Stat Papers
DOI 10.1007/s00362-017-0893-9

REGULAR ARTICLE

Generalized difference-based weighted mixed almost unbiased ridge estimator in partially linear models

Fikri Akdeniz[1] · **Mahdi Roozbeh**[2]

Received: 9 March 2016 / Revised: 10 February 2017

Abstract In this paper, a generalized difference-based estimator is introduced for the vector parameter β in partially linear model when the errors are correlated. A generalized difference-based almost unbiased ridge estimator is defined for the vector parameter β. Under the linear stochastic constraint $r = R\beta + e$, a new generalized difference-based weighted mixed almost unbiased ridge estimator is proposed. The performance of this estimator over the generalized difference-based weighted mixed estimator, the generalized difference-based estimator, and the generalized difference-based almost unbiased ridge estimator in terms of the mean square error matrix criterion is investigated. Then, a method to select the biasing parameter k and non-stochastic weight ω is considered. The efficiency properties of the new estimator is illustrated by a simulation study. Finally, the performance of the new estimator is evaluated for a real dataset.

Keywords Difference-based estimator · Generalized ridge estimator · Generalized difference-based weighted mixed almost unbiased ridge estimator · Partially linear model · Weighted mixed estimator

Mathematics Subject Classification Primary 62 G05 · Seconday 62J05 · 62J07

✉ Fikri Akdeniz
fikriakdeniz@gmail.com

Mahdi Roozbeh
mahdi.roozbeh@semnan.ac.ir

1 Department of Mathematics and Computer Science, Çağ University, 33800 Tarsus-Mersin, Turkey

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

Published online: 04 March 2017

1 Introduction

Partially linear models have received considerable attention in statistics and econometrics. They have a wide range of applications. In these models, some of the relations are believed to be of certain parametric form while others are not easily parameterized. Consider the partially linear model (PLM)

$$y_i = x_i'\beta + f(u_i) + \varepsilon_i, \quad i = 1, 2, \ldots, n \tag{1}$$

where $x_i' = (x_{i1}, x_{i2}, \ldots, x_{ip})$ is a vector of explanatory variables, $\beta = (\beta_1, \beta_2, \ldots, \beta_p)'$ is an unknown p-dimensional parameter vector, the u_i are known and nonrandom in some bounded domain $D \subset \Re$, $f(.)$ is an unknown smooth function, and ε_i 's are independent and identically distributed random errors with $E(\varepsilon_i) = 0$, $Var(\varepsilon_i) = \sigma^2$ and are independent of (x_i, u_i).

The explanatory variables are represented separately in two parts: the nonparametric part $(f(u_i))$ and the parametric linear part $(x_i'\beta)$. We shall call $f(u_i)$ the smooth part of the model and assume that it represents a smooth unparametrized functional relationship. The u_i have bounded support, say the unit interval, and have been arranged so that $u_1 \le u_2 \le \cdots \le u_n$. The goal is to estimate the unknown parameter vector β and nonparametric function $f(u)$ from the data $\{y_i, x_i, u_i\}$. In vector/matrix notation, the Model (1) can be written as

$$y = X\beta + f + \varepsilon, \tag{2}$$

where $y = (y_1, \ldots, y_n)'$, $X = [x_1, \ldots, x_n]'$, $f = (f(u_1), \ldots, f(u_n))'$, $\varepsilon = (\varepsilon_1, \ldots, \varepsilon_n)'$.

Partially linear regression models are more flexible than the standard linear regression models since they combine both parametric and nonparametric components when it is believed that the response variable y depends on variable X in a linear way but is nonlinearly related to other independent variable U. Due to its flexibility, PLM has been widely used in econometrics, finance, biology, sociology and so on. In Model (2), Yatchew (1997) concentrated on estimation of the linear component and used differencing to eliminate bias induced from the presence of the nonparametric component. Yatchew's method does not require an estimator of the function $f(.)$ and are often called difference-based estimation procedure, provided that $f(.)$ is differentiable and the $u_i's$ are closely spaced, it is possible to remove the effect of the function $f(.)$ by differencing the data appropriately (Yatchew 2003).

In regression analysis, researchers often encounter the problem of multicollinearity. The least squares estimator performs poorly in the presence of multicollinearity. Condition number is a measure of the presence of multicollinearity. If $X'X$ is ill conditioned with a large condition number, ridge regression estimator (Hoerl and Kennard 1970) can be used to estimate β.

Applying the shrinkage estimators is well-known as an efficient method to solve the problems caused by the multicollinearity. We assume that the condition number of the parametric component is large indicating that a biased estimation procedure is desirable. Its parametric part has the same structural form as the classical methods.

Akdeniz and Tabakan (2009) introduced a ridge estimator for the vector of parameters in a semiparametric regression model when additional linear restrictions on the parameter vector are assumed to hold. A difference-based ridge regression estimator of regression parameters in the partially linear model is given in Tabakan and Akdeniz (2010). The difference-based estimation procedure is optimal in the sense that the estimator of the linear component is asymptotically efficient and the estimator of the nonparametric component is asymptotically minimax rate optimal for the partial linear model (Wang et al. 2011).

In this paper, a generalized difference-based restricted estimator is introduced for the vector parameter β in the partially linear model when the errors are correlated. Under the linear stochastic constraint $r = R\beta + e$, a new generalized difference-based weighted mixed almost unbiased Ridge estimator (GDWMAURE) is proposed. The performance of this estimator over the generalized difference-based weighted mixed estimator (GDWME), the generalized difference-based estimator (GDE), and the generalized difference-based almost unbiased Ridge estimator (GDAURE) in terms of the mean square error matrix (MSEM) criterion is investigated. Then, a method to select the biasing parameter k and non-stochastic weight ω is considered.

The paper is organized as follows. In Sect. 2, the model and difference-based estimator is defined. In Sect. 3, the generalized difference-based estimator is introduced when the errors are correlated. The generalized difference-based weighted mixed estimator and generalized difference-based weighted mixed almost unbiased Ridge estimator of β is given in Sect. 4. The efficiency properties of the generalized difference-based weighted mixed estimator are given in Sect. 5. In Sect. 6, we propose a method to choose k and w. Finally, in Sect. 7 the performance of the new estimator is illustrated by a simulation study and a realdata example. Some conclusion remarks are given in Sect. 8.

2 Difference-based estimator

In this section we use a difference-based technique to estimate the linear regression coefficient vector β. This technique has been used to remove the nonparametric component in semiparametric regression model by various authors (e.g. Yatchew 1997, 2003; Klipple and Eubank 2007) . Consider the following partially linear model

$$y = X\beta + f + \varepsilon, \tag{3}$$

Yatchew (1997) suggested estimating β on the basis of the m-th order differencing equation

$$\sum_{j=0}^{m} d_j y_{k-j} = \left(\sum_{j=0}^{m} d_j x_{k-j}\right)\beta + \left(\sum_{j=0}^{m} d_j f(u_{k-j})\right) + \left(\sum_{j=0}^{m} d_j \varepsilon_{k-j}\right) \quad k = m+1, \ldots, n \tag{4}$$

Now let $d = (d_0, d_1, \ldots, d_m)'$ be a $(m+1)$-vector, where m is the order of differencing and $d_0, d_1, \ldots, d_m$ are differencing weights minimize $\min_{d_0,\ldots,d_m} \delta = \sum_{l=1}^{m} (\sum_{j=0}^{m-l} d_j d_{l+j})^2$ satisfying the conditions

$$\sum_{j=0}^{m} d_j = 0 \quad \text{and} \quad \sum_{j=0}^{m} d_j^2 = 1. \tag{5}$$

Let us define the $(n-m) \times n$ differencing matrix D to have first and last rows $\left[d', 0'_{n-m-1}\right]$, $\left[0'_{n-m-1}, d'\right]$ respectively, with i-th row $\left[0_i, d', 0'_{n-m-i-1}\right]$, $i = 2, \ldots, (n-m-1)$, where 0_r indicates a r-vector with all zero elements.

$$D = \begin{bmatrix} d_0 & d_1 & d_2 & \cdot & \cdot & \cdot & d_m & 0 & 0 & \cdot & \cdot & \cdot & 0 \\ 0 & d_0 & d_1 & d_2 & \cdot & \cdot & \cdot & d_m & 0 & 0 & \cdot & \cdot & 0 \\ \cdot & \cdot & \cdot & & & & & & & & & & \\ \cdot & \cdot & \cdot & & & & & & & & & & \\ \cdot & \cdot & \cdot & & & & & & & & & & \\ 0 & 0 & 0 & \cdot & \cdot & \cdot & & d_0 & d_1 & \cdot & \cdot & d_m & 0 \\ 0 & 0 & 0 & \cdot & \cdot & \cdot & & & d_0 & d_1 & \cdot & \cdot & d_m \end{bmatrix}.$$

Applying the differencing matrix to Model (3), permits direct estimation of the parametric effect. As a result of developments in Roozbeh et al. (2011), it is known that the parameter vector β in (3) can be estimated with parametric efficiency. Since the data have been ordered so that the values of the nonparametric variable(s) are close, the application of the differencing matrix D in Model (3) removes the non-parametric effect in large samples (Yatchew 2000). If $f(.)$ is an unknown function that is the inferential object and has a bounded first derivative, then Df is close to 0, so that by applying the differencing matrix we have

$$Dy = DX\ \beta + Df + D\varepsilon, \tag{6}$$

which is approximately equal to $DX\beta + D\varepsilon$, or

$$\tilde{y} \cong \tilde{X}\beta + \tilde{\varepsilon}, \tag{7}$$

where $\tilde{y} = Dy$, $\tilde{X} = DX$ and $\tilde{\varepsilon} = D\varepsilon$. So that the role of the constraints (5) is now evident (Yatchew 2003, p. 57; Klipple and Eubank 2007). Yatchew (2003) defines a simple differencing estimator of the parameter β in the semiparametric regression model. Thus, standart linear models considerations suggest estimating β by

$$\hat{\beta}_{diff} = \left[(DX)'(DX)\right]^{-1} (DX)'(Dy). \tag{8}$$

This estimator was first proposed in Yatchew (1997). Thus, differencing allows one to perform inferences on β as if there is no nonparametric component f in the Model (7) (see Yatchew 1997).

