Taylor & Francis
Taylor & Francis Group

Communications in Statistics - Simulation and Computation

ISSN: 0361-0918 (Print) 1532-4141 (Online) Journal homepage: http://www.tandfonline.com/loi/lssp20

Robust Group-lasso for Functional Regression Model

Jasdeep Pannu & Nedret Billor

To cite this article: Jasdeep Pannu & Nedret Billor (2015): Robust Group-lasso for Functional Regression Model, Communications in Statistics - Simulation and Computation, DOI: 10.1080/03610918.2015.1096375

To link to this article: http://dx.doi.org/10.1080/03610918.2015.1096375

Accepted author version posted online: 23 Oct 2015.

Robust Group-Lasso for Functional Regression Model

Jasdeep Pannu *and Nedret Billor

Department of Mathematics & Statistics
Auburn University, 221 Parker Hall
Auburn AL 36849 USA

Abstract

In this paper, 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, Least Absolute Deviation (*LAD*) 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 predictors based on *LASSO* fails since multiple parameters exist for a functional predictor. Therefore *group LASSO* is used for selecting functional predictors since *group LASSO* selects grouped variables rather than individual variables. In this study, we propose a robust functional predictor selection method, the *LAD-group LASSO*, for a functional linear regression model with a scalar response and functional predictors. We illustrate the performance of the *LAD-group LASSO* on both simulated and real data.

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

*Electronic address: jkp0008@auburn.edu; Corresponding author

1 INTRODUCTION

Functional data analysis has become increasingly frequent and important in diverse fields of sciences, engineering, and humanities. Imperative data pertaining to these fields 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. (2004), Denhere & Billor (2014), Boente & Fraiman (1999), Gervini (2008), Bali et al. (2011), Sawant et al. (2012), Goldsmith et al. (2011) and Ogden & Reiss (2010) 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 accuracy and therefore, enhances the model interpretability. However, in the presence of outliers, that are curves deviating from the remaining of functional data, the effective and correct selection of significant functional covariates become even more challenging.

Not much work has been done in the area of variable selection for functional predictors in functional regression models. Gertheiss et al. (2013), Matsui & Konishi (2011), Lian (2013) and Zhu & Cox (2009) proposed some variable selection techniques for functional predictors via *L1* and *L2* regularizations, for instance, using various roughness penalties like *group LASSO*, Wavelet based-*LASSO*, *group SCAD* for the generalized functional linear models. However, these methods do not work well in the presence of outliers. Since these variable selection techniques are all based on the estimation of the coefficient functions which the estimates are obtained by minimizing the penalized residual sum of squares, it is known to be non-robust in nature. Thus, there is a need for a robust variable selection method which is resistant to outliers.

Lilly and Billor (2013) 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. (2013), that minimizes the effect of outliers in the estimation and selection of the functional covariates in functional linear models. This paper is different from others because we propose a new robust functional variable selection technique for functional covariates in functional linear models. In this study, we consider the problem of selecting functional predictors using the *L1* regularization in a functional linear regression model with a scalar response and functional predictors in the presence of outliers. The first step that we take in this paper is to reformulate the functional linear model as a multiple linear one by approximating the functional covariates as a linear combination of an appropriate basis as discussed in Ramsay & Silverman (2005). Then we apply the *LAD-group LASSO* procedure for selection of grouped variables which each functional predictor is assumed to have grouped parameters.

This article is organized as follows. In sections 2 and 3, we provide the methodology which contains the formulation of the functional linear regression, penalty settings and selection and, functional *LAD-group LASSO* methodology. In order to show the goodness of the proposed methodology, numerical study consisting of toy example and simulation, is presented in section 4. In section 5, a real data application of the proposed method is provided. Finally, summary and conclusion are given in section 6.

2 REFORMULATION OF FUNCTIONAL REGRESSION MODEL

In order to estimate the parameter functions based on multivariate variable selection idea we follow two steps. The first step is to formulate the model given in (1) in a usual multiple regression model form to overcome infinite dimensionality issue which is inherent with functional data. The second step is to apply a robust variable selection method based on robust version of *group LASSO* that would select the influential functional predictors on the response.

In this section we will first give a description for a functional regression model with a scalar response and functional predictors and present a method to reformulate this model in ordinary multiple regression model form.

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 ith subject and $X_1, X_2, \ldots, X_p$ are the squared integrable random curves, $X_j : \mathcal{T}_I \subset \Re \to \Re$ and $X_{i1}, X_{i2}, \ldots, X_{ip}$ denote their independent realizations, respectively.

We also assume that the mean function of the underlying trajectories, X_j is equal to zero. For the sake of simplicity, each X_{ij} is considered to be observed without measurement error at a dense grid of time points $\{t_{j1}, t_{j2}, \ldots, t_{jN_j}\}$. Then a functional linear model with the scalar response and p functional predictors can be defined as:

$$Y_i = \alpha + \Sigma_{j=1}^{p} \int_{\mathcal{T}_I} X_{ij}(t)\beta_j(t)dt + \epsilon_i, \quad i = 1, \ldots, N. \tag{1}$$

The main object of our interest in this model is the regression coefficient functions which are assumed to be smooth and squared integrable. The random error terms ϵ_i are assumed to be independent normally distributed with mean 0 and variance

ACCEPTED MANUSCRIPT

σ^2. α is a scalar parameter and $\beta_j(t)$ is a parameter function for $j = i, \ldots, p$.

To overcome infinite dimensionality problem, we use basis approximation method. This requires the use of pre-set basis functions expansion for approximation of the parameter functions, $\beta_j(t)$ as well as for approximation of the functional predictors, $X_{ij}(.)$. The choices of basis functions are associated with characteristics of the parameter functions and functional predictors and they do not have to be the same basis functions. Then the integral in (1) can be approximated by Riemann sum as

$$\int X_{ij}(t)\beta_j(t)dt \approx \Sigma_m X_{ij}(t_m)\beta_j(t_m), \tag{2}$$

where

$$\beta_j(t) = \Sigma_{b=1}^{l} c_{jb}\phi_{jb}(t). \tag{3}$$

Here $\mathbf{\Phi_j}(t) = (\Phi_{j1}(t), \ldots, \Phi_{jl}(t))$ is a finite basis and c_{jb} are the corresponding basis coefficients.

Using (2) and (3), the integral on 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 $i = 1, \ldots, N$, $j = 1, \ldots, p$, $\delta_j = t_{jm} - t_{j,m-1}$, $\mathbf{c_j} = (c_{j1}, \ldots, c_{jl})^T$, $\mathbf{\Phi_{ij}} = (\Phi_{ij1}, \ldots, \Phi_{ijl})^T$ and $\Phi_{ijb} = \delta_j\Sigma_m X_{ij}(t_{jm})\phi_{jb}(t_{jm})$.

The new model in the usual multiple regression form is then obtained as

$$Y_i = \alpha + \Sigma_{j=1}^{p}\mathbf{\Phi_{ij}^T c_j} + \epsilon_i, \quad i = 1, \ldots, N \tag{5}$$

ACCEPTED MANUSCRIPT

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

3 Functional *LAD-group LASSO* Method

The second part of the problem is to consider ways to produce robust estimator through penalized estimation technique, such as *group LASSO* of the parameter functions in the presence of outliers.

