB}_0(t) - CB_0(t)\}$, where $Z_{\alpha/2}$ is the $\alpha/2$ upper quantile of the standard normal distribution and $\text{SE}\{\widehat{CB}_0(t) - CB_0(t)\} = [\widehat{\text{var}}\{\widehat{CB}_0(t) - CB_0(t)\}]^{1/2}$ is obtained by bootstrapping as in Algorithm 2.

Algorithm 2 Inference in the Presence of Multiple Predictors

1: **for** $b = 1$ **to** B **do**

2: Bootstrap the subjects with replacement. Let $\{b_1, \ldots, b_n\}$ be the subject index of the bootstrap sample.

3: Define the covariate and the response curves in the b^{th} bootstrap sample as $\{Temp_i^{(b)}(\cdot) = Temp_{b_i}(\cdot)\}_{i=1}^n$ and $\{CB_i^{(b)}(\cdot) = CB_{b_i}(\cdot)\}_{i=1}^n$, respectively. Also, define $\{AHum_i^{(b)} = AHum_{b_i}\}_{i=1}^n$ for the additional covariate. The bootstrap data for the i^{th} subject is obtained by collecting the trajectories $\{Temp_i^{(b)}(s_k), s_k\}_{k=1}^{m_W}$, $\{AHum_i^{(b)}\}$ and $\{CB_i^{(b)}(t_j), t_j\}_{j=1}^{m_Y}$.

4: Apply FPCA to $\{CB_i^{(b)}(\cdot)\}_{i=1}^n$ and obtain estimate of the eigenbasis $\{\phi_k^{(b)}(\cdot)\}_{k=1}^{K^{(b)}}$, where $K^{(b)}$ is the finite truncation estimated using PVE and the pre-specified threshold level.

5: For $l = 1, \ldots, K_x$, $l' = 1, \ldots, K_s$, and $k = 1, \ldots, K^{(b)}$, obtain parameter estimates $\widehat{\beta}_k^{(b)}$, $\widehat{\zeta}_k^{(b)}$ and $\widehat{\theta}_{l,l',k}^{(b)}$ by fitting the AFF-PC model based on $\{Temp_i^{(b)}(s_k), s_k\}_{k=1}^{m_W}$, $\{AHum_i^{(b)}\}$ and $\{CB_i^{(b)}(t_j), t_j\}_{j=1}^{m_Y}$.

6: For new covariates $Temp_0(s)$ and $AHum_0$, obtain the predicted response by $\widehat{CB}_0^{(b)}(t) = \sum_{k=1}^{K^{(b)}} \widetilde{\phi}_k^{(b)}(t) \left[\widehat{\beta}_k^{(b)} + AHum_0 \widehat{\zeta}_k^{(b)} + \sum_{l=1}^{K_x}\sum_{l'=1}^{K_s} \widehat{\theta}_{l,l',k}^{(b)} \int_{\mathcal{T}_X} B_{X,l}\{Temp_0(s)\} B_{S,l'}(s) ds\right]$.

7: Compute $V^{(b)}(t) = \widehat{\text{var}}\{\widehat{CB}_0^{(b)}(t) | \widetilde{\eta}_b\}$ using the model-based formula in (10) in the main paper.

8: **end for**

9: Approximate the marginal variance of predicted response by

$$\widehat{\text{var}}\{\widehat{CB}_0(t)\} \approx \frac{1}{B}\sum_{b=1}^B V^{(b)}(t) + \frac{1}{B}\sum_{b=1}^B \{\widehat{CB}_0^{(b)}(t) - \overline{CB_0(t)}\}^2,$$

where $\overline{CB_0(t)}$ is the sample mean of $\widehat{CB}_0^{(b)}(t)$'s.

F Yield Curves Data

Here, we consider an application to yield curves. The yield at maturity T years is the average interest rate that is earned on a bond maturing in T years (Ruppert and Matteson 2015). The yield curve for a given type of bond is the plot of yield against maturity, and the shape of a yield curve changes each day. Maturities range from short-term (e.g., less than 2 years) to long-term (e.g., 20 years, 30 years). The curves can take three primary shapes: positive slope (so-called "normal"), negative slope (so-called "inverted"), and flat curve. The slope of the yield curve is a powerful indication of the future economic growth (Plosser and Rouwenhorst 1994). If the slopes are positive, one can expect strong

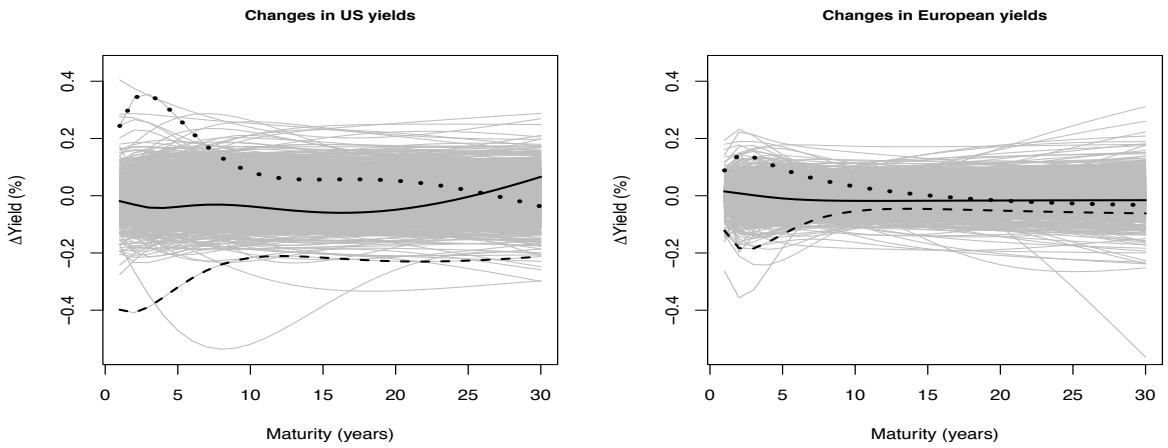

Figure 3: The gray lines are changes in the US yield curves (left panel) and in the European yield curves (right panel) during the period from January 2, 2006 to December 30, 2011. The measurements taken in three different days are indicated by solid, dashed, and dotted black lines.

