APPLICATION OF MONTE CARLO SIMULATION TO RIDGE REGRESSION

Wen-Fu P. Shih, Florida Atlantic University
Sun-Fu Shih, University of Florida, IFAS, AREC

ABSTRACT

Ridge regression has been introduced to solve the multicollinearity in multiple linear regression. The performance of approximate ridge estimators (AOPT) and Optimume ridge coefficients (OPT) are compared with ordinary least square (OLS) estimators by using the Monte Carlo simulation technique. The results indicated that the correlation coefficient r between two independent variables is an important factor to choose the method. For instance, when the r value is less than 0.5, the OLS performs as good as AOPT and OPT methods. When the r lies around 0.6 to 0.8, OPT is the best technique among those three methods, and AOPT is better than OLS. For those data with high correlations such as 0.9, both OPT and AOPT are all good methods to use. In general conclusion, the performance of OPT is better than AOPT, and the AOPT is better than OLS in terms of minimizing the mean square error of estimation in regression analysis to solve the multicollinearity among the independent variables.

INTRODUCTION

Multiple linear regression technique has been extensively used by the fields of engineering, science, technology, economics and social science for data analysis. But, the estimation of regression coefficients can present problems when multicollinearity (highly correlated independent variables) exists among variables. For discussion of problems of multicollinearity, see Althauser, 1971; Blalock, 1963, 1944; Christ, 1966; Gordon, 1963; Johnston, 1972; Rockwell, 1975.

The problems of multicollinearity have been solved by one statistical technique called ridge regression which was introduced by Hoerl and Kennard (1970a, 1970b) and applied by others (Deegan, 1975; Dempster, Schartzoff and Wermuth, 1977; McDonald and Schwing, 1973; McDonald and Galarneau, 1973; Vinod, 1975; etc.) These authors showed that by adding a small non-negative constant k to the diagonal of the correlation matrix of independent variables to substantially reduce error variance and thereby control for the general instability of ordinary least square (OLS) estimates.

The question remains, however, of the appropriate amount of bias to introduce as the ridge analysis increment. In recent papers Hoerl, et.al. (1975) have suggested an approximation to the optimum value k (AOPT) so that ridge regression produces a smaller square error than OLS. Shih and Kasarda (1977) have proposed a method for selecting the optimal k (OPT) for ridge analysis in terms of minimizing the mean square error of estimation.

The purpose of this paper is based on the Monte Carlo simulation technique to compare the performance of ordinary least square (OLS) with AOPT and OPT ridge regressions by simulating the different patterns of regression coefficients with different degrees of collinearity or multicollinearity among independent variables.

METHODOLOGY DESCRIPTION

Ridge Regression and Optimization

Consider the standard model for multiple linear regression:

$$Y = X\beta + \varepsilon \tag{1}$$

where Y is a nx1 vector of observations on a dependent variable, X is a nxp matrix of nonstochastic regressors with rank p, β is a px1 vector of unknown regression coefficients, and ε is a nx1 vector of unknown disturbances. Assuming $E(\varepsilon)=0$, and $E(\varepsilon\varepsilon')=\sigma^2 I_n$, the ordinary least square estimator (OLS) of β is

$$\hat{\beta} = (X'X)^{-1}X'Y \tag{2}$$

with $\mathrm{Var}(\hat{\beta}) = \sigma^2(X'X)^{-1}$ (3)

where $\hat{\beta}$ is an unbiased estimator of β and has the minimum variance within the class of unbiased estimators (Goldberger, 1964). As we noted, if the X's are highly collinear, the variance of $\hat{\beta}$ tends to become large, and little confidence can be placed in $\hat{\beta}$ as an estimator of β. By adding positive constant k to each of the diagonal elements of $X'X$ one can reduce the variance of the regression estimate, but at the expense of some bias. The resulting is the ridge estimator

$$\beta^* = (X'X + kI)^{-1}X'Y \tag{4}$$

where k is a positive scalar, and β^* is a biased estimator of β with

$$\mathrm{Var}(\beta^*) = (X'X+kI)^{-1}X'X(X'X+kI)^{-1}\sigma^2 \tag{5}$$

β^* is equal to β when k equals zero.

