Pak. J. Statist.
2015 Vol. 30(2), 187-210

ON THE PERFORMANCE OF SOME RIDGE ESTIMATORS IN PARTIALLY LINEAR MODELS WITH HETEROSKEDASTIC AND AUTOCORRELATED ERRORS

Gülin Tabakan[1] and **Asuman S. Turkmen**[2]
[1] Department of Mathematics, Aksaray University
68100, Aksaray, Turkey. Email: gtabakan@aksaray.edu.tr
[2] Department of Statistics, The Ohio State University
Columbus, OH, USA. Email: turkmen.2@osu.edu

ABSTRACT

This paper is concerned with a partially linear regression model with unknown regression coefficients, an unknown nonparametric function for the nonlinear component with correlated and uncorrelated random errors. The estimation of covariance matrices of parameter estimates are modeled by Newey-West heteroscedasticity and autocorrelation consistent estimator when the errors are dependent. Real and simulated data sets are utilized to demonstrate the performance of the biased estimators.

KEYWORDS

Autocorrelation; consistent estimator; difference-based estimator; heteroscedasticity; multicollinearity; partially linear regression model.

AMS Classification: (Primary: 62G08); (Secondary: 62J07)

1. INTRODUCTION

Partially linear models are popular semiparametric modeling techniques assuming that the response variable of interest depends linearly on some covariates, while its relation to other additional variables is characterized by a nonparametric function. In this paper, we consider the following partially linear model

$$y_i = Z_i\beta + f\left(x_i\right) + \varepsilon_i, \quad i = 1,\ldots,n, \tag{1}$$

where y_i's are observations, Z_i's are known p-dimensional covariate vectors, $\beta = \left(\beta_1, \beta_2, \ldots, \beta_p\right)'$ is an unknown p-dimensional parameter vector. All we know about $f(.)$ is that its first derivative is bounded by a constant, say L. The $x's$ have a bounded support, say the unit interval, and have been reordered so that $x_1 \le x_2 \le \ldots \le x_n$, where n is the number of observations in the sample and ε_i's are random errors assumed to be independently and identically distributed (i.i.d.) with $N\left(0, s^2\right)$. The main goal is to estimate the unknown parameter vector β and the nonparametric function f from the data $\{y_i, Z_i, x_i\}_{i=1}^{n}$. The estimating of β and f in model (1) has been studied by several authors including Engle et al. (1986), Speckman (1988), Eubank et al. (1988), Eubank (1999). Further examples and discussions of model (1) may be found in Ruppert et al. (2003), Härdle (2004), Yatchew (1997; 1999; 2003), Wang et al., 2011.

The idea of differencing in nonparametric and partially linear models is not new. The simplicity of this procedure makes this methodology practical for empirical researchers. The differencing device provides a convenient means for introducing nonparametric techniques to practitioners in a way which parallels their knowledge of parametric techniques. The reason is that it allows one to remove the nonparametric effect and to analyze the parametric portion of the model as if the nonparametric portion was not there to begin with. Such methods are often called difference-based estimators. This procedure may be easily combined with other procedures. Such as, Yatchew used this technique to estimate the parametric component of the partially linear model (1) (Yatchew, 1997; 2003). Conventional estimators require one to use the nonparametric regression techniques and they require the selection of a smoothing parameter. In contrast, differencing procedure can be performed even if nonparametric regression procedures are not available within the software being used and does not require the selection of a smoothing parameter.

In this article, a difference-based estimation method is considered (Yatchew, 1997; 1999; 2003). Following the approach suggested by Yatchew (1997; 2003), to fit the partially linear model (1), we first rearrange the data such that $0 \le x_1 \le x_2 \le ... \le x_n \le 1$. Suppose you are given data $(y_1, Z_1, x_1)...(y_i, Z_i, x_i)$ on the model $y = Z\beta + f(x) + \varepsilon$ and the conditional mean of Z is a smooth function of x, say $E(Z|x) = r(x)$ where r' is bounded and $Var(Z|x) = \sigma_u^2$. Then we may rewrite $Z = r(x) + u$. If we first difference the ordered observations we obtain,

$$y_i - y_{i-1} = (Z_i - Z_{i-1})\beta + (f(x_i) - f(x_{i-1})) + \varepsilon_i - \varepsilon_{i-1},\ i = 2,...,n.$$
$$= (r(x_i) - r(x_{i-1}) + u_i - u_{i-1})\beta + (f(x_i) - f(x_{i-1})) + \varepsilon_i - \varepsilon_{i-1} \qquad (2)$$

If the observations are sufficiently close and functions $r(x)$ and $f(x)$ are smooth with bounded derivatives then a consistent estimate for β can be derived using ordinary least squares (OLS) and the vector of estimated parameters, denoted by $\hat{\beta}_{diff}$, is

$$\hat{\beta}_{diff} = \frac{\sum (y_i - y_{i-1})(x_i - x_{i-1})}{\sum (x_i - x_{i-1})^2}.$$

It is important to note that we assume here that data have already been ordered according to the x's, but in general, one should start difference-based estimation by ordering the data. Permutation matrices (see, Yatchew, 2003) can be used to reorder data.

Since differencing introduces first-order moving average structure in the error term, the OLS estimator is expected to lose efficiency. Fortunately, the efficiency of the estimator can be improved substantially by using higher order differencing and appropriate differencing weights. Therefore, the general form of (2) for the m th-order differencing can be written as

$$\sum_{j=1}^{m} d_j y_{i-j} = \beta\left(\sum_{j=1}^{m} d_j Z_{i-j}\right) + \sum_{j=1}^{m} d_j f\left(x_{i-j}\right) + \sum_{j=1}^{m} d_j \varepsilon_{i-j} \quad i = 2,\ldots,n \tag{3}$$

where m is the order of differencing and $d_0,\ldots,d_m$ are differencing weights.

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

$$\sum_{j=1}^{m} d_j = 0, \qquad \sum_{j=1}^{m} d_j^2 = 1 . \tag{4}$$

The first condition in (4) ensures that the differencing removes the nonparametric component in (3) as the sample size increases, while the second normalization condition implies that the residuals in (3) have variance of σ_ε^2 . When the differencing weights are chosen optimally, the difference-based estimator, $\hat{\beta}_{diff}$, obtained by regressing Dy on DZ approaches asymptotic efficiency by selecting m sufficiently large. With the optimal choice of weights in (3), one can estimate β using OLS (Yatchew, 1997; 2003). Let us define to be the vector with elements Dy to be the $\left(n-m\right)\times 1$ vector with elements and DZ to be the $\left(n-m\right)\times p$ matrix with elements $\left(DZ\right)_i = \sum_{j=1}^{m} d_j Z_{i-j}$. In matrix notation, (3) can be rewritten as

$$\tilde{y} \approx \tilde{Z}\beta + \tilde{\varepsilon}, \tag{5}$$

where $\tilde{y} = Dy, \tilde{Z} = DZ$, $\tilde{\varepsilon} = D\varepsilon$ and $\tilde{Z}$ is a full-rank matrix. Thus, from the least square perspective, a simple difference-based estimator of the parameter β can be written as

$$\hat{\beta}_{diff} = \left(\tilde{Z}'\tilde{Z}\right)^{-1} \tilde{Z}'\tilde{y} = U\tilde{Z}'\tilde{y}, \tag{6}$$

where $U = \left(\tilde{Z}'\tilde{Z}\right)^{-1}$. Once $\hat{\beta}_{diff}$ is estimated, a variety of nonparametric techniques could be applied to estimate f as if β were known. Formally, subtracting the estimated parametric part from both sides of (1), we get:

$$y_i - Z_i\hat{\beta}_{diff} = Z_i\left(\beta - \hat{\beta}_{diff}\right) + f\left(x_i\right) + \varepsilon_i \cong f\left(x_i\right) + \varepsilon_i, \quad i = 1,\ldots,n$$

Because $\hat{\beta}_{diff}$ converges sufficiently quickly to true β, the consistency, optimal rate of convergence, and construction of confidence intervals for f remain valid and f could be estimated by the standard smoothing methods. Through the paper, main focus is the estimation of β and statistical inference for the nonparametric function f is omitted.

In practical applications, there is often some degree of multicollinearity problem among the covariates which can be measured by different tools such as the condition number. If $\tilde{Z}'\tilde{Z}$ is ill conditioned with a large condition number, then classical difference-based estimator generally produces poor estimates of parameters, therefore some biased estimators can be utilized to estimate β. To avoid that problem, several biased estimation methods have been proposed for linear models (Hoerl and Kennard, 1970; Hoerl et al. 1975; Liu, 1993; 2003) and some of them adopted for partially linear models (Tabakan and Akdeniz, 2010; Akdeniz and Duran, 2010; Duran and Akdeniz, 2011; Haibing and Jinhong, 2011; Duran et al., 2012; Luo, 2012; Tabakan, 2013; Roozbeh and Arashi, 2014; Wu, 2014).

