Biometrics 000, 000–000 DOI: 000
000 0000

Variable Selection in Nonlinear Function-on-scalar Regression

Rahul Ghosal

Department of Biostatistics, Johns Hopkins Bloomberg School of Public Health.

email: rghosal@ncsu.edu

and

Arnab Maity

Department of Statistics, North Carolina State University.

Summary: We develop a new method for variable selection in a nonlinear additive function-on-scalar regression model. Existing methods for variable selection in function-on-scalar regression have focused on the linear effects of scalar predictors, which can be a restrictive assumption in the presence of multiple continuously measured covariates. We propose a computationally efficient approach for variable selection in existing linear function-on-scalar regression using functional principal component scores of the functional response and extend this framework to a nonlinear additive function-on-scalar model. The proposed method provides a unified and flexible framework for variable selection in function-on-scalar regression, allowing nonlinear effects of the covariates. Numerical analysis using simulation study illustrates the advantages of the proposed method over existing variable selection methods in function-on-scalar regression even when the underlying covariate effects are all linear. The proposed procedure is demonstrated on accelerometer data from the 2003–2004 cohorts of the National Health and Nutrition Examination Survey (NHANES) in understanding the association between diurnal patterns of physical activity and demographic, lifestyle and health characteristics of the participants.

Key words: Functional Data Analysis; Functional Principal Component Analysis; Function-on-scalar Regression; NHANES; Nonlinear Regression; Variable Selection.

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.13564

1. Introduction

Function-on-scalar regression (FOSR) is an increasingly popular area of research in functional data analysis with diverse applications in biological science such as genome-wide association studies (GWAS) (Barber et al., 2017; Fan and Reimherr, 2017), physical activity research (Goldsmith et al., 2016; Kowal and Bourgeois, 2020), study of stroke severity on motor control (Goldsmith and Kitago, 2016; Chen et al., 2016) among many others. In FOSR, a continuously varying functional response is modelled using additive effects of scalar predictors of interest, which are captured using smooth univariate coefficient functions. Several methods for estimation and forecasting exist (Ramsay and Silverman, 2005; Reiss et al., 2010) in the traditional function-on-scalar regression model. The advent of modern high-throughput technologies, sensors and wearable devices in the last decade have made functional data structures complex and high-dimensional, which have motivated several variable selection methods (Chen et al., 2016; Barber et al., 2017; Fan and Reimherr, 2017; Parodi and Reimherr, 2018; Kowal and Bourgeois, 2020) in FOSR models to simultaneously identify and estimate the effects of the influential scalar predictors.

Most of these methods rely on modeling the unknown coefficient functions in terms of known basis function expansion and subsequently using appropriate penalty on the coefficient functions to induce smoothness and sparsity. Chen et al. (2016) used a group minimax concave penalty (MCP) (Zhang, 2010) with a penalized least squares criterion to perform variable selection in FOSR, where linear effects of the scalar predictors are considered. They further suggested using a "pre-whitening" method to take into account temporal correlation within curves. Such a method can be counter-productive if the model is misspecified, mainly if there are nonlinear effects of the scalar predictors. Moreover, this method is computationally costly and cannot be applied in the high dimensional case where the number of predictors is larger than the available sample size. Barber et al. (2017) proposed a function-on-scalar

lasso (FSL), where they used a penalized least squares objective function with the LASSO (Tibshirani, 1996) penalty on the basis coefficients. Like the classical lasso this method suffers from a non-negligible asymptotic bias, and hence an adaptive version of function on scalar lasso (AFSL) was developed by Fan and Reimherr (2017). FSL and AFSL, although imposing sparsity in number of predictors, does not guarantee smoothness in the estimated coefficient functions, which is a desirable property in many functional regression applications. Parodi and Reimherr (2018) introduced the FLAME method, which overcomes this problem by simultaneously introducing sparsity and smoothness in the estimated coefficient functions, is computationally fast, and is applicable in the high dimensional setting ($p > n$).

So far, all of these methods developed for estimation and variable selection in FOSR have focused on the linear effects of the scalar predictors. In the presence of multiple continuous scalar predictors, the assumption of linearity might be restrictive in many real-world applications. In this article, we develop a new method for variable selection in additive nonlinear function-on-scalar regression, which is more flexible and capable of capturing nonlinear dynamics between scalar predictors and continuously observed functional response. As a motivating application, we consider studying diurnal patterns of physical activity of the participants in the 2003–2004 cohorts of accelerometry data from the National Health and Nutrition Examination Survey (NHANES). We have minute level accelerometry data on each participant for seven days (Monday-Sunday) along with their demographic, lifestyle and health characteristics. The objective is to capture the important drivers of physical activity while simultaneously estimating their time-varying effects. Interestingly, past studies have indicated nonlinear dependence of physical activity on scalar predictors such as age (Maas et al., 2008; Varma et al., 2017) and hence a nonlinear function on scalar regression model might be more suitable in this scenario.

Our contribution in this article is two-fold. First, we develop a new, computationally

efficient method for variable selection in existing linear function-on-scalar regression using functional principal component scores of the functional response. Second, and more importantly, we extend this framework to nonlinear additive function-on-scalar regression, which is more general and provides a unified framework for variable selection in function-on-scalar regression. We use functional principal component analysis (Yao et al., 2005; Hall et al., 2006) to reformulate the FOSR model as a linear model of the principal component scores of the functional response on the scalar covariates or their nonlinear functions. The unknown nonlinear functions are further modeled using univariate basis expansions, and subsequently, row and column group-minimax concave penalty (MCP) are employed for performing the variable selection. The estimated coefficients are projected back using the univariate basis functions and the functional response's eigenfunctions to recover the effects of the scalar predictors. Numerical analyses using simulations illustrate satisfactory and competitive performance of the proposed method compared to the existing techniques for variable selection in linear function-on-scalar regression, even in very high dimensions ($p > n$). In nonlinear function-on-scalar regression, the proposed method is illustrated to be much more superior with negligible false positive and false negative rate and also providing better predictive performance than existing methods.

The rest of this article is organized in the following way. We present our modeling framework and illustrate the proposed variable selection method for additive function-on-scalar regression in Section 2. In Section 3, we evaluate the performance of the proposed method via simulations and compare it with existing methods of variable selection in FOSR. We demonstrate real data application of the proposed method on the NHANES 2003-04 data in Section 4. Section 5 presents a brief discussion on the contributions of our proposed method and some possible extensions of this work.

2. Methodology

2.1 *Modeling Framework*

We assume that the observed data for the ith subject is $\{Y_i(t), M_{i1}, M_{i2}, \ldots, M_{iq}\}$ $(i = 1, 2, \ldots, n)$, where $Y_i(\cdot)$ represents the functional response and M_{i1},M_{i2},$\ldots$, M_{iq} are the corresponding scalar predictors of interest. Sometimes there might be additional control or confounding variables (e.g. demographic variables) X_{i1},X_{i2},$\ldots$, X_{ip}, which we want to adjust for. In practice, we observe the functional response only on finitely many time points over some closed and bounded interval. In developing our method we assume the functions are observed on a dense and regular grid of points $S = \{t_1, t_2, \ldots, t_m\} \subset \mathcal{T} = [a, b]$ for some $a, b \in \mathbb{R}$, although this can be extended to accommodate more general scenario where the functional response is observed on irregular and sparse domain. The commonly used linear FOSR model is

$$Y_i(t) = \mu(t) + \sum_{j=1}^{p} X_{ij}\beta_j(t) + \sum_{\ell=1}^{q} M_{i\ell}\theta_\ell(t) + \epsilon_i(t), \quad \text{for } i = 1, \ldots, n, \quad t \in \mathcal{T}. \tag{1}$$