As have been shown in Hoerl and Kennard (1970 a, b), the mean square error of ridge estimator β^* can be written as

$$\mathrm{MSE}(\beta^*) = \sum_i \mathrm{VAR}(\beta_i^*) + \mathrm{Bias}^2(\beta^*) \tag{6}$$

where

$$\sum_i \mathrm{VAR}(\beta_i^*) = \sigma^2 T_r(Z) \tag{7}$$

with

$$Z = (X'X + kI)^{-1}X'X(X'X + kI)^{-1} \tag{8}$$

and

$$\mathrm{Bias}^2(\beta^*) = [E(\beta^*-\beta)]'[E(\beta^*-\beta)] \tag{9}$$

In equation 6, the total variance decreases as k increases, while the square bias increases with k. Based on these monotonic properties and existence of minimum point which also shown by Hoerl and Kennard (1970a, b), Shih and Kasarda (1977) have also shown that by computerized iteration procedures one can locate the optimum point k, which minimizes a consisten estimator of MSE (β^*), namely mse($\beta^*(k)$). Where

$$\mathrm{mse}(\beta^*(k)) = \hat{\sigma}^2 T_r(Z) + (\beta^*(k)-\hat{\beta})'(\beta^*(k)-\hat{\beta}) \tag{10}$$

where

$$\hat{\sigma}^2 = (Y'Y - \hat{\beta}X'Y)/(n-p) \tag{11}$$

Let V(k) denote an estimator of total variance of $\beta^*(k)$ and BS(k) denote an estimator of the bias square, then equation (10) becomes

$$\mathrm{mse}(\beta^*(k)) = V(k) + BS(k) \tag{12}$$

where

$V(k) = \hat{\sigma}^2 T_r(Z)$,

$BS(k) = (\beta^*(k)-\hat{\beta})'(\beta^*(k)-\hat{\beta})$,

$V(0) = \hat{\sigma}^2 T_r(X'X)^{-1}$, and

$S(0) = 0 \quad \text{for} \quad k \geq 0$

For given $k_1 > k_2 \geq 0$, we know that

$$V(k_2) \geq V(k_1) \tag{13}$$

and

$$S(k_2) \leq S(k_1) \tag{14}$$

The existence of a minimum point shows that

$$mse[\beta^*(k-c)] > mse[\beta^*(k)] < mse[\beta^*(k+c)] \tag{15}$$

for a small constant c and leads to the conclusion that k is a point which gives the minimum mse(β^*).

The iteration procedures to obtain this point can be summarized as follows:

(a) Read input data and desired tolerance of accuracy.

(b) Compute $\hat{\sigma}^2$ and $\hat{\beta}$ from equations 11 and 2, respectively.

(c) Initiate the k = 0 and an increment $\Delta k = 0.1$.

(d) Compute V(0) and BS(0).

(e) Let a new variable k1 = k + Δk.

(f) Compute V(k1) and BS(k1).

(g) Check the relationship between BS(k1) - BS(k) and V(k) - V(k1).

(h) If BS(k1) - BS(k) < V(k) - V(k1), then let k = k1, V(k) = V(k1), BS(k) = BS(k1), and the procedures of e, f, and g are repeated.

(i) If BS(k1) - BS(k) $\geq$ V(k) - V(k1), then check if the Δk is less than a desired tolerance. If the answer is no, the Δk is replaced by Δk/10 and the procedures of e, f, and g are repeated. If the answer is yes, the iteration procedures are complete, and k1 is the optimal value.

The above procedures have been converted to a computer program with Fortran IV language.

Hoerl, et. al. (1975) also suggested that an approximation method to obtain the optimal value k_a (AOPT) can be expressed as follows:

$$k_a \doteq P\hat{\sigma}^2/\hat{\beta}'\hat{\beta} \tag{16}$$

By simulation technique they showed that k_a can produce a smaller average square error than OLS, the distribution of squared errors for the regression coefficients has a smaller variance than does that for OLS, and the probability that the ridge regression produces a smaller square error than OLS is greater than 0.50.

