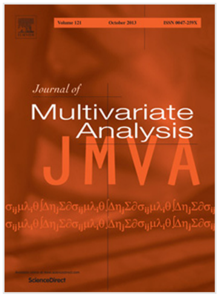

Robust ridge estimator in restricted semiparametric regression models

M. Roozbeh

PII: S0047-259X(16)00008-7
DOI: http://dx.doi.org/10.1016/j.jmva.2016.01.005
Reference: YJMVA 4074

To appear in: *Journal of Multivariate Analysis*

Received date: 15 June 2015

Robust Ridge Estimator in Restricted Semiparametric Regression Models

M. Roozbeh*

Faculty of Mathematics, Statistics and Computer Science,
Semnan University, P.O. Box 35195-363, Semnan, Iran

Abstract

In this paper, ridge and non-ridge type estimators and their robust forms are defined in the semiparametric regression model when the errors are dependent and some non-stochastic linear restrictions are imposed under a multicollinearity setting. In the context of ridge regression, the estimation of shrinkage parameter plays an important role in analyzing data. Another common problem in applied statistics is the presence of outliers in the data besides multicollinearity. In this respect, we propose some robust estimators for shrinkage parameter based on least trimmed squares (LTS) method. Given a set of n observations and the integer trimming parameter $h \leq n$, the LTS estimator involves computing the hyperplane that minimizes the sum of the smallest h squared residuals. The LTS estimator is closely related to the well-known least median squares (LMS) estimator in which the objective is to minimize the median squared residual. Although LTS estimator has the advantage of being statistically more efficient than LMS estimator, the computational complexity of LTS is less understood than LMS. Here, we extract the robust estimators for linear and non linear parts of the model based on robust shrinkage estimators. It is shown that these estimators perform better than ordinary ridge estimator. For our proposal, via a Monté-Carlo simulation and a real data example, performance of the ridge type of robust estimators are compared with the classical ones in restricted semiparametric regression models.

Key words and phrases: Breakdown point; Generalized restricted ridge estimator; Kernel smoothing; Least trimmed squares estimator; Linear restrictions; Multicollinearity; Outlier; Robust estimation; Semiparametric regression model

AMS Subject Classifications: primary: 62G08; 62G35 secondary: 62J05, 62J07

1 Introduction

Semiparametric regression models (SRMs) or are appropriate models when a suitable link function of the mean response is assumed to have a linear parametric relationship to some explanatory variables while its relationship to the other variables has an unknown form. Let $(y_1, x_1^\top, t_1),\ldots, (y_n, x_n^\top, t_n)$ be the observations that follow the semiparametric regression model, that is,

$$y_i = \boldsymbol{x}_i^\top \boldsymbol{\beta} + f(t_i) + \epsilon_i, \quad i = 1, \ldots, n, \tag{1.1}$$

*Emails: `mahdi.roozbeh@profs.semnan.ac.ir` & `m.roozbeh.stat@gmail.com`

where $\boldsymbol{x}_i^\top = (x_{i1}, \ldots, x_{ip})$ is a vector of explanatory variables, $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_p)^\top$ is an unknown p-dimensional vector parameter, t_i's are design points which belong to some bounded domain $D \subset \mathbb{R}$, $f(t)$ is an unknown smooth function and ϵ_i's are random errors which are assumed to be independent of $(\boldsymbol{x}_i, t_i)$.

Surveys regarding the estimation and application of the model (1.1) can be found in the monograph of Härdle et al. (2000). Speckman (1988) studied partial residual estimation of $\boldsymbol{\beta}$ and $f(\cdot)$ in (1.1), and obtained asymptotic bias and variance of the estimators. He showed that these estimators are less biased compared to the partial smoothing spline estimators. Bunea (2004) proposed a consistent covariate selection technique in an SRM through penalized least squares criterion. He showed that the selected estimator of the linear part is asymptotically normal. You and Chen (2007) considered the problem of estimation in model (1.1) with serially correlated errors, obtained the semiparametric generalized least-squares estimator of the parametric component and studied the asymptotic properties of it. You et al. (2007) developed statistical inference for the model (1.1) for both heteroscedastic and/or correlated errors under general assumption $Var(\boldsymbol{\epsilon}) = \sigma^2 \boldsymbol{V}$, with a positive definite matrix $\boldsymbol{V}$, is supposed to hold. For bandwidth selection in the context of kernel-based estimation in model (1.1), Li et al. (2011) used cross-validation criteria for optimal bandwidth selection.

Now consider a semiparametric regression model in the presence of multicollinearity. The existence of multicollinearity may lead to wide confidence intervals for the individual parameters or linear combination of the parameters and may produce estimates with wrong signs. For our purpose we only employ the ridge regression concept due to Hoerl and Kennard (1970), to combat multicollinearity. There are a lot of works adopting ridge regression methodology to overcome the multicollinearity problem. To mention a few recent researches in full-parametric regression, see Saleh and Kibria (1993), Hassanzadeh Bashtian et al. (2011), Kaciranlar et al. (2011), Kibria and Saleh (2011), Arashi et al. (2014) and Arashi et al. (2015). However Akdenïz and Tabakan (2009), Roozbeh et al. (2011), Akdenïz Duran et al. (2012), Roozbeh and Arashi (2013), Amini and Roozbeh (2015), Arashi and Valizadeh (2015), and Roozbeh (2015) adopted this approach in facing with semiparametric regression model. The main focus of this approach is to develop necessary tools for computing the risk function of regression coefficient in a semiparametric regression model based on the eigenvalues of design matrix and then, estimating it based on robust approach.

The restricted models are widely applicable in the problem of general hypothesis testing specially the generalized likelihood ratio (GLR) tests in regression models. Norouzirad et al. (2015) defined a restricted LASSO estimator and configure three classes of LASSO type estimators to fulfill both variable selection and restricted estimation in regression model. Akdenïz and Tabakan (2009) and Akdenïz et al. (2015) developed the restricted ridge and Liu estimators in semiparametric regression models. The problem of restricted ridge partial residual estimation in a semiparametric regression model with correlated errors is studied by Amini and Roozbeh (2015) used generalized cross-validation (GCV) criteria for optimal bandwidth and ridge parameter selection in model (1.1), simultaneously.

Besides multicollinearity, outliers (points that fail to follow partial linear pattern of the majority of the points) are another common problem in the regression analysis. Robust regression methods are used to overcome the effects of outliers (inflated sum of squares, bias or distortion of estimation, distortion of p-values, etc.). Here, we only employ the least trimmed squares semiparametric regression estimators for both parts of our model. It is well-known that the ordinary least-squares estimator is very sensitive to outliers. This motivated the researchers to focus on robust estimators. Examples of recent researches in full-parametric regression include the studies made by Edelsbrunner and Souvaine (1990), Bernholt (2005), Jung (2005), Erickson et al. (2006), Bremner et al. (2008), Nguyen and Welsch (2010), Mount et al. (2007, 2014) and Roozbeh and Babaie-Kafaki (2016).

The basic measure of the robustness of an estimator is its breakdown point, that is, the fraction (up to 50%) of outlying data points that can corrupt the estimator arbitrarily. The study of efficient algorithms for robust statistical estimators have been an active area of research in computational geometry. Many researchers studied Rousseeuw's least median of squares estimator which is defined to be the hyperplane that minimizes the median squared residual (for example, see Rousseeuw; 1984). Although the vast majority of works on robust linear estimation in the field of computational geometry has been devoted to the study of the LMS estimator, it has been observed by Rousseeuw and Leroy (1987) that LMS is not the estimator of choice from the perspective of statistical properties. They argued that a better choice is the least trimmed squares. The breakdown point of LTS and LMS are the same. Like LMS, the LTS estimator is a robust estimator with a 50%-breakdown point which means that the estimator is insensitive to the corruption made by outliers, provided that the outliers constitute less than 50% of the set. However, LTS has a number of advantages in contrast to LMS. The LTS objective function is smoother than that of LMS. LTS has better statistical efficiency because it is asymptotically normal (see Rousseeuw; 1984) and converges faster. Rousseeuw and van Driessen (2006) remarked that, for these reasons, LTS is more suitable as a starting point for two-step robust estimators such as the MM-estimator (see Yohai; 1987) and generalized M-estimators (see Simpson et al.; 1992).

The main focus of this paper is to study a robust generalized least squares ridge estimator in restricted semiparametric regression model. The organization of the paper is as follows: Section 2 contains the classical estimator of restricted semiparametric regression model based on kernel approach and related assumptions. The properties of generalized restricted ridge estimator of linear part are exactly derived in section 3. We review least trimmed squares estimators in semiparametric regression model and then, propose a new robust estimator in restricted semiparametric regression model together with its theoretical properties in Section 4. We estimate the shrinkage parameter based on robust methods in Section 5 and then, the proposed generalized least squares ridge estimator will be reconstructed based on robust estimators of shrinkage parameter. In Section 6, the efficiencies of robust estimators relative to nonrobust estimator are evaluated for both ridge and nonridge types through the Monté-Carlo simulation studies as same as a real data example. Finally, some concluding important results are stated in Section 7.

2 The Classical Estimators under Restriction

Consider the following semiparametric regression model

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{f}(\boldsymbol{t}) + \boldsymbol{\epsilon}, \tag{2.1}$$

where $\boldsymbol{y} = (y_1, \ldots, y_n)^\top$, $\boldsymbol{X} = (\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n)^\top$ is a $n \times p$ matrix, $\boldsymbol{f}(\boldsymbol{t}) = (f(t_1), \ldots, f(t_n))^\top$ and $\boldsymbol{\epsilon} = (\epsilon_1, \ldots, \epsilon_n)^\top$. We assume that in general, $\boldsymbol{\epsilon}$ is a vector of disturbances, which is distributed with $\mathrm{E}(\boldsymbol{\epsilon}) = \boldsymbol{0}$ and $\mathrm{E}(\boldsymbol{\epsilon}^\top\boldsymbol{\epsilon}) = \sigma^2\boldsymbol{V}$, where σ^2 is an unknown parameter and $\boldsymbol{V}$ is a symmetric, positive definite known matrix.

In this paper we confine ourselves to the semiparametric kernel smoothing estimator of $\boldsymbol{\beta}$, which attains the usual parametric convergence rate $n^{1/2}$ without under smoothing the nonparametric component $f(\cdot)$ (see Speckman; 1988). Assume that $(y_i, \boldsymbol{x}_i^\top, t_i)$, $i = 1, \ldots, n$ satisfy model (2.1). Since $\mathrm{E}(\varepsilon_i) = 0$, we have $f(t_i) = \mathrm{E}(y_i - \boldsymbol{x}_i^\top\boldsymbol{\beta})$ for $i = 1, \ldots, n$. Hence, if we know $\boldsymbol{\beta}$, a natural nonparametric estimator of $f(\cdot)$ is

$$\hat{f}(t, \boldsymbol{\beta}) = \sum_{i=1}^{n} W_{ni}(t)(y_i - \boldsymbol{x}_i^\top\boldsymbol{\beta}), \tag{2.2}$$

where the positive weight functions $W_{ni}(\cdot)$ satisfy three conditions below:
(*i*) $\max_{1\le i\le n}\sum_{j=1}^{n} W_{ni}(t_j) = O(1)$,
(*ii*) $\max_{1\le i,j\le n} W_{ni}(t_j) = O(n^{-2/3})$,
(*iii*) $\max_{1\le i\le n}\sum_{j=1}^{n} W_{ni}(t_j)I(|t_i - t_j| > c_n) = O(d_n)$,
where I is the indicator function, c_n satisfies $\limsup_{n\to\infty} nc_n^3 < \infty$, and d_n satisfies $\limsup_{n\to\infty} nd_n^3 < \infty$.

The above assumptions guarantee the existence of $\hat{f}(t, \boldsymbol{\beta})$ at the optimal convergence rate $n^{-4/5}$, in semiparametric regression models with probability one. See Müller (2000) for more details.

To estimate $\boldsymbol{\beta}$, we use the generalized least-squares estimator (GLSE), the best linear unbiased estimator, given by

$$\begin{aligned}\hat{\boldsymbol{\beta}}_G^{LS} &= \arg\min_{\boldsymbol{\beta}}\ (\widetilde{\boldsymbol{y}} - \widetilde{\boldsymbol{X}}\boldsymbol{\beta})^\top\boldsymbol{V}^{-1}(\widetilde{\boldsymbol{y}} - \widetilde{\boldsymbol{X}}\boldsymbol{\beta}) \\ &= \boldsymbol{C}^{-1}\widetilde{\boldsymbol{X}}^\top\boldsymbol{V}^{-1}\widetilde{\boldsymbol{y}},\end{aligned} \tag{2.3}$$

where $\boldsymbol{C} = \widetilde{\boldsymbol{X}}^\top\boldsymbol{V}^{-1}\widetilde{\boldsymbol{X}}$, $\widetilde{\boldsymbol{y}} = (\widetilde{y}_1, \ldots, \widetilde{y}_n)^\top$, $\widetilde{\boldsymbol{X}} = (\widetilde{\boldsymbol{x}}_1, \ldots, \widetilde{\boldsymbol{x}}_n)^\top$, $\widetilde{y}_i = y_i - \sum_{j=1}^{n} W_{nj}(t_i)y_j$ and $\widetilde{\boldsymbol{x}}_i = \boldsymbol{x}_i - \sum_{j=1}^{n} W_{nj}(t_i)\boldsymbol{x}_j$ for $i = 1, \ldots, n$.

Simultaneously, we assume that $\boldsymbol{\beta}$ satisfies a linear non stochastic constraint, i.e.,

$$\boldsymbol{R}\boldsymbol{\beta} = \boldsymbol{r}, \tag{2.4}$$

for a given $q \times p$ matrix $\boldsymbol{R}$ with rank $q < p$ and a given $q \times 1$ vector $\boldsymbol{r}$. In this article, we refer restricted semiparametric regression model (RSRM) to (2.4). The full row rank assumption is chosen for convenience and can be justified by the fact that every consistent linear equation can be transformed into an equivalent equation with a coefficient matrix of full row rank. Subject to the imposed linear restriction, the generalized least-squares restricted estimator (GLSRE) is given by

$$\begin{aligned}
\hat{\boldsymbol{\beta}}_{G\boldsymbol{R}}^{LS} &= \arg\min_{\boldsymbol{\beta}} \ (\widetilde{\boldsymbol{y}} - \widetilde{\boldsymbol{X}}\boldsymbol{\beta})^\top \boldsymbol{V}^{-1}(\widetilde{\boldsymbol{y}} - \widetilde{\boldsymbol{X}}\boldsymbol{\beta}) \\
& \quad\ s.t. \quad \boldsymbol{R}\boldsymbol{\beta} = \boldsymbol{r}, \\
&= \hat{\boldsymbol{\beta}}_G^{LS} - \boldsymbol{C}^{-1}\boldsymbol{R}^\top \left(\boldsymbol{R}\boldsymbol{C}^{-1}\boldsymbol{R}^\top\right)^{-1}\left(\boldsymbol{R}\hat{\boldsymbol{\beta}}_G^{LS} - \boldsymbol{r}\right). \qquad (2.5)
\end{aligned}$$

In this section, we will be discussing a biased estimation technique under multicollinearity, for RSRM. The covariance matrix of $\hat{\boldsymbol{\beta}}_G^{LS}$ is equal to $\sigma^2\boldsymbol{C}^{-1}$. As it can be seen, both GLSE and its covariance matrix heavily depend on the characteristics of the matrix $\boldsymbol{C}$. If $\boldsymbol{C}$ is ill-conditioned, the GLS estimators are sensitive to a number of errors. For example, some of the regression coefficients may be statistically insignificant or have wrong signs, and they may result in wide confidence intervals for individual parameters (which are called unstable estimators). With these errors, it is difficult to make valid statistical inference.

