〈**Research Paper**〉

Journal of the Korean
Statistical Society
Vol. 12, No. 2, 1983

The m^{th} Moment of Generalized Ridge Estimators

Ju Sung Kim*

ABSTRACT

Dwivedi, Srivastava and Hall (1980) derived the first and second moments of generalized ridge estimators. In this paper we consider the m^{th} moment of a generalized ridge estimator and tabulate its skewness measure.

1. Introduction

Hoerl and kennard (1970 a,b) proposed the ridge regression method of estimation and defined a class of estimators characterized by a scalar. By reducing the regression model to its canonical form, they defined the generalized ridge regression estimator and suggested the existence theorem of the scalar and an iterative procedure. Dwivedi, Srivastava and Hall (1980) derived the exact expressions for the first and second moments of the generalized ridge estimator employing the initial choice of characterizing scalars as recommended by Hoerl and Kennard (1970 a), and tabulated the values of the relative bias, the relative mean squared error and the efficiency with respect to the ordinary least squares estimator for a few selected values of the noncentrality parameter and the degree of freedom.

In this paper we consider the m^{th} moment of a generalized ridge estimator and tabulate the skewness of its distribution.

2. The Model and the Estimators

Consider the canonical form of a multiple linear regression model

* Dept. of Computer Science and Statistics, Chung-buk National University

$$y=X\beta+e. \tag{2.1}$$

where y is an $n\times1$ vector of observations, X is an $n\times p$ matrix, with full column rank, of n observations on p explanatory variables, β is a $p\times1$ vector of regression coefficients and e is an $n\times1$ vector of errors with

$$E(e)=0 \text{ and } E(ee')=\sigma^2 I_n. \tag{2.2}$$

Following Hoerl and Kennard (1970 a), we assume that $X'X=\Lambda$ where $\Lambda=\text{diag}\ (\lambda_1, \cdots\cdots, \lambda_p)$ is a $p\times p$ diagonal matrix. Then the generalized ridge regression estimator of β is given by

$$\beta^*=(\Lambda+K)^{-1}X'y \tag{2.3}$$

where K is a diagonal matrix with nonnegative elements $k_1, \cdots\cdots, k_p$ as the characterizing scalars. The least squares estimator of β is given by

$$b=\Lambda^{-1}X'y \tag{2.4}$$

Then

$$E(\beta^*-\beta)=[(\Lambda+K)^{-1}\Lambda-I_p]\beta, \tag{2.5}$$

$$\begin{aligned} E(\beta^*-\beta)(\beta^*-\beta)' &= \sigma^2\Lambda(\Lambda+K)^{-2}+E(\beta^*-\beta)E(\beta^*-\beta)' \\ &= \sigma^2\Lambda(\Lambda+K)^{-2}+(I_p-\Lambda(\Lambda+K)^{-1})\beta\beta'(I_p-\Lambda(\Lambda+K)^{-1}) \end{aligned} \tag{2.6}$$

and

$$E(\beta^*-\beta)'(\beta^*-\beta)=\sum_{i=1}^{p}\frac{\sigma^2\lambda_i+\beta_i^2k_i^2}{(\lambda_i+k_i)^2} \tag{2.7}$$

provided the k_i's are nonstochastic.

Differentiation of (2.7) with respect to k_i yields

$$\frac{\partial}{\partial k_i}E(\beta^*-\beta)'(\beta^*-\beta)=\frac{2\lambda_i(\lambda_i+k_i)(k_i\beta_i^2-\sigma^2)}{(\lambda_i+k_i)^4}=0, \quad i=1, 2, \cdots, p. \tag{2.8}$$

So the optimum values $k_i(\text{opt})=\sigma^2/\beta_i^2,\ i=1, 2, \cdots, p.$

Hoerl and Kennard (1970a) suggest an iterative procedure to estimate k_i. This prccedure may be described by

$$k_{i(j)}=S^2/(\beta^*_{i(j)}),\ i=1, 2, \cdots, p, \tag{2.9}$$

where the bracketed subscript j is used to denote the j^{th} iteration and b is the least squares estimate of β_i and

$$S^2=(y-Xb)'(y-Xb)/\nu \tag{2.10}$$

is an unbiased estimator of σ^2, where ν is the degree of freedom. As an initial estimate of β_i take

$$\beta^*_{i(0)}=b_i,\quad i=1,2,\cdots,p. \tag{2.11}$$

Replacing k_i in K by $\hat{k}_i$ to form $\hat{K}$ and substituting it in (2.3) leads to an adaptive estimator of β, i.e.,

