Robust Variable Selection in Functional Linear Models

Jasdeep Pannu *   Nedret Billor†

Abstract

We consider the problem of selecting functional variables using the L1 regularization in a functional linear regression model with a scalar response and functional predictors in the presence of outliers. Since the LASSO is a special case of the penalized least squares regression with L1-penalty function it suffers from the heavy-tailed errors and/or outliers in data. Recently, the LAD regression and the LASSO methods have been combined (the LAD-LASSO regression method) to carry out robust parameter estimation and variable selection simultaneously for a multiple linear regression model. However variable selection of the functional predictor based on LASSO fails since multiple parameters exist for a functional predictor . Therefore group LASSO is used for selecting grouped variables rather than individual variables. In this study we extend the LAD- group LASSO to a functional linear regression model with a scalar response and functional predictors. We illustrate the LAD- group LASSO on both simulated and real data.

Keywords: **Functional Regression Model; LASSO, LAD-LASSO, Outliers, Variable selection**

1 INTRODUCTION

Functional data analysis has become increasingly frequent and important in diverse fields of sciences, engineering, and humanities because most of the data collected these days is functional in nature, for instance, genomics data, fMRI data, DTI, weather data. There has been an evolving literature devoted to understanding the performance of estimation of functional predictors. Escabias et al. [4], Denhere & Billor [3], Boente & Fraiman [2], Gervini [5], Bali et al. [1], Sawant et al. [10], Goldsmith et al. [8] and Ogden & Reiss [9] proposed some robust parameter estimation techniques in functional logistic regression model, functional principal component analysis and generalized functional linear models, respectively.
Just as in ordinary data analysis, variable selection is also an important aspect of functional data analysis. The functional data suffer from high dimensionality and multicollinearity among functional predictors. This could lead us to wrong model selection and hence wrong scientific conclusions. Collinearity also gives rise to issues of over fitting and model misidentification. So it is very important to perform variable selection on functional covariates. With sparsity, variable selection effectively identifies the subset of significant predictors, which improves the estimation

*Department of Mathematics & Statistics, Auburn University, 221 Parker Hall, Auburn University, Alabama 36849, E-mail: jkp0008@auburn.edu

†Department of Mathematics & Statistics, Auburn University, 221 Parker Hall, Auburn University, Alabama 36849, E-mail: billone@auburn.edu

accuracy and therefore, enhances the model interpretability. Not much work has been done in the area of variable selection for functional predictors in functional regression models. Gertheiss et al. [6], Matsui & Konishi [8], Lian [9] and Zhu & Cox [16] proposed some variable selection techniques for functional predictors via L1 and L2 regularizations for instance, using various roughness penalties like groupLASSO, Wavelet Based- LASSO, gSCAD for the generalized functional linear models. However, these methods do not work well in the presence of functional outliers that is, the curves that deviate from the overall pattern of the data. Since these variable selection techniques are not robust in nature thus there is a need for robust variable selection method which is resistant to outlying curves.
Lilly and Billor [10] have proposed group LAD-LASSO for multiple regression model, but to our knowledge, there is no work that has been done in the area of robust variable selection of the functional linear model. In this article we propose a new methodology, by extending the ideas of functional group-LASSO by Gertheiss et al. [6], that minimizes the effect of outlying curves in the estimation and selection of the functional covariates in functional linear models.

2 METHODOLOGY

In this section we will briefly describe functional regression model, estimation of regression parameters and introduce the proposed method.

Functional Regression Model
Functional data are usually sampled discretely over a continuum, usually time and we assume that there is an underlying curve describing data. In the usual functional regression modeling setup, we assume that the response Y_i is scalar for the *i*th subject and $X_1, X_2, ..., X_p$ are the random curves where $X_{i1}, X_{i2}, ..., X_{ip}$ denote their independent realizations, respectively.
For the sake of simplicity, each X_{ij} is considered to be observed without measurement error at a dense grid of time points $\{t_{j1}, t_{j2}, ..., t_{jN_j}\}$.
Then a functional linear model with the scalar response and a p-functional predictors can be defined as :

$$\mathrm{Y}_i = \alpha + \Sigma_{j=1}^{p} \int X_{ij}(t)\beta_j(t)dt + \epsilon_i. \tag{1}$$

The random error terms ϵ_i are assumed to be independent normally distributed with mean 0 and variance σ^2 for i=1,....., n. α is a scalar parameter and $\beta_j(t)$ is a parameter function for j= i,....., p.