The problem of multicollinearity can be solved by collecting additional data, re-parameterizing the model and reselecting variables. There are two well-known mathematical methods to overcome multicollinearity: the principal components regression method and the ridge regression method. In this article, we will discuss the ridge regression method. A brief review of the literature reveals an abundance of works related to the ridge regression method. Hoerl and Kennard (1970) first proposed this method to solve the multicollinearity problem. They suggested a small positive number to be added to the diagonal elements of $\boldsymbol{C}$ matrix; and the resulting estimator has form

$$\hat{\boldsymbol{\beta}}_G^{LS}(k) = \boldsymbol{C}_k^{-1}\widetilde{\boldsymbol{X}}^\top \boldsymbol{V}^{-1}\widetilde{\boldsymbol{y}}, \ \boldsymbol{C}_k = \boldsymbol{C} + k\boldsymbol{I}_p, \qquad (2.6)$$

which is known as a generalized least-squares ridge estimator. For a positive value of k, this estimator provides a smaller mean squared error (MSE) compared to the GLSE. The constant k ($k \geq 0$) is called the "ridge" or "shrinkage" parameter, and it must be estimated using the data. Although, the ridge estimator is the most popular method for dealing with multicollinearity, it has some drawbacks. Dependency on the ridge parameter k tends to result in either instability or bias. However, as $k \to \infty$, $\hat{\boldsymbol{\beta}}_G^{LS}(k) \to \boldsymbol{0}$ and one obtains a stable, but biased estimator of $\boldsymbol{\beta}$. As $k \to 0$, $\hat{\boldsymbol{\beta}}_G^{LS}(k) \to \hat{\boldsymbol{\beta}}_G^{LS}$ and one obtains an unbiased, but unstable estimator of $\boldsymbol{\beta}$. The expected distance between $\hat{\boldsymbol{\beta}}_G^{LS}(k)$ and $\boldsymbol{\beta}$ must decrease as k increases from the origin. The value of k that produces the best estimator, however, is not clear. It is realized that the estimator $\hat{\boldsymbol{\beta}}_G^{LS}(k)$ is a complicated function of k.

It is clear that for the semi-positive definite matrix $\boldsymbol{C}$, there exists an orthogonal matrix $\boldsymbol{\Gamma}$ such that $\boldsymbol{C} = \boldsymbol{\Gamma}\boldsymbol{\Lambda}\boldsymbol{\Gamma}^\top$, where $\boldsymbol{\Lambda} = diag(\lambda_1, \ldots, \lambda_p)$ contains the eigenvalues of matrix $\boldsymbol{C}$. Therefore, the orthogonal (canonical) version of the model (2.1) is given by

$$\widetilde{\boldsymbol{y}} = \widetilde{\boldsymbol{X}}^*\boldsymbol{\alpha} + \boldsymbol{\epsilon}, \qquad (2.7)$$

where $\widetilde{\boldsymbol{X}}^* = \widetilde{\boldsymbol{X}}\boldsymbol{\Gamma}$ and $\boldsymbol{\alpha} = \boldsymbol{\Gamma}^\top\boldsymbol{\beta}$.

But, when the matrix $\boldsymbol{C}$ is ill-conditioned (in the sense that there is a near linear dependency among the columns of matrix), the GLSE of $\boldsymbol{\beta}$ has a large variance, and multicollinearity is said to be present. If multicollinearity is present, at least for one eigenvalue, $\lambda_i \doteq 0$. The more closeness of the smallest eigenvalue to the origin, the more strength of linear multicollinearity. To make the behavior of $\boldsymbol{C}$ matrix more like the canonical form, we need to increase the eigenvalues. Ridge regression replaces $\boldsymbol{C}$ with $\boldsymbol{C}_k = \boldsymbol{C} + k\boldsymbol{I}_p$, $(k > 0)$, which is the same as replacing the λ_i by $\lambda_i + k$. This replacement counters the damaging effect of the smallest eigenvalue.

Based on Hoerl et al. (1975), the ridge parameter k can be estimated using GLSRE in RSRM as follows:

$$\hat{k}_{LS} = \frac{p\hat{\sigma}^2_{LS}}{\hat{\boldsymbol{\beta}}^{LS\,\top}_{\boldsymbol{GR}} \hat{\boldsymbol{\beta}}^{LS}_{\boldsymbol{GR}}},\ \hat{\sigma}^2_{LS} = \frac{1}{n-(p+q)}(\widetilde{\boldsymbol{y}} - \widetilde{\boldsymbol{X}}\hat{\boldsymbol{\beta}}^{LS}_{\boldsymbol{GR}})^\top \boldsymbol{V}^{-1}(\widetilde{\boldsymbol{y}} - \widetilde{\boldsymbol{X}}\hat{\boldsymbol{\beta}}^{LS}_{\boldsymbol{GR}}). \tag{2.8}$$

Following Swamy et al. (1978), the restricted ridge estimator can be obtained by minimizing the sum of squared residuals with a spherical restriction and a linear restriction (2.4), i.e., the restricted linear semiparametric regression is transformed into an optimal problem with two restrictions:

$$\begin{aligned} &\min(\widetilde{\boldsymbol{y}} - \widetilde{\boldsymbol{X}}\boldsymbol{\beta})^\top \boldsymbol{V}^{-1}(\widetilde{\boldsymbol{y}} - \widetilde{\boldsymbol{X}}\boldsymbol{\beta}) \\ &\text{s.t.}\quad \boldsymbol{\beta}^\top \top \boldsymbol{\beta} \le \phi^2, \\ &\qquad \boldsymbol{R}\boldsymbol{\beta} = \boldsymbol{r}. \end{aligned}$$

The resulting estimator is given by

$$\begin{aligned} \hat{\boldsymbol{\beta}}^{LS}_{\boldsymbol{GR}}(k) &= (\boldsymbol{C} + k\boldsymbol{I})^{-1}\widetilde{\boldsymbol{X}}^\top \boldsymbol{V}^{-1}\widetilde{\boldsymbol{y}} \\ &\quad -(\boldsymbol{C} + k\boldsymbol{I})^{-1}\boldsymbol{R}^\top\big(\boldsymbol{R}(\boldsymbol{C} + k\boldsymbol{I})^{-1}\boldsymbol{R}^\top\big)^{-1}\big(\boldsymbol{R}(\boldsymbol{C} + k\boldsymbol{I})^{-1}\widetilde{\boldsymbol{X}}^\top \boldsymbol{V}^{-1}_{\boldsymbol{D}}\widetilde{\boldsymbol{y}} - \boldsymbol{r}\big) \\ &= \hat{\boldsymbol{\beta}}^{LS}_{G}(k) - \boldsymbol{C}^{-1}_k\boldsymbol{R}^\top\big(\boldsymbol{R}\boldsymbol{C}^{-1}_k\boldsymbol{R}^\top\big)^{-1}\big(\boldsymbol{R}\hat{\boldsymbol{\beta}}^{LS}_{G}(k) - \boldsymbol{r}\big). \end{aligned} \tag{2.9}$$

We designate generalized least-squares restricted ridge estimator (GLSRRE) to (2.9). Comparing GLSRRE with GLSRE, we can easily find that the form of our method, where $\hat{\boldsymbol{\beta}}^{LS}_{G}(k)$ and $\boldsymbol{C}_k$, taking the places of $\hat{\boldsymbol{\beta}}^{LS}_{G}$ and $\boldsymbol{C}$ of GLSRE, respectively, is completely consistent with that of GLSRE. Furthermore, we can give another explanation to the GLSRRE by transforming (2.9) into

$$\hat{\boldsymbol{\beta}}^{LS}_{\boldsymbol{GR}}(k) = \Big(\boldsymbol{I} - \boldsymbol{C}^{-1}_k\boldsymbol{R}^\top(\boldsymbol{R}\boldsymbol{C}^{-1}_k\boldsymbol{R}^\top)^{-1}\boldsymbol{R}\Big)\hat{\boldsymbol{\beta}}^{LS}_{G}(k) + \boldsymbol{C}_k\boldsymbol{R}^\top\big(\boldsymbol{R}\boldsymbol{C}^{-1}_k\boldsymbol{R}^\top\big)^{-1}\boldsymbol{r}. \tag{2.10}$$

Since $\boldsymbol{R}\Big(\boldsymbol{C}^{-1}_k\boldsymbol{R}^\top(\boldsymbol{R}\boldsymbol{C}^{-1}_k\boldsymbol{R}^\top)^{-1}\Big)\boldsymbol{R} = \boldsymbol{R}$ and the definition of generalized inverse, it is easy to see that $\boldsymbol{C}^{-1}_k\boldsymbol{R}^\top\big(\boldsymbol{R}\boldsymbol{C}^{-1}_k\boldsymbol{R}^\top\big)^{-1}$ is a generalized inverse of $\boldsymbol{R}$. Define $\boldsymbol{R}^- = \boldsymbol{C}^{-1}_k\boldsymbol{R}^\top\big(\boldsymbol{R}\boldsymbol{C}^{-1}_k\boldsymbol{R}^\top\big)^{-1}$, (2.10) is equivalent to

$$\hat{\boldsymbol{\beta}}^{LS}_{\boldsymbol{GR}}(k) = (\boldsymbol{I} - \boldsymbol{R}^-\boldsymbol{R})\hat{\boldsymbol{\beta}}^{LS}_{G}(k) + \boldsymbol{R}^-\boldsymbol{r} \tag{2.11}$$

It is easy to see that the $\hat{\boldsymbol{\beta}}^{LS}_{\boldsymbol{GR}}$ and $\hat{\boldsymbol{\beta}}^{LS}_{\boldsymbol{GR}}(k)$ are restricted with respect to $\boldsymbol{R}\boldsymbol{\beta} = \boldsymbol{r}$. It is also clear that for $k = 0$ we get $\hat{\boldsymbol{\beta}}^{LS}_{\boldsymbol{GR}}(0) = \hat{\boldsymbol{\beta}}^{LS}_{\boldsymbol{GR}}$.

3 Computing the Risk Function

For any particular estimator $\hat{\boldsymbol{\beta}}$ of $\boldsymbol{\beta}$, the risk function under square error loss is measured by

$$\mathrm{R}(\hat{\boldsymbol{\beta}},\boldsymbol{\beta}) = \mathrm{E}\big\{(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta})^\top(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta})\big\}.$$

Lemma 3.1. *(Roozbeh and Arashi (2013)) If $\boldsymbol{\beta}$ satisfies the linear restriction $\boldsymbol{R\beta}=\boldsymbol{r}$, then the bias, covariance matrix and risk functions of proposed estimator can be evaluated as follows:*

$$\begin{aligned}
\mathbf{b}\big(\hat{\boldsymbol{\beta}}^{LS}_{G\boldsymbol{R}}(k)\big) &= \mathrm{E}\left(\hat{\boldsymbol{\beta}}^{LS}_{G\boldsymbol{R}}(k)-\boldsymbol{\beta}\right)\\
&= -k\boldsymbol{M}_k\boldsymbol{\beta}, && (3.1)\\
\mathrm{Cov}\left(\hat{\boldsymbol{\beta}}^{LS}_{G\boldsymbol{R}}(k)\right) &= \sigma^2\boldsymbol{M}_k\boldsymbol{C}\boldsymbol{M}_k, && (3.2)\\
\mathrm{R}\big(\hat{\boldsymbol{\beta}}^{\mathrm{LS}}_{\mathrm{G}\boldsymbol{R}}(\mathrm{k}),\boldsymbol{\beta}\big) &= \sigma^2\,\mathrm{tr}\left(\boldsymbol{M}_k\boldsymbol{C}\boldsymbol{M}_k\right)+k^2\boldsymbol{\beta}^\top\boldsymbol{M}_k^2\boldsymbol{\beta}, && (3.3)
\end{aligned}$$

where $\boldsymbol{M}_k=\boldsymbol{C}_k^{-1}-\boldsymbol{C}_k^{-1}\boldsymbol{R}^\top\left(\boldsymbol{R}\boldsymbol{C}_k^{-1}\boldsymbol{R}^\top\right)^{-1}\boldsymbol{R}\boldsymbol{C}_k^{-1}$.

Then the properties of $\hat{\boldsymbol{\beta}}^{LS}_{G\boldsymbol{R}}$ is obtained by letting $k=0$ in the above Lemma as follows:

$$\begin{aligned}
\mathbf{b}\big(\hat{\boldsymbol{\beta}}^{LS}_{G\boldsymbol{R}}\big) &= \mathbf{0}, && (3.4)\\
\mathrm{Cov}\left(\hat{\boldsymbol{\beta}}^{LS}_{G\boldsymbol{R}}\right) &= \sigma^2\boldsymbol{M}_0\boldsymbol{C}\boldsymbol{M}_0, && (3.5)\\
\mathrm{R}\big(\hat{\boldsymbol{\beta}}^{\mathrm{LS}}_{\mathrm{G}\boldsymbol{R}},\boldsymbol{\beta}\big) &= \sigma^2\,\mathrm{tr}\left(\boldsymbol{M}_0\boldsymbol{C}\boldsymbol{M}_0\right). && (3.6)
\end{aligned}$$

Theorem 3.1. *The risk function of the estimator under study can be given by*

$$\mathrm{R}\big(\hat{\boldsymbol{\beta}}^{\mathrm{LS}}_{\mathrm{G}\boldsymbol{R}}(\mathrm{k}),\boldsymbol{\beta}\big) = \sigma^2\sum_{i=1}^{p}\frac{\lambda_i(\lambda_i+k-r^*_{ii})^2}{(\lambda_i+k)^4}+k^2\sum_{i=1}^{p}\left[\frac{\alpha_i(\lambda_i+k-r^*_{ii})}{(\lambda_i+k)^2}\right]^2, \qquad (3.7)$$

where $\boldsymbol{R}^*=\boldsymbol{\Gamma}^{top}\boldsymbol{R}^\top\left(\boldsymbol{R}\boldsymbol{C}_k^{-1}\boldsymbol{R}^\top\right)^{-1}\boldsymbol{R}\boldsymbol{\Gamma}$ and $Diag(\boldsymbol{R}^*)=r^*_{ii}$.

