Latent group detection in functional partially linear regression models

Biometrics 64, 1–22 DOI: 10.1111/j.1541-0420.2005.00454.x
December 2020

Latent group detection in functional partially linear regression models

Wu Wang[1,*], Ying Sun[2,], and Huixia Judy Wang[3,***]**

[1]Center for Applied Statistics and School of Statistics, Renmin University of China, Beijing, China

[2]Statistics Program, King Abdullah University of Science and Technology, Thuwal, Saudi Arabia

[3]Department of Statistics, The George Washington University, Washington, DC, U.S.A.

email: wu.wang@ruc.edu.cn

**email:* ying.sun@kaust.edu.sa

***email:* judywang@gwu.edu

Summary: In this paper, we propose a functional partially linear regression model with latent group structures to accommodate the heterogeneous relationship between a scalar response and functional covariates. The proposed model is motivated by a salinity tolerance study of barley families, whose main objective is to detect salinity tolerant barley plants. Our model is flexible, allowing for heterogeneous functional coefficients while being efficient by pooling information within a group for estimation. We develop an algorithm in the spirit of the K-means clustering to identify latent groups of the subjects under study. We establish the consistency of the proposed estimator, derive the convergence rate and the asymptotic distribution, and develop inference procedures. We show by simulation studies that the proposed method has higher accuracy for recovering latent groups and for estimating the functional coefficients than existing methods. The analysis of the barley data shows that the proposed method can help identify groups of barley families with different salinity tolerant abilities.

Key words: Functional data analysis; Homogeneity pursuit; Latent structure; Longitudinal data analysis; Model-based clustering.

This paper has been submitted for consideration for publication in *Biometrics*

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process, which may lead to differences between this version and the Version of Record. Please cite this article as doi: 10.1111/biom.13557

1. Introduction

As new technologies rapidly develop, more and more processes can be monitored dynamically over time and space. Densely observed data can be modeled as realizations of random curves or surfaces, which are examples of functional data. Functional data analysis (FDA) has been successfully applied in fields such as neuroimaging (Yu et al., 2016), plant science (Meng et al., 2017), and medical science (Kong et al., 2018). See Ramsay and Silverman (2005) for a comprehensive review.

Functional linear regression is an elegant statistical framework that links functional covariates with response variables. Because the space of functions is infinite dimensional, dimension reduction techniques, especially the functional principal component analysis (FPCA), are widely applied before further modeling. Yao et al. (2005), Hall and Horowitz (2007), and Hall and Hosseini-Nasab (2009) established sound theoretical properties of the FPCA-based estimators for functional linear models. When scalar covariates are present, Shin (2009), Kong et al. (2016), Kong et al. (2016), and Kong et al. (2018) studied the functional partially linear regression model for possibly high-dimensional covariates. Another line of work approximates functional coefficients by fixed basis functions. Popular basis functions are B-splines (Cardot et al., 2003), smoothing splines (Crambes et al., 2009), and reproducing kernel Hilbert spaces (Yuan and Cai, 2010). Functional linear regression has been generalized to deal with discrete responses (Müller and Stadtmüller, 2005), to model conditional quantiles of the response (Kato, 2012), and to handle complex nested structures in functional data (Xu et al., 2018).

Our work is motivated by the research on salinity tolerance of barley plants. Salinity is the primary environmental stress that limits growth and productivity of crops (Munns and Tester, 2008). Plant scientists are dedicated to finding salinity tolerant barley plants and understanding the underlying mechanism (Meng et al., 2017). In a barley salinity tolerance experiment conducted in The Plant Accelerator®, a facility of smart houses with automated

phenotyping technologies at King Abdullah University of Science and Technology, Thuwal, Saudi Arabia, plant scientists recorded daily growth rates of barley plants stressed with salt water for a period of 32 days. The growth rates were naturally modeled as functional data. The barley plants were from different families; those from the same family are treated as replicates because of their genetic similarity. To understand how the barley families react to the stress imposed by saltwater, we model the relationship between the final biomass of the barley plants and the relative growth rate using a functional regression model. The ability to tolerate saltwater was found inhomogeneous among barley families. Some barley families were more salt-tolerant than others (Meng et al., 2017). This means that there may exist heterogeneous latent groups of barley plants that have different salinity tolerant abilities.

Subject heterogeneity is a fundamental model specification problem. It arises due to individual characteristics, either unobserved or unknown. For example, in precision medicine, the ability to benefit from treatments is often different across different subgroups of patients based on their health conditions or genes (Wang et al., 2019). Because heterogeneity is induced by the studied subjects, functional data is not exempted from this issue. Exploring heterogeneity offers us an opportunity to gain insight into the underlying scientific problem, whereas failing to account for heterogeneity leads to biased estimates and inference.

Models that properly handle subject heterogeneity have not been extensively studied in functional data analysis. Yao et al. (2011) and Wang et al. (2016) proposed a functional mixture model that allows the regression structure to vary across latent groups of subjects. Their estimator exploited FPCA for dimension reduction and mixture regression methods for recovering latent group structures. A mixture model requires stringent distribution assumptions and suffers from high computational complexity. For non-functional data, one line of work penalizes the pairwise differences of the subject-specific coefficients for group recovery (Ma and Huang, 2017). Another line of research extends the clustering algorithms to detect

latent groups; see Bonhomme and Manresa (2015) and Zhang et al. (2019). Lastly, Ke et al. (2016) proposed a method based on change-point detection algorithms. These studies are restricted to scalar data, and functional data are not allowed.

Motivated by the barley data, we propose a functional partially linear regression model with latent group structures to account for subject heterogeneity. The regression coefficients are shared within the same group, whereas they are distinct across groups. Our model does not assume a particular relationship between the latent group membership and the observed covariates, as it usually does in mixture modeling. The latent group structure can thus be driven by arbitrary combinations of observed covariates and unobserved features. The analysis of the barley data demonstrates that the proposed method can help detect barley families that are more tolerant to saline conditions. Potential applications of the proposed methodology are treatment regime estimation with functional covariates (Ciarleglio et al., 2018) and assessment of heterogeneous effects of air pollution on health across different regions and age groups (Kong et al., 2016).

We propose a new method to identify latent groups based on FPCA and the idea of K-means clustering algorithm. In our experiment, the proposed algorithm is very fast, often converges within ten iterations and is ten times faster than the fused penalization approach. Compared to the competing methods, the proposed estimator has higher accuracy for recovering latent groups and for estimating the functional coefficients. We develop confidence sets of the parameters, including both the scalar coefficients and the functional coefficients. We prove the consistency of the group membership estimator, derive the asymptotic distribution of the estimator for the scalar coefficients, and obtain the convergence rate of the functional coefficient estimator.

The remainder of the paper is organized as follows. In Section 2, we present the proposed model and the estimator, establish the theoretical properties and discuss inference problems.

We examine the performance of the proposed methodology in a simulation study in Section 3 and present the analysis of the motivating barley growth data in Section 4. Section 5 concludes the paper.

2. The proposed method

2.1 *Motivation and model*

We propose a functional partially linear regression model with latent group structures. For $i = 1, \ldots, N$ and $j = 1, \ldots, T_i$, denote by y_{ij} the response variable, by $\mathbf{z}_{ij}$ a vector of scalar covariates, and by $x_{ij}(t) \in \mathcal{L}^2(I)$ the functional covariate, where $I \subset \mathbb{R}$ and $\mathcal{L}^2(I)$ is the space of square integrable functions on I. For simplicity, we assume the design is balanced with $T_i = T$ for all i. We model y_{ij} by a functional partially linear regression model,

$$y_{ij} = \alpha_i + \mathbf{z}_{ij}'\boldsymbol{\gamma} + \int_I x_{ij}(t)\beta_{g_i}(t)\mathbf{d}t + \epsilon_{ij}, \tag{1}$$

where α_i is a family-wise fixed effect, $\boldsymbol{\gamma}$ is the effect of the scalar covariates, and $\beta_{g_i}(t)$ is the coefficient of the functional predictor. A parameter with a superscript ^{0}, e.g., $\boldsymbol{\gamma}^0$, refers to its true value. In the model, we allow clustered patterns of heterogeneity in the coefficient of the functional predictor. The group membership $g_i \in \{1, \ldots, G\}$ and the number of potential groups G are unknown and shall be estimated from the data. Within a group, subjects share the same functional coefficient, whereas subjects have distinct functional coefficients in different groups. Denote $\mathbf{y}_i = (y_{i1}, \ldots, y_{iT})^\top$ and $\mathbf{z}_i = (\mathbf{z}_{i1}^\top, \ldots, \mathbf{z}_{iT}^\top)^\top$. Let $\mathbf{y} = (\mathbf{y}_1^\top, \ldots, \mathbf{y}_N^\top)^\top$ and $\mathbf{Z} = (\mathbf{z}_1^\top, \ldots, \mathbf{z}_N^\top)^\top$ denote the stacked observations. In addition, denote by $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_N)^\top$ the vector of fixed effects and $\boldsymbol{\epsilon} = (\epsilon_{11}, \ldots, \epsilon_{NT})^\top$ the vector of errors.