For the simultaneous estimation of the parameter functions and sparseness of the solution, Gertheiss et al. (2013) proposed a sparsity-smoothness penalty technique which is based on the penalty function proposed Meier et al. (2009) given as

$$\Sigma_{i=1}^{n}(Y_i - \alpha - \Sigma_{j=1}^{p}\mathbf{\Phi_{ij}}^T\mathbf{c_j})^2 + P_{\lambda,\varphi}(\beta_j), \tag{6}$$

where $P_{\lambda,\varphi}(\beta_j)$ is the penalty function defined by Meier et al. (2009). However, this method discussed by Gertheiss et al. (2013) along with other classical existing functional variable selection methods like *group SCAD*, simple *group LASSO* proposed by Lian (2013) and Zhu & Cox (2009), respectively, is merely a special case of penalized least squares method and thus suffers from the presence of outliers, therefore necessitating a different type of approach to handle this issue. We consider a new criterion called functional *LAD-group LASSO* to take into account the effect of outliers.

According to this criterion, α and $c_j(t)$ can be estimated by minimizing the follow-

ACCEPTED MANUSCRIPT

ing:

$$\Sigma_{i=1}^{n}|Y_i - \alpha - \Sigma_{j=1}^{p}\mathbf{\Phi_{ij}}^T\mathbf{c_j}| + P_{\lambda,\varphi}(\beta_j), \tag{7}$$

where $P_{\lambda,\varphi}(\beta_j)$ is the penalty function as introduced by Meier et al. (2009) and used by Gertheiss et al. (2013) for functional variable selection. Specifically,

$$P_{\lambda,\varphi}(\beta_j) = \lambda(||\beta_j||^2 + \varphi||\beta_j^{''}||^2)^{1/2}, \tag{8}$$

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

Here λ is the parameter that controls sparseness and φ is the smoothing parameter that controls smoothness of the coefficients. As the sparsity parameter λ increases, the estimated coefficient functions $\beta(t)'$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, in which the prediction error of the model is minimized. The most commonly used values of K are 5 and 10. Then we redefine the penalty function $P_{\lambda,\varphi}(\beta_j)$ in (8), as proposed by Gertheiss et al. (2013),

$$P_{\lambda,\varphi}(\beta_j) = \lambda(\mathbf{c_j}^T(C_{\varphi,j})\mathbf{c_j})^{1/2}, \tag{9}$$

where $C_{\varphi,j} = \Psi_j + \varphi\Omega_j$ is a $l \times l$ symmetric and positive definite matrix, $\Psi_j$is a $l \times l$ matrix whose $(b, k)th$ element is $\int \phi_{jb}(t)\phi_{jk}(t)dt$ and Ω_j is a $l \times l$ matrix whose $(b, k)th$ element is $\int \phi''_{jb}(t)\phi''_{jk}(t)dt$ for $b, k = 1, \ldots, l$.

Further $C_{\varphi,j}$ can be decomposed using Cholesky decomposition as following:

$$C_{\varphi,j} = L_{\varphi,j}L^T_{\varphi,j}, \tag{10}$$

where $L_{\varphi,j}$ is non-singular lower triangular matrix. Now using (9) and (10) equation (7) reduces to the following:

$$\Sigma^n_{i=1}|Y_i - \alpha - \tilde{\Phi}^T_{ij}\tilde{\mathbf{c}}_\mathbf{j}| + \lambda\Sigma^p_{j=1}||\tilde{\mathbf{c}}_\mathbf{j}||, \tag{11}$$

where $\tilde{\mathbf{c}}_\mathbf{j} = L^T_{\varphi,j}\mathbf{c}_\mathbf{j}$ and $\tilde{\Phi}_j = L^{-1}_{\varphi,j}\Phi_j$. Now $\hat{\alpha}$ and $\hat{\mathbf{c}}_\mathbf{j}$'s are the minimizers of (11) and the coefficient function $\beta(t)$ is estimated estimated by $\hat{\beta}_j(t) = \Sigma^l_{b=1}\phi_{jb}(t)\hat{c}_{jb}$ for $j = 1, \ldots, p$.

The *LAD-group LASSO* estimator can be computed easily. Basically, the *LAD-group LASSO* criterion combines the *LAD* criterion and the *group LASSO* penalty. Specifically, we can consider an augmented dataset $(Y^*_i, \mathbf{X}^*_\mathbf{i})$ with $i = 1, \ldots, N + p$, where $(Y^*_i, \mathbf{X}^*_\mathbf{i}) = (Y_i, \mathbf{X}_\mathbf{i})$ for $1 \le i \le N$, $(Y^*_{n+j}, \mathbf{X}^*_\mathbf{n+j}) = (\mathbf{0}, \lambda\mathbf{e}_\mathbf{j})$ for $1 \le j \le p$, and $\mathbf{e}_j$ is a $p-$dimensional vector with the jth component equal to 1 and all others equal to 0. It can be easily verified that the objective function in equation (11) becomes

$$\Sigma_{i=1}^{n}|Y_i^* - \alpha - \tilde{\Phi}_{ij}^{*T}\tilde{\mathbf{c}}_{\mathbf{j}}|. \tag{12}$$

This becomes a penalized *LAD* problem for which there are many available algorithms. Specifically we use the one based on penalized quantile regression. It fits a quantile regression model with the *LASSO* penalty. The algorithm is similar to *LASSO* code presented in Koenker & Mizera (2014) and specifically uses the function *rq.fit.lasso ()* function available in the R package *quantreg* (Koenker (2013)).

4 NUMERICAL STUDY

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

- **Model (0)**: Presence of no outliers in the scalar response Y and the functional predictors $X(t)$.
- **Model (1)**: Presence of outliers in the scalar response Y only.
- **Model (2)**: Presence of outliers both in the scalar response Y and the functional predictors $X(t)$.

We take following steps to carry out the numerical study:

A. Generating functional data:

Generating Functional Predictors $X(t)$: We consider two functional covariates $X_1(t)$ and $X_2(t)$ which are generated similarly as in Tutz & Gertheiss (2010) from:

$$X_{ij}(t) = [\sigma(t)]^{-1}\Sigma_{r=1}^{5}(a_{ijr} sin(\pi t(5 - a_{ijr})/150) - m_{ijr}), \tag{13}$$

where $i = 1, \ldots, 50,\ j = 1, 2,\ a_{ijr} \sim U(0, 5),\ m_{ijr} \sim U(0, 2 * \pi)$ and $\sigma(t)$ is defined so that $var[X_{ij}(t)] = 0.01$.

Generating Y: Response Y is generated for 50 functional curves from:

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

where $i = 1, \ldots, 50$, $\epsilon_i \sim N(0, 4)$ 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 to $X_2(t)$.

B. Contamination of data

Contamination of Y: In order to create outliers in response Y, the errors ϵ are generated from the standard normal distribution, the t-distribution with 2 degrees of freedom, and the t-distribution with 7 degrees of freedom. The contamination is done for 15% of Y.