The paper is organized as follows. Section 2 introduces difference-based ridge and restricted ridge estimators for the partially linear model. In section 3, we consider a simple version of the partially linear model in the case of independent errors with equal variance and give conditions under which the proposed estimators are superior to the difference-based estimation technique in the sense of mean squared error. Section 4 relaxes the assumption of i.i.d. errors and rederives the results of the previous section in the presence of heteroscedasticity and autocorrelation. Section 5 and 6 give a Monte Carlo simulation study and a numerical example to show the performance of the proposed estimator.

2. DIFFERENCE-BASED ESTIMATORS

From the least squares perspective, the coefficients β in (5) are chosen to minimize

$$L(\beta) = \left(\tilde{y} - \tilde{Z}\beta\right)'\left(\tilde{y} - \tilde{Z}\beta\right). \tag{7}$$

Adding a penalizing function, i.e. $k\|\beta\|^2$, to the least square objective (7), yields

$$L^*(\beta, k) = L(\beta) + k\beta'\beta, \tag{8}$$

where k is a pre-selected biasing parameter. Tabakan and Akdeniz (2010) have proposed the difference-based ridge estimator $\hat{\beta}(k)$ obtained by minimizing (8) and given as

$$\hat{\beta}(k) = S_k \tilde{Z}'\tilde{y}, \ (k \geq 0) \tag{9}$$

where $S_k = \left(\tilde{Z}'\tilde{Z} + kI\right)^{-1}$, I is the $p \times p$ identity matrix. Note that $\hat{\beta}(k) = \hat{\beta}_{diff}$ when $k = 0$.

Now, we consider the linear nonstochastic constraint

$$R\beta = r, \tag{10}$$

for a given $m \times p$ known matrix R with rank $m < p$ and a given $m \times 1$ known vector r. In this case the estimate of β in model (5), obtained by minimizing the Equation (8) under the condition (10), is equal to

$$\hat{\beta}_r(k) = \hat{\beta}(k) - S_k R' \left[R S_k R' \right]^{-1} \left(R\hat{\beta}(k) - r \right). \tag{11}$$

We call $\hat{\beta}_r(k)$ a restricted difference-based ridge estimator of the partially linear model (see, Tabakan, 2013).

In the following sections, we will compare classical difference-based estimator with these two ridge estimators using the mean square error matrix (MSEM) criterion. In general, if $\hat{\beta}_1$ and $\hat{\beta}_2$ are two estimators of β, $\hat{\beta}_2$ is called MSE superior to $\hat{\beta}_1$ if the difference of their MSEMs is nonnegative definite. Specifically, necessary and sufficient conditions for the difference-based ridge estimators to be MSE superior then the classical difference-based estimators are given in the Sections 3 and 4 for independent and dependent random errors, respectively. The comparison between difference-based ridge and restricted ridge estimators when using the same value of k, is also provided.

3. INDEPENDENT CASE

The aim of Section 3 is to compare the mean square error matrices of the estimators $\hat{\beta}(k)$, $\hat{\beta}_r(k)$ and $\hat{\beta}_{diff}$ for the simple version of the partially linear model (1) with independent errors.

We note that for any estimator $\tilde{\beta}$ of β in a linear model, MSEM is defined as $MSEM(\tilde{\beta}) = Cov(\tilde{\beta}) + Bias(\tilde{\beta}) Bias(\tilde{\beta})'$, where $Cov(\tilde{\beta})$ denotes the variance-covariance matrix and $Bias(\tilde{\beta}) = E(\tilde{\beta}) - \beta$.

3.1 Comparisons between Difference-Based Ridge Estimator, $\hat{\beta}(k)$ and Difference-Based Estimator $\hat{\beta}_{diff}$

Using the estimator $\hat{\beta}(k)$ in (9), the variance-covariance matrix and bias of $\hat{\beta}(k)$ are given by

$$Cov\left(\hat{\beta}(k)\right) = \sigma^2 S_k S S_k, \tag{12}$$

$$Bias\left(\hat{\beta}(k)\right) = -k S_k \beta, \tag{13}$$

respectively, and therefore MSEM of $\hat{\beta}(k)$ is

$$MSEM\left(\hat{\beta}(k)\right) = \sigma^2 S_k S S_k + k^2 S_k \beta' \beta S_k. \tag{14}$$

Similarly, the variance-covariance matrix and MSEM of unbiased estimator $\hat{\beta}_{diff}$ in (6) are given as

$$Cov\left(\hat{\beta}_{diff}\right) = MSEM\left(\hat{\beta}_{diff}\right) = \sigma^2 USU, \tag{15}$$

where $S = \left(D'\tilde{Z}\right)'\left(D'\tilde{Z}\right)$. From (12) and (15), one write the difference

$$V_1 = Cov\left(\hat{\beta}_{diff}\right) - Cov\left(\hat{\beta}(k)\right) = \sigma^2\left(USU - S_k S S_k\right)$$
$$= k^2\sigma^2 S_k\left(USU + \frac{1}{k}\left(SU + US\right)\right)S_k. \tag{16}$$

The necessary and sufficient conditions at which V_1 is a positive definite (p.d.) matrix (that is the estimator $\hat{\beta}(k)$ has a smaller variance compared to one of estimator $\hat{\beta}_{diff}$) are given in Tabakan and Akdeniz (2010). They also proved the following theorem for the difference-based ridge estimator stating the conditions that $\hat{\beta}(k)$ is MSE superior to the difference-based estimator, $\hat{\beta}_{diff}$ when V_1 is a positive definite matrix.

Theorem 3.1.

Consider two competing estimators $\hat{\beta}(k)$ and $\hat{\beta}_{diff}$ of β. Let the difference $Cov\left(\hat{\beta}_{diff}\right) - Cov\left(\hat{\beta}(k)\right)$ be p.d. Then the biased estimator $\hat{\beta}(k)$ is MSE superior over the $\hat{\beta}_{diff}$ if and only if the following inequality holds:

$$\beta' W_1^{-1}\beta \le 1 \tag{17}$$

where $W_1 = \sigma^2\left(USU + \frac{1}{k}\left(SU + US\right)\right)$.

Proof:

See Tabakan and Akdeniz (2010).

3.2 Comparisons between Difference-Based Estimator $\hat{\beta}_{diff}$ and Restricted Difference-Based Ridge Estimator $\hat{\beta}_r(k)$

Let us now compare the sampling variance-covariance matrices of $\hat{\beta}_{diff}$ and $\hat{\beta}_r(k)$. When the restrictions in (10) are assumed to be true, the variance-covariance matrix and bias of $\hat{\beta}_r(k)$ are given by

$$Cov\left(\hat{\beta}_r(k)\right) = \sigma^2 FSF', \tag{18}$$

$$Bias\left(\hat{\beta}_r(k)\right) = -kF\beta, \tag{19}$$

respectively, where

$$\begin{aligned} F &= S_k - S_k R'\left(RS_k R'\right)^{-1} RS_k \\ &= S_k\left[I - R'\left(RS_k R'\right)^{-1} RS_k\right] = S_k B' \\ &= \left[I - S_k R'\left(RS_k R'\right)^{-1} R\right] S_k = BS_k, \end{aligned}$$

with $B = I - S_k R'\left(RS_k R'\right)^{-1} R.$ The MSEM of $\hat{\beta}_r(k)$ in Equation (11) is given by

$$MSEM\left(\hat{\beta}_r(k)\right) = \sigma^2 FSF' + k^2 F\beta\beta' F'. \tag{20}$$

From, (15) and (18) the difference $V_2 = Cov\left(\hat{\beta}_{diff}\right) - Cov\left(\hat{\beta}_r(k)\right)$ can be written as

$$\begin{aligned} V_2 &= Cov\left(\hat{\beta}_{diff}\right) - Cov\left(\hat{\beta}_r(k)\right) \\ &= \sigma^2 USU - \sigma^2 FSF' \\ &= \sigma^2 U\left[S - E'SE\right]U, \end{aligned}$$

where $E = B\left(I + kU\right)^{-1}$. V_2 is a p.d. matrix under the conditions described in the next theorem.

Theorem 3.2.

The sampling variance of $\hat{\beta}_r(k)$ is less than that of the $\hat{\beta}_{diff}$ if and only if

$$\lambda_{\max}\left[S^{-1}\left(E'SE\right)\right] < 1, \tag{21}$$

where E is as defined before.