2.2 *Proposed estimation method*

To estimate the parameters of the model, we first apply FPCA for dimension reduction. Let $(y, z, x(t), \epsilon)$ denote a generic $(y_{ij}, z_{ij}, x_{ij}(t), \epsilon_{ij})$. Denote the covariance function of $x(t)$ by

$K(u,v) = cov\{x(u), x(v)\}$. To simplify notations, we assume the mean of $x(t)$ is zero and use K to denote the covariance operator associated with the covariance function $K(u,v)$, i.e., for any $\phi \in \mathcal{L}^2(I)$, $(K\phi)(u) = \int_I K(u,v)\phi(v)\mathbf{d}v$. By the Mercer's Lemma (Hall and Hosseini-Nasab, 2006), the covariance function can be decomposed as $K(u,v) = \sum_{k=1}^{\infty} \kappa_k \phi_k(u)\phi_k(v)$, where $\kappa_1 > \kappa_2 > ... > 0$ and $\phi_1, \phi_2, ...$ are the eigenvalues and normalized eigenfunctions of the population covariance operator, respectively. We assume that there are no ties in the eigenvalues. The eigenfunctions $\phi_1, \phi_2, ...$ form an orthonormal basis of $\mathcal{L}^2(I)$ (Hall and Hosseini-Nasab, 2006). With the decomposition of the covariance function, we have the Karhunen-Loève (K-L) expansion, $x(t) = \sum_{k=1}^{\infty} f_k \phi_k(t)$, where $f_k = \int_I x(t)\phi_k(t)\mathbf{d}t, k = 1, 2, \ldots,$ are principal component scores.

The covariance function $K(u,v)$ can be estimated by $\widehat{K}(u,v) = (NT)^{-1}\sum_{i=1}^{N}\sum_{j=1}^{T}(x_{ij}(u) - \overline{x}(u))(x_{ij}(v) - \overline{x}(v))$, where $\overline{x}(u) = (NT)^{-1}\sum_{i=1}^{N}\sum_{j=1}^{T} x_{ij}(u)$. Let $\widehat{K}(u,v) = \sum_{k=1}^{\infty} \widehat{\kappa}_k \widehat{\phi}_k(u)\widehat{\phi}_k(v)$ be the spectral decomposition of $\widehat{K}(u,v)$, where $\widehat{\kappa}_1 \geqslant \widehat{\kappa}_2 \geqslant ... \geqslant 0$ and $\widehat{\phi}_1, \widehat{\phi}_2, ...$ are the corresponding estimators of the eigenvalues and eigenfunctions, respectively. To estimate the parameters in model (1), we need to truncate the K-L expansion (Hall and Horowitz, 2007). Let $f_{ijk} = \int_I x_{ij}(t)\phi_k(t)dt, k = 1, \ldots, m$, and denote $\mathbf{f}_{ij} = (f_{ij1}, \ldots, f_{ijm})^\top$, where m is the truncation parameter. Similarly, we denote the empirical principal component scores by $\widehat{f}_{ijk} = \int_I x_{ij}(t)\widehat{\phi}_k(t)dt, k = 1, \ldots, m$, and denote $\widehat{\mathbf{f}}_{ij} = (\widehat{f}_{ij1}, \ldots, \widehat{f}_{ijm})^\top$. The functional slope $\beta_{g_i}(t)$ can also be expanded as $\beta_{g_i}(t) = \sum_{k=1}^{\infty} b_{g_i,k}\widehat{\phi}_k(t)$. After applying FPCA, model (1) can be approximated by

$$y_{ij} \approx \alpha_i + \mathbf{z}_{ij}^\top \boldsymbol{\gamma} + \widehat{\mathbf{f}}_{ij}^\top \mathbf{b}_{g_i} + \epsilon_{ij},\ i = 1, \ldots, N, j = 1, \ldots, T. \tag{2}$$

where $\mathbf{b}_{g_i} = (b_{g_i,1}, \ldots, b_{g_i,m})$. We transform the functional covariates into scalar principal component scores, which facilitates further estimation steps. Although model (2) bears some resemblance to a linear model, the interpretation and the statistical theory are totally different. In model (1), the problem of estimating $\beta_{g_i}(t)$ is related to the ill-posed inverse

problem in operator theory (Hall and Hosseini-Nasab, 2006), and the solution by truncating FPCA is also called regularization in the literature (Shin, 2009).

In practice, the trajectories of the functional covariate $x_{ij}(t)$ may not be fully observed. In this paper, we consider the case where $x_{ij}(t)$ is observed on a dense grid. Smoothing techniques, such as the spline smoother (Ramsay and Silverman, 2005), kernel smoother (Kong et al., 2016), local constant or local linear interpolation (Kato, 2012), can be used to estimate the trajectories of $x_{ij}(t)$. Once we estimated trajectories of $x_{ij}(t)$, the estimation steps are the same with fully observed trajectories.

We temporarily assume that the number of groups G is known, and in Section 2.5 we propose a criterion for selecting G from the data. The parameters are estimated by minimizing the least squares objective function,

$$\left(\widehat{\boldsymbol{\alpha}}, \widehat{\boldsymbol{\gamma}}, \widehat{\mathbf{B}}, \widehat{\mathcal{G}}\right) = \arg\min \sum_{i=1}^{N} \sum_{j=1}^{T} (y_{ij} - \alpha_i - \mathbf{z}_{ij}^{\top}\boldsymbol{\gamma} - \widehat{\mathbf{f}}_{ij}^{\top}\mathbf{b}_{g_i})^2, \tag{3}$$

where $\widehat{\mathcal{G}} = \{\widehat{g}_1, \ldots, \widehat{g}_N\}$, and $\widehat{\mathbf{B}} = \{\widehat{\mathbf{b}}_1, \ldots, \widehat{\mathbf{b}}_G\}$. Because the group membership $\mathcal{G}$ is discrete and takes values in a finite set $\{1, \ldots, G\}$, algorithms for solving least squares problems can not be implemented directly. We optimize the objective function (3) by iterating between optimizing $g_i, i = 1, \ldots, N$, and other parameters which take continuous values.

Algorithm 1:

Iterate between the following two steps until convergence,

1. Group membership assignment, $g_i^{(s+1)} = \underset{g \in \{1,\ldots,G\}}{\arg\min} \sum_{j=1}^{T} (y_{ij} - \alpha_i^{(s)} - z_{ij}'\gamma^{(s)} - \widehat{f}_{ij}' b_g^{(s)})^2$.
2. Update continuous parameters,

$$\left(\alpha^{(s+1)}, \gamma^{(s+1)}, b_{g_i}^{(s+1)}\right) = \underset{\alpha, \gamma, b_{g_i}}{\arg\min} \sum_{i=1}^{N} \sum_{j=1}^{T} (y_{ij} - \alpha_i - z_{ij}'\gamma - \widehat{f}_{ij}' b_{g_i^{(s+1)}})^2.$$

The initial values can be set by first estimating model (1) without the functional covariate. We then apply the classical K-means algorithm to the residuals and initialize the group membership as the clustering results. Algorithm 1 is similar in spirit to the algorithm 1 of

Bonhomme and Manresa (2015) and can be considered as a generalization of the K-means algorithm. Because the minimum sum-of-squares clustering problem is NP-hard (Aloise et al., 2009), we can only expect that our algorithm converges to a local minima.

PROPOSITION 1: The series of estimates generated by Algorithm 1 converge to a local minima of the objective function (3) in finite many steps.

Proposition 1 confirms that Algorithm 1 converges to a local minima of the objective function (3). The proof of Proposition 1 is collected in the Web Appendix A. For large datasets with many latent groups, e.g., $G > 10$, Algorithm 1 can be improved by exploiting the ideas in recent advances in clustering algorithms, such as the variable neighborhood search method (Hansen and Mladenovic, 2001) and careful selection of initial values (Ordin and Bagirov, 2015). We stick to the simple and clear Algorithm 1 to convey ideas.