Proof. We can write

$$\begin{aligned}
\boldsymbol{M}_k\boldsymbol{C}\boldsymbol{M}_k =& \Big(\boldsymbol{C}_k^{-1}-\boldsymbol{C}_k^{-1}\boldsymbol{R}^\top\left(\boldsymbol{R}\boldsymbol{C}_k^{-1}\boldsymbol{R}^\top\right)^{-1}\boldsymbol{R}\boldsymbol{C}_k^{-1}\Big)\,\boldsymbol{C}\,\Big(\boldsymbol{C}_k^{-1}-\boldsymbol{C}_k^{-1}\boldsymbol{R}^\top\left(\boldsymbol{R}\boldsymbol{C}_k^{-1}\boldsymbol{R}^\top\right)^{-1}\boldsymbol{R}\boldsymbol{C}_k^{-1}\Big)\\
=& \left(\boldsymbol{C}_k^{-1}\boldsymbol{C}\boldsymbol{C}_k^{-1}\right)-\Big(\boldsymbol{C}_k^{-1}\boldsymbol{C}\boldsymbol{C}_k^{-1}\boldsymbol{R}^\top\left(\boldsymbol{R}\boldsymbol{C}_k^{-1}\boldsymbol{R}^\top\right)^{-1}\boldsymbol{R}\boldsymbol{C}_k^{-1}\Big)\\
&-\Big(\boldsymbol{C}_k^{-1}\boldsymbol{R}^\top\left(\boldsymbol{R}\boldsymbol{C}_k^{-1}\boldsymbol{R}^\top\right)^{-1}\boldsymbol{R}\boldsymbol{C}_k^{-1}\boldsymbol{C}\boldsymbol{C}_k^{-1}\Big)\\
&+\Big(\boldsymbol{C}_k^{-1}\boldsymbol{R}^\top\left(\boldsymbol{R}\boldsymbol{C}_k^{-1}\boldsymbol{R}^\top\right)^{-1}\boldsymbol{R}\boldsymbol{C}_k^{-1}\boldsymbol{C}\boldsymbol{C}_k^{-1}\boldsymbol{R}^\top\left(\boldsymbol{R}\boldsymbol{C}_k^{-1}\boldsymbol{R}^\top\right)^{-1}\boldsymbol{R}\boldsymbol{C}_k^{-1}\Big).
\end{aligned}$$

Also, from $\boldsymbol{C}_k^{-1}=\boldsymbol{\Gamma}(\boldsymbol{\Lambda}+k\boldsymbol{I})^{-1}\boldsymbol{\Gamma}^\top$, we have

$$\begin{aligned}
\mathrm{tr}\left\{\boldsymbol{C}_k^{-1}\boldsymbol{C}\boldsymbol{C}_k^{-1}\right\} &= \mathrm{tr}\Big\{\boldsymbol{\Gamma}(\boldsymbol{\Lambda}+k\boldsymbol{I})^{-1}\boldsymbol{\Gamma}^\top\boldsymbol{\Gamma}\boldsymbol{\Lambda}\boldsymbol{\Gamma}^\top\boldsymbol{\Gamma}(\boldsymbol{\Lambda}+k\boldsymbol{I})^{-1}\boldsymbol{\Gamma}^\top\Big\}\\
&= \mathrm{tr}\left\{\boldsymbol{\Lambda}(\boldsymbol{\Lambda}+k\boldsymbol{I})^{-2}\right\}\\
&= \sum_{i=1}^{p}\frac{\lambda_i}{(\lambda_i+k)^2}, && (3.8)
\end{aligned}$$

$$
\begin{aligned}
\operatorname{tr}&\left\{\boldsymbol{C}_k^{-1}\boldsymbol{C}\boldsymbol{C}_k^{-1}\boldsymbol{R}^\top\left(\boldsymbol{R}\boldsymbol{C}_k^{-1}\boldsymbol{R}^\top\right)^{-1}\boldsymbol{R}\boldsymbol{C}_k^{-1}\right\}\\
&= \operatorname{tr}\left\{\boldsymbol{\Gamma}(\boldsymbol{\Lambda}+k\boldsymbol{I})^{-1}\boldsymbol{\Gamma}^\top\boldsymbol{\Gamma}\boldsymbol{\Lambda}\boldsymbol{\Gamma}^\top\boldsymbol{\Gamma}(\boldsymbol{\Lambda}+k\boldsymbol{I})^{-1}\boldsymbol{\Gamma}^\top\boldsymbol{R}^\top\left(\boldsymbol{R}\boldsymbol{C}_k^{-1}\boldsymbol{R}^\top\right)^{-1}\boldsymbol{R}\boldsymbol{\Gamma}(\boldsymbol{\Lambda}+k\boldsymbol{I})^{-1}\boldsymbol{\Gamma}^\top\right\}\\
&= \operatorname{tr}\left\{(\boldsymbol{\Lambda}+k\boldsymbol{I})^{-2}\boldsymbol{\Lambda}(\boldsymbol{\Lambda}+k\boldsymbol{I})^{-1}\boldsymbol{R}^*\right\}\\
&= \sum_{i=1}^{p}\frac{r_{ii}^*\lambda_i}{(\lambda_i+k)^3}, \qquad (3.9)
\end{aligned}
$$

$$
\begin{aligned}
\operatorname{tr}&\left\{\boldsymbol{C}_k^{-1}\boldsymbol{R}^\top\left(\boldsymbol{R}\boldsymbol{C}_k^{-1}\boldsymbol{R}^\top\right)^{-1}\boldsymbol{R}\boldsymbol{C}_k^{-1}\boldsymbol{C}\boldsymbol{C}_k^{-1}\right\}\\
&= \operatorname{tr}\left\{\boldsymbol{\Gamma}(\boldsymbol{\Lambda}+k\boldsymbol{I})^{-1}\boldsymbol{\Gamma}^\top\boldsymbol{R}^\top\left(\boldsymbol{R}\boldsymbol{C}_k^{-1}\boldsymbol{R}^\top\right)^{-1}\boldsymbol{R}\boldsymbol{\Gamma}(\boldsymbol{\Lambda}+k\boldsymbol{I})^{-1}\boldsymbol{\Gamma}^\top\boldsymbol{\Gamma}\boldsymbol{\Lambda}\boldsymbol{\Gamma}^\top\boldsymbol{\Gamma}(\boldsymbol{\Lambda}+k\boldsymbol{I})^{-1}\boldsymbol{\Gamma}^\top\right\}\\
&= \operatorname{tr}\left\{(\boldsymbol{\Lambda}+k\boldsymbol{I})^{-2}\boldsymbol{R}^*(\boldsymbol{\Lambda}+k\boldsymbol{I})^{-1}\boldsymbol{\Lambda}\right\}\\
&= \sum_{i=1}^{p}\frac{r_{ii}^*\lambda_i}{(\lambda_i+k)^3} \qquad (3.10)
\end{aligned}
$$

and

$$
\begin{aligned}
\operatorname{tr}&\left\{\boldsymbol{C}_k^{-1}\boldsymbol{R}^\top(\boldsymbol{R}\boldsymbol{C}_k^{-1}\boldsymbol{R}^\top)^{-1}\boldsymbol{R}\boldsymbol{C}_k^{-1}\boldsymbol{C}\boldsymbol{C}_k^{-1}\boldsymbol{R}^\top\left(\boldsymbol{R}\boldsymbol{C}_k^{-1}\boldsymbol{R}^\top\right)^{-1}\boldsymbol{R}\boldsymbol{C}_k^{-1}\right\}\\
&= \operatorname{tr}\left\{\boldsymbol{\Gamma}(\boldsymbol{\Lambda}+k\boldsymbol{I})^{-1}\boldsymbol{\Gamma}^\top\boldsymbol{R}^\top\left(\boldsymbol{R}\boldsymbol{C}_k^{-1}\boldsymbol{R}^\top\right)^{-1}\boldsymbol{R}\boldsymbol{\Gamma}(\boldsymbol{\Lambda}+k\boldsymbol{I})^{-1}\boldsymbol{\Gamma}^\top\boldsymbol{\Gamma}\boldsymbol{\Lambda}\boldsymbol{\Gamma}^\top\boldsymbol{\Gamma}(\boldsymbol{\Lambda}+k\boldsymbol{I})^{-1}\boldsymbol{\Gamma}^\top\boldsymbol{R}^\top\left(\boldsymbol{R}\boldsymbol{C}_k^{-1}\boldsymbol{R}^\top\right)^{-1}\boldsymbol{R}\boldsymbol{\Gamma}(\boldsymbol{\Lambda}+k\boldsymbol{I})^{-1}\boldsymbol{\Gamma}^\top\right\}\\
&= \operatorname{tr}\left\{(\boldsymbol{\Lambda}+k\boldsymbol{I})^{-2}\boldsymbol{R}^*(\boldsymbol{\Lambda}+k\boldsymbol{I})^{-1}\boldsymbol{\Lambda}(\boldsymbol{\Lambda}+k\boldsymbol{I})^{-1}\boldsymbol{R}^*\right\}\\
&= \sum_{i=1}^{p}\frac{{r_{ii}^*}^2\lambda_i}{(\lambda_i+k)^4}. \qquad (3.11)
\end{aligned}
$$

Thus, we get

$$
\begin{aligned}
\operatorname{tr}\left(\boldsymbol{M}_k\boldsymbol{C}\boldsymbol{M}_k\right) &= \sum_{i=1}^{p}\frac{\lambda_i}{(\lambda_i+k)^2}-2\sum_{i=1}^{p}\frac{r_{ii}^*\lambda_i}{(\lambda_i+k)^3}+\sum_{i=1}^{p}\frac{{r_{ii}^*}^2\lambda_i}{(\lambda_i+k)^4}\\
&= \sum_{i=1}^{p}\frac{\lambda_i(\lambda_i+k-r_{ii}^*)^2}{(\lambda_i+k)^4}. \qquad (3.12)
\end{aligned}
$$

Also,

$$
\begin{aligned}
\boldsymbol{M}_k^2 &= \left(\boldsymbol{C}_k^{-1}-\boldsymbol{C}_k^{-1}\boldsymbol{R}^\top\left(\boldsymbol{R}\boldsymbol{C}_k^{-1}\boldsymbol{R}^\top\right)^{-1}\boldsymbol{R}\boldsymbol{C}_k^{-1}\right)\left(\boldsymbol{C}_k^{-1}-\boldsymbol{C}_k^{-1}\boldsymbol{R}^\top\left(\boldsymbol{R}\boldsymbol{C}_k^{-1}\boldsymbol{R}^\top\right)^{-1}\boldsymbol{R}\boldsymbol{C}_k^{-1}\right)\\
&= \boldsymbol{C}_k^{-2}-\boldsymbol{C}_k^{-2}\boldsymbol{R}^\top\left(\boldsymbol{R}\boldsymbol{C}_k^{-1}\boldsymbol{R}^\top\right)^{-1}\boldsymbol{R}\boldsymbol{C}_k^{-1}-\boldsymbol{C}_k^{-1}\boldsymbol{R}^\top\left(\boldsymbol{R}\boldsymbol{C}_k^{-1}\boldsymbol{R}^\top\right)^{-1}\boldsymbol{R}\boldsymbol{C}_k^{-2}\\
&\quad +\boldsymbol{C}_k^{-1}\boldsymbol{R}^\top\left(\boldsymbol{R}\boldsymbol{C}_k^{-1}\boldsymbol{R}^\top\right)^{-1}\boldsymbol{R}\boldsymbol{C}_k^{-2}\boldsymbol{R}^\top\left(\boldsymbol{R}\boldsymbol{C}_k^{-1}\boldsymbol{R}^\top\right)^{-1}\boldsymbol{R}\boldsymbol{C}_k^{-1}. \qquad (3.13)
\end{aligned}
$$

By some algebraic manipulations, we can obtain

$$
\operatorname{tr}\left(\boldsymbol{C}_k^{-2}\right)=\operatorname{tr}\left\{\boldsymbol{\Gamma}(\boldsymbol{\Lambda}+k\boldsymbol{I})^{-1}\boldsymbol{\Gamma}^\top\boldsymbol{\Gamma}(\boldsymbol{\Lambda}+k\boldsymbol{I})^{-1}\boldsymbol{\Gamma}^\top\right\}=\sum_{i=1}^{p}\frac{1}{(\lambda_i+k)^2}, \qquad (3.14)
$$

$$
\begin{aligned}
\operatorname{tr}\left\{\boldsymbol{C}_k^{-2}\boldsymbol{R}^\top\left(\boldsymbol{R}\boldsymbol{C}_k^{-1}\boldsymbol{R}^\top\right)^{-1}\boldsymbol{R}\boldsymbol{C}_k^{-1}\right\} &= \operatorname{tr}\left\{\boldsymbol{C}_k^{-1}\boldsymbol{R}^\top\left(\boldsymbol{R}\boldsymbol{C}_k^{-1}\boldsymbol{R}^\top\right)^{-1}\boldsymbol{R}\boldsymbol{C}_k^{-2}\right\}\\
&= \operatorname{tr}\left\{\boldsymbol{\Gamma}(\boldsymbol{\Lambda}+k\boldsymbol{I})^{-1}\boldsymbol{R}^*(\boldsymbol{\Lambda}+k\boldsymbol{I})^{-2}\boldsymbol{\Gamma}^\top\right\}\\
&= \sum_{i=1}^{p}\frac{r_{ii}^*}{(\lambda_i+k)^3}. \qquad (3.15)
\end{aligned}
$$

In a similar fashion, we have

$$\operatorname{tr}\left\{\boldsymbol{C}_k^{-1}\boldsymbol{R}^\top\left(\boldsymbol{R}\boldsymbol{C}_k^{-1}\boldsymbol{R}^\top\right)^{-1}\boldsymbol{R}\boldsymbol{C}_k^{-2}\boldsymbol{R}^\top\left(\boldsymbol{R}\boldsymbol{C}_k^{-1}\boldsymbol{R}^\top\right)^{-1}\boldsymbol{R}\boldsymbol{C}_k^{-1}\right\} = \sum_{i=1}^p \frac{{r_{ii}^*}^2}{(\lambda_i+k)^4}. \tag{3.16}$$

Using equations (3.14) - (3.16), we have the point

$$\operatorname{tr}\left(\boldsymbol{M}_k^2\right) = \sum_{i=1}^p \left(\frac{1}{(\lambda_i+k)^2} - \frac{2r_{ii}^*}{(\lambda_i+k)^3} + \frac{{r_{ii}^*}^2}{(\lambda_i+k)^4}\right) = \sum_{i=1}^p \frac{(\lambda_i+k-r_{ii}^*)^2}{(\lambda_i+k)^4}.$$

The result follows by applying Lemma 4.1 and the fact that

$$\alpha^\top \boldsymbol{M}_k^2 \alpha \quad = \quad \sum_{i=1}^p \left\{\frac{\alpha_i(\lambda_i+k-r_{ii}^*)}{(\lambda_i+k)^2}\right\}^2,$$

where $\boldsymbol{\alpha} = \boldsymbol{\Gamma}^\top\boldsymbol{\beta} = (\alpha_1, \cdots, \alpha_p)^\top$. Finally, we have

$$\begin{aligned} \mathrm{R}\big(\hat{\boldsymbol{\beta}}_{\boldsymbol{GR}}^{LS}(k), \boldsymbol{\beta}\big) &= \sigma^2 \operatorname{tr}\left(\boldsymbol{M}_k\boldsymbol{C}\boldsymbol{M}_k\right) + k^2\boldsymbol{\beta}^\top\boldsymbol{M}_k^2(k)\boldsymbol{\beta} \\ &= \sigma^2\sum_{i=1}^p \frac{\lambda_i(\lambda_i+k-r_{ii}^*)^2}{(\lambda_i+k)^4} + k^2\sum_{i=1}^p\left\{\frac{\alpha_i(\lambda_i+k-r_{ii}^*)}{(\lambda_i+k)^2}\right\}^2. \end{aligned} \tag{3.17}$$

So, the proof is completed. □

4 Robust Approaches

We have mentioned that outliers can strongly corrupt the least-squares fit due to their dominant effect on the objective function. Least trimmed squares attempts to solve this problem by minimizing the sum of the smallest h squared residuals rather than the complete sum of squares. Here, h is a threshold such that the ratio $\alpha = (n-h)/n$ represents the percentage of the outlying observations.

Let z_i be the indicator whether observation i is a good observation or not. Consider the LTS problem in RSRM as follows:

$$\begin{aligned} \min_{\boldsymbol{\beta},\boldsymbol{z}} \quad & \psi(\boldsymbol{\beta},\boldsymbol{z}) = (\widetilde{\boldsymbol{y}} - \widetilde{\boldsymbol{X}}\boldsymbol{\beta})^\top\boldsymbol{V}^{-1/2}\boldsymbol{Z}\boldsymbol{V}^{-1/2}(\widetilde{\boldsymbol{y}} - \widetilde{\boldsymbol{X}}\boldsymbol{\beta}) \\ s.t. \quad & \boldsymbol{R}\boldsymbol{\beta} = \boldsymbol{r}, \\ & \boldsymbol{e}^\top\boldsymbol{z} = h, \\ & z_i \in \{0,1\},\ i = 1,\ldots,n, \end{aligned} \tag{4.1}$$

where $\boldsymbol{Z}$ is the diagonal matrix with diagonal elements $\boldsymbol{z} = (z_1,\ldots,z_n)^\top$ and $\boldsymbol{e} = (1,\ldots,1)_{n\times1}^\top$. The resulting estimator is generalized least trimmed squares restricted estimator (GLTSRE), which is given by

$$\hat{\boldsymbol{\beta}}_{\boldsymbol{GR}}^{LTS}(\boldsymbol{z}) = \hat{\boldsymbol{\beta}}_G^{LTS}(\boldsymbol{z}) - \boldsymbol{C}(\boldsymbol{z})^{-1}\boldsymbol{R}^\top\left(\boldsymbol{R}\boldsymbol{C}(\boldsymbol{z})^{-1}\boldsymbol{R}^\top\right)^{-1}\left(\boldsymbol{R}\hat{\boldsymbol{\beta}}_G^{LTS}(\boldsymbol{z}) - \boldsymbol{r}\right), \tag{4.2}$$

where $\boldsymbol{C}(\boldsymbol{z}) = \widetilde{\boldsymbol{X}}^\top\boldsymbol{V}^{-1/2}\boldsymbol{Z}\boldsymbol{V}^{-1/2}\widetilde{\boldsymbol{X}}$ and $\hat{\boldsymbol{\beta}}_G^{LTS}(\boldsymbol{z}) = \boldsymbol{C}(\boldsymbol{z})^{-1}\widetilde{\boldsymbol{X}}^\top\boldsymbol{V}^{-1/2}\boldsymbol{Z}\boldsymbol{V}^{-1/2}\widetilde{\boldsymbol{y}}$.