Monte Carlo Simulations

Applying simulation techniques to examine the performance of ridge regression has been found in many recent publications (McDonald and Galarnean, 1973; Hoerl, et.al., 1975; and Dempster, et.al., 1977). The basic Monte Carlo simulation used in this study is described as follows. The observations x_{ij} are generated based on the following simulation generator:

$$x_{ij} = (1-\alpha^2)^{0.5} Z_{ij} + \alpha Z_{i(p+1)}; \quad i = 1,\dots,n;\ j = 1,\dots,p. \tag{17}$$

where, n is the number of observations for each explanatory variable; p is the number of independent variables; Z_{ij} are independent standard normal pseudo-random numbers and α is specified so that the correlation between any two independent variables is given by α^2. The X's are then standardized so that $X'X$ is in correlation form. A true regression coefficient β is chosen as a normalized eigenvector corresponding to the largest eigenvalues of the $X'X$ matrix. Newhouse and Oman (1971) have noted that MSE(β^*) is minimized when β is such eigenvector subject to the constraint that $||\beta|| = 1$ when k is fixed.

Observations on the dependent variable are determined by

$$y_i = \beta_0 + \beta_1 x_{i1} + \dots + \beta_p x_{ip} + \varepsilon_i \tag{18}$$

where x's are computed from equation 17, ε_i are independent normal $(0, \sigma^2)$ psuedo-random numbers and β_0 is taken to be zero. The variables are standardized so that $X'Y$ represents the vector of correlations of the dependent variable with each independent variable.

Based on the Central Limit Theory, the major error involved in Monte Carlo simulation is a statistical sample error which is proportional to the $(1/\sqrt{N})$, where N is the total number of trials (Shih and Hamrick, 1974). In other words, one must increase the sample size by a factor of 4 in order to halve the possible error. Therefore, an additional set of samples W must be generated to increase the accuracy of simulation.

Comparisons of Different Methods

The observations generated by Monte Carlo simulation are used to perform ordinary least square (OLS) estimators, optimum ridge coefficients (OPT), and approximate ridge estimators (AOPT). The symbols $\hat{\beta}$, $\beta^*(k)$, and $\beta^{**}(k)$ represent the standardized coefficients of OLS, OPT and AOPT, respectively. Those standardized coefficients are then transformed back to the original coefficients. The constant terms are then computed as:

$$\hat{\beta}_0 = \bar{y} - \sum_{j=1}^{p} \hat{\beta}_j \bar{x}_j \quad \text{for OLS} \tag{19}$$

$$\beta_0^*(k) = \bar{y} - \sum_{j=1}^{p} \beta_j^*(k) \bar{x}_j \quad \text{for OPT} \tag{20}$$

and

$$\beta^{**}(k) = \bar{y} - \sum_{j=1}^{p} \beta_j^{**}(k) \bar{x}_j \quad \text{for AOPT} \tag{21}$$

where

$$\bar{y} = (1/n)\left(\sum_{i=1}^{n} y_i\right)$$

$$\bar{x}_j = (1/n)\left(\sum_{i=1}^{n} x_{ij}\right)$$

The total mean square errors are computed as

$$L = L(0) = \sum_{i=0}^{p} (\beta_i - \hat{\beta}_i)^2 \quad \text{for OLS} \tag{22}$$

$$L^* = L[\beta^*(k)] = \sum_{i=0}^{p} [\beta_i - \beta_i^*(k)]^2 \quad \text{for OPT} \tag{23}$$

$$L^{**} = L[\beta^{**}(k)] = \sum_{i=0}^{p} [\beta_i - \beta_i^{**}(k)]^2 \quad \text{for AOPT} \tag{24}$$

As noted above β_0 equals zero.

In order to find the regression coefficients which can produce a smaller mean square error than the corresponding least squares estimator a measure of the improvement can be obtained by

$$M^* = E[L^*(k)]/E[L(0)] \tag{25}$$

and

$$M^{**} = E[L^{**}(k)]/E[L(0)] \tag{26}$$

where E[L(k)] is the average sum of mean square error of specific ridge estimator.

$$E[L^*(k)] = (1/W) \sum_{w=1}^{W} L_w^* \tag{27}$$

$$E[L^{**}(k)] = (1/W) \sum_{w=1}^{W} L_w^{**} \quad (28)$$

and

$$E[L(0)] = (1/W) \sum_{w=1}^{W} L_w \quad (29)$$

where L_w, L_w^*, and L_w^{**} are total mean square error of sample w for OLS, OPT, and AOPT, respectively. W is the number of sample sets as indicated in the section of Monte Carlo simulation. The smaller value of M* and M** implies that the method used has a better solution.

