Ridge Stochastic Restricted Estimators in Semiparametric Linear Measurement Error Models

Hadi Emami

Department of Statistics, University of Zanjan, Zanjan, Iran.

Abstract. In this article we consider the stochastic restricted ridge estimation in semiparametric linear models when the covariates are measured with additive errors. The development of penalized corrected likelihood method in such model is the basis for derivation of ridge estimates. The asymptotic normality of the resulting estimates is established. Also, necessary and sufficient conditions, for the superiority of the proposed estimator over its counterpart, for selecting the ridge parameter k are obtained. A Monte Carlo simulation study is also performed to illustrate the finite sample performance of the proposed procedures. Finally theoretical results are applied to Egyptian pottery industry data set.

1 Introduction

One of the standard assumptions in the linear regression analysis is that all the explanatory variables are linearly independent. When this assumption is violated, the

Hadi Emami (h.emami@znu.ac.ir)

problem of multicollinearity enters into the data. Multicollinearity is troublesome and the effects of its presence on different aspects of linear models are well known, see Belsley (1991) for more details. When there is multicollinearity in data alternative estimators has been proposed which are generally biased. Among them ridge estimators have received a great deal of attention in statistical literature since the seminal work of Horel and Kennard (1970). An alternative method to combat the multicollinearity problem is to consider parameter estimation in addition to the sample information, such as some exact or stochastic restrictions on the unknown parameter vector. The incorporation of prior information available in the form of exact or stochastic restrictions provide better estimators than the ordinary estimator (Rao *et al.* , 2008). When the additional information of unknown vector of parameter is of the form of stochastic linear restrictions some mixed estimators are proposed by, Durbin (1953), Theil (1963), Ozkale (2009), Liang *et al.* (1999) and so on. When the additional information is of the form of exact linear restriction some other mixed ridge estimator are introduced; see Swamy and Mehta (1978), Grob (2003), Zhong and Yang (2007) and Siray and Torker (2014) among others. In semiparametric linear regression models there are a few studies that have looked at overcoming multicollinearity problems. Among them, Hu (2005) extended the ridge estimator of Horel and Kennard (1970), Roozbeh and Arashi (2013) and Roozbeh (2015) generalized the method of Zhong and Yang (2007) in semiparametric regression models. Akdeniz *et al.* (2012) considered a difference based ridge regression estimator and a Liu type estimator of the parameters, Akdeniz and Tabakan (2009) discussed restricted ridge estimation in semiparametric regression models. For recent developments in this field see Roozbeh (2018), Jibo and Asar (2017), Aydin *et al.* (2016) and Yuzbasi and Ejaz (2016). Another fundamental assumption in all statistical analyses is that all the observations are correctly observed. When this assumption is violated, the measurement errors creep into the data. Then the usual statistical tools tend to lose their validity, see Fuller (1987) and Cheng and Van (1999) for more details. When the problem of multicollinearity is present in the measurement error ridden data, then the main subject is how to obtain the consistent estimators of regression coefficients. The suggested idea is to use the ridge regression approach over the measurement error ridden data. In full-parametric linear measurement error models Rasekh (2001), Saleh (2014) and Ghapani and Babadi (2016) have considered collinearity in terms of a relationship between the elements of unobservable values and generalized ridge technique estimation to combat multicollinearity. However, there does not seem to be a work considering possible effects that collinearity can have on the different aspects of the estimates of parameters in semiparametric linear measurement error models (SLMeMs), nor to ridge estimation procedure as an alternative in

the presence of collinearity and ill conditioning. In this paper employing the corrected score function of Nakamura (1990) and Zhong *et al.* (2002) we concentrate on SLMeMs and consider the ridge estimates under a stochastic linear combination of parameters. In Section (2), the model and penalized corrected likelihood estimation are given, while the ridge stochastic restricted estimators are driven in Section (3), in continuation the asymptotic properties of ridge stochastic restricted estimators are summarized in this section. In Section (4) some comparison results are given. A simulation study with real data analysis are given in Sections (5) and (6) respectively. The final conclusions are given in Section (7).

2   The Model Description

Suppose that $\{(\mathbf{W}_i, \mathbf{X}_i, Y_i), i = 1, \ldots, n\}$ is a random sample from the semiparametric linear measurement error model (SLMeM),

$$\begin{cases} Y = \mathbf{X}\beta + \mathbf{f}(t) + \epsilon \\ \mathbf{W} = \mathbf{X} + \mathbf{U}, \end{cases} \tag{2.1}$$

where $Y = [Y_1, \ldots, Y_n]^\top$ is the n-vector of observations, $\mathbf{X} = [\mathbf{X}_1, \ldots, \mathbf{X}_n]^\top$ is the $n \times p$ matrix of unobserved covariates, which can be observed through the matrix $\mathbf{W} = [\mathbf{W}_1, \ldots, \mathbf{W}_n]^\top$ with the measurement error $\mathbf{U} = [\mathbf{U}_1, \ldots, \mathbf{U}_n]^\top$ where $\mathbf{U}_i, i = 1, \ldots, n$ are $p \times 1$ uncorrelated random vectors from p-variate normal distribution with $E(\mathbf{U}_i) = 0$ and $var(\mathbf{U}_i) = \Sigma$. Also, $\mathbf{f}(t) = [\mathbf{f}(t_1), \ldots, \mathbf{f}(t_n)]^\top$ is a vector of evaluations of the unknown nonparametric function and ϵ is the n-vector of the random errors from $N(0, \sigma^2 \mathbf{I}_n)$. The t_i is a scalar $a \le t_1 \le t_2 \le \ldots \le t_n \le b$, and $t_i's$ are not all idential and have been reordered so that $t_1 \le t_2 \le \ldots \le t_n$. The nonparametric part is to be estimated assuming only that f is an element of the $\mathcal{S}_2^2[a, b]$ of functions f that have first order continuous derivatives and square-integratable second order derivatives f'' in $[a, b]$. We assume that $\mathbf{U}$ is independent of $(Y, \mathbf{X}, t)$ and Σ is a $p \times p$ matrix of known values with nonnegative diagonal elements (Fuller , 1987) and an unbiased estimate of Σ is available using replicated observations on the independent variables (Nagelkerke , 1992, see Section 3.VII for more details). Let the ordered distinct values among $t_1, \ldots, t_n$ be denoted by $s_1, \ldots, s_q$. The connection between $t_1, \ldots, t_n$ and $s_1, \ldots, s_q$ is captured by means of $n \times q$ incidence matrix $\mathbf{N}$, with entries $\mathbf{N}_{ij} = 1$ if $t_i = s_j$ and 0 otherwise. Let $\mathbf{f}$ be the vector of values $a_i = \mathbf{f}(s_i)$. For model (2.1) the penalized log-likelihood function is defined as

$$l(\Theta; \mathbf{X}, Y) = -\frac{n}{2} log(2\pi\sigma^2) - \frac{\|Y - \mathbf{X}\beta - \mathbf{N}\mathbf{f}\|^2}{2\sigma^2} - \frac{\lambda}{2\sigma^2} \int \mathbf{f}''(t)^2 dt, \tag{2.2}$$

where $\Theta = (\beta, \mathbf{f})$ and λ is a smoothing parameter which is determined by minimizing the generalized cross-validation criterion *GCV*. There exists a matrix $\mathbf{M}$ depending only on the knots $\{s_j\}$, such that the minimized value of $\int \mathbf{f}''(t)^2 dt$ is $\mathbf{f}^\top \mathbf{M} \mathbf{f}$ (Green and Silverman (1994), p.66). The equation in (2.2) is therefore of the form

$$l(\Theta; \mathbf{X}, Y) = -\frac{n}{2} log(2\pi\sigma^2) - \frac{\|Y - \mathbf{X}\beta - \mathbf{N}\mathbf{g}\|^2}{2\sigma^2} - \frac{\lambda}{2\sigma^2} \mathbf{f}^\top \mathbf{M} \mathbf{f}. \tag{2.3}$$

In cases that measurement error is negligible, as shown in (2.1), we replace $\mathbf{X}$ by $\mathbf{W}$ in (2.3) so that

$$U(\Theta; \mathbf{W}, Y) = \frac{\partial}{\partial\Theta} l(\Theta; \mathbf{W}, Y), \tag{2.4}$$

which is typically called naive score function. Hence, the expectations of $U(\Theta; \mathbf{W}, Y)$ with respect to Y evaluated at the true parameter value Θ_0, typically, are not equal to zero. Furthermore, in general, estimators obtained using naive score functions are not consistent. To correct this using approach of Nakamura (1990) the corrected likelihood $l^*(\Theta; \mathbf{W}, Y)$ is

$$l^*(\Theta; \mathbf{W}, Y) = -\frac{1}{2\sigma^2} \left\{ \sigma^2 n log(2\pi\sigma^2) + \|Y - \mathbf{W}\beta - \mathbf{N}\mathbf{f}\|^2 - n\beta^\top \Sigma\beta + \lambda \mathbf{f}' \mathbf{M} \mathbf{f} \right\}. \tag{2.5}$$

Since

$$E^*(\mathbf{W}^\top \mathbf{W}) = \mathbf{X}^\top \mathbf{X} + n\Sigma,$$

$l^*(\Theta; \mathbf{W}, Y)$ satisfies

$$E^*\{l^*(\Theta; \mathbf{W}, Y)\} = l(\Theta; \mathbf{X}, Y),$$

where $\mathbf{E}^*$ denotes the conditional mean with respect to $\mathbf{W}$ given $\mathbf{X}$ and Y.