Here $\beta_j(t)$ and $\theta_\ell(t)$ are smooth coefficient functions which represent the dynamic effects of the scalar predictors X_{ij} and $M_{i\ell}$, respectively. The coefficient functions $\beta_j(\cdot)$,$\theta_\ell(\cdot)$, functional response $Y_i(\cdot)$ and the error process $\epsilon_i(\cdot)$ are assumed to lie in a real separable Hilbert space $\mathcal{H}$ (Barber et al., 2017; Parodi and Reimherr, 2018). In this paper, we focus our attention to $L^2(\mathcal{T})$, although the model and the proposed approach can be used in other functional spaces as well. We assume that the covariates are centered and $\mu(\cdot)$ is a smooth function capturing marginal mean of the functional response $Y(\cdot)$. We further assume the error functions $\epsilon_i(\cdot)$ are i.i.d. copies of $\epsilon(\cdot)$ which is a mean zero stochastic process with unknown nontrivial covariance structure. The goal in such model is to simultaneously estimate $\theta(\cdot)$ and have sparsity in the variables of interest. As mentioned in Section 1, model (1) only considers linear effects of the the scalar predictors of interest, whereas in many applications (e.g., child growth curve) the effects might as well be non linear. We propose the following generalized

function-on-scalar regression (GFOSR) model as a generalization of model (1),

$$Y_i(t) = \mu(t) + \sum_{j=1}^{p} X_{ij}\beta_j(t) + \sum_{\ell=1}^{q} \theta_\ell(M_{i\ell}, t) + \epsilon_i(t). \tag{2}$$

Here $\theta_\ell(\cdot,\cdot)$ are unknown smooth functions (twice differentiable in both arguments) on $\mathcal{R}\times\mathcal{T}$ capturing the dynamic effects of the predictors $M_{i\ell}$. Similar to model (1), here the objective is to simultaneously estimate the nonparametic functions $\theta_\ell(\cdot,\cdot)$ and have sparsity in the variables of interest, $M_{i\ell}$. We illustrate the variable selection approaches for model (1) (FOSR) and model (2) (GFOSR) in the following section.

2.2 *Variable Selection Method*

Our strategy is to consider the marginal functional principal component analysis (FPCA) of $Y(\cdot)$ and reduce the model in (1) by projecting it onto the resulting eigenfunctions. In particular, we approximate the true curves $Y_i(\cdot)$ using the truncated Karhunen-Loève expansion approximation (Karhunen, Loeve 1946) as $Y_i(t) \approx \mu(t) + \sum_{k=1}^{K} \zeta_{ik}\psi_k(t)$, where ζ_{ik} are the functional principal component (FPC) scores and $\psi_k(t)$ are the orthogonal eigenfunctions.

Specifically, we can see that $\zeta_{ik} = \int_{\mathcal{T}}(Y_i(t) - \mu(t))\psi_k(t)dt$. Thus the model in (1) becomes

$$\begin{aligned}
\zeta_{ik} &= \sum_{j=1}^{p} X_{ij}\int_{\mathcal{T}} \beta_j(t)\psi_k(t)dt + \sum_{\ell=1}^{q} M_{i\ell}\int_{\mathcal{T}} \theta_\ell(t)\psi_k(t)dt + \int_{\mathcal{T}} \epsilon_i(t)\psi_k(t)dt \\
&= \sum_{j=1}^{p} X_{ij}\beta_{jk} + \sum_{\ell=1}^{q} M_{i\ell}\theta_{\ell k} + \epsilon_{ik},
\end{aligned}$$

for $k = 1,\ldots,K$, where we define $\beta_{jk} = \int_{\mathcal{T}} \beta_j(t)\psi_k(t)dt$, and similarly $\theta_{\ell k}$. Denote $\mathbf{X}_i^T = (X_{i1}, X_{i2}, \ldots, X_{ip})$ and similarly $\mathbf{M}_i^T$. Thus we can write the FOSR model as

$$\boldsymbol{\zeta}_i = \mathbb{U}_i\boldsymbol{\beta} + \mathbb{V}_i\boldsymbol{\theta} + \boldsymbol{\epsilon}_i, \tag{3}$$

where $\mathbb{U}_i = I_{K\times K} \otimes \mathbf{X}_i^T$, $\mathbb{V}_i = I_{K\times K} \otimes \mathbf{M}_i^T$, $\boldsymbol{\zeta}_i = (\zeta_{i1}, \zeta_{i2}, \ldots, \zeta_{iK})$, $\boldsymbol{\epsilon}_i = (\epsilon_{i1}, \epsilon_{i2}, \ldots, \epsilon_{iK})$, $\boldsymbol{\beta}_k^T = (\beta_{1k}, \ldots, \beta_{pk})$, and similarly $\boldsymbol{\theta}_k$. Here $\otimes$ denotes the Kronecker product. This reduces the linear FOSR model (1) to a linear model of functional principal component scores on

the scalar predictors. Sparsity can now be introduced in the model through appropriate sparsity penalty on the parameter $\boldsymbol{\theta} = (\boldsymbol{\theta}_1^T, \boldsymbol{\theta}_2^T, \ldots, \boldsymbol{\theta}_K^T)^T$, where $\boldsymbol{\theta}_k^T = (\theta_{1k}, \theta_{2k}, \ldots, \theta_{qk})$. In particular, we use the group minimax concave penalty (MCP) (Zhang, 2010) and propose the following penalized least square criterion for performing variable selection.

$$\hat{\boldsymbol{\gamma}} = (\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\theta}}) = \underset{\boldsymbol{\beta},\boldsymbol{\theta}}{\operatorname{argmin}} \sum_{i=1}^{n} ||\boldsymbol{\zeta}_i - \mathbb{U}_i\boldsymbol{\beta} - \mathbb{V}_i\boldsymbol{\theta}||_2^2 + n\sum_{l=1}^{q} P_{MCP,\lambda,\phi}(||\boldsymbol{\theta}_{\ell\cdot}||_2), \tag{4}$$

where $\boldsymbol{\theta}_{\ell\cdot}^T = (\theta_{\ell 1}, \theta_{\ell 2}, \ldots, \theta_{\ell K})$. The MCP term $P_{MCP,\lambda,\phi}(||\boldsymbol{\theta}_{\ell\cdot}||_2$ is defined in the following way,

$$P_{MCP,\lambda,\phi}(||\boldsymbol{\theta}_{\ell\cdot}||_2) = \begin{cases} \lambda||\boldsymbol{\theta}_{\ell\cdot}||_2 - \frac{||\boldsymbol{\theta}_{\ell\cdot}||_2^2}{2\phi} & \text{if } ||\boldsymbol{\theta}_{\ell\cdot}||_2 \leqslant \lambda\phi. \\ .5\lambda^2\phi & \text{if } ||\boldsymbol{\theta}_{\ell\cdot}||_2 > \lambda\phi. \end{cases}$$

MCP (Zhang, 2010) has been shown to ensure selection consistency and estimation consistency under standard assumptions in the traditional scalar regression case and it also overcomes the high bias problem of LASSO, which makes it particularly suitable for doing simultaneous variable selection and estimation. Next, we extend this variable selection approach to the generalized function-on-scalar regression (GFOSR) model. We follow a similar strategy using the truncated KL expansion of $Y(\cdot)$ and reformulate the GFOSR model (2) using the FPC scores of the functional response as follows,

$$\begin{aligned} \int_{\mathcal{T}} (Y_i(t) - \mu(t))\psi_k(t)dt &= \sum_{j=1}^{p} X_{ij}\int_{\mathcal{T}} \beta_j(t)\psi_k(t)dt + \sum_{\ell=1}^{q}\int_{\mathcal{T}} \theta_\ell(M_{i\ell}, t)\psi_k(t)dt + \int_{\mathcal{T}} \epsilon_i(t)\psi_k(t)dt \\ \zeta_{ik} &= \sum_{j=1}^{p} X_{ij}\beta_{jk} + \sum_{\ell=1}^{q}\theta_{\ell k}(M_{i\ell}) + \epsilon_{ik}, \end{aligned}$$

for $k = 1, \ldots, K$, where we define $\beta_{jk} = \int_S \beta_j(t)\psi_k(t)dt$, and $\theta_{\ell k}(u) = \int_{\mathcal{T}} \theta_\ell(u,t)\psi_k(t)dt$. Now denoting $\boldsymbol{\beta}_k^T = (\beta_{1k}, \ldots, \beta_{pk})$ and $\mathbf{X}_i^T = (X_{i1}, X_{i2}, \ldots, X_{ip})$, the GFOSR model becomes

$$\zeta_{ik} = \mathbf{X}_i^T\boldsymbol{\beta}_k + \sum_{\ell=1}^{q}\theta_{\ell k}(M_{i\ell}) + \epsilon_{ik}. \tag{5}$$