Proof:

Let $S = \left(D'\tilde{Z}\right)'\left(D'\tilde{Z}\right) = \left(D'DZ\right)'\left(D'DZ\right) = H'H$, where S is a $p \times p$ positive definite matrix, $rank\left(D'DZ\right) = rank(H) = p < n - m$ $\left(m < n\right)$, and $U = \left(\tilde{Z}'\tilde{Z}\right)^{-1}$ is a nonsingular and symmetric matrix as defined (Tabakan and Akdeniz, 2010; Duran and Akdeniz, 2011). Since S is a p.d. matrix and $K = E'SE$ is a symmetric matrix, there exists a nonsingular matrix Q such that $Q'SQ = I$ and $Q'KQ = \Lambda$, where Λ is a diagonal matrix and its diagonal elements are the eigenvalues of $S^{-1}K$ (Schott, 2005). Thus, the difference V_2 can be rewritten as

$$V_2 = Cov\left(\hat{\beta}_{diff}\right) - Cov\left(\hat{\beta}_r(k)\right)$$
$$= \sigma^2 U Q^{-1}\left[Q'SQ - Q'KQ\right]Q^{-1}U = \sigma^2 U Q^{-1}\left[I - \Lambda\right]Q^{-1}U.$$

It is seen that the estimator $\hat{\beta}_r(k)$ has a smaller variance than $\hat{\beta}_{diff}$, i.e. V_2 becomes a p.d. matrix, if and only if $\lambda_{\max}\left(S^{-1}K\right) < 1$ or equivalently $\lambda_{\max}[S^{-1}\left(E'SE\right)] < 1$.

The following lemma from Farebrother (1976) is needed to prove Theorem 3.3 that allows one to compare the MSE matrices of $\hat{\beta}_{diff}$ and $\hat{\beta}_r(k)$.

Lemma 3.1.

Let A be a positive definite $p \times p$ matrix, b a $p \times 1$ nonzero vector, and α a positive scalar. Then $\alpha A - bb'$ is positive (semi-) definite if and only if $b'A^{-1}b$ is less than (or equal to) α (Farebrother, 1976).

Theorem 3.3.

Consider two competing estimators $\hat{\beta}_r(k)$ and $\hat{\beta}_{diff}$ of β. Let the difference $Cov\left(\hat{\beta}_{diff}\right) - Cov\left(\hat{\beta}_r(k)\right)$ be p.d. Then the biased estimator $\hat{\beta}_r(k)$ is MSE superior over the $\hat{\beta}_{diff}$ if and only if the following inequality holds:

$$\beta' W_2^{-1} \beta \le 1, \qquad (22)$$

where $W_2 = \dfrac{\sigma^2}{k^2} F^{-1} U\left[S - E'SE\right]U\left(F^{-1}\right)'$.

Proof:

Using (15) and (20), we find that

$$\Delta_2 = MSEM\left(\hat{\beta}_{diff}\right) - MSEM\left(\hat{\beta}_r(k)\right)$$
$$= V_2 - k^2 F\beta\beta' F'$$
$$= k^2 F\left\{\frac{\sigma^2}{k^2} F^{-1} U\left[S - E'SE\right]U\left(F^{-1}\right)' - \beta\beta'\right\}F'$$
$$= k^2 F\left[W_2 - \beta\beta'\right]F'.$$

where $W_2 = \dfrac{\sigma^2}{k^2} F^{-1} U\left[S - E'SE\right]U\left(F^{-1}\right)'$. Using Lemma 3.1, the stated result is obtained.

3.3 Comparisons between Difference-Based Ridge Estimator $\hat{\beta}(k)$ and Restricted Difference-Based Ridge Estimator $\hat{\beta}_r(k)$

In this section, our objective is to examine the difference between MSEMs of two estimators $\hat{\beta}(k)$ and $\hat{\beta}_r(k)$. Using (12) and (18),

$$\begin{aligned} V_3 &= Cov\left(\hat{\beta}(k)\right) - Cov\left(\hat{\beta}_r(k)\right) \\ &= \sigma^2 S_k S S_k - \sigma^2 F S F' = \sigma^2 S_k \left(S - B'\ SB\right) S_k, \end{aligned}$$

where $B = I - S_k R'\left(R S_k R'\right)^{-1} R$, S is a $p \times p$ positive definite matrix and S_k is a nonsingular and symmetric matrix. Now we can give the following theorem describing the conditions for V_3 to be p.d.

Theorem 3.4.

The sampling variance of $\hat{\beta}_r(k)$ is less than that of the $\hat{\beta}(k)$ if and only if

$$\lambda_{\max}\left[S^{-1}\left(B' S B\right)\right] < 1. \tag{23}$$

Proof:

The difference V_3 can be rewritten as

$$\begin{aligned} V_3 &= Cov\left(\hat{\beta}(k)\right) - Cov\left(\hat{\beta}_r(k)\right) \\ &= \sigma^2 S_k O^{-1}\left[O' S O - O' \mathsf{C} O\right] O^{-1} S_k \\ &= \sigma^2 S_k O^{-1}\left[I - \Lambda_1\right] O^{-1} S_k \end{aligned}$$

where $\mathsf{C} = B' S B$ is a symmetric matrix, O is a nonsingular matrix such that $O' S O = I$ and $O' \mathsf{C} O = \Lambda_1$, and Λ_1 is a diagonal matrix and its diagonal elements are the eigenvalues of $S^{-1}\mathsf{C}$. Therefore, V_3 becomes a p.d. matrix, if and only if $\lambda_{\max}\left(S^{-1}\mathsf{C}\right) < 1$ or $\lambda_{\max}\left[S^{-1}\left(B'\ SB\right)\right] < 1$.

Theorem 3.4 states that $\hat{\beta}_r(k)$ is superior to $\hat{\beta}(k)$ in terms of MSEM under certain conditions which can be proved by direct application of the following Lemma 3.2.

Lemma 3.2.

Let $\tilde{\beta}_j = A_j y,\ j = 1, 2$ be two linear estimators of β. Suppose that the difference $Cov\left(\tilde{\beta}_1\right) - Cov\left(\tilde{\beta}_2\right) > 0$, where $Cov\left(\tilde{\beta}_j\right), j = 1, 2$ denotes the covariance matrix of $\tilde{\beta}_j$. Then $MSEM\left(\tilde{\beta}_1\right) - MSEM\left(\tilde{\beta}_2\right) \geq 0$ if and only if $d_2'\left(Cov\left(\tilde{\beta}_1\right) - Cov\left(\tilde{\beta}_2\right) + d_1 d_1'\right)^{-1} d_2 \leq 1,$

where $MSEM\left(\tilde{\beta}_j\right)$, d_j, $j=1,2$ denote the MSEM and bias vector of $\tilde{\beta}_j$, respectively (see, Trenkler and Toutenburg, 1990).

Theorem 3.5.

Consider two competing estimators $\hat{\beta}_r(k)$ and $\hat{\beta}(k)$ of β Let the difference $Cov\left(\hat{\beta}(k)\right)-Cov\left(\hat{\beta}_r(k)\right)$ be p.d. Then the biased estimator $\hat{\beta}_r(k)$ is MSE superior over $\hat{\beta}(k)$ if and only if the following inequality holds:

$$d_2'\left(Cov\left(\hat{\beta}(k)\right)-Cov\left(\hat{\beta}_r(k)\right)+d_1d_1'\right)^{-1}d_2 \le 1, \tag{24}$$

with $d_1=-kS_k\beta,\ d_2=-kF\beta.$

Proof:

The difference between the MSEM of $\hat{\beta}(k)$ given in (14) and that of $\hat{\beta}_r(k)$ in (20) is denoted by Δ_3 and is equal to

$$\begin{aligned}\Delta_3 &= MSEM\left(\hat{\beta}(k)\right)-MSEM\left(\hat{\beta}_r(k)\right)\\ &=\sigma^2S_k\left(S-B'SB\right)S_k+k^2S_k\beta\beta'S_k-k^2F\beta\beta'F'\\ &=Cov\left(\hat{\beta}(k)\right)-Cov\left(\hat{\beta}_r(k)\right)+d_1d_1'-d_2d_2'\end{aligned}$$

where $Cov\left(\hat{\beta}(k)\right)-Cov\left(\hat{\beta}_r(k)\right)=\sigma^2S_k\left(S-B'SB\right)S_k$, $d_1=-kS_k\beta$ and $d_2=-kF\beta.$ Applying Lemma 3.2 on Δ_3 finalizes the proof.

4. DEPENDENT CASE

In the previous sections, it has been assumed that the original observations are independent of each other. However, autocorrelation (dependence) among the errors in the regression model is an important problem faced in applications. When autocorrelation is present in the errors the OLS estimator will not be efficient and the usual estimator of the variance-covariance matrix will be biased. To combat these effects, more general case of heteroscedasticity and autocorrelation in the error terms is considered.