Let $U^*(\Theta, \mathbf{W}, Y) = \partial l^*(\Theta; \mathbf{W}, Y)/\partial\Theta$ and $\mathbf{J}^*(\Theta, \mathbf{W}, Y) = -\partial U^*(\Theta, \mathbf{W}, Y)/\partial\Theta$ be the corrected score function and corrected observed function respectively. Then we have

$$E^*\{U^*(\Theta; \mathbf{W}, Y)\} = U(\Theta; \mathbf{X}, Y), \quad E^*\{\mathbf{J}^*(\Theta; \mathbf{W}, Y)\} = \mathbf{J}(\Theta; \mathbf{X}, Y).$$

The value of $\hat{\Theta}$ satisfying $U^*(\hat{\Theta}; \mathbf{W}, Y) = 0$ is called the penalized corrected likelihood estimate (*PCLE*) of Θ. By simple calculus, it follows that (2.5) is maximized when β and $\mathbf{f}$ satisfy the equation:

$$\begin{pmatrix} \mathbf{W}^\top \mathbf{W} - n\Sigma & \mathbf{W}^\top \mathbf{N} \\ \mathbf{N}^\top \mathbf{W} & \mathbf{N}^\top \mathbf{N} + \lambda \mathbf{M} \end{pmatrix} \begin{pmatrix} \beta \\ \mathbf{f} \end{pmatrix} = \begin{pmatrix} \mathbf{W}^\top \\ \mathbf{N}^\top \end{pmatrix} Y.$$

This equation gives the *PCLE*s of β and $\mathbf{f}$ as

$$\hat{\beta} = \{\mathbf{W}^\top(\mathbf{I}_n - \mathbf{S})\mathbf{W} - n\Sigma\}^{-1}\mathbf{W}^\top(\mathbf{I}_n - \mathbf{S})Y, \tag{2.6}$$

$$\hat{\mathbf{f}} = (\mathbf{N}^\top\mathbf{N} + \lambda\mathbf{M})^{-1}\mathbf{N}^\top(Y - \mathbf{W}\hat{\beta}), \tag{2.7}$$

respectively, where $\mathbf{S} = \mathbf{N}(\mathbf{N}^\top\mathbf{N} + \lambda\mathbf{M})^{-1}\mathbf{N}^\top$. Using (2.6) and (2.7) the vector of fitted values is $\hat{Y} = \mathbf{W}\hat{\beta} + \mathbf{N}\hat{\mathbf{f}}$ and the estimate of σ^2 is $\hat{\sigma}^2 = \frac{e^\top e}{n-p}$, where $e = Y - \hat{Y}$.

The incorporation of prior information available in the form of restrictions may provide better estimators than the $\hat{\beta}$. Since uncertainty may exist about the prior information on a particular parameter or a linear combination of parameters, one suggestion is to make use of stochastic linear restrictions of the following form

$$\mathbf{r} = \mathbf{R}\beta + \mathbf{e}, \tag{2.8}$$

where $\mathbf{r}$ is a $q \times 1$ observable random vector, $\mathbf{R}$ is a $q \times p$ known prior information on the vector of regression coefficients of rank $q \leq p$, and $\mathbf{e}$ is an error vector, independent of $\mathbf{U}$ and ϵ, with $E(\mathbf{e}) = 0$ and $V(\mathbf{e}) = \sigma^2\mathbf{V}$, where $\mathbf{V}$ is an $q \times q$ known positive definite (p.d) matrix. If $\mathbf{V}$ is unknown, it may sometimes be more realistic to suppose that $E(\mathbf{e}\mathbf{e}^\top) = \mathbf{V}$. In this case since we have $E_{ry}(r - \mathbf{R}\hat{\beta})(r - \mathbf{R}\hat{\beta})^\top = \mathbf{V}$, where E_{ry} is the expectation with respect to r and Y, then $\frac{1}{q}(r - \mathbf{R}\hat{\beta})(r - \mathbf{R}\hat{\beta})^\top$ can be regarded as a rough estimate of $\mathbf{V}$. Subject to the imposed stochastic linear restriction (2.8), the stochastic restricted PCLE (*SRE*, say) are given by

$$\hat{\beta}_{sr} = (\mathbf{M}_0 + \mathbf{R}^\top\mathbf{V}\mathbf{R})^{-1}\{\mathbf{W}^\top(\mathbf{I}_n - \mathbf{S})Y + \mathbf{R}^\top\mathbf{V}r\},$$

and

$$\hat{\mathbf{f}}_{sr} = (\mathbf{N}^\top\mathbf{N} + \lambda\mathbf{M})^{-1}\mathbf{N}^\top(Y - \mathbf{W}\hat{\beta}_{sr}),$$

respectively, where $\mathbf{M}_0 = \mathbf{W}^\top(\mathbf{I}_n - \mathbf{S})\mathbf{W} - n\Sigma$.

3 Ridge Stochastic Restricted Estimators

In fact, the coefficient parameter β can be regarded as a vector in p dimensional space. If there exists multicollinearity in $\mathbf{X}$ (or equivalently, $\mathbf{X}^\top(\mathbf{I}_n - \mathbf{S})\mathbf{X}$ is ill conditioned), the coefficients which are estimated by *PCLE*, would be badly apart from the actual coefficient parameter in some directions of p dimensional space, since the asymptomatic variance of $\hat{\beta}$ depends on matrix $[\mathbf{X}^\top(\mathbf{I}_n - \mathbf{S})\mathbf{X}]^{-1}$ (see Corollary (2)). In order to overcome the multicollinearity we maximize the penalized corrected likelihood with a spherical

restriction ($\beta^\top \beta \leq \rho^2$) and stochastic linear restriction (2.8). Therefore, the stochastic restricted SLMeM is transformed into an optimization problem with two restrictions:

$$\max_\Theta l^*(\Theta, \mathbf{W}, Y),$$

$$\text{subject to} \quad \beta^\top \beta \leq \rho^2,$$

$$\mathbf{r} = \mathbf{R}\beta + \mathbf{e}.$$

In order to incorporate the above restrictions in the estimation of parameters, we maximize

$$l^*(\Theta; \mathbf{W}, Y) + (\mathbf{r} - \mathbf{R}\beta)^\top \mathbf{V}^{-1}(\mathbf{r} - \mathbf{R}\beta) + k(\beta^\top \beta - \rho^2), \tag{3.1}$$

where k is a Lagrange multiplier. Differentiating (3.1) with respect to $\Theta = (\beta, \mathbf{f})$ and setting the result to zero, we obtain the ridge *SREs* (*RSREs*), as

$$\hat{\beta}_{sr}^{(k)} = (\mathbf{M}_0 + \mathbf{R}^\top \mathbf{V}^{-1}\mathbf{R} + k\mathbf{I}_p)^{-1}\{\mathbf{W}^\top(\mathbf{I}_n - \mathbf{S})Y + \mathbf{R}^\top \mathbf{V}^{-1}\mathbf{r}\}, \tag{3.2}$$

and

$$\hat{\mathbf{f}}_{sr}^{(k)} = (\mathbf{N}^\top \mathbf{N} + \lambda \mathbf{M})^{-1}\mathbf{N}^\top(Y - \mathbf{W}\hat{\beta}_{sr}^{(k)}), \tag{3.3}$$

respectively. It is clear that for $k = 0$, we get $\hat{\beta}_{sr}^{(0)} = \hat{\beta}_{sr}$ and $\hat{\mathbf{f}}_{sr}^{(0)} = \hat{\mathbf{f}}_{sr}$.

3.1 Asymptotic Characteristics

To investigate the asymptotic behavior of $\hat{\beta}_{sr}^{(k)}$, we assume that the derivatives related to the likelihood exist. It is also assumed that as $n \to \infty$, the following limits exist: $n^{-1}\mathbf{X}^\top(\mathbf{I}_n - \mathbf{S})\mathbf{X}$, $n^{-1}\mathbf{X}^\top(\mathbf{I}_n - \mathbf{S})^2\mathbf{X}$, $n^{-1}\mathbf{X}^\top(\mathbf{I}_n - \mathbf{S})\mathbf{U}$, $n^{-1}tr(\mathbf{I}_n - \mathbf{S})$ and $n^{-1}tr(\mathbf{I}_n - \mathbf{S})^2$. The following lemma is useful for the next asymptotic results.

Lemma 3.1. Under the above assumption we have

$$\mathbf{W}^\top(\mathbf{I}_n - \mathbf{S})\mathbf{W} = \mathbf{X}^\top(\mathbf{I}_n - \mathbf{S})\mathbf{X} + tr(\mathbf{I}_n - \mathbf{S})\Sigma + O_p(n^{\frac{1}{2}}). \tag{3.4}$$

Proof. From (2.1) we can write

$$n^{-1}[\mathbf{W}^\top(\mathbf{I}_n - \mathbf{S})\mathbf{W} - \mathbf{X}^\top(\mathbf{I}_n - \mathbf{S})\mathbf{X} - tr(\mathbf{I}_n - \mathbf{S})\Sigma] = n^{-1}[\mathbf{X}^\top(\mathbf{I}_n - \mathbf{S})\mathbf{U} + \mathbf{U}^\top(\mathbf{I}_n - \mathbf{S})\mathbf{X} + \Pi],$$

where $\Pi = \mathbf{U}^\top(\mathbf{I}_n - \mathbf{S})\mathbf{U} - tr(\mathbf{I}_n - \mathbf{S})\Sigma$. Since by assumption as $n \to \infty$, $n^{-1}\mathbf{X}^\top(\mathbf{I}_n - \mathbf{S})^2\mathbf{X}$ exists we have $n^{-1/2}\mathbf{X}^\top(\mathbf{I}_n - \mathbf{S})\mathbf{U} \sim N(0, n^{-1}\mathbf{X}^\top(\mathbf{I}_n - \mathbf{S})^2\mathbf{X}\Sigma)$. Then $n^{-1/2}\mathbf{X}^\top(\mathbf{I}_n - \mathbf{S})\mathbf{U} =$

