Shrinkage estimators for semiparametric regression model

Hadi Salman Mohammed\textsuperscript{1} and Zakariya Yahya Algamal\textsuperscript{2}

\textsuperscript{1}Department of Statistics and Informatics, University of Mosul, Mosul, Iraq  
\textsuperscript{2}Department of Statistics and Informatics, University of Mosul, Mosul, Iraq  
\textsuperscript{1}Email: hadisalman366@gmail.com  
\textsuperscript{2}zakariya.algamal@uomosul.edu.iq

Abstract. Semiparametric regression models are extensions of linear regression models to include a nonparametric function of some explanatory variables. In semiparametric regression model researchers often encounter the problem of multicollinearity. In the context of ridge estimator, the estimation of shrinkage parameter plays an important role in analyzing data. In this paper, numerous selection methods of the shrinkage parameter of ridge estimator are explored and investigated. Our Monte Carlo simulation results suggest that some estimators can bring significant improvement relative to others, in terms of mean squared error.

1. Introduction

Semiparametric regression models have received considerable attention in statics and econometrics, because of their flexibility in modeling events \cite{1, 2}. Consider a semiparametric regression model given by

\[ y_i = x_i \beta + f(t_i) + \varepsilon_i, \quad i = 1, 2, \ldots, n, \]

where \( x = (x_{i1}, x_{i2}, \ldots, x_{ip})' \) is a vector of explanatory variable, \( \beta = (\beta_1, \beta_2, \ldots, \beta_p)' \) is an unknown p-dimensional parameter vector, the \( t_i \)'s are known and non-random in some bounded domain \( D \subset \mathbb{R}, f(t) \) is an unknown smooth function and \( \varepsilon_i \)'s are independent and identically distributed random error with mean 0, variance \( \sigma^2 \), which are independent of \((x_i, t_i)\) \cite{3}.

Most of the approaches for the semiparametric regression model are based on different nonparametric regression procedures. The have been several approaches to estimating \( \beta \) and \( f(\cdot) \). An alternative approach to nonparametric procedure is differencing methodology.

This incoming used differences to remove the trend in the data that arises from the function \( f(\cdot) \) and does not require an estimator of the function \( f(\cdot) \) and often called difference-based procedure. Provided that \( f(\cdot) \) is differentiable and the \( t \) ordinates are closely spaced, it is possible to remove the effect of the function \( f(\cdot) \) by differencing the data appropriately. In model \cite{1, 5} concentrated on estimation of the linear component and used 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 used higher order differences for optimal efficiency in estimating the linear party by using special class of difference sequences.

Now consider a semiparametric regression model in the presence of multicollinearity. The existence of multicollinearity may lead to wide confidence intervals for the individual parameters or linear combination of the parameters and signs. For our purpose we only employ the ridge regression concept due to Hoerl and Kennard \cite{3}, to combat multicollinearity. There are a lot of work adopting ridge regression methodology to overcome the multicollinearity problem.
2. Differencing Approach

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

$$y = x \beta + f(t) + \varepsilon,$$

where $y = (y_1, y_2, \ldots, y_n)'$, $x = (x_1, x_2, \ldots, x_n)'$ is then $n \times p$ matrix, $f(t) = (f(t_1), \ldots, f(t_n))'$, and $\varepsilon = (\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_n)'$. We assume that in general, $\varepsilon$ is a vector of disturbances distributed with $E(\varepsilon) = 0$ and $E(\varepsilon \varepsilon') = \sigma^2 V$, where $V$ is symmetric, positive definite known matrix and $\sigma^2$ is an unknown parameter.

Yatchew (1997) suggested estimating $\beta$ on the basis of the $m^{th}$ order differencing equation when $V = I_p$ as

$$\sum_{j=0}^{m} d_j y_{j-i} = \left( \sum_{j=0}^{m} d_j x_{j-i} \right) \beta + \sum_{j=0}^{m} d_j f(t_{j-i}) + \sum_{j=0}^{m} d_j \varepsilon_{j-i},$$

where $d_0, d_1, \ldots, d_m$ are differencing weights.