2.3 *Large sample theory*

In this section, we study the asymptotic properties of the proposed estimator. The main results are the consistency of the group membership estimator, the asymptotic distribution of the estimator for the scalar coefficients and the convergence rate of the functional coefficient estimator. These results are established through the asymptotic equivalence of the proposed estimator and an infeasible estimator that assumes the knowledge of group memberships. We consider the setup where both N and T go to infinity and the number of groups is known.

The difficulty in establishing the asymptotic properties is two-fold. On the one hand, the group membership estimators, $\widehat{g}_i, i = 1, \ldots, N$, only take discrete values, whose number is increasing to infinity as N increases. On the other hand, the approximation error in the principal component analysis needs to be carefully bounded, and we need to establish accurate bounds for both the approximation error and the observation error.

We first fix notations and introduce some regularity conditions. There are three types of

parameters in the model (1): the parameters $\alpha_i^0 \in \mathcal{A} \subset \mathbb{R}, i = 1, \ldots, N$, and $\boldsymbol{\gamma}^0 \in \Gamma \subset \mathbb{R}^d$ are finite dimensional; The group memberships $g_i^0, i = 1, \ldots, N$, are discrete, and they take values in $\{1, \ldots, G\}$; The functional coefficients $\beta_1^0(t), \ldots, \beta_G^0(t)$ are elements of $\mathcal{L}_2(I)$, and we assume that $\beta_g^0(t) = \sum_{k=1}^{\infty} b_{gk}^0 \phi_k(t) \in \Xi \subset \mathcal{L}_2(I), g = 1, \ldots, G$, where $\Xi = \{f(t) : f(t) = \sum_{k=1}^{\infty} b_k \phi_k(t), \sum_{k=1}^{\infty} b_k^2 < \infty\}$. Using non-standard notations, we use $\|\mathbf{z}\|$ to denote the Euclidean norm of $\mathbf{z} \in \mathbb{R}^d$, we also use $\|x\|^2 = \int_I x(t)^2 \mathbf{d}t$ to denote the L_2 norm of $x(t)$. The meaning should be clear in the context. We consider the following assumptions.

(**A1**) The spaces $\mathcal{A}, \Gamma$, and Ξ are bounded subsets of $\mathbb{R}, \mathbb{R}^d$, and $\mathcal{L}_2(I)$, respectively.

(**A2**) The data $\{y_{ij}, \mathbf{z}_{ij}, x_{ij}(t)\}$ are independent both within and across families.

The assumption (A1) is standard in the literature. Since the objective function (3) is not convex, the boundedness assumption ensures that the estimator does not drift away from the truth asymptotically. In the assumption (A2), we require that the data are independent both within and across subjects. This is reasonable in the plant data analysis. The plants were fully randomized to the positions of the Smarthouse using a split-plot design (Meng et al., 2017) and grown separately in pots. Possible sources of dependency between plants are the unobserved features of the barley plants both within and across families and the microclimate. The effects of microclimate can be removed by the method used in Meng et al. (2017). The latent group structure model is used to accommodate the salinity tolerance heterogeneity. The fixed effect is adopted to capture unobserved family-wise heterogeneity.

(**A3**) For some $r \geqslant 2$, $\mathbb{E}\|x\|^{2r} < \infty$, $\mathbb{E}\|z\|^{2r} < \infty$, and $\mathbb{E}|\epsilon|^{2r} < \infty$. For all $k \geqslant 2$, $\mathbb{E}\left(\int x(t)\phi_k(t)\mathbf{d}t\right)^{2r} < C\kappa_k^r$ for some constant $C > 0$.

(**A4**) For some $\chi > 1$ and $\nu > \chi/2 + 1$, $C^{-1}k^{-\chi} \leqslant \kappa_k \leqslant Ck^{-\chi}$, and $\max_{g=1,\ldots,G} |\int \beta_g^0(t)\phi_k(t)\mathbf{d}t| \leqslant Ck^{-\nu}$ for some constant $C > 0$. The truncation parameter $m \propto n^{1/(\chi+2\nu)}$.

(**A5**) For all $g \in \{1, \ldots, G\}$, $\lim_{N\to\infty} N^{-1} \sum_{i=1}^{N} I\{g_i^0 = g\} = \pi_g > 0$.

(**A6**) For all $\{g, g'\} \in \{1, \ldots, G\}$, $\lim_{m\to\infty} \min_{g\neq g'} \sum_{k=1}^{m} (b_{gk}^0 - b_{g'k}^0)^2 \kappa_k > 0$.

Lastly, we need to restrict the dependence between the finite dimensional covariate $\mathbf{z}$ and the functional covariate $x(t)$, which is a common assumption in semi-parametric regression analysis; see Shin (2009) and Wang et al. (2009). Let $\mathcal{H} = \{\sum_{k=1}^{\infty} h_k f_k, \sum_{k=1}^{\infty} h_k^2 < \infty\}$, where $f_k = \int_I x(t)\phi_k \mathbf{d}t$. Let $z_k, k = 1, \ldots, d$, denote the the kth element of $\mathbf{z}$. For $k = 1, \ldots, d$, let ζ_k be the projection of z_k to $\mathcal{H}$. That is $\zeta_k = \arg\min_{\zeta \in \mathcal{H}} \mathbb{E}(z_k - \zeta)^2 = \arg\min_{\zeta \in \mathcal{H}} \mathbb{E}(\mathbb{E}(z_k|x) - \zeta)^2$. Let $\widetilde{z}_k = z_k - \zeta_k$, and $\widetilde{\mathbf{z}} = (\widetilde{z}_1, \ldots, \widetilde{z}_d)'$, we make the following additional assumption.

(**A7**) The covariance matrix of $\widetilde{\mathbf{z}}$ is positive definite.

The moment conditions which appear in the assumption (A3) on the functional observations also appear in Hall and Hosseini-Nasab (2006, 2009). The assumption (A4) is adapted from Hall and Horowitz (2007). Specifically, the assumption (A4) means that the eigenvalues of the functional covariate are positive and well separated from each other, and the functional coefficients are smooth relative to the covariance function $K(u, v)$. Under this assumption, the eigenfunctions $\phi_k, k = 1, 2, \ldots,$ can be identified and estimated with a reasonable rate. The assumption (A5) requires that the number of observations for all the groups are in the same order as N grows.

The identification condition of a general functional linear model has been discussed in detail by Cardot et al. (2003), Shin (2009), and Scheipl and Greven (2016). Essentially, a functional linear model is identifiable if the cross-covariance function, which is defined as the covariance of the response and the functional covariate, is in the range of the covariance operator, and the Picard's condition is satisfied (Cardot et al., 2003; Shin, 2009). We state the identifiability of model (1) in Proposition 2.

PROPOSITION 2: Under the assumptions (A1)-(A7), the parameters $\alpha_i^0, i = 1, \ldots, N$, $\boldsymbol{\gamma}^0$,$\beta_g^0, g = 1, \ldots, G$, are identifiable. The group membership parameters $g_i^0, i = 1, \ldots, N$, are asymptotically identifiable up to a permutation.

In Proposition 2, the asymptotic identifiability of the group membership parameters $g_i^0, i =$

$1, \ldots, N$, is defined as follows. Denote $\widehat{g}_i = \underset{g\in\{1,\ldots,G\}}{\arg\min} \sum_{j=1}^{T}(y_{ij} - \alpha_i^0 - \mathbf{z}_{ij}'\boldsymbol{\gamma}^0 - \mathbf{f}_{ij}'\mathbf{b}_g^0)^2$. The group membership parameters $g_i^0, i = 1, \ldots, N$, are asymptotically identifiable if $\lim_{T\to\infty} \mathbb{P}(\widehat{g}_i \neq g_i^0) = 0$ for all $i = 1, \ldots, N$. The assumption (A6) is key to the asymptotic identifiability of g_i^0, which says that the group-specific functional coefficients are well separated in the sense that $E(\mathbf{f}_{ij}'b_g^0 - \mathbf{f}_{ij}'b_{g'}^0)^2 = \sum_{k=1}^{m}(b_{gk}^0 - b_{g'k}^0)^2\kappa_k > 0$. That is, the contribution of the functional coefficients is different across groups. A slightly stronger condition which induces the assumption (A6) is that $\|\beta_g^0 - \beta_{g'}^0\| = \sum_{k=1}^{\infty}(b_{gk}^0 - b_{g'k}^0)^2 > C$ for all g and g' for some constant $C > 0$.