$$\hat{\beta}^*=(\Lambda+\hat{K})^{-1}X'y. \tag{2.12}$$

3. The Main Theorem

We assume that e follows a multivariate normal distribution $N(0,\sigma^2 I_n)$. If we partition X as

$$X=(x_1,x_2,\cdots,x_p), \tag{3.1}$$

where x_i, $i=1,2,\cdots,p$, is the i^{th} column vector of X, then the i^{th} element of $\hat{\beta}^*$ is given by $\hat{\beta}_i^*=x_i'y/(\lambda_i+\hat{k}_i)$. (3.2)

THEOREM: The m^{th} moment of $\hat{\beta}_i^*$ of (3.2) is given by

$$E(\hat{\beta}_i^{*m})=\frac{\beta_i^m e^{-\tau_i^2/2}}{\tau_i^m(m-1)!}\sum_{y=0}^{\infty}(y+m-1)!\left(\frac{\nu-1}{\nu}\right)^y\frac{2^{1/2}\,\Gamma\left(y+\frac{\nu}{2}\right)}{y!\,\Gamma(\nu/2)}\cdot$$

$$\sum_{j=0}^{\infty}\frac{\Gamma\left(j+\frac{m+\nu+1}{2}\right)\Gamma\left(j+\frac{3m+1}{2}\right)}{\Gamma\left(y+j+\frac{3m+\nu+1}{2}\right)\Gamma\left(j+\frac{1}{2}\right)}\frac{\left(\frac{1}{2}\tau_i^2\right)^j}{j!} \tag{3.3}$$

if m is an even integer and

$$E(\hat{\beta}_i^{*m})=\frac{\beta_i^m.\,e^{-\tau_i^2/2}}{\tau_i^m(m-1)!}\sum_{y=0}^{\infty}(y+m-1)!\left(\frac{\nu-1}{\nu}\right)^y\frac{2^{2m-\frac{1}{2}}\tau_i\;\Gamma\left(y+\frac{\nu}{2}\right)}{y!\;\Gamma\left(\frac{\nu}{2}\right)}\cdot$$

$$\sum_{j=0}^{\infty}\frac{\Gamma\left(j+\frac{m+\nu+2}{2}\right)\Gamma\left(j+\frac{3m+2}{2}\right)}{\Gamma\left(y+j+\frac{3m+\nu+2}{2}\right)\Gamma\left(j+\frac{3}{2}\right)}\frac{\left(\frac{\tau_i^2}{2}\right)^j}{j!} \tag{3.4}$$

if m is an odd integer, where $\tau_i=\sqrt{\lambda_i}\beta/\sigma$.

PROOF: When e follows a multivariate normal distribution $N(0,\sigma^2 I_n)$, it is well known that the distribution of $b=\Lambda^{-1}X'y$ is a multivariate normal $N(\beta,\sigma^2\Lambda^{-1})$. Let

$$Z_i=\frac{\sqrt{\lambda_i}b_i}{\sigma}=\frac{x_i'y}{\sigma\sqrt{\lambda_i}}$$

then the distribution of Z_i is a normal with mean τ_i and variance unity. From (3.2), we can write

$$\hat{\beta}_i^* = \frac{x_i'y}{\lambda_i + S^2/b_i^2} = \frac{\sigma}{\sqrt{\lambda_i}} \frac{Z_i^3}{Z_i^2 + S^2/\sigma^2}$$

$$= \frac{\sigma}{\sqrt{\lambda_i}} \frac{Z_i^3}{(Z_i^2 + V)} \left[1 + \left(\frac{\nu-1}{\nu}\right)\left(\frac{V}{Z_i^2 + V}\right)\right]^{-1}$$

where $b_i = x_i'y/\lambda_i$ and $V = \nu S^2/\sigma^2$.