Suppose $t_i$ are equally spaced on the unit interval and $f'(\cdot) \leq L$. By the mean value theorem, for some $t_i^* \in [t_{i-1}, t_i]$ we have

$$f(t_i) - f(t_{i-1}) = f'(t_i^*)(t_i - t_{i-1}) \leq \frac{L}{n}.$$

Note that with $m = p = 1$ from (2.2) we have

$$y_i - y_{i-1} = (x_i - x_{i-1}) \beta + f(t_i) - f(t_{i-1}) + \varepsilon_i - \varepsilon_{i-1}$$

$$= (x_i - x_{i-1}) \beta + O\left(\frac{1}{n}\right) + \varepsilon_i - \varepsilon_{i-1}$$

$$\approx (x_i - x_{i-1}) \beta + \varepsilon_i - \varepsilon_{i-1}.$$

We then estimate the linear regression coefficient $\beta$ by the ordinary least-square estimators based on the differences. Then we obtain the least-squares estimate

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

Now let $d = (d_0, \ldots, d_m)$ be a $(m+1)$-vector, where $m$ is the order of differencing and $d_0, d_1, \ldots, d_m$ are differencing weights minimizing the variance of the estimators i.e.,

$$\min_{d_0, d_1, \ldots, d_m} \left( \sum_{j=0}^{m} d_j^2 \right)$$

Satisfying the conditions
The role of constraints (Eq. (3)) is now evident. The first condition ensures that, as the $t$'s become close, the nonparametric effect is removed and the second one ensures that the variance of the sum of weighted residuals remains equals to $\sigma^2$ in Eq. (2).

Now, we define the $(n-m) \times n$ differencing matrix $D$ whose element satisfy Eq. (3) as

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

This and related matrices are given, for example, in [4, 6-9].

Applying the differencing matrix to model (Eq. (2)) permits direct estimation of the parametric effect. As a result of development in Speck man (1988) it is know that the parameter vector $\beta$ in (Eq. (1)) can be estimated with parametric efficiency. We now show the difference-based estimators that can be used for this purpose. 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 (Eq. (2)) removes the nonparametric effect in large samples. If $f(t)$ is an unknown function, i.e., the inferential object and has a bound first derivative, then $Df(t)$ is close to 0, so that by applying the differencing matrix, we may rewrite (Eq. (1)) as

$$Dy = DX \beta + D \varepsilon,$$

Or

$$y_D = X_D \beta + D \varepsilon,$$

where $y_D = D'y$, $X_D = D'X$, $D_D = D'$. So, $D_D$ is a(n-m)-vector of disturbances distributed with $E(D_D) = 0$ and $E(D_D D_D') = \sigma^2 V_D$ where $V_D = D' D' \neq I_{n-m}$. For arbitrary differencing coefficients satisfying (2.3), Yatchew (1997) defined a simple differencing estimator of the parameter $\beta$ in a semiparametric regression model (SPRM) when $V = I_{n}$ as

$$\hat{\beta} D = (X_D' X_D)^{-1} X_D' y_D.$$  

Thus, differencing allows one to perform inferences on $\beta$ as if there were no nonparametric component $f(.)$ in the model (Eq. (1)), (Yatchew 2003). Once $\beta$ is estimated, a variety of nonparametric techniques could be applied to estimate $f(.)$ as if $\beta$ were known.
To account the parameter $\beta$ in Eq. (4), we introduce the modified estimator of $\sigma^2$, defined as

$$\sigma^2 D = \frac{y^\top V_D^{-\frac{1}{2}} (I - P V_D^{-\frac{1}{2}}) y D}{\text{tr}(D'(I - P) D)}$$

where $\text{tr}(.)$ is the trace function for a squared matrix and $P$ is the projection matrix defined as

$$p = V_D^{-\frac{1}{2}} X_D (X_D V_D^{-1} X_D)^{-1} X_D^\top V_D^{-\frac{1}{2}}.$$

3. Ridge Estimator

To overcome the effect of multicollinearity, ridge estimator is usually utilized. The ridge estimator for the semiparametric regression model (RE) is defined as

$$\hat{\beta}_{RE} = (X_D^\top X_D + k I)^{-1} X_D^\top y_D,$$  \hspace{1cm} (7)

where $k > 0$ is the shrinkage parameter. Several methods were proposed to estimate the value of $k$.