To state the consistency results, we use the Hausdorff distance. For two sets of $\mathbb{R}^m$ vectors $\mathbf{B}^{(1)} = \{\mathbf{b}_1^{(1)}, \ldots, \mathbf{b}_G^{(1)}\}$ and $\mathbf{B}^{(2)} = \{\mathbf{b}_1^{(2)}, \ldots, \mathbf{b}_G^{(2)}\}$, the Hausdorff distance between $\mathbf{B}^{(1)}$ and $\mathbf{B}^{(2)}$ is defined as

$$d_H(\mathbf{B}^{(1)}, \mathbf{B}^{(2)}) = \max\left\{\max_{g'}\min_{g}\|\mathbf{b}_g^{(1)} - \mathbf{b}_{g'}^{(2)}\|, \max_{g}\min_{g'}\|\mathbf{b}_g^{(1)} - \mathbf{b}_{g'}^{(2)}\|\right\}.$$

Let the true values of the functional coefficient truncated at m be $\mathbf{B}^{(0)} = \{\mathbf{b}_1^{(0)}, \ldots, \mathbf{b}_G^{(0)}\}$, where $\mathbf{b}_g^{(0)} = (b_{g,1}^0, \ldots, b_{g,m}^0)$ and $b_{g,k}^0 = \int_I \beta_g^0(t)\phi_k(t)\mathbf{d}t$. Theorem 1 establishes the consistency of the proposed estimators $(\widehat{\boldsymbol{\alpha}}, \widehat{\boldsymbol{\gamma}}, \widehat{\mathbf{B}})$ with $\widehat{\mathbf{B}} = \{\widehat{\mathbf{b}}_1, \ldots, \widehat{\mathbf{b}}_G\}$.

THEOREM 1: *Under the assumptions (A1)-(A6), the proposed estimator is consistent:* $\sup_{i=1,\ldots,N}|\widehat{\alpha}_i - \alpha_i^0| = o_p(1)$, $\|\widehat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}^0\| = o_p(1)$, *and* $d_H(\mathbf{B}^0, \widehat{\mathbf{B}}) = o_p(1)$.

For the functional coefficients $\beta_g^0(t), g = 1, \ldots, G$, we only proved consistency of its coefficients expanded on the eigenfunctions $\phi_1, \phi_2, \ldots$. This result will be refined in Corollary 1, where we establish the convergence rate of $\widehat{\beta}_g(t)$. As argued in Bonhomme and Manresa (2015), consistency under the Hausdorff distance along with Assumption (A6) implies that there exists a permutation of group labels $\sigma : \{1, \ldots, G\} \to \{1, \ldots, G\}$, such that $\|\widehat{\mathbf{b}}_{\sigma(g)} - \mathbf{b}_g^{(0)}\| \to 0$ in probability. Without loss of generality, we assume that $\sigma(g) = g$, $g = 1, \ldots, G$, in the rest of the article.

Based on the consistency of the estimator in Theorem 1, we can establish the asymptotic

equivalence between the proposed estimator and an infeasible estimator which assumes the knowledge of the group memberships $g_i^0, i = 1, \ldots, N$, defined as

$$\left(\widetilde{\boldsymbol{\alpha}}, \widetilde{\boldsymbol{\gamma}}, \widetilde{\mathbf{B}}\right) = \sup_{(\boldsymbol{\alpha},\boldsymbol{\gamma},\mathbf{B})} \frac{1}{NT} \sum_{i=1}^{N} \sum_{j=1}^{T} (y_{ij} - \alpha_i - \mathbf{z}_{ij}'\boldsymbol{\gamma} - \widehat{\mathbf{f}}_{ij}' \mathbf{b}_{g_i^0})^2. \tag{4}$$

THEOREM 2: *Let $a(n;r) = T^{1-r} + (\log n)^r m^{3r} n^{-r}$, where $n = NT$ and $r \geqslant 2$ is defined in the assumption (A3). Under the assumptions (A1)-(A6), we have*

(i) $\widehat{\boldsymbol{\alpha}} = \widetilde{\boldsymbol{\alpha}} + O_p(a(n;r))$, $\widehat{\boldsymbol{\gamma}} = \widetilde{\boldsymbol{\gamma}} + O_p(a(n;r))$, *and* $\widehat{\mathbf{B}} = \widetilde{\mathbf{B}} + O_p(a(n;r))$

(ii) $\mathbb{P}\left(\sup_{i=1,\ldots,N} |\widehat{g}_i - g_i^0| > 0\right) = o(1) + o(Na(n;r))$.

The convergence rate $a(n;r)$ has two parts. The first part T^{1-r} quantifies the observation error. The second part bounds the approximation error of FPCA to the functional covariate. Theorem 2 (i) establishes that the proposed estimator $(\widehat{\boldsymbol{\alpha}}, \widehat{\boldsymbol{\gamma}}, \widehat{\mathbf{B}})$ is equal to the infeasible estimator $(\widetilde{\boldsymbol{\alpha}}, \widetilde{\boldsymbol{\gamma}}, \widetilde{\mathbf{B}})$ up to an error of order $O_p(a(n;r))$. Under the assumption (A4), it is easy to see that $a(n;r)$ converges to zero as both N and T go to infinity. The proposed estimator and the infeasible estimator are thus asymptotically equivalent. Theorem 2 (ii) implies the consistency of the group membership estimator when $Na(n;r)$ goes to zero as both N and T go to infinity. Under the assumption (A4), the rate $Na(n;r)$ goes to zero if $(\log(NT))^{(\chi+2\nu)(r-1)/(\chi+2\nu-3)} NT^{1-r}$ converges to zero. Using the asymptotic equivalence established in Theorem 2, we can derive the asymptotic distribution of $\widehat{\boldsymbol{\gamma}}$ and the convergence rate of $\widehat{\beta}_g(t), t = 1, \ldots, G$, where $\widehat{\beta}_g(t) = \sum_{k=1}^{m} \widehat{b}_{g,k} \widehat{\phi}(t)$.

COROLLARY 1: *Suppose that the assumptions (A1)-(A7) hold. Let N and T go to infinity such that $N/T^{2r-3} \to 0$, then we have $\sqrt{n}(\widehat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}^0) \xrightarrow{d} N(\mathbf{0}, \Sigma)$, where $\Sigma = \mathrm{cov}(\widetilde{z})$. Furthermore, $\|\widehat{\beta}_g(t) - \beta_g^0(t)\| = O_p\left(n^{-\frac{2\nu-1}{2(\chi+2\nu)}}\right)$, for all $g = 1, \ldots, G$.*

Under the condition that the observations have certain moments and T is relatively large compared to N, i.e., $N/T^{2r-3} \to 0$, Corollary 1 shows that the proposed estimator for $\boldsymbol{\gamma}^0$ has the same asymptotic distribution as the infeasible estimator defined in (4). The convergence

rate of the estimator for the functional coefficients $\beta_g^0(t), g = 1, \ldots, G$, is the same as that obtained by Hall and Horowitz (2007) and Shin (2009). The convergence rate attains the minimax lower bound derived by Hall and Horowitz (2007).

Remark: In the plant data analysis, the number of families $N = 17$ is relatively small. Following the same arguments in the proofs we can show that all the theoretical results still hold with a finite N as long as T goes to infinity. The consistency of the parameters and the group memberships crucially depends on a large T but not a large N.

2.4 *Inference*

In Corollary 1 of Section 2.3, we established the asymptotic distribution of the estimator $\widehat{\boldsymbol{\gamma}}$. As in Shin (2009), we utilize the asymptotic distribution for constructing confidence intervals and conducting hypothesis testings. In practice, we estimate the asymptotic variance of $\widehat{\boldsymbol{\gamma}}$ as follows. Conditional on the estimated group structure, we denote $\widehat{\mathbf{F}}_i = (\widehat{f}_{i1}', \ldots, \widehat{f}_{iT}')', i = 1, \ldots, N$, and define $\widehat{\mathbf{F}}$ as a block matrix with the block at the ith row and kth column being $\widehat{\mathbf{F}}_i I\{\widehat{g}_i = k\}, i = 1, \ldots, N, k = 1, \ldots, G$. The asymptotic variance of $\widehat{\boldsymbol{\gamma}}$ is estimated as $\widehat{\sigma}^2 \left(\mathbf{Z}'(\mathbf{I}_{NT} - \mathbf{P}_{\widehat{\mathbf{F}}})\mathbf{Z}\right)^{-1}$, where $\mathbf{I}_{NT}$ is the identity matrix of size $NT \times NT$, $\mathbf{P}_{\widehat{\mathbf{F}}} = \widehat{\mathbf{F}}(\widehat{\mathbf{F}}'\widehat{\mathbf{F}})^{-1}\widehat{\mathbf{F}}'$, $\widehat{\sigma}^2 = 1/(NT)\sum_{i=1}^{N}\sum_{j=1}^{T}(\widehat{\epsilon}_{ij} - \bar{\widehat{\epsilon}})^2$, $\widehat{\epsilon}_{ij}, i = 1, \ldots, N, j = 1, \ldots, T$ are the residuals, and $\bar{\widehat{\epsilon}}$ is the mean of the residuals. Confidence intervals for $\boldsymbol{\gamma}^0$ can be formulated using the estimated variance and the normal approximation.