We model the unknown nonparametric functions $\theta_{\ell k}(\cdot)$ in terms of known basis expansion

as $\theta_{\ell k}(u) = \sum_{j=1}^{J} B_j(u)\delta_{\ell,k,j}$. Model (5) can then be reformulated using this expansion as,

$$\begin{aligned}\zeta_{ik} &= \mathbf{X}_i^T\boldsymbol{\beta}_k + \sum_{\ell=1}^{q}\sum_{j=1}^{J} B_j(M_{i\ell})\delta_{\ell,k,j} + \epsilon_{ik} \\ &= \mathbf{X}_i^T\boldsymbol{\beta}_k + \boldsymbol{\eta}_i^T\boldsymbol{\delta}_{.k.} + \epsilon_{ik},\end{aligned}$$

where we define $\boldsymbol{\delta}_{.k.}^T = (\delta_{1,k,1}, \ldots, \delta_{1,k,J}, \ldots, \delta_{q,k,J})$ and $\boldsymbol{\eta}_i^T = \{B_j(M_{i\ell})\}_{j=1,\ell=1}^{J,q}$. Hence the GFOSR model can be written similarly as the FOSR model as,

$$\boldsymbol{\zeta}_i = \mathbb{U}_i\boldsymbol{\beta} + \mathbb{V}_i\boldsymbol{\delta} + \boldsymbol{\epsilon}_i, \tag{6}$$

where $\mathbb{U}_i = I_{K\times K} \otimes \mathbf{X}_i^T$, $\mathbb{V}_i = I_{K\times K} \otimes \boldsymbol{\eta}_i^T$ and $\boldsymbol{\delta} = (\boldsymbol{\delta}_{.1.}^T, \boldsymbol{\delta}_{.2.}^T, \ldots, \boldsymbol{\delta}_{.K.}^T)^T$. Similarly as in model (3), sparsity in the variables $M_{i\ell}$ can be introduced now though a group MCP on the parameter $\boldsymbol{\delta} = (\boldsymbol{\delta}_{.1.}^T, \boldsymbol{\delta}_{.2.}^T, \ldots, \boldsymbol{\delta}_{.K.}^T)^T$. The parameters are estimated in a similar manner using a penalized least square criterion given by

$$\hat{\boldsymbol{\gamma}} = (\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\delta}}) = \underset{\boldsymbol{\beta},\boldsymbol{\delta}}{\operatorname{argmin}} \sum_{i=1}^{n} ||\boldsymbol{\zeta}_i - \mathbb{U}_i\boldsymbol{\beta} - \mathbb{V}_i\boldsymbol{\delta}||_2^2 + n\sum_{l=1}^{q} P_{MCP,\lambda,\phi}(||\boldsymbol{\delta}_{\ell.}||_2), \tag{7}$$

where $\boldsymbol{\delta}_{\ell.}^T = (\delta_{\ell,1,1}, \ldots, \delta_{\ell,1,J}, \delta_{\ell,2,1} \ldots, \delta_{\ell,2,J}, \ldots, \delta_{\ell,K,J})$.

We use FPCA methods of Yao et al. (2005) for estimation of the scores ζ_{ik}, functional mean $\mu(\cdot)$ and the eigenfunctions $\psi_k(\cdot)$. The methods are based on estimating the mean and covariance functions using local linear smoothing and subsequently estimating the eigenvalues and eigenfunctions from a spectral decomposition of the estimated covariance matrix. Subsequently the scores $\zeta_{ik} = \int_S (Y_i(t) - \mu(t))\psi_k(t)dt$ can be estimated using numerical integration. For sparse longitudinal data, possibly observed with measurement error the scores can be obtained using the PACE (principal analysis by conditional estimation) method (Yao et al., 2005) which ensures the estimated scores asymptotically goes to the BLUP of the original scores. These estimates are then put together using Karhunen-Loève expansion to get estimates $\hat{Y}_i(\cdot)$ of the true curve $Y_i(\cdot)$.

The number of eigenbasis K and the number of basis functions J (B-splines in this article)

work as a tuning parameter of the minimization problem, which controls the smoothness of the function $\theta_l(\cdot,\cdot)$ in u and t directions respectively and we implicitly control them following a truncated basis approach (Ramsay and Silverman, 2005; Fan et al., 2015). We choose them in a data-driven way using V-fold ($V = 5$ in this article) cross-validation. We use the Extended BIC (EBIC) (Chen and Chen, 2008) criterion of a corresponding Gaussian likelihood for choosing the penalty parameter λ. Chen and Chen (2008) established selection consistency of EBIC and also illustrated its superiority over other methods, e.g., cross-validation, AIC, and BIC, which tend to over select the variables. Although we make no distributional assumption, this has shown satisfactory performance in our empirical analysis in terms of the reported true positive and false positive rates. For the tuning parameter ϕ, we use the value 3 for MCP, on the recommendation by the original author (Zhang, 2010) and as the default value for the MCP based variable selection methods in FOSR (Chen et al., 2016). The `grpreg` package (Breheny and Huang, 2015) in R is used (R Core Team, 2018) for implementation of the group MCP in FOSR and GFOSR. The FPC scores and other FPC objects are estimated using the "fpca.sc" function within the `Refund` package (Goldsmith et al., 2018) in R.

Remark 1:

The proposed variable selection methods for FOSR and GFOSR provide a uniform framework for variable selection in additive function-on-scalar regression. In the linear FOSR model, we allow for a few confounding factors which are not penalized and denote them as X. If there is no need for such adjustment, there would be no X, while M will include all the independent variables. The X variables offer an additional option to have predictors without any penalization. Next, the GFOSR model builds upon the linear FOSR model to allow the penalized variables M to have nonlinear effects, modeled using unknown bivariate functions. It is plausible that there could be other combinations of the above two scenarios where

we would want to penalize both linear (e.g., categorical covariates) and nonlinear effects of covariates and allow the confounding factors to have linear or nonlinear effects. One can extend the proposed variable selection framework for FOSR and GFOSR to perform variable selection even in such scenarios. We illustrate a specific example of this type in our NHANES data application.

3. Simulation Study

3.1 *Simulation Setup*

In this section, we investigate the performance of the proposed variable selection method using simulations. We evaluate our method in terms of selection accuracy, estimation accuracy, and out of sample prediction performance. We first consider the FPC based linear FOSR method proposed in this article and compare it with existing methods for FOSR (Chen et al., 2016; Fan and Reimherr, 2017; Parodi and Reimherr, 2018). We then present a comparison of the proposed variable selection method for nonlinear FOSR with linear FOSR to illustrate its advantages. We consider the following simulation designs.

