International Journal of Soft Computing 11 (5): 295-298, 2016
ISSN: 1816-9503

Ridge Estimation in Semiparametric Partial Linear Regression Models Using Differencing Approach

Mahdi Roozbeh
Department of Mathematics, Statistics and Computer Sciences,
School of Sciences, Semnan University, P.O. Box, 35195-363 Semnan, Iran

Abstract: A common problem in applied sciences is multicollinearity between variables. Multicollinearity is frequently encountered problems in practice that produce undesirable effects on classical Ordinary Least-Squares (OLS) regression estimator. The ridge estimation is an important tool to reduce the effects of multicollinearity. Also, it is suspected that some additional linear constraints may hold on to the whole parameter space. This restriction is based on either additional information or prior knowledge. The proposed estimators based on restricted estimator performs fairly well than the other estimators based on ordinary least-squares estimator. In this study, by some theorems, necessary and sufficient conditions for the superiority of the new estimator over the restricted least-squares estimator for selecting the ridge parameter k are derived. For illustrating the usefulness of the proposed result, the performance of this estimator is compared to the classic estimator via a simulation study in restricted partial linear regression models.

Key words: Differencing approach, linear restrictions, multicollinearity, partial linear regression model, ridge estimation

INTRODUCTION

Consider the partial linear model given by:

$$y = X\beta + f + \varepsilon \qquad (1)$$

where, $y = (y_1, ..., y_n)'$, $X = (x_1, ..., x_n)'$, $f = (f(u_1), ...,f(u_n))'$ $\varepsilon = (\varepsilon_1, ..., \varepsilon_n)'$. We assume that in general, $f(\cdot)$ is an unknown function, the u's ave bounded support, say the unit interval and have been reordered so that $u_1 \leq ... \leq u_n$. In our study, ϵ is a n-vector of disturbances with the characteristics $E(\varepsilon) = 0$ and the dispersion matrix $D(\varepsilon) = \sigma^2 I$ where, σ^2 is an unknown parameter. Partial linear model is more flexible than standard linear model since it has a parametric and a nonparametric component. It can be a suitable choice when one suspects that the response y linearly depends on x but that it is nonlinearly related to u. This model is first considered by Engle *et al*. (1986) to study the effect of weather on electricity demand in which they assumed that the mean relationship between temperature and electricity usage was unknown while other related factors such as income and price were parameterized linearly.

For the main purposes of this study we will employ the ridge regression concept that was proposed in the 1970's to combat the multicollinearity in the partial linear model. The existence of multicollinearity may lead to wide confidence intervals for individual parameters or linear combination of the parameters and may produce estimates with wrong signs, etc. Most of the literature judges the performance of ridge regression estimators on the basis of the concentration of estimates around the true value of the parameter (Grob, 2003; Kibria and Saleh, 2004; Tabakan and Akdeniz, 2010; Akdeniz *et al*., 2015; Roozbeh *et al*., 2010; Arashi *et al*., 2015; Arashi and Valizadeh, 2015; Roozbeh, 2015).

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 semiparametric model (Wang *et al*., 2011).

In what follows, researchers present an explanation made by Akdeniz *et al*. (2015), demonstrating how the approximation works. Let $d = (d_0, ..., d_m)$ be a (m+1) vector where m is the order of differencing and $d_0, ..., d_m$ are differencing weights satisfying the conditions:

$$\sum_{j=0}^{m} d_j = 0 \text{ and } \sum_{j=0}^{m} d_j^2 = 1 \qquad (2)$$

A differencing matrix denoted by D is a (n-m)×n known matrix with the elements satisfying Eq. 2 (Yatchew, 2003). Imposing the differencing matrix to the model Eq. 1, permits direct estimation of the parametric effect. In particular, it takes:

$$Dy = DX\beta + Df + D\varepsilon \qquad (3)$$

Since, the data have been reordered so that the u's are close, the application of the differencing matrix D in model Eq. 3 removes the nonparametric effect in large samples. Thus, the underlying model is rewritten as:

$$\tilde{y} \cong \tilde{X}\beta + \tilde{\varepsilon} \tag{4}$$

where, $\tilde{y} = Dy, \tilde{X} = DX$ and $\tilde{\varepsilon} = D\varepsilon$.