Based on Nguyen and Welsch (2010), we can consider the following relaxation, here called relaxed least trimmed squares (RLTS) problem, in RSRM as follows:

$$\begin{aligned}
\min_{\boldsymbol{\beta},\boldsymbol{z}^*} \quad & \psi(\boldsymbol{\beta},\boldsymbol{z}^*) = (\widetilde{\boldsymbol{y}} - \widetilde{\boldsymbol{X}}\boldsymbol{\beta})^\top \boldsymbol{V}^{-1/2}\boldsymbol{Z}^*\boldsymbol{V}^{-1/2}(\widetilde{\boldsymbol{y}} - \widetilde{\boldsymbol{X}}\boldsymbol{\beta}) \\
s.t. \quad & \boldsymbol{R}\boldsymbol{\beta} = \boldsymbol{r}, \\
& \boldsymbol{e}^\top \boldsymbol{z}^* = h, \\
& 0 \le z_i^* \le 1,\ i = 1,\ldots,n,
\end{aligned} \tag{4.3}$$

where $\boldsymbol{Z}^*$ is the diagonal matrix with diagonal elements $\boldsymbol{z}^* = (z_1^*,\ldots,z_n^*)^\top$ and h is a positive integer. This optimization problem lead to the generalized relaxed least trimmed squares restricted estimator (GRLTSRE) of the regression coefficients $\boldsymbol{\beta}$, obtained in the RSRM as

$$\hat{\boldsymbol{\beta}}_{G\boldsymbol{R}}^{RLTS}(\boldsymbol{z}^*) = \hat{\boldsymbol{\beta}}_G^{RLTS}(\boldsymbol{z}^*) - \boldsymbol{C}(\boldsymbol{z}^*)^{-1}\boldsymbol{R}^\top\big(\boldsymbol{R}\boldsymbol{C}(\boldsymbol{z}^*)^{-1}\boldsymbol{R}^\top\big)^{-1}\big(\boldsymbol{R}\hat{\boldsymbol{\beta}}_G^{RLTS}(\boldsymbol{z}^*) - \boldsymbol{r}\big), \tag{4.4}$$

where $\boldsymbol{C}(\boldsymbol{z}^*) = \widetilde{\boldsymbol{X}}^\top \boldsymbol{V}^{-1/2}\boldsymbol{Z}^*\boldsymbol{V}^{-1/2}\widetilde{\boldsymbol{X}}$ and $\hat{\boldsymbol{\beta}}_G^{RLTS}(\boldsymbol{z}^*) = \boldsymbol{C}(\boldsymbol{z}^*)^{-1}\widetilde{\boldsymbol{X}}^\top \boldsymbol{V}^{-1/2}\boldsymbol{Z}^*\boldsymbol{V}^{-1/2}\widetilde{\boldsymbol{y}}$.

Here, we propose an extension of the RLTS problem in RSRM, called ERLTS, as follows:

$$\begin{aligned}
\min_{\boldsymbol{\beta},\boldsymbol{z}^{**}} \quad & \psi(\boldsymbol{\beta},\boldsymbol{z}^{**}) = (\widetilde{\boldsymbol{y}} - \widetilde{\boldsymbol{X}}\boldsymbol{\beta})^\top \boldsymbol{V}^{-1/2}\boldsymbol{Z}^{**}\boldsymbol{V}^{-1/2}(\widetilde{\boldsymbol{y}} - \widetilde{\boldsymbol{X}}\boldsymbol{\beta}) \\
s.t. \quad & \boldsymbol{R}\boldsymbol{\beta} = \boldsymbol{r}, \\
& h_1 \le \boldsymbol{e}^\top \boldsymbol{z}^{**} \le h_2, \\
& 0 \le z_i^{**} \le 1,\ i = 1,\ldots,n,
\end{aligned} \tag{4.5}$$

where $\boldsymbol{Z}^{**}$ is the diagonal matrix with diagonal elements $\boldsymbol{z}^{**} = (z_1^{**},\ldots,z_n^{**})^\top$, h_1 and h_2 are positive integers such that $h_1 \le h \le h_2$. The solution of this optimization problem is called the generalized extended relaxed least trimmed squares restricted estimator (GERLTSRE) of the regression coefficients $\boldsymbol{\beta}$, obtained in the RSRM as

$$\hat{\boldsymbol{\beta}}_{G\boldsymbol{R}}^{ERLTS}(\boldsymbol{z}^{**}) = \hat{\boldsymbol{\beta}}_G^{ERLTS}(\boldsymbol{z}^{**}) - \boldsymbol{C}(\boldsymbol{z}^{**})^{-1}\boldsymbol{R}^\top\big(\boldsymbol{R}\boldsymbol{C}(\boldsymbol{z}^{**})^{-1}\boldsymbol{R}^\top\big)^{-1}\big(\boldsymbol{R}\hat{\boldsymbol{\beta}}_G^{ERLTS}(\boldsymbol{z}^{**}) - \boldsymbol{r}\big), \tag{4.6}$$

where $\boldsymbol{C}(\boldsymbol{z}^{**}) = \widetilde{\boldsymbol{X}}^\top \boldsymbol{V}^{-1/2}\boldsymbol{Z}^{**}\boldsymbol{V}^{-1/2}\widetilde{\boldsymbol{X}}$ and $\hat{\boldsymbol{\beta}}_G^{ERLTS}(\boldsymbol{z}^{**}) = \boldsymbol{C}(\boldsymbol{z}^{**})^{-1}\widetilde{\boldsymbol{X}}^\top \boldsymbol{V}^{-1/2}\boldsymbol{Z}^{**}\boldsymbol{V}^{-1/2}\widetilde{\boldsymbol{y}}$.

In order to theoretically justify the suggested extension of RLTS given by ERLTS, let $\mathcal{F}^*$ and $\mathcal{F}^{**}$ respectively denote the regions of RLTS and ERLTS. Since $h \in [h_1, h_2]$, we have $\mathcal{F}^* \subseteq \mathcal{F}^{**}$. Hence, the optimal solution of the ERLTS problem is at least as good as that of the RLTS problem. Also, in real applications the exact number of suitable points to be used in a regression may not be determined. So, to make the problem more realistic, it is proper to relax optimization problem (4.3) in the sense of optimization problem (4.5) to allow uncertainty in the input data.

Theorem 4.1. *At most $p + q + 2$ decision variables z_i^{**}, $i = 1,\ldots,n$, are noninteger in the optimal solution of the extended relaxed problem.*

Proof. Consider the following form of the ERLTS problem:

$$\begin{aligned}
\min_{\boldsymbol{\beta},\boldsymbol{z}^{**}} \quad & \psi(\boldsymbol{\beta},\boldsymbol{z}^{**}) = (\widetilde{\boldsymbol{y}} - \widetilde{\boldsymbol{X}}\boldsymbol{\beta})^\top \boldsymbol{V}^{-1/2}\boldsymbol{Z}^{**}\boldsymbol{V}^{-1/2}(\widetilde{\boldsymbol{y}} - \widetilde{\boldsymbol{X}}\boldsymbol{\beta}) - \boldsymbol{\lambda}^\top(\boldsymbol{R}\boldsymbol{\beta} - \boldsymbol{r}) \\
s.t. \quad & \boldsymbol{z}^{**} \geq 0, \\
& -\boldsymbol{z}^{**} + 1 \geq 0, \\
& \boldsymbol{e}^\top \boldsymbol{z}^{**} - h_1 \geq 0, \\
& -\boldsymbol{e}^\top \boldsymbol{z}^{**} + h_2 \geq 0,
\end{aligned}$$

where $\boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_q)^\top \in \mathbb{R}^q$. Let $\hat{\boldsymbol{z}}^{**}$ be the optimal solution of the ERLTS problem. Thus, from the first-order optimality conditions (see chapter 8 of Sun and Yuan (2006)), there exist Lagrange multipliers $\boldsymbol{\alpha}, \boldsymbol{\xi} \in \mathbb{R}^n$ and $\gamma_1, \gamma_2 \in \mathbb{R}$ such that

$$\begin{aligned}
\frac{\partial\psi\big(\boldsymbol{\beta}(\hat{\boldsymbol{z}}^{**}), \hat{\boldsymbol{z}}^{**}\big)}{\partial z_i} &= \alpha_i - \xi_i + \gamma_1 - \gamma_2 - \sum_{j=1}^{p} \frac{\partial\boldsymbol{\beta}(\hat{\boldsymbol{z}}^{**})}{\partial z_i}\boldsymbol{\lambda}^\top \boldsymbol{R}_j, \; i = 1, \ldots, n, && (4.7) \\
\alpha_i \hat{z}_i^{**} &= 0, \; i = 1, \ldots, n, && (4.8) \\
\xi_i(-\hat{z}_i^{**} + 1) &= 0, \; i = 1, \ldots, n, && (4.9) \\
\gamma_1(\boldsymbol{e}^\top \hat{\boldsymbol{z}}^{**} - h_1) &= 0, \\
\gamma_2(-\boldsymbol{e}^\top \hat{\boldsymbol{z}}^{**} + h_2) &= 0, \\
\boldsymbol{\alpha}, \boldsymbol{\xi}, \gamma_1, \gamma_2 &\geq 0,
\end{aligned}$$

where $\boldsymbol{R}_j, \; j = 1, \ldots, p$, is the jth column of $\boldsymbol{R}$

Now, define $\mathcal{G} = \{i : \; \hat{z}_i^{**} \in (0, 1)\}$. Hence, for all $i \in \mathcal{G}$ from the complementary slackness conditions (4.8) and (4.9) we have $\alpha_i = \xi_i = 0$, and consequently from (4.7) we get

$$\frac{\partial\psi\big(\boldsymbol{\beta}(\hat{\boldsymbol{z}}^{**}), \hat{\boldsymbol{z}}^{**}\big)}{\partial z_i} = \gamma_1 - \gamma_2 - \sum_{j=1}^{p} \frac{\partial\boldsymbol{\beta}(\hat{\boldsymbol{z}}^{**})}{\partial z_i}\boldsymbol{\lambda}^\top \boldsymbol{R}_j, \; \forall i \in \mathcal{G}. \tag{4.10}$$

Let $u_i \in \mathbb{R}^n$, $i = 1, \ldots, n$, be the ith column of the matrix $\boldsymbol{V}^{-1/2}$. Since $\boldsymbol{V}^{-1/2}$ is a symmetric matrix, $u_i^\top$ is its ith row. Hence,

$$\begin{aligned}
\frac{\partial\psi\big(\boldsymbol{\beta}(\hat{\boldsymbol{z}}^{**}), \hat{\boldsymbol{z}}^{**}\big)}{\partial z_i} = & \left(u_i^\top \left(\widetilde{\boldsymbol{y}} - \widetilde{\boldsymbol{X}}\boldsymbol{\beta}(\hat{\boldsymbol{z}}^{**})\right)\right)^2 \\
& -2\frac{\partial\boldsymbol{\beta}^\top(\hat{\boldsymbol{z}}^{**})}{\partial z_i}\widetilde{\boldsymbol{X}}^\top \boldsymbol{V}^{-1/2}\boldsymbol{Z}^{**}\boldsymbol{V}^{-1/2}\big(\widetilde{\boldsymbol{y}} - \widetilde{\boldsymbol{X}}\boldsymbol{\beta}(\hat{\boldsymbol{z}}^{**})\big). \qquad (4.11)
\end{aligned}$$

Therefore, since

$$\widetilde{\boldsymbol{X}}^\top \boldsymbol{V}^{-1/2}\boldsymbol{Z}^{**}\boldsymbol{V}^{-1/2}\big(\widetilde{\boldsymbol{y}} - \widetilde{\boldsymbol{X}}\boldsymbol{\beta}(\hat{\boldsymbol{z}}^{**})\big) = 0,$$

from (4.10) and (4.11) we have

$$\left(u_i^\top \left(\widetilde{\boldsymbol{y}} - \widetilde{\boldsymbol{X}}\boldsymbol{\beta}(\hat{\boldsymbol{z}}^{**})\right)\right)^2 = \gamma_1 - \gamma_2 - \sum_{j=1}^{p} \frac{\partial\boldsymbol{\beta}(\hat{\boldsymbol{z}}^{**})}{\partial z_i}\boldsymbol{\lambda}^\top \boldsymbol{R}_j, \; \forall i \in \mathcal{G}. \tag{4.12}$$

By a small randomization on the difference-based data $(\widetilde{\boldsymbol{X}}, \widetilde{\boldsymbol{y}})$, since the system of equations (4.12) has $p+q+2$ degrees of freedom on $\boldsymbol{\beta}$, γ_1, γ_2 and $\boldsymbol{\lambda}$, it can be ensured that the system holds for no more than $p+q+2$ observations $(\widetilde{\boldsymbol{x}}_i, \widetilde{y}_i)$. That is, the cardinality of the set $\mathcal{G}$ is at most $p+q+2$. Hence, the proof is complete. $\square$

5 Ridge Estimator based on Robust Approaches

We estimate the ridge parameter k and $\boldsymbol{\beta}$ by using robust methods which were introduced in last section. Indeed, we propose robust estimators for ridge parameter by substituting σ^2 and $\boldsymbol{\beta}$ by their robust estimators in (2.8). After substituting these estimators into (2.8), we estimate the ridge parameter and then, we obtain two stage estimators for $\boldsymbol{\beta}$ as follows:

1. Estimators for k and $\boldsymbol{\beta}$ based on LTS method in RSRM (thereafter GLTSRR method)

$$\hat{k}_{LTS} = \frac{p\hat{\sigma}^2_{LTS}}{\hat{\boldsymbol{\beta}}^{LTS}_{\boldsymbol{GR}}(\boldsymbol{z})^\top \hat{\boldsymbol{\beta}}^{LTS}_{\boldsymbol{GR}}(\boldsymbol{z})}, \ \hat{\sigma}^2_{LTS} = \frac{1}{n-(p+q)}\big(\widetilde{\boldsymbol{y}} - \widetilde{\boldsymbol{X}}\hat{\boldsymbol{\beta}}^{LTS}_{\boldsymbol{GR}}(\boldsymbol{z})\big)^\top \boldsymbol{V}^{-1/2}\boldsymbol{Z}\boldsymbol{V}^{-1/2}\big(\widetilde{\boldsymbol{y}} - \widetilde{\boldsymbol{X}}\hat{\boldsymbol{\beta}}^{LTS}_{\boldsymbol{GR}}(\boldsymbol{z})\big), \quad (5.1)$$

$$\hat{\boldsymbol{\beta}}^{LTS}_{\boldsymbol{GR}}(\hat{k}_{LTS}, \boldsymbol{z}) = \hat{\boldsymbol{\beta}}^{LTS}_{G}(\hat{k}_{LTS}, \boldsymbol{z}) - \boldsymbol{C}(\hat{k}_{LTS}, \boldsymbol{z})^{-1}\boldsymbol{R}^\top\big(\boldsymbol{R}\boldsymbol{C}(\hat{k}_{LTS}, \boldsymbol{z})^{-1}\boldsymbol{R}^\top\big)^{-1}\big(\boldsymbol{R}\hat{\boldsymbol{\beta}}^{LTS}_{G}(\hat{k}_{LTS}, \boldsymbol{z}) - \boldsymbol{r}\big), \quad (5.2)$$

where $\boldsymbol{C}(\hat{k}_{LTS}, \boldsymbol{z}) = \boldsymbol{C}(\boldsymbol{z}) + \hat{k}_{LTS}\boldsymbol{I}$ and $\hat{\boldsymbol{\beta}}^{LTS}_{G}(\hat{k}_{LTS}, \boldsymbol{z}) = \boldsymbol{C}(\hat{k}_{LTS}, \boldsymbol{z})^{-1}\widetilde{\boldsymbol{X}}^\top\boldsymbol{V}^{-1/2}\boldsymbol{Z}\boldsymbol{V}^{-1/2}\widetilde{\boldsymbol{y}}$.