Scenario A1, Linear FOSR, $n > p$:

We generate data from the model,

$$Y_i(t) = \mu(t) + \sum_{j=1}^{2} X_{ij}\beta_j(t) + \sum_{\ell=1}^{20} M_{i\ell}\theta_\ell(t) + \epsilon_i(t), \text{for } \; i = 1, \ldots, n, \;\; t \in [0, 100],$$

where we have the coefficient functions given by $\mu(t) = 8\, sin(\pi t/50)$, $\beta_1(t) = 3 + 5t/100$, $\beta_2(t) = 4sin(\pi t/50) + 4cos(\pi t/50)$, $\theta_1(t) = 25\, exp(-t/100)$, $\theta_2(t) = 1 + 2(t/100) + (t/100)^2$, $\theta_3(t) = 5 + 7(t/100)$, and $\theta_j(t) = 0$ for $j = 4, 5, 6, \ldots, 20$, i.e., the last 17 predictors are not relevant. This is exactly the linear FOSR model (1) considered in this article. The exogenous covariates $X_{ij} \overset{i.i.d}{\sim} Unif(-2j, 2j)$, and the predictors of interest $M_{ij} \overset{i.i.d}{\sim} N(0, 5^2)$. The error process $\epsilon_i(t)$ is generated as follows;

$$\epsilon_i(t) = \xi_{i1}cos(t) + \xi_{i2}sin(t) + N(0, 1^2 I_{m_i}),$$

where $\xi_{i1} \overset{iid}{\sim} \mathcal{N}(0, .5^2)$ and $\xi_{i2} \overset{iid}{\sim} \mathcal{N}(0, 0.75^2)$. The response $Y_i(t)$ is observed on a grid of $m = 81$ equidistant time points in $S = [0, 100]$. Sample size $n \in \{100, 200\}$ are considered for this design. Web Figure 1 in the supporting information displays the coefficient functions $\beta_j(t), \theta_\ell(t)$ and few observed trajectories of the response function $Y_i(t)$.

Scenario A2, Linear FOSR, $p > n$:

Next, we consider a high dimensional ($p > n$) linear FOSR model to illustrate the performance of the proposed method for linear FOSR in high dimensions. We follow the high dimensional ("rough") simulation design in Fan and Reimherr (2017); Parodi and Reimherr (2018),

$$Y_i(t) = \sum_{\ell=1}^{1000} M_{i\ell}\theta_\ell(t) + \epsilon_i(t), \text{for } \; i = 1, \ldots, n, \;\; t \in [0, 100].$$

Here, the coefficient functions $\theta_\ell(\cdot)$ are generated from a Matérn process (mean zero Gaussian process) with smoothness parameter $\nu = 2.5$, point-wise variance $\sigma^2 = 1$ and the range $\tau = 0.25$. The covariance kernel of a general Matérn process with parameters ν, σ^2, τ is given by $C(t, s) = \sigma^2 \frac{2^{1-\nu}}{\Gamma(\nu)} (\sqrt{2\nu}\frac{d}{\tau})^\nu K_\nu(\sqrt{2\nu}\frac{d}{\tau}), \;\; d = |t-s|$, where K_ν is the modified Bessel function of the second kind. In Particular, for $\nu = 2.5$, the Matérn covariance kernel has a simplified expression given by

$$C(t, s) = \sigma^2(1 + \frac{\sqrt{5}d}{\tau} + \frac{5d^2}{3\tau^2})\, exp(-\frac{\sqrt{5}d}{\tau}), \;\; d = |t - s|, \;\; \nu = 2.5$$

Among the 1000 scalar predictors only $I_0 = 10$ are significant ones, rest of the $\theta_\ell(t)$ are set to zero. The predictors of interest $M_{ij} \overset{i.i.d}{\sim} N(0, 1^2)$. The error process $\epsilon(\cdot)$ are generated from a similar Matérn process as $\theta_\ell(\cdot)$ with $\nu = 1.5$ and rest of the parameters exactly the same. In this case, the covariance kernel is given by,

$$C(t, s) = \sigma^2(1 + \frac{\sqrt{3}d}{\tau})\, exp(-\frac{\sqrt{3}d}{\tau}), \;\; d = |t - s|, \;\; \nu = 1.5.$$

Sample size $n = 500$ is considered for this simulation design.

Scenario B, Nonlinear FOSR:

In this scenario, We generate data from the nonlinear FOSR model,

$$Y_i(t) = \mu(t) + \sum_{j=1}^{2} X_{ij}\beta_j(t) + \sum_{\ell=1}^{20} \theta_\ell(M_{i\ell}, t) + \epsilon_i(t).$$

Here, we have the coefficient functions given by $\mu(t) = 8\, sin(\pi t/50)$, $\beta_1(t) = 3 + 5t/100$, $\beta_2(t) = 4sin(\pi t/50) + 4cos(\pi t/50)$, the nonparametric functions are given by $\theta_1(x,t) = 2x^3\, exp(t/100)$, $\theta_2(x,t) = 5(x + x^3)sin(\pi t/100)$, $\theta_3(x,t) = 16sin(xt/100)$. Rest of the nonlinear functions $\theta_l(x,t)$ are set to zero. The exogenous covariates $X_{ij} \overset{i.i.d}{\sim} Unif(-2j, 2j)$, and the predictors of interest $M_{ij} \overset{i.i.d}{\sim} N(0, 1^2)$. The error process $\epsilon_i(t)$ is generated exactly as in scenario A1 for linear FOSR. The response $Y_i(t)$ is observed on a grid of $m = 81$ equidistant time points in $S = [0, 100]$. Sample size $n = 200$ is considered for this design, out of which 80% is used for model training and rest as a test set for comparing out of sample prediction performance. Web Figure 2 and 3 in the supporting information display the coefficient functions $\beta_j(t)$,nonlinear effects $\theta_\ell(x,t)$ and few observed trajectories of the response function $Y_i(t)$.

We generate 500 replicated data sets from scenario A1 and 100 replicated datasets from scenario A2 and Scenario B to assess the performance of the proposed variable selection methods. For the linear FOSR method, the number of eigenbasis K is chosen using cross-validation on a grid $K \in \{2, 3, 4, \ldots, 20\}$. For the GFOSR method the values of K and J (number of B-spline basis) were chosen from a two dimensional grid with $K \in \{3, 4, 5, 6, 7\}$ and $J \in \{5, 7, 9\}$. We use $V = 5$ fold cross-validation for choosing the tuning parameters K, J throughout this article. Also we note that the eigenfunctions $\psi_k(t)$ were estimated only using the training samples.

3.2 *Simulation Results*

Scenario A1:

We evaluate the performance of the proposed variable selection method in terms of selection

accuracy and estimation accuracy. We compare the proposed method for linear FOSR (denoted henceforth as FPC-FOSR) with the variable selection method for FOSR by Chen et al. (2016) (implemented using `fosr.vs` function within `Refund` package in R). For the `fosr.vs` implementation (Chen et al., 2016), the penalty parameter λ is chosen by cross-validation and number of basis functions $nbasis = 10$ is used as per the default option. The tuning parameter ϕ is set to 3 as in our analysis. Table 1 reports the selection performances of the variable selection methods in terms of the average number of true positives, false positives, and average model size in the 500 Monte-Carlo replications. Web Table 1 in the supporting information reports the selection percentage in details for each of the scalar predictor.

[Table 1 about here.]

We can observe, the proposed selection method for linear FOSR performs similarly as "fosr.vs" method with negligible false positive rate and practically zero false negative rate. The average model sizes for the proposed method are marginally lower and closer to the true model size 3, illustrating more parsimonious and accurate model selection.