The idea behind these used methods is obtained from the work by Hoerl and Kennard [10], Kibria [11], Kibria, Månsson [12], Dorugade and Kashid [13], Asar, Karaibrahimoğlu [14], and Bhat [15].

1. Hoerl and Kennard [10] (HK1 and HK2), which are, respectively, defined as

$$\text{HK1} = \frac{p \hat{\sigma}^2}{\hat{\alpha}^2 \hat{\alpha}},$$

$$\text{HK2} = \frac{\hat{\sigma}^2}{\hat{\alpha}^2 \max},$$

2. Kibria, Månsson [12] used several methods which were proposed by Kibria, Månsson [16] and Muniz and Kibria [17] (K1 – K12). They are, respectively, defined as

$$\text{K1} = \max \left\{ \frac{1}{m_j} \right\}, \hspace{0.5cm} j = 1, 2, ..., p,$$

$$\text{K2} = \max \left\{ m_j \right\}, \hspace{0.5cm} j = 1, 2, ..., p,$$

$$\text{K3} = \prod_{j=1}^{p} \left\{ \frac{1}{m_j} \right\}^{1/p}, \hspace{0.5cm} j = 1, 2, ..., p,$$

$$\text{K4} = \prod_{j=1}^{p} \left\{ m_j \right\}^{1/p}, \hspace{0.5cm} j = 1, 2, ..., p,$$

$$\text{K5} = \text{median} \left\{ \frac{1}{m_j} \right\}, \hspace{0.5cm} j = 1, 2, ..., p.$$
\[ K_6 = \text{median} \left\{ m_j \right\}, \quad j = 1, 2, \ldots, p, \]  
(15)

\[ K_7 = \max \left\{ \frac{1}{q_j} \right\}, \quad j = 1, 2, \ldots, p, \]  
(16)

\[ K_8 = \max \left\{ q_j \right\}, \quad j = 1, 2, \ldots, p, \]  
(17)

\[ K_9 = \prod_{j=1}^{p} \left( \frac{1}{q_j} \right)^{1/p}, \quad j = 1, 2, \ldots, p, \]  
(18)

\[ K_{10} = \prod_{j=1}^{p} \left( q_j \right)^{1/p}, \quad j = 1, 2, \ldots, p, \]  
(19)

\[ K_{11} = \text{median} \left\{ q_j \right\}, \quad j = 1, 2, \ldots, p, \]  
(20)

where \( m_j = \sqrt{\hat{\sigma}^2 / \hat{\alpha}_j^2} \) and \( q_j = \lambda_{\max} / (n - p) \hat{\sigma}^2 + \lambda_{\max} \hat{\alpha}_j^2 \).

4. Simulation Results

A Monte Carlo simulation scheme to evaluate the performance of the estimating methods for the ridge estimator shrinkage parameter. “The explanatory variables \( x_i = (x_{i1}, x_{i2}, \ldots, x_{in}) \) have been generated from the following formula

\[ x_{ij} = (1 - \rho^2)^{1/2} w_i + \rho w_j, \quad i = 1, 2, \ldots, n, \quad j = 1, 2, \ldots, p, \]  
(21)

where \( \rho \) represents the correlation between the explanatory variables and \( w_i \)'s are independent standard normal pseudo-random numbers. Because the sample size has direct impact on the prediction accuracy, three representative values of the sample size are considered: 30, 50 and 100. Further, because we are interested in the effect of multicollinearity, in which the degrees of correlation considered more important, three values of the pairwise correlation are considered with \( \rho = \{0.90, 0.95, 0.99\} \). Then \( n \) observations for the dependent variable are determined by

\[ y_i = \sum_{j=1}^{6} x_{ij} \beta_j + f(t) + \epsilon, \quad i = 1, \ldots, n, \]  
(22)

where \( \beta = (3, 1, -3, 2, -5, 4)' \) and 

\[ f(t) = \frac{1}{3} \left[ \Phi(t; -3, 0.81) + \Phi(t; 0, 0.36) + \Phi(t; 3, 0.81) \right], \]