Dimension Reduction
The first part of the problem is to consider ways that minimize multicollinearity and reduce the high dimensionality which is inherent with functional data.
We consider the coefficient functions, β_j as belonging to a finite-dimension space generated by the some basis, $\phi_{jb}(t)$, as

$$\beta_j(t) = \Sigma_{b=1}^{q} c_{jb}\phi_{jb}(t). \tag{2}$$

The curves $X_{ij}(t)$ can be also discretized as Riemann Integration as described by Gertheiss et al. [6]

$$\int X_{ij}(t)\beta_j(t)dt \approx \Sigma_m X_{ij}(t_m)\beta_j(t_m). \tag{3}$$

Using (2) and (3), the right side of the model equation in (1) approximates to the following

$$\int X_{ij}(t)\beta_j(t)dt \approx \Sigma_b\{\delta_j \Sigma_m X_{ij}(t_{jm})\phi_j(t_{jm})\}c_{jb} = \Sigma_b \Phi_{ijb} c_{jb} = \mathbf{\Phi_{ij}^T c_j}. \tag{4}$$

where,
i= 1,..., n; j= 1, ..., p; $\delta_j = t_{jm} - t_{j,m-1}$, $\mathbf{c_j} = (c_{j1}, ..., c_{jq})^T$, $\mathbf{\Phi_{ij}} = (\Phi_{ij1}, ..., \Phi_{ijq})^T$, $\Phi_{ijb} = \delta_j \Sigma_m X_{ij}(t_{jm})\phi_j(t_{jm})$.

Then the new model is then obtained as

$$\mathrm{Y}_i = \alpha + \Sigma_{j=1}^p \mathbf{\Phi_{ij}^T c_j} + \epsilon_i. \tag{5}$$

where $\mathbf{\Phi_{ij}}$ are known and α and $\mathbf{c_j}$'s are the unknown regression coefficients that need to be estimated.

Functional LAD-groupLASSO Criterion

The second part of the problem is to consider ways that allow for the parameter selector and estimator to be resistant to outliers.

The classical existing functional variable selection methods like group SCAD, simple group LASSO proposed by Lian [9] and Zhu & Cox [16], respectively, have already been outperformed by group LASSO developed by Gertheiss [6]. However, the method discussed by Gertheiss [6] suffers from the presence of outlying curves, therefore necessitating a different type of approach to handle this issue. We consider a new criterion called functional LAD-groupLASSO to take into account the effect of outliers.

According to this criterion, α and $\mathbf{c_j}(t)$ can be estimated by minimizing

$$\Sigma_{i=1}^n |Y_i - \alpha - \Sigma_{j=1}^p \mathbf{\Phi_{ij}}^T \mathbf{c_j}| + P_{\lambda,\varphi}(\beta_j).$$

where, $P_{\lambda,\varphi}(\beta_j)$ is the penalty function as used by J. Gerthesis et al. [6]

$$P_{\lambda,\varphi}(\beta_j) = \lambda(||\beta_j||^2 + \varphi||\beta_j^{''}||^2)^{1/2}.$$

where

$||.||^2 = \int(.)^2 dt$ is the L^2 norm and $\beta_j^{''}$ is, the second derivative of β_j.

λ is the parameter that controls sparseness and φ is the smoothing parameter that controls smoothness of the coefficients. As the sparseness parameter λ increases, the estimated coefficient functions β(t)'s are shrunk and at some value, set to zero. As the smoothing parameter φ increases, the departure from linearity is penalized stronger and thus the estimated curves become closer to a linear function. Smaller values for φ result in very wiggly and difficult to interpret estimated coefficient functions. For optimal estimates (in terms of accuracy and interpretability), an adequate (λ, φ) combination has to be chosen. λ and φ are selected via *K*-fold cross-validation.