To construct confidence bands for the functional coefficients $\beta_g^0(t), g = 1, \ldots, G$, we adapt the strategy in Imaizumi and Kato (2019) to our model. Instead of constructing a confidence band that uniformly covers $\beta_g^0(t)$, a more practical strategy is constructing a band such that with probability at least $1 - \xi_1$, the proportion of $t \in I$ where $\beta_g^0(t)$ lies outside the band is less than a small number $\xi_2 \in (0, 1)$. This strategy is also adopted in Juditsky and Lambert-Lacroix (2003) for nonparametric function estimation. We first derive an asymptotic expansion of $\widehat{\beta}_g(t) - \beta_g^0(t), g = 1, \ldots, G$. Let $\mathbf{E} = (\widehat{\mathbf{F}}'\widehat{\mathbf{F}})^{-1}\widehat{\mathbf{F}}'(\boldsymbol{\epsilon} - \mathbf{Z}(\mathbf{Z}'(\mathbf{I}_{NT} - \mathbf{P}_{\widehat{\mathbf{F}}})\mathbf{Z})\mathbf{Z}'(\mathbf{I}_{NT} -$

$\mathbf{P}_{\widehat{\mathbf{F}}})\boldsymbol{\epsilon}$), and let $\mathbf{e}_g = (0, \ldots, 1, \ldots, 0), g = 1, \ldots, G$, be unit vectors of length G where the gth element is 1. Let $\mathbf{1}_m$ be a vector of length m with elements all equal to 1. Then

$$\|\widehat{\beta}_g - \beta_g^0\|^2 \approx (\mathbf{e}_g \otimes \mathbf{1}_m)'\mathbf{E}\mathbf{E}'(\mathbf{e}_g \otimes \mathbf{1}_m), \tag{5}$$

The approximation (5) is derived in the Web Appendix A. We use a simple bootstrap procedure to estimate the quantiles of $\|\widehat{\beta}_g - \beta_g^0\|^2, g = 1, \ldots, G$, based on (5). Let $\boldsymbol{\epsilon}_b^*, b = 1, \ldots, B$, be B bootstrap samples from the residuals $\widehat{\boldsymbol{\epsilon}}$, and let $\mathbf{E}_b^*$, be the corresponding bootstrap version of $\mathbf{E}$. Denote by $\widehat{c}_g(1-\xi_1)$ the $(1-\xi_1)$th sample quantile of $(\mathbf{e}_g \otimes \mathbf{1}_m)'\mathbf{E}_b^*\mathbf{E}_b'^*(\mathbf{e}_g \otimes \mathbf{1}_m), b = 1, \ldots, B$. As in Juditsky and Lambert-Lacroix (2003) and Imaizumi and Kato (2019), a confidence band that covers $1-\xi_2$ proportions of $t \in I$ of $\beta_g^0(t)$ is

$$\left[\widehat{\beta}_g(t) - \widehat{c}_g(1-\xi_1)\sqrt{\frac{1}{\xi_2\lambda(I)}}, \widehat{\beta}_g(t) + \widehat{c}_g(1-\xi_1)\sqrt{\frac{1}{\xi_2\lambda(I)}}\right], g = 1, \ldots, G, \tag{6}$$

where $\lambda(I)$ is the length of I. The confidence band (6) has a constant length. One shortcoming of the proposed confidence band is that the stochastic error in the group membership estimation is ignored. The resulting effect on empirical coverage probabilities is small as shown in the simulation studies in Section 3, especially when different groups are well separated and the sample size is large.

2.5 *Selection of G and m*

To implement the proposed estimator, we need to determine the number of groups G and the number of principal components m used in the truncation of the K-L expansion. We propose to select these parameters by minimizing a Bayesian Information Criterion,

$$\begin{aligned} BIC(m, G) = &\log\left(\frac{1}{NT}\sum_{i=1}^{N}\sum_{j=1}^{T}(y_{ij} - \widehat{\alpha}_i - \mathbf{z}_{ij}'\widehat{\boldsymbol{\gamma}} - \widehat{\mathbf{f}}_{ij}'\widehat{\mathbf{b}}_{\widehat{g}_i})^2\right) \\ &+ \frac{Gm + N + d}{NT}\log(NT). \end{aligned} \tag{7}$$

Note that the estimators $(\widehat{\boldsymbol{\alpha}}, \widehat{\boldsymbol{\gamma}}, \widehat{\mathbf{b}}, \widehat{\mathcal{G}})$ are understood to be a function of the parameters m and G implicitly. The BIC criterion balances the model fit and the parsimony of the parameters. In practice, we minimize the BIC criterion on a grid of the parameter values.

3. Simulation study

In this section, we investigate the finite sample performance of the proposed estimator through simulation studies. We simulated two settings with two latent groups, wherein setting one the functional coefficients depend on the first four principal components, and in setting two the generalized Fourier coefficients decrease smoothly. In each setting, we simulated five cases (C1)-(C5), such that the functional coefficients of the two groups deviate from each other gradually from the case (C1) to the case (C5). The detailed setup of the simulation study is presented in the Web Appendix B.1.

In the simulation, the sample size is $T = 40$ or $T = 60$, and $N = 25$, which is close to the sample size in the motivating barley data. The simulations are repeated 500 times for each case. The adjusted Rand index (RI) of Rand (1971) is used to assess similarities between the truth and the estimated groups. Larger values of RI indicate that the truth and the estimated groups are more similar, where 1 means complete recovery. We also report the percentage that the estimated number of groups equals the truth (P, in %) and the averaged number of estimated groups (M). The averaged number of selected PCs and the computing time are reported in the Web Table A.1.

For comparisons, we adapt the penalization methods of Ma and Huang (2017) to our setting with the SCAD and LASSO penalty. We also compare with a two-step naive K-means (NKMEAN) approach, where in the first step the family-specific functional coefficients are estimated, and in the second step, group membership is recovered by the K-means algorithms. The details of these methods are described in the Web Appendix B.1.

Generally, the proposed method performs the best in all the cases because it has a larger RI and P, and the averaged number of estimated groups (M) is closer to the truth (Table 1). For example, in case C5 of setting 1, RIs for the proposed method are 0.96 and 0.99 when the sample size is $T = 40$ and $T = 60$, respectively. In contrast, RIs are much smaller, 0.51,

0.32, and 0.77 for SCAD, LASSO, and NKMEAN, receptively, when $T = 40$, and increase to 0.73, 0.77, and 0.87 when T increases to 60 (Table 1). Moreover, the averaged number of estimated groups (M) for the proposed method is close to the truth (G=2). In contrast, the number of groups are overestimated for other methods.

[Table 1 about here.]

Next, we compare the estimation error of the proposed method with other methods. Besides the methods discussed above, we compare the proposed method with the pooled model (POOL), where the functional coefficients are constant across families, and the family-wise model (FAMILY), where each family has its own functional coefficient. Table 2 shows the integrated mean squared error (IMSE), the integrated variance (IVAR), and the integrated squared bias (IBIAS) of the functional coefficient for all the estimators. We first examine the estimation error for setting 1. In most cases, the proposed method has a smaller IMSE than all other methods. For example, IMSE is 0.075 for the proposed method when $T = 60$ in case C5, whereas IMSEs for all other methods are greater than 0.5 (Table 2, the last column). The proposed method has the smallest bias, compared with all the methods, whereas the pooled method has the smallest variance over all the methods (Table 2).

[Table 2 about here.]

Next, we examine the estimation error for setting 2. The proposed method performs the best in terms of IMSE and IBIAS, in all the cases. For example, IMSE is 0.281 for the proposed method in case C5 when $T = 40$, and decreases to 0.199 when T increases to 60. In contrast, IMSEs are larger, 0.322 and 0.790 for SCAD and LASSO, respectively, in case C5 when $T = 40$, and 0.217 and 0.614 when T increases to 60 (Table 2).