Contamination of $X(t)$: We consider contaminating both $X_1(t)$ and $X_2(t)$ at 15% level to produce functional outliers. The contamination process is carried out as described by Fraiman & Muniz (2001). The following three cases of contamination are considered:

- **Case (1)**: Asymmetric contamination $Z_j(t) = X_j(t) + cM$ 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 $X_j(t)$ is as defined in (12).

- **Case (2)**: Symmetric contamination $Z_j(t) = X_j(t) + c\sigma M$ where $X_j(t)$, 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 (3)**: Partial contamination $Z_j(t) = X_j(t)+c\sigma M$ if $t > T$ and $Z_j(t) = X_j(t)$ if $t < T$, where $T \sim U[0, 10]$.

The effects of these different types of contamination on $X_1(t)$ at 15 % level are shown in Figure 2.

4.1 Toy Example

Model (0): Presence of no outliers in the scalar response Y and the functional predictors $X(t)$.

First we apply our proposed method to Model (0). Model (0) has neither outliers in scalar response Y nor in the functional predictors $X_1(t)$ and $X_2(t)$. The response Y is dependent only on the first predictor $X_1(t)$. Figure 3 shows the fitting results of the classical functional *group LASSO* and the new robust functional *LAD-group LASSO* method. We used *rq.fit.lasso ()* function from the R package *quantreg* (Koenker (2013)) to implement our proposed method and the R package *grplasso* (Meier (2013)) for the classical *group LASSO*. In Figure 3, the green solid curves display the true functions $\beta_1(t)$ and $\beta_2(t)$, the red and blue dashed lines display the estimations done by classical functional *group LASSO* and robust functional *LAD-group LASSO*, respectively. The combinations of λ and φ for the robust functional *LAD-group LASSO* and the classical functional *group LASSO* are ($\lambda = 10$, $\varphi = 10$) and ($\lambda = 15$, $\varphi = 10^3$), respectively. We can see in Figure 3 that both methods perform equally well in terms of estimation of relevant predictor $X_1(t)$ and elimination of irrelevant predictor $X_2(t)$ in the model.

Model (1): Presence of outliers in the scalar response Y only.

Secondly, we apply our proposed method to the Model (1). Model (1) has outliers only in scalar response Y and the functional predictors $X_1(t)$ and $X_2(t)$ are free of outliers. Also the response Y depends only the first predictor $X_1(t)$ and not on $X_2(t)$. Since $X_2(t)$ is irrelevant to the true model, so it should be excluded from the model by the applied method. Figure 4 shows the comparison of the classical functional *group LASSO* with the new proposed method functional *LAD-group LASSO* which is robust in nature. R package *quantreg* (Koenker (2013)) was employed again to execute our proposed method. In Figure 4, the green solid curves display the true functions $\beta_1(t)$ and $\beta_2(t)$, the red and blue dashed lines display the estimations done by classical functional *group LASSO* (λ = 5, φ = 10) and robust functional *LAD-group LASSO* (λ = 10, φ = 10^3), respectively. We can see in Figure 4 that our robust proposed method is not only able to exclude the irrelevant predictor $X_2(t)$ from the model, but also is able to provide good estimation of relevant predictor $X_1(t)$. Whereas the classical method does poor estimation and shrinkage, in its comparison to the our proposed robust method.

Model (2): Presence of outliers both in the scalar response Y and the functional predictors $X(t)$.

Thirdly, we apply our proposed method to the Model (2). Model (2) has outliers both in scalar response Y and the functional predictors $X_1(t)$ and $X_2(t)$. All three cases of contamination are considered for the functional covariates and only the first covariate $X_1(t)$ is relevant to the true model. $X_2(t)$ is irrelevant and should be excluded from the model. Figure 5 shows the fitting results of the classical functional *group LASSO* and the new robust functional *LAD-group LASSO* method. Again R package *quantreg* (Koenker (2013)) is utilized to execute the proposed method. In Figure 5, the green solid curves are the true coefficient functions $\beta_1(t)$ and $\beta_2(t)$, the red dashed lines represent the estimation done by classical functional *group LASSO* and the blue lines represent the estimation done by robust functional *LAD-group LASSO*. Figure 5 shows that the robust functional *LAD-group LASSO* excludes the irrelevant predictor $X_2(t)$ from the estimated model, and estimates relevant predictor $X_1(t)$ close to its true value at fixed combinations of ($\lambda = 0.5$, $\varphi = 10^3$), ($\lambda = 1$, $\varphi = 10$) and ($\lambda = 1$, $\varphi = 10^3$) for three cases of contamination, respectively. The combinations of λ and φ for the classical functional *group LASSO* are ($\lambda = 0.5$, $\varphi = 10^2$), ($\lambda = 10$, $\varphi = 10^2$) and ($\lambda = 0.5$, $\varphi = 10$) for three cases of contamination, respectively. In contrast, the classical method performs poorly in both variable estimation and selection.

4.2 Simulation Study

To elucidate the performance of the proposed method, we conducted simulation study in a variety of settings. We used the same technique as described above to generate the scalar response Y and the functional predictors $X(t)$. The functional predictors were contaminated the same way as described in three cases above. We consider the following:

1) 300 observations for the scalar response.

2) Ten functional predictors are considered. We generated 300 sample curves for

each $X_j(t)$ which are observed at 300 equidistant time points.

3) The true model is

$$Y_i = \alpha + \sum_{j=1}^{5} \int_0^{300} \beta_j(t) X_{ij}(t) dt + \epsilon_i, \tag{14}$$

where $\epsilon_i \sim N(0, 4)$, and the parameter functions $\beta_j(t)$ are observed at 300 equidistant points in (0, 300). The shapes of $\beta_j(t)$ are as shown in Figure 6. We can see in Figure 6 that the $\beta_6(t)$-$\beta_{10}(t)$ are essentially 0. The true model (13) depends only on $\beta_1(t)$-$\beta_5(t)$.

In simulation study, we compare the performance of the proposed method with the classical functional *group LASSO* in terms of estimation and selection of variables for three different model scenarios Model (0), Model (1) and Model (2), as described in Toy example. Again the contamination is done for 15% in Model (1) and Model (2). We use five-fold cross validation to choose the tuning parameters in the simulation study, in which the prediction error minimized in terms of MSE is minimized. The the signal-to-noise ratio is 0.8

First we consider the squared errors (SE) to assess the performance of the proposed method. The Squared Error (SE) = $\int (\hat{\beta}_j(t) - \beta_j(t))^2 dt$, where $\hat{\beta}_j(t)$ and $\beta_j(t)$ are the estimated and true coefficient functions, respectively. Squared errors are observed in 50 independent simulation runs for Model (0), Model (1) and Model (2). Figures 6 and 7 show the boxplots of the squared errors for Model (0) and Model (1), respectively. Figures 8-10 show the boxplots for all three cases of contamination for Model (2). The blue and red boxplots in these figures correspond to the robust functional *LAD-group LASSO* and classical functional *group LASSO*, respectively. Then we consider the Mean Squared Errors (MSE) of prediction $\frac{1}{n}\sum_i (Y_i - \hat{Y}_i)^2$ and the Mean Absolute Error (MAE) $\frac{1}{n}\sum_i |Y_i - \hat{Y}_i|$ of prediction to assess the predictive ability of the proposed method. This time we generated a data set with 5000 observations. The MSE and the MAE of prediction are observed in 50 independent simulation runs for Model (0), Model (1) and Model (2). Figures 11 and 12 show the boxplots of the MSE and MAE of prediction for Model (0) and Model (1), respectively. Figures 13 and 14 show the boxplots for all three cases of contamination for Model (2).