3 NUMERICAL STUDY

In order to show the goodness of the proposed method we We first applied the method to a toy example and then conducted a simulation study. In this section we considered following three cases:

- Case (i): Presence of outliers in the scalar response Y only.
- Case (ii): Presence of outliers both in the scalar response Y and the functional predictors X(t).
- Case (iii): Presence of outliers in the functional predictors X(t) only.

Due to limitation on space, we present only Case (i) and Case (iii) in this article.

Generating data as below:
STEP 1: Generating Functional Predictors $X_j(t)$
We consider only two functional covariates $X_1(t)$ and $X_2(t)$ and generated 50 sample curves for each $X_j(t)$ which are observed at 50 equidistant time points.
Functional Predictors $X_j(t)$s are generated similar to Tutz and Gerthesis (15)

$$X_{ij}(t) = [\sigma(t)]^{-1}\Sigma_{r=1}^{5}(a_{ijr}sin(\pi t(5-a_{ijr})/150) - m_{ijr}).$$
$$where \quad i = 1, ..., 50 \quad and \quad j = 1, 2.$$

Here a_{ijr}~U(0,5), m_{ijr}~ U(0,2*π), $\sigma(t)$ is defined so that $var[X_{ij}(t)] = 0.01$.

STEP 2: Generating Y
Response Y is defined as:

$$Y_i = \alpha + \int_0^{50} \beta_1(t)X_{ij}(t)dt + \epsilon_i.$$

where i= 1,..., 50, ϵ_i ~ N(0,4) and the parameter function $\beta_1(t)$ has a sine-wave function shape as shown in Figure 1.
We set up the model where the response is related to the $X_1(t)$ and not on $X_2(t)$.

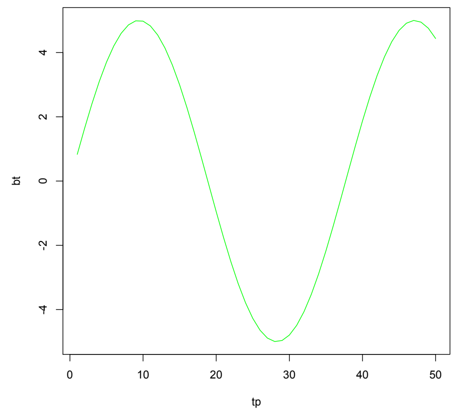

Figure 1. $\beta_1(t)$

Contamination of $X_{ij}(t)$
We consider contaminating only $X_{i1}(t)$ to produce functional outliers while $X_{i2}(t)$ is not contaminated. The contamination process is carried out as described by Fraiman & Muniz [5]. The following five cases of contamination are considered:

- Case (1): No contamination $X_{i1}(t)$ are generated as described previously.
- Case (2): Asymmetric contamination $Z_{ij}(t) = X_{i1}(t) + cM$ where c is 1 with probability q and 0 with probability 1- q and $q = \{0\%; 5\%; 10\%; 15\%; 20\%\}$; M is the contamination constant size equal to 10 and $X_{ij}(t)$ is as defined in Case (1).
- Case (3): Symmetric contamination $Z_{ij}(t) = X_{ij}(t) + c\sigma M$ where $X_{ij}(t)$, c and M are as defined before and σ is a sequence of random variables independent of c that takes the values 1 and - 1 with probability 0.5.
- Case (4): Partial contamination $Z_{ij}(t) = X_{ij}(t) + c\sigma M$ if t >T and $Z_{ij}(t) = X_{ij}(t)$ if t <T, where T∼ U [0,10].
- Case (5): Peak contamination $Z_{ij}(t) = X_{ij}(t) + c\sigma M$ if $T \leq T \leq +l$ and $Z_{ij}(t) = X_{ij}(t)$ if $t \notin [T, T + l]$ where$l = 2$ and $T \sim U[0, 10 - l]$.

The effects of these different types of contamination are shown in Figure 2.