Comparison of estimation accuracy of the methods in terms of Monte-Carlo mean square error of the estimated coefficient functions $\hat{\theta}_\ell(\cdot)$ $(\ell = 1, 2, 3)$ and $\hat{\beta}_j(\cdot)$ $(j = 1, 2)$ is reported in Web Table 2. We notice the proposed method for linear FOSR (FPC-FOSR) performs competitively and provides smaller and negligible mean square error for the majority of the coefficient functions. The performance of the variable selection method improves with increasing sample size indicating consistency of the proposed method. The Monte Carlo (MC) mean estimates (averaged estimated coefficient function over 500 replications) of the coefficient functions $\theta_j(t)$ $(j = 1, 2, 3)$ are displayed in Figure 1. The estimated coefficient functions are superimposed on the true curves indicating satisfactory performance of the proposed method in terms of estimation.

[Figure 1 about here.]

Finally, we compare the two methods in terms of their runtime (benchmarked over 100 replications) to illustrate their computational cost. For $n = 100$ (200), the proposed FPC-FOSR method on an average takes 6.56 (10.95) seconds compared to 19.22 (32.11) seconds for fosr.vs. In particular, the average runtime for fosr.vs is found to be around 90% longer illustrating the computational gain of the proposed method. Overall, our simulation results illustrate competitive performance of the FPC-FOSR method while also highlighting its computational efficiency.

Additional Simulations: Web Table 3 in the supporting information illustrates a comparison among three different information criteria in terms of their selection performance for the proposed FPC-FOSR method in the paper. We notice a superior performance of EBIC compared to BIC and AIC in terms of the reported average number of true positives (higher) and false positives (smaller). Web Table 4 presents the selection results from the FPC-FOSR method for different choices of the tuning parameter ϕ. The selection performance is not much sensitive to the choice of ϕ particularly around the recommended value of $\phi = 3$. Web Table 5 reports the selection results from the FPC-FOSR method and the fosr.vs method for different values of σ_w, the standard deviation of the white noise in the error process. We observe an increased false positive rate of the `fosr.vs` method by (Chen et al., 2016), while the results from the proposed FPC-FOSR method remain robust.

Scenario A2, Linear FOSR, $p > n$:

In this scenario, the goal is to demonstrate applicability of the proposed variable selection method for high dimensional linear FOSR. The "fosr.vs" method by Chen et al. (2016) no longer works in this high dimensional setting. We provide comparison with the "AFSL" method by (Fan and Reimherr, 2017) and "FLAME" method by (Parodi and Reimherr, 2018) developed specifically for such scenario. For the FLAME and AFSL method, we directly use the results from Fan and Reimherr (2017) and Parodi and Reimherr (2018) for their

high dimensional setting. Table 2 displays comparison of the selection methods in terms of average number of true positives, false positives and prediction error on data defined as the L_2 distance between the true response function and the prediction; $E = \sum_{i=1}^{n} ||Y_i^*(\cdot) - \hat{Y}_i(\cdot)||_{L_2}$.

[Table 2 about here.]

We notice the proposed method FPC-FOSR has higher true positive rate and smaller false positive rate compared to AFSL and FLAME while giving lower prediction error than both the methods. The results illustrate satisfactory performance of the proposed variable selection method for linear FOSR even in high dimensions.

Scenario B, Nonlinear FOSR:

In this scenario, the effects of the predictors of interest are nonlinear. We evaluate the performance of the proposed method for variable selection in GFOSR (denoted as FPC-GFOSR) in terms of selection accuracy and out of sample prediction performance. We provide a comparison with existing variable selection methods for linear FOSR; "fosr.vs" (Chen et al., 2016) and FPC-FOSR developed in this article to illustrate the advantage of the proposed method. Table 3 reports the selection performances of the variable selection methods in terms of the average number of true positives, false positives and average model size in the 100 Monte-Carlo replications. Web Table 6 in the supporting information reports the selection percentage in details of each of the scalar predictor.

[Table 3 about here.]

We notice that the linear FOSR method "fosr.vs" by Chen et al. (2016) has much higher false positive rate and average model size compared to the fpc based FOSR approahes proposed in this article. FPC-FOSR, the method developed for linear FOSR does reasonably well even in the presence of nonlinear effects in terms of variable selection. The FPC-GFOSR method outperforms both these methods and produces oracle performance in terms of identifying the true model in all the replications. A Comparison of of the variable selection

methods in terms of out of sample prediction accuracy using the measure R^2 defined as

$$R^2 = 1 - \sum_{i=1}^{n}\sum_{j=1}^{m}\{Y_i(t_{ij}) - \hat{Y}_i(t_{ij})\}^2 / \sum_{i=1}^{n}\sum_{j=1}^{m}\{Y_i(t_{ij}) - \hat{\mu}(t_{ij})\}^2$$

is displayed in Table 4.

[Table 4 about here.]

The FPC-GFOSR method produce much higher out of sample R^2 (median) than the other two methods, which can be attributed to its capability of successfully identifying the nonlinear effects of the predictors. The simulation results demonstrate the advantages of the proposed method for variable selection in FOSR, particularly in the presence of nonlinear time-varying effects of the predictors. The FPC-GFSOR method provides a unified and generalized framework for variable selection in FOSR.

4. NHANES Accelerometry Data Application

We study associations between diurnal patterns of physical activity (PA) and demographic, lifestyle and health characteristics of the participants from the 2003–2004 cohorts of accelerometry data from the National Health and Nutrition Examination Survey (NHANES). Data is obtained from the `rnhanesdata` (Leroux et al., 2020) R package. We have minute level accelerometry data on each of the participants for seven days (Monday-Sunday), along with information on device calibration, data reliability, and wear non-wear flags calculated by an algorithm in Troiano et al. (2008). The data was originally collected using a hip worn Actigraph AM-7164 (Actigraph, Ft. Walton Beach, FL) uni-axial accelerometer following an existing set of protocols (Troiano et al., 2008). As covariates, we have information on age, gender, BMI, Race, and various other clinical and demographic characteristics of the participants. The complete list of the predictors along with their description and statistical summaries are given in Web Table 7.

We consider only the reliable wear-time data for our analysis and derive minute level

averages of physical activity count over the seven days. This gives minute level PA data on 3627 participants. We restrict our attention to activity data between 8 a.m to 10 p.m because of a large number of missing values outside this window, and since this is the time-period we are most interested in to understand the diurnal patterns of physical activity. The raw minute level accelerometry data is noisy, to extract and understand diurnal patterns, we pre-process the data using 2-hour moving windows (8 a.m-10 a.m, 8.15 a.m- 10.15 a.m, ..., 8 p.m-10 p.m) with 15 minutes of move length. This creates 49 windows between 8 a.m-10 p.m with an overlap of 90 minutes between two consecutive windows, ensuring smooth transition between them. Finally, we log-transform the activity profiles and obtain daily activity profiles $Y_i(t)$, where t belongs to 49 equispaced grid points between 9 (midpoint of the first window, 8-10 hours in the morning) to 21 (midpoint of last window, 20-22 hours at night). Figure 2 displays the raw minute level physical activity data and the pre-processed smooth data (both log-transformed) for three representative study subjects along with the mean activity profile of the participants. We notice the diurnal mean activity profile to be decreasing, particularly late in the day, which is expected.

[Figure 2 about here.]