2. Estimators for k and $\boldsymbol{\beta}$ based on RLTS method in RSRM (thereafter GRLTSRR method)

$$\hat{k}_{RLTS} = \frac{p\hat{\sigma}^2_{RLTS}}{\hat{\boldsymbol{\beta}}^{RLTS}_{\boldsymbol{GR}}(\boldsymbol{z}^*)^\top \hat{\boldsymbol{\beta}}^{RLTS}_{\boldsymbol{GR}}(\boldsymbol{z}^*)}, \ \hat{\sigma}^2_{RLTS} = \frac{1}{n-(p+q)}\big(\widetilde{\boldsymbol{y}} - \widetilde{\boldsymbol{X}}\hat{\boldsymbol{\beta}}^{RLTS}_{\boldsymbol{GR}}(\boldsymbol{z}^*)\big)^\top \boldsymbol{V}^{-1/2}\boldsymbol{Z}^*\boldsymbol{V}^{-1/2}\big(\widetilde{\boldsymbol{y}} - \widetilde{\boldsymbol{X}}\hat{\boldsymbol{\beta}}^{RLTS}_{\boldsymbol{GR}}(\boldsymbol{z}^*)\big), (5.3)$$

$$\hat{\boldsymbol{\beta}}^{RLTS}_{\boldsymbol{GR}}(\hat{k}_{RLTS}, \boldsymbol{z}^*) = \hat{\boldsymbol{\beta}}^{RLTS}_{G}(\hat{k}_{RLTS}, \boldsymbol{z}^*) - \boldsymbol{C}(\hat{k}_{RLTS}, \boldsymbol{z}^*)^{-1}\boldsymbol{R}^\top\big(\boldsymbol{R}\boldsymbol{C}(\hat{k}_{RLTS}, \boldsymbol{z}^*)^{-1}\boldsymbol{R}^\top\big)^{-1}\big(\boldsymbol{R}\hat{\boldsymbol{\beta}}^{RLTS}_{G}(\hat{k}_{RLTS}, \boldsymbol{z}^*) - \boldsymbol{r}\big), (5.4)$$

where $\boldsymbol{C}(\hat{k}_{RLTS}, \boldsymbol{z}^*) = \boldsymbol{C}(\boldsymbol{z}^*) + \hat{k}_{RLTS}\boldsymbol{I}$ and $\hat{\boldsymbol{\beta}}^{RLTS}_{G}(\hat{k}_{RLTS}, \boldsymbol{z}^*) = \boldsymbol{C}(\hat{k}_{RLTS}, \boldsymbol{z}^*)^{-1}\widetilde{\boldsymbol{X}}^\top\boldsymbol{V}^{-1/2}\boldsymbol{Z}^*\boldsymbol{V}^{-1/2}\widetilde{\boldsymbol{y}}$,

3. Estimators for k and $\boldsymbol{\beta}$ based on ERLTS method in RSRM (thereafter GERLTSRR method)

$$\begin{aligned}\hat{k}_{ERLTS} &= \frac{p\hat{\sigma}^2_{ERLTS}}{\hat{\boldsymbol{\beta}}^{ERLTS}_{\boldsymbol{GR}}(\boldsymbol{z}^{**})^\top \hat{\boldsymbol{\beta}}^{ERLTS}_{\boldsymbol{GR}}(\boldsymbol{z}^{**})}, \qquad (5.5)\\ \hat{\sigma}^2_{ERLTS} &= \frac{1}{n-(p+q)}\big(\widetilde{\boldsymbol{y}} - \widetilde{\boldsymbol{X}}\hat{\boldsymbol{\beta}}^{ERLTS}_{\boldsymbol{GR}}(\boldsymbol{z})\big)^\top \boldsymbol{V}^{-1/2}\boldsymbol{Z}^{**}\boldsymbol{V}^{-1/2}\big(\widetilde{\boldsymbol{y}} - \widetilde{\boldsymbol{X}}\hat{\boldsymbol{\beta}}^{ERLTS}_{\boldsymbol{GR}}(\boldsymbol{z})\big),\end{aligned}$$

$$\begin{aligned}\hat{\boldsymbol{\beta}}^{ERLTS}_{\boldsymbol{GR}}(\hat{k}_{ERLTS}, \boldsymbol{z}^{**}) &= \hat{\boldsymbol{\beta}}^{ERLTS}_{G}(\hat{k}_{ERLTS}, \boldsymbol{z}^{**})\\ &\quad - \boldsymbol{C}(\hat{k}_{ERLTS}, \boldsymbol{z}^{**})^{-1}\boldsymbol{R}^\top\big(\boldsymbol{R}\boldsymbol{C}(\hat{k}_{ERLTS}, \boldsymbol{z}^{**})^{-1}\boldsymbol{R}^\top\big)^{-1}\big(\boldsymbol{R}\hat{\boldsymbol{\beta}}^{ERLTS}_{G}(\hat{k}_{ERLTS}, \boldsymbol{z}^{**}) - \boldsymbol{r}\big), (5.6)\end{aligned}$$

where $\boldsymbol{C}(\hat{k}_{ERLTS}, \boldsymbol{z}^{**}) = \boldsymbol{C}(\boldsymbol{z}^{**}) + \hat{k}_{ERLTS}\boldsymbol{I}$ and $\hat{\boldsymbol{\beta}}_G^{ERLTS}(\hat{k}_{ERLTS}, \boldsymbol{z}^{**}) = \boldsymbol{C}(\hat{k}_{ERLTS}, \boldsymbol{z}^{**})^{-1}\widetilde{\boldsymbol{X}}^\top \boldsymbol{V}^{-1/2}\boldsymbol{Z}^{**}\boldsymbol{V}^{-1/2}\widetilde{\boldsymbol{y}}$.

To evaluate the performance of risk function we use different values of ridge parameter k to evaluate the proposed estimators. We compare the nonridge and ridge estimators based on LS, LTS, RLTS and ERLTS methods. Since, theoretically, these estimators are very difficult to compare, the Monté-Carlo simulation studies have been conducted to compare the efficiency of the estimators as well as real data example in the following section.

6 Numerical Results

In this section we proceed with some numerical computations as proofs of our assertions. First we consider the Monté-Carlo simulation schemes to evaluate the performance of the proposed ridge estimators and then, we consider a real data example.

6.1 The Monté-Carlo Simulation

Here, we numerically examine the accuracy of our robust estimators for RSRM with contaminated data. The regressors are drawn anew in every replication. The efficiencies of $\hat{\boldsymbol{\beta}}_2$ relative to $\hat{\boldsymbol{\beta}}_1$ are defined based on the Euclidean norm by

$$\mathrm{Eff}(\hat{\boldsymbol{\beta}}_2, \hat{\boldsymbol{\beta}}_1) = \frac{\hat{\mathrm{R}}(\hat{\boldsymbol{\beta}}_1, \boldsymbol{\beta})}{\hat{\mathrm{R}}(\hat{\boldsymbol{\beta}}_2, \boldsymbol{\beta})} = \frac{\frac{1}{\mathrm{M}}\sum_{\mathrm{m}=1}^{\mathrm{M}} \left\|\hat{\boldsymbol{\beta}}_1^{(\mathrm{m})} - \boldsymbol{\beta}\right\|_2^2}{\frac{1}{\mathrm{M}}\sum_{\mathrm{m}=1}^{\mathrm{M}} \left\|\hat{\boldsymbol{\beta}}_2(\mathrm{m}) - \boldsymbol{\beta}\right\|_2^2}, \tag{6.1}$$

where M is the number of iterations, $\hat{\boldsymbol{\beta}}_i^{(m)}$ is the ith estimator of $\boldsymbol{\beta}$ in the mth iteration and $\|\boldsymbol{x}\|_2$ denotes the Euclidean norm of vector $\boldsymbol{x}$. To achieve different degrees of collinearity, following McDonald and Galarneau (1975) and Gibbons (1981) the explanatory variables were generated using the following device for $n = 150$ with 10^3 iteration from the following model:

$$x_{ij} = (1-\gamma^2)^{\frac{1}{2}} z_{ij} + \gamma z_{ip}, \ i = 1, \ldots, n, \ j = 1, \ldots, p, \tag{6.2}$$

where z_{ij} are independent standard normal pseudo-random numbers, and γ is specified so that the correlation between any two explanatory variables is given by γ^2. These variables are then standardized so that $\boldsymbol{X}^\top\boldsymbol{X}$ and $\boldsymbol{X}^\top\boldsymbol{y}$ are in correlation forms. Three different sets of correlation corresponding to $\gamma = 0.90, 0.95$ and 0.99 are considered. Then n observations for the dependent variable are determined by

$$y_i = \sum_{j=1}^{6} x_{ij}\beta_j + f(t_i) + \epsilon_i, \ i = 1, \ldots, n, \tag{6.3}$$

where

$$\boldsymbol{\beta} = (-1, 4, 2, -5, -3)^\top,$$

$$f(t) = \exp\left\{\sin(2t)\cos(5t) + \sqrt{t}\right\},$$

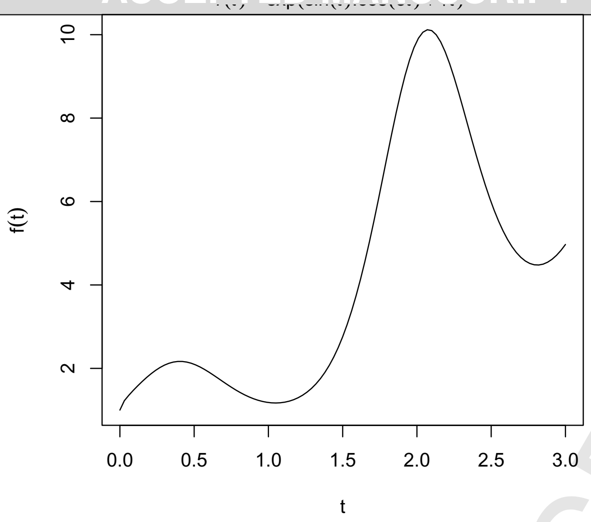

Figure 1: The nonparametric function.

for $t \in [0,3]$ and $\boldsymbol{\epsilon} = (\boldsymbol{\epsilon}_1^\top, \boldsymbol{\epsilon}_2^\top)^\top$ where

$$\boldsymbol{\epsilon}_{1\ (h\times 1)} \sim \mathcal{N}_h(\mathbf{0}, \sigma^2 \boldsymbol{V}),\ \sigma^2 = 1.44,\ v_{ij} = \exp\big(-9|i-j|\big),$$
$$\boldsymbol{\epsilon}_{2\ ((n-h)\times 1)} \overset{i.i.d.}{\sim} \chi_1^2(15),$$

where $\chi_\nu^2(\delta)$ is the noncentral Chi-squared distribution with ν degrees of freedom and non centrality parameter δ. The main reason of selecting such structure for errors is to contaminate the data to check the efficiency of the robust estimators. We set the first h error terms as dependent normal random variables and the last $(n-h)$ error terms as independent noncentral Chi-squared random variables. The non centrality causes the outliers lie on one side of the true regression model while pull the nonrobust estimation toward themselves.

For estimating the nonparametric part of the model, $f(.)$, we use

$$W_{ni}(t_j) = \frac{1}{nh_n} K\Big(\frac{t_i - tj}{h_n}\Big) = \frac{1}{nh_n}.\frac{1}{\sqrt{2\pi}} \exp\Big\{-\frac{(t_i - t_j)^2}{2h_n^2}\Big\},$$

which is Priestley and Chao's weight with the Gaussian kernel. We also apply the cross-validation (C.V.) method to select the optimal bandwidth h_n, which minimizes the following C.V. function

$$\text{C.V.}(h_n) = \frac{1}{n} \sum_{i=1}^{n} \Big(\widetilde{\boldsymbol{y}}^{(-i)} - \widetilde{\boldsymbol{X}}^{(-i)} \hat{\boldsymbol{\beta}}^{(-i)}\Big)^2,$$

where $\hat{\boldsymbol{\beta}}^{(-i)}$ is obtained by replacing $\widetilde{\boldsymbol{X}}$ and $\widetilde{\boldsymbol{y}}$ with $\widetilde{\boldsymbol{X}}^{(-i)} = \Big(\tilde{x}_{jk}^{(-i)}\Big)$, $\tilde{x}_{sk}^{(-i)} = x_{sk} - \sum_{j\neq i}^{n} W_{nj}(t_i) x_{sj}$, $1 \leq k \leq n$, $1 \leq j \leq p$, $\boldsymbol{y}^{(-i)} = \Big(\tilde{y}_1^{(-i)}, \ldots, \tilde{y}_n^{(-i)}\Big)$, $\tilde{y}_k^{(-i)} = y_k - \sum_{j\neq i}^{n} W_{nj}(t_i) y_j$. Here $\boldsymbol{y}^{(-i)}$ is the predicted value of $\boldsymbol{y} = (y_1, \ldots, y_n)$ at $\boldsymbol{x}_i = (x_{1i}, \ldots, x_{pi})$ with y_i and x_i left out of the estimation of the $\boldsymbol{\beta}$.