Thus

$$\hat{\beta}_i^{*m} = \frac{\sigma^m \, Z_i^{3m}}{\lambda_i^{\frac{1}{2}m} \, (Z_i^2 + V)^m} \left[1 - \left(\frac{\nu-1}{\nu}\right)\left(\frac{V}{Z_i^2 + V}\right)\right]^{-m}.$$

Since $\left(\frac{\nu-1}{\nu}\right)\left(\frac{V}{Z_i^2+V}\right) < 1$, by Taylor series expansion, we can expand

$$\left[1 - \left(\frac{\nu-1}{\nu}\right)\left(\frac{V}{Z_i^2 + V}\right)\right]^{-m}$$

so that

$$\hat{\beta}_i^{*m} = \frac{\sigma^m}{\lambda_i^{\frac{1}{2}m}} \frac{Z_i^{3m}}{(Z_i^2 + V)^m} \sum_{y=0}^{\infty} \frac{(y+m-1)!}{y!(m-1)!} \left(\frac{\nu-1}{\nu}\right)^y \left(\frac{V}{Z_i^2 + V}\right)$$

$$= \frac{\beta_i^m}{\tau_i^m (m-1)!} \sum_{y=0}^{\infty} \frac{(y+m-1)!}{y!} \left(\frac{\nu-1}{\nu}\right)^y \frac{Z_i^{3m} V^y}{(Z_i^2 + V)^{m+y}} .$$

Therefore

$$E(\hat{\beta}_i^{*m}) = \frac{\beta_i^m}{\tau_i^m (m-1)!} \sum_{y=0}^{\infty} \frac{(y+m-1)!}{y!} \left(\frac{\nu-1}{\nu}\right)^y E\left[\frac{Z_i^{3m} V^y}{(Z_i^2 + V)^{m+y}}\right] .$$

Due to Dwivedi *et al.* (1980), if m is an even integer

$$E(\hat{\beta}_i^{*m}) = \frac{\beta_i^m}{\tau_i^m (m-1)!} \sum_{y=0}^{\infty} \frac{(y+m-1)!}{y!} \left(\frac{\nu-1}{\nu}\right)^y \frac{2^{\frac{1}{2}m} \Gamma\left(y + \frac{\nu}{2}\right) e^{-\frac{1}{2}\tau_i^2}}{\Gamma(\nu/2)}$$

$$\cdot \sum_{j=0}^{\infty} \frac{\Gamma\left(j + \frac{m+\nu+1}{2}\right) \Gamma\left(j + \frac{3m+1}{2}\right) \left(\frac{1}{2}\tau_i^2\right)^j}{\Gamma\left(y + j + \frac{3m+\nu+1}{2}\right) \Gamma\left(j + \frac{1}{2}\right) \quad j!}$$

$$= \frac{\beta_i^m}{\tau_i^m (m-1)!} \sum_{y=0}^{\infty} \sum_{j=0}^{\infty} \frac{(y+m-1)! \left(\frac{\nu-1}{\nu}\right)^y 2^{\frac{1}{2}m} \Gamma\left(y + \frac{\nu}{2}\right) e^{-\frac{1}{2}\tau_i^2}}{y! \Gamma(\nu/2) \, \Gamma\left(y + j + \frac{3m+\nu+1}{2}\right) \Gamma\left(j + \frac{1}{2}\right)}$$

$$\cdot \frac{\Gamma\left(j + \frac{m+\nu+1}{2}\right) \Gamma\left(j + \frac{3m+1}{2}\right) \left(\frac{1}{2}\tau_i^2\right)^j}{j!}$$

and if m is an odd integer

$$E(\hat{\beta}_i^{*m}) = \frac{\beta_i^m}{\tau_i^m (m-1)!} \sum_{y=0}^{\infty} \frac{(y+m-1)!}{y!} \left(\frac{\nu-1}{\nu}\right)^y \frac{2^{2m - \frac{1}{2}}\tau_i \, \Gamma\left(y + \frac{\nu}{2}\right) e^{-\frac{1}{2}\tau_i^2}}{\Gamma\left(\frac{\nu}{2}\right)}$$

$$\cdot \sum_{j=0}^{\infty} \frac{\Gamma\left(j+\frac{m+\nu+2}{2}\right) \Gamma\left(j+\frac{3m+2}{2}\right) \left(\frac{1}{2}\tau_i^2\right)^y}{\Gamma\left(y+j+\frac{3m+\nu+2}{2}\right) \Gamma\left(j+\frac{3}{2}\right) j!}$$

$$= \frac{\beta_i^m}{\tau_i^{m-1}(m-1)!} \sum_{y=0}^{\infty} \sum_{j=0}^{\infty} \frac{(y+m-1)!\left(\frac{\nu-1}{\nu}\right)^y 2^{2m-\frac{1}{2}} \Gamma\left(y+\frac{\nu}{2}\right)}{y!\ \Gamma\left(\frac{\nu}{2}\right)}$$

$$\cdot \frac{\Gamma\left(j+\frac{m+\nu+2}{2}\right) \Gamma\left(j+\frac{3m+2}{2}\right) \left(\frac{1}{2}\tau_i^2\right)^j e^{-\frac{1}{2}\tau_i^2}}{\Gamma\left(y+j+\frac{3m+\nu+2}{2}\right) j!\ \Gamma\left(y+\frac{3}{2}\right)} .$$

4. Skewness

Dwivedi *et al.* (1980) showed that

(i) $\hat{\beta}_i^*$ is biased in the direction which is opposite to the sign of β_i.