Suppose we have data $\{y_i, Z_i, x_i\}_{i=1}^{n}$ as described earlier on the semiparametric model $y=Z\beta+f(x)+\varepsilon$. Let $Var(\varepsilon\varepsilon')=\Omega$ is not necessarily diagonal as assumed so far. In other words, here we relax homoscedasticity and independence assumptions on errors. If $Var(\varepsilon\varepsilon')=\Omega$, then the covariance matrix for the difference-based estimation becomes

$$Cov\left(\hat{\beta}_{diff}\right)=U\tilde{S}U\ , \tag{25}$$

where $U = \left(\tilde{Z}'\tilde{Z}\right)^{-1}$ and $\tilde{S} = \tilde{Z}'D\Omega D'\tilde{Z}$. We now need a consistent estimate of $D\Omega D'$ to have a consistent estimator of $Cov\left(\hat{\beta}_{diff}\right)$. For this purpose, we will use Newey-West heteroscedasticity and autocorrelation consistent estimator (Newey and West, 1987). Let L be the maximum lag which exhibits non-zero autocorrelation in the matrix Ω and L_ℓ be a matrix with ones on the l-th diagonal. Define matrices H^ℓ, $l = 0,...,L$ as follows. Let H^0 be the identity matrix. For $l = 1,...,$ L let

$$H_{ij}^{\ell} = \begin{cases} 1, & if \left[D\left(L_\ell + L_\ell'\right)D' \right]_{ij} \neq 0 \\ 0, & otherwise \end{cases}.$$

Thus we define

$$D\hat{\Omega}D' = \left(\hat{D\varepsilon}\left(\hat{D\varepsilon}\right)'\right)\Theta\left(\sum_{\ell=0}^{L}\left(\frac{\ell}{L+1}\right)H^{\ell}\right), \tag{26}$$

with $\hat{D\varepsilon} = Dy - DZ\hat{\beta}_{diff}$, Θ denoting the element-wise matrix product. Plugging (26) in (25), we have a consistent estimator for $Cov\left(\hat{\beta}_{diff}\right)$; (see, Yatchew, 1999 for details).

Similarly, we can write down $Cov\left(\hat{\beta}(k)\right)$ and $Cov\left(\hat{\beta}_r(k)\right)$ as:

$$Cov\left(\hat{\beta}(k)\right) = S_k\tilde{S}S_k \tag{27}$$

$$Cov\left(\hat{\beta}_r(k)\right) = F\tilde{S}F',$$

(28)

where S_k and F as defined in Section 3. Using (25) and (27), the difference $V_1 = Cov\left(\hat{\beta}_{diff}\right) - Cov\left(\hat{\beta}(k)\right)$ can be given as

$$\begin{aligned} V_1 &= \left(U\tilde{S}U - S_k\tilde{S}S_k\right) \\ &= k^2 S_k\left(U\tilde{S}U + \frac{1}{k}\left(\tilde{S}U + U\tilde{S}\right)\right)S_k \\ &= k^2 S_k\left(\tilde{M} + \eta\tilde{N}\right)S_k, \end{aligned}$$

where $\eta = \frac{1}{k} > 0$, $\tilde{M} = U\tilde{S}U$, $\tilde{N} = U\tilde{S} + \tilde{S}U$. Since $\tilde{M}$ is $\left(p \times p\right)$ positive definite matrix and $\tilde{N}$ is a symmetric matrix there exists a nonsingular matrix T such that $T'\tilde{M}T = I$ and $T'\tilde{N}T = \tilde{G}$, where $\tilde{G}$ is a diagonal matrix and its diagonal elements are the roots λ of the polynomial equation $\left|\tilde{N}-\lambda\tilde{M}\right|=\left|\tilde{M}^{-1}\tilde{N}-\lambda I\right|=0$ (Graybill, 1993; Haville, 1997) and we may write V_1 as in the following form:

$$V_1 = k^2 S_k \left(T^{-1}\right)' \left[T'\ \tilde{M}T + \eta T'\ \tilde{N}T\right] T^{-1} S_k ,$$

$$= k^2 S_k \left(T^{-1}\right)' \left[I + \eta\tilde{G}\right] T^{-1} S_k .$$

where $I+\eta\tilde{G} = diag\left(1+\eta\tilde{g}_{11},...,1+\eta\tilde{g}_{pp}\right)$. Since $\tilde{N} = U\tilde{S} + \tilde{S}U \neq 0$ there is at least one diagonal element of $\tilde{G}$ that is nonzero. Let $\tilde{g}_{ii} \neq 0$,

$$0 < \eta < \min_{\tilde{g}_{ii}\neq 0}\left|\frac{1}{\tilde{g}_{ii}}\right| \tag{29}$$

and hence $1+\eta\tilde{g}_{ii} > 0$ for all $i = 1,..., p$ and $I+\eta\tilde{G}$ is a positive definite matrix. Hence, V_1 becomes positive definite matrix. It is now evident that the estimator $\hat{\beta}(k)$ has a smaller variance compared with the estimator $\hat{\beta}_{diff}$ if and only if (29) is satisfied. Thus we can give the following theorem.

Theorem 4.1.

Consider the estimators $\hat{\beta}_{diff}$ and $\hat{\beta}(k)$ of β. Let $Cov\left(\hat{\beta}_{diff}\right)-Cov\left(\hat{\beta}(k)\right)$ be a positive definite matrix. Then the biased estimator $\hat{\beta}(k)$ is MSE superior to $\hat{\beta}_{diff}$ if and only if the following inequality holds:

$$\beta'\tilde{W}_1^{-1}\beta \leq 1 \tag{30}$$

where $\tilde{W}_1 = \left(U\tilde{S}U + \frac{1}{k}\left(\tilde{S}U + U\tilde{S}\right)\right)$.

Proof:

Consider the differences

$$\Delta_1' = MSEM\left(\hat{\beta}_{diff}\right) - MSEM\left(\hat{\beta}(k)\right)$$

$$= Cov\left(\hat{\beta}_{diff}\right) - Cov\left(\hat{\beta}(k)\right) - bias\left(\hat{\beta}(k)\right) bias\left(\hat{\beta}(k)\right)'$$

$$= S_k \left\{k\left(\tilde{S}U + U\tilde{S}\right) + k^2 U\tilde{S}U - k^2\beta\beta'\right\} S_k$$

$$= k^2 S_k \left(\eta\tilde{N} + \tilde{M} - \beta\beta'\right) S_k$$

$$= k^2 S_k \left(\tilde{W}_1 - \beta\beta'\right) S_k ,$$

where $\tilde{W}_1 = \tilde{M} + \eta \tilde{N}$, $\tilde{M} = U\tilde{S}U$, $N = U\tilde{S} + \tilde{S}U$ and $\eta = \frac{1}{k}$. Applying Lemma 3.1, the assertation follows (see, Duran et.al., 2012).

Theorem 4.1 gives conditions under which the biased estimator $\hat{\beta}(k)$ is superior to $\hat{\beta}_{diff}$ in the presence of heteroscedasticity and autocorrelation in the data.

With similar argumentation as above obtained in Section 3.2, Theorem 3.2 can be extended straight forwardly to the general case by exchanging S by $\tilde{S} = \tilde{Z}'D\Omega D'\tilde{Z}$. Hence, the estimator $\hat{\beta}_r(k)$ has a smaller variance than $\hat{\beta}_{diff}$ if and only if $\lambda_{\max}\left[\tilde{S}^{-1}\left(E'\tilde{S}E\right)\right] < 1$, where $\lambda_{\max}$ is the maximum eigenvalue of $\tilde{S}^{-1}\left(E'\tilde{S}E\right)$.

Now, we can give a generalized version of Theorem 3.3 to compare the MSE matrices of $\hat{\beta}_{diff}$ and $\hat{\beta}_r(k)$.

Theorem 4.2.

Consider the estimators $\hat{\beta}_{diff}$ and $\hat{\beta}_r(k)$ of β. Let $Cov\left(\hat{\beta}_{diff}\right) - Cov\left(\hat{\beta}(k)\right)$ be a positive definite matrix. Then the biased estimator $\hat{\beta}_r(k)$ is MSE superior to $\hat{\beta}_{diff}$ if and only if the following inequality holds:

$$\beta' \tilde{W}_2^{-1} \beta \leq 1 \qquad (31)$$

where $\tilde{W}_2 = \frac{1}{k^2} F^{-1} U \left[\tilde{S} - E'\tilde{S}E\right] U \left(F^{-1}\right)'$.

Proof:

Similar to the proof of Theorem 3.3.

Note that for comparison of biased estimators $\hat{\beta}_r(k)$ and $\hat{\beta}(k)$, Theorem 3.4 can be extended for heteroscedastic and autocorrelated errors by exchanging S by $\tilde{S}$. Hence, the sampling variance of $\hat{\beta}(k)$ is always smaller than that of $\hat{\beta}_r(k)$, if and only if $\lambda_{\max}\left[\tilde{S}^{-1}\left(B'\tilde{S}B\right)\right] < 1$, where $\lambda_{\max}$ is the maximum eigenvalue of $\tilde{S}^{-1}\left(B'\tilde{S}B\right)$. Thus, the next theorem extends the results of Theorem 3.5 of Section 3.3 to the more general case of (24).