We can see in Figures 6 and 11 that robust functional *LAD-group LASSO*, which is represented by blue color, performs equally well as the classical functional *group LASSO*, which is represented by red color for Model (0) scenario. Also for Model (1) and Model (2) scenarios, we see in Figures 7-10 and Figures 12-14 that robust functional *LAD-group LASSO* represented by blue color, outperforms the classical functional *group LASSO* represented by red color. We also observe in these figures that the proposed method performs better for Model (1) compared to Model (2). Both Squared Error and Mean Squared Errors of prediction for Model (1) are low in their contrast to the ones for Model (2).

Furthermore, we examined how many times each variable was selected in 50 independent simulation runs for Model (0), Model (1) and Model (2) using both robust functional *LAD-group LASSO* and classical functional *group LASSO*. Tables 1 and 2 show the proportion of simulation runs for which each predictor is selected using robust functional *LAD-group LASSO* and classical functional *group LASSO*, respectively. We see in Tables 1 and 2 that the true predictors $X_1(t)$-$X_5(t)$ were selected most frequently and predictors $X_6(t)$-$X_{10}(t)$ which are irrelevant to the true model are less frequently selected by the robust functional *LAD-group LASSO* compared to the classical functional *group LASSO*. Table 1 also shows that the percentage of false positives is low for Model (1) compared to Model (2), when robust functional *LAD-group LASSO* is used.

5 REAL DATA APPLICATION

We applied functional *LAD- groupLASSO* to the analysis of weather data used by Matsui & Konishi (2011), available in Chronological Scientific Tables 2005, selecting variables concerning weather information. We used weather data observed at 79 stations in Japan. The data set includes monthly and annual total observations averaged from 1971 to 2000: monthly observed average temperatures (TEMP), average atmospheric pressure (PRESSURE), time of daylight (DAYLIGHT), average humidity (HUMIDITY) and annual total precipitation. The aim of the anal-

ACCEPTED MANUSCRIPT

ysis is to select and estimate the variables that have a relationship with annual total precipitation. Since the data is collected over time for 12 months, it can be treated as functional data. Figure 15 shows weather data represented by functions observed at 12 points. In Figure 15, the group of curves shows presence of a few outliers i.e. trajectories that are in some way different from the rest in all the four predictor variables. Specifically Figure 16 shows that curves 78 and 79 in TEMP and PRESSURE and curves with shapes 1, 2 and 3 in HUMIDITY are the outlying curves, as detected by Sawant (2012) using robust functional principal component analysis. Figure 17 shows the outliers in scalar response annual total precipitation. The response (annual total precipitation) is continuous. We use the functional linear model (2) with our approach to determine the most useful variables. Figure 18 shows the estimated coefficient functions when using the standard penalty (11). Tuning parameters were chosen via five-fold cross-validation. The Humidity and Daylight are excluded from the model. The results indicate that there is no relationship between these variables and the precipitation. The remaining variables, the temperature and pressure, can be considered to relate to the precipitation. We also applied classical functional *group LASSO* to this data set. The results are shown in Figure 19. It is clear from this figure that classical functional *group LASSO* is not able to exclude any variable(s) from the model.

Furthermore, we generated 50 bootstrap samples from the weather data. For each bootstrap sample, functional regression modeling was performed; then we examined how many times each variable was selected. The results are shown in Table 3. The mean PRESSURE was selected most frequently among the four variables, followed by the TEMPERATURE. This reveals the relationships of these variables to the precipitation. On the other hand, the average HUMIDITY and the DAYLIGHT are less frequently selected. From the results, there seems to be less of a relationship between these variables and the precipitation.

6 SUMMARY AND DISCUSSION

We considered a robust variable selection procedure for functional linear regression models in the presence of outliers, where various functional predictors are considered but only a few of these are actually related to the scalar 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 and estimates the important functional variables.

We found that our proposed method performs well in terms of prediction error as well as mean squared errors for the estimated coefficient functions compared to classically fitting a model without taking outliers into consideration. Also we found that the false positive and false negative rates are quite low for our method.

Also in this paper we focused only on the LAD, which is one of the first and easiest robust methodologies available. However, there are many different types of loss functions which can be used in this framework. We hope to explore more of these and extend our presented method to quantile regression in the future. Furthermore, our future work will also focus on applying the *LAD-group LASSO* to functional regression models with outliers in functional responses and predictors and explore its theoretical properties such as oracle sparsity property, persistence property and consistency property. We believe that the proposed method may be an efficient solution for analyzing functional data in the presence of outliers.

REFERENCES

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

Boente, G. & Fraiman. (1999). Discussion of robust principal components for functional data. Test 8: 28- 35.

Denhere, M. & Billor, N. (2014). Robust principal component functional logistic regression, Communications in Statistics - Simulation and Computation.10.1080/03610918.2013.861628.

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

Fraiman, R. & Muniz, G. 2001. Trimmed means for functional data. Test.10: 419- 440.

Gertheiss J. Maity, A. & Staicu, A-M. (2013). Variable Selection in Generalized Functional Linear Models. Stat. 2: 86101.

Gervini, D. (2008). Robust functional estimation using the median and spherical principal components. Biometrika. 95: 587- 600.

Goldsmith, J. Bobb, J. Crainiceanu, CM. Caffo, B. & Reich, D. (2011). Penal-

ACCEPTED MANUSCRIPT

ized functional regression. Journal of Computational and Graphical Statistics. 20: 830- 851.

Koenker, R. (2013). quantreg: Quantile Regression. R package version 5.05. http://CRAN.R-project.org/package=quantreg.

Koenker, R. & Mizera, I. (2014). Convex optimization in R. Journal of Statistical Software, 60: 1-23.

Lian, H. (2013). Shrinkage estimation and selection for multiple functional regression. Statistica Sinica. 23: 51-74.

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

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

Meier, L. (2013). grplasso: Fitting user specified models with Group Lasso penalty. R package version 0.4-3. http://CRAN.R-project.org/package=grplasso.

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

Ogden, RT. & Reiss, PT. (2010). Functional generalized linear models with

ACCEPTED MANUSCRIPT

images as predictors. Biometrics. 66: 61- 69.

Ramsay, JO. & Silverman, BW. (2005). Functional Data Analysis. Springer.

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

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. Journal of the Royal Statistical Society B. 58: 267- 288.

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

Zhu, H. & Cox, DD. (2009). 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.

	$X_1(t)$	$X_2(t)$	$X_3(t)$	$X_4(t)$	$X_5(t)$	$X_6(t)$	$X_7(t)$	$X_8(t)$	$X_9(t)$	$X_{10}(t)$	Avg. model size
Model (0)	1	1	1	0.98	0.96	0.58	0.62	0.68	0.72	0.60	8.14
Model (1)	1	1	1	0.96	0.94	0.72	0.74	0.60	0.78	0.62	8.36
Model (2)	1	1	1	0.94	0.90	0.80	0.82	0.78	0.82	0.70	8.76

Table 1: Proportion of the functional predictors being selected with respect to simulation runs using robust functional *LAD-group LASSO*.