Lastly, we investigate the performance of the proposed confidence sets. For $\beta^0(t), t \in I$, we

define the empirical coverage probability of the confidence bands for $\beta_g^0(t), g = 1, 2$ as

$$\frac{1}{N}\sum_{i=1}^{N}\mathbb{P}\left(\lambda\left(t \in [0,1] : \beta_{g_i^0}^0(t) \in \widehat{\mathcal{C}}_{\widehat{g}_i}(t)\right) \leqslant 1 - \xi_2\right),$$

where $\widehat{\mathcal{C}}_{\widehat{g}_i}(t)$ is the confidence band for the functional coefficient of group $\widehat{g}_i$. The confidence level is 95% for both $\boldsymbol{\gamma}^0$ and $\beta_g^0(t), g = 1, 2$; we set $\xi_2 = 0.1$ for $\beta^0(t)$ as previously done by Imaizumi and Kato (2019). The data presented in Table 3 show that the empirical coverage percentages for $\boldsymbol{\gamma}^0$ are close to 95%, in all the cases. The empirical coverage percentages of the confidence band for $\beta_g^0(t), g = 1, 2$, are greater than 95%, except for cases C1 and C2 in which the groups are not well separated. The proposed confidence band for $\beta_g^0(t), g = 1, 2$, is conservative, as previously reported by Imaizumi and Kato (2019).

[Table 3 about here.]

4. Plant data analysis

In this section, we study the effect of salinity stress on the growth of barley plants. After data cleaning, the data from 17 barley families, and in total, 725 barley plants are available for analysis. For more details on the experiment design and how the data were collected, we refer the reader to Meng et al. (2017). Let $\omega_{ij}(t)$ denote the projected shoot area of barley j in family i at the tth day. The response is defined as $y_{ij} = \log(\omega_{ij}(32)) - \log(\omega_{ij}(19))$. The response measures the growth of the plant during the period when saltwater was applied. The relative growth rate $x_{ij}(t)$ is defined as $x_{ij}(t) = \omega'_{ij}(t)/\omega_{ij}(t), t \in [21, 30]$. We obtain $x_{ij}(t)$ by smoothing the discrete observations of $\omega_{ij}(t)$ using cubic splines. The model used in the analysis is $y_{ij} = \alpha_i + \mathrm{Na}_{ij}^{+}\gamma + \int x_{ij}(t)\beta_{g_i}(t)\mathbf{d}t + \epsilon_{ij}$, where Na^{+} is the concentration of Na^{+} in the leaves of the barley plants. In the literature, there is no evidence of heterogeneity for the effects of Na^{+} across barley families (Meng et al., 2017). The main interest lies in the growth pattern under salinity stress.

We apply the proposed algorithm to the data. By the BIC criterion, two groups of barley

families are detected, and four PCs are used in the estimation. The detected group 1 contains ten families of barley, and group 2 contains seven families. We also apply the penalization methods with the SCAD or LASSO penalty to the data, but these two methods do not detect any groups. The marginal distribution of the response shows two modes (Figure 1 (a)). The figures appear in color in the electronic version of this article, and any mention of color refers to that version. After fitting the proposed model, the residuals do not show a bi-modal structure anymore (Figure 1 (b)). The bi-modal structure of the response in Figure 1 (a) is well accommodated by the proposed latent group structure model. Figure 2 (b) shows the pooled estimate and the family-wise estimates of the functional coefficients. The effect of Na^+ is -0.135, with a standard error 0.128. The negative relationship of Na^+ is consistent with the findings reported in the literature (Munns and Tester, 2008).

[Figure 1 about here.]

The average final biomass is 1.04 and 1.34 for groups 1 and 2, respectively. The higher biomass of group 2 is reflected in the estimated functional coefficients. Overall, the estimated coefficient $\widehat{\beta}_2(t)$ of group 2 lies above the coefficient $\widehat{\beta}_1(t)$ of group 1 (Figure 2 (a)), which means that the growth of the barley plants in group 2 contributes more to the final biomass than that of the barley plants in group 1 at every stage under salinity stress.

To assess the prediction accuracy of the proposed method and other methods, we randomly split the data in each family 100 times, and use 90% of the data for training, 10% of the data for prediction. We compare the mean absolute prediction error (MAPE), which is defined as the mean absolute difference between the true and estimated responses. The average MAPE is 0.0695 for the proposed method, 0.0759, 0.1167, 0.0719, and 0.0718, respectively, for the pooled model, family-wise model, penalized method with the SCAD penalty and LASSO penalty. The standard error of MAPE is less than 0.0012 for all the methods. The proposed method has a lower prediction error than other methods.

[Figure 2 about here.]

In the following, we further explore differences between the detected barley groups. From Figure 3 (a), group 2 accumulates less Na^+ than group 1. The average Na^+ are 174 and 167 μmol per gram of dry mass for groups 1 and 2, respectively. The difference is significant, with p-values of the one-sided t-test and Wilcoxon rank sum test being 0.002, and 0.008, respectively. Accumulating less Na^+ means a higher ability for Na^+ exclusion, which is a main aspect of salinity tolerance in plants (Munns and Tester, 2008).

Next, we compare the barley groups in terms of osmotic tolerance. As suggested by Munns and Tester (2008), the growth reduction in the first few days after applying saltwater can be attributed to osmotic stress. Since $\widehat{\beta}_2(t)$ reaches a plateau around the 25th day, we define an index for osmotic tolerance as $\int_{21}^{25} x_{ij}(t)\widehat{\beta}_{\widehat{g}_i}(t)\mathbf{d}t$. The barley families in group 2 exhibit a higher osmotic tolerance than those in group 1 (Figure 3 (b)), because the contribution of growth in the first four days for group 2 is higher than that of group 1, with a mean of 0.21 and 0.36 for groups 1 and 2, respectively. The above analyses suggest that the barley plants in group 2 have a better salinity tolerance ability than that of group 1.

[Figure 3 about here.]

5. Conclusion

Motivated by the barley salinity tolerance study, we proposed a functional partially linear regression model with latent group structures. The proposed method provides a new tool to account for heterogeneity in functional linear regression. The developed algorithm is accurate for recovering the latent groups compared to competing methods in simulation studies. This paper considers latent group structures in the conditional mean of the response given the covariates. It is interesting to extend the current methodology to model discrete responses, e.g., counts or binary data, using the generalized linear model. Another promising extension

is to consider conditional quantiles of the response rather than the conditional mean, which might give a more complete description of the heterogeneity in the conditional distribution of the response.

Acknowledgments

The authors would like to thank Professor Mark Tester for providing the data. Y. Sun and W. Wang would like to thank the funding from King Abdullah University of Science and Technology, Thuwal, Saudi Arabia. H. Wang would like to acknowledge the IR/D program from the US National Science Foundation (NSF) and the NSF grant DMS-1712760. Any opinion, findings, and conclusions, or recommendations expressed in this material are those of the authors and do not necessarily the views of the NSF.

Data Availability Statement

The data that support the findings in this paper were obtained from The Plant Accelerator®, King Abdullah University of Science and Technology, Thuwal, Saudi Arabia. The data are not publicly available due to privacy or ethical restrictions.

References

Aloise, D., Deshpande, A., Hansen, P., and Popat, P. (2009). NP-hardness of Euclidean sum-of-squares clustering. *Machine Learning* **75,** 245–248.

Bonhomme, S. and Manresa, E. (2015). Grouped patterns of heterogeneity in panel data. *Econometrica* **83,** 1147–1184.

Cardot, H., Ferraty, F., and Sarda, P. (2003). Spline estimators for the functional linear model. *Statistica Sinica* **13,** 571–591.

Ciarleglio, A., Petkova, E., Ogden, T., and Tarpey, T. (2018). Constructing treatment

decision rules based on scalar and functional predictors when moderators of treatment effect are unknown. *Journal of the Royal Statistical Society. Series C* **67,** 1331.

Crambes, C., Kneip, A., and Sarda, P. (2009). Smoothing splines estimators for functional linear regression. *The Annals of Statistics* **37,** 35–72.

Hall, P. and Horowitz, J. (2007). Methodology and convergence rates for functional linear regression. *The Annals of Statistics* **35,** 70–91.

Hall, P. and Hosseini-Nasab, M. (2006). On properties of functional principal components analysis. *Journal of the Royal Statistical Society: Series B* **68,** 109–126.

Hall, P. and Hosseini-Nasab, M. (2009). Theory for high-order bounds in functional principal components analysis. In *Mathematical Proceedings of the Cambridge Philosophical Society*, volume 146, pages 225–256. Cambridge University Press.

Hansen, P. and Mladenovic, N. (2001). J-MEANS: a new local search heuristic for minimum sum of squares clustering. *Pattern Recognition* **34,** 405–413.