We want to identify the variables influencing diurnal patterns of physical activity and estimate their possible time-varying effects. We use the following function-on-scalar regression model combining model (1) (FOSR) and model (2) (GFOSR) of this article for modeling smoothed diurnal patterns of physical activity using linear effects of the categorical predictors and nonlinear effects of continuous predictors.

$$Y_i(t) = \mu(t) + G_i\eta(t) + \sum_{j=1}^{p} X_{ij}\beta_j(t) + \sum_{\ell=1}^{q} \theta_\ell(M_{i\ell}, t) + \epsilon_i(t). \tag{8}$$

Here G_i is the indicator of gender (female) of the ith participant, which we treat as a confounding variable (not penalized). Prior research in the literature (Varma et al., 2017; Xiao et al., 2015) has indicated significant gender specific differences in PA. Here X_{ij} ($j =$

$1, 2, \ldots, p = 16$) are the categorical predictors (dummy encoded) with linear effects $\beta_j(t)$, and $M_{i\ell}$ ($\ell = 1, 2, \ldots, q = 3$) are the continuous predictors. We use the proposed variable selection method for GFOSR in this article combining it with the FPC-FOSR approach for linear effects. The number of basis functions K and J were chosen to be 3, 9, respectively by five-fold cross-validation,

The variables Race (Mexican American, Other Hispanic), "Diabetes Yes", "CHF Yes" (congestive heart failure), "Stroke No" and "Mobility Problem Any Difficulty" are selected among the categorical predictors, from the continuous variables only age is selected as an important driver of PA. The linear functional effects $\beta_j(t)$ of the selected categorical predictors are displayed in Figure 3.

[Figure 3 about here.]

We notice that the coefficient function corresponding to gender (female) is negative through most of the day (8 a.m- 7 p.m) indicating lower levels of PA for females compared to males during these time-period. Interestingly the females are found to be more active in the late evening ($\approx$ 7 p.m onwards). The coefficient functions corresponding to Race "Mexican American" and "Other Hispanic" are positive, indicating they have higher physical activity compared to the reference group, although the differences again decrease as the day progress. The coefficient function corresponding to "Diabetes Yes" is negative, indicating people with diabetes has lower levels of PA through out the day and the coefficient function attains a minimum around afternoon indicating this is where the largest difference in PA occurs (compared to subjects with no Diabetes). We see a similar coefficient function for subjects having congestive heart failure in the past, with the strongest difference between the groups (CHF Yes/No) occurring in the late afternoon. Coefficient functions corresponding to 'Stroke NO" group has the opposite trend, higher PA for subjects with no stroke, maximum difference occurring again in the afternoon. Subjects with any difficulty in mobility, as expected, have

lower levels of PA as illustrated by the coefficient function of the predictor "Mobility Problem Any Difficulty".

The nonparametric effect of the continuous predictor age on diurnal mean activity is displayed in Figure 4.

[Figure 4 about here.]

It can be noticed at any particular time of day, the levels of physical activity increase during adulthood (age: 20-30 years), stabilizes during mid-years (age: 31- 59) and starts decreasing after age 60. The dependence is highly nonlinear, illustrating the importance of the GFOSR model, which can capture the nonlinear dependence of diurnal pattern of physical activity on age. Our findings match with Varma et al. (2017), where similar dependence of physical activity on age was noticed while modelling scalar summaries of physical activity such as TLAC (total log-transformed activity count). Our analysis in this section, using the proposed variable selection method for GFOSR, captures the important drivers influencing diurnal patterns of physical activity while simultaneously estimating their time-varying effects.

Remark 2:

The variable selection method for linear FOSR (FPC-FOSR) was also applied for comparison and yielded an in-sample R-square 0.22, in comparison, model (8) which combines both FOSR and GFOSR, provided an in-sample R-square 0.26 outlining the nonlinear dependence of physical activity on age.

5. Discussion

In this article, we have provided a uniform framework for variable selection in additive function-on-scalar regression. The proposed work constitutes a new approach for variable selection in linear function-on-scalar regression and a new method for variable selection in nonlinear function-on-scalar regression. Through numerical simulations, we have shown the

proposed variable selection method for FOSR and GFOSR identifies the true underlying variables with minimal false positive and minuscule false negative rate even in higher dimensions and fares extremely well in terms of out of sample prediction accuracy against the competing methods. Results from the simulation Scenario B illustrate that usual pre-whitening of linear FOSR (Chen et al., 2016) can be counter-productive in the presence of nonlinear effects, as the initial model for estimating the covariance matrix becomes misspecified in such a situation. We have demonstrated application of the proposed method on accelerometer data from the NHANES 2003–2004 cohorts in identifying clinical and demographic drivers of diurnal patterns of physical activity. Our analysis using GFOSR reveals a nonlinear dependence of physical activity on age which is interesting from a physiological viewpoint.

In developing our method, we have assumed the functional response $Y(\cdot)$ is observed on a dense and equispaced grid. The proposed method for variable selection extends to a more general scenario when the response is observed on an irregular and sparse domain using functional principal component analysis (FPCA) methods for sparse longitudinal data (Yao et al., 2005). In particular, the scores can be obtained using the "PACE" method (Principal Analysis by Conditional Expectation). Subsequently, FPC-FOSR and FPC-GFOSR method developed in this article can be applied with these scores for performing variable selection. In this article, we have controlled the smoothness of the coefficient functions following a truncated basis approach (Ramsay and Silverman, 2005; Fan et al., 2015). An alternative could be directly penalizing the roughness of the coefficient functions via added smoothness penalties. For example, an elastic net type procedure can be used with a mixture of ridge and L_1 penalty on the basis coefficients.

Multiple research directions could be explored based on our current work. The models proposed for variable selection for function-on-scalar regression have primarily focused on additive effects of scalar predictors, while in reality there might be significant interaction

among them. Functional single index model (Jiang et al., 2011) with possibly time-varying indexes (Luo et al., 2016) will be an appropriate tool for modeling in such cases. Extending the proposed variable selection method to such general class of function-on-scalar regression models can have diverse applications and remain areas for future research.

Acknowledgement

We thank the Editor, an Associate Editor, and two anonymous reviewers for their constructive and helpful suggestions which led to an improved version of our article.

Data Availability Statement

The data that support the findings in this paper are available in the `rnhanesdata` (Leroux et al., 2020) R package at https://github.com/andrew-leroux/rnhanesdata.

References

Barber, R. F., Reimherr, M., Schill, T., et al. (2017). The function-on-scalar LASSO with applications to longitudinal GWAS. *Electronic Journal of Statistics* **11,** 1351–1389.

Breheny, P. and Huang, J. (2015). Group descent algorithms for nonconvex penalized linear and logistic regression models with grouped predictors. *Statistics and Computing* **25,** 173–187.

Chen, J. and Chen, Z. (2008). Extended Bayesian information criteria for model selection with large model spaces. *Biometrika* **95,** 759–771.

Chen, Y., Goldsmith, J., and Ogden, R. T. (2016). Variable selection in function-on-scalar regression. *Stat* **5,** 88–101.

Fan, Y., James, G. M., and Radchenko, P. (2015). Functional additive regression. *The Annals of Statistics* **43,** 2296–2325.

Fan, Z. and Reimherr, M. (2017). High-dimensional adaptive function-on-scalar regression. *Econometrics and Statistics* **1,** 167–183.

Goldsmith, J. and Kitago, T. (2016). Assessing systematic effects of stroke on motorcontrol by using hierarchical function-on-scalar regression. *Journal of the Royal Statistical Society. Series C, Applied Statistics* **65,** 215.

Goldsmith, J., Liu, X., Jacobson, J., and Rundle, A. (2016). New insights into activity patterns in children, found using functional data analyses. *Medicine and Science in Sports and Exercise* **48,** 1723.

Goldsmith, J., Scheipl, F., Huang, L., Wrobel, J., Gellar, J., Harezlak, J., McLean, M. W., Swihart, B., Xiao, L., Crainiceanu, C., and Reiss, P. T. (2018). *refund: Regression with Functional Data*. R package version 0.1-17.

Hall, P., Müller, H.-G., and Wang, J.-L. (2006). Properties of principal component methods for functional and longitudinal data analysis. *The Annals of Statistics* **34,** 1493–1517.

Jiang, C.-R., Wang, J.-L., et al. (2011). Functional single index models for longitudinal data. *The Annals of Statistics* **39,** 362–388.

Kowal, D. R. and Bourgeois, D. C. (2020). Bayesian function-on-scalars regression for high-dimensional data. *Journal of Computational and Graphical Statistics* **29,** 629–638.