$O_p(n^{-1/2})$ and similarly $n^{-1}\mathbf{U}^\top(\mathbf{I}_n - \mathbf{S})\mathbf{X} = O_p(n^{-1/2})$. Now let the (a, b)th element of Π be is denoted by Π_{ab}. Then

$$\Pi_{ab} = \sum_{i=1}^{n}\sum_{j=1}^{n}\mathbf{U}_{ia}(\mathbf{I}_n - \mathbf{S})_{ij}\mathbf{U}_{jb} - tr(\mathbf{I}_n - \mathbf{S})\Sigma_{ab},$$

where $\mathbf{U} = (\mathbf{U}_{ij})$, $\mathbf{I}_n - \mathbf{S} = (\mathbf{I}_n - \mathbf{S})_{ij}$, $\Sigma = (\Sigma_{ab})$, $i, j = 1, 2, \ldots, n$ and $a, b = 1, 2, \ldots, p$. It is easily seen that $E(\mathbf{U}_{ia}\mathbf{U}_{ib}) = \Sigma_{ab}$, so we obtain $E(\Pi_{ab}) = 0$. Further, we have

$$E(\Pi_{ab}^2) = \sum_{i,j}\sum_{k,l}E(\mathbf{U}_{ia}\mathbf{U}_{jb}\mathbf{U}_{ka}\mathbf{U}_{ln})(\mathbf{I}_n - \mathbf{S})_{ij}(\mathbf{I}_n - \mathbf{S})_{kl} + tr(\mathbf{I}_n - \mathbf{S})^2\Sigma_{ab}^2$$

$$= tr(\mathbf{I}_n - \mathbf{S})^2(\Sigma_{aa}\Sigma_{bb} + \Sigma_{ab}^2).$$

The above result follows considering the moments of the normal distribution. In this case

- $E(\mathbf{U}_{ia}^2\mathbf{U}_{ib}^2) = (\Sigma_{aa}\Sigma_{bb} + \Sigma_{ab}^2), i = 1, \ldots, n.$

- $E(\mathbf{U}_{ia}^2\mathbf{U}_{jb}^2) = \Sigma_{aa}\Sigma_{bb}, i \neq j, i, j = 1, \ldots, n.$

- $E(\mathbf{U}_{ia}\mathbf{U}_{jb}\mathbf{U}_{ka}\mathbf{U}_{lb}) = \Sigma_{ab}^2, i \neq j, i, j = 1, \ldots, n.$

- It can be verified that the remaining terms $E(\mathbf{U}_{ia}\mathbf{U}_{jb}\mathbf{U}_{ka}\mathbf{U}_{lb}) = 0$ are all equal to zero.

By assumption, as $n \to \infty$, $n^{-1}tr(\mathbf{I}_n - \mathbf{S})^2$ exist, so we have $E(n^{-1/2}\Pi_{ab})^2 = O(1)$ as $n \to \infty$ and then $n^{-1}\Pi = O(n^{-1/2})$. Combining all the above results, we get Lemma (3.1). ☐

Theorem 1. $\hat{\beta}_{sr}^{(k)}$ is asymptotically normally distributed. The asymptotic mean and variance are respectively given as:

$$E(\hat{\beta}_{sr}^{(k)}) = \mathbf{T}_k^{-1}\mathbf{T}_0\beta,$$

$$avar(\hat{\beta}_{sr}^{(k)}) = \mathbf{T}_k^{-1}(\mathbf{D} + [\mathbf{X}^\top(\mathbf{I}_n - \mathbf{S})^2\mathbf{X} + \mathbf{R}^\top\mathbf{V}^{-1}\mathbf{R}]\sigma^2)\mathbf{T}_k^{-1},$$

where $\mathbf{T}_0 = \mathbf{X}^\top(\mathbf{I}_n - \mathbf{S})\mathbf{X} + \mathbf{R}^\top\mathbf{V}^{-1}\mathbf{R}$, $\mathbf{T}_k = \mathbf{T}_0 + k\mathbf{I}_p$ and $\mathbf{D} = \{(\mathbf{X}\beta + \mathbf{N}\mathbf{f})^\top(\mathbf{I}_n - \mathbf{S})^2(\mathbf{X}\beta + \mathbf{N}\mathbf{f}) + tr(\mathbf{I}_n - \mathbf{S})\sigma^2\}\Sigma.$

Proof. Let $\eta = n^{\frac{-1}{2}}\{\mathbf{W}^\top(\mathbf{I}_n - \mathbf{S})Y + \mathbf{R}^\top\mathbf{V}^{-1}\mathbf{r}\} = n^{\frac{-1}{2}}\{\mathbf{W}^\top(\mathbf{I}_n - \mathbf{S})Y + \mathbf{R}^\top\mathbf{V}^{-1}\mathbf{R}\beta + \mathbf{R}^\top\mathbf{V}^{-1}\mathbf{e}\}$. It is enough to obtain the asymptotic properties of η, for this let $(\mathbf{I}_n - \mathbf{S}) = \Gamma\mathbf{P}\Gamma^\top$ and $\mathbf{V}^{\frac{-1}{2}} = \Upsilon\Psi\Upsilon^\top$ be the spectral decompositions of $(\mathbf{I}_n - \mathbf{S})$ and $\mathbf{V}^{\frac{-1}{2}}$, respectively, $\Gamma\Gamma^\top = \mathbf{I}_n$, $\Upsilon\Upsilon^\top = \mathbf{I}_q$, $\mathbf{P} = diag(\gamma_1, \ldots, \gamma_n)$, $\Psi = diag(\lambda_1, \ldots, \lambda_q)$, γ_i and λ_i are eigenvalues of $\mathbf{I}_n - \mathbf{S}$ and $\mathbf{V}^{\frac{-1}{2}}$, respectively.

Let $\alpha = \mathbf{R}^\top\mathbf{V}^{-1}\mathbf{R}\beta$, then η can be rewritten as

$$\eta = n^{\frac{-1}{2}}[\alpha + \mathbf{W}^{*\top}\mathbf{P}\Upsilon^* + \mathbf{R}^{*\top}\Psi\mathbf{e}^*], \tag{3.5}$$

where $\mathbf{W}^* = \Gamma^\top\mathbf{W} \sim N(\Gamma^\top\mathbf{W}, \Sigma)$, $Y^* = \Gamma^\top Y \sim N(\Gamma^\top\mathbf{X}\beta + \Gamma^\top\mathbf{Nf}, \sigma^2\mathbf{I}_n)$, $\mathbf{R}^* = \Upsilon^\top\mathbf{R}$ and $\mathbf{e}^* = \Upsilon^\top\mathbf{e} \sim N(0, \sigma^2\mathbf{I}_q)$. The sth element of η is given by

$$\eta_s = \frac{1}{\sqrt{n}}[\alpha_i + \sum_{i=1}^{n}(\mathbf{W}_{is}^*\gamma_i Y_i^* + \mathbf{R}_{is}^*\lambda_i\mathbf{e}_i^*)] = \frac{\alpha_i}{\sqrt{n}} + \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\varsigma_i.$$

Since ς_i's are independent and since follow $var(\varsigma_k)$ exists as $n \to \infty$, by the central limit theorem, ς_k is asymptotically normal. It follows from $E(\mathbf{W}^\top(\mathbf{I}_n - \mathbf{S})Y + \mathbf{R}^\top\mathbf{V}^{-1}\mathbf{e}) = [\mathbf{X}^\top(\mathbf{I}_n - \mathbf{S})\mathbf{X} + \mathbf{R}^\top\mathbf{V}^{-1}\mathbf{R}]\beta + \mathbf{X}^\top(\mathbf{I}_n - \mathbf{S})\mathbf{Nf}$, that

$$E(\eta) = \frac{1}{\sqrt{n}}\mathbf{T}_0\beta + O_n(1/\sqrt{n}). \tag{3.6}$$

Let E^+ and var^+ denote the expectation and variance with respect to the random vector $(Y, \mathbf{r})^\top$, respectively. The variance of η can be obtained by

$$\begin{aligned}
var(\eta) &= E^+\{var^*(\eta)\} + var^+\{E^*(\eta)\} \\
&= n^{-1}E^+\{Y^\top(\mathbf{I}_n - \mathbf{S})^2 Y\Sigma\} + n^{-1}var^+\{\mathbf{X}^\top(\mathbf{I}_n - \mathbf{S})Y + \mathbf{R}^\top\mathbf{V}^{-1}\mathbf{r}\} \\
&= n^{-1}\{(\mathbf{X}\beta + \mathbf{Nf})^\top(\mathbf{I}_n - \mathbf{S})^2(\mathbf{X}\beta + \mathbf{Nf}) + tr(\mathbf{I}_n - \mathbf{S})\sigma^2\}\Sigma + n^{-1}\{\mathbf{X}^\top(\mathbf{I}_n - \mathbf{S})^2\mathbf{X} \\
&\quad + \mathbf{R}^\top\mathbf{V}^{-1}\mathbf{R}\}\sigma^2 \\
&= n^{-1}\{\mathbf{D} + [\mathbf{X}^\top(\mathbf{I}_n - \mathbf{S})^2\mathbf{X} + \mathbf{R}^\top\mathbf{V}^{-1}\mathbf{R}]\sigma^2\}. \tag{3.7}
\end{aligned}$$

Now from Lemma (3.1) we have

$$\begin{aligned}
\sqrt{n}\hat{\beta}_{sr}^{(k)} &= \{n^{-1}[\mathbf{X}^\top(\mathbf{I}_n - \mathbf{S})\mathbf{X} + \mathbf{R}^\top\mathbf{V}^{-1}\mathbf{R} + k\mathbf{I}_p] + O_p(1/n)\}^{-1}n^{\frac{-1}{2}}\{\mathbf{X}^\top(\mathbf{I}_n - \mathbf{S})Y + \mathbf{R}^\top\mathbf{V}^{-1}\mathbf{r}\} \\
&= \{\mathbf{I}_p + O_p(1/n)\}^{-1}\{n^{-1}[\mathbf{X}^\top(\mathbf{I}_n - \mathbf{S})\mathbf{X} + \mathbf{R}^\top\mathbf{V}^{-1}\mathbf{R} + k\mathbf{I}_p]\}^{-1}n^{\frac{-1}{2}}\{\mathbf{X}^\top(\mathbf{I}_n - \mathbf{S})Y + \mathbf{R}^\top\mathbf{V}^{-1}\mathbf{r}\} \\
&= \{\mathbf{I}_p + O_p(1/n)\}\{n^{-1}[\mathbf{X}^\top(\mathbf{I}_n - \mathbf{S})\mathbf{X} + \mathbf{R}^\top\mathbf{V}^{-1}\mathbf{R} + k\mathbf{I}_p]\}^{-1}n^{\frac{-1}{2}}\{\mathbf{X}^\top(\mathbf{I}_n - \mathbf{S})Y + \mathbf{R}^\top\mathbf{V}^{-1}\mathbf{r}\}. \tag{3.8}
\end{aligned}$$