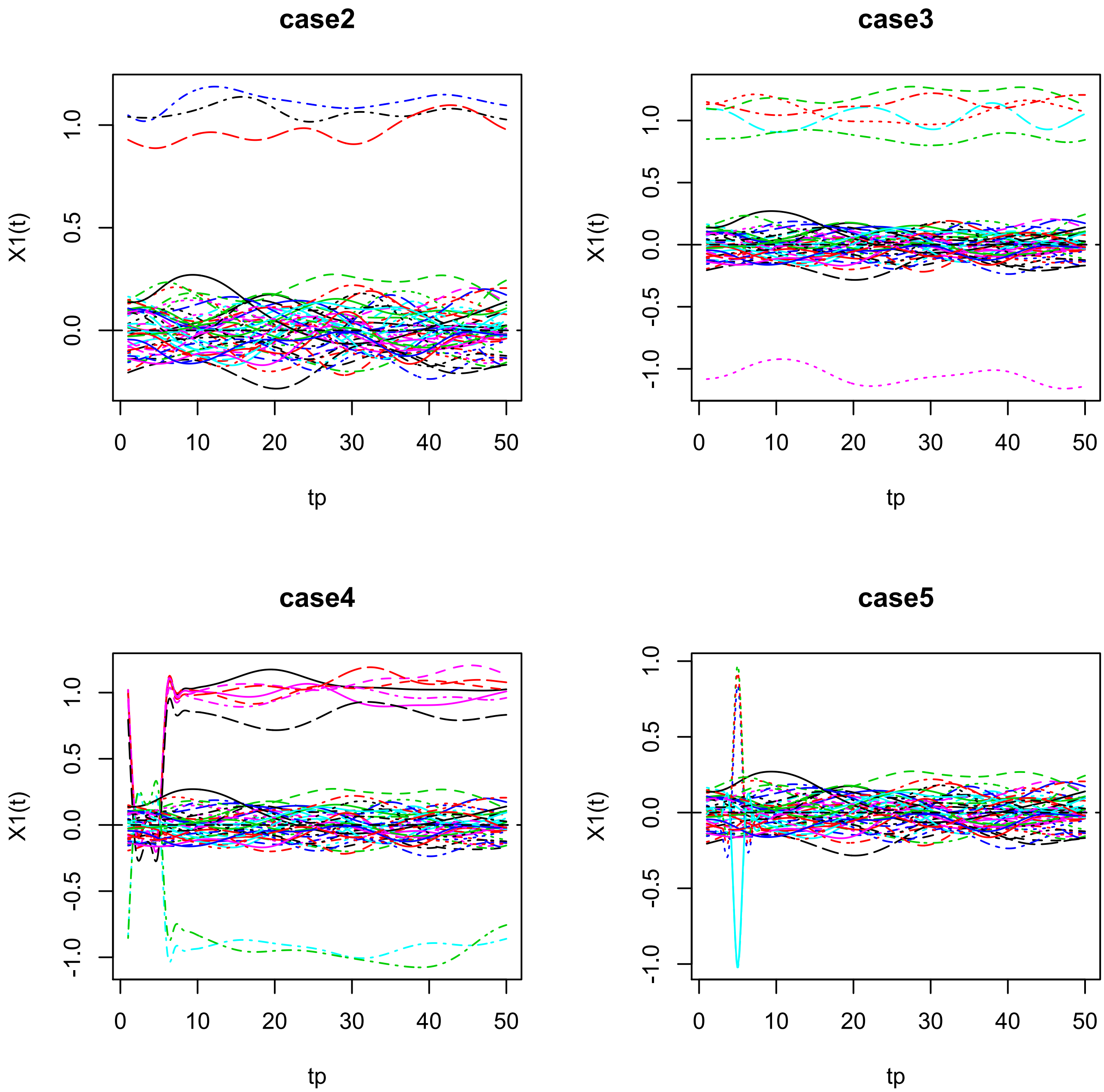

Figure 2. The contaminated $X_{i1}(t)$ curves for cases 2-5 (q = 15%)

Contamination of Y_i

We consider contaminating Y at 15% contamination level by generating the errors from the standard normal distribution, the t-distribution with 2 degrees of freedom, and the t-distribution with 7 degrees of freedom. This will allow for heavy-tail error distributions and some outliers in the response direction.