In Figure 1, the nonparametric part of the model (6.3) is plotted. This function is difficult to be estimated and provides a good test case for the nonparametric regression method. All computations were conducted using the statistical package R 3.1.2. In Tables 1 to 9, we computed the proposed estimators, respectively. We numerically estimated the R(.), efficiencies of robust estimators relative to nonrobust estimators for ridge types

Table 1: Evaluation of parameters for proposed estimators with $\gamma = 0.9$ and $\alpha = 25\%$ breakdown

Method / Coefficients	GLSRE	GLTSRE	GRLTSRE	GERLTSRE	GLSRRE	GLTSRRE	GRLTSRRE	GERLTSRRE
$\hat{\beta}_1$	-1.0053	-1.0029	-1.0044	-1.0029	-1.0015	-1.0018	-1.0023	-1.0014
$\hat{\beta}_2$	3.8837	3.9351	3.9033	3.9372	3.9670	3.9600	3.9499	3.9701
$\hat{\beta}_3$	1.6510	1.8054	1.7099	1.8116	1.9010	1.8799	1.8496	1.9103
$\hat{\beta}_4$	-4.6986	-4.8320	-4.7495	-4.8372	-4.9145	-4.8962	-4.8701	-4.9225
$\hat{\beta}_5$	-2.8414	-2.9116	-2.8681	-2.9143	-2.9550	-2.9454	-2.9316	-2.9592
$\hat{k}$	0.0000	0.0000	0.0000	0.0000	0.7168	0.2200	0.4409	0.2456
$\hat{\mathrm{R}}(\hat{\boldsymbol{\beta}}, \boldsymbol{\beta})$	1.0338	0.4477	0.7035	0.4265	0.7756	0.3985	0.5824	0.3804
$\mathrm{Eff}(\hat{\boldsymbol{\beta}}(\hat{\mathrm{k}}), \hat{\boldsymbol{\beta}})$	1.0000	2.3091	1.4695	2.4238	1.0000	1.9465	1.3316	2.0388
$\hat{\mathrm{mse}}(\hat{f}(\boldsymbol{t}), f(\boldsymbol{t}))$	0.5139	0.3384	0.4055	0.3293	0.4457	0.3255	0.3739	0.3180

Table 2: Evaluation of parameters for proposed estimators with $\gamma = 0.95$ and $\alpha = 25\%$ breakdown

Method / Coefficients	GLSRE	GLTSRE	GRLTSRE	GERLTSRE	GLSRRE	GLTSRRE	GRLTSRRE	GERLTSRRE
$\hat{\beta}_1$	-1.0120	-1.0079	-1.0087	-1.0069	-1.0051	-1.0056	-1.0045	-1.0041
$\hat{\beta}_2$	3.7358	3.8258	3.8097	3.8488	3.8881	3.8760	3.9010	3.9096
$\hat{\beta}_3$	1.2074	1.4774	1.4290	1.5464	1.6643	1.6279	1.7029	1.7289
$\hat{\beta}_4$	-4.3155	-4.5486	-4.5069	-4.6083	-4.7101	-4.6786	-4.7434	-4.7659
$\hat{\beta}_5$	-2.6397	-2.7624	-2.7405	-2.7938	-2.8474	-2.8309	-2.8650	-2.8768
$\hat{k}$	0.0000	0.0000	0.0000	0.0000	0.6386	0.2053	0.3493	0.2131
$\hat{\mathrm{R}}(\hat{\boldsymbol{\beta}}, \boldsymbol{\beta})$	2.7769	1.1724	1.6454	1.1101	1.7160	0.8745	1.0845	0.8476
$\mathrm{Eff}(\hat{\boldsymbol{\beta}}(\hat{\mathrm{k}}), \hat{\boldsymbol{\beta}})$	1.0000	2.3685	1.6876	2.5016	1.0000	1.9622	1.5823	2.0246
$\hat{\mathrm{mse}}(\hat{f}(\boldsymbol{t}), f(\boldsymbol{t}))$	0.5522	0.3750	0.4160	0.3635	0.4150	0.3364	0.3472	0.3266

and nonridge types, separately, and $\hat{\mathrm{mse}}(\hat{f}(\boldsymbol{t}), f(\boldsymbol{t})) = \frac{1}{hM}\sum_{i=1}^{M} \|\hat{f}(\boldsymbol{t}) - f(\boldsymbol{t})\|_2^2$ for proposed estimators, which heavily depends on $\hat{k}, \gamma$ and α. The Figure 2 shows the fitted function by kernel smoothing after estimating the linear part of the model by proposed estimators for $\alpha = 25\%$ and $\gamma = 0.90,\ 0.95,\ 0.99$.

Table 3: Evaluation of parameters for proposed estimators with $\gamma = 0.99$ and $\alpha = 25\%$ breakdown

Method / Coefficients	GLSRE	GLTSRE	GRLTSRE	GERLTSRE	GLSRRE	GLTSRRE	GRLTSRRE	GERLTSRRE
$\hat{\beta}_1$	-1.0305	-1.0158	-1.0212	-1.0166	-1.0075	-1.0068	-1.0063	-1.0050
$\hat{\beta}_2$	3.3279	3.6535	3.5330	3.6340	3.8353	3.8510	3.8624	3.8896
$\hat{\beta}_3$	-0.0163	0.9604	0.5989	0.9021	1.5060	1.5529	1.5871	1.6688
$\hat{\beta}_4$	-3.2587	-4.1021	-3.7899	-4.0518	-4.5733	-4.6139	-4.6434	-4.7140
$\hat{\beta}_5$	-2.0835	-2.5274	-2.3631	-2.5009	-2.7754	-2.7968	-2.8123	-2.8495
$\hat{k}$	0.0000	0.0000	0.0000	0.0000	4.5216	0.5244	1.9424	0.8638
$\hat{\mathrm{R}}(\hat{\boldsymbol{\beta}}, \boldsymbol{\beta})$	11.3412	5.1621	7.2416	4.4436	4.6362	3.3157	4.2489	2.2428
$\mathrm{Eff}(\hat{\boldsymbol{\beta}}(\hat{\mathrm{k}}), \hat{\boldsymbol{\beta}})$	1.0000	2.1970	1.5661	2.5522	1.0000	1.3983	1.0912	2.0672
$\widehat{\mathrm{mse}}(\hat{f}(\boldsymbol{t}), f(\boldsymbol{t}))$	0.3575	0.2759	0.2988	0.2626	0.2829	0.2520	0.2632	0.2351

Table 4: Evaluation of parameters for proposed estimators with $\gamma = 0.9$ and $\alpha = 33\%$ breakdown

Method / Coefficients	GLSRE	GLTSRE	GRLTSRE	GERLTSRE	GLSRRE	GLTSRRE	GRLTSRRE	GERLTSRRE
$\hat{\beta}_1$	-1.0064	-1.0031	-1.0059	-1.0043	-1.0016	-1.0012	-1.0035	-1.0027
$\hat{\beta}_2$	3.8588	3.9318	3.8697	3.9044	3.9638	3.9741	3.9237	3.9414
$\hat{\beta}_3$	1.5764	1.7953	1.6091	1.7132	1.8913	1.9223	1.7711	1.8243
$\hat{\beta}_4$	-4.6342	-4.8232	-4.6624	-4.7523	-4.9061	-4.9329	-4.8023	-4.8483
$\hat{\beta}_5$	-2.8075	-2.9069	-2.8223	-2.8696	-2.9506	-2.9647	-2.8959	-2.9201
$\hat{k}$	0.0000	0.0000	0.0000	0.0000	1.4631	0.4550	0.4852	0.3879
$\hat{\mathrm{R}}(\hat{\boldsymbol{\beta}}, \boldsymbol{\beta})$	1.1776	0.4522	0.7872	0.4351	0.8152	0.3671	0.5453	0.3161
$\mathrm{Eff}(\hat{\boldsymbol{\beta}}(\hat{\mathrm{k}}), \hat{\boldsymbol{\beta}})$	1.0000	2.6042	1.4959	2.7067	1.0000	2.2204	1.4950	2.5787
$\widehat{\mathrm{mse}}(\hat{f}(\boldsymbol{t}), f(\boldsymbol{t}))$	0.5562	0.3269	0.4411	0.3347	0.4403	0.2959	0.3677	0.2992

Table 5: Evaluation of parameters for proposed estimators with $\gamma = 0.95$ and $\alpha = 33\%$ breakdown

Method / Coefficients	GLSRE	GLTSRE	GRLTSRE	GERLTSRE	GLSRRE	GLTSRRE	GRLTSRRE	GERLTSRRE
$\hat{\beta}_1$	-1.0119	-1.0059	-1.0063	-1.0054	-1.0034	-1.0025	-1.0018	-1.0024
$\hat{\beta}_2$	3.7377	3.8698	3.8619	3.8815	3.9262	3.9440	3.9596	3.9463
$\hat{\beta}_3$	1.2131	1.6093	1.5857	1.6445	1.7786	1.8320	1.8787	1.8388
$\hat{\beta}_4$	-4.3204	-4.6626	-4.6422	-4.6930	-4.8087	-4.8549	-4.8952	-4.8608
$\hat{\beta}_5$	-2.6423	-2.8224	-2.8117	-2.8384	-2.8993	-2.9236	-2.9448	-2.9267
$\hat{k}$	0.0000	0.0000	0.0000	0.0000	1.6417	0.4172	0.4864	0.3774
$\hat{\mathrm{R}}(\hat{\boldsymbol{\beta}}, \boldsymbol{\beta})$	2.3571	1.2820	1.3432	0.9283	1.1347	0.9903	0.9968	0.7044
$\mathrm{Eff}(\hat{\boldsymbol{\beta}}(\hat{\mathrm{k}}), \hat{\boldsymbol{\beta}})$	1.0000	1.8386	1.7549	2.5391	1.0000	1.1458	1.1383	1.6108
$\widehat{\mathrm{mse}}(\hat{f}(\boldsymbol{t}), f(\boldsymbol{t}))$	0.4729	0.3523	0.3552	0.3051	0.3318	0.3139	0.3181	0.2780

Table 6: Evaluation of parameters for proposed estimators with $\gamma = 0.99$ and $\alpha = 33\%$ breakdown

Method / Coefficients	GLSRE	GLTSRE	GRLTSRE	GERLTSRE	GLSRRE	GLTSRRE	GRLTSRRE	GERLTSRRE
$\hat{\beta}_1$	-1.0284	-1.0189	-1.0180	-1.0138	-1.0061	-1.0064	-1.0029	-1.0021
$\hat{\beta}_2$	3.3749	3.5838	3.6045	3.6968	3.8664	3.8588	3.9364	3.9548
$\hat{\beta}_3$	0.1247	0.7513	0.8135	1.0905	1.5991	1.5764	1.8092	1.8643
$\hat{\beta}_4$	-3.3804	-3.9216	-3.9753	-4.2145	-4.6538	-4.6341	-4.8352	-4.8828
$\hat{\beta}_5$	-2.1476	-2.4324	-2.4607	-2.5866	-2.8178	-2.8074	-2.9133	-2.9383
$\hat{k}$	0.0000	0.0000	0.0000	0.0000	4.3002	0.3844	1.2217	0.2575
$\hat{\mathrm{R}}(\hat{\boldsymbol{\beta}}, \boldsymbol{\beta})$	14.7570	8.7656	8.0113	5.2058	10.8751	6.8073	6.5814	3.8510
$\mathrm{Eff}(\hat{\boldsymbol{\beta}}(\hat{\mathrm{k}}), \hat{\boldsymbol{\beta}})$	1.0000	1.6835	1.8420	2.8347	1.0000	1.5976	1.6524	2.8240
$\hat{\mathrm{mse}}(\hat{f}(\boldsymbol{t}), f(\boldsymbol{t}))$	0.3430	0.2851	0.2685	0.2472	0.3105	0.2692	0.2572	0.2432

Table 7: Evaluation of parameters for proposed estimators with $\gamma = 0.9$ and $\alpha = 50\%$ breakdown

Method / Coefficients	GLSRE	GLTSRE	GRLTSRE	GERLTSRE	GLSRRE	GLTSRRE	GRLTSRRE	GERLTSRRE
$\hat{\beta}_1$	-1.0081	-1.0060	-1.0038	-1.0029	-1.0016	-1.0022	-1.0008	-1.0008
$\hat{\beta}_2$	3.8209	3.8682	3.9159	3.9356	3.9659	3.9516	3.9833	3.9817
$\hat{\beta}_3$	1.4628	1.6047	1.7477	1.8069	1.8977	1.8548	1.9500	1.9450
$\hat{\beta}_4$	-4.5361	-4.6586	-4.7821	-4.8332	-4.9116	-4.8746	-4.9568	-4.9525
$\hat{\beta}_5$	-2.7558	-2.8203	-2.8853	-2.9122	-2.9535	-2.9340	-2.9773	-2.9750
$\hat{k}$	0.0000	0.0000	0.0000	0.0000	2.9139	1.4506	0.5676	0.5010
$\hat{\mathrm{R}}(\hat{\boldsymbol{\beta}}, \boldsymbol{\beta})$	1.8595	1.3729	0.9263	0.8069	1.2753	1.0716	0.8122	0.7199
$\mathrm{Eff}(\hat{\boldsymbol{\beta}}(\hat{\mathrm{k}}), \hat{\boldsymbol{\beta}})$	1.0000	1.3545	2.0075	2.3046	1.0000	1.1901	1.5702	1.7715
$\hat{\mathrm{mse}}(\hat{f}(\boldsymbol{t}), f(\boldsymbol{t}))$	0.8091	0.6377	0.5045	0.4878	0.6202	0.5485	0.4643	0.4581

Table 8: Evaluation of parameters for proposed estimators with $\gamma = 0.95$ and $\alpha = 50\%$ breakdown

Method / Coefficients	GLSRE	GLTSRE	GRLTSRE	GERLTSRE	GLSRRE	GLTSRRE	GRLTSRRE	GERLTSRRE
$\hat{\beta}_1$	-1.0157	-1.0093	-1.0084	-1.0070	-1.0038	-1.0019	-1.0025	-1.0028
$\hat{\beta}_2$	3.6555	3.7950	3.8154	3.8469	3.9165	3.9587	3.9452	3.9390
$\hat{\beta}_3$	0.9665	1.3850	1.4462	1.5408	1.7496	1.8761	1.8357	1.8170
$\hat{\beta}_4$	-4.1075	-4.4688	-4.5217	-4.6034	-4.7837	-4.8930	-4.8581	-4.8420
$\hat{\beta}_5$	-2.5302	-2.7204	-2.7483	-2.7913	-2.8862	-2.9437	-2.9253	-2.9168
$\hat{k}$	0.0000	0.0000	0.0000	0.0000	1.4943	1.0051	0.8013	0.4644
$\hat{\mathrm{R}}(\hat{\boldsymbol{\beta}}, \boldsymbol{\beta})$	4.0294	2.7950	2.5772	1.8992	1.8742	1.8018	1.8950	1.4603
$\mathrm{Eff}(\hat{\boldsymbol{\beta}}(\hat{\mathrm{k}}), \hat{\boldsymbol{\beta}})$	1.0000	1.4417	1.5635	2.1217	1.0000	1.0402	0.9890	1.2834
$\hat{\mathrm{mse}}(\hat{f}(\boldsymbol{t}), f(\boldsymbol{t}))$	0.7299	0.5717	0.5258	0.4511	0.4470	0.4340	0.4328	0.3877

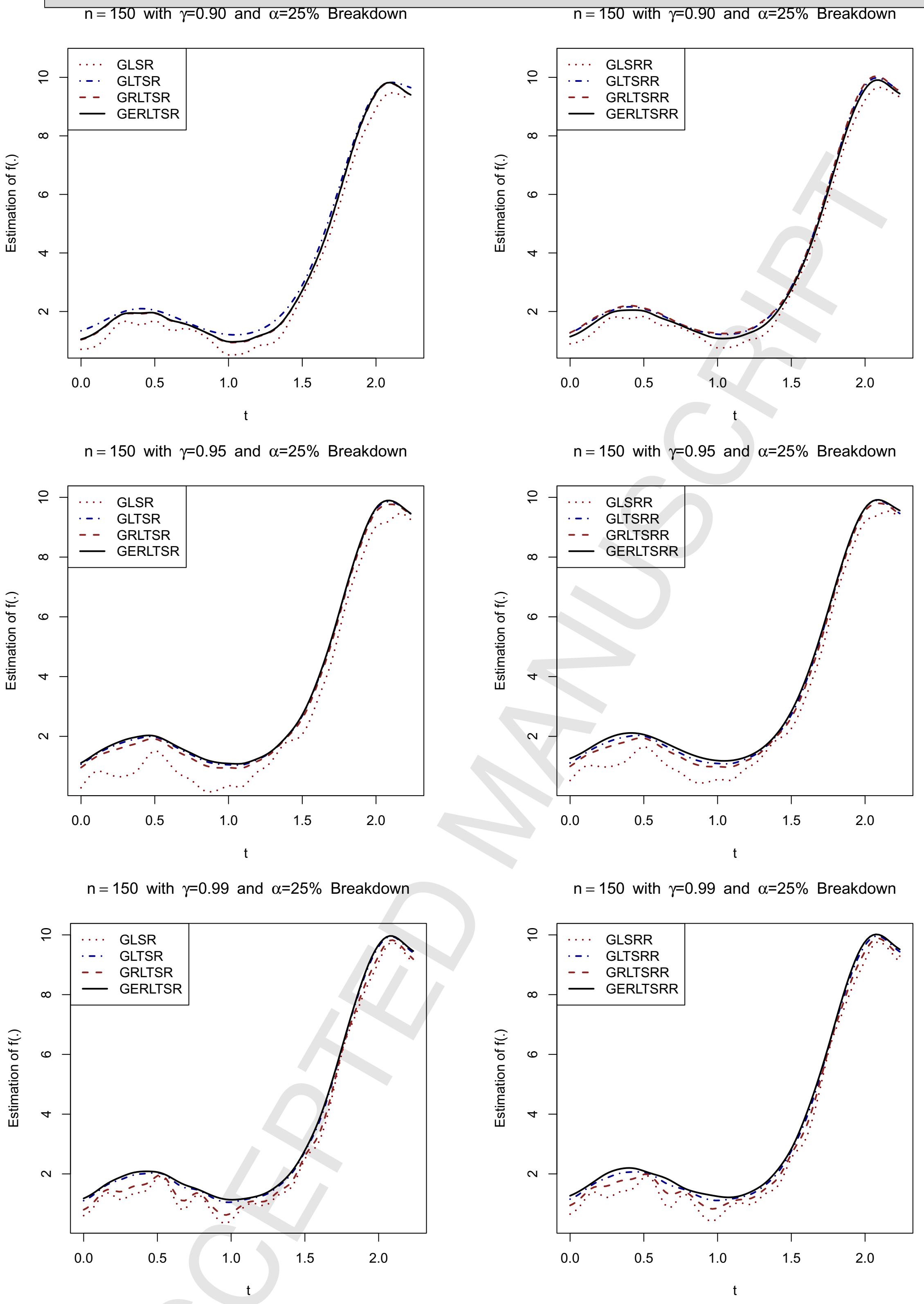

Figure 2: Estimation of the function under study by kernel approach for $n = 150$ and $\alpha = 25\%$. Non ridge and ridge estimations for $\gamma = 0.90$ (top), estimations for $\gamma = 0.95$ (middle), estimations for $\gamma = 0.99$ (bottom).