Using (3.6) we can write

$$\sqrt{n}(\hat{\beta}_{sr}^{(k)} - \mathbf{T}_k^{-1}\mathbf{T}_0\beta) = \mathbf{T}_k^{-1}\{\eta - E(\eta)\} + O_p(1/\sqrt{n}). \tag{3.9}$$

By (3.7)-(3.9) we have $avar(\sqrt{n}\hat{\beta}_{sr}^{(k)}) = \mathbf{T}_k^{-1}\mathrm{var}(\eta)\mathbf{T}_k^{-1}$. Thus, $avar(\hat{\beta}_{sr}^{(k)}) = \mathbf{T}_k^{-1}(\mathbf{D} + [\mathbf{X}^\top(\mathbf{I}_n - \mathbf{S})^2\mathbf{X} + \mathbf{R}^\top\mathbf{V}^{-1}\mathbf{R}]\sigma^2)\mathbf{T}_k^{-1}$. □

From theorem (1) we have the following results:

Corollary 1. $\hat{\beta}_{sr}$ is asymptotically normally distributed. The asymptotic mean and variance are respectively given as

$$E(\hat{\beta}_{sr}) = \beta.$$
$$avar(\hat{\beta}_{sr}) = \mathbf{T}_0^{-1}\Lambda\mathbf{T}_0^{-1},$$

where $\Lambda = (\mathbf{D} + [\mathbf{X}^\top(\mathbf{I}_n - \mathbf{S})^2\mathbf{X} + \mathbf{R}^\top\mathbf{V}^{-1}\mathbf{R}]\sigma^2)$.

Corollary 2. $\hat{\beta}$ is asymptotically normally distributed. The asymptotic mean and variance are respectively given as

$$E(\hat{\beta}) = \beta.$$
$$avar(\hat{\beta}) = [\mathbf{X}^\top(\mathbf{I}_n - \mathbf{S})\mathbf{X}]^{-1}\Lambda_0[\mathbf{X}^\top(\mathbf{I}_n - \mathbf{S})\mathbf{X}]^{-1},$$

where $\Lambda_0 = (\mathbf{D} + [\mathbf{X}^\top(\mathbf{I}_n - \mathbf{S})^2\mathbf{X}]\sigma^2)$.

Now, we provide expressions for the asymptotic distributional bias ($\mathcal{ADB}$) and the asymptotic distributional risk $\mathcal{ADR}$ of aforementioned estimator. To do so, let us consider the following weighted quadratic loss function

$$\mathcal{L}(\hat{\beta}, \beta) = n(\hat{\beta} - \beta)^\top\mathbf{Q}(\hat{\beta} - \beta),$$

where $\mathbf{Q}$ is a symmetric positive definite matrix. Assume that we have $F(x) = \lim_{n\to\infty} P(\sqrt{n}(\hat{\beta} - \beta) \le x)$. Now we define the asymptotic distributional bias ($\mathcal{ADB}$) and asymptotic distributional risk ($\mathcal{ADR}$) by

$$\mathcal{ADB}(\hat{\beta}_n) = \lim_{n\to\infty} \sqrt{n}(\hat{\beta}_n - \beta).$$
$$\mathcal{ADR}_\mathbf{Q}(\hat{\beta}_n) = tr[\mathbf{Q}\int_{\Re^p}(x - \mu_x)(x - \mu_x)^\tau dF(x)] = tr(\mathbf{Q}\mathbf{V}_x).$$

where $\mathbf{V}_x$ is the dispersion matrix for the distribution $F(x)$. Thus we get

$$\mathcal{ADB}(\hat{\beta}_{sr}^{(k)}) = \lim_{n\to\infty} E\left(\sqrt{n}(\hat{\beta}_{sr}^{(k)} - \beta)\right)$$
$$= (\mathbf{T}_k\mathbf{T}_0 - \mathbf{I}_p)\beta = -k\mathbf{T}_k^{-1}\beta,$$
$$\mathcal{ADR}_Q(\hat{\beta}_{sr}^{(k)}) = tr\left(\mathbf{Q}^{\frac{1}{2}}\mathbf{T}_k^{-1}\Lambda\mathbf{T}_k^{-1}\mathbf{Q}^{\frac{1}{2}}\right).$$

Using the above definitions the $\mathcal{AMSE}$ function of $\hat{\beta}_{sr}^{(k)}$ is

$$\mathcal{AMSE}_Q(\hat{\beta}_{sr}^{(k)}) = \mathcal{ADB}(\hat{\beta}_{sr}^{(k)})^\top \mathcal{ADB}(\hat{\beta}_{sr}^{(k)}) + \mathcal{ADR}_Q(\hat{\beta}_{sr}^{(k)})$$
$$= k^2\beta^\top\mathbf{T}_k^{-2}\beta + tr\left(\mathbf{Q}^{\frac{1}{2}}\mathbf{T}_k^{-1}\Lambda\mathbf{T}_k^{-1}\mathbf{Q}^{\frac{1}{2}}\right), \qquad (3.10)$$

and the $\mathcal{AMSE}$ function of $\hat{\beta}_{sr}$ is

$$\mathcal{AMSE}_Q(\hat{\beta}_{sr}) = tr\left(\mathbf{Q}^{\frac{1}{2}}\mathbf{T}_0^{-1}\Lambda\mathbf{T}_0^{-1}\mathbf{Q}^{\frac{1}{2}}\right). \qquad (3.11)$$

4 Superiority Conditions

In this section, we provide necessary and sufficient conditions for the $\mathcal{AMSE}$ superiority of the $\hat{\beta}_{sr}^{(k)}$ over $\hat{\beta}_{sr}$. From (3.10) and (3.11), the difference $\delta = \mathcal{AMSE}(\hat{\beta}_{sr}) - \mathcal{AMSE}(\hat{\beta}_{sr}^{(k)})$ is given by

$$\delta = tr\left(\mathbf{Q}^{\frac{1}{2}}[\mathbf{T}_0^{-1}\Lambda\mathbf{T}_0^{-1} + \mathbf{Q}^{\frac{1}{2}}\mathbf{T}_k^{-1}\Lambda\mathbf{T}_k^{-1}]\mathbf{Q}^{\frac{1}{2}}\right) - k^2\beta^\top\mathbf{T}_k^{-2}\beta. \qquad (4.1)$$

Let $\mathcal{ADR}_Q(\hat{\beta}_{sr}^{(k)}) = \mathbf{Q}^{\frac{1}{2}}\mathbf{T}_k^{-1}\Lambda\mathbf{T}_k^{-1}\mathbf{Q}^{\frac{1}{2}}$ be the matrix form of $\mathcal{ADR}_Q(\hat{\beta}_{sr}^{(k)})$ then we have the following lemma.

Lemma 4.1. $\mathcal{ADR}_Q(\hat{\beta}_{sr}^{(k)})$ is consistently smaller than $\mathcal{ADR}_Q(\hat{\beta}_{sr})$ therefore, the following inequality always holds for arbitrary $k > 0$

$$\Delta = \mathcal{ADR}_Q(\hat{\beta}_{sr}) - \mathcal{ADR}_Q(\hat{\beta}_{sr}^{(k)}) > 0.$$

Proof. Let $\mathbf{Q} = \mathbf{I}_p$, from $\mathbf{T}_k - \mathbf{T}_0 = k\mathbf{I}_p > 0$, we have $\mathbf{T}_0^{-1} > \mathbf{T}_k^{-1}$, since $\Lambda > 0$ then $\mathbf{T}_0^{-1}\Lambda\mathbf{T}_0^{-1} > \mathbf{T}_k^{-1}\Lambda\mathbf{T}_k^{-1}$ and this completes the proof. □

Theorem 2. Given $\hat{\beta}_{sr}^{(k)}$ under stochastic linear combination in (2.8), the $\mathcal{AMSE}$ difference δ is non negative if only if

$$0 \le k \le \frac{1}{\sqrt{\beta^\top\mathbf{T}_k^{-1}\Delta^-\mathbf{T}_k^{-1}\beta}}.$$

Proof. Here we prove the necessary and sufficient conditions for the $\mathcal{AMSE}_Q$ difference Δ^*. Using Lemma (4.1) and Lemmas (7.1)-(7.2) in Appendix we can write

$$
\begin{aligned}
\Delta^* &= \mathcal{AMSE}_Q(\hat{\beta}_{sr}) - \mathcal{AMSE}_Q(\hat{\beta}_{sr}^{(k)}) > 0 \\
&\Leftrightarrow \mathcal{ADR}_Q(\hat{\beta}_{sr}) - \mathcal{ADR}_Q(\hat{\beta}_{sr}^{(k)}) - k^2 \beta^\top \mathbf{T}_k^{-1} \mathbf{T}_k^{-1} \beta > 0 \\
&\Leftrightarrow \Delta - k^2 \beta^\top \mathbf{T}_k^{-1} \mathbf{T}_k^{-1} \beta > 0 \\
&\Leftrightarrow k^2 \beta^\top \mathbf{T}_k^{-1} \Delta^+ \mathbf{T}_k^{-1} \beta < 1,
\end{aligned}
$$

which completes the proof. □

4.1 Finding k

Now we need to find a k which can make $\hat{\beta}_{sr}^{(k)}$ superior to the $\hat{\beta}_{sr}$ in the sense of $\mathcal{AMSE}$. Based on Ozkale (2009), we suggest k such that $\Delta > 0$. After some algebraic calculations we can write Δ as

$$
\begin{aligned}
\Delta = \mathbf{Q}^{\frac{1}{2}} \mathbf{T}_k^{-1} \{ k^2 \mathbf{T}_0^{-1} \mathbf{D} \mathbf{T}_0^{-1} &+ k^2 \sigma^2 [\mathbf{X}^\top (\mathbf{I}_n - \mathbf{S})^2 \mathbf{X} + \mathbf{R}^\top \mathbf{V}^{-1} \mathbf{R}] + k(\mathbf{T}_0^{-1} \mathbf{D} + \mathbf{D} \mathbf{T}_0^{-1}) \\
&+ 2k\sigma^2 [\mathbf{X}^\top (\mathbf{I}_n - \mathbf{S})^2 \mathbf{X} + \mathbf{R}^\top \mathbf{V}^{-1} \mathbf{R}] \mathbf{T}_0^{-1} - k^2 \beta^\top \beta \} \mathbf{T}_k^{-1} \mathbf{Q}^{\frac{1}{2}}.
\end{aligned}
$$