Toy Example

Case(i): Presence of outliers in the scalar response Y only.

Figure 3 below shows the comparison of the classical functional group LASSO with new proposed method functional LAD-group LASSO which is robust in nature in the presence of outliers in the scalar response Y. The green curve is the true function $\beta_1(t)$. The red line in the plots represents the estimation done by classical functional LASSO and the blue line represents the estimation done by robust functional LAD-group LASSO at fixed combination of (λ= 0.5, φ=100). We can see from the Figure 3 that in the presence of outliers in scalar response, the classical method does poor estimation and shrinkage, whereas new robust method does good both in estimation and shrinkage.

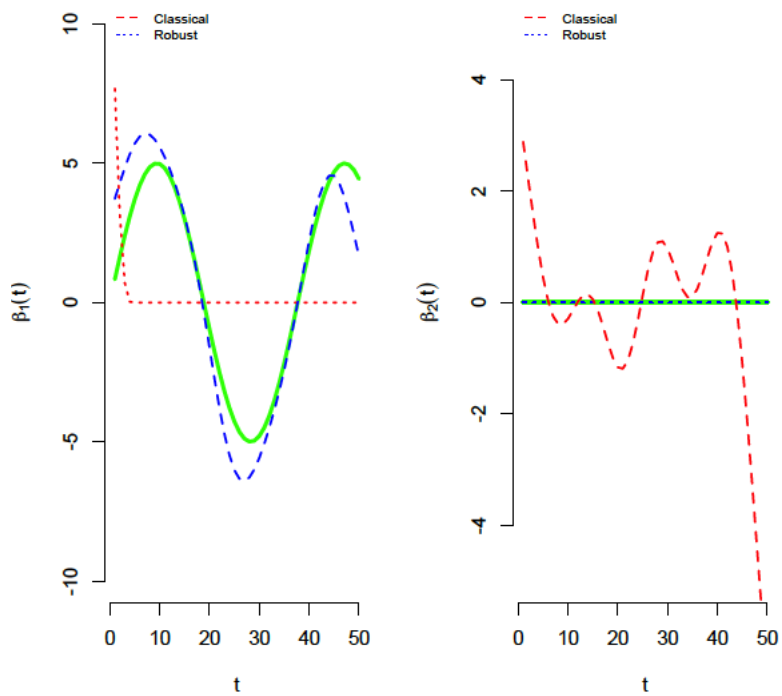

Figure 3. Fitting results for the comparison classical of functional groupLasso and robust functional LAD-groupLASSO for Case (i).

Case(iii): Presence of outliers in the functional predictors X(t) only.

Figure 4 below shows the comparison of the classical functional group LASSO with new proposed method functional LAD-group LASSO which is robust in nature in the presence of functional outliers. The green curve is the true function $\beta_1(t)$. The red line in the plots represents the estimation done by classical functional LASSO and the blue line represents the estimation done by robust functional LAD- group LASSO at fixed combination of (λ=0.4, φ=100). We can see from the Figure 4 that in the presence of outliers, classical method does poor estimation and shrinkage, whereas new robust method does good both in estimation and shrinkage. Due to the limitation on the length of the manuscript, we only give results based on the q=15% level.

