Statistica Sinica **36** (2026), 463-483
doi:https://doi.org/10.5705/ss.202023.0091

FUNCTIONAL ADAPTIVE DOUBLE-SPARSITY ESTIMATOR FOR FUNCTIONAL LINEAR REGRESSION MODEL WITH MULTIPLE FUNCTIONAL COVARIATES

Cheng Cao[1], Jiguo Cao[2], Hailiang Wang[3], Kwok-Leung Tsui[4] and Xinyue Li*[1]

[1] *City University of Hong Kong,* [2] *Simon Fraser University,* [3] *The Hong Kong Polytechnic University and* [4] *Virginia Polytechnic Institute and State University*

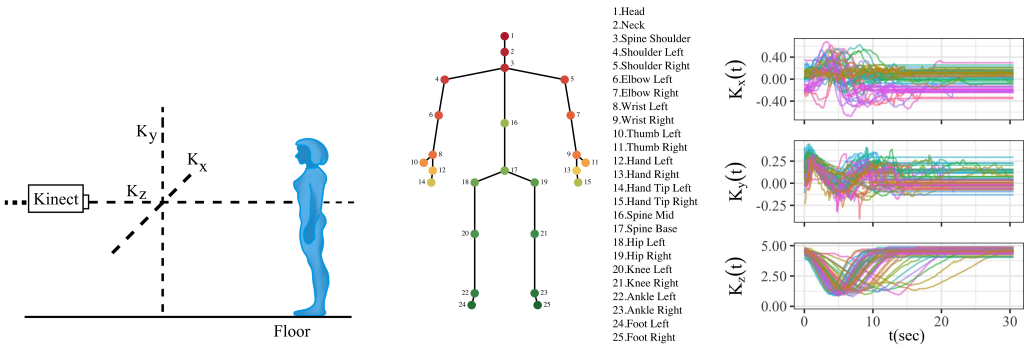
Abstract: Sensor devices have been increasingly used in engineering and health studies recently, and the captured multi-dimensional activity and vital sign signals can be studied in association with health outcomes to inform public health. The common approach is the scalar-on-function regression model, in which health outcomes are the scalar responses while high-dimensional sensor signals are the functional covariates, but how to effectively interpret results becomes difficult. In this study, we propose a new Functional Adaptive Double-Sparsity (FadDoS) estimator based on functional regularization of sparse group Lasso with multiple functional predictors, which can achieve global sparsity via functional variable selection and local sparsity via zero-subinterval identification within coefficient functions. We prove that the FadDoS estimator converges at a bounded rate and satisfies the oracle property under mild conditions. Extensive simulation studies confirm the theoretical properties and exhibit excellent performances compared to existing approaches. Application to a Kinect sensor study that utilized an advanced motion sensing device tracking human multiple joint movements and conducted among community-dwelling elderly demonstrates how the FadDoS estimator can effectively characterize the detailed association between joint movements and physical health assessments. The proposed method is not only effective in Kinect sensor analysis but also applicable to broader fields, where multi-dimensional sensor signals are collected simultaneously, to expand the use of sensor devices in health studies and facilitate sensor data analysis.

Key words and phrases: Kinect sensor, scalar-on-function regression, sensor device data, sparse group Lasso.

1. Introduction

Wearable devices have been transforming healthcare, medical, and rehabilitation fields through continuous monitoring of physical activity and vital signs. With recent advances in sensing technology, a novel sensor device called

*Corresponding author. E-mail: xinyueli@cityu.edu.hk



(a) (b) (c)

Figure 1. (a) Setup of the Kinect sensor. (b) The 25 joints of human skeleton captured by the Kinect sensor. (c) Three-dimensional displacement of the sacrum (joint 17: Spine Base) for all participants in the Timed Up and Go (TUG) Test. Each curve represents the time-dependent sacrum position of a participant during the entire test. Due to the different durations of finishing, we used the last observation to pad the shorter signals until all signals have equal time domain. Participant IDs are color-coded. Top panel: Horizontal direction. Middle panel: Vertical direction. Bottom panel: Walking direction.

Kinect has been increasingly used to perform precise three-dimensional movement tracking of 25 skeleton joints, which can unlock the potential of automated health monitoring and enable enhanced healthcare decisions (Gasparrini et al., 2014; Cippitelli et al., 2015; Kohout et al., 2021). In an elderly mobility study conducted in Hong Kong local communities, we collected multi-joint movements using Kinect sensor among the community-dwelling elderly to evaluate their dynamic balance capabilities. Figure 1(a) displays the setup of the Kinect camera and its three-dimensional coordinates for motion capturing. Figure 1(b) depicts the set of all skeletal joints captured. Figure 1(c) provides an example of the captured movement of the sacrum at the base of the spine for 50 participants. The three-dimensional displacement signals $K_x(t)$, $K_y(t)$ and $K_z(t)$ denote the horizontal direction representing swaying motion, the vertical direction representing up-and-down motion, and the walking direction respectively. The sensor records displacement data, from which velocity, and frequency domain data can be calculated to enrich the Kinect sensor dataset. Accurate multi-joint movements provide unprecedented opportunities to analyze the relationship between joint movement patterns and physical health assessments.

However, the analysis of the complex high-dimensional Kinect data, same as other types of multi-dimensional sensor device data, remains challenging. Given a large number of functional covariates and measured scalar responses, we consider scalar-on-function regression model with multiple functional covariates (Ramsay and Silverman, 2005). The estimates of the coefficient functions can yield insights into the detailed associations between functional covariates and the outcome

over the time domain. To enhance the interpretability of the complex model, it is crucial to conduct functional variable selection and further obtain sparse coefficient estimates, which we refer to as global sparsity and local sparsity respectively, to obtain meaningful results and relevant interpretation of the associations.

Several estimators targeting scalar-on-function regression model with multiple functional covariates address global sparsity to achieve variable selection (Pannu and Billor, 2017; Collazos, Dias and Zambom, 2016; Cheng, Shi and Eyre, 2020). Gertheiss, Maity and Staicu (2013) proposed a sparsity-smoothness estimator to obtain a parsimonious model based on functional generalization of group Lasso penalty (Meier, Van de Geer and Bühlmann, 2009), while Matsui and Konishi (2011) and Mingotti, Lillo and Romo (2013) suggested alternative functional regularizations. However, these estimators without concerning local sparsity can only facilitate the interpretability with respect to variable selection. With increasing attention to coefficient interpretability, more methods aim to preserve local sparsity of functional coefficient (James, Wang and Zhu, 2009; Zhou, Wang and Wang, 2013). The sparse estimation of functional coefficient is reformulated to parametric variable selection in basis function representation of linear regression with various regularizations. Wang and Kai (2015) and Lin et al. (2017) proposed an estimator based on the group bridge and functional generalization of SCAD penalty, respectively. These approaches can identify zero subregions with nice asymptotic properties, but both primarily target a single functional covariate. Tu et al. (2012) and Tu, Park and Wang (2020) extended the group bridge penalty to multiple functional covariates with binary and functional responses rather than scalar responses, while the suboptimal local solution due to the non-convex penalty could be a concern.

In this paper, we propose a novel functional adaptive double-sparsity (FadDoS) estimator, which can achieve global sparsity via functional variable selection and local sparsity via sparse coefficient estimation simultaneously in scalar-on-function regression models with multiple functional covariates. The combination of global and local sparsity is termed double-sparsity. The proposed estimator achieves double-sparsity through functional generalization of sparse group Lasso and adaptive penalization. By solving an optimization problem that incorporates regularization and smoothing splines in a single convex objective function, we can provide the estimates of coefficient functions with nice double-sparsity and smoothness control. The benefit of convexity is the guarantee of converging to a local minimum during model fitting.

The major contribution of our work is three-fold. First, we defined an important property, double-sparsity, of functional linear regression with multiple functional covariates, which leads to sparse and consistent functional estimates when not all functional covariates over the entire time domain are associated with the scalar response. Second, we provided a one-stage estimation procedure for the

FadDoS estimator to achieve double-sparsity, which is computationally efficient and easy to implement. Third, the accuracy of the estimation is theoretically stable and solid as we prove that the FadDoS estimator converges at an optimal rate and enjoys oracle properties. Our method can remarkably enhance model parsimony and interpretability for high-dimensional sensor device data that existing methods in the literature lack. As an example, the FadDoS estimator can be applied to complex Kinect sensor data to effectively examine the association between joint movements and mobility, based on which automated assessment tools can be further developed for health monitoring and early intervention.

The rest of this paper is organized as follows. In Section 2, we introduce the double-sparsity structure of the scalar-on-function regression model with multiple functional covariates and computation of the FadDoS estimator. Section 3 presents asymptotic properties of the proposed estimator. Extensive simulation studies are conducted in Section 4 to evaluate the performance of our estimator and compare the results with existing approaches. Section 5 describes the application for the Kinect elderly study. Conclusion and discussion are in Section 6. All code for model implementation and simulation is available in the Supplementary Material and at `https://github.com/Cheng-0621/FadDoS`.