MATERIALS AND METHODS

Restricted difference-based ridge estimator: It is well known that adopting the linear model (1.4), the unbiased estimator of β is the following difference-based estimator given by:

$$\hat{\beta}_D = C_D^{-1}\tilde{X}'\tilde{y}, C_D = \tilde{X}'\tilde{X} \tag{5}$$

It is observed from Eq. 5 that the properties of the difference-based estimator of β depends heavily on the characteristics of the information matrix C_D. If the C_D matrix is ill-conditioned, then some of the regression coefficients may be statistically insignificant with wrong sign and meaningful statistical inference become difficult for the researcher. As a remedy following by Hoerl and Kennard (1970) we suggest to use the following estimator, namely, difference-based ridge estimator:

$$\hat{\beta}_D(k) = T(k)\hat{\beta}_D, T(k) = \left(kC_D^{-1} + I_p\right)^{-1} \tag{6}$$

where, $k \geq 0$ is the shrinking parameter. Now we consider the linear non-stochastic constraint:

$$R\beta = r \tag{7}$$

for a given $q \times p$ matrix R with rank $q<p$ and a given $q \times 1$ vector r. The full row rank assumption is chosen for convenience and can be justified by the fact that every consistent linear equation can be transformed into an equivalent equation with a coefficient matrix of full row rank. Subject to the linear restriction Eq. 7, the restricted difference-based estimator is given by:

$$\hat{\beta}_{RD} = \hat{\beta}_D - C_D^{-1}R'\left(RC_D^{-1}R'\right)^{-1}\left(R\hat{\beta}_D - r\right)^{-1} \tag{8}$$

So, the restricted difference-based ridge estimator can be written as:

$$\hat{\beta}_{RD}(k) = \hat{\beta}_D(k) - C_D(k)^{-1}R'\left(RC_D(k)^{-1}R'\right)^{-1}\left(R\hat{\beta}_D(k) - r\right) \tag{9}$$

$$C_D(k) = C_D + kI_p \text{ and } \hat{\beta}_D(k) = C_D(k)^{-1}\tilde{X}'\tilde{y}$$

RESULTS AND DISCUSSION

Evaluation of risk functions: In this researchers calculate the risk function for the proposed estimator given in previous section. Before deriving the risk function of β_{RD} = (k) we propose a new formula for β_{RD} = (k) which simplifies the calculation of risk function as follows:

$$\hat{\beta}_{RD}(k) = N_D(k)\tilde{X}'\tilde{y} - N_D(k)C_D(k)\beta_0 + \beta_0 \tag{10}$$

$$\beta_0 = R'\left(RR'\right)^{-1}r$$

and:

$$N_D(k) = C_D(k)^{-1} - C_D(k)^{-1}R'\left(RC_D(k)^{-1}R'\right)^{-1}RC_D(k)^{-1}$$

Thus, the risk functions of the proposed estimators are:

$$MSE\left(\hat{\beta}_{RD}(k)\right) = s^2N_D(k) - ks^2N_D(k)^2 + k^2N_D(k)\beta\beta'N_D(k) \tag{11}$$

and:

$$MSE\left(\hat{\beta}_{RD}\right) = \sigma^2 N_D(0) \tag{12}$$

Theorem 1: There exists at least a $k^*>0$ such that $\beta_{RD}(k)$ dominates $\beta_{RD}(k)$ in the sense of MSE.

Proof: It is enough to show that there exists $k^*>0$ such that:

$$MSE\left(\hat{\beta}_{RD}\right) - MSE\left(\hat{\beta}_{RD}(k)\right) > 0$$

The partial derivative of Eq. 11 with respect to k is:

$$\left.\frac{\partial MSE\left(\hat{\beta}_{RD}(k)\right)}{\partial k}\right|_{k=0} = -2\sigma^2 N_D(0)^2 < 0$$

Since, $\sigma^2>0$ and $N_D(0)^2$ is non zero we conclude that the $MSE(\beta_{RD}(k))$ has decreasing trend at k = 0. This implies that there exists at least $k^*>0$, satisfying $MSE(\beta_{RD})$ = $MSE(\beta_{RD}(k))>0$. Therefore researchers can select a suitable positive number k to let the estimator $\beta_{RD}(k)$ performs better than β_{RD} in the sense of MSE.