Leroux, A., Crainiceanu, C., Smirnova, E., and Cao, Q. (2020). *rnhanesdata: NHANES Accelerometry Data Pipeline*. R package version 1.02.

Luo, X., Zhu, L., and Zhu, H. (2016). Single-index varying coefficient model for functional responses. *Biometrics* **72,** 1275–1284.

Maas, J., Verheij, R. A., Spreeuwenberg, P., and Groenewegen, P. P. (2008). Physical activity as a possible mechanism behind the relationship between green space and health: a multilevel analysis. *BMC Public Health* **8,** 206.

Parodi, A. and Reimherr, M. (2018). Simultaneous variable selection and smoothing for high-

dimensional function-on-scalar regression. *Electronic Journal of Statistics* **12,** 4602–4639.

R Core Team (2018). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

Ramsay, J. and Silverman, B. (2005). *Functional Data Analysis*. Springer-Verlag, New York.

Reiss, P. T., Huang, L., and Mennes, M. (2010). Fast function-on-scalar regression with penalized basis expansions. *The International Journal of Biostatistics* **6,** 28.

Tibshirani, R. (1996). Regression shrinkage and selection via the Lasso. *Journal of the Royal Statistical Society: Series B (Methodological)* **58,** 267–288.

Troiano, R. P., Berrigan, D., Dodd, K. W., Masse, L. C., Tilert, T., and McDowell, M. (2008). Physical activity in the United States measured by accelerometer. *Medicine and Science in Sports and Exercise* **40,** 181.

Varma, V. R., Dey, D., Leroux, A., Di, J., Urbanek, J., Xiao, L., and Zipunnikov, V. (2017). Re-evaluating the effect of age on physical activity over the lifespan. *Preventive Medicine* **101,** 102–108.

Xiao, L., Huang, L., Schrack, J. A., Ferrucci, L., Zipunnikov, V., and Crainiceanu, C. M. (2015). Quantifying the lifetime circadian rhythm of physical activity: a covariate-dependent functional approach. *Biostatistics* **16,** 352–367.

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.

Zhang, C.-H. (2010). Nearly unbiased variable selection under minimax concave penalty. *The Annals of Statistics* **38,** 894–942.

Supporting Information

Web Tables 1-7 and Web Figures 1-3 referenced in Section 3 and 4 are available with this paper at the Biometrics website on Wiley Online Library. Software illustration of the

proposed method is provided with this paper and also available at https://github.com/rahulfrodo/GFOSR_Selection.

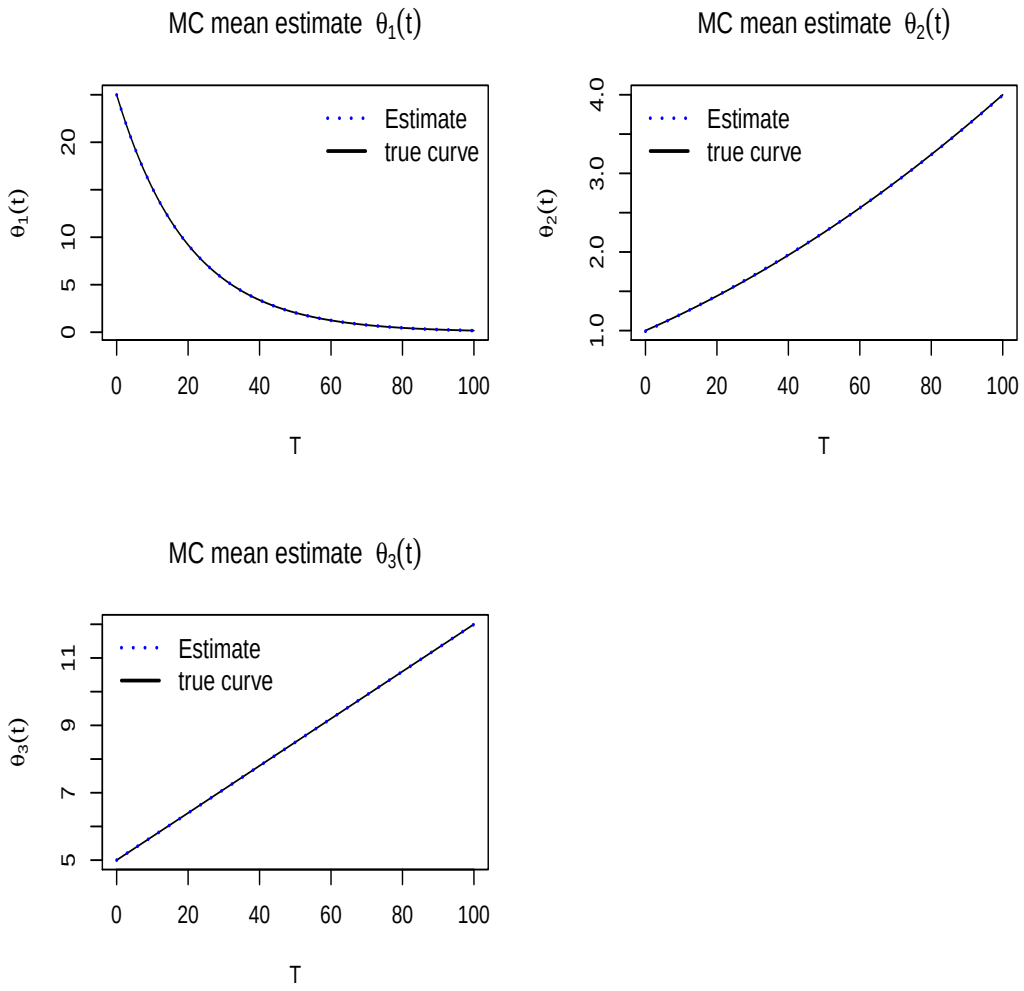

Figure 1: Displayed are the average estimated coefficient functions $\theta_j(t)$ (dashed line) ($j = 1, 2, 3$) (averaged over 500 M.C replications) overlaid on the true curves (solid line), simulation scenario A1.

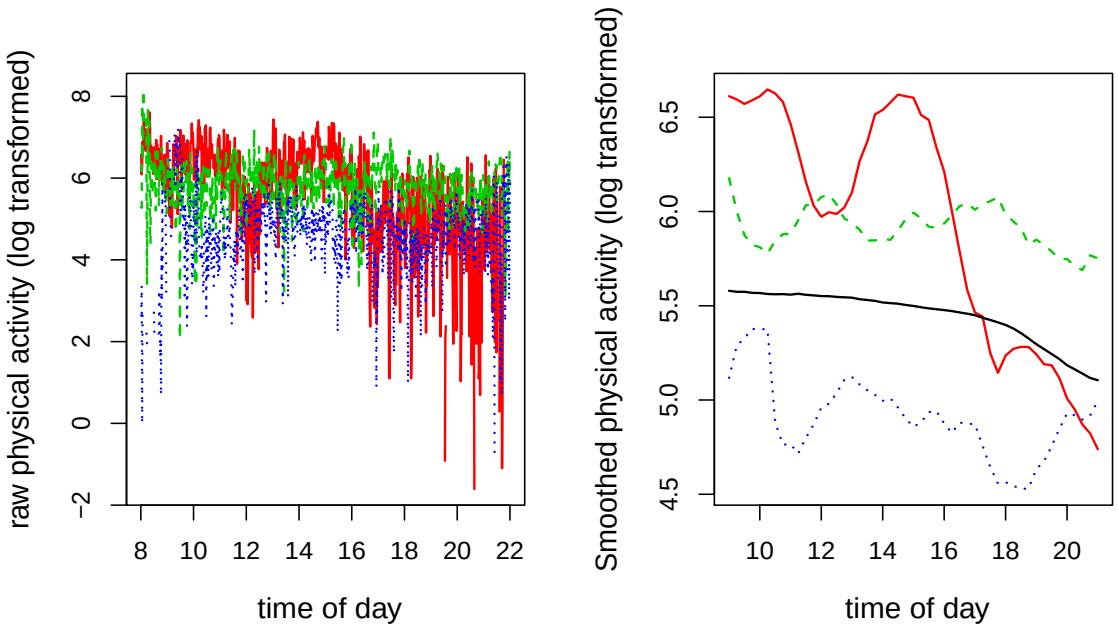

Figure 2: Displayed are the raw minute level activity profile (left column) and smoothed activity profile (right column) of three NHANES subjects. Mean activity profile across all the subjects is shown by the solid black line (right column). This figure appears in color in the electronic version of this article, and any mention of color refers to that version.