2. Methodology

2.1. Global and local sparsity in scalar-on-function regression model with multiple functional covariates

For the ith subject, suppose that Y_i is the scalar response and $X_{ij}(t) \in L^2(\mathcal{T})$ be the jth functional covariate. The functional linear model with multiple functional covariates is thus

$$Y_i = \mu + \sum_{j=1}^{J} \int_{\mathcal{T}} X_{ij}(t)\beta_j(t)dt + \epsilon_i, \quad i = 1, \ldots, n, \tag{2.1}$$

where μ is the intercept term and $\epsilon_i \sim N(0, \sigma^2)$. The smooth coefficient function $\beta_j(t)$ provides a time-dependent contribution of $X_{ij}(t)$ to the corresponding response Y_i.

We use the following notation: for all measurable functions f whose absolute value raised to the q-th power has a finite integral, which is $||f||_q = (\int_{\mathcal{T}} |f(t)|^q dt)^{1/q} < \infty$, and the supremum norm of a function f is defined as $||f||_\infty = \sup_{t\in\mathcal{T}} |f(t)|$. The double-sparsity structure of the functional linear model (2.1) contains both global sparsity and local sparsity. Mathematically, the global sparsity is characterized by the set of indices for which functional covariates have no contribution to the response over the entire domain, i.e. $\{j \in \{1, \ldots, J\} : \beta_j(t) = 0, \text{for all } t \in \mathcal{T}\}$, while for those functional covariates not belonging to this set, the local sparsity is depicted by the partition of zero and

nonzero subregion of the corresponding coefficient functions. Suppose a subregion $I \subset \mathcal{T}$, the zero subregion is defined $N_0(\beta_j) = \{I \subset \mathcal{T} : \beta_j(t) = 0, \text{for all } t \in I\}$ and thus the nonzero subregion $N_1(\beta_j)$ is the complement of $N_0(\beta_j)$. It implies that within the subinterval $N_0(\beta_j)$, the functional covariate has no effect on the response. With the double-sparsity structure, important functional predictors can be selected and nonzero subintervals can be further identified.

2.2. Functional sparse group Lasso penalization

To achieve both global and local sparsity, we propose a novel regularization method based on functional generalization of sparse group Lasso (fSGL), where SGL is initially proposed by Friedman, Hastie and Tibshirani (2010) and Simon et al. (2013). The penalty function is

$$P_{\lambda_1,\lambda_2,\varphi}(\boldsymbol{\beta}) = \lambda_1||\boldsymbol{\beta}||_1 + \lambda_2 \sum_{j=1}^{J} \left(||\beta_j||_2^2 + \varphi||\mathcal{D}^m\beta_j||_2^2\right)^{1/2}, \tag{2.2}$$

where $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_J)^T$ is the vector of coefficient functions, $||\boldsymbol{\beta}||_1 = \sum_{j=1}^{J} ||\beta_j||_1$ with $||\beta_j||_1 = \int_{\mathcal{T}} |\beta_j(t)|dt$, $||\beta_j||_2^2 = \int_{\mathcal{T}} |\beta_j(t)|^2 dt$ and $||\mathcal{D}^m\beta_j||_2^2 = \int_{\mathcal{T}} |\partial^m\beta_j(t)/\partial t^m|^2 dt$ such that $m \leq d$, where d is the degree of polynomial functions, $\lambda_1, \lambda_2, \varphi \geq 0$. The additive penalties contain two convex regularizations. The functional generalization of ℓ_1 (functional ℓ_1) penalty regularizes the magnitude of local sparseness, while functional generalization of $\ell_{1,2}$ (functional $\ell_{1,2}$) penalty manages global sparseness. Within functional $\ell_{1,2}$ penalty, the roughness penalty $\varphi||\mathcal{D}^m\beta_j||_2$ is employed to apply penalization to the m-th order derivative of the jth coefficient function to control the smoothness. It is worth noting that achieving this goal is also possible by placing $\varphi||\mathcal{D}^m\beta_j||_2$ outside the square root, whereas our functional $\ell_{1,2}$ penalty offers an advantage of computational efficiency for solving a simpler optimization problem, same as discussed in Meier, Van de Geer and Bühlmann (2009). Thus, the minimizer of mean squared loss function regularized by fSGL in Equation (2.2) is called functional double-sparsity (FDoS) estimator.

However, one drawback of this regularization method is that the magnitude of penalization is the same for all functional covariates in the model, which may result in either insufficient suppression of zero functional estimates or underestimation of nonzero functional estimates. Theoretically, the equal degrees of penalization fail to satisfy the oracle property suggested in Fan and Li (2001), unless a proper adaptive method is applied to enable importance re-weighting of different coefficient functions (Zou, 2006).

2.3. Functional adaptive double-sparsity estimator

We further enhance the fSGL penalization with adaptive weights to allow for flexible regularization to different coefficient functions. The adaptive penalty

function is

$$P_{\lambda_1,\lambda_2,\varphi}(\boldsymbol{\beta}) = \lambda_1 \sum_{j=1}^{J} w_j^{(1)} ||\beta_j||_1 + \lambda_2 \sum_{j=1}^{J} w_j^{(2)} \left(||\beta_j||_2^2 + \varphi ||\mathcal{D}^m \beta_j||_2^2 \right)^{1/2}, \tag{2.3}$$

where $w_j^{(1)} = ||\check{\beta}_j||_1^{-a}$ and $w_j^{(2)} = ||\check{\beta}_j||_2^{-a}$ are known nonnegative weights and $\check{\beta}_j$ is the initial estimator, $a > 0$ is an adjustment of the adaptive weights and usually $a = 1$.

Intuitively, the adaptive weights can incorporate prior insights into functional covariates: a larger value implies less importance in the model. The adaptive approach can effectively penalize zero functions to reduce false negatives without over-regularizing nonzero functions. Additionally, the adaptive weights in Equation (2.3) are predictor-wise rather than subregion-wise, i.e., $||w_j^{(1)}\beta_j||_1$ with known nonnegative weight function $w_j^{(1)}(t) = |\check{\beta}_j(t)|^{-a}$. We do not choose subregion-wise weight because it is computationally intensive and more importantly, prone to errors, as the coefficient estimation is largely affected by choices of initial estimators. While the adaptive estimation allows for different initial estimators, we choose the smoothing spline estimator as in Cardot, Ferraty and Sarda (2003) with the second-order differential operator in the penalty derived from functional generalization of ridge penalization with generally no sparse solution, to avoid explosion to infinity when calculating the reciprocal of the norm even for those less significant functional covariates.

Given the adaptive fSGL penalization in Equation (2.3), we propose the FadDoS estimator, which minimizes the penalized mean squared loss defined by

$$\begin{aligned} L_n(\boldsymbol{\beta}, \mu) =& \frac{1}{2} \sum_{i=1}^{n} \left\{ Y_i - \mu - \sum_{j=1}^{J} \int_{\mathcal{T}} X_{ij}(t)\beta_j(t)dt \right\}^2 \\ &+ \lambda_1 \sum_{j=1}^{J} w_j^{(1)} ||\beta_j||_1 + \lambda_2 \sum_{j=1}^{J} w_j^{(2)} \left(||\beta_j||_2^2 + \varphi ||\mathcal{D}^m \beta_j||_2^2 \right)^{1/2}. \end{aligned} \tag{2.4}$$

It is obvious that FDoS is a special case of FadDoS if fixing $w_j^{(1)} = w_j^{(2)} = 1$ for all $j = 1, \ldots, J$. Moreover, $m = 2$ is commonly chosen and we adopt it for the rest of the paper.

The computational procedure of the FadDoS estimator is shown below. We first express the coefficient functions $\beta_j(t)$ with B-splines. Suppose that there are $M_n + 1$ equally spaced knots $0 = t_{j0} < t_{j1} < \cdots < t_{jM_n} = T$ in the domain $\mathcal{T}$ and d degree of polynomial functions, and let $\mathcal{S}_j$ be the linear space spanned by basis functions $\{B_{jk}(t) : k = 1, \ldots, M_n + d\}$ in $\mathcal{T}$, we have $\beta_j(t) = \sum_{k=1}^{M_n+d} B_{jk}(t)b_{jk} = \boldsymbol{B}_j^T(t)\boldsymbol{b}_j$, where $\boldsymbol{B}_j(t) = (B_{j1}(t), \ldots, B_{j(M_n+d)}(t))^T$, $\boldsymbol{b}_j = (b_{j1}, \ldots, b_{j(M_n+d)})^T$. With the compact support property of B-splines, the basis function can characterize the sparseness of coefficient functions as it is a

nonzero polynomial over no more than $d+1$ adjacent subintervals $[t_{j(r-d-1)}, t_{jr})$, $r \in \{1, \ldots, M_n\}$, implying only $d+1$ basis functions have supports within a subinterval $[t_{j(r-1)}, t_{jr})$.