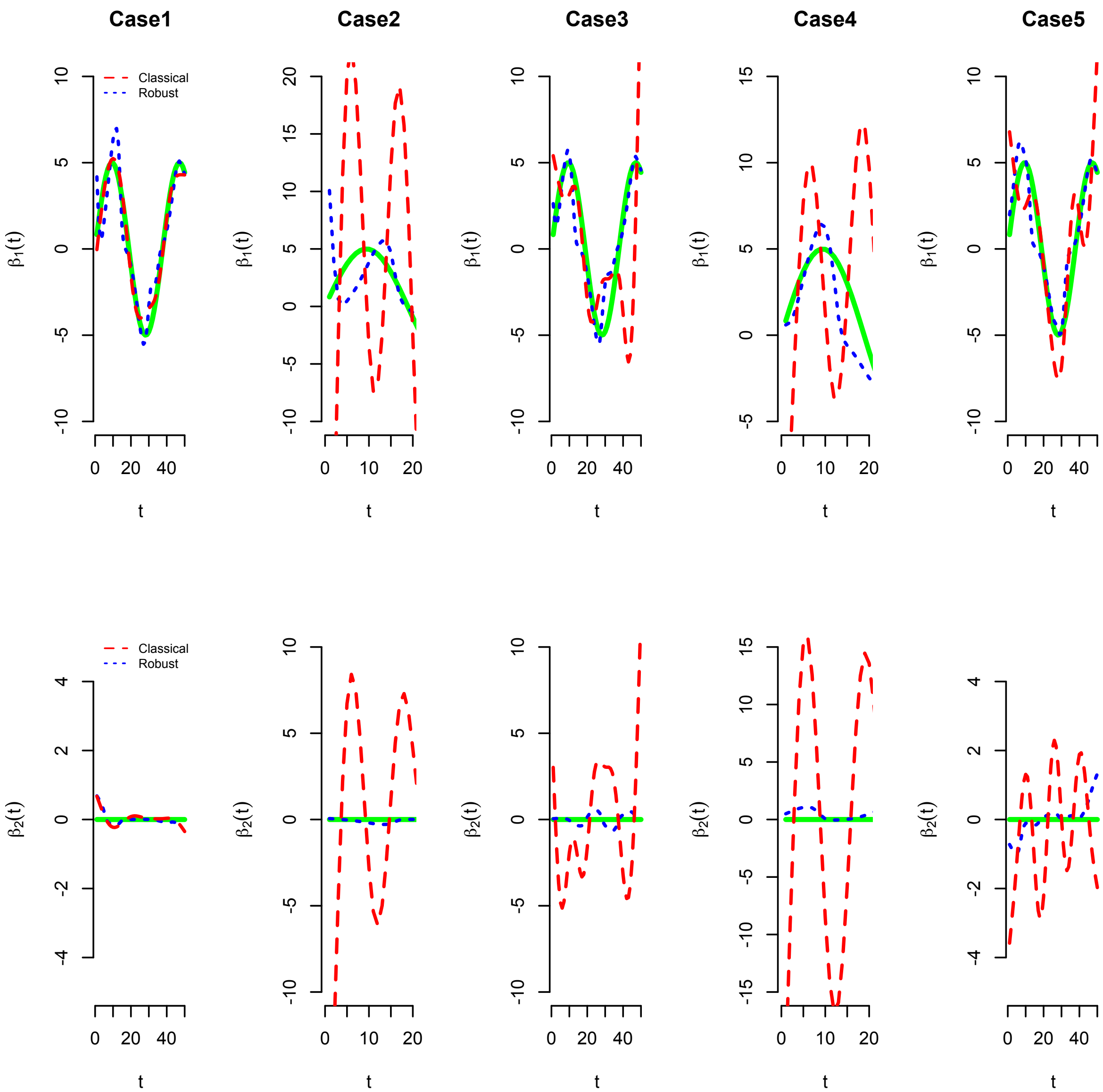

Figure 4. Fitting results for the comparison classical of functional groupLasso and robust functional LAD-groupLASSO for Case (iii)(15% contamination).

Simulation Study
We used the same technique as described above to generate the scalar responses y, the coefficient functions and the functional predictors. The functional predictors were contaminated the same way as described in five cases above. We consider:
1) 1000 observations for the scalar response.
2) Two functional Predictors are considered. We generated 1000 sample curves for each $X_j(t)$ which are observed at 1000 equidistant time points.
3) The true model is

$$Y_i = \alpha + \int_0^{1000} \beta_1(t) X_{ij}(t) dt + \epsilon_i.$$

where, $\epsilon_i \sim$ N(0,4) and the parameter function $\beta_1(t)$ is observed at 1000 equidistant points in (0, 1000) and has a sine-wave function shape as shown in Figure 1 above. So the $\beta_2(t)$ is essentially 0.