MSE-superiority of the difference-based ridge estimator over the differencing estimator: In this reserchers provide necessary and sufficient conditions for which the

Table 1: Evaluation of estimators at different values k for simulated model

k coefficients	0	1	2	3	4	5
β_1	-1.499217	-1.499220	-1.499223	-1.499226	-1.499229	-1.499233
β_2	-2.008458	-2.008424	-2.008390	-2.008356	-2.008321	-2.008287
β_3	-3.005952	-3.005928	-3.005904	-3.005880	-3.005856	-3.005832
β_4	4.996084	4.996100	4.996116	4.996132	4.996147	4.996163
β_5	-3.999217	-3.999220	-3.999223	-3.999226	-3.999229	-3.999233
tr (MSE)	0.0001915	0.00019155	0.00019154	0.00019153	0.00019153	0.00019153
Δ	0	1.614e-08	2.789e-08	3.523e-08	3.818e-08	3.673e-08
k coefficients	6	7	8	9	10	11
β_1	-1.499236	-1.499239	-1.499242	-1.499245	-1.499249	-1.499252
β_2	-2.008253	-2.008219	-2.008185	-2.008150	-2.008116	-2.008082
β_3	-3.005808	-3.005784	-3.005759	-3.005735	-3.005711	-3.005687
β_4	4.996179	4.996195	4.996211	4.996227	4.996243	4.996258
β_5	-3.999236	-3.999239	-3.999242	-3.999245	-3.999249	-3.999252
tr (MSE)	0.00019153	0.00019154	0.00019156	0.00019158	0.00019160	0.00019163
Δ	3.088e-08	2.063e-08	5.993e-09	-1.304e-08	-3.646e-08	-5.428e-08

estimator $\beta_{RD}(k)$ performs better than β_{RD} in the sense of $MSE(\beta_{RD}(k)) \leq MSE(\beta_{RD})$. From Eq. 11 and 12, the difference matrix $DM = MSE\beta_{RD} - MSE(\beta_{RD}(k))$ is given by:

$$DM = s^2 N_D(0) - s^2 N_D(k) + ks^2 N_D(k)^2 - k^2 N_D(k)\beta\beta' N_D(k)$$

Theorem 2: Let the estimator $\beta_{RD}(k)$ given by under the linear regression model with true restrictions $R\beta = r$. If $k>0$, then the MSE difference DM is nonnegative definite if and only if:

$$k\left(s^2\beta\beta' - \left(PC_D P\right)^+\right) \leq P \tag{13}$$

where, $P = I_p - R'(RR')^{-1}R$. Note that by a "+" superscript we denote the unique Moore-Penrose inverse (Roozbeh and Arashi, 2013).

Choice of the biasing parameter: In the process of determining k, on one side we must control the condition number of $C_D(k)$ to a lesser level if we want to avoid the instability of estimated coefficients brought by the morbidity of C_D. Hence, we must do our best to let the ridge parameter k be big. As stated in Theorem 2 we do not need to find out the best k in the practice. That is to say we just need to find a k which can make $\beta_{RD}(k)$ be superior to the β_{RD} in the sense of MSE.

Although, the criterion mentioned above is simple, our problem to select k is not yet completely solved. Therefore, we give a range to select k in Theorem 2.

Theorem 3: Let us be given the estimator $\beta_{RD}(k)$ under the linear regression model with true restrictions $R\beta = r$ and $\beta \neq \beta$. The MSE difference matrix DM is nonnegative definite if:

$$0 < k \leq \frac{2s^2}{\beta' P\beta} \tag{14}$$

Roozbeh and Arashi (2013).

Numerical study: In this researchers proceed with the comparison of the proposed estimators by some numerical computations. In the scalar comparison, the trace of MSE. That is we will compare the trace of $MSE\beta_{RD}(k)$ and MSE $\beta_{RD}(k)$ and define scalar Δ as:

$$\Delta = tr(DM) = s^2 tr\left(N_D(0)\right) - s^2 tr\left(N_D(k)\right) + ks^2 tr\left(N_D(k)^2\right) - k^2\beta' N_D(k)^2\beta$$