Using B-spline basis functions, we reparameterize Equation (2.4) to facilitate computation. First, $||\beta_j||_1$ can be represented by basis coefficients given by property illustrated in De Boor (2001) and Mingotti, Lillo and Romo (2013). Mathematically, let $\Delta_n = t_{jr} - t_{j(r-1)}$, by Riemann Integral and the B-spline property, we have

$$||\beta_j||_1 = \int_{\mathcal{T}} |\beta_j(t)|\, dt \approx \Delta_n \sum_{r=1}^{M_n} |\beta_j(t_{jr})| \approx \Delta_n \sum_{k=1}^{M_n+d} |b_{jk}| = \Delta_n ||\boldsymbol{b}_j||_1.$$

Hence, functional ℓ_1 penalty term can be replaced by $\Delta_n \sum_{j=1}^{J} ||\boldsymbol{b}_j||_1$. Additionally, the functional $\ell_{1,2}$ penalty term can be rewritten as $\sum_{j=1}^{J}(||\beta_j||_2^2 + \varphi||\mathcal{D}^2\beta_j||_2^2)^{1/2} = \sum_{j=1}^{J}\{\boldsymbol{b}_j^T(\boldsymbol{\Phi}_j + \varphi\boldsymbol{\Omega}_j)\boldsymbol{b}_j\}^{1/2}$, where $\boldsymbol{\Phi}_j$ is a $(M_n+d)\times(M_n+d)$ matrix with the (p,q) element equal to $(\boldsymbol{\Phi}_j)_{pq} = \int B_{jp}(t)B_{jq}(t)dt$, and $\boldsymbol{\Omega}_j$ is also a $(M_n+d)\times(M_n+d)$ matrix with the (p,q) element equal to $(\boldsymbol{\Omega}_j)_{pq} = \int B_{jp}^{(2)}(t)B_{jq}^{(2)}(t)dt$. In practice, with dense grids observed over the domain $\mathcal{T}$, we compute $(\boldsymbol{\Phi}_j)_{pq} \approx \Delta_n \sum_r B_{jp}(t_{jr})B_{jq}(t_{jr})$ and $(\boldsymbol{\Omega}_j)_{pq} \approx \Delta_n \sum_r B_{jp}^{(2)}(t_{jr})B_{jq}^{(2)}(t_{jr})$. Let $\boldsymbol{K}_{\varphi,j,\Delta_n} = (\boldsymbol{\Phi}_j + \varphi\boldsymbol{\Omega}_j)/\Delta_n^2$, and we can use Cholesky decomposition to have $\boldsymbol{K}_{\varphi,j,\Delta_n} = \boldsymbol{L}_j\boldsymbol{L}_j^T$, where $\boldsymbol{L}_j$ is a $(M_n+d)\times(M_n+d)$ non-singular lower triangular matrix. By defining a new vector of coefficients $\tilde{\boldsymbol{b}}_j = \Delta_n\boldsymbol{L}_j^T\boldsymbol{b}_j$ and $\tilde{\boldsymbol{b}} = (\tilde{\boldsymbol{b}}_1^T, \ldots, \tilde{\boldsymbol{b}}_J^T)^T$, we obtain the equivalent matrix form of the objective function in Equation (2.4),

$$L_n(\tilde{\boldsymbol{b}}, \boldsymbol{\mu}) = \frac{1}{2}\left\|\boldsymbol{Y} - \boldsymbol{\mu} - \sum_{j=1}^{J}\widetilde{\boldsymbol{U}}_j\tilde{\boldsymbol{b}}_j\right\|_2^2 + \lambda_1\sum_{j=1}^{J} w_j^{(1)}||(\boldsymbol{L}_j^T)^{-1}\tilde{\boldsymbol{b}}_j||_1 + \lambda_2\sum_{j=1}^{J} w_j^{(2)}||\tilde{\boldsymbol{b}}_j||_2, \tag{2.5}$$

where $\boldsymbol{Y} = (Y_1, \ldots, Y_n)^T$, $\boldsymbol{\mu} = \mu\mathbf{1}_n$, $\mathbf{1}_n = (1, \ldots, 1)^T$ denotes a n-vector of ones. $\widetilde{\boldsymbol{U}}_j = \boldsymbol{U}_j(\boldsymbol{L}_j^T)^{-1}/\Delta_n$ where $\boldsymbol{U}_j$ is an $n\times(M_n+d)$ matrix with entries $(\boldsymbol{U}_j)_{ip} = \int_{\mathcal{T}} X_{ij}(t)B_{jp}(t)dt$ that can be approximated by $\Delta_n\sum_r X_{ij}(t_{jr})B_{jp}(t_{jr})$. The solution to the FadDoS estimator of $\hat{\beta}_j(t)$ is thus $\hat{\beta}_j(t) = \Delta_n^{-1}\boldsymbol{B}_j^T(\boldsymbol{L}_j^T)^{-1}\tilde{\boldsymbol{b}}_j^*$ and the objective turns to estimate $\tilde{\boldsymbol{b}}_j^*$.

2.4. Algorithm

The algorithm for solving Equation (2.5) was developed based on Alternating Direction Method of Multipliers (ADMM). ADMM is effective in Lasso and group Lasso, and has been recently extended to non-separable and non-smooth convex problems such as fused Lasso, as it is powerful in leveraging the complicated structure of ℓ_1 norm (Boyd et al., 2010; Li et al., 2014; Beer et al., 2019). Here we choose ADMM because it does not require a differentiable objective function

and at the same time supports decomposition to split the objective into solvable subproblems.

We first define two common operators: for $\lambda, \kappa > 0$ and $y \in \mathbb{R}^p$, the soft thresholding operator $S_{1,\lambda/\kappa}(y) = \text{sgn}(y)(|y| - \lambda/\kappa)_+$ and $S_{2,\lambda/\kappa}(y) = y/||y||_2(||y||_2 - \lambda/\kappa)_+$. For ease of notation, we assume no intercept, i.e., $\mu = 0$ and consider $w_j^{(1)} = w_j^{(2)} = 1$ for all j to illustrate the main idea, which is equivalent to the FDoS estimator. The estimate of μ for the case $\mu \neq 0$ is shown at the end of the section. Let $\boldsymbol{D}$ be a block diagonal matrix, $\boldsymbol{D} = \text{diag}\{(\boldsymbol{L}_1^T)^{-1}, (\boldsymbol{L}_2^T)^{-1}, \ldots, (\boldsymbol{L}_J^T)^{-1}\}$, the optimization problem in Equation (2.5) can be rewritten as

$$\min_{\tilde{\boldsymbol{b}}_j \in \mathbb{R}^{M_n+d}, j=1,\ldots,J} \frac{1}{2} \left\| \boldsymbol{Y} - \sum_{j=1}^{J} \widetilde{\boldsymbol{U}}_j \tilde{\boldsymbol{b}}_j \right\|_2^2 + \lambda_1 \left\| \boldsymbol{z} \right\|_1 + \lambda_2 \sum_{j=1}^{J} \left\| \tilde{\boldsymbol{b}}_j \right\|_2 \text{ s.t. } \boldsymbol{z} = \boldsymbol{D}\tilde{\boldsymbol{b}} \quad (2.6)$$

and the corresponding augmented Lagrangian function $\mathcal{L}(\tilde{\boldsymbol{b}}, \boldsymbol{z}, \boldsymbol{u}) = 1/2||\boldsymbol{Y} - \sum_{j=1}^{J} \widetilde{\boldsymbol{U}}_j \tilde{\boldsymbol{b}}_j||_2^2 + \lambda_1||\boldsymbol{z}||_1 + \lambda_2 \sum_{j=1}^{J} ||\tilde{\boldsymbol{b}}_j||_2 + \boldsymbol{u}^T(\boldsymbol{D}\tilde{\boldsymbol{b}} - \boldsymbol{z}) + \rho/2||\boldsymbol{D}\tilde{\boldsymbol{b}} - \boldsymbol{z}||_2^2$, where $\rho > 0$ is the prespecified augmented Lagrangian parameter, namely the step size to update the dual variable. The iteration scheme of ADMM consists of three sub-problems shown below.

First, we deal with the $\tilde{\boldsymbol{b}}$-update at the kth iteration. Since the penalty terms are separable, we can update $\tilde{\boldsymbol{b}}_j^{k+1}$ parallelly at the kth iteration. Assuming other coefficients fixed, we define $\boldsymbol{r}_{(-l)} = \boldsymbol{Y} - \sum_{j \neq l} \widetilde{\boldsymbol{U}}_j \tilde{\boldsymbol{b}}_j^k$, $\boldsymbol{z}^k = \{(\boldsymbol{z}_1^k)^T, \ldots, (\boldsymbol{z}_J^k)^T\}^T$ and $\boldsymbol{u}^k = \{(\boldsymbol{u}_1^k)^T, \ldots, (\boldsymbol{u}_J^k)^T\}^T$, then

$$\begin{aligned} \tilde{\boldsymbol{b}}_l^{k+1} = \underset{\tilde{\boldsymbol{b}}_l \in \mathbb{R}^{M_n+d}}{\arg\min} & \frac{1}{2} \left\| \boldsymbol{r}_{(-l)} - \widetilde{\boldsymbol{U}}_l \tilde{\boldsymbol{b}}_l \right\|_2^2 + \lambda_2 ||\tilde{\boldsymbol{b}}_l||_2 + (\boldsymbol{u}_l^k)^T \{(\boldsymbol{L}_l^T)^{-1} \tilde{\boldsymbol{b}}_l - \boldsymbol{z}_l^k\} \\ & + \frac{\rho}{2} \left\| (\boldsymbol{L}_l^T)^{-1} \tilde{\boldsymbol{b}}_l - \boldsymbol{z}_l^k \right\|_2^2. \end{aligned} \quad (2.7)$$