EXAMPLE OF APPLICATIONS

The values of $n = 100$; $p = 3$; $\alpha = 0.6, 0.7, 0.8, 0.9, 0.95$, and 0.99; and $\sigma = 0.01, 0.21, 0.41, 0.61, 0.81$, and 1.01 are used in this study. The coefficients of β corresponding to each α value are given in Table 1.

TABLE 1: Values of α and Coefficient Vectors β Used in Simulation

α	β_1	β_2	β_3
0.60	.575	.573	.583
0.70	.576	.574	.582
0.80	.576	.575	.581
0.90	.577	.576	.580
0.95	.577	.577	.579
0.99	.577	.577	.578

Combining the six coefficients of α with six standard deviations of σ, thirty-six sets of data are generated. The values of standardized coefficients and total mean square are computed based on equations 19, 20, 21, 22, 23, and 24.

As mentioned in previous sections of Monte Carlo simulation, the accuracy of estimations can be improved by increasing the sample size so that additional 50 samples of observations with $n = 100$ and $p = 3$ are generated to each of the 36 different sets of data. The independent variables and true coefficient β are unchanged, while the random error term is varied, so that the dependent variable is changed. The average of optimum k for OPT and AOPT; and average mean square for OLS, OPT and AOPT in these 50 samples are computed. The results are also listed in Table 2. The measure of M* and M** used to compare the different methods are computed based on equations 25 and 26. The results are also listed in Table 2.

RESULTS AND DISCUSSIONS

As the example given in previous sections the results of the performance of MSE in each method of OLS, OPT and AOPT with $\alpha = 0.6, 0.7, 0.8, 0.9, 0.95$, and 0.99 corresponding to the six error terms σ were plotted on Figures 1, 2, 3, 4, 5, and 6, respectively. The following conclusions can be drawn.

First, when the correlation coefficient between independent variables are less than 0.5 (i.e. the cases of $\alpha = 0.6$ and 0.7 as shown in Figure 1 and 2), the MSE of OLS is close to the ridge estimators of OPT and AOPT. The deviation of result between OPT and AOPT is also negligible. As Table 2 shows, the value of M* and M** are close to 1 when the r value equal 0.36 and 0.49. This implies that the OLS method is as good as OPT and AOPT methods when the correlation coefficient is less than 0.5.

Second, when the correlation coefficient exist between 0.6 and 0.8 (i.e. the cases of $\alpha = 0.8$ and 0.9 plotted on Figures 3 and 4), the MSE of OLS is much larger than OPT and AOPT, and OPT is much smaller than AOPT. As Table 2 shows, the value of M* is much less than M**. These imply that the ridge regression analysis is required when correlation coefficients exist between 0.6 and 0.8, and the OPT method is much better than AOPT method.

Third, when the correlation coefficient is greater than 0.9 (i.e. the cases of $\alpha = 0.95$ and 0.99 as shown in Figures 5 and 6), the MSE of OPT is much larger than both OPT and AOPT methods and the performance of the two ridge type estimators gave an approximate same solution. As Table 2 shows, the value of M* is similar to the M** when α equals 0.95 and 0.99. These imply that the ridge analysis is required when the correlations is greater than 0.9 and both OPT and AOPT methods give a similar solution.

Fourth, all M* and M** in Table 2 are decreasing while error terms σ are increasing. This implies that higher the error term σ the better ridge estimator is performed. For instance, M* equals .972 when $\sigma = 0.01$ and equals .379 when $\sigma = 1.01$ for the data with $\alpha = .95$. This means that the MSE of OPT is 97% of OLS when error is 0.01, but it is only 38% of the OLS when error becomes 1.01.

Fifth, as Table 2 shows, the value of M* is much smaller than M** and M** is smaller than or equal to one. This concludes that the performance of OPT is better than AOPT, and the AOPT is better than OLS in terms of minimizing the mean square error of estimation in regression analysis to solve the multicollinearity among the independent variables.