We note that the matrices $\mathbf{T}_0^{-1} \mathbf{D} \mathbf{T}_0^{-1}$, $[\mathbf{X}^\top (\mathbf{I}_n - \mathbf{S})^2 \mathbf{X} + \mathbf{R}^\top \mathbf{V}^{-1} \mathbf{R}]$, $(\mathbf{T}_0^{-1} \mathbf{D} + \mathbf{D} \mathbf{T}_0^{-1})$ and $[\mathbf{X}^\top (\mathbf{I}_n - \mathbf{S})^2 \mathbf{X} + \mathbf{R}^\top \mathbf{V}^{-1} \mathbf{R}] \mathbf{T}_0^{-1}$ are positive definite (p.d). Therefore, using Farebrother (1976) we have that Δ is positive definite if $2k\sigma^2 [\mathbf{X}^\top (\mathbf{I}_n - \mathbf{S})^2 \mathbf{X} + \mathbf{R}^\top \mathbf{V}^{-1} \mathbf{R}] \mathbf{T}_0^{-1} - k^2 \beta^\top \beta$ is positive semi definite. Thus a sufficient condition for $\hat{\beta}_{sr}^{(k)}$ to be superior over $\hat{\beta}_{sr}$ is

$$
0 < k < 2\sigma^2 \left(\beta^\top \mathbf{T}_0 [\mathbf{X}^\top (\mathbf{I}_n - \mathbf{S})^2 \mathbf{X} + \mathbf{R}^\top \mathbf{V}^{-1} \mathbf{R}]^{-1} \beta \right)^{-1}. \tag{4.2}
$$

To facilitate the computation, we suggest to select the median of the range (4.2). We need only to substitute the separate estimates for the unknowns and obtain the estimate of k. Let $y^* = (\mathbf{I}_n - \mathbf{S})y$, $\mathbf{X}^* = (\mathbf{I}_n - \mathbf{S})\mathbf{X}$ and $\mathbf{W}^* = (\mathbf{I}_n - \mathbf{S})\mathbf{W}$ then using methods of Fuller (1987) an estimate of $\mathbf{X}^*$ can be derived as $\hat{\mathbf{X}}^* = \mathbf{W}^* + \hat{\sigma}^{2*} e^* \hat{\beta}^\top \Sigma$ and hence $\hat{\mathbf{X}} = (\mathbf{I}_n - \mathbf{S})^- \mathbf{X}^*$, where $e^* = Y^* - \mathbf{X}^* \hat{\beta}$ and $\hat{\sigma}^{2*} = \hat{\sigma}^2 + \hat{\beta}^\top \Sigma \hat{\beta}$. Now, replacing σ^2, β and $\mathbf{X}$ respectively, with their estimates $\hat{\sigma}^2$, $\hat{\beta}$ and $\hat{\mathbf{X}}$, a practical solution for k can be suggested as

$$
\hat{k} = \frac{\hat{\sigma}^2}{\hat{\beta}^\top \hat{\mathbf{T}}_0 [\hat{\mathbf{X}}^\top (\mathbf{I}_n - \mathbf{S}) \hat{\mathbf{X}} + \mathbf{R}^\top \mathbf{V}^{-1} \mathbf{R}]^{-1} \hat{\beta}}. \tag{4.3}
$$

where $\hat{\mathbf{T}}_0 = \hat{\mathbf{X}}^\top (\mathbf{I}_n - \mathbf{S}) \hat{\mathbf{X}} + \mathbf{R}^\top \mathbf{V}^{-1} \mathbf{R}$.

4.2 Estimating Smoothing Parameter λ

The cross-validation or the generalized cross-validation, which might be used to estimate the tuning parameter λ can be adapted here. The GCV has been applied for obtaining the optimal bandwidth of the kernel smoother in smoothing spline (Craven and Wahba (1979)) as well as partial linear models and also for obtaining the optimal ridge parameter in a ridge regression model (Golub *et al.* (1979)). For this let $\hat{\beta}_{sr(i)}^{(k)}$ and $\hat{\mathbf{f}}_{sr(i)}^{(k)}$ be the RSRE*s* at $k = \hat{k}$ in (4.3), with ith point Y_i omitted. The argument be that if λ is a good choice then the ith component $[\mathbf{W}\hat{\beta}_{sr(i)}^{(k)} + \mathbf{N}\hat{\mathbf{f}}_{sr(i)}^{(k)}]_i$ of $\mathbf{W}\hat{\beta}_{sr(i)}^{(k)} + \mathbf{N}\hat{\mathbf{f}}_{sr(i)}^{(k)}$ should be a good predictor of Y_i. Therefore the estimate of λ is the minimizer of

$$GCV(\lambda) = n^{-1} \sum_{i=1}^{n} \left(Y_i - [\mathbf{W}\hat{\beta}_{sr(i)}^{(k)} + \mathbf{N}\hat{\mathbf{f}}_{sr(i)}^{(k)}]_i\right)_{k=\hat{k}}^{2}.$$

Recently Roozbeh (2018) and Amini and Roozbeh (2015) have used the GCV criterion for selecting the optimal values of both ridge and bandwith parameters (k and λ) simultaneously, in the presence of multicollinearity for the semiparametric regression models. We can also apply the GCV method to select the optimal bandwidth λ and k simultaneously, which minimizes the following GCV function

$$GCV(\lambda, k) = n^{-1} \sum_{i=1}^{n} \left(Y_i - [\mathbf{W}\hat{\beta}_{sr(i)}^{(k)} + \mathbf{N}\hat{\mathbf{f}}_{sr(i)}^{(k)}]_i\right)^{2}.$$

5 Simulation

In this section, we carry out a simulation study to demonstrate the finite sample performance of the proposed procedures. Our sampling experiment consists of different values of k, i.e., $k = 0, 4, 8, 16, 20$. The data are generated from the model (2.1). For this we assume that $\beta = (1, 2, -2, -5, 4)$ and $\mathbf{X} = N_5(\mu, \Sigma_x)$ with

$$\mu^\tau = [2, 1, -1, 3, 4], \quad \Sigma_x = \begin{pmatrix} 1.90 & 1.79 & 2.25 & 1.50 & 1.22 \\ 1.79 & 1.80 & 1.92 & 1.25 & 1.00 \\ 2.25 & 1.92 & 4.20 & 2.53 & 2.12 \\ 1.50 & 1.25 & 2.53 & 2.4 & 1.38 \\ 1.22 & 1.00 & 2.12 & 1.38 & 2.20 \end{pmatrix}, \quad (5.1)$$

and $f(t_i) = sin2\pi t_i$, where $t_i \sim U(0, 1)$. The random errors ϵ and $\mathbf{U}$ are generated from $N(0, 0.5)$ and $N_5(0, \Sigma)$ receptively. We consider two cases: (I) $\Sigma = \text{diag}(0.05,$
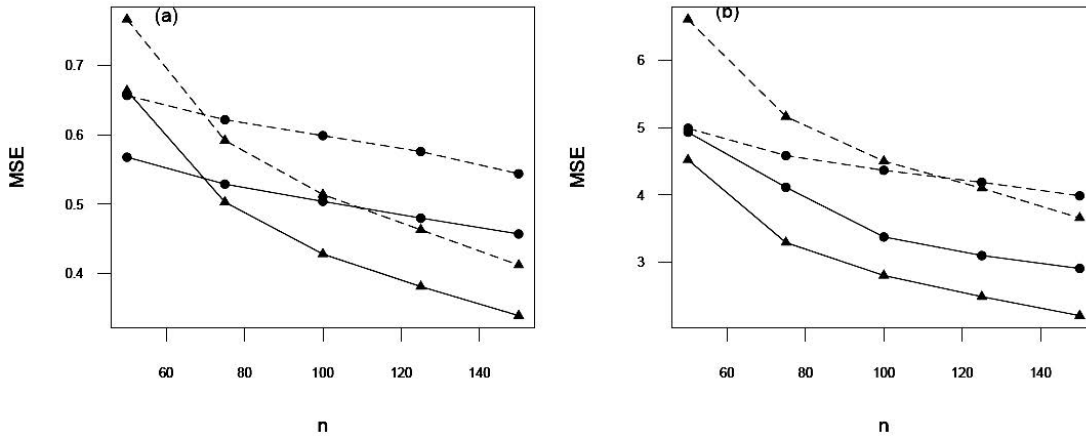
Figure 1: (a) The MSE_{β^*} of $\hat{\beta}_{sr}^{(k)}$ (solid bullet line), and MSE_{f^*} of $\hat{\mathbf{f}}_{sr}^{(k)}$ (solid triangular line) $\hat{\beta}_{sr}^{(k)}$, The MSE_{β^*} of $\tilde{\beta}_{sr}^{(k)}$ (dashed bullet line), and MSE_{f^*} of $\tilde{\mathbf{f}}_{sr}^{(k)}$ (dashed triangular line) vs sample size for case (I). (b) Same as part (a) for case (II).