To further solve the optimization problem, we employ a linearization technique to approximate the quadratic term efficiently (Wang and Yuan, 2012), and details of derivation are provided in the online Supplementary Material. We let $\widehat{\boldsymbol{U}}_l = (\widetilde{\boldsymbol{U}}_l^T, \rho^{1/2}\boldsymbol{L}_l^{-1})^T$ and $\hat{\boldsymbol{r}}_{(-l)}^k = \{\boldsymbol{r}_{(-l)}^T, \rho^{1/2}(\boldsymbol{z}_l^k + \boldsymbol{u}_l^k/\rho)^T\}^T$, and thus the $\tilde{\boldsymbol{b}}$-update solution is computed by the second operator such that $\tilde{\boldsymbol{b}}_l^{k+1} = S_{2,\lambda_2/\nu_l}\{\tilde{\boldsymbol{b}}_l^k - \widehat{\boldsymbol{U}}_l^T(\widehat{\boldsymbol{U}}_l \tilde{\boldsymbol{b}}_l^k - \hat{\boldsymbol{r}}_{(-l)}^k)/\nu_l\}$, where $\nu_l > 0$ is the proximal parameter that controls the proximity to $\tilde{\boldsymbol{b}}_l^k$. Note that ν_l is required to be larger than the spectral radius of $\widetilde{\boldsymbol{U}}_l^T \widetilde{\boldsymbol{U}}_l + \rho \mathbf{I}$ for convergence (Wang and Yuan, 2012).

Second, the $\boldsymbol{z}$-update at the kth iteration is

$$\boldsymbol{z}^{k+1} = \underset{\boldsymbol{z} \in \mathbb{R}^{J \times (M_n+d)}}{\arg\min} \lambda_1 ||\boldsymbol{z}||_1 + (\boldsymbol{u}^k)^T(\boldsymbol{D}\tilde{\boldsymbol{b}}^{k+1} - \boldsymbol{z}) + \frac{\rho}{2} \left\| \boldsymbol{D}\tilde{\boldsymbol{b}}^{k+1} - \boldsymbol{z} \right\|_2^2. \quad (2.8)$$

The $\boldsymbol{z}$-update solution is computed by the first operator such that $\boldsymbol{z}^{k+1} =$

$S_{1,\lambda_1/\rho}(\boldsymbol{D}\tilde{\boldsymbol{b}}^{k+1} + \boldsymbol{u}^k/\rho)$. Details of derivation are given in the Supplementary Material.

Third, the dual variable update at the kth iteration is $\boldsymbol{u}^{k+1} = \boldsymbol{u}^k + \rho(\boldsymbol{D}\tilde{\boldsymbol{b}}^{k+1} - \boldsymbol{z}^{k+1})$.

The full algorithm is described in Algorithm 1. When $\mu \neq 0$, we can estimate $\hat{\mu}^{k+1}$ after $\boldsymbol{z}$-update at the kth iteration by computing $\hat{\mu}^{k+1} = n^{-1}\mathbf{1}_n^T(\boldsymbol{Y} - \widetilde{\boldsymbol{U}}\boldsymbol{D}^{-1}\boldsymbol{z}^{k+1})$ until the stopping criteria is reached. In addition, the set of three tuning parameters is determined by cross-validation, assuming ρ prespecified, and M_n and d fixed.

Algorithm 1 ADMM FDoS.

Hyperparameters Choose tuning parameters λ_1, λ_2, $\varphi > 0$; step size $\rho > 0$; B-splines parameters M_n and d;
Initialize $\tilde{\boldsymbol{b}}^1$, $\boldsymbol{z}^1$, $\boldsymbol{u}^1$
  for $k = 1, 2, \ldots,$ **do**
    for $l = 1, 2, \ldots, J$ **do**
      Let ν_l be the spectral radius of $\widetilde{\boldsymbol{U}}_l^T\widetilde{\boldsymbol{U}}_l + \rho\mathbf{I}$ multiplied by 5.
      Compute $\tilde{\boldsymbol{b}}_l^{k+1} = S_{2,\lambda_2/\nu_l}\{\tilde{\boldsymbol{b}}_l^k - \widehat{\boldsymbol{U}}_l^T(\widehat{\boldsymbol{U}}_l\tilde{\boldsymbol{b}}_l^k - \hat{\boldsymbol{r}}_{(-l)}^k)/\nu_l\}$.
    end for
    Compute $\boldsymbol{z}^{k+1} = S_{1,\lambda_1/\rho}(\boldsymbol{D}\tilde{\boldsymbol{b}}^{k+1} + \boldsymbol{u}^k/\rho)$.
    Compute $\boldsymbol{u}^{k+1} = \boldsymbol{u}^k + \rho(\boldsymbol{D}\tilde{\boldsymbol{b}}^{k+1} - \boldsymbol{z}^{k+1})$.
  end for until stopping criteria with a pre-specified tolerance ϵ_{tol} such that

$$\frac{||\tilde{\boldsymbol{b}}^{k+1} - \tilde{\boldsymbol{b}}^k||_2}{\max\{||\tilde{\boldsymbol{b}}^{k+1}||_2, 1\}} \leq \epsilon_{\text{tol}}$$

  is reached.
  Output $\boldsymbol{z}^*$ and $\tilde{\boldsymbol{b}}^* = \boldsymbol{D}^{-1}\boldsymbol{z}^*$.

3. Asymptotic Properties

We study the asymptotic properties of our proposed estimators in the fixed dimensional models, where the number of functional covariates does not depend on the sample size. Under some regularity conditions, we show that the proposed estimator converges to the true coefficient function in a bounded rate. As established in prior work (Zou, 2006; Nardi and Rinaldo, 2008; Poignard, 2018), Lasso-type estimators including group Lasso and sparse group Lasso cannot achieve the oracle property without appropriate adaptive penalization. Our theoretical analysis confirms that the FDoS estimator follows the same results and shows that only the FadDoS estimator fulfills the oracle property under mild conditions requiring an initial estimator. To demonstrate the asymptotic behaviors, the following assumptions are required:

(A.1) $||X||_2 \leq c_1 < \infty$ almost surely for some constant c_1.

(A.2) $\beta(t)$ satisfies Hölder condition: for some integer ξ and $\zeta \in [0,1]$, $|\beta^{\xi}(t_1) - \beta^{\xi}(t_2)| \leq c_2|t_1 - t_2|^{\zeta}$ for some constant c_2, and $t_1, t_2 \in \mathcal{T}$. We denote $\delta \stackrel{\text{def}}{=} \xi + \zeta$ and assume that $5/2 < \delta \leq d$, where d is the degree of B-splines.

Assumptions (A.1) and (A.2) are identical with (H1) and (H3) of Cardot, Ferraty and Sarda (2003). (A.2) requires the coefficient function $\beta(t)$ to be sufficiently smooth and have δ derivatives.

Recall that λ_1, λ_2, and φ are tuning parameters for the proposed estimators. For simplicity of the conditions, we choose $\varphi = \lambda_2^2$ to reduce the number of tuning parameters. In addition, we assume the following condition of choosing values of M_n.

(A.3) $M_n \asymp n^{1/(2\delta-1)}$, which is defined such that both $M_n/n^{1/(2\delta-1)}$ and $n^{1/(2\delta-1)}/M_n$ are bounded.

We also provide some notations for illustrating the following properties. Coefficient function $\beta_l(t)$ is divided into two regions: $I_1(\beta_l) = \{I \subset \mathcal{T} : \sup|\beta_l(t)| > D_l M_n^{-\delta}, \text{for all } t \in I\}$ for some positive constant D_l, and the complementary part is $I_0(\beta_l) = \{I \subset \mathcal{T} : 0 \leq \sup|\beta_l(t)| \leq D_l M_n^{-\delta}, \text{for all } t \in I\}$. It is clear that the zero subregion $N_0(\beta_l)$ is the subset of $I_0(\beta_l)$ for sufficient large D_l. We further define $\mathcal{B}$ as the set consisting of k basis coefficients whose corresponding basis functions $B_k(t)$ have support inside $I_1(\beta_l)$, i.e., $\{k \in \{1, \ldots, M_n + d\} : b_k \neq 0\}$. Let $\mathcal{A} = \{l = \{1, \ldots, J\} : \sup|\beta_l(t)| > D_l M_n^{-\delta}, \text{for all } t \in \mathcal{T}\}$ be the group of nonzero coefficient functions.

The first theoretical result demonstrates that our proposed estimator converges at an optimal rate in approximating the true function. It is proved through the estimation error between the proposed estimator and B-spline approximant $\beta_l^{\alpha}(t)$ from the family of $\mathcal{S}_l$ with a corresponding vector of basis coefficient $\boldsymbol{\alpha}_l$, while $\beta_l^{\alpha}(t)$ converges to the true function as stated in the following lemma, which is the same as De Boor (2001, Thm. XII(6)).

Lemma 1. *There exists $\beta^{\alpha}(t)$ such that $||\beta - \beta^{\alpha}||_{\infty} \leq C_1 M_n^{-\delta}$ for some constant $C_1 \geq 0$, where $\beta^{\alpha}(t) \stackrel{def}{=} \boldsymbol{B}^T(t)\boldsymbol{\alpha}$.*

It is further noted that the B-spline approximant can preserve the sparsity of the true function because the basis coefficients can be equal to zero if the support of the corresponding basis function falls in the zero subregion $N_0(\beta_l)$. Thus, the asymptotic behavior of the proposed estimator is stated in the following theorem.