Table 9: Evaluation of parameters for proposed estimators with $\gamma = 0.99$ and $\alpha = 50\%$ breakdown

Method / Coefficients	GLSRE	GLTSRE	GRLTSRE	GERLTSRE	GLSRRE	GLTSRRE	GRLTSRRE	GERLTSRRE
$\hat{\beta}_1$	-1.0273	-1.0186	-1.0193	-1.0146	-0.9977	-0.9977	-0.9996	-0.9997
$\hat{\beta}_2$	3.4002	3.5904	3.5751	3.6792	4.0515	4.0511	4.0086	4.0061
$\hat{\beta}_3$	0.2006	0.7711	0.7253	1.0376	2.1545	2.1534	2.0257	2.0182
$\hat{\beta}_4$	-3.4459	-3.9387	-3.8991	-4.1688	-5.1334	-5.1324	-5.0222	-5.0157
$\hat{\beta}_5$	-2.1821	-2.4414	-2.4206	-2.5625	-3.0702	-3.0697	-3.0117	-3.0083
$\hat{k}$	0.0000	0.0000	0.0000	0.0000	1.4395	1.2383	0.2498	0.2802
$\hat{\mathrm{R}}(\hat{\boldsymbol{\beta}}, \boldsymbol{\beta})$	14.4475	13.1786	9.2513	6.7690	12.9781	11.5284	7.0926	4.8948
$\mathrm{Eff}(\hat{\boldsymbol{\beta}}(\hat{\mathrm{k}}), \hat{\boldsymbol{\beta}})$	1.0000	1.0963	1.5617	2.1343	1.0000	1.1257	1.8298	2.6514
$\hat{\mathrm{mse}}(\hat{f}(\boldsymbol{t}), f(\boldsymbol{t}))$	0.4031	0.3416	0.3220	0.2851	0.3534	0.3169	0.2793	0.2615

Table 10: Correlation Matrix

Variable	SP	LT	SFH	FP	DHW	GAR	ANI
SP	1.0000	0.14591	0.4774	0.3367	-0.1001	0.2995	0.3415
LT	0.1459	1.00000	0.1544	0.1595	0.0814	0.1699	0.1334
SFH	0.4774	0.15445	1.0000	0.4633	0.0229	0.2230	0.2779
FP	0.3367	0.15953	0.4633	1.0000	0.1085	0.1456	0.3814
DHW	-0.1001	0.08145	0.0229	0.1085	1.0000	0.0579	-0.1096
GAR	0.2995	0.16991	0.2230	0.1456	0.0579	1.0000	0.0270
ANI	0.3415	0.13344	0.2779	0.3814	-0.1096	0.0270	1.0000

6.2 Real Data Example

To motivate the problem of linearly constrained estimation in the semiparametric regression model, we consider the hedonic prices of housing attributes. Housing prices are very much affected by lot size. The semiparametric regression model that follows was estimated by Ho (1995) using semiparametric least squares. The data consist of 92 detached homes in the Ottawa area that were sold in 1987. The variables are defined as follows: The dependent variable y is sale price (SP), the independent variables include lot size (lot area = LT), square footage of housing (SFH), average neighborhood income (ANI), distance to highway (DHW), presence of garage (GAR) and fireplace (FP). We first consider the pure parametric model:

$$(SP)_i = \beta_0 + \beta_1(LT)_i + \beta_2(SFH)_i + \beta_3(FP)_i + \beta_4(DHW)_i + \beta_5(GAR)_i + \beta_6(ANI)_i + \epsilon_i. \tag{6.4}$$

In order to detect the correlation between variables, we can take a look at correlation matrix given by table 10. As it can be investigated, there exists a potential multicollinearity between SFH & FP and DHW & ANI. Also, the eigenvalues of $\boldsymbol{X}^\top \boldsymbol{X}$ are as $\lambda_7 = 2.385563e+05$, $\lambda_6 = 2.302869e+02$, $\lambda_5 = 2.390021e+01$, $\lambda_4 = 1.882493e+01$, $\lambda_3 = 1.561710e+01$, $\lambda_2 = 6.658370$ and $\lambda_1 = 1.682933$. It is easy to see that the condition number is approximately equal to $1.4175e+005$. So, the design matrix $\boldsymbol{X}$ is morbidity badly.

An appropriate approach is to replace the pure parametric model with semiparametric model. We use the added-variable plots (except for the binary explanatory variables) to identify the parametric and nonparametric components of the model. Added-variable plots enable us to visually assess the effect of each predictor, having adjusted for the effects of the other predictors. By looking at added-variable plot (Figure 3), we consider ANI

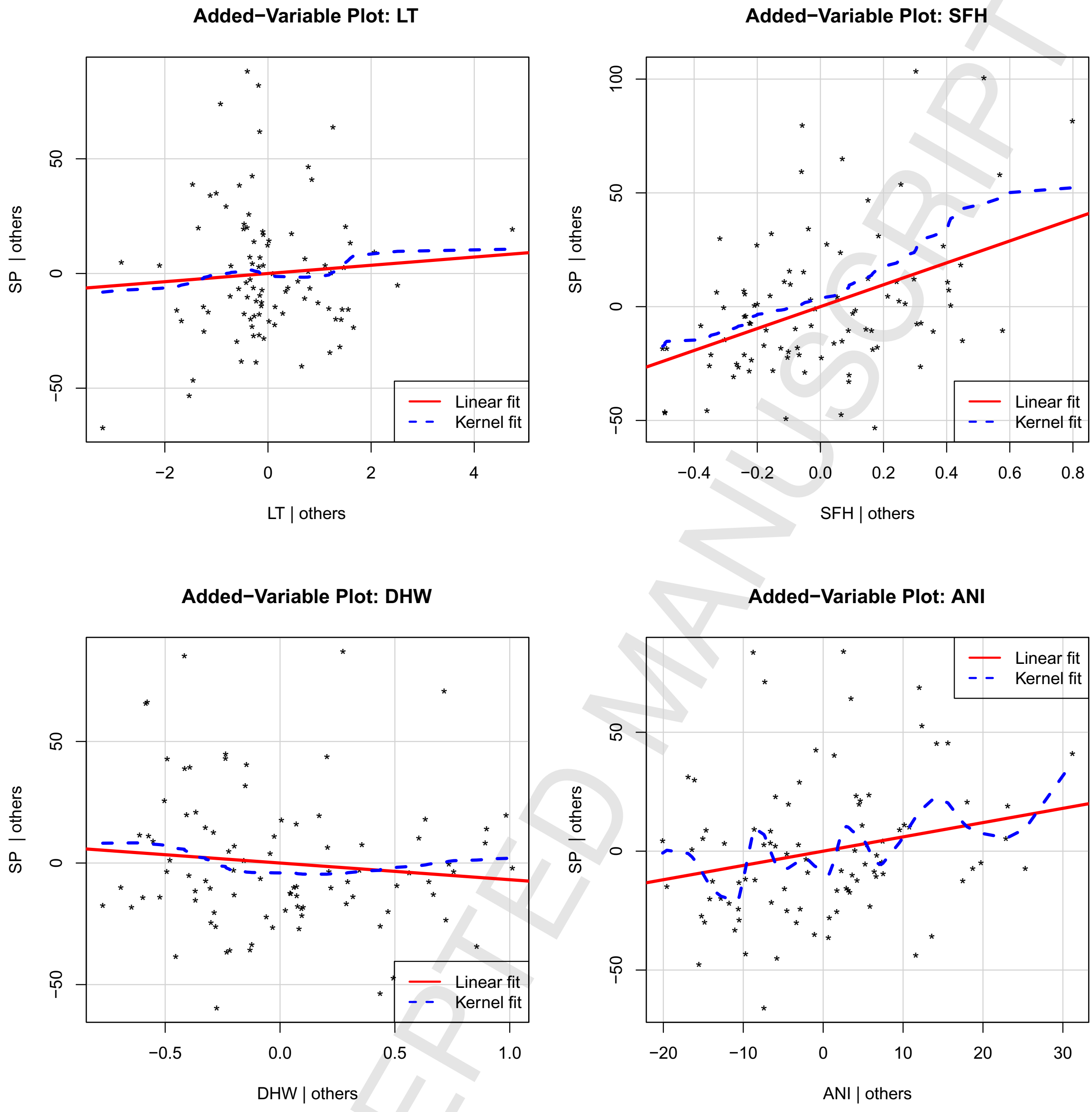

Figure 3: Added-variable plots of individual explanatory variables vs. dependent variable, linear fit (red solid line) and kernel fit (blue dashed line).

Table 11: Evaluation of parameters for proposed estimators for real data set

Method / Coefficients	GLSRE	GLTSRE	GRLTSRE	GERLTSRE	GLSRRE	GLTSRRE	GRLTSRRE	GERLTSRRE
intercept	-	-	-	-	-	-	-	-
LT	0.7018	1.0509	0.1522	0.1507	0.8019	1.1349	0.2636	0.2568
SFH	46.7515	33.5686	30.9487	30.2085	39.2135	26.8318	25.1185	24.4659
FP	3.9311	2.5740	2.7443	2.6777	3.2004	1.9234	2.1637	2.1075
DHW	-1.6147	-0.7616	-1.2961	-1.2635	-1.1992	-0.3942	-0.9500	-0.9254
GAR	6.2476	4.3865	4.1926	4.0919	5.2015	3.4525	3.3773	3.2896
$\boldsymbol{e}^\top(\boldsymbol{z} \vee \boldsymbol{z}^* \vee \boldsymbol{z}^{**})$	92.000	86.0000	71.9956	74.0379	92.000	86.0000	71.9956	74.0379
$\hat{k}$	0.0000	0.0000	0.0000	0.0000	1.6605	1.8761	1.4598	1.4851
RSS	77851.32	39094.07	30964.67	25337.26	70034.08	28293.90	23897.44	17590.95
R^2	0.2346	0.6156	0.6956	0.7509	0.3114	0.7218	0.7650	0.8271

as a nonparametric part and so, the specification of the semiparametric regression model is

$$(SP)_i = \beta_0 + \beta_1(LT)_i + \beta_2(SFH)_i + \beta_3(FP)_i + \beta_4(DHW)_i + \beta_5(GAR)_i + f(ANI)_i + \epsilon_i. \tag{6.5}$$

The ratio of largest eigenvalue to smallest eigenvalue for new design matrix in model (6.5) is approximately $\lambda_5/\lambda_1 = 427.9926$ and so, there exists a potential multicollinearity between the columns of design matrix. As it can be seen, the ridge type of proposed estimators perform better than non-ridge forms.

To compare the performance of the proposed restricted estimators, we consider the parametric restriction $\boldsymbol{R\beta} = \boldsymbol{0}$, where

$$\boldsymbol{R} = \begin{pmatrix} -1 & 0 & -1 & -1 & 1 \\ 1 & 0 & -1 & 2 & 0 \\ 0 & -1 & 0 & -2 & 8 \end{pmatrix}.$$

We test the linear hypothesis $H_0 : \boldsymbol{R\beta} = \boldsymbol{0}$ in the framework of our semiparametric model (2.1). The test statistic for H_0, given our observations, is

$$\chi^2_{rank(\boldsymbol{R})} \simeq (\boldsymbol{R}\hat{\boldsymbol{\beta}}_G^{LS} - \boldsymbol{r})^\top \left(\boldsymbol{R}\hat{\Sigma}\boldsymbol{R}^\top\right)^{-1} (\boldsymbol{R}\hat{\boldsymbol{\beta}}_G^{LS} - \boldsymbol{r}) = 0.4781,$$

where $\hat{\Sigma} = \hat{\sigma}^2 \big(\widetilde{\boldsymbol{X}}^\top \widetilde{\boldsymbol{X}}\big)^{-1}$. Thus we conclude that the null hypothesis H_0 is not rejected.

Table 11 shows a summery of the results (the intercept term was not significant for the model and so, we considered an RSRM without intercept). In this Table, the RSS and R^2 respectively are the residual sum of squares and coefficient of determination of the model, i.e., $\mathrm{RSS} = \sum_{i=1}^n (y_i - \hat{y}_i)^2$, $\hat{y}_i = \boldsymbol{x}_i\hat{\boldsymbol{\beta}} + \hat{f}(t_i)$ and $\mathrm{R}^2 = 1 - \mathrm{RSS}/\mathrm{S_{yy}}$, which calculated for eight proposed estimators of $\boldsymbol{\beta}$. For estimation of nonparametric effect, at first we estimated the parametric effects by one of the proposed methods and then, local polynomial approach was applied to fit $SP_i - \boldsymbol{x}_i^\top\hat{\boldsymbol{\beta}}$ on ANI_i, $i = 1, ..., n$ for all proposed linear estimators, where $\boldsymbol{x}_i^\top = (LT_i, SFH_i, FP_i, DHW_i, GAR_i)$ (the results have been displayed in Figures 4 & 5).