ACCEPTED MANUSCRIPT

	$X_1(t)$	$X_2(t)$	$X_3(t)$	$X_4(t)$	$X_5(t)$	$X_6(t)$	$X_7(t)$	$X_8(t)$	$X_9(t)$	$X_{10}(t)$	Avg. model size
Model (0)	1	1	1	0.92	0.94	0.70	0.74	0.70	0.62	0.66	8.28
Model (1)	1	1	1	0.90	0.94	0.98	0.82	0.80	0.86	0.82	9.12
Model (2)	1	1	1	0.84	0.86	0.96	0.94	0.98	0.92	0.90	9.40

Table 2: Proportion of the functional predictors being selected with respect to simulation runs using classical functional *group LASSO*.

ACCEPTED MANUSCRIPT

Variable	TEMP	PRESSURE	HUMIDITY	DAYLIGHT
Select	31	42	11	15

Table 3: Number of times each variable is selected.

ACCEPTED MANUSCRIPT

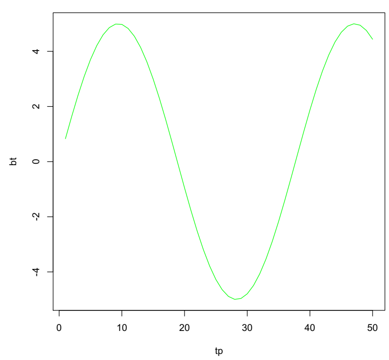

Figure 1: $\beta_1(t)$.

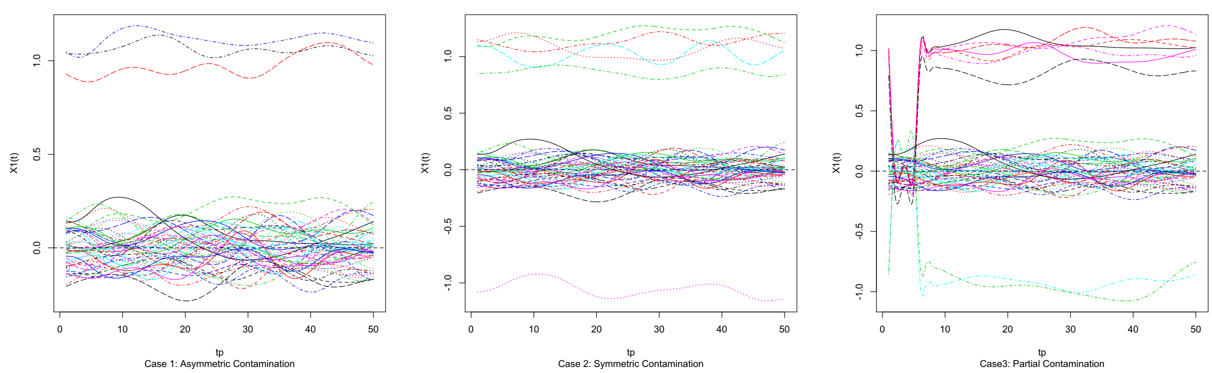

Figure 2: The contaminated $X_1(t)$ curves for contamination cases 1- 3 (q = 15%).

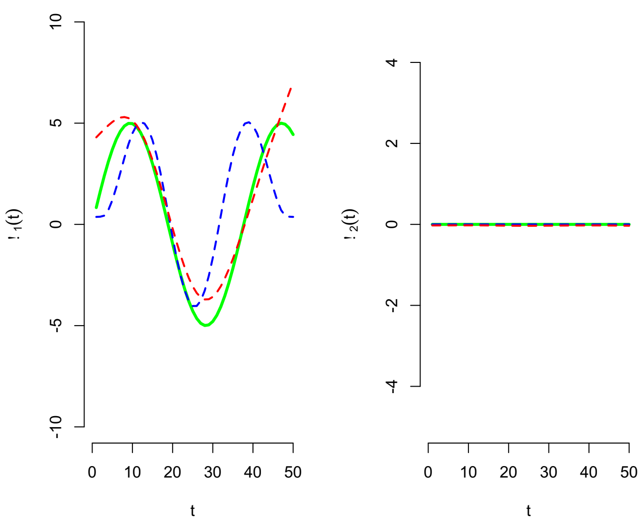

Figure 3: Fitting results for the comparison of classical functional *group LASSO* (red) and robust functional *LAD-group LASSO* (blue) for Model (0) (0% contamination).

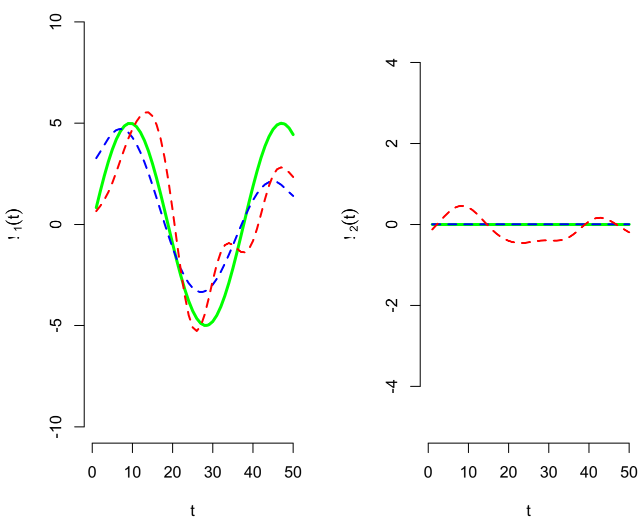

Figure 4: Fitting results for the comparison of classical functional *group LASSO* (red) and robust functional *LAD-group LASSO* (blue) for Model (1) (15% contamination).

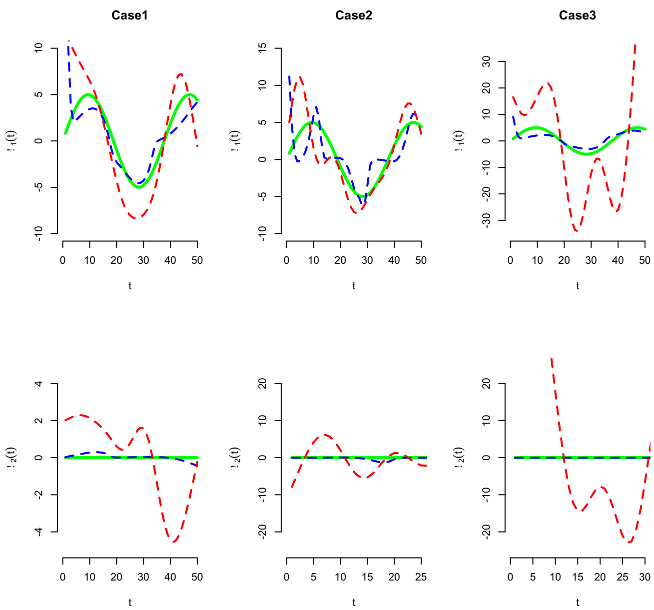

Figure 5: Fitting results for the comparison classical of functional *group LASSO* (red) and robust functional *LAD-group LASSO* (blue) for Model (2) (15% contamination).