Theorem 1 (Convergence). *Under* (A.1)–(A.3)*, the penalized estimator satisfies*

$$||\hat{\beta}_l - \beta_l||_{\infty} = O_p(M_n n^{-1/2}), \quad l = 1, \ldots, J,$$

if $\lambda_1 = O(M_n^{1/2} n^{-1/2})$ and $\lambda_2^2 = O(n^{-1/2})$ for FDoS*; or if $\lambda_1 \phi_1 = O_p(M_n^{1/2} n^{-1/2})$ and $\lambda_2^2 \phi_2 = O_p(n^{-1/2})$ for* FadDoS*, where $\phi_1 = \sup_{l \in \mathcal{A}} ||\check{\beta}_l||_1^{-a}$ and $\phi_2 = \sup_{l \in \mathcal{A}} ||\check{\beta}_l||_2^{-a}$ are stochastic quantities.*

The above theorem highlights that there always exists a local solution $\hat{\beta}_l(t)$ for the lth functional covariate with probability approaching one, while the convergence rate is bounded by $O_p(M_n n^{-1/2})$.

The oracle property of the functional linear model with multiple functional covariates can be stated under the double-sparsity structure. For global sparsity, the estimator can identify all true nonzero coefficient functions with probability approaching one. For local sparsity, the estimator further recovers zero subregions with probability approaching one and is asymptotically pointwise normally distributed.

The following theorem is developed for the asymptotic distribution of the FDoS estimator. The proof is through finding the limiting distribution of the vector of B-spline basis coefficient as n goes to infinity, which is also the minimizer of a convex function $V_1^{(l)}$. Additionally, we assume a positive-definite matrix, $\boldsymbol{C}_l = \lim_{n\to\infty}(M_n/n)\boldsymbol{U}_l^T\boldsymbol{U}_l$, whose minimum and maximum eigenvalues are bounded away from zero and infinity with probability approaching one; see Lemma 2 in the Supplementary Material for more details.

Theorem 2. *Under* (A.1)–(A.3), *if* $\lambda_1(n/M_n)^{1/2} \to \gamma_1 \geq 0$ *and* $\lambda_2^2 n^{1/2} \to \gamma_2 \geq 0$,

$$\left(\frac{n}{M_n}\right)^{1/2}\{\hat{\beta}_l(t) - \beta_l(t)\} \xrightarrow{d} \boldsymbol{B}_l^T(t)\boldsymbol{v}_l^*, \quad l = 1, \ldots, J,$$

where $\boldsymbol{v}_l^* = \arg\min_{\boldsymbol{v}\in\mathbb{R}^{M_n+d}} V_1^{(l)}(\boldsymbol{v})$ *such that*

$$V_1^{(l)}(\boldsymbol{v}) = \boldsymbol{v}^T\boldsymbol{C}_l\boldsymbol{v} - 2\boldsymbol{W}_l^T\boldsymbol{v} + \gamma_1\Gamma_1^{(l)}(\boldsymbol{v}) + \gamma_2\Gamma_2^{(l)}(\boldsymbol{v}),$$

and some random vector $\boldsymbol{W}_l \sim N(\boldsymbol{0}, \sigma^2\boldsymbol{C}_l)$, $\Gamma_1^{(l)}(\boldsymbol{v}) = \sum_{k=1}^{M_n+d}\big\{|v_k|\mathbb{I}(\alpha_{lk} = 0) + v_k sgn\big(\alpha_{lk}\big)\mathbb{I}(\alpha_{lk} \neq 0)\big\}$ *and* $\Gamma_2^{(l)}(\boldsymbol{v}) = ||\boldsymbol{v}||_2\mathbb{I}(\boldsymbol{\alpha}_l = \boldsymbol{0}) + (\boldsymbol{v}^T\boldsymbol{\alpha}_l/||\boldsymbol{\alpha}_l||_2)\mathbb{I}(\boldsymbol{\alpha}_l \neq \boldsymbol{0})$.

In Thoerem 2, $\Gamma_1^{(l)}$ and $\Gamma_2^{(l)}$ are brought by estimation error with respect to functional ℓ_1 and $\ell_{1,2}$ respectively. The above theorems imply that when $\lambda_1 = O(M_n^{1/2}n^{-1/2})$ and $\lambda_2 = O(n^{-1/4})$, the FDoS estimate is $\sqrt{n}/M_n$-consistent, but neither global nor local sparsity structures can be recovered with high probability, as stated in the following Proposition 1 under the assumption that $\gamma_1 < \infty$ and $\gamma_2 < \infty$.

Proposition 1. *Under* (A.1)–(A.3), *if* $\lambda_1(n/M_n)^{1/2} \to \gamma_1 \geq 0$ *and* $\lambda_2^2 n^{1/2} \to \gamma_2 \geq 0$, *the* FDoS *estimator satisfies the following:*

1. $\lim_{n\to\infty} P(\hat{\mathcal{A}}_n = \mathcal{A}) < 1$;
2. *For* $t \in I_0(\beta_l)$ *such that* $l \in \mathcal{A}$, $P(\hat{\beta}_l(t) = 0) < 1$.

Proposition 1 shows that the FDoS estimator fails to enjoy the oracle property. An appropriate adaptive penalization discussed in Section 2.3 and termed as FadDoS is a promising way to address this issue. The below theorem

shows that the FadDoS estimator enjoys the oracle property if we can ensure that adaptive weights satisfy mild conditions. For example, the initial estimator is chosen to be the smoothing spline estimator proposed by Cardot, Ferraty and Sarda (2003), which is a $\sqrt{n}/M_n$-consistent estimator for $\beta_l(t)$. The following results show that the FadDoS estimator achieves the oracle property with proper convergence rates of λ_1 and λ_2.

Theorem 3 (Oracle Property of FadDoS). *Under* (A.1)–(A.3), *if* $\lambda_1(n/M_n)^{1/2} \to 0$ *and* $\lambda_2^2 n^{1/2} \to 0$; $\lambda_1 n^{(a+1)/2}/M_n^{a+1/2} \to \infty$ *and* $\lambda_2^2 n^{(a+1)/2}/M_n^a \to \infty$, *the* FadDoS *estimator satisfies the following:*

1. (Consistency of global variable selection) $\lim_{n\to\infty} P(\hat{\mathcal{A}}_n = \mathcal{A}) = 1$.
   (Consistency of local zero-subregion identification) *For* $t \in I_0(\beta_l)$ *such that* $l \in \mathcal{A}$, $\lim_{n\to\infty} P(\hat{\beta}_l(t) = 0) = 1$.

2. *For* $t \in I_1(\beta_l)$ *such that* $l \in \mathcal{A}$, *we have* $(n/M_n)^{1/2}\{\hat{\beta}_l(t) - \beta_l(t)\} \xrightarrow{d} N\{0, \sigma_l^2(t)\}$, *where* $\sigma_l^2(t) = \lim_{n\to\infty} \sigma^2 (\boldsymbol{B}_l)_{\mathcal{B}}^T(t)(\boldsymbol{C}_l)_{\mathcal{B}\mathcal{B}}^{-1}(\boldsymbol{B}_l)_{\mathcal{B}}(t)$. $(\boldsymbol{C}_l)_{\mathcal{B}\mathcal{B}}$ *and* $(\boldsymbol{B}_l)_{\mathcal{B}}(t)$ *are sub-matrix and the corresponding sub-vector only with column indices in* $\mathcal{B}$, *respectively.*

4. Simulation

We conduct extensive simulation studies to illustrate the global sparsity and local sparsity of the proposed method and compare it with existing methods to evaluate the model performance.

The true functional linear model is $Y_i = \sum_{j=1}^{10} \int_0^1 X_{ij}(t)\beta_j(t)dt + \epsilon_i, \quad i = 1, \ldots, n$, where $\epsilon_i \sim N(0, \sigma_\epsilon^2)$. The measurement error σ_ϵ is chosen so that signal-to-noise ratio equals to 4. We generate functional covariates $X_{ij}(t)$ by letting $X_{ij}(t) = \sum_k a_{ijk}B_{jk}(t), t \in [0, 1]$, where $a_{ijk} \sim N(0, 1)$ and $B_{jk}(t)$ is a B-spline basis function defined by order 4 and 50 equally spaced knots. Three different types of coefficient functions are considered, each representing a unique condition:

(i) $\beta_1(t)$ has a zero subregion $\beta_1(t) = 2\sin(3\pi t)$ for $0 \le t \le 1/3$; $\beta_1(t) = 0$ for $1/3 < t < 2/3$; $\beta_1(t) = -2\sin(3\pi t)$, for $2/3 \le t \le 1$;

(ii) $\beta_2(t)$ has no zero subregion but two crossings at zero, such that $\beta_2(t) = 1.5t^2 + 2\sin(3\pi t)$;

(iii) $\beta_j(t) = 0$ for $j = 3, \ldots, 10$, indicating that the functional covariate has no contribution to the response throughout the entire time domain. We independently simulate 100 replicates. Some examples of generated data are provided in the Supplementary Material.