Then we consider the Mean squared Errors of prediction 1/n$\sum_i (Y_i - \hat{Y}_i)^2$ to assess the predictive ability of the proposed method. Figure 5 shows the boxplots of MSE from 50 simulation runs. We can see from this figure that robust functional LAD-group LASSO, which is represented by blue color, outperforms the classical functional group LASSO, which is represented by red color.

As pointed out earlier due to limited space, we could not present simulation results for Case (i) and Case (ii), but we get satisfactory results for these two cases as well.

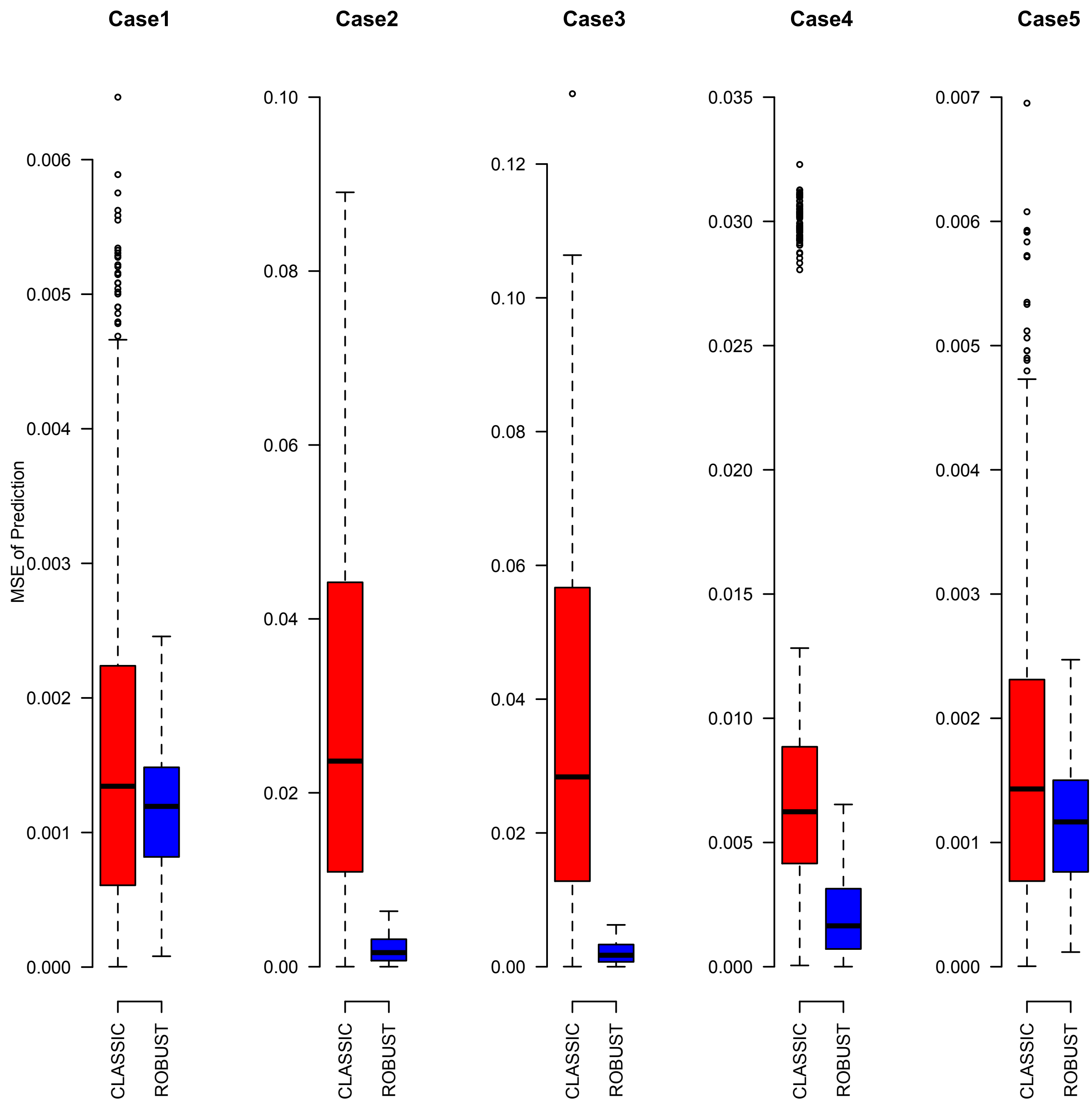

Figure 5. Comparison of MSE's of prediction of Functional groupLasso and robust functional LAD-groupLASSO (15%contamination).