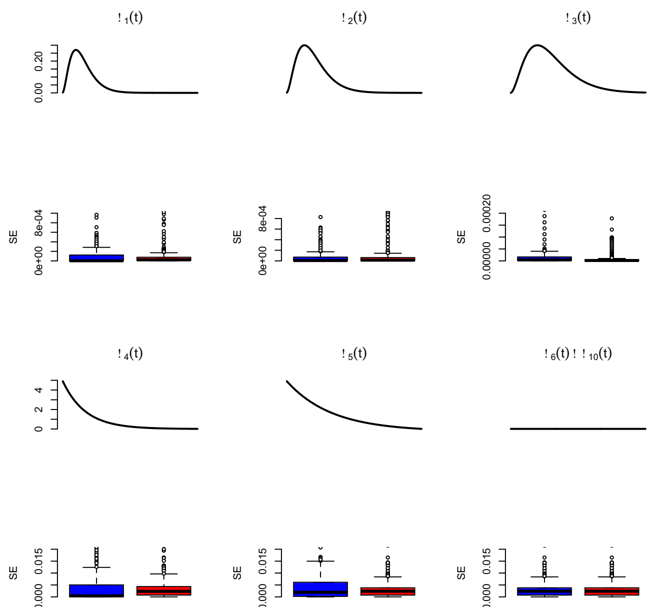

Figure 6: Comparison of SE of classical functional *group Lasso* (red) and robust functional *LAD-group LASSO* (blue) at 0% contamination for Model (0).

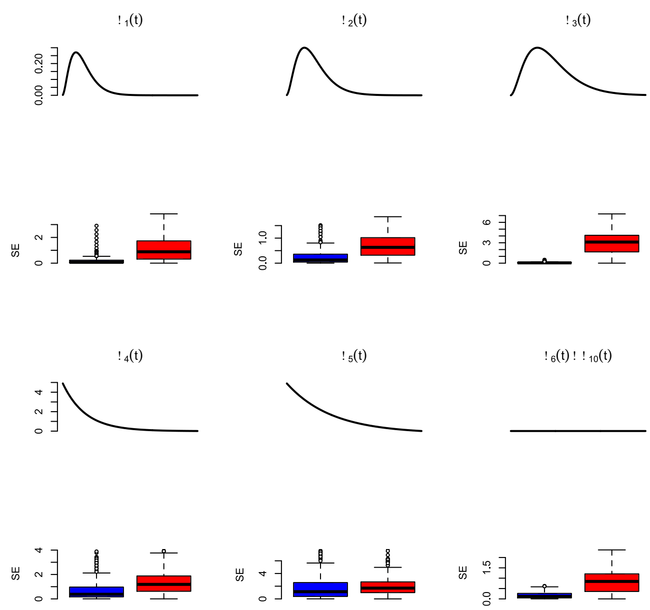

Figure 7: Comparison of SE of classical functional *group LASSO* (red) and robust functional *LAD-group LASSO* (blue) at 15% contamination for Model (1).

ACCEPTED MANUSCRIPT

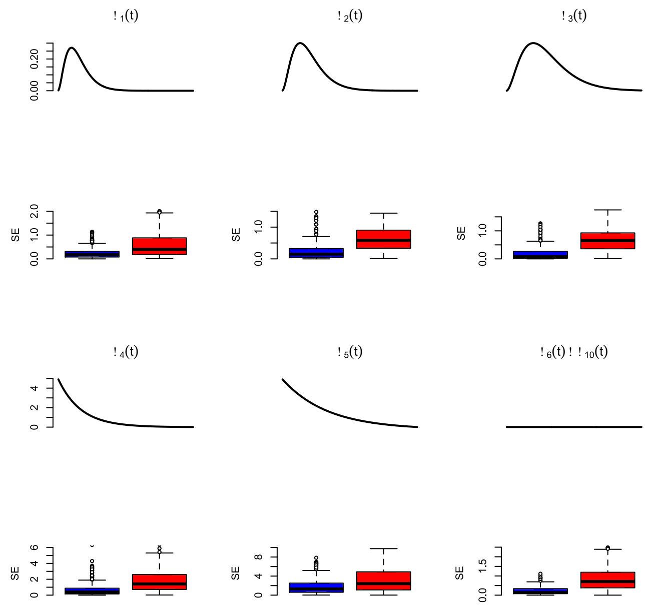

Figure 8: Comparison of SE of classical functional *group LASSO* (red) and robust functional *LAD-group LASSO* (blue) at 15% asymmetric contamination (Case 1) for Model (2).

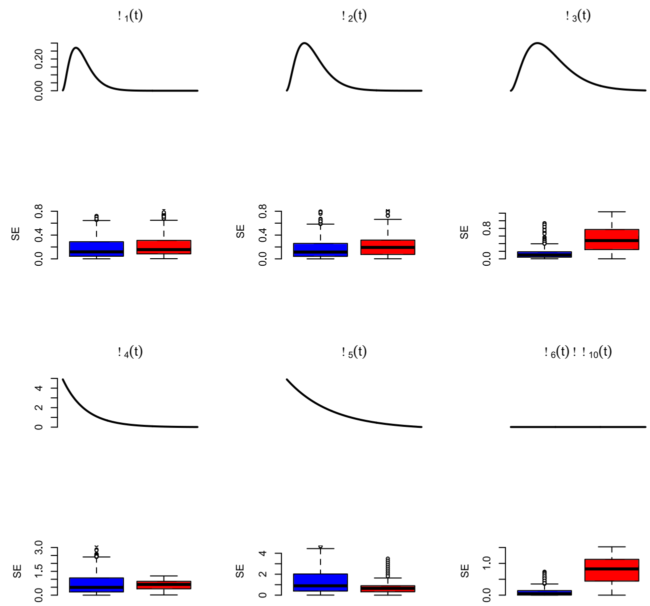

Figure 9: Comparison of SE of classical functional *group LASSO* (red) and robust functional *LAD-group LASSO* (blue) at 15% symmetric contamination (Case 2) for Model (2).

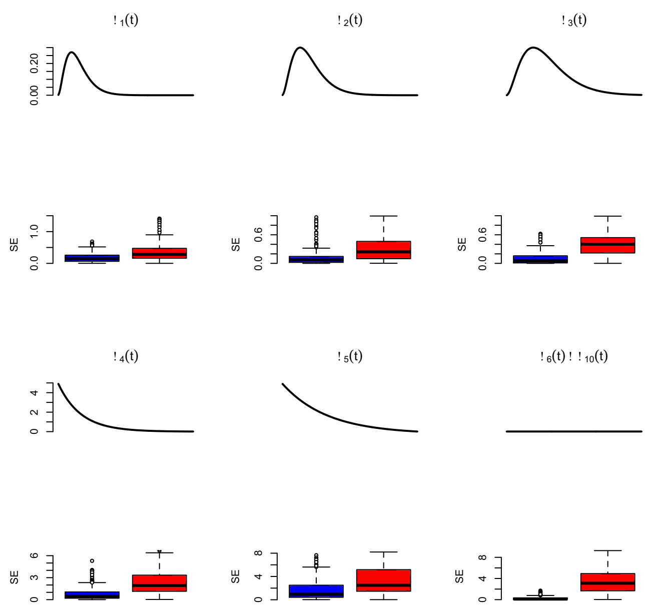

Figure 10: Comparison of SE of classical functional *group LASSO* (red) and robust functional *LAD-group LASSO* (blue) at 15% partial contamination (Case 3) for Model (2).