The performance of the estimator is evaluated from two aspects reflecting the realization of global sparsity and local sparsity respectively. The global sparsity, which is equivalent to functional variable selection, can be assessed by the true

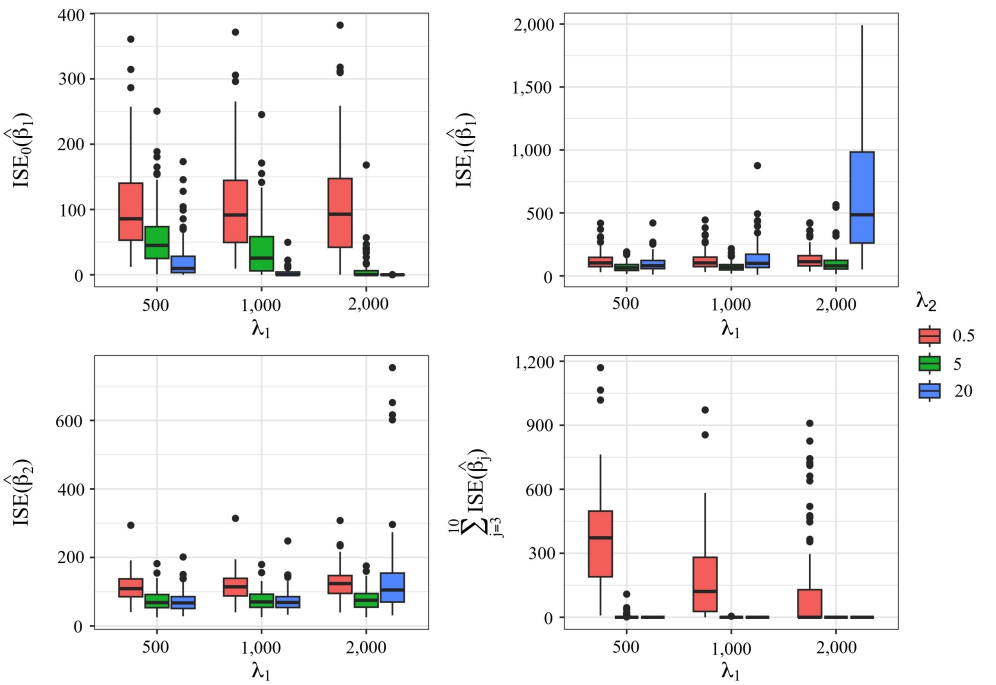


Figure 2. ISE performance of the FadDoS estimator with varying λ_1 and λ_2, φ = 1e-5 over 100 simulation replicates. The training sample size is 200.

positive rate (TPR) and the true negative rate (TNR) respectively presenting the proportion of significant or insignificant functional covariates accurately retained or excluded from the model. For local sparsity, we use the integrated squared error (ISE) to measure the difference between the estimate and the underlying truth, defined as $\text{ISE}(\hat{\beta}_j) = (1/|\mathcal{T}|) \int_{\mathcal{T}} \{\hat{\beta}_j(t) - \beta_j(t)\}^2 dt$. It can be further decomposed into $\text{ISE}_0(\hat{\beta}_j)$ and $\text{ISE}_1(\hat{\beta}_j)$ for local zero subregion $N_0(\beta_j)$ and nonzero subregion $N_1(\beta_j)$. The local sparsity can be evaluated via ISE_0. Additionally, the predictive performance measured by the prediction mean squared error (PMSE) on the testing data set is also critical to examine the effectiveness of the estimator.

4.1. Tuning parameter effects

Figure 2 shows the effect of tuning parameters λ_1 and λ_2 on the behavior of the FadDoS estimator by controlling the strength of penalization. Increasing λ_1 and λ_2 shrink $\hat{\beta}_1(t)$ in zero subregion and all $\hat{\beta}_j(t)$, $j = 3, \ldots, 10$ towards zero, while the procedure delivers exactly the opposite results regarding nonzero subregions of $\hat{\beta}_1(t)$and $\hat{\beta}_2(t)$. Thus, they provide an elastic-net-like estimation balancing zero and nonzero regions of coefficient functions. Table 1 further provides quantitative evaluation of tuning parameter effects. The minimization of PMSE is achieved under the trade-off between local and global sparsity parameters. Moreover, the global sparsity parameter λ_2 primarily administers

Table 1. PMSE ($\times 10^{-2}$), average TPR (avgTNR), and average TNR (avgTNR) performance of the FadDoS estimator with varying λ_1 and λ_2, φ =1e-5. Each entry is the average of 1,000 test examples throughout 100 simulation replicates. The entry in the parenthesis corresponds to the standard deviation. The training sample size is 200 and the test sample size is 1,000.

	λ_1	λ_2		
		0.5	5	20
PMSE($\times 10^{-2}$)	500	3.29 (0.50)	2.51 (0.18)	2.52 (0.18)
	1,000	2.95 (0.42)	2.49 (0.18)	2.56 (0.22)
	2,000	2.85 (0.42)	2.51 (0.21)	3.41 (0.87)
avgTPR	500	1.00 (0.00)	1.00 (0.00)	1.00 (0.00)
	1,000	1.00 (0.00)	1.00 (0.00)	1.00 (0.00)
	2,000	1.00 (0.00)	1.00 (0.00)	0.96 (0.14)
avgTNR	500	0.33 (0.21)	0.98 (0.05)	1.00 (0.00)
	1,000	0.67 (0.27)	1.00 (0.01)	1.00 (0.00)
	2,000	0.84 (0.24)	1.00 (0.00)	1.00 (0.00)

the overall sparseness of the functional estimates, thereby achieving functional variable selection by successfully identifying all zero coefficient functions. Further discussion of the effect and choice of φ can be found in the Supplementary Material.

4.2. Comparison with existing methods

We compare our proposed estimators with the aforementioned estimation methods fGL (Gertheiss, Maity and Staicu, 2013), fL (Mingotti, Lillo and Romo, 2013), and SLoS (Lin et al., 2017). Tuning parameters for all estimators are selected by five-fold cross-validation to ensure optimal performance of each estimator. Although the software of fL can not be identified, we can still implement it by our algorithm. It is a special case of the FDoS estimator when $\lambda_2 = 0$.

Figure 3 shows the ISE performance of all estimators for different types of coefficient functions, while Table 2 summarizes PMSE and variable selection accuracy. The proposed FDoS and FadDoS demonstrate comparable results to SLoS, and outperform fGL and fL overall. This highlights the value of modeling both global and local sparsity simultaneously, which fGL and fL fail to achieve, limiting their ability to handle such complex coefficient functions. In particular, utilizing only functional $\ell_{1,2}$ (fGL) or functional ℓ_1 (fL) penalties results in either under- or over-shrinkage.

More importantly, the use of proper adaptive weights in the FadDoS estimator can lead to more accurate estimation of zero and nonzero regions of $\beta_1(t)$ and $\beta_2(t)$ simultaneously, and at the same time reduce the number of false negatives, namely avoiding nonzero estimates of true zero functions, as suggested by the

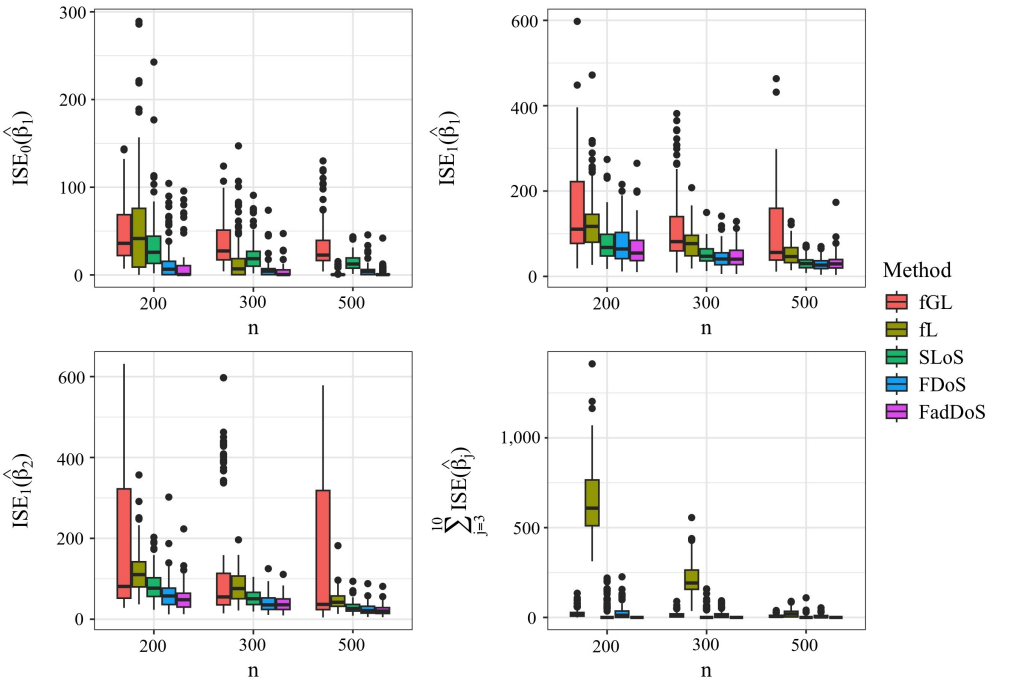


Figure 3. Comparison of ISE performance for fGL (Gertheiss, Maity and Staicu, 2013), fL(Mingotti, Lillo and Romo, 2013), SLoS (Lin et al., 2017), FDoS, and FadDoS.

higher average TNR for the eight true zero coefficients. It is noteworthy that FadDoS slightly outperforms SLoS. One explanation is that the penalization terms in FadDoS are weighted by the initial coefficient function estimates to control both global and local sparsity, while the sparsity parameters for every functional covariate designed in SLoS are identical. Therefore, the SLoS estimator might not regularize both global and local shrinkage simultaneously. Additionally, SLoS demonstrates worse estimation at the last coefficient function. While tail fluctuations when using polynomial splines are expected, FadDoS alleviates this wiggliness issue by embedding the roughness penalty within the functional $\ell_{1,2}$ penalty term rather than placing it independently as SLoS.