Imaizumi, M. and Kato, K. (2019). A simple method to construct confidence bands in functional linear regression. *Statistica Sinica* **29,** 2055–2081.

Juditsky, A. and Lambert-Lacroix, S. (2003). Nonparametric confidence set estimation. *Mathematical Methods of Statistics* **12,** 410–428.

Kato, K. (2012). Estimation in functional linear quantile regression. *The Annals of Statistics* **40,** 3108–3136.

Ke, Y., Li, J., and Zhang, W. (2016). Structure identification in panel data analysis. *The Annals of Statistics* **44,** 1193–1233.

Kong, D., Ibrahim, J., Lee, E., and Zhu, H. (2018). FLCRM: Functional linear Cox regression model. *Biometrics* **74,** 109–117.

Kong, D., Staicu, A.-M., and Maity, A. (2016). Classical testing in functional linear models. *Journal of Nonparametric Statistics* **28,** 813–838.

Kong, D., Xue, K., Yao, F., and Zhang, H. H. (2016). Partially functional linear regression in high dimensions. *Biometrika* **103,** 147–159.

Ma, S. and Huang, J. (2017). A concave pairwise fusion approach to subgroup analysis. *Journal of the American Statistical Association* **112,** 410–423.

Meng, R., Saade, S., Kurtek, S., Berger, B., Brien, C., Pillen, K., Tester, M., and Sun, Y. (2017). Growth curve registration for evaluating salinity tolerance in barley. *Plant Methods* **13,** 1–18.

Müller, H.-G. and Stadtmüller, U. (2005). Generalized functional linear models. *The Annals of Statistics* **33,** 774–805.

Munns, R. and Tester, M. (2008). Mechanisms of salinity tolerance. *Annual Review of Plant Biology* **59,** 651–681.

Ordin, B. and Bagirov, A. M. (2015). A heuristic algorithm for solving the minimum sum-of-squares clustering problems. *Journal of Global Optimization* **61,** 341–361.

Ramsay, J. O. and Silverman, B. (2005). *Functional Data Analysis*. Wiley Online Library.

Rand, W. M. (1971). Objective criteria for the evaluation of clustering methods. *Journal of the American Statistical Association* **66,** 846–850.

Scheipl, F. and Greven, S. (2016). Identifiability in penalized function-on-function regression models. *Electronic Journal of Statistics* **10,** 495–526.

Shin, H. (2009). Partial functional linear regression. *Journal of Statistical Planning and Inference* **139,** 3405–3418.

Wang, H. J., Zhu, Z., and Zhou, J. (2009). Quantile regression in partially linear varying coefficient models. *The Annals of Statistics* **37,** 3841–3866.

Wang, J., Li, J., Li, Y., and Wong, W. K. (2019). A model-based multithreshold method for subgroup identification. *Statistics in Medicine* **38,** 2605–2631.

Wang, S., Huang, M., Wu, X., and Yao, W. (2016). Mixture of functional linear models and

its application to CO2-GDP functional data. *Computational Statistics & Data Analysis* **97,** 1–15.

Xu, Y., Li, Y., and Nettleton, D. (2018). Nested hierarchical functional data modeling and inference for the analysis of functional plant phenotypes. *Journal of the American Statistical Association* **113,** 593–606.

Yao, F., Fu, Y., and Lee, T. C. (2011). Functional mixture regression. *Biostatistics* **12,** 341–353.

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

Yu, D., Kong, L., and Mizera, I. (2016). Partial functional linear quantile regression for neuroimaging data analysis. *Neurocomputing* **195,** 74–87.

Yuan, M. and Cai, T. (2010). A reproducing kernel Hilbert space approach to functional linear regression. *The Annals of Statistics* **38,** 3412–3444.

Zhang, Y., Wang, H. J., and Zhu, Z. (2019). Quantile-regression-based clustering for panel data. *Journal of Econometrics* **213,** 54–67.

Supporting Information

Web Appendix A and B, referenced in Section 3 and 4, are available with this paper at the Biometrics website on Wiley Online Library. An R package for the proposed estimator is available at `https://github.com/wangwustat/fdagroup`. The barley data are not publicly available due to privacy or ethical restrictions.

Received October 2020. *Revised February* 2020. *Accepted March* 2020.

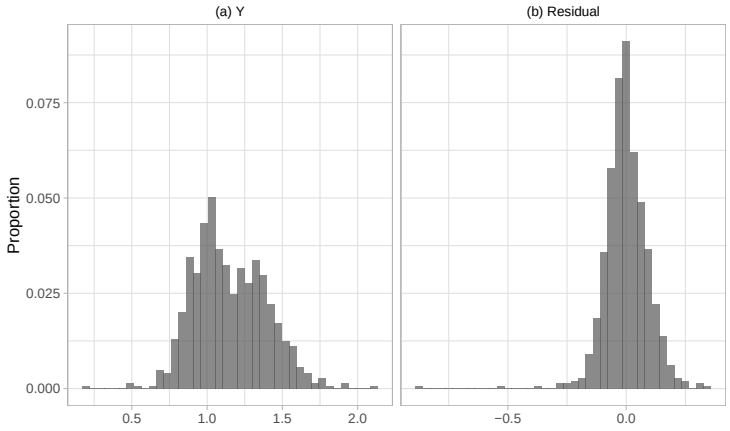

Figure 1. (a), the histogram of the response Y. (b), the histogram of the residuals from fitting the proposed model. This figure appears in color in the electronic version of this article, and any mention of color refers to that version.

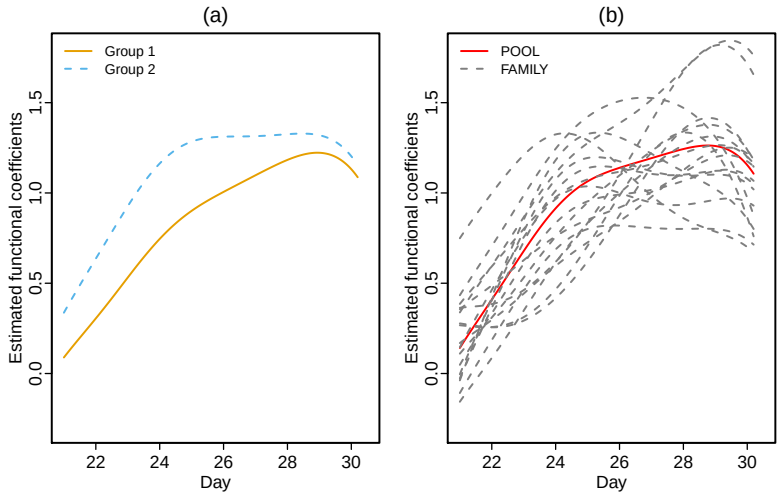

Figure 2. (a), the estimated functional coefficients of the two groups of barley. (b), the functional coefficients estimated by the family-wise model (FAMILY) and the pooled model (POOL). This figure appears in color in the electronic version of this article, and any mention of color refers to that version.

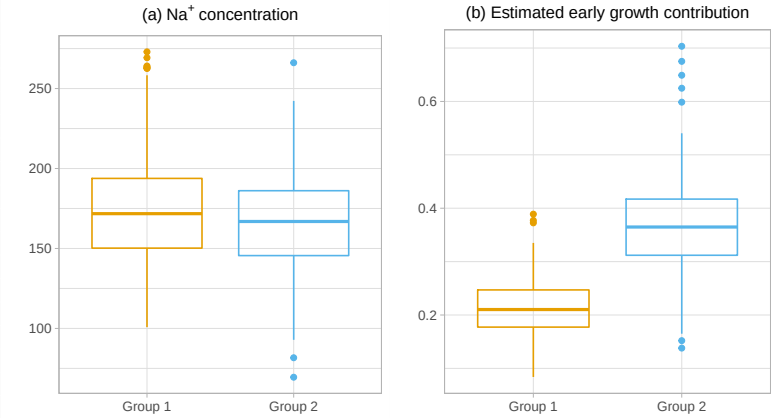

Figure 3. (a), the boxplot of Na^+ concentration of the two groups. (b), the boxplot of $\int_{21}^{25} x_{ij}(t)\widehat{\beta}_{\widehat{g}_i}(t)\mathbf{d}t$ of the two groups as an index for osmotic tolerance. This figure appears in color in the electronic version of this article, and any mention of color refers to that version.

Table 1

The rand index (RI), the percentage that the estimated number of groups equals to the truth (P, in %), and the averaged number of estimated groups (M). C1-C5 refer to cases 1-5.