Then,

$$S^2_{diff} = \frac{1}{n}(Dy - DX\hat{\beta}_{diff})'(Dy - DX\hat{\beta}_{diff})$$
$$= \frac{1}{n}(\tilde{y} - \tilde{X}\hat{\beta}_{diff})'(\tilde{y} - \tilde{X}\hat{\beta}_{diff}), \quad (9)$$

(see Yatchew 2003 p. 71). To account the parameter β in Eq. (7), we introduce the modified estimator of σ^2, defined by

$$\hat{\sigma}^2 = \frac{\tilde{y}'(I-P)\tilde{y}}{tr(D'(I-P)D)}, \quad (10)$$

where *tr*(.) is the trace function for a square matrix and P is the projection matrix and defined by Yatchew (2003)

$$P = \tilde{X}(\tilde{X}'\tilde{X})^{-1}\tilde{X}'.$$

3 Partially linear models with correlated errors

In this section we consider the following partially linear model

$$y = X\beta + f + \varepsilon$$

with $E(\varepsilon) = 0$, and $E(\varepsilon\varepsilon') = \sigma^2 V$. So, $\tilde{\varepsilon} = D\varepsilon$ is a $(n-m)-$vector of disturbances distributed with zero mean and

$$E(\tilde{\varepsilon}\tilde{\varepsilon}') = \sigma^2 DVD' = \sigma^2 V_D, \quad (11)$$

where $V_D = DVD' \neq I_{n-m}$ is a known $(n-m)\times(n-m)$ symmetric positive definite (p.d.) matrix and $\sigma^2 > 0$ is an unknown parameter (see Roozbeh et al. 2011). It is well known that adopting the linear Model (7), the unbiased estimator of β is the following *generalized difference-based estimator* given by

$$\hat{\beta}_{GDE} = (\tilde{X}'V_D^{-1}\tilde{X})^{-1}\tilde{X}'V_D^{-1}\tilde{y}. \quad (12)$$

It is observed from Eq. (12) that the properties of the generalized difference-based estimator of β depends on the characteristics of the information matrix $G = \tilde{X}'V_D^{-1}\tilde{X}$.

The estimate of the σ^2 is

$$\hat{\sigma}^2 = \frac{1}{n}(\tilde{y} - \tilde{X}\hat{\beta}_{GDE})'V_D^{-1}(\tilde{y} - \tilde{X}\hat{\beta}_{GDE}).$$

It is easy to show that

$$s^2 = \frac{1}{n-p}(\tilde{y} - \tilde{X}\hat{\beta}_{GDE})'V_D^{-1}(\tilde{y} - \tilde{X}\hat{\beta}_{GDE})$$

is an unbiased estimator of σ^2. If G: $p \times p$, $p << n - m$ matrix is ill-conditioned with a large condition number, then the $\hat{\beta}_{GDE}$ produces large sample variances. Moreover, some regression coefficients may be statistically insignificant and meaningful statistical inference becomes difficult for the researcher. We assume that the condition number of the G matrix is large indicating that a biased estimation procedure is desirable.

4 Weighted mixed regression and estimation of parameters

The use of prior information in linear regression analysis is well known to provide more efficient estimators of regression coefficients. The available prior information sometimes can be expressed in the form of exact, stochastic or inequality restrictions.

We consider the Model (7):

$$\tilde{y} \cong \tilde{X}\beta + \tilde{\varepsilon}, \quad \tilde{\varepsilon} \sim \left(0, \sigma^2 DVD'\right) = \left(0, \sigma^2 V_D\right). \tag{13}$$

When a set of stochastic linear constraints binding the regression coefficients in a linear regression model is available, Theil and Goldberger (1961) have proposed the method of mixed regression estimation. Their method typically assumes that the prior information in the form of stochastic linear constaints and sample information in the form of observations on the study variable and explanatory vaiables are equally important and therefore receive equal weights in the estimation procedure.

Totally, we do not have exact prior information such as $R\beta = r$, involving estimation of economic relations, industrial structures, production planning, etc. Therefore, stochastic uncertainty occurs in specifying linear programming due to economic and financial studies.

In addition to the sample Model (13), it is supposed that a set of stochastic linear constraints binding the regression coefficients is available in the form of independent prior information:

$$r = R\beta + e, \quad e \sim \left(0, \sigma^2 W\right), \tag{14}$$

where R is a $q \times p$ known matrix with rank$(R) = q$, e is a $q \times 1$ vector of disturbances. W is assumed to be known and positive definite, the $q \times 1$ vector r can be interpreted as a random variable with expectation $E(r) = R\beta$. Therefore, the restriction (14) does not hold exactly but in the mean and we assume r to be known, that is to be a realized value of the random vector, so that all expectations are conditional on r as, for example, $E(\hat{\beta}|r)$ (Rao et al. 2008). In order to take the information (14) into account while constructing estimators $\hat{\beta}$ for β, we require that $E(R\hat{\beta}|r) = r$ (see Toutenburg et al. 2003).

In Model (14), we have assumed $E(ee') = \sigma^2 W$, that is, with the same factor of proportionality σ^2 as occured in the sample model. Therefore, it may some times be more realistic to suppose that $E(ee') = W$. It is also assumed that the random vector ε is stochastically independent of e.

When the sample information given by (13) and prior information is described by (14) are to be assigned not necessarily equal weights on the basis of some extraneous considerations in the estimation of regression parameters, Schaffrin and Toutenburg (1990) have proposed the method of weighted mixed regression estimation. Following their technique, we obtain the generalized difference-based weighted mixed esimator of β. In order to incorporate the restrictions (14) in the estimation of parameters, we minimize

$$(\tilde{y} - \tilde{X}\beta)' V_D^{-1} (\tilde{y} - \tilde{X}\beta) + \omega (r - R\beta)' W^{-1} (r - R\beta), \tag{15}$$

with respect to β. This leads to the following solution for β :

$$\hat{\beta}_{GDWME}(\omega) = \hat{\beta}(\omega) = (G + \omega R' W^{-1} R)^{-1} (\tilde{X}' V_D^{-1} \tilde{y} + \omega R' W^{-1} r), \tag{16}$$

where ω is a non-stochastic and non-negative scalar weight with $0 \leq \omega \leq 1$ ($\omega = 0$ would lead to $\hat{\beta}_{GDE}$). It is seen that a value of ω between 0 and 1 specifies an estimator in which the prior information receives less weight in comparison to the sample information. On the other hand, a value of ω greater than 1 implies higher weight to the prior information which, of course, may be of little practical interest. Since

$$(G + \omega R' W^{-1} R)^{-1} = G^{-1} - \omega G^{-1} R' (W + \omega R G^{-1} R')^{-1} R G^{-1}, \tag{17}$$

we have

$$\hat{\beta}_{GDWME}(\omega) = \hat{\beta}_{GDE} + \omega G^{-1} R' (W + \omega R G^{-1} R')^{-1} (r - R\hat{\beta}_{GDE}). \tag{18}$$

If we substitute $\omega = 1$ in (16), we get

$$\hat{\beta}_{GDME} = (G + R' W^{-1} R)^{-1} (\tilde{X}' V_D^{-1} \tilde{y} + R' W^{-1} r), \tag{19}$$

which is the *generalized difference-based mixed estimator*. The ordinary mixed estimator is proposed by Theil and Goldberger (1961). This estimator gives equal weight to sample and prior information.

Generalized Ridge estimator proposed by Hoerl and Kennard (1970) is defined as

$$\begin{aligned} \hat{\beta}_{GDRE}(k) &= (\tilde{X}' V_D^{-1} \tilde{X} + kI)^{-1} \tilde{X}' V_D^{-1} \tilde{y} \\ &= (G + kI)^{-1} \tilde{X}' V_D^{-1} \tilde{y} = G_k^{-1} \tilde{X}' V_D^{-1} \tilde{y} \\ &= (I + kG^{-1})^{-1} \hat{\beta}_{GDE} = T_k \hat{\beta}_{GDE}, \end{aligned} \tag{20}$$

where $G_k = G + kI$, $T_k = G_k^{-1} G = G G_k^{-1}$.