Theorem 4.3.

Consider two competing estimators $\hat{\beta}_r(k)$ and $\hat{\beta}(k)$ of β Let the difference $Cov\left(\hat{\beta}(k)\right) - Cov\left(\hat{\beta}_r(k)\right)$ be p.d. Then $\hat{\beta}_r(k)$ is MSE superior over the $\hat{\beta}(k)$ if and only if the following inequality holds:

$$d_2'\left(Cov\left(\hat{\beta}(k)\right)-Cov\left(\hat{\beta}_r(k)\right)+d_1d_1'\right)^{-1}d_2\le 1, \quad (32)$$

where $Cov\left(\hat{\beta}(k)\right)-Cov\left(\hat{\beta}_r(k)\right)=S_k\left(\tilde{S}-B'\tilde{S}B\right)S_k$, $d_1=-kS_k\beta$ and $d_2=-kF\beta$.

Proof:

Similar to the proof of Theorem 3.5.

In the following sections, we investigate the numerical performances of the estimators using both simulated and real data sets.

5. A MONTE CARLO EXPERIMENT

In this section, the theoretical comparisons are extended with a Monte Carlo study. We compare the performances of classical difference-based, ridge and the restricted ridge estimators in terms of simulated MSE values. The MSE estimates of $\hat{\beta}_{diff}$, $\hat{\beta}(k)$, and $\hat{\beta}_r(k)$ are obtained under different degrees of collinearity as well as different levels of error variances. All computations were conducted using R version 3.0.0.

To compare the three estimators, following McDonald and Galarneau (1975), the explanatory variables and the observations on the dependent variable are generated by

$$z_{ij}=\left(1-\gamma^2\right)^{1/2}w_{ij}+\gamma w_{ip+1},\quad i=1,\dots,n;\ j=1,\dots,p$$

where w_{ij} are independent standard normal pseudo-random numbers and γ^2 is the theoretical correlation between any two explanatory variables. For the nonparametric component in Equation (1), we selected a smooth function of the form $f(x_i)=2\sin(4\pi x_i)$ for $x_i=i/n$ on $[0, 1]$.

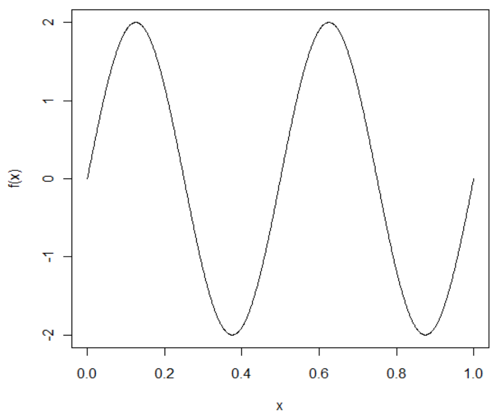

Fig. 1: Plot for the Nonparametric Sine Function

Observations on the dependent variable are then generated by the following model

$$y_i = \beta_1 z_{i1} + \beta_2 z_{i2} + ... + \beta_p z_{ip} + f(x_i) + \varepsilon_i, \quad i = 1, ..., n$$

where ε_i are independent standard normal pseudo-random numbers with mean zero and variance σ^2, $\varepsilon \sim N(0, \sigma^2)$. The parameter vector is taken as $\beta = [1 \ \ 2 \ \ ... \ \ p]'$ and fixed through the replications. Once $Z, \beta,$ f and ε are generated, the dependent variable vector y is simulated by $y = Z\beta + f + \varepsilon$. The data are standardized to obtain $Z'Z$ in correlation form. The resulting Z matrix, β and f are kept fixed through the replications.

For the independent errors with constant variance, the least squares estimate of error variance proposed by Eubank et al. (1988) is used and calculated as

$$\hat{\sigma}^2_{diff} = \frac{(Dy)'(I-H)Dy}{tr(D'(I-H)D)} \tag{33}$$

with $tr(.)$ the trace function for a square matrix and H the projection matrix

$$H = DZ\left[\left((DZ)'(DZ)\right)\right]^{-1}(DZ)'.$$

We consider three different correlation coefficients denoted by γ (i.e., 0.7, 0.95 and 0.995). To see the effect of the number of observations and the number of variables, two different combinations of n and $p: (n, p) = \{(20, 4), (60, 10)\}$ are used. The restriction for the restricted ridge estimator is $R\beta = r$ where $R = [1 \ \ 1 \ \ ... \ \ 1]$, $r = p(p+1)/2 \times c$ and c is a constant. The usage of c allows one to control whether the restriction is true or not and how illuminating the prior information for the parameter of interest and it is selected as 0.5, 1, or 1.5 throughout the replications. It is obvious that the restriction $\beta_1 + \beta_2 + ... + \beta_p = p(p+1)/2 \times c$ is true and prior information is exactly equal to the actual parameter when $c = 1$ yielding the optimal results for both estimators. The approaches of Hoerl et al. (1975) is used to specify the value of k. That is, the biasing parameter is set to $\hat{k}_{HKB} = \left(p\hat{\sigma}^2_{diff}\right) / \left(\hat{\beta}'_{diff}\hat{\beta}_{diff}\right) = 0.676$ where $\hat{\sigma}^2_{diff}$ is the classical least squares estimates of error variance given in (33). Although original error terms are independent and homoscedastic, differencing can introduce a correlation structure and heteroscedasticity in the model. Therefore, standard errors are calculated using classical least squares estimates and Newey-West estimators of the variances.

For the simulation study, we used order $m = 3$. For each choice of c, γ, n and p, the experiment is replicated 5000 times by generating new error terms for the fixed Z, β vector and f. Once 5000 samples are generated, we calculated $\hat{\beta}_{diff}$, $\hat{\beta}(k)$, $\hat{\beta}_r(k)$ and computed their respective simulated MSE values defined by

$$MSE\left(\beta^*\right)=\frac{1}{5000}\sum_{j=1}^{5000}\sum_{i=1}^{p}\left(\beta_{ij}^*-\beta_i\right)^2, \tag{34}$$

where β_{ij}^* denotes the estimate of the ith parameter in jth replication and β_i, $i=1,...,p$ are the true parameter values. Monte Carlo experiments conducted to examine the relative efficiency of competing estimators. In general, the relative efficiency of β_1^* and β_2^* is given as

$$eff\left(\beta_1^*,\beta_2^*\right)=\frac{MSE\left(\beta_2^*\right)}{MSE\left(\beta_1^*\right)}, \tag{35}$$

where β_1^* and β_2^* are two estimators of β. If the relative efficiency value obtained from (35) is smaller than 1, then it means that $\tilde{\beta}_2$ has a smaller MSE than β_1^*.

We obtain the simulated MSE values for the $\hat{\beta}_{diff}$, $\hat{\beta}(k)$ and $\hat{\beta}_r(k)$ using the Equation (34). Then, the relative efficiencies of $\hat{\beta}(k)$ with respect to $\hat{\beta}_{diff}$, $eff\left(\hat{\beta}_{diff},\hat{\beta}(k)\right)$, and $\hat{\beta}_r(k)$, $eff\left(\hat{\beta}_r(k),\hat{\beta}(k)\right)$, as well as the relative efficiency of $\hat{\beta}_r(k)$ with respect to $\hat{\beta}_{diff}$, $eff\left(\hat{\beta}_{diff},\hat{\beta}_r(k)\right)$, are calculated using (35) and summarized in Table 1 and Table 2.

Table 1
Relative Efficiencies of Estimators for $n=20$, $p=4$

γ	c	$\sigma=0.01$			$\sigma=0.1$			$\sigma=1$		
		0.5	1	1.5	0.5	1	1.5	0.5	1	1.5
0.7	$eff\left(\hat{\beta}_{diff},\hat{\beta}(k)\right)$	0.130	0.130	0.130	0.301	0.301	0.301	0.524	0.52	0.524
	$eff\left(\hat{\beta}_{diff},\hat{\beta}_r(k)\right)$	0.218	0.144	0.398	0.402	0.298	0.567	0.763	0.514	0.528
	$eff\left(\hat{\beta}(k),\hat{\beta}_r(k)\right)$	1.677	1.108	3.062	1.336	0.990	1.884	1.456	0.981	1.008
0.9	$eff\left(\hat{\beta}_{diff},\hat{\beta}(k)\right)$	0.107	0.107	0.107	0.107	0.107	0.107	0.322	0.322	0.322
	$eff\left(\hat{\beta}_{diff},\hat{\beta}_r(k)\right)$	0.107	0.089	0.214	0.152	0.111	0.231	0.369	0.322	0.328
	$eff\left(\hat{\beta}(k),\hat{\beta}_r(k)\right)$	1.005	0.833	2.008	1.422	1.042	2.156	1.146	0.999	1.019
0.995	$eff\left(\hat{\beta}_{diff},\hat{\beta}(k)\right)$	0.068	0.068	0.068	0.109	0.109	0.109	0.148	0.148	0.148
	$eff\left(\hat{\beta}_{diff},\hat{\beta}_r(k)\right)$	0.064	0.034	0.056	0.127	0.115	0.171	0.191	0.148	0.202
	$eff\left(\hat{\beta}(k),\hat{\beta}_r(k)\right)$	0.939	0.501	0.825	1.163	1.060	1.566	1.288	1.000	1.362