0.05,0.05, 0.05,0.05) and (II) $\Sigma = \text{diag}(0.15,0.15,0.15,0.15,0.15)$. For the linear restriction and covariance matrix of **e**, we consider the pre-specified matrix **R** and covarince matrix **V** as

$$\mathbf{R} = \begin{pmatrix} 1 & 5 & -3 & -1 & -1 \\ -2 & -1 & 0 & -2 & 3 \\ 1 & 2 & 1 & 3 & -2 \\ 1 & 2 & 1 & 3 & -2 \\ 4 & -1 & 2 & 2 & 0 \end{pmatrix}, \qquad \mathbf{V} = \mathbf{I}_5.$$

The generalized cross-validation, $GCV(\lambda)$, which might be used to estimate the tuning parameter λ is adapted here. The simulation study was conducted using the R software. The number of simulated realizations with a sample size of 50 is 5000. First, we estimate the $\hat{\beta}_{sr}^{(k)}$ and naive stochastic restricted ridge estimator which is denoted by $\tilde{\beta}_{sr}^{(k)}$ (stochastic restricted ridge estimators are obtained by ignoring measurement errors in **X**) with their corresponding estimates of non-parametric terms and next we compute their respective estimated MSE values which is defined by

$$MSE_{\beta^*} = \frac{1}{5000} \sum_{j=1}^{5000} \sum_{l=1}^{5} (\beta_{jl}^* - \beta)^2,$$
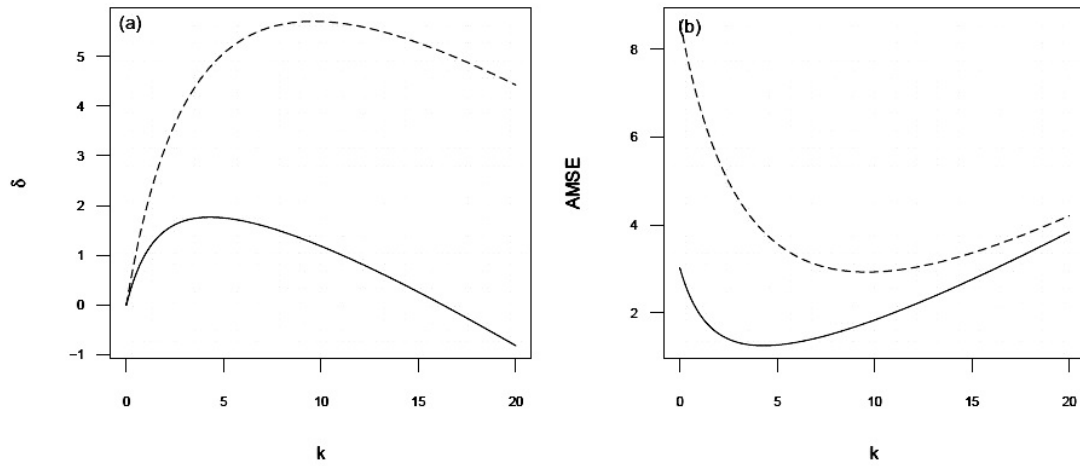
Figure 2: (a) The values of δ for case (I) (solid line) and for case (II) (dashed line) vs k.
(b) The values of $\mathcal{AMSE}$ for case (I) (solid line) and for case (II) (dashed line) vs k.

$$MSE_{f^*} = \frac{1}{5000} \sum_{j=1}^{5000} \sum_{i=1}^{n} (f^*(t_i) - f(t_i))^2,$$

respectively. For a different sample size, the MSEs of $\hat{\beta}_{sr}^{(k)}$ and $\tilde{\beta}_{sr}^{(k)}$ and MSEs of their corresponding non-parametric component estimates are plotted in Figure (1). It can be seen that for the two case measurement error model as the sample size increases the estimated MSE values of the $\hat{\beta}_{sr}^{(k)}$ and $\tilde{\beta}_{sr}^{(k)}$ decrease in general. Moreover, we find out that for all cases, the $\hat{\beta}_{sr}^{(k)}$ has smaller estimated MSE values than the $\tilde{\beta}_{sr}^{(k)}$.

In Table (1), we compute the $\hat{\beta}_{sr}^{(k)}$ for cases (I) and (II). We numerically estimate the $\mathcal{ADB}^\top \mathcal{ADB}$, $\mathcal{ADR}$, $\mathcal{AMSE}$ and $MSE_f = \frac{1}{n} \sum_{i=1}^{n} [\hat{\mathbf{f}}_{sr}^{(k)}(t_i) - \mathbf{f}(t_i)]^2$ for different values of k with 5000 samples of size 50. The ratio of the largest eigenvalue to the smallest eigenvalue of matrix $\mathbf{T}_0$ is $\lambda_5/\lambda_1 = 228.167$ which implies the existence of multicollinearity in the data set. In Figure (2) panel (a) the δ for case (I) and (II) are plotted with solid and dashed lines respectively, while panel (b) shows the $\mathcal{AMSE}$ of $\hat{\beta}_{sr}^{(k)}$ for case (I) and case (II) versus ridge parameter k. As it can be seen from Table (1) and Figure (2), for the two cases the δ increases ($\mathcal{AMSE}$ decreases) at first and then decreases (increases), which provides a reason for the assertion in (4.2). Furthermore, the maximum of δ

Table 1: The finite sample performance of the $\hat{\beta}_{sr}^{(k)}$ for cases (I) and (II).The standard errors are in parentheses.

k	0.00		4		8	
Case	*I*	*II*	*I*	*II*	*I*	*II*
$\hat{\beta}_1$	1.073(0.110)	0.583(0.185)	1.382(0.056)	1.359(0.122)	1.406(0.038)	1.241(0.092)
$\hat{\beta}_2$	2.948(0.099)	3.359(0.178)	2.502(0.061)	2.805(0.130)	2.297(0.035)	2.260(0.088)
$\hat{\beta}_3$	-2.088(0.108)	-1.851(0.180)	-2.120(0.065)	-2.027(0.127)	-2.074(0.037)	-2.034(0.093)
$\hat{\beta}_4$	-5.029(0.111)	-5.578(0.175)	-4.655(0.054)	-5.138(0.118)	-4.359(0.034)	-4.799 (0.090)
$\hat{\beta}_5$	4.044(0.109)	4.283(0.183)	3.831(0.051)	3.921(0.121)	3.649(0.310)	3.631(0.089)
$\|\mathcal{ADB}(\hat{\beta}_{sr}^{(k)})\|^2$	0	0	0.494	0.262	1.182	0.837
$\mathcal{ADR}_Q(\hat{\beta}_{sr}^{(k)})$	3.024	8.634	0.767	3.731	0.364	2.153
$\mathcal{AMSE}$	3.024	8.634	1.261	3.993	1.547	2.991
δ	0	0	1.762	4.641	1.476	5.643
$MSE_{\hat{f}}$	0.123	0.831	0.002	0.046	0.032	0.006
k	12		16		20	
Case	*I*	*II*	*I*	*II*	*I*	*II*
$\hat{\beta}_1$	1.377(0.029)	1.217(0.075)	1.334(0.024)	1.173(0.064)	1.287(0.021)	1.124(0.056)
$\hat{\beta}_2$	2.153(0.031)	2.468(0.079)	2.036(0.026)	2.348(0.061)	1.936(0.019)	2.242(0.051)
$\hat{\beta}_3$	-2.012(0.028)	-2.014(0.080)	-1.947(0.023)	-1.985(0.065)	-1.881(0.021)	-1.951(0.058)
$\hat{\beta}_4$	-4.104(0.029)	-4.511(0.070)	-3.881(0.025)	-4.262(0.066)	-3.682(0.024)	-4.042(0.055)
$\hat{\beta}_5$	3.471(0.023)	3.381(0.074)	3.315(0.022)	3.163(0.060)	3.172(0.020)	2.971(0.056)
$\|\mathcal{ADB}(\hat{\beta}_{sr}^{(k)})\|^2$	1.961	1.593	2.816	2.462	3.725	3.401
$\mathcal{ADR}_Q(\hat{\beta}_{sr}^{(k)})$	0.221	1.438	0.152	1.048	0.114	0.807
$\mathcal{AMSE}$	2.183	3.031	2.969	3.510	3.839	4.209
δ	0.840	5.602	0.0548	5.124	-0.815	4.424
$MSE_{\hat{f}}$	0.154	0.083	0.351	0.219	0.614	0.359

(minimum of the $\mathcal{AMSE}$) is obtained when k equals (4.3) which are approximately equal to 4.174 and 7.924 for case (I) and (II) respectively. For estimating the nonlinear part, we simulate the response from our model for $n = 300$ again. In Figure (3), the nonparametric part of the model for case (I) and (II) are plotted in panels (a) and (b) using $\hat{\beta}_{sr}^{(k)}$ and $\tilde{\beta}_{sr}^{(k)}$, respectively. By an increase in the variance of measurement error the nonparametric estimates of $\tilde{\beta}_{sr}^{(k)}$ outperform $\mathbf{f}(t)$.

6   Real Data

To motivate the problem of stochastic linearly constrained estimation in the semiparametric regression model, we consider the Egyptian pottery data set. This data set was collected from an extensive archaeological survey of pottery production and distribution in the ancient Egyptian city of Al-Amarna. The data consist of measurements of chemical contents (mineral elements) made on many samples of pottery using two different techniques, NAA and ICP (see Smith *et al.* (1988)) for description of techniques).
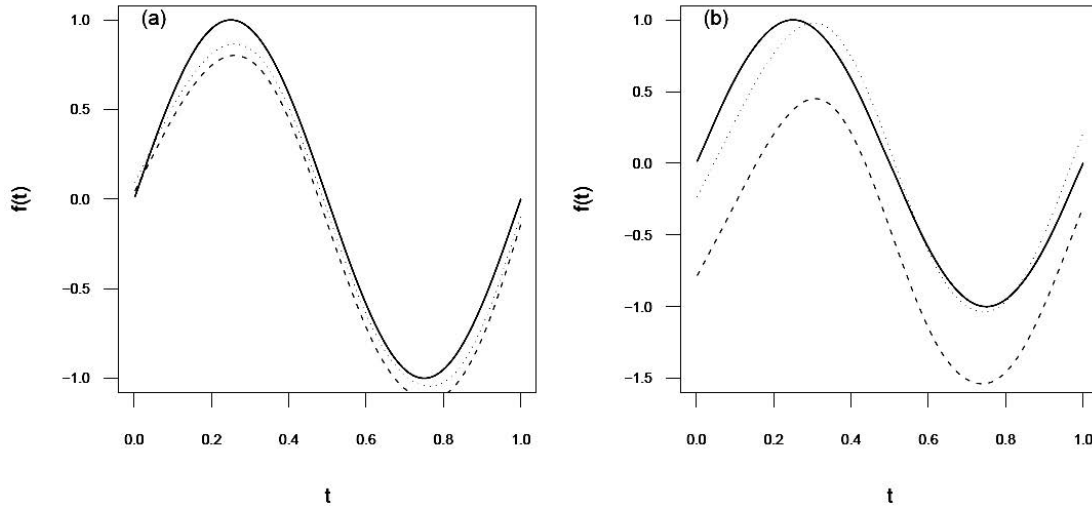
Figure 3: (a) For case (I): The estimates of $\mathbf{f}(.)$, $sin2\pi t$ (solid curve), the $\hat{\mathbf{f}}_{sr}^{(k)}$ (dash-dotted curve) and $\tilde{\mathbf{f}}_{sr}^{(k)}$ (dotted curve), with size $n = 300$, bandwidth $\lambda = 0.041$ and $k = 4.174$.(b) For case (II): The estimates of $\mathbf{f}(.)$ with size $n = 300$, bandwidth $\lambda = 0.277$ and $k = 7.924$.