Akdeniz and Erol (2003) discussed the almost unbiased Ridge regression estimator (AURE), which is given as follows:

$$\hat{\beta}_{\hat{A}URE}(k) = \left(I - k^2 S_k^{-2} \right) \hat{\beta}_{OLS}, \tag{21}$$

where $S_k = X'X + kI$, $\hat{\beta}_{OLS} = (X'X)^{-1}X'y$. Similarly, we define the generalized difference-based almost unbiased Ridge estimator (GDAURE) as follows:

$$\hat{\beta}_{GDAURE}(k) = \left(I - k^2 G_k^{-2}\right) \hat{\beta}_{GDE}. \tag{22}$$

Roozbeh and Arashi (2013) proposed the feasible Ridge estimator in partially linear model; Li and Yang (2010) proposed the stochastic mixed Ridge estimator, Arumairajan and Wijekoon (2014) proposed stochastic restricted ordinary Ridge estimator; the generalized diffrence-based ridge estimator (GDRE) defined by Wu (2016). Liu et al. (2013) introduced the weighted mixed almost unbiased Ridge estimator based on the weighted mixed estimator. Substituting $\hat{\beta}_{GDE}$ with $\hat{\beta}_{GDAURE}(k)$ in $\hat{\beta}_{GDWME}(\omega)$, we describe a generalized difference–based weighted mixed almost unbiased Ridge estimator (GDWMAURE), as follows:
Since

$$(W + \omega R G^{-1} R')^{-1} = W^{-1} - \omega W^{-1} R (G + \omega R' W^{-1} R)^{-1} R' W^{-1}, \tag{23}$$

and

$$\begin{aligned}
&\omega G^{-1} R' (W + \omega R G^{-1} R')^{-1} r \\
&\quad = \omega G^{-1} R' \left[W^{-1} - \omega W^{-1} R (G + \omega R' G^{-1} R)^{-1} R' W^{-1} \right] r \\
&\quad = \left[G^{-1} - \omega G^{-1} R' W^{-1} R (G + \omega R' W^{-1} R)^{-1} \right] \omega R' W^{-1} r \\
&\quad = \left[G^{-1} - G^{-1} \{ (G + \omega R' W^{-1} R) - G \} (G + \omega R' W^{-1} R)^{-1} \right] \omega R' W^{-1} r \\
&\quad = \left[G^{-1} - G^{-1} \left[I - G (G + \omega R' W^{-1} R)^{-1} \right] \right] \omega R' W^{-1} r \\
&\quad = (G + \omega R' W^{-1} R)^{-1} \omega R' W^{-1} r.
\end{aligned} \tag{24}$$

Using the equalities in (22) and (24), we have

$$\begin{aligned}
\hat{\beta}_{GDWMAURE}(\omega, k) &= \hat{\beta}_{AU}(\omega, k) \\
&= \hat{\beta}_{GDAURE}(k) + \omega G^{-1} R' (W + \omega R G^{-1} R')^{-1} (r - R \hat{\beta}_{GDAURE}(k)) \\
&= (I - k^2 G_k^{-2}) \hat{\beta}_{GDE} + \omega G^{-1} R' (W + \omega R G^{-1} R')^{-1} \\
&\quad \times (r - R (I - k^2 G_k^{-2}) \hat{\beta}_{GDE}) \\
&= (G + \omega R' W^{-1} R)^{-1} \left[(I - k^2 G_k^{-2}) \tilde{X}' V_D^{-1} \tilde{y} + \omega R' W^{-1} r \right].
\end{aligned} \tag{25}$$

In fact, from the definition of $\hat{\beta}_{GDWMAURE}(\omega, k) = \hat{\beta}_{AU}(\omega, k)$, we can see that $\hat{\beta}_{AU}(\omega, k)$ is a general estimator, which includes $\hat{\beta}_{GDE}$, $\hat{\beta}_{GDME}$, $\hat{\beta}_{GDAURE}(k)$ and $\hat{\beta}_{GDWME}(\omega)$ as special cases. Namely,

$$\text{if } k=0, \omega=0 \text{ then } \hat{\beta}_{AU}(0,0)=\hat{\beta}_{GDE}, \tag{26}$$

$$\text{if } k=0 \text{ and } \omega=1 \text{ then } \hat{\beta}_{AU}(\omega=1,k=0)=\hat{\beta}_{GDME}, \tag{27}$$

$$\text{if } \omega=0, \text{ then } \hat{\beta}_{AU}(\omega=0,k)=G^{-1}(I-k^2G_k^{-2})\tilde{X}'V_D^{-1}\tilde{y},$$

Observing that G^{-1} and $(I-k^2G_k^{-2})$ are commutative, we have

$$\hat{\beta}_{AU}(\omega=0,k)=(I-k^2G_k^{-2})G^{-1}\tilde{X}'V_D^{-1}\tilde{y}=(I-k^2G_k^{-2})\hat{\beta}_{GDE}=\hat{\beta}_{GDAURE}(k), \tag{28}$$

$$\text{if } k=0, \text{ then } \hat{\beta}_{AU}(\omega,k=0)=(G+\omega R'W^{-1}R)^{-1}\left[\tilde{X}'V_D^{-1}\tilde{y}+\omega R'W^{-1}r\right]$$
$$=\hat{\beta}_{GDWME}(\omega). \tag{29}$$

5 Mean squared error matrix comparisons of estimators

In this section, we compare the underlying estimators. For the convenience of the following discussions, we give some lemmas here.

Lemma 5.1 (Farebrother 1976) *Let A be a positive definite matrix, namely $A>0$, and let α be some vector, then $A-\alpha\alpha'\geq 0$ if and only if $\alpha'A^{-1}\alpha\leq 1$.*

Lemma 5.2 (Rao et al. 2008, Theorem A50, p. 504) *Let $n\times n$ matrices $M>0$ and $N>0$ (**or** $N\geq 0$), then $M>N$ if and only if $\lambda_{\max}(NM^{-1})<1$.*

Lemma 5.3 (Trenkler and Toutenburg 1990) *Let $\hat{\beta}_j=A_jy, j=1,2$ be two competing estimators of β. Suppose that $\Delta=Cov(\hat{\beta}_1)-Cov(\hat{\beta}_2)>0$. Then $MSEM(\hat{\beta}_1)-MSEM(\hat{\beta}_2)\geq 0$ if and only if $b_2'(\Delta+b_1b_1')^{-1}b_2\leq 1$, where b_j denotes bias vector of $\hat{\beta}_j$.*

The mean squared error matrix (MSEM) of an estimator $\tilde{\beta}$ is defined as

$$MSEM(\tilde{\beta})=E(\tilde{\beta}-\beta)(\tilde{\beta}-\beta)'=Cov(\tilde{\beta})+Bias(\tilde{\beta})Bias(\tilde{\beta})',$$

where $Cov(\tilde{\beta})$ is the dispersion matrix of $\tilde{\beta}$ and $Bias(\tilde{\beta})=E(\tilde{\beta})-\beta$ is the bias vector of $\tilde{\beta}$. If two estimators $\tilde{\beta}_1$ and $\tilde{\beta}_2$ of β are given, the estimator $\tilde{\beta}_2$ is said to be superior to $\tilde{\beta}_1$ with respect to the MSEM criterion, if and only if

$$MSEM(\tilde{\beta}_1)-MSEM(\tilde{\beta}_2)\geq 0.$$

Now we study the efficiency properties of generalized difference-based weighted mixed estimator, $\hat{\beta}_{GDWME}(\omega)$ and the dominance conditions for the mean squared error matrix (MSEM) superiority of $\hat{\beta}_{GDWME}(\omega)$ and $\hat{\beta}_{AU}(\omega,k)$ over $\hat{\beta}_{GDE}$. It is easy to compute that the expectation and covariance matrix of the $\hat{\beta}_{AU}(\omega,k)$ are

$$E(\hat{\beta}_{AU}(\omega,k))=B\left[(I-k^2G_k^{-2})G\beta+\omega R'W^{-1}R\beta\right]=BB^*\beta, \tag{30}$$

and

$$Cov((\hat{\beta}_{AU}(\omega, k)) = \sigma^2 B B^{**} B', \tag{31}$$

where

$$B =: (G + \omega R'W^{-1}R)^{-1}, \quad B^* =: (I - k^2 G_k^{-2})G + \omega R'W^{-1}R,$$
$$B^{**} =: \left((I - k^2 G_k^{-2})G(I - k^2 G_k^{-2})' + \omega^2 R'W^{-1}R\right).$$

The bias of $\hat{\beta}_{AU}(\omega, k)$ is

$$\text{Bias } (\hat{\beta}_{AU}(\omega, k)) = E(\hat{\beta}_{AU}(\omega, k)) - \beta = -k^2 B G_k^{-2} G\beta, \tag{32}$$

and the mean squared error matrix of $\hat{\beta}_{AU}(\omega, k)$ is

$$\begin{aligned} MSEM(\hat{\beta}_{AU}(\omega, k)) &= Cov((\hat{\beta}_{AU}(\omega, k)) + \text{Bias}(\hat{\beta}_{AU}(\omega, k))\,\text{Bias}(\hat{\beta}_{AU}(\omega, k))' \\ &= \sigma^2 B B^{**} B' + b_1 b_1', \end{aligned} \tag{33}$$

where $b_1 = -k^2 B G_k^{-2} G\beta$. So it is obvious that $\hat{\beta}_{AU}(\omega, k)$ is always biased unless $k = 0$.