SUMMARY AND CONCLUSIONS

The estimation of regression coefficients in multiple linear regression can present problems when multicollinearity exists among independent variables. This type of problem has been solved by one of the statistical methods called ridge regression. This technique shows that by adding a non-negative constant "k" to the diagonal of correlation matrix it is possible to substantially reduce error variance of estimators. The methods of optimum ridge coefficients (OPT) and approximate ridge estimators (AOPT) are used in this study to compare the performance of each technique with the ordinary least square (OLS) estimators.

The Monte Carlo simulations are used to generate the observations of independent variables. The simulations are performed based on different correlation coefficients and error terms. Comparisons of the AOPT and OPT methods are made with OLS technique. The results indicated that when correlation between two independent variables is less than 0.5, the OLS performs as good as ridge regression, i.e., the multicollinearity problem does not exist in this case. But, when correlation lies aroun 0.6 and 0.8, OPT method is considered the best among those three methods, and AOPT method is better than OLS method. For those data with high correlation such as 0.9,

TABLE 2: Comparisons of the Simulation Results of AOPT and OPT Methods with OLS Method in Different Correlation Coefficient and Standard Deviations.

α	Correl. Coeffi. r	Stand. Devia. σ	OLS L(0)	OPT L(B*(k))	OPT Ave. k	AOPT L(B**(k))	AOPT Ave. k	M*	M**
0.60	0.36	0.01	0.000	0.000	0.000	0.000	0.000	1.000	1.000
		0.21	0.026	0.026	0.106	0.025	0.009	1.000	0.962
		0.41	0.081	0.068	0.212	0.072	0.034	0.834	0.891
		0.61	0.224	0.164	0.273	0.183	0.071	0.733	0.816
		0.81	0.271	0.193	0.375	0.204	0.122	0.694	0.754
		1.01	0.596	0.413	0.384	0.440	0.184	0.693	0.739
0.70	0.49	0.01	0.000	0.000	0.000	0.000	0.000	1.000	1.000
		0.21	0.031	0.029	0.122	0.029	0.010	0.954	0.948
		0.41	0.096	0.072	0.227	0.082	0.037	0.746	0.856
		0.61	0.264	0.174	0.277	0.205	0.075	0.658	0.774
		0.81	0.319	0.200	0.376	0.222	0.129	0.627	0.693
		1.01	0.703	0.428	0.377	0.476	0.188	0.609	0.677
0.80	0.64	0.01	0.000	0.000	0.000	0.000	0.000	1.000	1.000
		0.21	0.041	0.033	0.136	0.038	0.011	0.805	0.927
		0.41	0.127	0.074	0.224	0.101	0.039	0.583	0.798
		0.61	0.349	0.189	0.276	0.242	0.079	0.541	0.693
		0.81	0.418	0.219	0.375	0.255	0.135	0.524	0.610
		1.01	0.926	0.404	0.353	0.548	0.186	0.501	0.592
0.90	0.81	0.01	0.000	0.000	0.001	0.002	0.000	1.032	1.000
		0.21	0.073	0.033	0.156	0.063	0.012	0.452	0.863
		0.41	0.273	0.133	0.200	0.190	0.042	0.487	0.693
		0.61	0.445	0.158	0.257	0.234	0.081	0.355	0.524
		0.81	0.872	0.368	0.324	0.437	0.137	0.422	0.501
		1.01	1.301	0.509	0.357	0.582	0.185	0.391	0.447
0.95	0.91	0.01	0.000	0.000	0.010	0.000	0.000	0.972	1.000
		0.21	0.136	0.055	0.122	0.106	0.013	0.402	0.780
		0.41	0.415	0.095	0.169	0.247	0.040	0.230	0.523
		0.61	1.147	0.437	0.177	0.542	0.072	0.381	0.472
		0.81	1.352	0.434	0.250	0.507	0.120	0.321	0.375
		1.01	2.987	1.133	0.194	1.201	0.128	0.379	0.402
0.99	0.98	0.01	0.002	0.001	0.031	0.002	0.000	0.795	0.800
		0.21	0.641	0.231	0.086	0.332	0.010	0.357	0.517
		0.41	1.960	0.495	0.100	0.606	0.026	0.253	0.309
		0.61	5.454	1.968	0.079	1.950	0.036	0.361	0.357
		0.81	6.418	1.488	0.130	1.797	0.066	0.226	0.281
		1.01	14.092	4.971	0.104	4.763	0.053	0.352	0.338

both OPT and AOPT methods are good techniques to solve the multicollinearity in linear regression models.