(ii) the relative mean squared error decreases as τ_i^2 increases, and as ν grows large, it decreases as long as $\tau_i^2 \leqslant 2$.

(iii) the magnitude of relative bias is a decreasing function of τ_i^2 while it is an increasing function of ν.

(iv) $\hat{\beta}_i^*$ is more efficient as long as $\tau_i^2 \leqslant 2$. Substantial gain is achieved where ν is large and τ_i^2 is small.

It is interesting to note that the relative m^{th} moment

$$RM(\hat{\beta}_i^{*m}) = E(\hat{\beta}_i^{*m}/\beta_i^m)$$

is a function of ν and τ_i^2 only. Let

$$d_1 = [E(\hat{\beta}_i^* - E(\hat{\beta}^*))^3]^2/[E(\hat{\beta}_i^* - E(\hat{\beta}_i^*))^2]^3$$

and

$$d_2 = E(\hat{\beta}_i^* - E(\hat{\beta}^*))^4/[E(\hat{\beta}_i^* - E(\hat{\beta}_i^*))^2]^2.$$

Then the Pearson skewness measure of the distribution of $\hat{\beta}_i^*$ [Kendall and Stuart (1969)]

$$Sk(\hat{\beta}_i^*) = \frac{\sqrt{d_1}(d_2+3)}{2(5d_2-6d_1-9)}$$

is a function of ν and τ_i^2 only, and can be evaluated for all values of ν and τ_i^2.

Table 1 shows the skewness measure of the distribution of $\hat{\beta}_i^*$ for a few selected values of ν and τ_i^2. The skewness decreases as τ_i^2 increases. And as ν increases, it increases, decreases, and then increases again.

5. Acknowledgement

The author wishes to thank the referees for helpful comments.

Table 1. **The Skewness measure of the distribution of $\hat{\beta}_i^*$**

$\frac{1}{2}\tau_i^2$ \ ν	1	2	5	10	15	20	30	40	50	60
.01	.371	.395	.391	.392	.394	.396	.400	.403	.405	.407
.05	.364	.391	.386	.387	.389	.391	.395	.398	.400	.402
.10	.345	.380	.375	.375	.377	.379	.383	.386	.388	.390
.20	.291	.346	.338	.337	.340	.342	.346	.350	.352	.354
.50	.157	.224	.213	.211	.213	.215	.219	.223	.225	.227
.70	.110	.166	.156	.154	.156	.157	.160	.162	.164	.166
.90	.083	.126	.118	.116	.117	.118	.120	.122	.123	.124
1.00	.073	.111	.104	.102	.103	.104	.105	.107	.108	.109
2.00	.030	.043	.040	.040	.039	.039	.039	.039	.039	.039
5.00	.010	.011	.011	.011	.011	.010	.010	.010	.010	.009
10.00	.017	.021	.019	.083	.186	-.340	-.546	.000	.000	.000
20.00	.003	.004	.004	.004	.004	.004	.004	.004	.005	.005
50.00	-223.605	-223.605	-223.605	-223.605	-223.605	-223.605	-223.605	-223.605	-223.605	-223.605

REFERENCES

(1) Dwivedi, T. D., Srivastava V. K. and Hall, R. L. (1980). Finite Sample Properties of Ridge Estimators, *Technometrics*, Vol. 22, No. 2, 205—212.

(2) Hemmerle, W. J. (1975). An Explicit Solution for Generalized Ridge Regression, *Technometrics*, Vol. 17, 309—314.

(3) Hoerl, A. E. and Kennard, R. W. (1970a). Ridge Regression; Biased Estimation for Nonorthogonal Problems, *Technometrics*, Vol. 12, 55—67.

(4) Hoerl A. E. and Kennard, R. W. (1970b). Ridge Regression; Applications to Nonorthogonal Problems, *Technometrics*, Vol. 12, 69—82.

(5) Kendall, M. G. and Stuart, A. (1969). *The Advanced Theory of Statistics*, Vol. 1. 3rd ed., Hafner Publishing Co.