We can easily obtain the MSEM of the estimators $\hat{\beta}_{GDE}$, $\hat{\beta}_{GDME}$, $\hat{\beta}_{GDRE}(k)$,$\hat{\beta}_{GDWME}$ (ω) as follows

$$MSEM(\hat{\beta}_{GDE}) = Cov(\hat{\beta}_{GDE}) = \sigma^2 G^{-1}, \tag{34}$$
$$MSEM(\hat{\beta}_{GDME}) = Cov(\hat{\beta}_{GDME}) = \sigma^2 (G + R'W^{-1}R)^{-1}, \tag{35}$$
$$MSEM(\hat{\beta}_{GDAURE}(k)) = \sigma^2 (I - k^2 G_k^{-2}) G^{-1} (I - k^2 G_k^{-2})' + b_2 b_2', \tag{36}$$

with $b_2 = Bias(\hat{\beta}_{GDAURE}(k)) = -k^2 G_k^{-2}\beta$. Also,

$$MSEM(\hat{\beta}_{GDWME}(\omega)) = MSEM(\hat{\beta}(\omega)) = Cov(\hat{\beta}(\omega)) = \sigma^2 B(G + \omega^2 R'W^{-1}R)B'. \tag{37}$$

5.1 MSEM comparison between $\hat{\boldsymbol{\beta}}(\boldsymbol{\omega})$ and $\hat{\boldsymbol{\beta}}_{AU}(\boldsymbol{\omega}, \boldsymbol{k})$

Theorem 5.1 *The generalized difference-based weighted mixed almost unbiased Ridge estimator $\hat{\beta}_{AU}(\omega, k)$ is superior to the generalized difference-based weighted mixed estimator $\hat{\beta}(\omega)$ in the MSEM sense, namely $MSEM(\hat{\beta}(\omega)) - MSEM(\hat{\beta}_{AU}(\omega, k)) \geq 0$, if and only if $\sigma^{-2} b_1' \tilde{B}^{-1} b_1 \leq 1$.*

Proof In order to compare $\hat{\beta}(\omega)$ with $\hat{\beta}_{AU}(\omega, k)$ in the MSEM sense we consider the difference

$$\begin{aligned} \Delta_1 &= MSEM(\hat{\beta}(\omega)) - MSEM(\hat{\beta}_{AU}(\omega, k)) \\ &= \sigma^2 B(G + \omega^2 R'W^{-1}R)B' \end{aligned}$$

$$-\left[\sigma^2 B((I-k^2G_k^{-2})G(I-k^2G_k^{-2})'+\omega^2 R'W^{-1}R)B'+b_1b_1'\right]$$
$$=\sigma^2 B(G-F_kGF_k')B'-b_1b_1'=\sigma^2 B\Delta B'-b_1b_1', \tag{38}$$

where $F_k=(I-k^2G_k^{-2})$ and $\Delta=G-F_kGF_k'$. For $G=\tilde{X}'V_D^{-1}\tilde{X}>0$, there exists some orthogonal matrix Q, such that $G=Q\Lambda Q'$, where $\Lambda=diag(\lambda_1,\ldots,\lambda_p)$. Therefore, we can easily compute that

$$G-F_kGF_k'=Qdiag(\delta_1,\delta_2,\ldots,\delta_p)Q',$$

where $\delta_i=\frac{\lambda_i(G)k^2(k^2+4\lambda_i(G)k+2\lambda_i(G)^2)}{(\lambda_i(G)+k)^4}$. Since $k>0$ and $\lambda_i(G)>0$, then $\delta_i>0$. Observing that $B=:(G+\omega R'W^{-1}R)^{-1}>0$, we get $\Delta>0$ and $\tilde{B}=:B\Delta B'>0$. By applying the Lemma 5.1, we get $MSEM(\hat{\beta}(\omega))-MSEM(\hat{\beta}_{AU}(\omega,k))\geq 0$ if and only if $\sigma^{-2}b_1'\tilde{B}^{-1}b_1\leq 1$. This completes the proof. □

5.2 MSEM comparison between $\hat{\beta}_{GDAURE}(k)$ and $\hat{\beta}_{AU}(\omega,k)$

Theorem 5.2 *When $\lambda_{\max}(NM^{-1})<1$, the generalized difference-based weighted mixed almost unbiased Ridge estimator $\hat{\beta}_{AU}(\omega,k)$ is superior to the generalized difference-based almost unbiased Ridge estimator $\hat{\beta}_{GDAURE}(k)$ in the MSEM sense, namely $MSEM(\hat{\beta}_{GDAURE}(k))-MSEM(\hat{\beta}_{AU}(\omega,k))\geq 0$, if and only if $b_1'(\sigma^2\tilde{\Delta}+b_2b_2')^{-1}b_1\leq 1$.*

Proof In order to compare $\hat{\beta}_{GDAURE}(k)$ with $\hat{\beta}_{AU}(\omega,k)$ in the MSEM sense, we similarly consider the difference

$$\begin{aligned}\Delta_2&=MSEM(\hat{\beta}_{GDAURE}(k))-MSEM(\hat{\beta}_{AU}(\omega,k))\\&=\sigma^2F_kG^{-1}F_k'+b_2b_2'-\left[\sigma^2B(F_kGF_k'+\omega^2R'W^{-1}R)B'+b_1b_1'\right]\\&=\sigma^2\left[F_kG^{-1}F_k'-B(F_kGF_k'+\omega^2R'W^{-1}R)B'\right]+b_2b_2'-b_1b_1'\\&=\sigma^2(M-N)+b_2b_2'-b_1b_1'=\sigma^2\tilde{\Delta}+b_2b_2'-b_1b_1'.\end{aligned} \tag{39}$$

where $M=F_kG^{-1}F_k'$, $N=B(FGF'+\omega^2R'W^{-1}R)B'$ and $\tilde{\Delta}=M-N$.

It is obvious that, $M=F_kG^{-1}F_k'>0$, $N=B(F_kGF_k'+\omega^2R'W^{-1}R)B'>0$. Therefore, when $\lambda_{\max}(NM^{-1})<1$, we get $\tilde{\Delta}>0$ by applying Lemma 5.2. Furthermore, by Lemma 5.3, we have $MSEM(\hat{\beta}_{GDAURE}(k))-MSEM(\hat{\beta}_{AU}(\omega,k))\geq 0$ if and only if $b_1'(\sigma^2\tilde{\Delta}+b_2b_2')^{-1}b_1\leq 1$. This assertation completes the proof. □

5.3 MSEM comparison between $\hat{\beta}_{GDE}$ and $\hat{\beta}_{GDWME}(\omega)$

Theorem 5.3 *The generalized difference-based weighted mixed estimator $\hat{\beta}_{GDWME}(\omega)$, is superior to the generalized difference based estimator, $\hat{\beta}_{GDE}$ in the sense that $Cov(\hat{\beta}_{GDE})-Cov(\hat{\beta}_{GDWME})>0$, if and only if $(\frac{2}{\omega}-1)W+RG^{-1}R'>0$.*

Proof The covariance matrix of $\hat{\beta}_{GDE}$ is

$$Cov(\hat{\beta}_{GDE}) = \sigma^2(\tilde{X}'V_D^{-1}\tilde{X})^{-1} = \sigma^2 G^{-1}. \quad (40)$$

The covariance matrix of $\hat{\beta}_{GDWME}$ is

$$Cov(\hat{\beta}_{GDWME}(\omega)) = \sigma^2 B(G + \omega^2 R'W^{-1}R)B', \quad (41)$$

where $B = (G + \omega R'W^{-1}R)^{-1}$. Since $MSEM(\hat{\beta}_{GDWME}(\omega)) = Cov(\hat{\beta}_{GDWME}(\omega))$ and $MSEM(\hat{\beta}_{GDE}) = Cov(\hat{\beta}_{GDE})$, we have

$$\begin{aligned}
\Delta_3 &= MESM(\hat{\beta}_{GDE}) - MSEM(\hat{\beta}_{GDWME}(\omega)) \\
&= \sigma^2 G^{-1} - \sigma^2 B(G + \omega^2 R'W^{-1}R)B' \\
&= \sigma^2 B\left[B^{-1}G^{-1}B'^{-1} - G - \omega^2 R'W^{-1}R\right]B' \\
&= \sigma^2\omega^2 BR'W^{-1}\left[(\frac{2}{\omega} - 1)W + RG^{-1}R'\right]W^{-1}RB'. \quad (42)
\end{aligned}$$

The difference is positive definite when B is positive definite and $\left[(\frac{2}{\omega} - 1)W + RG^{-1}R'\right] > 0$, which is possible as long as $\omega < 2$. ω is a nonstochastic and non-negative scalar weight with $0 \leq \omega \leq 1$, a value of w greater than 1 implies higher weight to the prior information which of course, may be little practical interest. When $q < p$, R' has full column rank and it follows that in this case we can only conclude that $\Delta_3 \geq 0$. This result completes the proof. □

6 Selection of biasing parameter k and non-stochastic weight ω

In this section, we give a method to choose k and ω. Firstly, a difference-based Model (7) can be transformed to a canonical form by the orthogonal transformation. Let Q be an orthogonal matrix such that

$$Q'GQ = Q'\tilde{X}'V_D^{-1}\tilde{X}Q = Q'\tilde{X}'V_D^{-1/2}V_D^{-!/2}\tilde{X}Q = \bar{X}'\bar{X} = \Lambda = diag(\lambda_1, \lambda_2, \ldots, \lambda_p),$$

where λ_i is the eigenvalue of G. Then we get canonical form of Model (7) as

$$V_D^{-1/2}\tilde{y} = V_D^{-1/2}\tilde{X}\beta + V_D^{-1/2}\tilde{\varepsilon},$$

and

$$y_* = X_*\beta + \varepsilon_* = X_*QQ'\beta + \varepsilon_* = \bar{X}\bar{\beta} + \varepsilon_*, \quad (43)$$

where $\bar{X} = X_*Q = V_D^{-1/2}\tilde{X}Q$, $\beta = Q'\beta$.