5. Application

In this section, the proposed FadDoS is applied to the motivating Kinect study to examine the association between elderly mobility assessments and their multi-joint movements. A total of 50 subjects aged between 68 and 89 years from four Neighbourhood Elderly Centres (NECs) in Hong Kong participated in the study. The participants performed a functional balance test called Timed Up and Go (TUG) test (Podsiadlo and Richardson, 1991), during which Kinect sensor monitored the three-dimensional displacements of 25 skeletal joints. The whole

Table 2. PMSE($\times 10^{-2}$), average TPR (avgTPR), and average TNR (avgTNR) of the proposed estimators FDoS and FadDoS as well as fGL (Gertheiss, Maity and Staicu, 2013), fL(Mingotti, Lillo and Romo, 2013), and SLoS (Lin et al., 2017). The test sample size is 1,000. The entry in the parenthesis corresponds to the standard deviation among 100 simulation replicates.

	Methods	PMSE($\times 10^{-2}$)	avgTPR	avgTNR
n=200	fGL	2.57(0.19)	1.00(0.00)	0.45(0.04)
	fL	3.76(0.51)	1.00(0.00)	0.01(0.01)
	SLoS	2.49(0.21)	1.00(0.00)	0.96(0.02)
	FDoS	2.52(0.91)	1.00(0.00)	0.78(0.04)
	FadDoS	2.44(0.18)	1.00(0.00)	1.00(0.00)
n=300	fGL	2.45(0.17)	1.00(0.00)	0.49(0.04)
	fL	2.81(0.25)	1.00(0.00)	0.17(0.04)
	SLoS	2.40(0.17)	1.00(0.00)	0.97(0.02)
	FDoS	2.41(0.17)	1.00(0.00)	0.81(0.04)
	FadDoS	2.38(0.16)	1.00(0.00)	1.00(0.00)
n=500	fGL	2.36(0.15)	1.00(0.00)	0.50(0.06)
	fL	2.38(0.16)	1.00(0.00)	0.77(0.03)
	SLoS	2.31(0.15)	1.00(0.00)	0.99(0.01)
	FDoS	2.32(0.15)	1.00(0.00)	0.86(0.04)
	FadDoS	2.31(0.15)	1.00(0.00)	1.00(0.00)

process of TUG test includes standing up from the chair, walking forward, turning around, walking back to the chair, and sitting down, as illustrated in Figure S3. We use their Balance Evaluation Systems Test (BESTest) scores as the measure of general functional balance and walking capability (Horak, Wrisley and Frank, 2009). In the pre-processing step, since the shorter completion time itself reflects better mobility (Nielsen et al., 2016), we padded the shorter signals till all observations have an equal time length to preserve a series of sequential raw actions, instead of warping time to align all movement signals. Additionally, all displacement signals are interpolated to ensure the same sampling rate. We calculate velocity by taking the first derivative of displacement signals. It is prioritized over displacement because they are less likely to be affected by different positioning of the Kinect sensor. In addition, previous research has shown interesting frequency patterns during walking (Kestenbaum et al., 2015; Morgan and Noehren, 2018), and therefore we include frequency domain knowledge to provide a more insightful understanding of joint movement patterns for the entire activity. The frequency domain information is obtained via Fourier transform. In scalar-on-function regression model, we regress the average BESTtest score on velocity and frequency of three dimensions respectively in order to identify associated joints and investigate in detail how joint movements are associated with mobility assessments over the time and frequency domains.

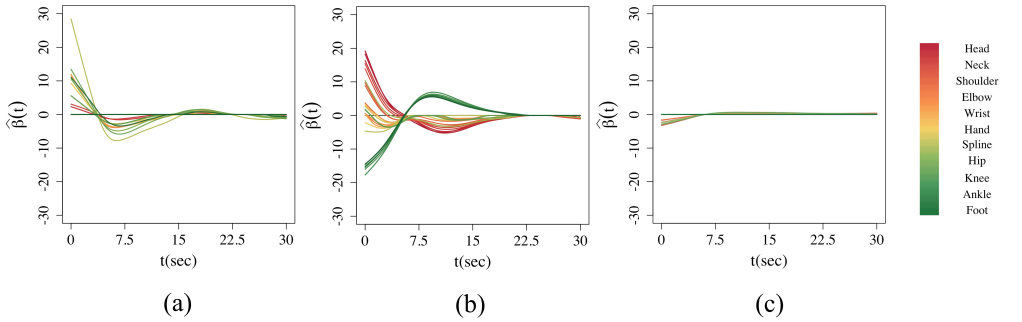
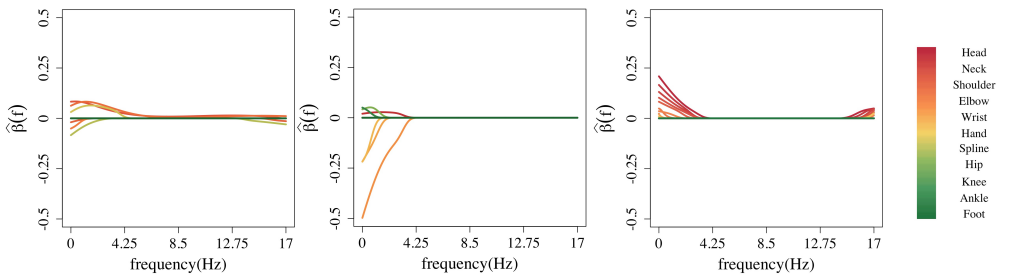


Figure 4. The estimated coefficient functions of (a) horizontal velocity, (b) vertical velocity, and (c) walking velocity. Each curve represents the estimated coefficient function of a joint and is color coded as Figure 1(b).

Figure 4(a) shows that the horizontal velocity of all joints shares an almost identical pace and therefore yields similar estimates. The estimates are positive till around 3 seconds when the healthy elderly stands up, which implies that faster ascending comes with good mobility. All estimates are close to zero around 18 seconds when the majority of the subjects completed the test. Figure 4(b) exhibits three clusters of behavior patterns related to joints in head, trunk and upper limbs, and lower limbs. The positive coefficients for the upward velocity from head to shoulder in the beginning also reveal that the swifter standing up the better mobility. Due to the setup of the sensor shown in Figure 1(a), the trunk and upper limbs are closer to the position of the Kinect camera, namely the origin of the three-dimensional coordinates, leading to relatively weaker effects. For lower limbs, high velocity during ascending and walking forward implies good physical capability. Moreover, the direction of vertical velocity is reversed when walking backward. It is reflected in the change of the signs in the estimates, all of which cross zero around 7 seconds as the healthy elderly turn. It is observed that the amplitude of the estimated coefficient is maximized at 12 seconds for head or 10 seconds for lower limbs when returning to the chair. The lag is due to the downward motion of head during sitting down. The nonzero estimates of most joints depicted in Figure 4(c) display consistent patterns. Since the stride velocity during walking forward is negative due to the Kinect camera setup, negative estimates before 7 seconds also indicate the faster pace the better mobility assessment.

Figure 5 shows the estimated coefficient functions in the frequency spectrum from 0 Hz to 17 Hz along three directions. Figure 5(a) demonstrates that only joints of shoulder, elbows, and hands have substantial nonzero effects in part of the horizontal frequency domain, where the low-frequency oscillation reflects the arm's horizontal swinging motion during walking, while the high-frequency oscillation represents the horizontally involuntary and rhythmic shaking probably



(a) (b) (c)

Figure 5. The estimated coefficient functions of (a) horizontal frequency, (b) vertical frequency, and (c) walking frequency. Each curve represents the estimated coefficient function of a joint and is color coded as Figure 1(b).

caused by some disorders and thus associated with lower mobility (Hess and Pullman, 2012). The estimates of vertical oscillations of the shoulder, spine, and ankles are positive only at low frequencies as shown in Figure 5(b), representing the natural up-and-down oscillation in the normal stride. However, the negative estimates imply that the vertical swinging movement of hands is generally seen among the low-mobility elderly who spend excessive energy vertically to assist with walking. Additionally, Figure 5(c) shows that only joints between the head and hands in the walking direction are associated with mobility. The nonzero estimates at the low-frequency domain confirm that the average stride frequency is 2 Hz (Henriksen et al., 2004). The magnitude of the effect decreases from the top joints to the bottom is likely because intensive head motions tilting forward are compensatory for trunk pitch to stabilize the body (Hirasaki et al., 1993, 1999). Moreover, the positive estimates of high-frequency oscillation are probably due to normal action tremors present in healthy individuals for active movements (Hess and Pullman, 2012).