Which is mixture of normal densities for \( t \in [-5, 5] \) and \( \Phi(x; \mu, \sigma^2) \) is a normal density function
with mean $\mu$ and variance $\sigma^2$. The main reason of selecting such structure for nonlinear part is to check the efficiency of nonparametric estimation for wavy function. Moreover, $\varepsilon \sim N (0, \sigma^2 V)$ for which the element of $V$ are $v_{ij} = \frac{1}{n} |i-j|$. Four values of $\sigma^2 = 0.01$ is investigated. We use a fourth-order differencing coefficients,

$$d_0 = 0.8873, \quad d_1 = 0.3099, \quad d_2 = -0.2426, \quad d_3 = -0.1910, \quad \text{and} \quad d_4 = -0.1409 \quad \text{in which} \quad m = 4.$$ 

The estimated MSE for all the different selection methods of $k$ and the combination of $n$ and $\rho$, are respectively summarized in Tables 1 – 3. Several observations can be obtained as follows:

1- In terms of $\rho$ values, there is increasing in the MSE values when the correlation degree increases regardless the value of $n$.
2- With respect to the value of $n$, The MSE values decrease when $n$ increases, regardless the value of $\rho$, in most cases.
3- All the selection methods of $K$ are superior to the SPRM estimator in terms of MSE.
4- Clearly, in terms of MSE, notice the superiority of the two methods, $K^8$ and $K^2$, respectively, when $\rho = 0.90$, and for all study sample sizes (small, medium and large), and that they improved the performance of the ridge estimator compared to other methods because they gave the lowest values for MSE.
5- As for the correlation coefficient $\rho = 0.95$, it was noticed the superiority of the $K^1$ method for all sizes, followed by the $K^8$ method in small and medium samples size, but in the case of large samples size, the $HK^2$ method came second.
6- The results showed that, when a correlation coefficient $\rho = 0.99$, the $K^3$ method was the best, and in the next rank was the $K^{11}$ method at different sizes of the study samples.

| Method | SPRM | HK1 | HK2 | K1 | K2 | K3 | K4 | K5 | K6 | K7 | K8 | K9 | K10 | K11 |
|--------|------|-----|-----|----|----|----|----|----|----|----|----|----|-----|-----|
| $\rho = 0.90$ | 1.63 | 1.498 | 1.549 | 1.564 | 1.276 | 1.608 | 1.45 | 1.585 | 1.492 | 1.621 | 1.268 | 1.627 | 1.301 | 1.625 |
| $\rho = 0.95$ | 2.991 | 3.058 | 3.151 | 1.579 | 2.476 | 2.808 | 2.65 | 2.785 | 2.692 | 2.821 | 2.468 | 2.827 | 2.501 | 2.825 |
| $\rho = 0.99$ | 3.64 | 3.231 | 3.298 | 3.101 | 3.155 | 2.868 | 2.894 | 2.955 | 3.056 | 3.235 | 3.026 | 2.918 | 2.872 |
Table 2: MSE values when $n = 50$

| Method | SPRM | HK1 | HK2 | K1  | K2  | K3  | K4  | K5  | K6  | K7  | K8  | K9  | K10 | K11 |
|--------|------|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| $\rho = 0.90$ | 1.641 | 1.439 | 1.49 | 1.505 | 1.217 | 1.549 | 1.391 | 1.526 | 1.433 | 1.562 | 1.209 | 1.568 | 1.242 | 1.566 |
| $\rho = 0.95$ | 2.836 | 2.991 | 2.986 | 1.412 | 2.309 | 2.641 | 2.483 | 2.618 | 2.525 | 2.658 | 2.301 | 2.66 | 2.336 | 2.658 |
| $\rho = 0.99$ | 3.429 | 2.978 | 3.035 | 2.838 | 2.892 | 2.605 | 2.631 | 2.692 | 2.867 | 2.793 | 2.972 | 2.763 | 2.655 | 2.609 |