			$T = 40$					$T = 60$				
			C1	C2	C3	C4	C5	C1	C2	C3	C4	C5
Setting 1	RI	Proposed	0.46	0.65	0.82	0.92	0.96	0.61	0.82	0.94	0.98	0.99
		SCAD	0.00	0.00	0.03	0.17	0.51	0.00	0.01	0.13	0.52	0.73
		LASSO	0.00	0.00	0.02	0.10	0.32	0.00	0.01	0.11	0.47	0.77
		NKMEAN	0.34	0.50	0.63	0.73	0.77	0.48	0.68	0.78	0.84	0.87
	M	Proposed	2.01	2.09	2.05	2.02	2.02	2.05	2.01	2.01	2.01	2.01
		SCAD	1.01	1.04	1.16	2.09	4.18	1.02	1.05	1.45	3.61	4.72
		LASSO	1.00	1.00	1.04	1.18	1.79	1.01	1.03	1.20	1.80	2.85
		NKMEAN	2.54	2.60	2.65	2.68	2.70	2.44	2.42	2.49	2.51	2.48
	P	Proposed	89.0	90.0	95.4	97.6	97.6	93.8	98.8	98.8	99.4	99.2
		SCAD	0.8	2.6	7.0	7.2	7.0	1.8	3.6	8.4	16.4	10.8
		LASSO	0.2	0.0	3.4	7.6	18.8	0.6	2.2	9.8	32.0	34.6
		NKMEAN	50.0	43.4	42.8	42.2	44.0	57.6	60.8	55.4	56.0	58.0
Setting 2	RI	Proposed	0.59	0.79	0.88	0.94	0.94	0.78	0.91	0.96	0.97	0.98
		SCAD	0.05	0.31	0.65	0.81	0.92	0.09	0.51	0.75	0.90	0.97
		LASSO	0.02	0.17	0.45	0.50	0.47	0.06	0.45	0.66	0.64	0.62
		NKMEAN	0.45	0.58	0.65	0.68	0.70	0.60	0.70	0.73	0.73	0.73
	M	Proposed	2.24	2.21	2.24	2.19	2.24	2.12	2.10	2.08	2.09	2.07
		SCAD	1.56	3.23	4.76	3.98	2.91	1.62	4.38	4.68	3.22	2.41
		LASSO	1.24	2.44	4.96	6.95	7.35	1.28	3.16	5.10	5.65	5.79
		NKMEAN	2.78	2.96	3.12	3.21	3.24	2.76	2.91	3.03	3.13	3.15
	P	Proposed	78.0	81.8	79.2	83.6	82.0	89.0	90.2	93.2	92.0	94.6
		SCAD	6.8	8.0	10.0	19.2	45.8	9.6	8.8	11.4	32.6	68.0
		LASSO	2.8	8.6	11.6	4.2	1.2	10.4	16.2	12.4	7.6	6.0
		NKMEAN	25.8	14.8	13.8	13.2	13.2	31.2	25.6	23.8	18.6	18.4

Table 2

The integrated mean square error(×10), integrated variance (×10), and integrated bias (×10) for the functional coefficients for setting 1. C1-C5 refer to cases 1-5.

			$T = 40$					$T = 60$				
			C1	C2	C3	C4	C5	C1	C2	C3	C4	C5
Setting 1	IMSE	Proposed	3.01	3.02	2.43	1.86	1.62	1.75	1.33	0.97	0.80	0.75
		SCAD	2.54	4.15	6.24	8.88	10.06	2.29	3.82	5.71	7.25	5.79
		LASSO	2.59	4.17	6.12	8.52	11.29	2.30	3.81	5.67	7.35	8.69
		FAMILY	19.15	18.41	17.63	17.27	17.17	14.53	14.83	15.07	14.87	14.83
		POOL	2.59	4.16	6.14	8.59	11.52	2.32	3.90	5.90	8.33	11.24
		NKMEAN	17.65	16.89	16.05	15.38	15.29	13.04	13.27	13.51	13.23	13.12
	IVAR	Proposed	2.86	2.91	2.37	1.84	1.61	1.69	1.31	0.96	0.80	0.74
		SCAD	0.57	0.69	1.10	3.60	5.84	0.36	0.45	1.18	4.25	3.37
		LASSO	0.62	0.67	0.73	0.97	1.94	0.36	0.39	0.60	1.11	1.89
		FAMILY	4.83	4.59	4.22	4.36	4.60	5.55	4.74	3.86	3.50	3.07
		POOL	0.60	0.61	0.61	0.62	0.67	0.33	0.36	0.37	0.37	0.40
		NKMEAN	3.17	3.03	2.60	2.46	2.65	3.99	3.16	2.31	1.82	1.36
	IBIAS	Proposed	0.15	0.11	0.05	0.02	0.02	0.06	0.02	0.01	0.00	0.00
		SCAD	1.96	3.46	5.13	5.26	4.19	1.93	3.37	4.51	3.00	2.41
		LASSO	1.97	3.49	5.39	7.54	9.36	1.94	3.41	5.07	6.24	6.80
		FAMILY	14.33	13.87	13.41	12.89	12.63	8.97	10.10	11.20	11.37	11.75
		POOL	1.99	3.54	5.53	7.97	10.85	1.99	3.54	5.53	7.97	10.84
		NKMEAN	14.48	13.87	13.45	12.92	12.65	9.05	10.11	11.21	11.41	11.79
Setting 2	IMSE	Proposed	4.20	3.63	3.30	2.91	2.81	2.96	2.55	2.30	2.11	1.99
		SCAD	5.01	6.66	5.96	4.24	3.22	4.38	5.16	3.90	2.79	2.17
		LASSO	4.74	6.61	7.95	8.05	7.90	4.30	5.74	6.28	6.23	6.14
		FAMILY	10.73	10.29	9.72	9.18	8.79	8.98	8.42	8.39	8.04	7.84
		POOL	4.73	6.69	9.17	12.28	16.05	4.43	6.33	8.88	11.96	15.67
		NKMEAN	10.42	10.02	9.35	8.73	8.26	8.59	7.91	7.89	7.55	7.33
	IVAR	Proposed	2.39	1.90	1.64	1.36	1.32	1.39	1.06	0.86	0.80	0.64
		SCAD	1.55	3.28	2.56	1.60	1.22	1.18	2.22	1.02	0.90	0.72
		LASSO	1.10	1.89	2.36	1.98	1.71	0.87	1.49	1.52	1.55	1.47
		FAMILY	2.74	2.40	2.20	2.06	1.94	2.65	2.44	2.10	1.87	1.69
		POOL	0.88	0.90	0.83	0.79	0.85	0.74	0.67	0.61	0.65	0.57
		NKMEAN	2.44	2.11	1.83	1.61	1.42	2.23	1.95	1.59	1.38	1.17
	IBIAS	Proposed	1.81	1.74	1.64	1.56	1.50	1.56	1.48	1.43	1.31	1.35
		SCAD	3.42	3.38	3.40	2.64	2.00	3.21	2.93	2.89	1.90	1.45
		LASSO	3.71	4.77	5.59	6.07	6.18	3.50	4.29	4.76	4.68	4.66
		FAMILY	7.97	7.88	7.52	7.11	6.85	6.34	5.98	6.30	6.17	6.15
		POOL	3.88	5.82	8.34	11.49	15.20	3.73	5.68	8.26	11.31	15.10
		NKMEAN	8.01	7.93	7.51	7.11	6.85	6.37	5.96	6.30	6.17	6.16

Table 3

The empirical coverage percentages (%) of the proposed 95% confidence interval for $\boldsymbol{\gamma}^0$ and the confidence band for the functional coefficients $\beta_g^0(t), g = 1, 2$.

	Setting 1				Setting 2			
	$T = 40$		$T = 60$		$T = 40$		$T = 60$	
Case	$\boldsymbol{\gamma}^0$	$\beta_g^0(t)$	$\boldsymbol{\gamma}^0$	$\beta_g^0(t)$	$\boldsymbol{\gamma}^0$	$\beta_g^0(t)$	$\boldsymbol{\gamma}^0$	$\beta_g^0(t)$
C1	93.0	85.9	94.4	89.7	94.1	93.1	94.4	95.2
C2	93.2	89.7	94.4	95.2	93.5	96.3	94.5	98.4
C3	94.5	94.9	95.0	98.3	94.2	98.3	94.1	99.3
C4	94.5	97.8	94.6	99.5	94.2	99.4	94.5	99.8
C5	94.8	99.1	93.9	99.6	94.5	99.7	95.5	99.8