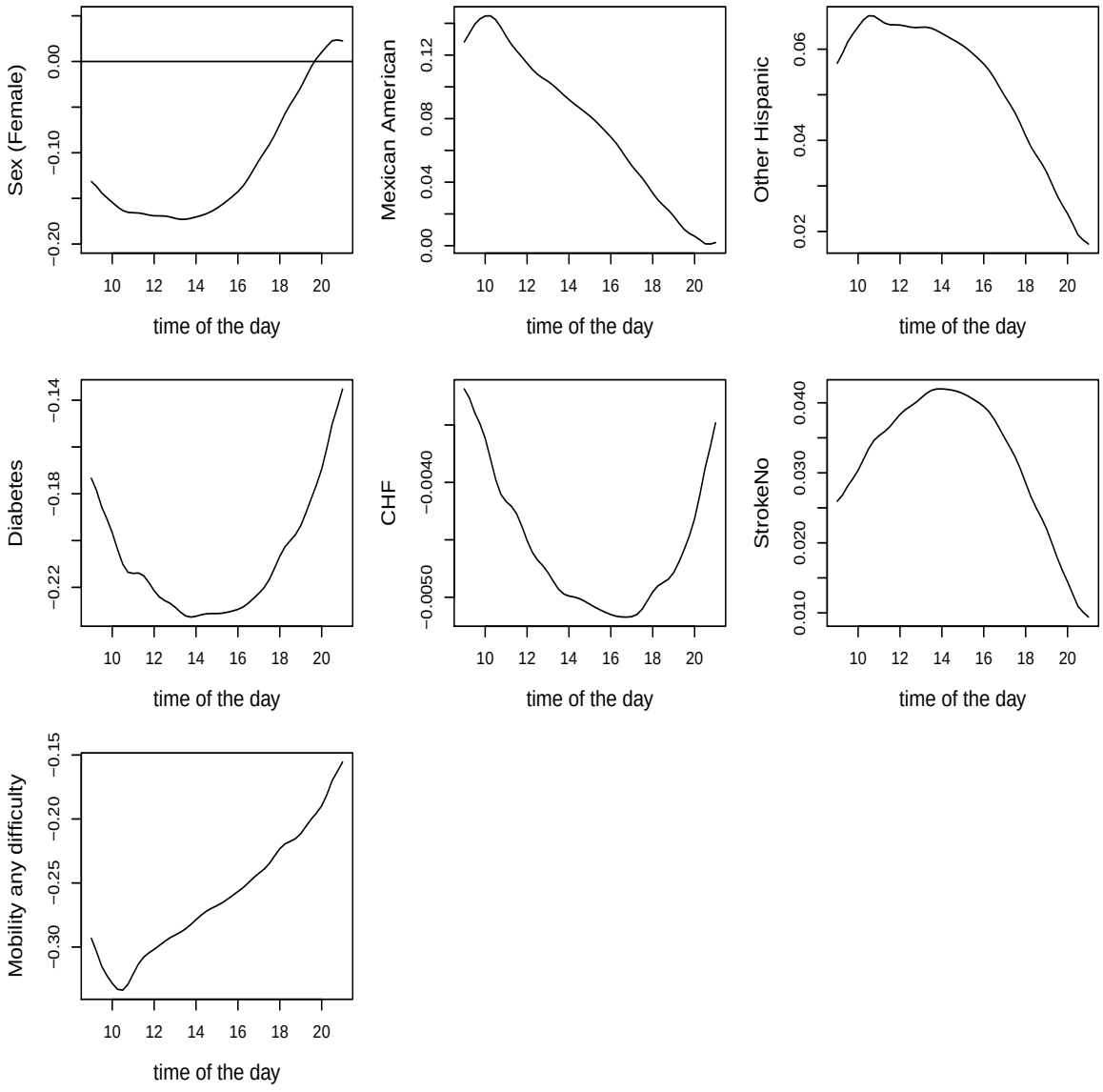

Figure 3: Estimated linear functional effects of gender (female) and the selected categorical predictors, Race (Mexican American), Race (Other Hispanic), Diabetes (Yes), CHF (Yes), Stroke (No) and Mobility Problem Any Difficulty.

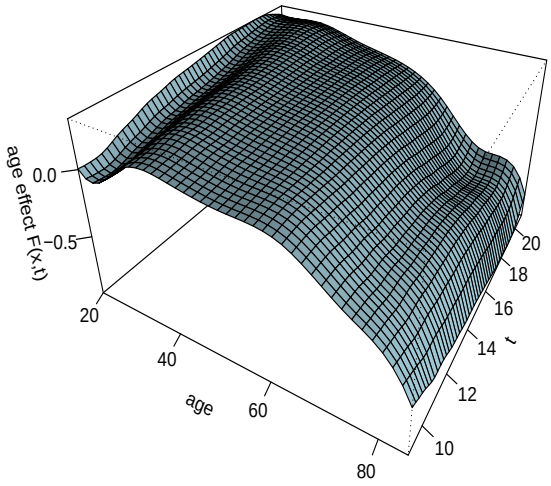

Figure 4: Estimated surface illustrating nonparametric diurnal effect ("t" denoting time of the day) of age on physical activity.

Table 1: Comparison of average number of true positives (TP), false positives (FP) and average model size, Scenario A1. The performance of the proposed method for linear FOSR is shown in FPC-FOSR, and that of competing method (Chen et al., 2016) is shown as fosr.vs.

Sample Size	Method	TP	FP	Model size
n=100	FPC-FOSR	3	0.016	3.016
	fosr.vs	3	0.06	3.06
n=200	FPC-FOSR	3	0.006	3.006
	fosr.vs	3	0.068	3.068

Table 2: Comparison of average number of true positives (TP), false positives (FP) and prediction error of the variable selection methods, n=500, p=1000, $I_0 = 10$, Scenario A2. Values for competing methods are reported from the rough setting of AFSL (Fan and Reimherr, 2017) and FLAME (Parodi and Reimherr, 2018) respectively. A range of values are reported for different choices of tuning parameter for FLAME.

Method	TP	FP	Prediction Error
FPC-FOSR	9.98	0	73.30454
FLAME	8-10	0	80-200
AFSL	9.92	0.08	352.51

Table 3: Comparison of average number of true positives (TP), false positives (FP) and average model size, Scenario B. The performance of the proposed method for nonlinear FOSR is shown in FPC-GFOSR.

Method	TP	FP	Model size
fosr.vs	3	3.63	6.63
FPC-FOSR	3	0.05	3.05
FPC-GFOSR	3	0	3

Table 4: Quantiles of out of sample R^2 based on MC simulation, Scenario B.

Sample Size (train)	R^2 (test)	10%	25%	50%	75%	90%
n=160	fosr.vs	0.6013	0.6904	0.7540	0.7930	0.8269
	FPC-FOSR	0.6691	0.7311	0.7952	0.8195	0.8546
	FPC-GFOSR	0.9960	0.9969	0.9974	0.9979	0.9982