future economic growth and higher interest rates. The negative slopes indicate slow future economic growth and a rise in unemployment.

The changes in US yield curves are often discussed in relation to the changes of yield curves in other countries (see e.g., Plosser and Rouwenhorst 1994; Mehl 2009). In particular, Plosser and Rouwenhorst (1994) studied linkages between the real economic growth in the United States and interest-rates across other countries including Germany and the UK. Motivated by these studies, we consider the problem of predicting the changes in the US daily yield curves using the changes in the European yield curves on the same day. Our data consist of US and European daily yields during the period from January 2, 2006 to December 30, 2011 - a total of 1286 days. Figure 3 presents changes in the US yield curves (left panel) and European yield curves (right panel), each curve corresponding to a particular day. Simple visual inspection reveals that the changing patterns between the US and the European yield curves are very similar.

Let $Y_{ij} = Y_i(t_j)$ be the changes in the US yield curves corresponding to i^{th} day ($i = 1, \ldots, 1286$) measured at maturity t_j ($j = 1, \ldots, 30$), where the maturities t_j range from 1 year to 30 years. Let $W_{ik} = X_i(s_k) + \delta_{ik}$ be the changes in the European yield curves

corresponding to i^{th} day measured at maturity s_k ($k = 1, \ldots, 30$), and δ_{ik} is the white noise random deviation. We considered two modeling approaches, the AFF-PC model and the functional linear model, and assessed both the in-sample and out-of-sample prediction accuracy based on RMSPE. For the AFF-PC model, the functional covariate was processed using the methods described in Section 4. The first 1000 days were taken as a training data set, and the last 286 days were taken as the test data. Thus, our models were estimated based on the past changes in the yield curves, while the model assessment was for future changes. To fit the AFF-PC model, we used $K_x = K_s = 7$ dimension cubic B-spline bases for x and s, and the number of eigenbasis K was selected to have 95% of the variance explained. In the functional linear model, the bivariate coefficient function, $\beta(s, t)$, was modeled using a 7 dimensional cubic B-spline basis for each variable. Both the AFF-PC model and the functional linear model provided the similar values for the in-sample and out-of-sample predication errors (RMSPE^{in}=0.059 and $\text{RMSPE}^{\text{out}}$=0.064), indicating that a simple linear association is appropriate. The coverage performance was also assessed for pointwise prediction intervals constructed using 1000 bootstrap replications. For the nominal levels of 0.85, 0.90, and 0.95, the average coverage probabilities were 0.847, 0.885, 0.920, respectively.

References

Huang, L., F. Scheipl, J. Goldsmith, J. Gellar, J. Harezlak, M. W. McLean, B. Swihart, L. Xiao, C. Crainiceanu, and P. Reiss (2015). *refund: Regression with Functional Data*. R package version 0.1-13.

Kauermann, G. and J. D. Opsomer (2011). Data-driven selection of the spline dimension in penalized spline regression. *Biometrika 98*(1), 225–230.

Kim, J., A. Maity, and A. M. Staicu (2016). General functional concurrent model. Unpublished manuscript (under review).

Li, Y. and D. Ruppert (2008). On the asymptotics of penalized splines. *Biometrika 95*(2), 415–436.

Mehl, A. (2009). The yield curve as a predictor and emerging economies. *Open Economies Review 20*, 683–716.

Müller, H.-G. and F. Yao (2008). Functional additive models. *Journal of the American Statistical Association 103*, 1534–1544.

Plosser, C. I. and K. G. Rouwenhorst (1994). International term structures and real economic growth. *Journal of Monetary Economics 33*, 133–155.

Pya, N. and S. N. Wood (2016). A note on basis dimension selection in generalized additive modelling. *arxiv*.

Ramsay, J. O. and B. W. Silverman (2005). *Functional Data Analysis* (2nd ed.). New York, Springer.

Ruppert, D. (2002). Selecting the number of knots for penalized splines. *Journal of computational and graphical statistics 11*(4), 735–757.

Ruppert, D. and D. Matteson (2015). *Statistics and data analysis for financial engineering* (2nd ed.). Springer.

Scheipl, F., A.-M. Staicu, and S. Greven (2015). Functional additive mixed models. *Journal of Computational and Graphical Statistics 24*, 477–501.

Wood, S. N. (2006). *Generalized Additive Models: An Introduction with R*. Boca Raton, Florida: Chapman and Hall/CRC.

Yao, F., H.-G. Müller, and J.-L. Wang (2005). Functional data analysis for sparse longitudinal data. *Journal of the American Statistical Association 100*, 577–590.

Zhang, J.-T. and J. Chen (2007). Statistical inference for functional data. *Annals of Statistics 35*, 1052–1079.