ACCEPTED MANUSCRIPT

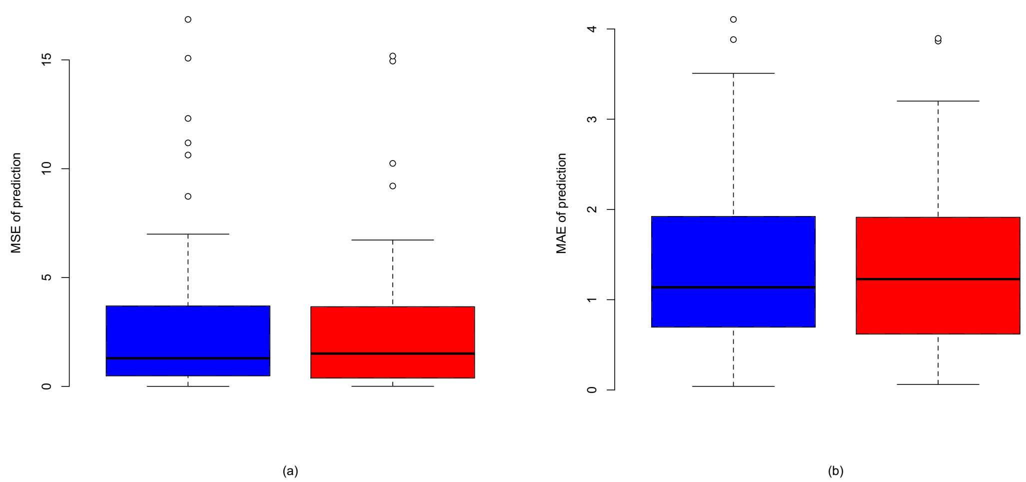

Figure 11: Comparison of MSE of prediction (Fig. (a)) and MAE of prediction (Fig. (b)) of classical functional *group LASSO* (red) and robust functional *LAD-group LASSO* (blue) at 0% contamination for Model (0).

ACCEPTED MANUSCRIPT

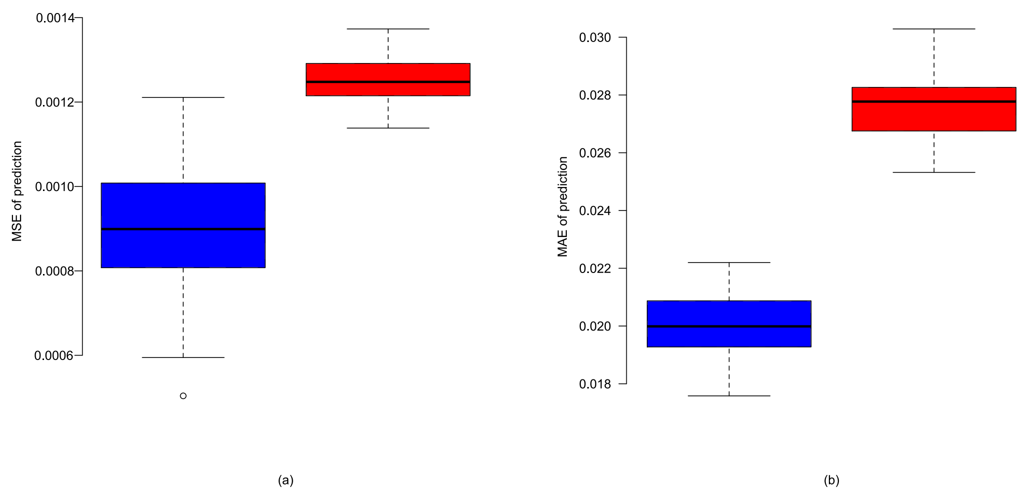

Figure 12: Comparison of MSE of prediction (Fig. (a)) and MAE of prediction (Fig. (b)) of classical functional *group LASSO* (red) and robust functional *LAD-group LASSO* (blue) at 15% contamination of Y for Model (1).

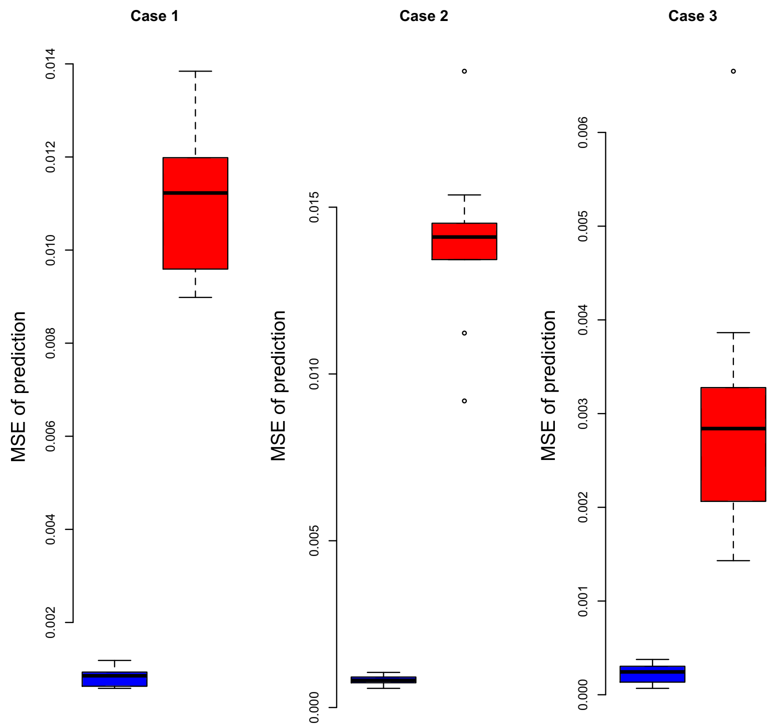

Figure 13: Comparison of MSE of prediction of classical functional *group LASSO* (red) and robust functional *LAD-group LASSO* (blue) at 15% contamination for Model (2).