Not that, $MSEM(\hat{\bar{\beta}}_{AU}(\omega, k)) = Q'MSEM(\hat{\beta}_{AU}(\omega, k))$Q. It is supposed that G and $R'W^{-1}R$ are commutative, then from equation (33) we see that

$$\begin{aligned}
&MSEM(\hat{\hat{\beta}}_{AU}(\omega,k)) = Q'(\sigma^2 BB^{**}B' + b_1^* b_1^{*'})Q \\
&= Q'\left\{\sigma^2\ (G+\omega R'W^{-1}R)^{-1}\left[(I-k^2G_k^{-2})G(I-k^2G_k^{-2}) + \omega^2 R'W^{-1}R\right]\right. \\
&\quad (G+\omega R'W^{-1}R)^{-1}Q + k^4 Q'BG_k^{-2}G\beta\beta'GG_k^{-2}BQ \\
&= \sigma^2(\Lambda+\omega\Psi)^{-1}\left\{\ \left[(I-k^2(\Lambda+kI)^{-2}\Lambda(I-k^2(\Lambda+kI)^{-2}\right] + \omega^2\Psi\ \right\}(\Lambda+\omega\Psi)^{-1} \\
&\quad + k^4(\Lambda+\omega\Psi)^{-1}(\Lambda+kI)^{-2}\Lambda\bar{\beta}\bar{\beta}'\Lambda(\Lambda+kI)^{-2}(\Lambda+\omega\Psi)^{-1} \qquad (44)
\end{aligned}$$

where $Q'R'W^{-1}RQ = \Psi = diag(\xi_1, \xi_2, \ldots, \xi_p)$; $G_k = (G+kI)$; $Q'G_k^{-2}Q = (\Lambda+kI)^{-2}$.
Since, $I - k^2(\Lambda+kI)^{-2} = (\Lambda+kI)^{-1}\Lambda\left[I + k(\Lambda+kI)^{-1}\right]$, we have

$$\begin{aligned}
&MSEM(\hat{\hat{\beta}}_{AU}(\omega,k)) \\
&= \sigma^2(\Lambda+\omega\Psi)^{-1}\left\{\left[(I+k(\Lambda+kI)^{-1})(\Lambda+\mathrm{kI})^{-1}\Lambda^3(\Lambda+kI)^{-1}(I+k(\Lambda+\mathrm{kI})^{-1})\right]\right. \\
&\quad \left. + \omega^2\Psi\right\}(\Lambda+\omega\Psi)^{-1} + k^4(\Lambda+\omega\Psi)^{-1}(\Lambda+kI)^{-2}\Lambda\bar{\beta}\bar{\beta}'\Lambda(\Lambda+kI)^{-2}(\Lambda+\omega\Psi)^{-1}
\end{aligned}$$

Optimal values for k and ω can be derived by minimizing

$$\begin{aligned}
g(\omega,k) &= trMSEM(\hat{\hat{\beta}}_{AU}(\omega,k)) \\
&= \sum_{i=1}^{p} \frac{\sigma^2\left[\lambda_i^3(2\mathrm{k}+\lambda_i)^2 + \omega^2\xi_i(k+\lambda_i)^4\right] + k^4\lambda_i^2\bar{\beta}_i^2}{(k+\lambda_i)^4(\lambda_i+\omega\xi_i)^2} \qquad (45)
\end{aligned}$$

First we give how to choose the value of k. It is easy to see that,

$$\frac{\partial g(\omega,k)}{\partial k} = 0 \qquad (46)$$

or

$$\sum_{i=1}^{p} \frac{4k\lambda_i^3(\bar{\beta}_i^2k^2 - 2\sigma^2k - \sigma^2\lambda_i)}{(k+\lambda_i)^5(\lambda_i+\omega\xi_i)^2} = 0,$$

Thus, the optimal choice of the paramater k is

$$k_{opt} = \frac{\sigma^2\left(\sum_{i=1}^{p}\lambda_i^3\right) \pm \sqrt{\sigma^4\left(\sum_{i=1}^{p}\lambda_i^3\right)^2 + \sigma^2\left(\sum_{i=1}^{p}\lambda_i^4\right)\left(\sum_{i=1}^{p}\bar{\beta}_i^2\lambda_i^3\right)}}{\left(\sum_{i=1}^{p}\bar{\beta}_i^2\lambda_i^3\right)}. \qquad (47)$$

After the unkown parameters σ^2 and $\bar{\beta}_i^2$ are replaced by their unbiased estimators, we get the optimal estimator of k for a fixed ω value as

$$\hat{k}_{opt} = \frac{\hat{\sigma}^2\left(\sum_{i=1}^{p}\lambda_i^3\right) + \sqrt{\hat{\sigma}^4\left(\sum_{i=1}^{p}\lambda_i^3\right)^2 + \hat{\sigma}^2\left(\sum_{i=1}^{p}\lambda_i^4\right)\left(\sum_{i=1}^{p}\hat{\bar{\beta}}_i^2\lambda_i^3\right)}}{\left(\sum_{i=1}^{p}\hat{\bar{\beta}}_i^2\lambda_i^3\right)}. \qquad (48)$$

Note that $\hat{k}_{\hat{o}pt} > 0$. The value of ω which minimizes the function $g(\omega, k)$ can be found by differentiating with respect to ω when k is fixed:

$$\frac{\partial g(\omega,k)}{\partial \omega} = \sum_{i=1}^{p} \frac{2\left\{\sigma^2 \omega \xi_i \lambda_i (k+\lambda_i)^4 - k^4 \lambda_i^2 \bar{\beta}_i^2 \xi_i - \sigma^2 \lambda_i^3 \xi_i (2k+\lambda_i)^2\right\}}{(k+\lambda_i)^4 (\lambda_i + \omega \xi_i)^3}, \tag{49}$$

and equating it zero. After unknown parameters σ^2 and $\bar{\beta}_i^2$ are replaced by their unbiased estimators, we get the optimal estimator of ω for a fixed k-value as

$$\hat{\omega}_{\hat{o}pt} = \frac{\sum_{i=1}^{p} \hat{e}_{2i}}{\sum_{i=1}^{p} \hat{e}_{1i}}, \tag{50}$$

where $\hat{e}_{1i} = 2\hat{\sigma}^2 \xi_i \lambda_i (k+\lambda_i)^4$, $\hat{e}_{2i} = 2k^4 \lambda_i^2 \hat{\bar{\beta}}_i^2 \xi_i + 2\hat{\sigma}^2 \lambda_i^3 \xi_i (2k+\lambda_i)^2$.

7 Illustrative examples

In this section, we present some numerical examples to support our assertions. The process is categorized into two setups: the first part is devoted to the Monte-Carlo simulation studies and the second one is application of our proposed estimation method to the electricity consumption dataset collected in Germany (Akdeniz Duran et al. 2012).

7.1 The Monte-Carlo simulation studies

In this section, we continue the comparison of proposed estimators based on the scalar values of mean squared error matrix by some simulations and graphical results. Since, theoretically, these estimators are very difficult to compare, the Monte-Carlo simulation studies have been conducted to compare the efficiency of the estimators. The scalar-valued mean squared error (SMSE) for any estimator $\tilde{\beta}$ is defined as

$$SMSE(\tilde{\beta}) = tr(MSEM(\tilde{\beta})) = \mathrm{tr}(Cov(\tilde{\beta})) + Bias(\tilde{\beta})' Bias(\tilde{\beta}.)$$

The explanatory variables were generated for n=50 with 10,000 iteration from the following model:

$$y_i = \sum_{j=1}^{6} x_{ij}\beta_j + f(u_i) + \varepsilon_i, \quad i = 1, 2, \ldots, n, \tag{51}$$

where

$$\beta = (-3, 1, -3, 1, -7, -4)', \; f(u) = \sin(u)\cos(8u), \quad u \in [0,1],$$
$$X \sim N_p(\mu_x, \Sigma_x), \; \mu_x = (1, 1.5, 1.2, 3, -5, 2)', \; (\Sigma_x)_{ij} = \rho^{|\mathrm{i-j}|}, \quad \mathrm{i,j} = 1, \ldots, \mathrm{n}.$$

and

$$\varepsilon \sim N_n(0, \sigma^2 V), \sigma = 0.15, v_{ij} = \exp\{-\phi|\text{i-j}|\}, \phi = 1, \quad \text{i,j} = 1, \ldots, \text{n}$$

The parametric part of Model (53), i.e. β, is estimated by a third-order differencing coefficients, $d_0 = 0.8582, d_1 = -0.3832, d_2 = -0.2809$ and $d_3 = -0.1942$, and then, the nonparametric part is estimated by kernel methodology and cross-validation criteria. Optimal differencing weights do not have analytic expressions but may be calculated easily using an optimization routine. Hall et al. (1990) presented weights to order $m = 10$. These contain some minor errors. Now, we define the $(n-3) \times n$ differencing matrix as

$$D = \begin{pmatrix} d_0 & d_1 & d_2 & d_3 & 0 & 0 & \ldots & 0 \\ 0 & d_0 & d_1 & d_2 & d_3 & 0 & \ldots & 0 \\ \vdots & \ddots & & & & & & \vdots \\ 0 & 0 & \ldots & 0 & d_0 & d_1 & d_2 & d_3 \end{pmatrix}$$

For the restriction, we consider the following stochastic linear restrictions

$$r = R\beta + \text{e}, \; R = \begin{pmatrix} 1 & 5 & -3 & -1 & -1 & 0 \\ -2 & -1 & 0 & -2 & 3 & 1 \\ 1 & 2 & 1 & 3 & -2 & 0 \\ 4 & -1 & 2 & 2 & 0 & -2 \\ 5 & 3 & 4 & -5 & 1 & 0 \end{pmatrix},$$

where $e \sim N_q(0, \sigma_e^2 W), \quad \sigma_e = 0.15, \quad w_{ij} = \left(\frac{1}{\text{n}}\right)^{|\text{i}-\text{j}|}$, i, j = 1, ..., q.