REFERENCES

Althauser, R.P. (1971). "Multicollinearity and Non-Additive Regression Models," pp. 453-472 in H.M. Blalock, ed., Causal Models in the Social Sciences," Chicago: Aldine.

Blalock, H.M., Jr. (1963). "Correlated Independent Variables: The Problem of Multicollinearity." Social Forces 42:233-237.

______________. (1964). Causal Inferences in Nonexperimental Research. Chapel Hill: University of North Carolina Press.

Christ, C.F. (1966). Econometric Models and Methods. New York: Wiley.

Deegan, J., Jr. (1975). "The Process of Political Development.: Sociological Methods and Research 3:384-415.

Dempster, A.P., M. Schatzoff, and N. Wermuth. (1977). "A Simulation Study of Alternative to Ordinary Least Square." Journal of the American Statistical Association 72:77-90.

Goldberger, A.S. (1964). Econometric Theory. New York: Wiley.

Gordon, R.A. (1968). "Issues in Multiple Regression." American Journal of Sociology 73:592-616.

Hoerl, A.E. and R.W. Kennard (1970a). "Ridge Regression: Biased Estimation for Nonorthogonal Problems." Technometrics 12:55-68.

Hoerl, A.E. and R.W. Kennard (1970b). "Ridge Regression: Applications to Nonorthogonal Problems." Technometrics 12:69-82.

Hoerl, A.E., R.W. Kennard, and K.F. Baldwin (1975). "Ridge Regression: Some Simulations." Communications in Statistics 4(2):105-123.

Johnston, J. (1972). Econometric Methods, 2nd ed. New York: McGraw-Hill.

McDonald, G.C. and D.I. Galarneau (1973). "A Monte Carlo Evaluation of Some Ridge Type Estimators," Publication GMR-1322-B, General Motors Research Laboratories, Warren, Michigan.

McDonald, G.C. and R.C. Richard (1973). "Instabilities of Regression Estimates Relating Air Pollution to Mortality." Technometrics 15, 15:463-481.

Newhouse, J.P. and S.D. Oman (1971). "An Evaluation of Ridge Estimators," Rand Report No. R-716-PR:1-28.

Rockwell, R.C. (1975). "Assessment of Multicollinearity." Sociological Methods and Research 3:308-320.

Shih, S.F. and R.L. Hamrick (1974). "A Technique Used to Determine Random Point Position." Water Resource Bulletin 10(5):884-898.

Shih, W.F.P. and J.D. Kasarda (1977). "Optimal Bias in Ridge Regression Approach to Multicollinearity." Sociological Methods & Research, 5:461-470.

Vinod, H.D., (1976). "Application of New Ridge Regression Method to A Study of Bell System Scale Economics." Journal of the American Statistical Association 71:835-841.

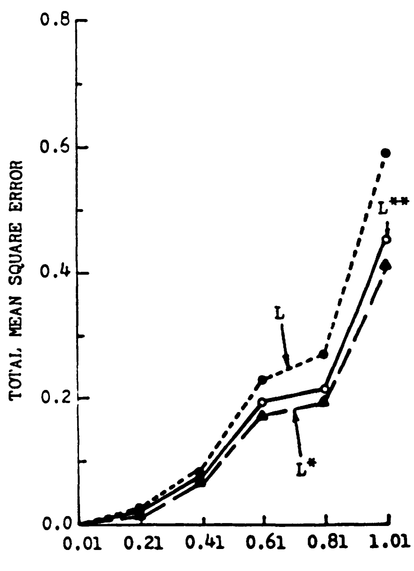

Fig. 1. Total mean square error related to standard deviation, α = 0.6.