In this study we simulate the response for n = 5000 from the following models:

$$y = X\beta + f(u) + \varepsilon, f(u) = \sqrt{u(1-u)}\sin\left(\frac{2.1\pi}{u+0.05}\right), u \in (0,1)$$

where $\beta = (-1.5, -2, -3, 5, -4)'$, $\varepsilon \approx N(0, \sigma^2 I_n)$ which $\sigma^2 = 4$ that is called the Doppler function and $x_i \approx N(\mu_x \Sigma_x)$ with:

$$\mu_x = \begin{pmatrix} 2.5 \\ 2.0 \\ 3.0 \\ 1.0 \\ -1.0 \end{pmatrix}, \Sigma_x = \begin{pmatrix} 1.9 & 1.8 & 1.8 & 1.0 & 1.0 \\ 1.8 & 1.8 & 1.8 & 1.0 & 1.0 \\ 1.8 & 1.8 & 4.25 & 1.0 & 1.0 \\ 1.0 & 1.0 & 1.0 & 2.49 & 1.0 \\ 1.0 & 1.0 & 1.0 & 1.0 & 2.25 \end{pmatrix},$$

$$R = \begin{pmatrix} 1 & -2 & 1 & 4 & 5 \\ -1 & 2 & -1 & -3 & 0 \\ 1 & 2 & -1 & -2 & 3 \\ 2 & -1 & 3 & -2 & 0 \end{pmatrix}$$

Table 1 gives several results including $\beta_{RD}(k)$, $tr(MSE(\beta_{RD}(k)))$ and Δ for different values of k. From this table, the Δ increases EMBED Eq. 3 decreases at first and then decreases (increases). Furthermore, the maximum of Δ minimum of the EMBED Eq. 3 is obtained when k equals to median range of Eq. 14 which is approximately equal to 4.17 in simulated model.

CONCLUSION

Grob (2003) considered restricted ridge estimator in regression model while Tabakan and Akdeniz (2010) studied unrestricted estimator in partial regression model. In this study, combining these approaches we proposed two new estimators in a partial linear model when some additional linear constraints held on the whole parameter space β. In the presence of multicollinearity in a partial linear model we introduced the restricted difference-based ridge regression estimator $\beta_{RD}(k)$ versus the non-ridge version under β_{RD} dependency among column vectors of the design matrix. The MSE functions of proposed estimators are driven. In this regard we continued the comparison study by some simulation strategy and graphical results. The experiment was taken for different values of ridge parameter k and nonparametric functions.

ACKNOWLEDGEMENTS

This research was supported by Research Council of Semnan University.

RECOMMENDATIONS

As some future works, one can consider:

- Restricted ridge estimator in high-dimensional partial regression model
- Ridge estimator in high-dimensional partial regression model under stochastic constraints

REFERENCES

Akdeniz, F., E.A. Duran, M. Roozbeh and M. Arashi, 2015. Efficiency of the generalized difference-based Liu estimators in semiparametric regression models with correlated errors. J. Stat. Comput. Simul., 85: 147-165.

Arashi, M. and T. Valizadeh, 2015. Performance of Kibria's methods in partial linear ridge regression model. Stat. Pap., 56: 231-246.

Engle, R.F., C.W. Granger, J. Rice and A. Weiss, 1986. Semiparametric estimates of the relation between weather and electricity sales. J. Am. Stat. Assoc., 81: 310-320.

Grob, J., 2003. Restricted ridge estimation. Stat. Probab. Lett., 65: 57-64.

Hoerl, A.E. and R.W. Kennard, 1970. Ridge regression: Biased estimation for nonorthogonal problems. Technometrics, 12: 55-67.

Kibria, B.G. and A.M.E. Saleh, 2004. Preliminary test ridge regression estimators with Student's errors and conflicting test-statistics. Metrika, 59: 105-124.

Roozbeh, M. and M. Arashi, 2013. Feasible ridge estimator in partially linear models. J. Multivariate Anal., 116: 35-44.

Roozbeh, M., 2015. Shrinkage ridge estimators in semiparametric regression models. J. Multivariate Anal., 136: 56-74.

Roozbeh, M., Arashi, M. and H.A. Niroumand, 2010. Semiparametric ridge regression approach in partially linear models. Commun. Stat. Simul. Comput., 39: 449-460.

Tabakan, G. and F. Akdeniz, 2010. Difference-based ridge estimator of parameters in partial linear model. Stat. Pap., 51: 357-368.

Wang, L., L.D. Brown and T.T. Cai, 2011. A difference based approach to the semiparametric partial linear model. Electron. J. Stat., 5: 619-641.

Yatchew, A., 2003. Semiparametric Regression for the Applied Econometrician. Cambridge University Press, Cambridge, England,.