The Monte-Carlo simulation is performed with M=10^4 replications, obtaining the estimators $\hat{\beta}_{(1)} = \hat{\beta}_{GDE}, \hat{\beta}_{(2)} = \hat{\beta}_{GDWME}(\hat{\omega}_{opt}), \hat{\beta}_{(3)} = \hat{\beta}_{GDME}, \hat{\beta}_{(4)} = \hat{\beta}_{GDRE}\left(\hat{k}_{opt}\right), \hat{\beta}_{(5)} = \hat{\beta}_{GDAURE}(\hat{k}_{opt})$ and $\hat{\beta}_{(6)} = \hat{\beta}_{GDWMAURE}\left(\hat{\omega}_{opt}, \hat{k}_{opt}\right)$ in the restricted semiparametric regression model.

The relative efficiencies of the above methods with respect to the first method are estimated as

$$\text{Eff}(\hat{\beta}_{(\text{i})}) = \frac{\frac{1}{\text{M}}\sum_{m=1}^{M} MSE\left(\hat{\beta}_{(1)}^{(\text{m})}\right)}{\frac{1}{\text{M}}\sum_{m=1}^{M} MSE\left(\hat{\beta}_{(\text{i})}^{(\text{m})}\right)} = \frac{\frac{1}{\text{M}}\sum_{m=1}^{M} \left\|\hat{\beta}_{(1)}^{(\text{m})} - \beta\right\|_2^2}{\frac{1}{\text{M}}\sum_{m=1}^{M} \left\|\hat{\beta}_{(i)}^{(\text{m})} - \beta\right\|_2^2}, \quad i = 1, \ldots, 6,$$

Where $\hat{\beta}_{(i)}^{(m)}$ is the estimator obtained in the m^{th} iteration and $\|v\|_2^2 = \sum_{i=1}^{q} v_i^2$ for $v = (v_1, \ldots, v_q)'$.

To achieve different degrees of collinearity, four different correlations corresponding to $\rho = 0.50, 0.75, 0.90$ and 0.95 are considered. In Fig. 1, the nonparametric part of the Model (53) is plotted. This function is wavy and provides a good test case for the nonparametric regression method. All computations were conducted using the statistical package R. In Tables 1, 2, 3, 4, we computed the proposed estimators at

Fig. 1 The nonparametric function of Model (53)

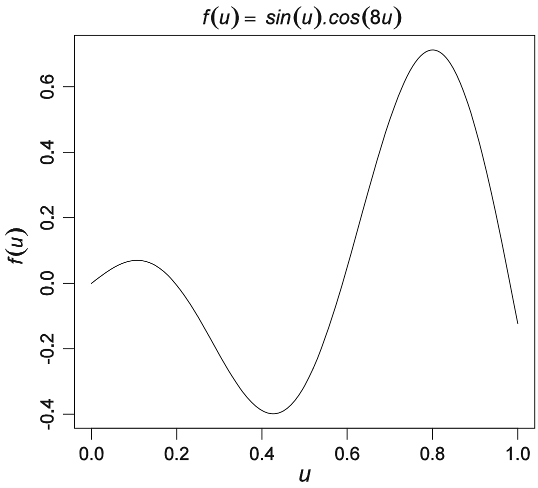

Table 1 Evaluation of parameters for proposed estimators with $\rho = 0.50$

Method coefficients	GDE	GDWME	GDME	GDRE	GDAURE	GDWMAURE
$\hat{\beta}_1$	−2.99359	−2.99903	−2.99847	−2.97742	−2.98015	−2.99330
$\hat{\beta}_2$	1.00147	1.00108	1.00118	0.97107	0.97212	0.99354
$\hat{\beta}_3$	−2.99886	−3.00365	−3.00366	−2.96704	−2.96235	−3.00276
$\hat{\beta}_4$	1.47496	1.49702	1.48871	1.42677	1.42906	1.49658
$\hat{\beta}_5$	−6.97396	−6.99540	−6.97734	−6.93061	−6.93473	−6.99410
$\hat{\beta}_6$	−4.01220	−4.00484	−4.01302	−4.02218	−4.02648	−4.00512
$S\hat{M}SE\left(\hat{\beta}_{(i)}\right)$	0.15195	0.03590	0.03935	0.12744	0.12693	0.03175
$Eff\left(\hat{\beta}_{(i)}\right)$	1.00000	4.23275	3.86166	1.19234	1.19708	4.78551
$\hat{\text{mse}}\left(\hat{f}_{(i)}, f\right)$	1.47895	0.19898	0.20299	1.12680	1.13552	0.17652

optimum values of parameters, $\hat{\omega}_{opt}$ and $\hat{k}_{opt}$ respectively. These optimum values of parameters are obtained using "`optim`" command with "`L-BFGS-B`" method. We numerically estimated the efficiencies of proposed estimators relative to GDE and $\hat{\text{mse}}(\hat{f}_{(\text{i})}, f) = \frac{1}{\text{nM}} \sum_{m=1}^{M} \left\| \hat{f}_{(\text{i})}^{(\text{m})} - f \right\|_2^2$ for all proposed estimators, where $\hat{f}_{(i)}^{(m)}$ is obtained in the m^{th} iteration using kernel method as $\hat{f}_{(i)}^{(m)} = K\left(y - X\hat{\beta}_{(i)}^{(m)}\right)$ for i=1,…,6 and K is the smoother matrix.

The 3D diagrams as well as the 2D slices of SMSE's versus parameters are plotted for proposed estimators in Fig. 2. Since the results were similar across cases, to save space we only reported the results for $\rho = 0.90$. As it can be seen from Fig. 2, the 2D (3D) diagrams of SMSE are convex functions (surface) and hence they have a global minimum. This guarantees the existence of optimum values of k and ω which

Table 2 Evaluation of parameters for proposed estimators with $\rho = 0.75$

Method coefficients	GDE	GDWME	GDME	GDRE	GDAURE	GDWMAURE
$\hat{\beta}_1$	−2.99739	−2.99645	−2.98574	−2.97182	−2.97601	−2.99654
$\hat{\beta}_2$	0.99784	0.99855	0.98778	0.94467	0.94702	0.99840
$\hat{\beta}_3$	−2.99821	−3.00642	−3.01130	−2.93893	−2.93115	−3.00555
$\hat{\beta}_4$	1.47989	1.49741	1.48650	1.39079	1.39500	1.49588
$\hat{\beta}_5$	−6.99142	−6.99983	−6.97494	−6.91205	−6.91769	−7.00030
$\hat{\beta}_6$	−3.99560	−3.99813	−4.00672	−4.01575	−4.02225	−3.99676
$S\hat{M}SE\left(\hat{\beta}_{(i)}\right)$	0.36378	0.05674	0.06279	0.29221	0.29138	0.04827
$Eff\left(\hat{\beta}_{(i)}\right)$	1.00000	6.41134	5.79387	1.24491	1.24846	7.53623
$m\hat{s}e\left(\hat{f}_{(i)}, f\right)$	3.49761	0.22877	0.22149	2.45599	2.48343	0.20882

Table 3 Evaluation of parameters for proposed estimators with $\rho = 0.90$

Method coefficients	GDE	GDWME	GDME	GDRE	GDAURE	GDWMAURE
$\hat{\beta}_1$	−3.00300	−3.00971	−2.98434	−2.95076	−2.95806	−3.00605
$\hat{\beta}_2$	1.01266	1.00682	0.98584	0.91059	0.91705	1.00528
$\hat{\beta}_3$	−2.99023	−2.99187	−3.01306	−2.88133	−2.86998	−2.99652
$\hat{\beta}_4$	1.47482	1.50290	1.49484	1.31050	1.31988	1.50357
$\hat{\beta}_5$	−6.97500	−6.99208	−6.97794	−6.82976	−6.84086	−6.99389
$\hat{\beta}_6$	−4.02091	−4.01743	−4.00576	−4.05900	−4.06919	−4.01377
$S\hat{M}SE\left(\hat{\beta}_{(i)}\right)$	1.04777	0.09615	0.11547	0.81087	0.81056	0.07849
$Eff\left(\hat{\beta}_{(i)}\right)$	1.00000	10.89775	9.07367	1.29216	1.29265	13.34882
$m\hat{s}e\left(\hat{f}_{(i)}, f\right)$	9.34131	0.24149	0.24444	6.62499	6.60640	0.23103

minimize the SMSE's. Figure 3 shows the fitted function by kernel smoothing after estimating the linear part of the model by proposed estimators for $\rho = 0.90$.

7.2 Application to electricity consumption data set

To motivate the problem of estimation in the partially linear model, we apply the electricity consumption, considered by Akdeniz Duran et al. (2012). The variables are defined for 177 items as follows:

The dependent variable y is the log monthly electricity consumption per person (LEC) and the independent variables include log income per person (LI), log rate of electricity price to the gas price (LREG) and cumulated average temperature index (Temp) for the corresponding month taken as average of 20 German cities computed from the data of German weather service.