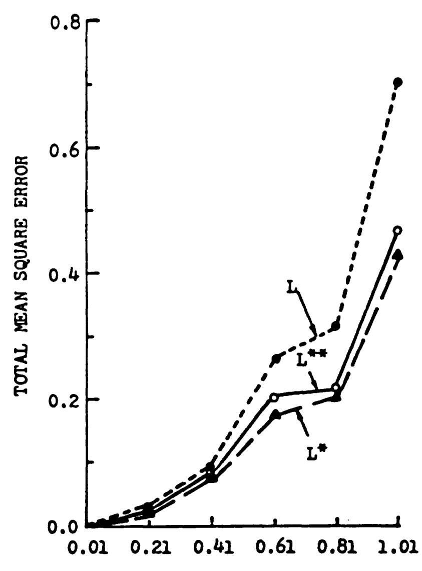

Fig. 2. Total mean square error related to standard deviation, α = 0.7.

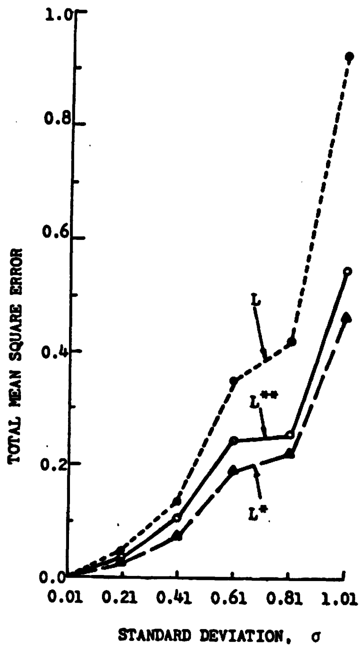

Fig. 3. Total mean square error related to standard deviation, $\alpha = 0.8$.

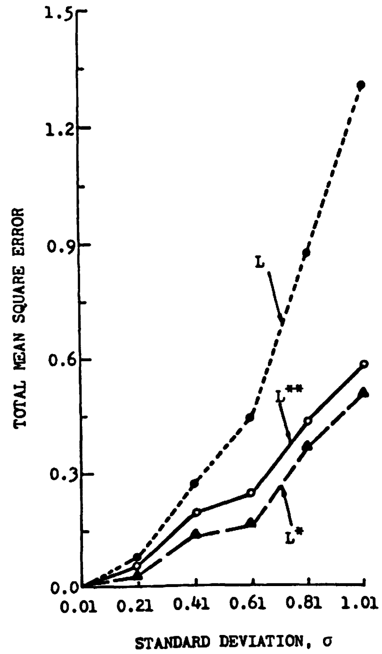

Fig. 4. Total mean square error related to standard deviation, $\alpha = 0.9$.

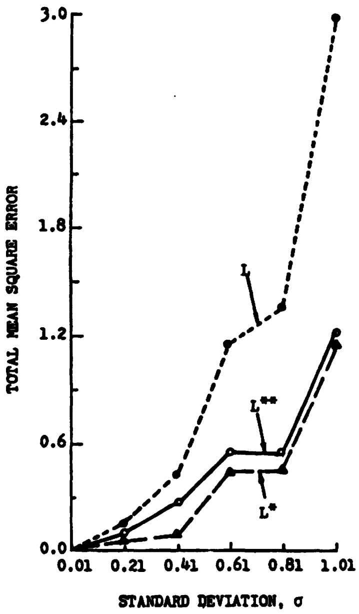

Fig. 5. Total mean square error related to standard deviation, $\alpha = 0.95$.

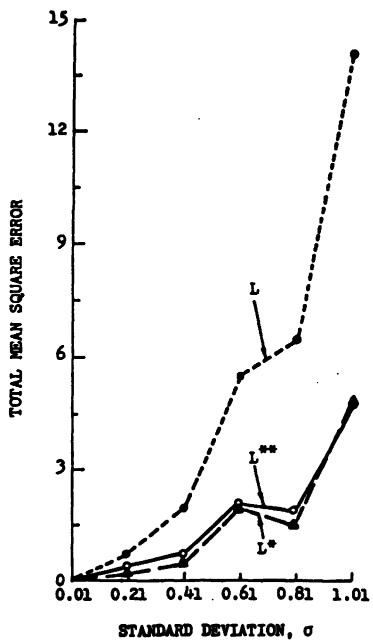

Fig. 6. Total mean square error related to standard deviation, $\alpha = 0.99$.