The set of pottery was collected from different locations around the city. The group structure among the objects arises from two main sources, fabric code and location of pottery. Consequently, according to this group structure, the selected vessels have been divided into 28 groups. In each group, there are different numbers of vessels from the same fabric code and provenance, which can essentially be regarded as replicated observations. Ghapani and Babadi (2016) and Rasekh (2001) analyzed the same data to find anomalous observations in full parametric linear measurement error models using the methods of shift outlier model and local influence approach, respectively. They considered Na measured with NAA as response variable versus mineral elements Na, Al,K, V, Cr and Mn measured by ICP as predictor variables. Using the same data we suspected in the type of relation (linear or non linear) between dependent variable NAA versus explanatory variable Mn. By Yatchew (2003), the test statistic for the null hypothesis that the regression function has the parametric form, i.e., $H_0 : \mathbf{f}(t) = h(t, \beta)$ for a known function $h(.)$, against the nonparametric alternative is

$$Z_0 = \frac{n^{1/2}(s_{res}^2 - s_{diff}^2)}{s_{diff}^2} \xrightarrow{D} N(0, 1),\tag{6.1}$$

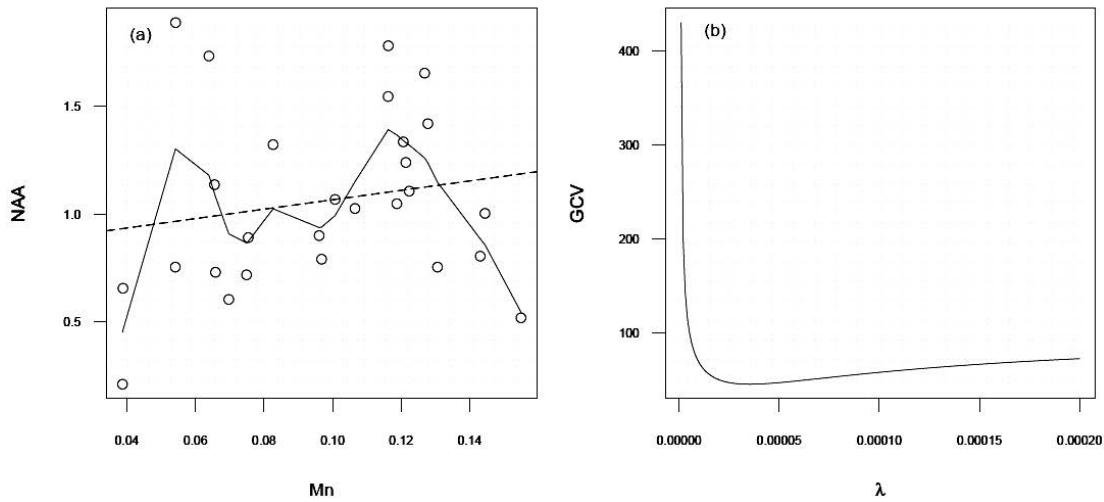Figure 4: (a) NAA versus Mn, linear fit (dashed line) smooth spline fit (solid line). (b).Plot of GCV vs λ.

where s^2_{res} the usual estimate of the residual variance obtained from linear regression and $s^2_{diff} = \frac{1}{2n} \sum_{i=2}^{n} (y_i - y_{i-1})^2$. We consider the average Mn as a nonparametric part (assuming no measurement error for this variable), because it has the largest value of nonparametric significance test statistics among those of other independent variables. Nonparametric significance test of the Mn effect yields the value $Z_0 = 6.723$, which indicates that non linearly relation of the Mn variable to dependent variable is significant (see Figure (4)(a)). Finally the semiparametric linear measurement error model (2.1) with y=NAA, $\mathbf{W}$=[Al,K,V,Cr] and f(Mn) is specified. To estimate Σ, since there are the replicated observations $\mathbf{W}_{ij} = \mathbf{X}_i + \mathbf{U}_{ij}, j = 1, \ldots, m_i$, an unbiased method of moments estimate for Σ can be considered as

$$\hat{\Sigma} = \frac{\sum_{i=1}^{n} \sum_{j=1}^{m_i} (\mathbf{W}_{ij} - \bar{\mathbf{W}}_i)(\mathbf{W}_{ij} - \bar{\mathbf{W}}_i)^{\top}}{\sum_{i=1}^{n} (m_i - 1)}.$$

To investigate the performance of the proposed stochastic restricted estimators, from previous studies (see, Ghapani and Babadi (2016)) we consider the parametric
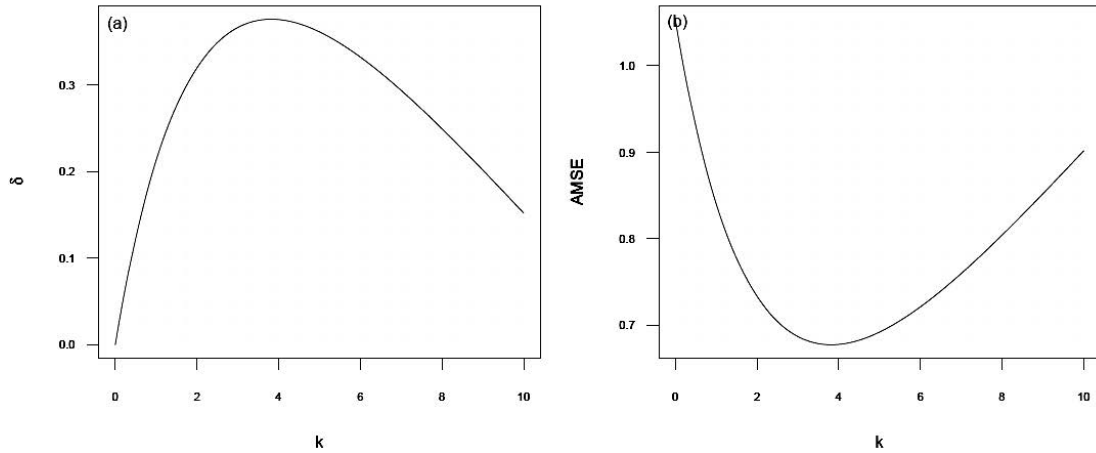
Figure 5: (a) The values of δ vs k. (b) The values of $\mathcal{AMSE}$ vs k for Egyptian pottery data.

restriction $r = \mathbf{R}\beta + e$, where

$$
\mathbf{R} = \begin{pmatrix} 1 & 0 & -3 & 0 & 0 \\ 0 & -1 & 0 & 8 & 0 \\ 0 & 0 & 0 & 1 & 6 \end{pmatrix}.
$$

We test the linear hypothesis $H_0 : r = \mathbf{R}\beta$ in the framework of our proposed semiparametric model. The test statistic for H_0, given our observations, is

$$
\chi^2_{[rank(\mathbf{R})]} \simeq (\mathbf{R}\hat{\beta} - r)^\top (\mathbf{R}\hat{\Sigma}_\beta \mathbf{R}^\top)^{-1}(\mathbf{R}\hat{\beta} - r),
$$

where $\hat{\Sigma}_\beta = (1 + 1/2m)\hat{\sigma}^2(\mathbf{W}^\top\mathbf{W})$. Thus we conclude that the null hypothesis H_0 is not rejected and this can be regarded as a prior information for β. The ratio of the largest eigenvalue of smallest eigenvalue for new design matrix in model is approximately $\lambda_4/\lambda_1 = 4424.752$ and so, there exists a high multicollinearity between the columns of $\mathbf{T}_0$. Now in order to overcome the multicollinearity for better performance of the estimators, we use the proposed ridge estimators for our model. From Figure (5) we can select the value of k to be 4 and from Figure (4)(b) the minimum of GCV occurred at $\lambda = 0.0003$. The stochastic restricted ridge estimator for the proposed value of ridge parameter are given in Table (2). As it can be seen, $\hat{\beta}_{sr}^{(k=4)}$ is the best estimator for linear part of the SLMeM in the sense of risk. Finally, we estimated the non-parametric effect

Table 2: The restricted ridge estimator (RRE) for Egyptian pottery dataset. The standard errors are in parentheses.

k	0	2	4	6	8	10
$\hat{\beta}_1$	-0.399(0.018)	-0.085(0.018)	0.042(0.016)	0.108(0.015)	0.147(0.013)	0.172(0.011)
$\hat{\beta}_2$	-0.014(0.806)	-0.048(0.666)	-0.055(0.561)	-0.052(0.490)	-0.047(0.438)	-0.041(0.397)
$\hat{\beta}_3$	1.130(0.063)	1.145(0.061)	1.048(0.057)	0.983(0.050)	0.932(0.047)	0.891(0.045)
$\hat{\beta}_4$	0.040(0.011)	0.039(0.010)	0.031(0.009)	0.022(0.009)	0.012(0.009)	0.003(0.008)
$\hat{\beta}_5$	0.001(0.624)	0.001(0.471)	-0.002(0.397)	-0.003(0.346)	-0.005(0.310)	-0.008(0.280)
$\|\mathcal{ADB}(\hat{\beta}_{sr}^{(k)})\|^2$	0	0.065	0.200	0.356	0.513	0.662
$\mathcal{ADR}_Q(\hat{\beta}_{sr}^{(k)})$	1.053	0.667	0.477	0.364	0.291	0.239
$\mathcal{AMSE}$	1.053	0.733	0.677	0.721	0.804	0.901
δ	0	0.320	0.375	0.332	0.249	0.152
$MSE_{\hat{f}}$	0	0.320	0.081	3.724	13.109	27.964

$f(Mn)$ after estimating the linear part by $\hat{\beta}_{sr}^{(k=4)}$ by a smooth spline method in Figure (6), i.e., we used smooth spline fit to regress $f_{mn} = \text{NAA-}\mathbf{W}\hat{\beta}_{sr}^{(k=4)}$ on Mn with $\lambda = 0.092$.