Table 3: MSE values when $n = 100$

| Method | SPRM | HK1 | HK2 | K1  | K2  | K3  | K4  | K5  | K6  | K7  | K8  | K9  | K10 | K11 |
|--------|------|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| $\rho = 0.90$ | 3.612 | 1.39 | 1.431 | 1.446 | 1.158 | 1.49 | 1.332 | 1.467 | 1.374 | 1.503 | 1.15 | 1.509 | 1.183 | 1.507 |
| $\rho = 0.95$ | 4.699 | 2.695 | 1.819 | 1.245 | 2.142 | 2.474 | 2.316 | 2.451 | 2.358 | 2.491 | 2.135 | 2.493 | 2.167 | 2.491 |
| $\rho = 0.99$ | 6.216 | 2.905 | 2.772 | 2.575 | 2.629 | 2.342 | 2.368 | 2.429 | 2.604 | 2.53 | 2.709 | 2.5 | 2.392 | 2.346 |

5. Conclusion

In this paper, numerous selection methods of the shrinkage parameter are explored and investigated of ridge estimator for semiparametric regression model. According to Monte Carlo simulation studies, it has been seen that some estimator can bring significant improvement relative to others, in terms of MSE. The $K_2$ and $K_8$ improved the performance of the semiparametric ridge regression compared to SPRM estimator when $\rho = 0.90$, as well as $K_3$ and $K_{11}$ when $\rho = 0.99$, and $K_1$ at $\rho = 0.95$, in all the cases without any domination. In conclusion, the use of these estimators is recommended when multicollinearity is present in the semiparametric regression model.

REFERENCES

1. Akdeniz, F., et al., Efficiency of the generalized difference-based Liu estimators in semiparametric regression models with correlated errors. *Journal of Statistical Computation and Simulation*, 2013. 85(1): p. 147-165.
2. Akdeniz, F. and E.A. Duran, Liu-type estimator in semiparametric regression models. *Journal of Statistical Computation and Simulation*, 2010. 80(8): p. 853-871.
3. Emami, H., Local influence for Liu estimators in semiparametric linear models. *Statistical Papers*, 2016. 59(2): p. 529-544.
4. Yatchew, A.J.E.l, An elementary estimator of the partial linear model. 1997. 57(2): p. 135-143.
5. Wei, C. and X. Wang, Liu-type estimator in semiparametric partially linear additive models. *Journal of Nonparametric Statistics*, 2016. 28(3): p. 459-468.

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

7. Roozbeh, M. and M. Arashi, New Ridge Regression Estimator in Semiparametric Regression Models. *Communications in Statistics - Simulation and Computation*, 2015. **45(10)**: p. 3683-3715.

8. Roozbeh, M., M. Arashi, and H.A. Niroumand, Semiparametric Ridge Regression Approach in Partially Linear Models. *Communications in Statistics - Simulation and Computation*, 2010. **39(3)**: p. 449-460.

9. Ullah, A., et al., A semiparametric generalized ridge estimator and link with model averaging. *Econometric Reviews*, 2015. **36(1-3)**: p. 370-384.

10. Hoerl, A.E. and R.W. Kennard, *Ridge regression: Biased estimation for nonorthogonal problems*. *Technometrics*, 1970. **12(1)**: p. 55-67.

11. Kibria, B.M.G., Performance of some new ridge regression estimators. *Communications in Statistics - Simulation and Computation*, 2003. **32(2)**: p. 419-435.

12. Kibria, B.M.G., K. Månsson, and G. Shukur, A Simulation Study of Some Biasing Parameters for the Ridge Type Estimation of Poisson Regression. *Communications in Statistics - Simulation and Computation*, 2015. **44(4)**: p. 943-957.

13. Dorugade, A. and D. Kashid, Alternative method for choosing ridge parameter for regression. *Applied Mathematical Sciences*, 2010. **4(9)**: p. 447-456.

14. Asar, Y., A. Karaibrahimoğlu, and A. Genç, Modified ridge regression parameters: A comparative Monte Carlo study. *Hacettepe Journal of Mathematics and Statistics*, 2014. **43(5)**: p. 827-841.

15. Bhat, S.S., A comparative study on the performance of new ridge estimators. *Pakistan Journal of Statistics and Operation Research*, 2016. **12(2)**: p. 317-325.

16. Kibria, B.M.G., K. Månsson, and G. Shukur, Performance of some logistic ridge regression estimators. *Computational Economics*, 2012. **40(4)**: p. pp 401–414.

17. Muniz, G. and B.G. Kibria, On some ridge regression estimators: An empirical comparisons. *Communications in Statistics—Simulation and Computation*, 2009. **38(3)**: p. 621-630.