Table 2
Relative Efficiencies of Estimators for $n = 60$, $p = 10$

γ	c	$\sigma = 0.01$			$\sigma = 0.1$			$\sigma = 1$		
		0.5	1	1.5	0.5	1	1.5	0.5	1	1.5
0.7	$eff\left(\hat{\beta}_{diff}, \hat{\beta}(k)\right)$	1.498	1.498	1.498	0.962	0.962	0.962	0.682	0.682	0.682
	$eff\left(\hat{\beta}_{diff}, \hat{\beta}_r(k)\right)$	640.076	1.423	662.653	56.247	0.939	48.481	3.455	0.673	2.889
	$eff\left(\hat{\beta}(k), \hat{\beta}_r(k)\right)$	427.303	0.950	442.375	58.489	0.976	50.413	5.068	0.987	4.239
0.9	$eff\left(\hat{\beta}_{diff}, \hat{\beta}(k)\right)$	0.318	0.318	0.318	0.328	0.328	0.328	0.340	0.340	0.340
	$eff\left(\hat{\beta}_{diff}, \hat{\beta}_r(k)\right)$	0.514	0.318	0.542	8.782	0.327	7.398	6.882	0.340	5.586
	$eff\left(\hat{\beta}(k), \hat{\beta}_r(k)\right)$	1.616	0.999	1.705	26.750	0.997	22.536	20.240	1.000	16.427
0.995	$eff\left(\hat{\beta}_{diff}, \hat{\beta}(k)\right)$	0.034	0.034	0.034	0.042	0.042	0.042	0.247	0.247	0.247
	$eff\left(\hat{\beta}_{diff}, \hat{\beta}_r(k)\right)$	1.706	0.034	1.674	2.775	0.040	2.450	0.418	0.248	0.399
	$eff\left(\hat{\beta}(k), \hat{\beta}_r(k)\right)$	49.963	1.000	49.034	65.811	0.939	58.114	1.689	1.001	1.615

In all of the settings for $n = 20$, $p = 4$ both $\hat{\beta}(k)$ and $\hat{\beta}_r(k)$ perform better than the $\hat{\beta}_{diff}$ estimator which is not surprising given the existence of multicollinearity. When $n = 60$, $p = 10$, $\hat{\beta}_{diff}$ is better than both $\hat{\beta}(k)$ and $\hat{\beta}_r(k)$ for the smallest σ and the smallest correlation. However as σ and correlation increases, $\hat{\beta}_{diff}$ loses its optimality. In general, the efficiency of $\hat{\beta}_{diff}$ clearly decreases as the correlation increases.

From Table 1 where $n = 20$ and $p = 4$, we can see that for small correlation and large error variance $\hat{\beta}_r(k)$ seems to be either slightly more efficient than $\hat{\beta}(k)$ when $c = 1$ that is the linear restriction is true or as efficient as $\hat{\beta}(k)$. Specifically, for $\sigma = 0.01$ as correlation increases, $\hat{\beta}_r(k)$ seems to be doing better. However, when the linear restriction is not satisfied, then $\hat{\beta}(k)$ beats $\hat{\beta}_r(k)$. The results for the $n = 60$ and $p = 10$ given in Table 2 indicate that $\hat{\beta}_r(k)$ performs even worse than former case when the linear restriction does not meet. Once again $\hat{\beta}(k)$ and $\hat{\beta}_r(k)$ perform at par for $c = 1$ and they are better than $\hat{\beta}_{diff}$ especially for larger correlations.

The optimal k value is ranged between 0 to 0.05 for all simulations when $n = 20$, $p = 4$, $\sigma = 0.1$ and $\gamma = 0.9$. Figure 2(a) and 2(b) illustrate the trace of the covariance matrices (i.e. sum of variances) and sMSEs for $\hat{\beta}_{diff}$, $\hat{\beta}(k)$ and $\hat{\beta}_r(k)$ for $0 < k < 0.05$. It can be seen from Figure 2 that $\hat{\beta}(k)$ has smaller sum of variances and smaller sMSE values than $\hat{\beta}_r(k)$. sMSE (which is same as sum of variances) value for $\hat{\beta}_{diff}$ is equal to 3.07 and it is higher than ones both obtained from $\hat{\beta}(k)$ and $\hat{\beta}_r(k)$. These results are consistent with the Table 1. Furthermore, if we apply Theorem 3.1 (Theorem 4.1) for $k = 0.03$, we observe also that V_1 is positive definite and $\hat{\beta}(k)$ is MSE superior to $\hat{\beta}_{diff}$ since inequalities in (17). (30) hold for $k = 0.03$. Similarly, Theorem 3.3 (Theorem 4.2) indicates $\hat{\beta}_r(k)$ is MSE superior to $\hat{\beta}_{diff}$. Theorem 3.5 and 4.3 are not applicable since the difference between the covariance matrices of $\hat{\beta}(k)$ and $\hat{\beta}_r(k)$ is not p.d. when $k = 0.03$.

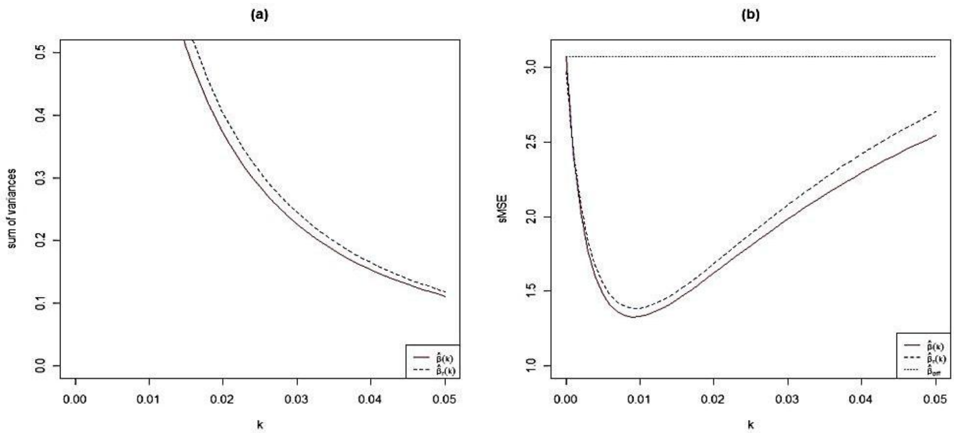

Fig. 2: Trace of (a) Covariance and (b) MSE Matrices of Coefficient Estimates for $n = 20$, $p = 4$, $\sigma = 0.1$, $\gamma = 0.9$

To evaluate the performances of the classical and Newey-West estimators, we compared these estimates with the actual values of variances again for $n = 20$, $p = 4$ and $\gamma = 0.9$. More specifically, we calculated the standard deviations of the coefficient estimates using the real $\Omega = \sigma^2 I$ (where $\sigma = 0.1$) and standard errors obtained by the classical and Newey-West estimators. The sum of the absolute differences between true and estimated standard deviations for $p = 4$ coefficients is used a measure of accuracy. Figure 3 demonstrates boxplots obtained for 5000 simulations indicating that standard errors calculated using Newey-West estimator yielded smaller differences, thus it is better than the classical estimator of variance. This might lead the conclusion that inference based on Newey-West estimator of variances can be more appropriate if one uses a difference-based estimator especially for small sample size at which differencing may cause violating the assumption of independent errors with equal variance.