7   Conclusion

Ridge estimation of statistical models with measurement errors seems to be overlooked in the literature. In this paper, we derive the *RSRE* based penalized corrected likelihood for the vector of parameters in a SLMeM, when additional stochastic linear restrictions on the parameter vector are assumed to hold. We constructed *RSRE* for the parametric, nonparametric cases by taking the measurement errors into account, and showed that they were consistent and asymptotically normal. The performance of *RSRE* over *SRE* based on ridge parameter k was proved. Applying kernel smoothing and cross-validation methods, we estimated the nonlinear functions of the proposed model. We see that the δ increases ($\mathcal{AMSE}$ decreases) at first and then decreases (increases), which provides a reason for the assertion in (4.2). Furthermore, the maximum of δ (minimum of the $\mathcal{AMSE}$) is obtained when k equals the median range of (4.2) i.e., $\hat{k} = \frac{\hat{\sigma}^2}{\hat{\beta}^\top \hat{\mathbf{T}}_0 [\hat{\mathbf{X}}^\top (\mathbf{I}_n - \mathbf{S})\hat{\mathbf{X}} + \mathbf{R}^\top \mathbf{V}^{-1}\mathbf{R}]^{-1}\hat{\beta}}$ which is approximately equal to 4.147, 7.947 and 3.985 respectively for case *I*, case *II* (in simulation data) and real data. As a final point, there was a clear difference between the *MSE*s of the naive *RSRE*s and those of the proposed estimators. We believe that this difference is caused by the measurement errors.
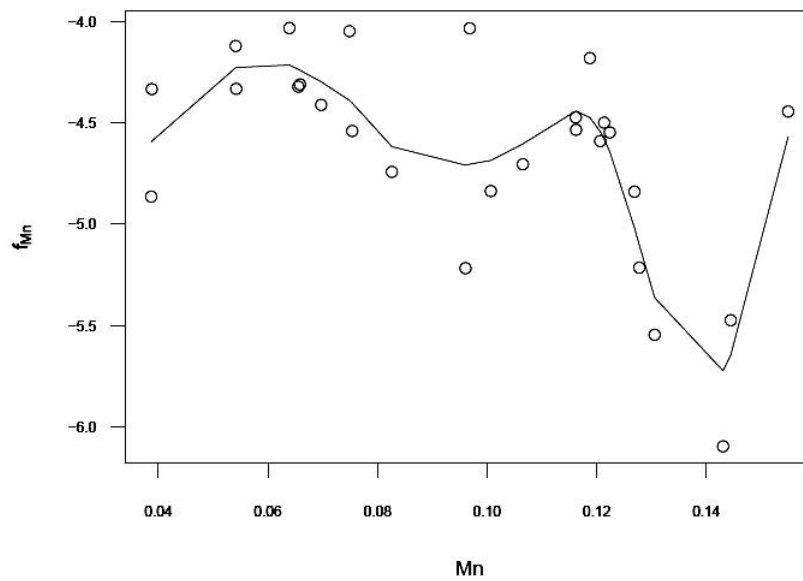
Figure 6: Estimation of the nonparametric part by smooth spline method with (λ = 0.092).

Appendix

Lemma 7.1. If $\hat{\beta}_2$ is superior to $\hat{\beta}_1$ with respect to $\mathcal{AMSE}$, then it is superior to $\hat{\beta}_1$ with respect to MSE and vice versa (Farebrother , 1976).

Lemma 7.2. Let $\mathbf{A}$ be a symmetric positive definite $n \times n$ matrix, α an $n \times 1$ vector. Then $\mathbf{A} - \alpha\alpha^\top$ is nonnegative definite if and only if $\alpha^\top \mathbf{A}^- \alpha \leq 1$ is satisfied, where $\mathbf{A}^-$ is the generalized inverse of $\mathbf{A}$ (Rao *et al.* , 2008).

Acknowledgements

References

Akdeniz, Duran E., Hardle, W. K. and Osipenko M. (2012), Difference-based ridge and Liu type estimators in semiparametric regression models. *Journal of Multivariate Analysis*, **105**, 164-175.

Akdeniz, F. and Tabakan, G. (2009), Restricted ridge estimators of the parameters in semiparametric regression model. *Communications in StatisticsTheory and Methods*, **38**, 1852-1869.

Amini, M. and Roozbeh, M. (2015), Optimal partial ridge estimation in restricted semi-parametric regression models. *Journal of Multivariate Analysis*, **136**, 26-40.

Aydn, D., Yuzbasi, B. and Ahmed, S. E. (2016), Modified Ridge Type Estimator in Partially Linear Regression Models and Numerical Comparisons. *Journal of Computational and Theoretical Nanoscience*, **13**(10), 7040-7053.

Belsley, A. D. (1991), *Conditioning diagnostics*. Wiley, New York.

Cheng, C., Van Ness and J. W. (1999), *Statistical Regression with Measurement Error*. Arnold, London.

Craven, P. and Wahba, G. (1979), Smoothing Noisy Data with Spline Functions. *Numerische Mathematik*, **31**, 377-403.

Durbin, J. (1953), A note on regression when there is extraneous information about one of the coefficients. *Journal of the American Statistical Association*, **48**, 799-808.

Farebrother, R. (1976), Further results on the mean square error of ridge regression. *ournal of the Royal Statistical Society. Series B (Methodological)*, **38**, 248-250.

Fuller, W. A.(1987), *Measurement Error Models*. Wiley, New York.

Ghapani, F. and Babadi, B. (2016), A New Ridge Estimator in Linear Measurement Error Model with Stochastic Linear Restrictions. *Journal of Iranaian Statistical Society*, **15**, 87-103.

Golub, G., Heath, M. and Wahba, G. (1979), Generalized cross validationas a method for choosing a good ridge parameter. *Technometrics*, **21**, 215-223.

Green, P. J. and Silverman, B. W. (1994), *Nonparometric Regression and Generalized Linear Models*. London, Chapman and Hall.

Grob, J. (2003), Restricted ridge estimation. *Statistics Probability Letters*, **65**, 57-64.

Hoerl, A. E. and Kennard, R. W. (1970), Ridge regression: biased estimation for nonorthogonal problems. *Thechnometrics*, **12**, 69-82.

Hu, H. (2005), Ridge estimation of a semiparametric regression model. *Journal of Computational and Applied Mathematics*, **176**, 215-222.

Jibo, Wu. and Y. Asar (2017), A weighted stochastic restricted ridge estimator in partially linear model. *Communications in Statistics-Theory and Methods*, **46**(18), 9274-9283.

Li, Y. L. and Yang Hu. (2011), A new ridge-type estimator in stochastic restricted linear regression. *Statistics*, **45**, 123-130.

Liang H, Hardle WK. and Carroll R.J . (1999), Estimation in a semiparametric partially linear errors-in-variables model, *The Annals of Statistics* , **27**, 1519-1535.

Nagelkerke, N. J. (1992), *Maximum likelihood estimation of functional relationships*. Springer, Berlin

Nakamura, T. (1990), Corrected score functions for error in variables models: methodology and application to generalized linear models. *Biometrika*, **77**, 127-137.

Ozkale, M. R. (2009), A stochastic restricted ridge regression estimator. *Journal of Multivariate Analysis*, **100**, 1706-1716.

Rao, C. R., Toutenburg H., Shalabh. H. C. (2008), *Linear models and generslizations*. Springer, Berlin.

Rasekh, A. (2001), Ridge estimation in functional measurement error models. In *Annales de l'ISUP* **45**(2-3), 47-59. Institut de statistique de l'Université de Paris.

Roozbeh, M. (2015), Shrinkage ridge estimators in semiparametric regression models. *Journal of Multivariate Analysis*, **136**, 56-74.

Roozbeh, M. and Arashi M. (2013), Feasible ridge estimator in partially linear models, *Journal of Multivariate Analysis*, **116**, 35-44.

Roozbeh, M. (2018), Optimal QR-based estimation in partially linear regression models with correlated errors using GCV criterion. *Journal of Computational Statistics and Data Analysis*, http://dx.doi.org/10.1016/j.csda.08.002.

Saleh, A. M. E. (2014), A ridge regression estimation approach to the measurement error model. *Journal of Multivariate Analysis*, **123**, 68-84.

Siray, S. and Torker, S. (2014), Restricted two parameter ridge estimators. *Australian and New Zealand Journal of Statistics*, **55**(4), 455-469.

Smith, D. M., Hart, F. A., Symond, R. D., and Walsh, J. N. (1987), Analysis of Roman pottery from Colchester by inductively coupled plasma spectrometry. *Science and archaeology Glasgow*, **196**, 41-55.

Swamy, P. A. V. B., Mehta, J. S., and Rappoport, P. N. (1978), Two methods of evaluating Hoerl and Kennard's ridge regression.*Communications in Statistics-Theory and Methods*, **7**(12), 1133-1155.

Theil, H. (1963), The use of incomplete prior information in regression analysis. *Journal of the American Statistical Association*, **58**, 401-414.

Yatchew, A. (2003), *Semiparametric regression for the applied econometrican*. Cambridge University press, Cambridge.

Y′uzba, B., and Ejaz Ahmed, S. (2016), Shrinkage and penalized estimation in semiparametric models with multicollinear data.*Journal of Statistical Computation and Simulation*, **86**(17), 3543-3561.

Xu-Ping, Z., Fung, W. K., and Bo-Cheng, W. (2002), Estimation in Linear Models with Random Effects and Errors-in-Variables. *Annals of the Institute of Statistical Mathematics*, **54**(3), 595.

Zhong, Z., and Yang, H. (2007), Ridge estimation to the restricted linear model. *Communications in Statistics- Theory and Methods*, **36**(11), 2099-2115.