4 Summary and Discussion

We considered a robust variable selection procedure for functional linear regression models in the presence of functional outliers, where various functional predictors are considered but only a few of these predictors are actually related to the response. Typical variable selection procedures for functional models do not consider the issue of outliers while selecting the useful predictors, and thus may suffer from wrong models. Our proposed procedure simultaneously selects functional variables and estimates the important regression coefficients functions.
We found that our proposed method performs well in terms mean squared errors of prediction for the estimated coefficient functions compared to classically fitting a model without taking outliers into consideration. Although we used contamination levels varying from 15% to 40% , but reported the results based on only 15% due to the limitation of space, we are happy to report that our method performed well at all contamination levels as well.
We believe that the proposed method may be an efficient solution for analyzing functional data in the presence of outliers in scalar response and functional predictors.

REFERENCES

[1] Bali, JL, Boente, G, Tyler, DE & Wang, JL, 'Robust functional principal components: A projection-pursuit approach,' *The Annals of Statistics*, **39**: 2852-2882, 2011.

[2] Boente, G & Fraiman, 'Discussion of robust principal components for functional data,' *Test*, **8**: 28-35, 1999.

[3] Denhere, M & Billor, N, 'Robust principal component functional logistic regression,' *Communications in Statistics - Simulation and Computation*, 2013

[4] Escabias, M, Aguilera, AM & Valderrama, MJ, 'Principal component estimation of functional logistic regression: Discussion of two different approaches,' *Journal of Nonparametric Statistics*, **16(3-4)**: 365-384, 2004.

[5] Fraiman, R & Muniz, G, 'Trimmed means for functional data,' *Test*, **10**: 419-440, 2001.

[6] Gertheiss J, Maity, A & Staicu, A-M, 'Variable Selection in Generalized Functional Linear Models,' *Stat*, **2**: 86-101, 2013.

[7] Gervini, D, 'Robust functional estimation using the median and spherical principal components,' *Biometrika*, **95**:587- 600, 2008.

[8] Goldsmith, J, Bobb, J, Crainiceanu, CM, Caffo, B & Reich, D, 'Penalized functional regression,' *Journal of Computational and Graphical Statistics*, **20**: 830-851, 2011

[9] Lian, H, Shrinkage estimation and selection for multiple functional regression,' *Statistica Sinica*, **23**: 51-74, 2013.

[10] Lilly, K & Billor, N., 'A robust variable selection method for grouped data', *JSM Proceedings, Section on Statistical Learning and Data Mining. Alexandria, VA: American Statistical Association*, 3334-3341, 2013.

[11] Matsui, H & Konishi, S, 'Variable Selection for Functional Regression Models via the L1 Regularization,' *Computational Statistics and Data Analysis*, **55**:3304-3310, 2011.

[12] Muller, HG & StadtMuller, U, 'Generalized functional linear models,' *The Annals of Statistics*, **33(2)**: 774-805, 2005.

[13] Ogden, RT & Reiss, PT, 'Functional generalized linear models with images as predictors,' *Biometrics*, **66**: 61-69, 2010.

[14] Sawant, P, Billor, N & Shin, H, 'Functional outlier detection with robust functional principal component analysis,' *Computational Statistics*, **27(1)**: 83-102, 2012.

[15] Tutz, G & Gertheiss, J, 'Feature extraction in signal regression: A boosting technique for functional data regression,' *Journal of Computational and Graphical Statistics*, **19**: 154-174, 2010.

[16] Zhu, H & Cox, DD, 'A functional generalized linear model with curve selection in cervical pre-cancer diagnosis using fluorescence spectroscopy,' *IMS Lecture Notes, Monograph Series - Optimality: The Third Erich L. Lehmann Symposium*, **57**: 173-189, 2009.