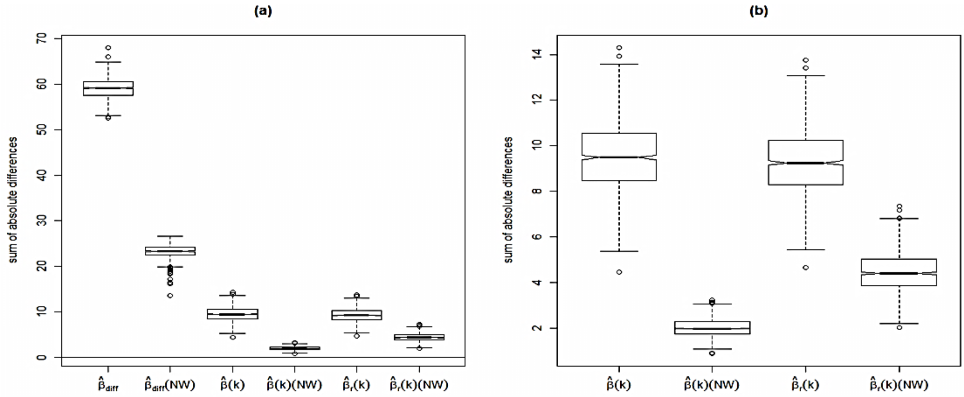

Fig. 3: Boxplots of Sum of the Absolute Differences between the True Standard Deviations and Standard Errors when $n = 20$, $p = 4$, $\sigma = 0.1$, $\gamma = 0.9$ for (a) all Estimators (b) only Ridge and Restricted Ridge where NW in the Parentheses Indicate Newey-West Estimate

6. EMPIRICAL APPLICATION: HEDONIC PRICING OF HOUSING ATTRIBUTES

To motivate the problem of restricted ridge estimation in partially linear regression model, we will fit a partially linear hedonic pricing model of house attributes using the data collected by Ho (1995). This data set includes 92 detached homes in the Ottawa area that were sold in 1987. The variables are defined as follows: y is the sale price (saleprix) of the house (dependent variable), fireplac=1 if the house has a fireplace, garage=1 if the house has a garage, luxbath=1 if the house has a luxury appointment, avginc is the average neighborhood income, disthwy is the distance to highway, lot area is a continuous variable showing the lot size of the property in square feet, usespace is the square footage of housing, nrbed is the number of bedrooms. The model aims to explain saleprix (response variable) as a function of the remaining variables. Initially, we can write in the following pure parametric model which gives the OLS estimates $\hat{\beta}_{OLS}$ of the parameters.

$$\begin{aligned}(saleprix)_i &= \beta_0 + \beta_1(fireplac) + \beta_2(garage)_i + \beta_3(luxbath)_i + \beta_4(avginc)_i \\ &+ \beta_5(disthwy)_i + \beta_6(nrbed)_i + \beta_7(usespace)_i + \beta_8(lotarea)_i + \varepsilon_i\end{aligned} \quad (36)$$

For checking multicollinearity among the variables in model (36), we calculated the condition number corresponding to the Z^* that is an 92×9 observation matrix consisting of the fireplac, garage, luxbath, avginc, disthwy, nrbed, usespace, lotarea in addition to a column of ones. The resulting condition number of Z^* is 250.069 which is large and consequently Z^* is considered as being “ill-conditioned”.

An alternative to the pure parametric model given in (36) is the partially linear model in (2). Standard scatter plots of the response variable versus other explanatory variables indicate that the effect of lotarea on saleprix is likely to be nonlinear, while the effects of others variables are roughly linear. In other words, lotarea is very effective on the sale price of the house, but it has no natural parametric specification; therefore, we include a nonparametric lot size effect, $f(lotarea)$, in our partially linear model given as

$$\begin{aligned}(saleprix)_i &= \beta_1(fireplac)+\beta_2(garage)_i+\beta_3(luxbath)_i+\beta_4(avginc)_i \\ &+\beta_5(disthwy)_i+\beta_6(nrbed)_i+\beta_7(usespace)_i+f(lotarea)_i+\varepsilon_i\end{aligned} \tag{37}$$

which includes both parametric effects and nonparametric effect. It should be noted that we assume here that data have already been ordered according to the lotarea.

Let Z denotes the 92×7 matrix of the following regressor variables: fireplac, garage, luxbath, avginc, disthwy, nrbed, usespace. Similar to the Z^*, Z is also " very ill-conditioned" due to the multicollinearity among the columns. Since classical difference-based estimates will have large variances in the presence of multicollinearity, we will also examine biased estimation techniques discussed in the paper to estimate the parametric components in (36).

Optimal difference sequences for $1\le m\le 10$ can be found in Yatchew (2003). For this study, we used order $m=5$ and the parametric effect, β, in model (37) is estimated by a differencing procedure. We consider the parametric restriction $R\beta=0$, where $R=\left(1-10-1\ -1-10\right)$ as in Akdeniz and A. Duran (2010) for the restricted ridge estimator.

Table 3 illustrates OLS estimates for the parameters of the benchmark parametric model given in (36); classical difference-based, restricted difference-based ridge and difference-based ridge estimates for the parameters of the partially linear model given in (37). It should be noted that in the specification of the partially linear model, the dummy variables enter the linear part of model (1).

Table 3

Coefficient Estimates of Parameters. $\hat{\beta}_r(k)$ and $\hat{\beta}(k)$ are Calculated using $k=0.22$ and $k=0.32$ (in the Parentheses).

	$\hat{\beta}_{OLS}$	$\hat{\beta}_{diff}$	$\hat{\beta}_r(k)$	$\hat{\beta}(k)$
(Intercept)	62.520	--	--	--
fireplac	6.428	6.582	6.163 (5.998)	6.721 (6.682)
garage	13.112	12.140	12.664 (12.703)	12.299 (12.256)
luxbath	66.426	66.943	62.338 (63.572)	62.589 (63.886)
avginc	0.607	0.646	0.669 (0.668)	0.659 (0.655)
disthwy	-11.171	-12.661	-11.734 (-11.797)	-12.142 (-12.298)
nrbed	3.688	3.368	4.565 (4.425)	4.162 (3.929)
usespace	26.691	30.159	27.942 (28.468)	28.175 (28.759)
lotarea	0.730	--	--	--

The plots constructed for two cases: under the independent errors with constant variance assumption (Figure 4a) and more general assumptions on the errors allowing heteroscedasticity and dependence (Figure 4b). For the latter case, Newey-West covariance estimator is utilized for standard error calculations. Figure 4 (a) indicates that the minimum sMSE achieved at k=0.32 (k=0.22) for independent (dependent) error setting. All estimates demonstrate that sale price is positively correlated with all covariates except for distance to highway. In this section, we can note from our theorems that the comparison result depend on the unknown parameters β and σ^2. Since they are unknown, we replaced them by $\hat{\beta}_{diff}$ and $\hat{\sigma}^2_{diff}$, respectively. The trace of the MSEM for a given estimator, called scalar MSE, is calculated and denoted by sMSE in the tables.

It can be seen in Table 4 that the OLS estimates of the parametric model provides larger sMSE estimate than the classical difference-based estimates for the partially linear model parameters meaning that the latter has a better performance. In general, we can see that OLS estimator performs the worst among all estimators in terms of the scalar MSE. One may also see in Table 4 that the restricted difference-based ridge estimator $\hat{\beta}_r(k)$ has smaller sMSE values than the ones from difference-based ridge estimator $\hat{\beta}(k)$. Therefore, by choosing the proper prior information, $\hat{\beta}_r(k)$ may always outperform $\hat{\beta}(k)$. To assess the significance of the covariates, the standard errors calculated using classical and Newey-West estimates of covariance and reported in Table 4.

Under the independent error assumption, all methods found presence of luxurious bath significant while both difference-based ridge methods also declared average neighbor income as an important variable. All difference-based methods identified these two variables along with garage and square footage of usable space under the dependent error assumption while yielding much smaller sMSE values compared to the ones obtained with dependent error assumption. In general, since difference-based estimators introduce a dependence structure among errors due to the differencing (especially for small m), considering heteroscedastic and autocorrelated errors might be beneficial.

Table 4
Standard Errors for Coefficients, Estimated Variance and sMSE Values of Parameters

	$k=0.32$				$k=0.22$			
	$\hat{\beta}_{OLS}$	$\hat{\beta}_{diff}$	$\hat{\beta}_r(k)$	$\hat{\beta}(k)$	$\hat{\beta}_{OLS}$	$\hat{\beta}_{diff}$	$\hat{\beta}_r(k)$	$\hat{\beta}(k)$
(Intercept)	19.081*	--	--	--	16.291*	--	--	--
fireplac	6.466	7.010	6.594	6.534	5.936	7.354	6.921	7.187
garage	5.371	5.764	5.561	5.528	4.246*	4.098*	3.977*	4.019*
luxbath	11.231*	11.613*	10.818*	10.085*	7.213*	7.161*	6.867*	6.831*
avginc	0.239	0.251	0.248*	0.248*	0.184*	0.179*	0.175*	0.178*
disthwy	5.711	6.008	5.784	5.769	5.769	6.738	6.561	6.659
nrbed	5.226	5.499	5.196	5.117	4.411	5.019	4.797	4.919
usespace	11.765	12.821	11.708	10.706	12.110	10.948*	10.294*	10.285*
lotarea	2.336	--	--	--	1.586	--	--	--
sMSE	764.720	448.002	412.649	419.277	572.607	312.630	294.574	300.587
tr(Var(b))	764.720	448.002	389.039	395.433	572.607	312.630	282.918	288.812

* indicates significance at 1% level.