ACCEPTED MANUSCRIPT

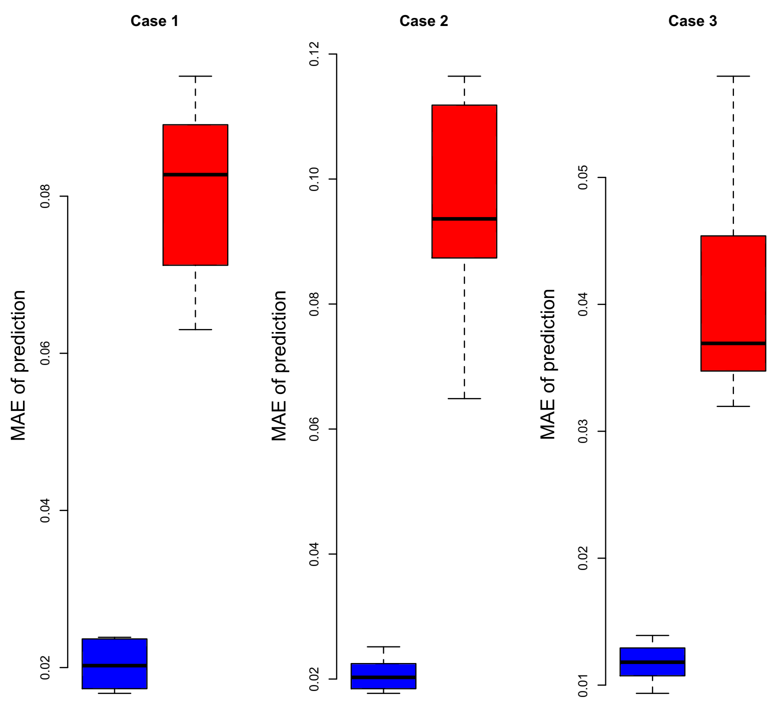

Figure 14: Comparison of MAE of prediction of classical functional *group LASSO* (red) and robust functional *LAD-group LASSO* (blue) at 15% contamination for Model (2).

ACCEPTED MANUSCRIPT

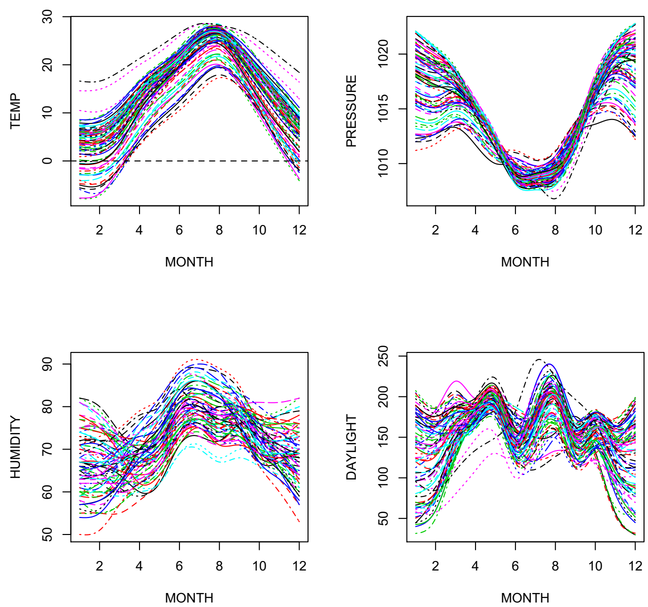

Figure 15: Weather Data.

ACCEPTED MANUSCRIPT

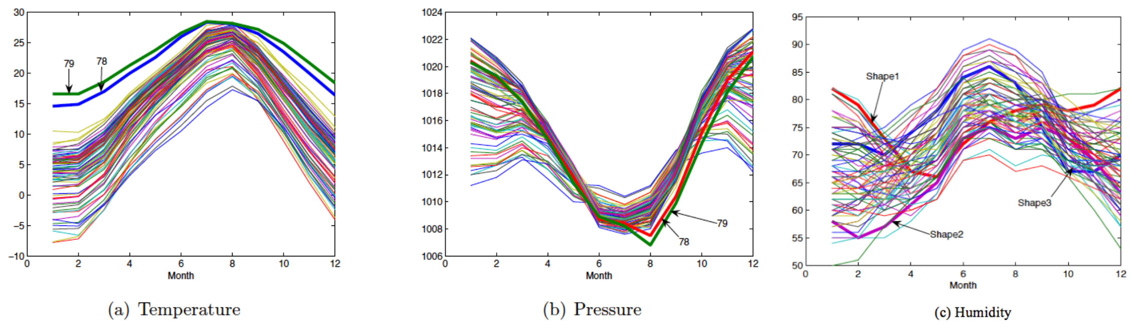

Figure 16: Outliers in Weather Data.

ACCEPTED MANUSCRIPT

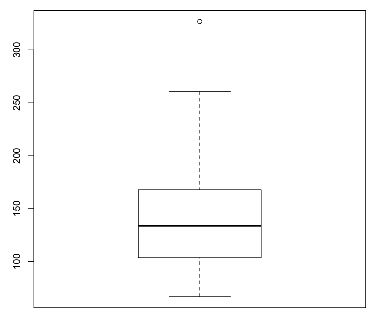

Figure 17: Outliers in response, annual total precipitation.

ACCEPTED MANUSCRIPT

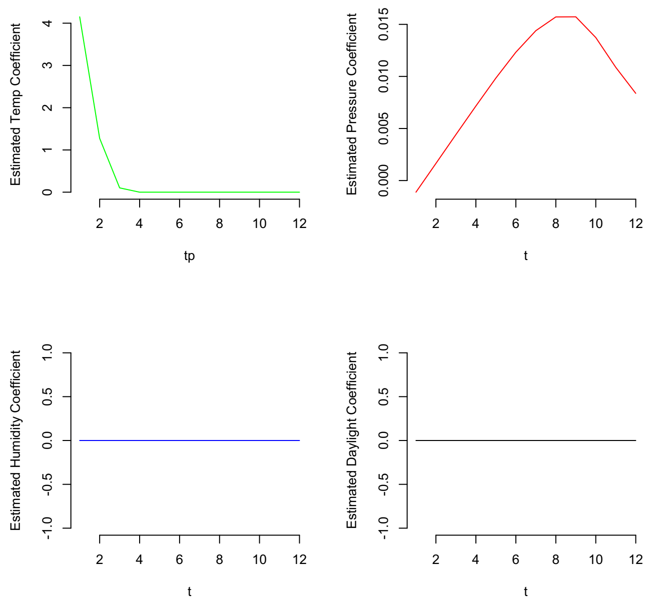

Figure 18: Estimated Variable Coefficients for Weather data using robust functional *LAD-group LASSO*.

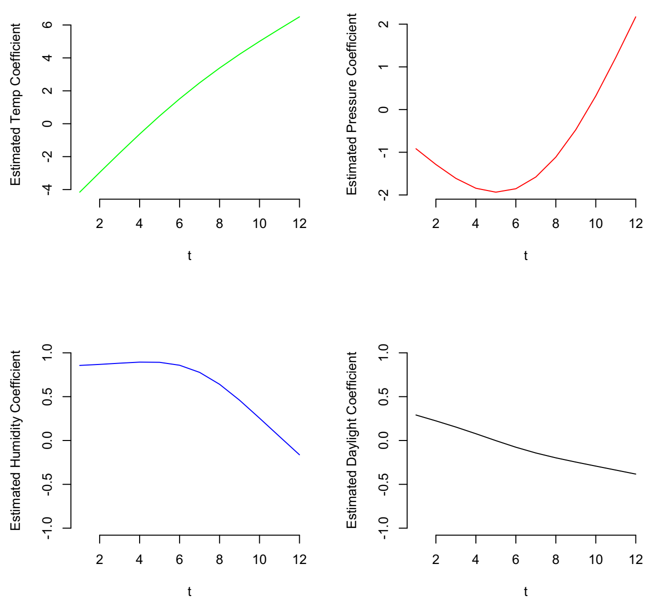

Figure 19: Estimated Variable Coefficients for Weather data using classical functional *group LASSO*.