6. Conclusion

While scalar-on-function regression model is appealing for multi-dimensional sensor data, where a large number of functional covariates are collected and scalar responses are measured, the double-sparsity property is critical for the interpretability of the model. Thus, we propose the novel FadDoS estimator, based on which we develop a nice double-sparsity model. In particular, one can concurrently identify important functional covariates and influential nonzero subregions throughout the entire domain. Our estimator is developed based on the functional generalization of sparse group Lasso to control local and global sparsity respectively, as well as smoothness. We further enhance it with adaptation where the penalty can be weighted by preliminary estimates. The FadDoS estimator enjoys the oracle property under mild conditions of the initial

estimator and therefore it is theoretically sounder than the nonadaptive version. Simulation studies also show that our estimator outperforms existing methods with optimal performance.

Our work has provided an effective method for analyzing wearable sensor data, especially those that capture high-dimensional functional signals. In the application of Kinect sensor study, we achieved satisfactory performance and identified the associated joints and their detailed association over both time and frequency domains. Furthermore, the FadDoS estimator is generalizable and applicable to other types of sensor devices that collect multi-dimensional signals including acceleration, heart rates, skin impedance, and more, thus promoting the use of advanced sensor devices in health research and having wide applications in the health field.

Supplementary Material

The online Supplementary Material contains the proofs of the theorems.

Acknowledgments

The authors thank all participants who took part in the study, and the researchers who assisted with data collection. This study was funded by the City University of Hong Kong, Hong Kong SAR, China internal research grants (9610473 and 7005892), and the Hong Kong Innovation and Technology Commission (InnoHK Project CIMDA). The funder did not have any influence over the study design, subsequent analysis, or drafting of the manuscript.

References

Beer, C., Aizenstein, J., Anderson, J. and Krafty, T. (2019). Incorporating prior information with fused sparse group Lasso: Application to prediction of clinical measures from neuroimages. *Biometrics* **75**, 1299–1309.

Boyd, S., Parikh, N., Chu, E., Peleato, B. and Eckstein, J. (2010). Distributed optimization and statistical learning via the alternating direction method of multipliers. *Foundations and Trends® in Machine Learning* **3**, 1–122.

Cardot, H., Ferraty, F. and Sarda, P. (2003). Spline estimators for the functional linear model. *Statistica Sinica* **13**, 571–791.

Cheng, Y., Shi, J. and Eyre, J. (2020). Nonlinear mixed-effects scalar-on-function models and variable selection. *Statistics and Computing* **30**, 129–140.

Cippitelli, E., Gasparrini, S., Spinsante, S. and Gambi, E. (2015). Kinect as a tool for gait analysis: Validation of a real time joints extraction algorithm working in side view. *Sensors* **15**, 1417–1434.

Collazos, A., Dias, R. and Zambom, Z. (2016). Consistent variable selection for functional regression models. *Journal of Multivariate Analysis* **146**, 63–71.

De Boor, C. (2001). *A Practical Guide to Splines*. Springer.

Fan, J. and Li, R. (2001). Variable selection via nonconcave penalized likelihood and its oracle properties. *Journal of the American Statistical Association* **96**, 1348–1360.

Friedman, J., Hastie, T. and Tibshirani, R. (2010). A note on the group Lasso and a sparse group Lasso. *arXiv:1001.0736*.

Gasparrini, S., Cippitelli, E., Spinsante, S. and Gambi, E. (2014). A depth-based fall detection system using a Kinect sensor. *Sensors* **14**, 2756–2775.

Gertheiss, J., Maity, A. and Staicu, A. (2013). Variable selection in generalized functional linear models. *Stat* **2**, 86–101.

Henriksen, M., Lund, H., Moe-Nilssen, R., Bliddal, H. and Danneskiod-Samsøe, B. (2004). Test-retest reliability of trunk accelerometric gait analysis. *Gait & Posture* **19**, 288–297.

Hess, C. W. and Pullman, S. L. (2012). Tremor: Clinical phenomenology and assessment techniques. *Tremor and other Hyperkinetic Movements* **2**, tre–02–65–365–1. DOI: 10.7916/D8WM1C41.

Hirasaki, E., Kubo, T., Nozawa, S., Matano, S. and Matsunaga, T. (1993). Analysis of head and body movements of elderly people during locomotion. *Acta Oto-Laryngologica. Supplementum* **501**, 25–30.

Hirasaki, E., Moore, S. T., Raphan, T. and Cohen, B. (1999). Effects of walking velocity on vertical head and body movements during locomotion. *Experimental Brain Research* **127**, 117–130.

Horak, F. B., Wrisley, D. M. and Frank, J. (2009). The balance evaluation systems test (bestest) to differentiate balance deficits. *Physical Therapy* **89**, 484–498.

James, G. M., Wang, J. and Zhu, J. (2009). Functional linear regression that's interpretable. *The Annals of Statistics* **37**, 2083–2108.

Kestenbaum, M., Michalec, M., Yu, Q., Pullman, S. L. and Louis, E. D. (2015). Intention tremor of the legs in essential tremor: Prevalence and clinical correlates. *Movement Disorders Clinical Practice* **2**, 24–28.

Kohout, J., Verešpejová, L., Kříž, P., Červená, L., Štícha, K., Crha, J. et al. (2021). Advanced statistical analysis of 3D Kinect data: Mimetic muscle rehabilitation following head and neck surgeries causing facial paresis. *Sensors* **21**, 103. DOI: 10.3390/s21010103.

Li, X., Mo, L., Yuan, X. and Zhang, J. (2014). Linearized alternating direction method of multipliers for sparse group and fused Lasso models. *Computational Statistics & Data Analysis* **79**, 203–221.

Lin, Z., Cao, J., Wang, L. and Wang H. (2017). Locally sparse estimator for functional linear regression models. *Journal of Computational and Graphical Statistics* **26**, 306–318.

Matsui, H. and Konishi, S. (2011). Variable selection for functional regression models via the l_1 regularization. *Computational Statistics & Data Analysis* **55**, 3304–3310.

Meier, L., Van de Geer, S. and Bühlmann, P. (2009). High-dimensional additive modeling. *The Annals of Statistics* **37**, 3779–3821.

Mingotti, N., Lillo, R. E. and Romo, J. (2013). Lasso variable selection in functional regression. Working paper 13–14, Statistics and Econometrics Series 13. Universidad Carlos III de Madrid. Web: `https://hdl.handle.net/10016/16959`.

Morgan, K. and Noehren, B. (2018). Identification of knee gait waveform pattern alterations in individuals with patellofemoral pain using fast fourier transform. *PLoS ONE* **13**, e0209015.

Nardi, Y. and Rinaldo, A. (2008). On the asymptotic properties of the group Lasso estimator for linear models. *Electronic Journal of Statistics* **2**, 603–633.

Nielsen, L. M., Kirkegaard, H., Østergaard, L. G., Bovbjerg, K., Breinholt, K. and Maribo, T. (2016). Comparison of self-reported and performance-based measures of functional ability in elderly patients in an emergency department: Implications for selection of clinical outcome measures. *BMC Geriatrics* **16**, 199. DOI: 10.1186/s12877-016-0376-1.

Pannu, J. and Billor, N. (2017). Robust group-Lasso for functional regression model. *Communications in Statistics - Simulation and Computation* **46**, 3356–3374.

Podsiadlo, D. and Richardson, S. (1991). The timed "Up & Go": A test of basic functional mobility for frail elderly persons. *Journal of the American Geriatrics Society* **39**, 142–148.

Poignard, B. (2018). Asymptotic theory of the adaptive sparse group Lasso. *Annals of the Institute of Statistical Mathematics* **72**, 297–328.

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

Simon, N., Friedman, J., Hastie, T. and Tibshirani, R. (2013). A sparse-group Lasso. *Journal of Computational and Graphical Statistics* **22**, 231–245.

Tu, C. Y., Park, J. and Wang, H. (2020). Estimation of functional sparsity in nonparametric varying coefficient models for longitudinal data analysis. *Statistica Sinica* **30**, 439–465.

Tu, C. Y., Song, D., Breidt, F. J., Berger, T. W. and Wang, H. (2012). Functional model selection for sparse binary time series with multiple inputs. In *Economic Time Series: Modeling and Seasonality*, 477–497. Chapman and Hall/CRC.

Wang, H. and Kai, B. (2015). Functional sparsity: Global versus local. *Statistica Sinica* **25**, 1337–1354.

Wang, X. and Yuan, X. (2012). The linearized alternating direction method for dantzig selector. *SIAM Journal on Scientific Computing* **34**, A2792–A2811.

Zhou, J., Wang, N. and Wang, N. (2013). Functional linear model with zero-value coefficient function at sub-regions. *Statistica Sinica* **23**, 25–50.

Zou, H. (2006). The adaptive Lasso and its oracle properties. *Journal of the American Statistical Association* **101**, 1418–1429.

(Received March 2023; accepted February 2024)