To further explore the performances of restricted difference-based ridge and ridge estimators for different values of biasing parameter, k, and make a connection between empirical and theoretical results derived earlier, we constructed the Figure 4. From Figure 4(a), one can see that the $\hat{\beta}_r(k)$ outperforms the $\hat{\beta}(k)$ for all values of *k*. However since $Cov(\hat{\beta}(k)) - Cov(\hat{\beta}_r(k))$ is not a p.d. matrix (i.e. $\lambda_{\max}[S^{-1}(B'SB)] \geq 1$) for $k \in [0,1]$, we cannot apply Theorem 3.5. On the other hand since both $Cov(\hat{\beta}_{diff}) - Cov(\hat{\beta}(k))$ and $Cov(\hat{\beta}_{diff}) - Cov(\hat{\beta}_r(k))$ are p.d. matrices for all $k > 0$, Theorem 3.1 and Theorem 3.3 are applicable. These theorems indicates that the conditions (17) and (22) are satisfied for difference-based ridge with 0<k<0.74 and restricted difference-based ridge with $0 < k < 0.79$. These results are also supported by sMSE plots in Figure 4(a). Similar arguments are hold for Figure 4(b).

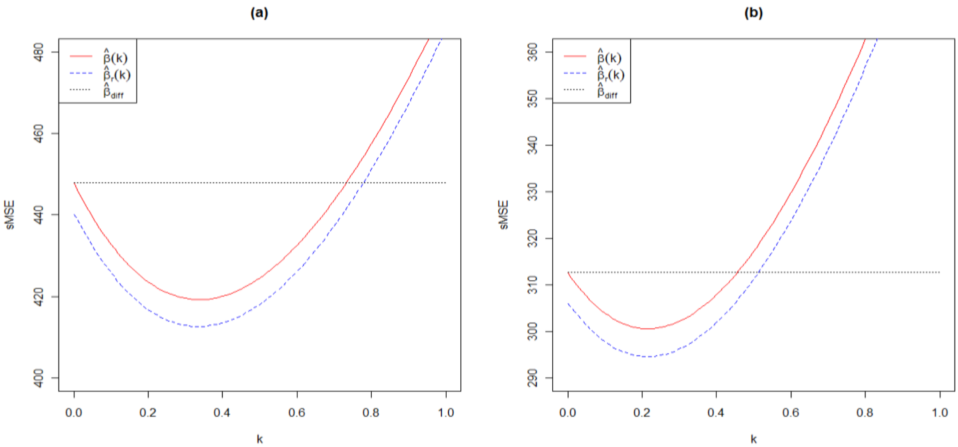

Fig. 4: Estimated MSE Values of $\hat{\beta}_r(k)$ and $\hat{\beta}(k)$ (a) under the Assumptions of Independent Errors with Constant Variance (b) Dependent Errors with Non-Constant Variance for Different Values of *k*.

7. CONCLUSIONS

In this paper, we focus on biased estimators for estimating the parametric component of the partially linear regression model with multicollinearity and correlated and uncorrelated random errors. We consider difference-based ridge, $\hat{\beta}(k)$ and restricted ridge, $\hat{\beta}_r(k)$, estimators for the partially linear model. The bias and mean square error matrix $(MSEM)$ expressions of the estimators are given. The theoretical properties of the $\hat{\beta}(k)$ and $\hat{\beta}_r(k)$ are discussed and their performances over the difference-based estimator, $\hat{\beta}_{diff}$, in terms of MSEM criterion are investigated. A real data example and a simulation study have been provided to evaluate the performance of these estimators based on the MSEM criterion. It is evident from the real data example and simulation results that the difference-based ridge estimators work well for the partially linear regression model under multicollinearity and our estimators are meaningful in practice. It is also shown that Newey-West estimator of variances can be more appropriate if one uses a difference-based estimator since differencing may cause violating the assumption of independent errors with equal variance especially when the sample size is small.

ACKNOWLEDGEMENTS

Gülin TABAKAN was supported by The Scientific and Research Council of Turkey (TUBITAK) under grant TUBITAK-BIDEB-2219 during this study at The Ohio State University in the USA. The authors are indebted to referees for their valuable comments that the contents of the paper.

REFERENCES

1. Akdeniz, F. and Duran, E.A. (2010). Liu-type estimator in semiparametric regression models. *J. Stat. Comput. Sim.,* 80(8), 853-871.
2. Duran, E.A. and Akdeniz, F. (2011). New difference-based estimator of parameters in semiparametric regression models. *J. Stat. Comput. Sim*., 1-15. DOI: 10.1080/ 00949655.2011.638633.
3. Duran, E.A., Hardle, W.K. and Osipenko, M. (2012). Difference based ridge and Liu type estimators in semiparametric regression models. *J. Mult. Anal*., 105, 164-175.
4. Engle, R.F., Granger, C.W.J., Rice, J. and Weiss, A. (1986). Semiparametric estimates of the relation between weather and electricity sales. *J. Amer. Statist. Assoc.*, 81, 310-320.
5. Eubank, R.L., Kambour, E.L., Kim, J.T., Klipple, K., Reese, C.S. and Schimek, M.G. (1988). Estimation in partially linear models. *Comput. Statist. Data Anal.,* 29, 27-34.
6. Eubank, R.L. (1999). *Nonparametric Regression and Spline Smoothing.* Marcel Dekker, NewYork.
7. Farebrother, R.W. (1976). Further results on the mean square error of ridge regression. *J. Roy. Statist. Soc. B.,* 38, 248-250.
8. Graybill, F. (1993). *Matrices with Applications in Statistics.* Duxbury Classic.
9. Haibing, Z. and Jinhong Y. (2011). Difference based estimation for partially linear regression models with measurement errors. *J. Mult. Anal.,*102(10), 1321-1338.
10. Härdle, W., Müller, M., Sperlich, S. and Werwatz, A. (2004). *Nonparametric and Semiparametric Models*. Springer, Berlin.
11. Haville, D. (1997). *Matrix Algebra from a Statistician's Perspective*. Springer Verlag, New York.
12. Hoerl, A.E. and Kennard, R.W. (1970). Ridge regression: biased estimation for orthogonal problems. *Technometrics,* 12, 55-67.
13. Hoerl, A.E., Kennard R.W. and Baldwin, K.F. (1975). Ridge regression: Some simulation. *Commun. Stat.,* 4, 105-123.
14. Ho, M. (1995). *Essays on the Housing Market*. Unpublished Ph.D. Dissertation, University of Toronto.
15. Liu, K. (1993). A new class of biased estimate in linear regression. *Commun. Stat. Theory Methods,* 22, 393-402.
16. Liu, K. (2003). Using Liu type estimator to combat multicollinearity. *Commun. Stat. Theory Methods,* 32(5), 1009-1020.
17. Luo, J. (2012) Asymptotic Efficiency of Ridge Estimator in Linear and Semiparametric Linear Models. *Stat. Prob. Lett.*, 82(1) 58-62.

18. McDonald, G.C. and Galarneau, D.I. (1975). A Monte Carlo evaluation of some ridge-type estimators. *J. Amer. Statist. Assoc.*, 70, 407-416.
19. Newey, W. and West, K. (1987). A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica*, 55(3), 703-708.
20. Roozbeh, M. and Arashi M. (2014). Feasible Ridge Estimator in Seemingly Unrelated Semiparametric Models. *Commun. Stat. Sim. and Comput.*, 43, 2593-2613.
21. Ruppert, D., Wand, M.P. and Carroll, R.C. (2003). *Semiparametric Regression*. Cambridge University Press, Cambridge, UK.
22. Schott, J. (2005). *Matrix Analysis for Statistics*. second ed., Wiley Inc., New Jersey.
23. Speckman, P. (1988). Kernel smoothing in partial linear models. *J. Roy. Statist. Soc. Ser. B.*, 50(3), 413-436.
24. Tabakan, G. (2013). Performance of the difference-based estimators in partially linear models. *Statistics,* 47(2), 329-347.
25. Tabakan, G. and Akdeniz, F. (2010). Difference-based ridge estimator of parameters in partial linear model. *Statist,. Pap*., 51, 357-368.
26. Trenkler, G. and Toutenburg, H. (1990). Mean square matrix comparisons between biased estimators-an overview of recent results. *Statist. Pap.*, 31, 165-179.
27. Wang, L., Brown, L.D. and Cai, T.T. (2011). A difference based approach to the semiparametric partial linear model. *Elect. J. Stat*., 5, 619-641.
28. Wu, J. (2014). The relative efficiency of Liu-type estimator in a partially linear model. *App. Math. and Comput*., 243, 349-357.
29. Yatchew, A. (1997). An elementary estimator of the partial linear model. *Econ. Lett.,* 57, 135-143.
30. Yatchew, A. (1999). Differencing methods in nonparametric regression: Simple techniques for the applied econometrician. *Manuscript, University of Toronto. http://www.economics.utoronto.ca/yatchew*/.
31. Yatchew, A. (2003). *Semiparametric Regression for the Applied Econometrican*. Cambridge University Press, Cambridge, UK.