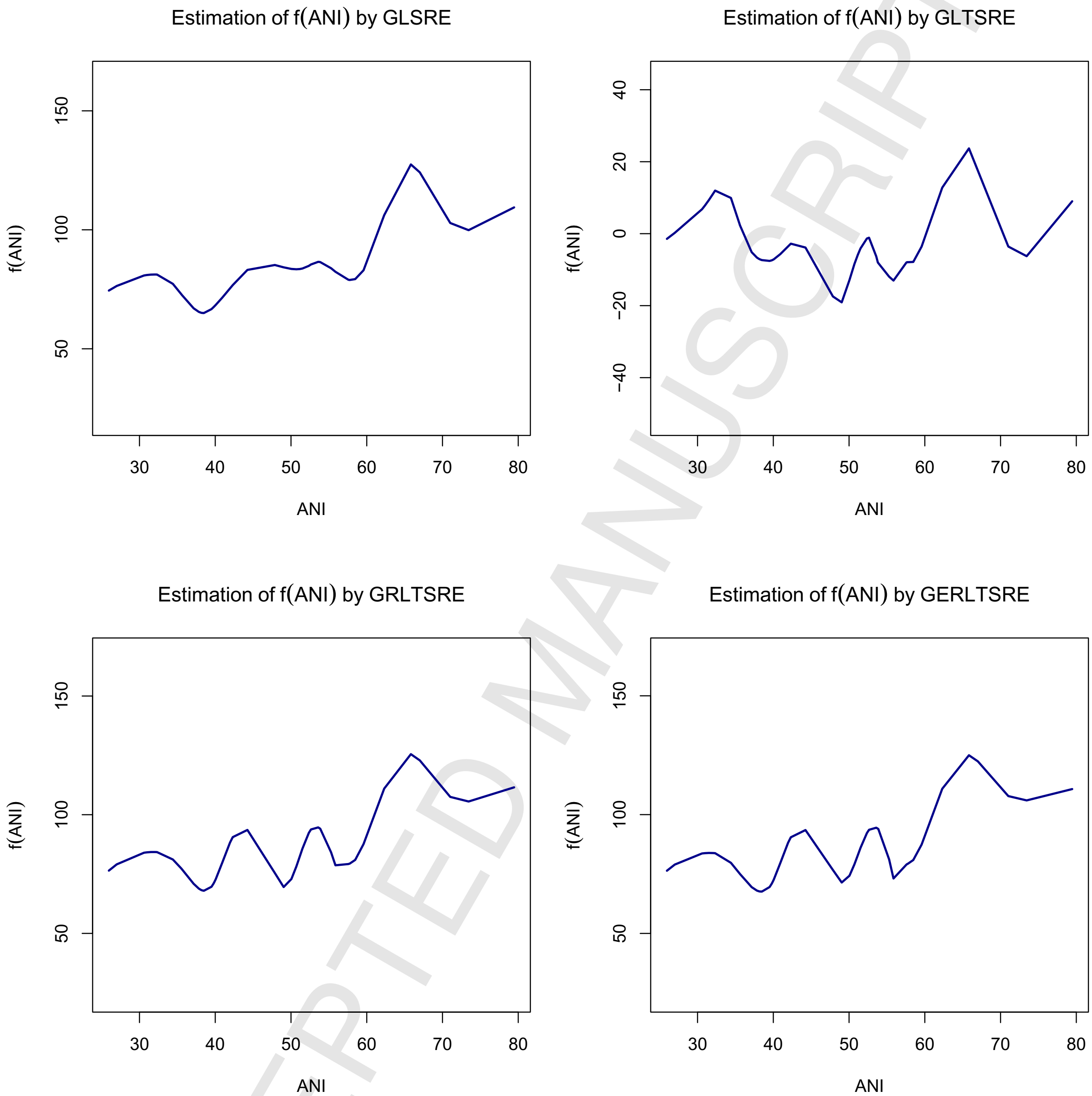

Figure 4: The robust and nonrobust estimations of nonparametric part of model (4.1).

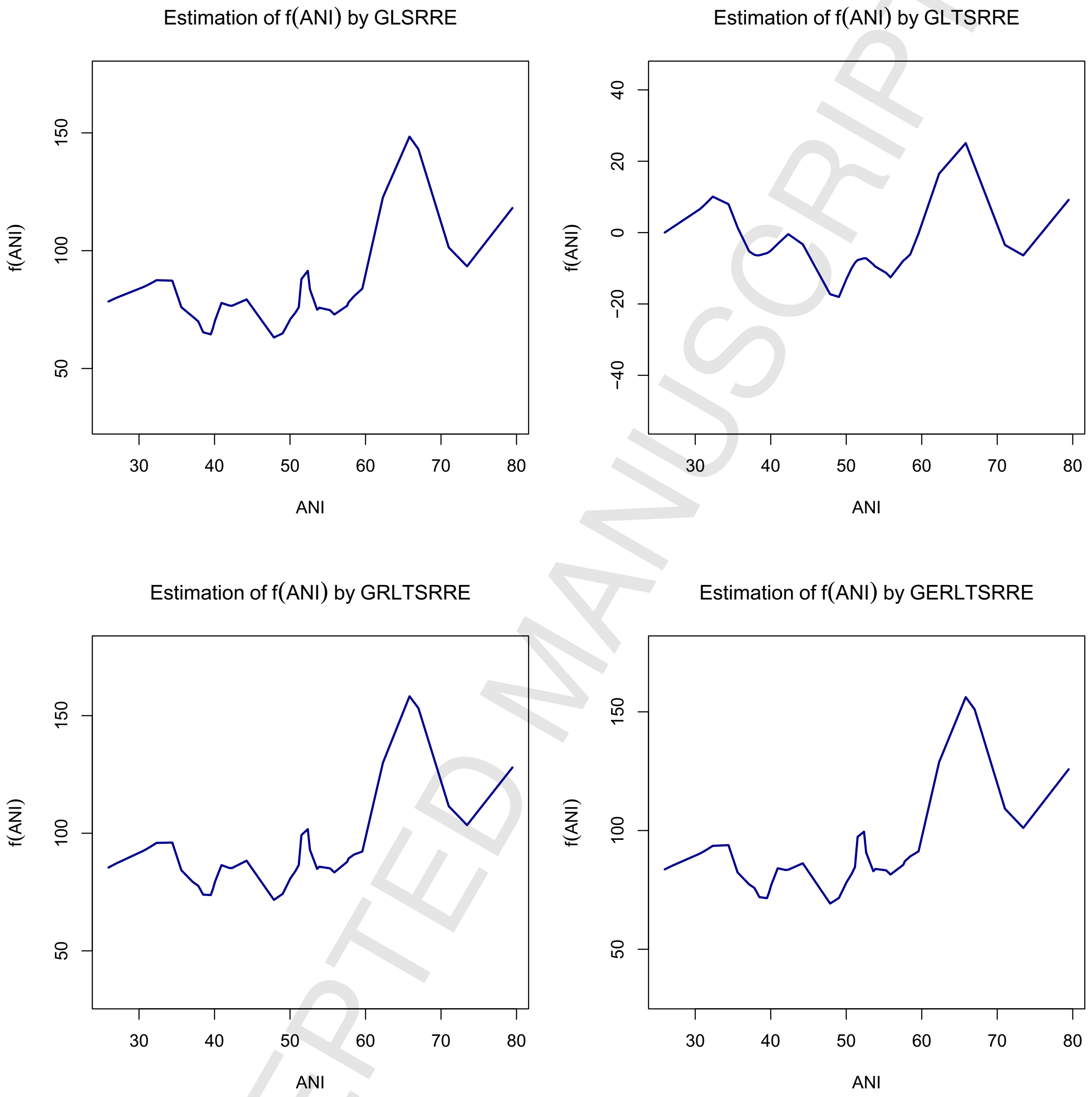

Figure 5: The robust and nonrobust ridge type estimations of nonparametric part of model (4.1).

As it can be found from Table 11 and Figures 4 & 5, because of multicollinearity between the columns of design matrix, the ridge type of robust and nonrobust estimators perform better than nonridge type in the sense of parametric and nonparametric estimations. Moreover, since there exist outliers in the data set, the robust estimators are more efficient than nonrobust estimators in the model fitting.

7 Summary and conclusions

In this paper, we proposed ridge and nonridge form of generalized restricted robust estimators in a semiparametric regression model when the errors were dependent and some additional linear constraints held on the parameter $\boldsymbol{\beta}$. In the presence of multicollinearity we introduced robust ridge type estimators under dependency among column vectors of the design matrix in a semiparametric regression model. Since, a theoretical comparison was not possible, the Monté-Carlo simulation studies and a real data example have been conducted to compare the performance of the proposed estimators numerically. The results from the Monté-Carlo simulations for $n = 150$, $P = 5$ and different ρ^2 and γ are presented in Tables 1 to 9 and Figure 2. From these tables it can be seen that the factors affecting the performance of the estimators are the degree of correlation (γ) and percentage of outliers (α). It can be concluded that GERLTSRRE is leading to be the best estimator among others for the parametric part of the model, since it offers smaller risk and mse values in all proposed estimators. Further GLSRE is the worst estimator for the parametric part in this examples. In general, the values of α and γ^2 have positive (negative) effect on the estimated risks (performance of the estimators). In the real example study, a near dependency among the column of $\boldsymbol{X}^\top\boldsymbol{X}$ identified from $\lambda_7/\lambda_1 = 141750.3$, that is, the design matrix may be considered as being very ill-conditioned and we had to consider the ridge form of proposed estimators in our study. As it can be seen from Figure 3, the nonlinear relation between sale price and average neighborhood income (ANI) can be detected and so, the pure parametric model does not fit to the data and semiparametric regression model fits more significantly. Further, from Table 11 and Figures 4 and 5, it can be deduced that GERLTSRRE is quite efficient in the sense that it has significant value of goodness of fit.

Acknowledgements

This research was supported by Iran National Science Foundation (INSF) and Research Council of Semnan University. The author is grateful to the anonymous reviewers, associate editor and editor-in-chief for their valuable comments helped to improve the quality of this paper.

References

[1] Akdenïz F, Akdenïz Duran H, Roozbeh M, Arashi M. Efficiency of the Generalized Difference-based Liu Estimators in Semiparametric Regression Models with Correlated Errors. *J. Statist. Comp. Sim.* 2015;

85: 147–165.

[2] Akdenïz F, Tabakan G. Restricted ridge estimators of the parameters in semiparametric regression model, *Comm. Statist. Theo. Meth.* 2009; **38**(11): 1852–1869.

[3] Akdenïz Duran E, Härdle WK, Osipenko M. Difference-based ridge and Liu type estimators in semiparametric regression models. *J. Mul. Anal.* 2012; **105**: 164–175.

[4] Amini M, Roozbeh M. Optimal partial ridge estimation in restricted semiparametric regression models. *J. Mul. Anal.* 2015; **136**: 26–40.

[5] Arashi M, Janfada M, Norouzirad M. Singular ridge regression with stochastic constraints, *Comm. Statist. Theo. Meth.* 2015; **44**: 1281–1292.

[6] Arashi M, Kibria BMG, Norouzirad M, Nadarajah S. Improved Preliminary test and Stein-Rule Liu Estimators for the Ill-Conditioned Elliptical Linear Regression Model. *J. Mul. Anal.* 2014; **124**: 53–74.

[7] Arashi M, Valizadeh T. Performance of Kibria's methods in partial linear ridge regression model. *Stat. Pap.* 2015; **56**: 231–246.

[8] Bunea F. Consistent covariate selection and post model selection inference in semiparametric regression. *Ann. Statist.* 2004; **32**: 898927.

[9] Bernholt T. 2005. Computing the least median of squares estimator in time $O(n^d)$. Proceedings of the International Conference on Computational Science and Its Applications. Springer LNCS. Berlin: Springer; Vol. 3480. p. 697–706.

[10] Bremner D, Chen D, Iacono J, Langerman S, Morin P. Output-sensitive algorithms for Tukey depth and related problems. *Stat. Comput.* 2008; **18**: 259–266.

[11] Edelsbrunner H, Souvaine DL. Computing median-of-squares regression lines and guided topological sweep. *J. Am. Stat. Assoc.* 1990; **85**: 115–119.

[12] Erickson J, Har-Peled S, Mount DM. On the least median square problem, *Discrete Comput. Geom.* 2006; **36**: 593–607.

[13] Gibbons DG. A simulation study of some ridge estimators, *J. Amer. Statist. Assoc.* 1981; **76**: 131-139.

[14] Härdle Wk, Liang H, Gao J. *Partially linear models. Heidelberg*: Physika Verlag; 2000.

[15] Hassanzadeh Bashtian M, Arashi M, Tabatabaey SMM. (2011). Using improved estimation strategies to combat multicollinearity. *J. Statist. Comp. Sim.* 2011; **81**(12): 1773-1797.

[16] Ho M. *Essays on the Housing Market.* Unpublished Ph.D. Dissertation University. 1995. of Toronto.

[17] Hoerl AE, Kennard RW. Ridge regression: biased estimation for non-orthogonal problems. *Thechnometrics*. 1970; **12**: 69–82.

[18] Hoerl A E, Kennard RW, Baldwin KF. Ridge regression: Some simulation. *Comm. Statist. Sim. Comp.* 1975; **4**: 105–123.

[19] Jung KM. Multivariate least-trimmed squares regression estimator. *Comput. Stat. Data Anal.* 2005; **48**: 307–316.

[20] Li J, Zhang W, Wu Z. Optimal zone for bandwidth selection in semiparametric models. *J. Nonparametric Statist.* 2011; **23**: 701–717.

[21] Kaciranlar S, Sakalliogu S, Ozkale MR, Guler H. More on the restricted ridge regression estimation. *J. Statist. Comp. Sim.* 2011; **23**: 1433–1448.

[22] Kibria BMG, Saleh AKMdE. Improving the Estimators of the Parameters of a Probit Regression Model: A Ridge Regression Approach. *J. Statist. Plann. Infer.* 2011; **142**(6), 1421–1435.

[23] Kibria BMG. Some Liu and Ridge Type Estimators and their Properties Under the ill-conditioned Gaussian Linear Regression Model. *J. Statist. Comp. Sim.* 2012; **82**: 1–17.

[24] McDonald GC, Galarneau DI. A monte carlo evaluation of some Ridge-type estimators. *J. Amer. Statist. Assoc.* 1975; **70**: 407–416.

[25] Mount DM, Netanyahu NS, Romanik K, Silverman R, Wu AY. A practical approximation algorithm for the LMS line estimator. *Comput. Stat. Data Anal.* 2007; **51**: 2461–2486.

[26] Mount DM, Netanyahu NS, Piatko CD, Silverman R,Wu AY. On the least trimmed squares estimator. *Algorithmica*. 2014; **69**: 148–183.

[27] Müller M. *Semiparametric Extensions to Generalized Linear Models.* Habilitationsschrift; 2000.

[28] Nguyen TD, Welsch R. Outlier detection and least trimmed squares approximation using semi-definite programming. *Comput. Stat. Data Anal.* 2010; **54**: 3212–3226.

[29] Norouzirad M, Arashi M, Saleh AKMdE. Restricted lASSO and double shrinking. *arXiv:1505.02913v1 [math.ST]* 2015.

[30] Roozbeh M, Arashi M, Niroumand HA. Ridge regression methodology in partial linear models with Correlated errors. *J. Statist. Comp. Sim.* 2011; **81**: 517–528.

[31] Roozbeh M. Shrinkage ridge estimators in semiparametric regression models. *J. Mul. Anal.* 2015; **136**: 56–74.

[32] Roozbeh M, Arashi M. Feasible ridge estimator in partially linear models. *J. Mul. Anal.* 2013; **116**: 35–44.

[33] Roozbeh M, Babaie-Kafaki S. Extended least trimmed squares estimator in semiparametric regression models with correlated errors. *J. Statist. Comp. Sim.* 2016; **86**: 357–372.

[34] Rousseeuw PJ. Least median of squares regression. *J. Am. Stat. Assoc.* 1984; **79**: 871–880.

[35] Rousseeuw PJ, Leroy AM. *Robust Regression and Outlier Detection.* New York: Wiley; 1987.

[36] Rousseeuw PJ, van Driessen K. Computing LTS regression for large data sets. *Data Min. Knowl. Disc.* 2006; **12**: 29–45.

[37] Saleh AKMdE, Kibria BMG. Performances of some new preliminary test ridge regression estimators and their properties. *Comm. Statist. Ser. A.* 1993; **22**: 2747–2764.

[38] Simpson DG, Ruppert D, Carroll RJ. On one-step GM estimates and stability of influences in linear regression. *J. Am. Stat. Assoc.* 1992; **87**: 439–450.

[39] Speckman P. Kernel somoothing in partial linear models. *J. R. Statist. Soc.* Ser. B. 1988; **50**: 413–436.

[40] Sun W, Yuan YX. *Optimization Theory and Methods: Nonlinear Programming.* New York: Springer; 2006.

[41] Swamy PAVB, Mehta JS, Roppoport PN. Two methods of evaluating Horel and Kennard's ridge regression. *Comm. Statist. Theo. Meth.* 1978; **12**: 1133–1155.

[42] Yohai VJ. High breakdown-point and high efficiency robust estimates for regression. *Ann. Stat.* 1987; **15**: 642–656.

[43] You J, Chen G, Zhou Y. Statistical inference of partially linear regression models with heteroscedastic errors. *J. Mul. Anal.* 2007; **98**: 1539–1557.

[44] You J, Chen G. Semiparametric generalized least-squares estimation in partially linear regression models with correlated errors, *J. Statist. Plann. Infer.* 2007; **137**: 117–132.