Table 4 Evaluation of parameters for proposed estimators with $\rho = 0.95$

Method coefficients	GDE	GDWME	GDME	GDRE	GDAURE	GDWMAURE
$\hat{\beta}_1$	−2.90359	−2.96549	−2.99494	−2.91271	−2.97424	−2.99398
$\hat{\beta}_2$	0.82283	0.97354	0.99851	0.83207	0.96629	0.99784
$\hat{\beta}_3$	−2.80692	−3.02548	−3.00296	−2.79289	−2.95991	−3.00217
$\hat{\beta}_4$	1.22227	1.48824	1.50039	1.23569	1.45264	1.49726
$\hat{\beta}_5$	−6.78489	−6.99168	−7.00153	−6.80125	−6.98940	−7.00219
$\hat{\beta}_6$	−4.04966	−3.97889	−4.00043	−4.06178	−3.99634	−3.99761
$S\hat{M}SE\left(\hat{\beta}_{(i)}\right)$	2.05247	0.14073	0.16843	1.59901	1.59891	0.10672
$Eff\left(\hat{\beta}_{(i)}\right)$	1.00000	14.58471	12.18624	1.28359	1.28367	19.23319
$m\hat{s}e\left(\hat{f}_{(i)}, f\right)$	17.96314	0.25270	0.26767	12.62382	12.56059	0.24046

To detect the nonparametric part of the model, by Yatchew (2000), the test statistic for the null hypothesis that the regression function has the parametric form i.e., $H_0 : f(u) = h(u; \beta)$ for a parametric function $h(.)$, against the nonparametric alternative $f(u)$, when one uses optimal differencing weights, is

$$Z_0 = \sqrt{nm}\frac{\hat{\sigma}^2 - \hat{\sigma}^2_{diff}}{\hat{\sigma}^2_{diff}} \xrightarrow{D} N(0, 1) \tag{52}$$

where $\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}\left(y_i - \mathrm{h}(\mathrm{u};\hat{\beta})\right)^2$, $\hat{\sigma}^2_{diff} = \frac{\tilde{y}'(I-P)\tilde{y}}{\mathrm{tr}(D'(I-P)D)}$, $P = \tilde{X}\left(\tilde{X}'\tilde{X}\right)^{-1}\tilde{X}$.

We consider Temp as a non-parametric part (using a third-order differencing coefficients), because, it has the largest value of nonparametric significance test statistics among those of other independent variables. The statistics of above test for all explanatory variables can be found in Table 5. We also use the added-variable plots 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 (Fig. 5), we consider Temp as a nonparametric part and so, the specification of the partially linear model is

$$(LEC)_i = \sum_{j=1}^{11}\beta_j x_{ij} + \beta_{12}(LI)_i + \beta_{13}(LREG)_i + f(Temp_i) + \varepsilon_i, \tag{53}$$

where $x_1, \ldots, x_{11}$ are dummy variables for the monthly effects. The ratio of largest eigenvalue to smallest eigenvalue for new design matrix in model (55) after applying differencing method is approximately $\lambda_{13}/\lambda_1 = 220.3069$ and so, there exists a potential multicollinearity between the columns of design matrix.

After a primary evaluation of model (55), one might consider the stochastic restriction $r \cong \mathrm{R}\beta$, where

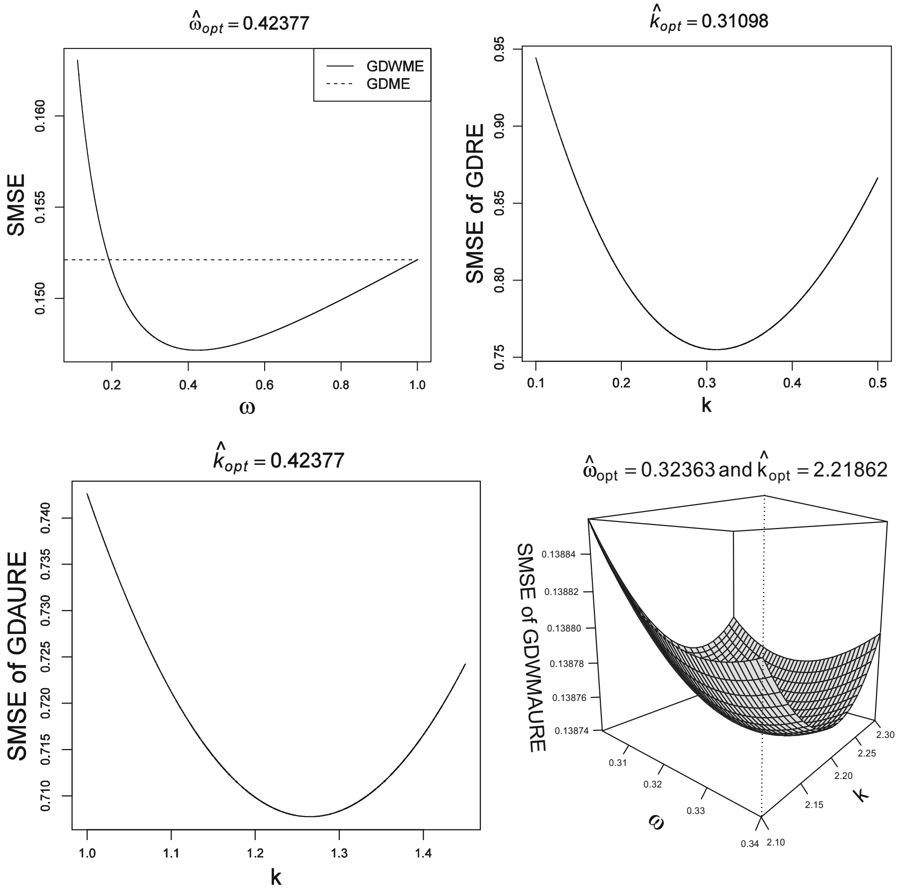

Fig. 2 The diagram of SMSE versus parameters for $\rho = 0.90$. *Top left* $SMSE\left(\hat{\beta}_{GDME}\right)$ *and* $SMSE\left(\hat{\beta}_{GDWME}\right)$; *Top right* $SMSE\left(\hat{\beta}_{GDRE}(k)\right)$; *Bottom left* $SMSE\left(\hat{\beta}_{GDAURE}(k)\right)$; *Bottom right* $SMSE\left(\hat{\beta}_{GDWMAURE}(\omega, k)\right)$

$$R = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 2 & 3 \\ 1 & 1 & 0 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & -1 \end{pmatrix}, \quad r = \begin{pmatrix} -0.03 \\ 0 \end{pmatrix}.$$

We test the linear hypothesis $H_0 : r \cong R\beta$ in the framework of our partially linear model (55). The test statistic for H_0, given our observations, is

$$\chi^2_{rank(R)} = \left(R\hat{\beta}_{diff} - r\right)' \left(R\hat{\Sigma}_{\hat{\beta}_{diff}} R'\right)^{-1} \left(R\hat{\beta}_{diff} - r\right) = 0.00015,$$

where $\hat{\Sigma}_{\hat{\beta}_{diff}} = \left(1 + \frac{1}{2m}\right) \hat{\sigma}^2_{diff} \left(\tilde{X}'\tilde{X}\right)^{-1}$ (see Yatchew 2000). The test statistic is not greater than upper α-quantile of chi-square distribution.Thus we conclude that the null hypothesis H_0 is not rejected.

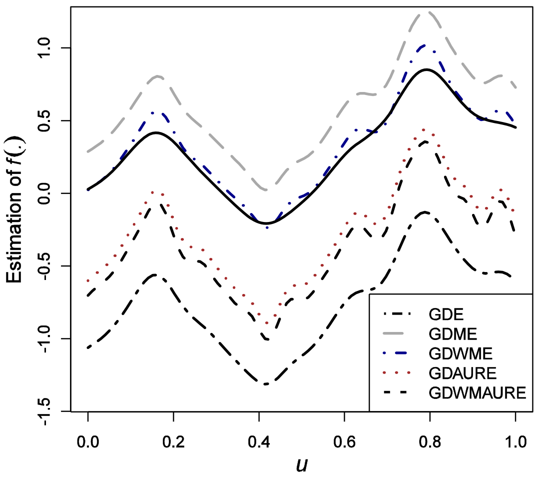

Fig. 3 The Estimation of the function under study by kernel approach for $\rho = 0.90$

Table 5 The values of test statistics (54)

Variable	Z_0
LI	1.99
LREG	2.06
Temp	2.95*

Table 6 shows a summary of the results. In this Table, the RSS and R^2 respectively are the residual sum of squares and coefficient of determination of the model, i.e., $\| y - \hat{y} \|_2^2, \hat{y} = X\hat{\beta}_{(i)} + \hat{f}(u)$ and $R^2 = 1 - RSS/S_{yy}$, which calculated for each proposed estimators of β. For estimation of nonparametric effect, at first we estimated the parametric effects by one of the proposed estimators and then, local polynomial approach was applied to fit $LEC - X\hat{\beta}_{(i)}$ on $u = Temp$, where $X = (x_1, x_2, x_3, x_4, x_5, x_6, x_7, x_8, x_9, x_{10}, x_{11}, LI, LREG)$ (Fig. 5).

8 Conclusions

We considered the method of weighted mixed regression estimation to estimate the regression coefficients in generalized difference-based semiparametric partially linear model. The generalized difference-based weighted mixed almost unbiased Ridge estimator, $\hat{\beta}_{AU}(\omega, k)$ is derived and its dominance over both the generalized difference-based weighted mixed estimator, $\hat{\beta}(\omega)$ and the generalized difference-based almost unbiased Ridge estimator, $\hat{\beta}_{GDAURE}(k)$ is studied under the criterion of mean squared error matrix. After some theorems, the Monte-Carlo simulation studies and a realdata example have been conducted to compare the performance of the proposed estimators numerically. The results from the Monte-Carlo simulations for n = 50, p = 6 and different ρ are presented in Tables 1, 2, 3, 4 and Figs. 2 and 3. From these tables it can be seen that the factor affecting the performance of the estimators is the degree

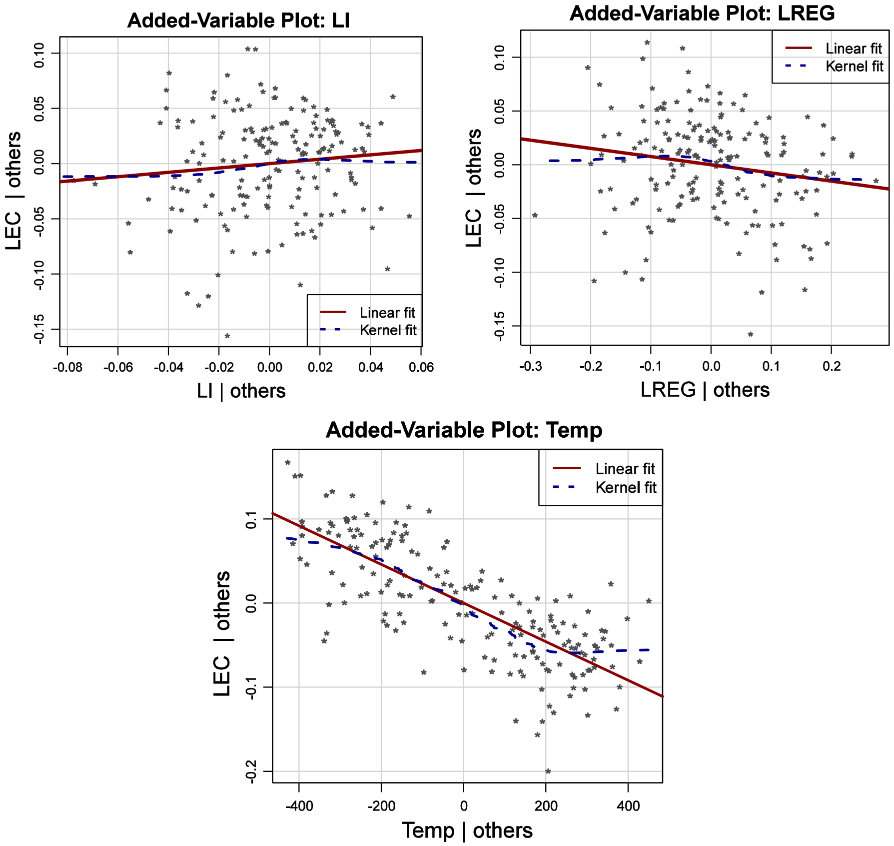

Fig. 4 Added-variable plots of individual explanatory variables versus dependent variable, linear fit (*red solid line*) and kernel fit (*blue dashed line*)

of correlation (ρ). It can be concluded that GDWMAURE is leading to be the best estimator among others for the parametric part of the model, since it offers bigger efficiency and smaler mse values in all proposed estimators. Further GDE is the worst estimator for the parametric part in this examples. In general, the value of ρ has positive effect on the performance of the proposed estimators with respect to GDE. In the real example study, a near dependency among the columns of $\tilde{X}'\tilde{X}$ identified from $\lambda_{13}/\lambda_1 = 220.3069$, that is, the design matrix may be considered as being very ill-conditioned and we had to consider the ridge or form of proposed estimators in our study. As it can be seen from Table 5 and Fig. 4, the nonlinear relation between log monthly electricity consumption per person (LEC) and cumulated average temperature index (Temp) can be detected and so, the pure parametric model does not fit to the data and semiparametric partially linear model fits more significantly. Further, from Table 6 and Fig. 5, it can be deduced that GDWMAURE is quite efficient in the sense that it has significant value of goodness of fit.

Table 6 Evaluation of parameters for proposed estimators for real data set

Method variables	GDE	GDWME	GDME	GDRE	GDAURE	GDWMAURE
x_1	−0.59549	−0.10495	−0.00466	−0.05732	−0.11217	−0.04961
x_2	−0.16763	−0.04680	−0.02483	−0.05139	−0.07615	−0.06107
x_3	−0.02016	−0.01367	−0.00788	−0.00847	−0.01168	−0.01193
x_4	−0.00801	−0.00753	−0.00300	−0.00364	−0.00581	−0.00680
x_5	−0.00619	−0.00286	0.00156	0.00104	−0.00086	−0.00143
x_6	−0.01977	−0.01484	−0.00933	−0.00939	−0.01236	−0.01280
x_7	−0.01178	−0.01346	−0.00913	−0.00784	−0.00937	−0.01059
x_8	−0.02579	−0.01568	−0.00924	−0.01036	−0.01441	−0.01424
x_9	−0.01444	−0.01017	−0.00558	−0.00550	−0.00826	−0.00875
x_{10}	−0.01230	−0.01166	−0.00790	−0.00756	−0.01053	−0.01153
x_{11}	−0.00878	−0.00112	0.00409	0.00240	−0.00011	−0.00015
LI	−0.00006	0.00534	0.01005	0.00778	0.00601	0.00571
LREG	−0.00631	−0.00279	0.00177	0.00164	0.00019	−0.00032
RSS	0.39933	0.36051	0.35987	0.35463	0.35216	0.35061
R^2	0.57175	0.61338	0.61406	0.61969	0.62233	0.62400

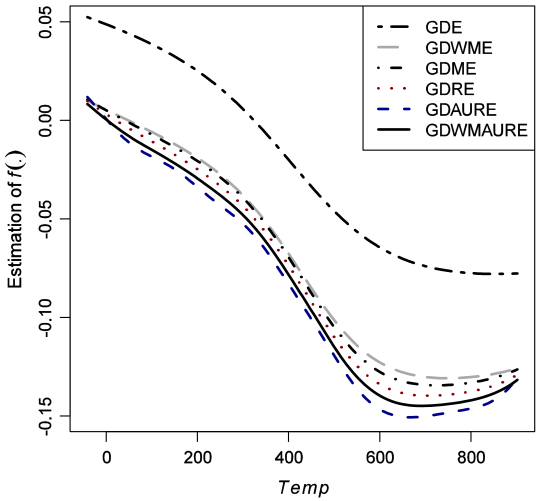

Fig. 5 The Estimations of nonparametric part of model (55)

References

Akdeniz F, Erol H (2003) Mean squared error comparisons of some biased estimators in linear regression. Commun Stat Theory Methods 32(12):2389–2413

Akdeniz F, Tabakan G (2009) Restricted ridge estimators of the parameters in semiparametric regression model. Commun Stat Theory Methods 38(11):1852–1869

Arumairajan S, Wijekoon P (2014) Improvement of ridge estimator when stochastic restrictions are available in the linear regression model. J Stat Econ Methods 3(1):35–48

Duran EA, Härdle WK, Osipenko M (2012) Difference-based ridge and Liu type estimators in semiparametric regression models. J Multivar Anal 105(1):164–175
Farebrother RW (1976) Further results on the mean square error of ridge regression. J R Stat Soc B 38:248–250
Hall P, Kay JW, Titterington DM (1990) On estimation of noise variance in two-dimensional signal processing. Adv Appl Probab 23:476–495
Hoerl AE, Kennard RW (1970) Ridge regression: biased estimation for orthogonal problems. Technometrics 12:55–67
Klipple K, Eubank RL (2007) Difference-based variance estimators for partially linear models. Festschrift in honor of Distinguished Professor Mir Masoom Ali on the occasion of his retirement. May 18-19, pp 313–323
Li Y, Yang H (2010) A new stochastic mixed ridge estimator in linear regression. Stat Pap 51(2):315–323
Liu C, Yang H, Wu J (2013) On the weighted mixed almost unbiased ridge estimator in stochastic restricted linear regression. J Appl Math 10 pages. Article ID: 902715
Rao CR, Toutenburg H, Shalab, Heumann C (2008) Linear models and generalizations: least squares and alternatives. Springer, New York
Roozbeh M, Arashi M (2013) Feasible ridge estimator in partially linear models. J Multivar Anal 116:35–44
Roozbeh M, Arashi M, Niroumand HA (2011) Ridge regression methodology in partial linear models with correlated errors. J Stat Comput Simul 81(4):517–528
Schaffrin B, Toutenburg H (1990) Weighted mixed regression. Z Angew Math Mech 70:735–738
Tabakan G, Akdeniz F (2010) Difference-based ridge estimator of parameters in partial linear model. Stat Pap 51:357–368
Theil H, Goldberger AS (1961) On pure and mixed statistical estimation in Economics. Int Econ Rev 2:65–78
Toutenburg H, Srivastava VK, Schaffrin B, Heumann C (2003) Efficiency properties of weighted mixed regression estimation. METRON - Int J Stat LXI(1):91–103
Trenkler G, Toutenburg H (1990) Mean squared error matrix comparisons between biased estimators: an overview of recent results. Stat Pap 31:165–179
Wang L, Brown LLD, Cai TT (2011) A difference-based approach to semiparametric partial linear model. Electron J Stat 5:619–641
Wu JB (2016) A jackknifed difference-based ridge estimator in the partial linear model with correlated errors. Statistics 50(6):1363–1375
Yatchew A (1997) An elemantary estimator of the partial linear model. Econ Lett 57:135–143
Yatchew A (2000) Scale economics in electricity distribution. J Appl Econ 15:187–210
Yatchew A (2003) Semiparametric regression for the applied econometrican. Cambridge University Press, Cambridge