THE EGYPTIAN STATISTICAL JOURNAL
ISSR, CAIRO UNIV. VOL.39 NO.1, 1995

A New Method for Generating the Design Matrix of a Linear Regression Model

By

Adel M. Hallawa
Faculty of Economics and administrative Sciences, U.A.E. Univ.*

Abdul-Mordy H. Azzam
Faculty of Economics and Management King Saud University, Saudi Arabia*

This paper introduces a new algorithm for generating the design matrix X, of a linear regression model, with prespecified simple correlation coefficients between each pair of its columns. Controlling the correlation coefficients among the regressors makes the proposed algorithm useful for simulation studies of biased estimation techniques under linear regression models where different degrees of multicollinearity have to be investigated to judge their performance. Unlike the existing algorithms, the new generated matrices have certain desirable features that will be discussed in the sequel. A small simulation study, reveals some differences between the results obtained from applying the new and the McDonald and Galarneau (1975) algorithms. Our main claim is that, the results associated with the new algorithm describe the true state of nature more precisely.

Key words: Correlation Matrix; Positive Definite Matrices; Eigenvalues; Orthonormal Eigenvectors; Ridge Regression.

1- Introduction

Biased estimation techniques are among the basic tools to handle the problem of multicollinearity, which arises from the ill structure of the design matrix X, in the linear regression model, $Y=X\beta+\sigma\epsilon$. Comparisons among these techniques are always done using simulation studies. In order to draw valid conclusions from such comparisons, the simulation studies must cover several situations that reflect different degrees of the severity of the multicollinearity problem. In particular, when using ridge regression technique, we find that, there are several alternatives related to the selection of the ridge parameters. Comparisons among these selection rules should be done under different degrees of multicollinearity, for the sake of determining whether there exists a best rule, or determining the conditions under which a certain rule is preferred to the others. Thus, it is important

---

The authors are on leave from Alex. Univ., Egypt

to execute the simulation studies under different settings of the design matrix X, or some of its characteristics that reflect different degrees of multicollinearity.

Simulation studies, in the literature of biased estimation techniques, can be classified into two main groups. Studies that belong to the first group, start with generating the matrix X'X, which we will always assume that it is written in a correlation form, e.g., Dempster, Schatzoff and Wermuth (1977), or start with selecting different hypothesized sets of eigenvalues for the matrix X'X, e.g., El Bassiouni and El Sayed (1986). The studies within this group rely completely on the normality assumption where, for each simulation run, estimates from the sampling distributions of the sufficient statistics X'Y, or some linear function of it, and $Y'(I-P_X)Y$, $P_X=X(X'X)^{-1}X'$, are generated and used to evaluate the required measurements. There are two limitations associated with starting the simulation by generating X'X or making certain selections of its eigenvalues. First, they are confined to the assumption that the error term has a normal distribution. Second, they do not allow a full analysis of the data. For example, it is not possible to compare between variable selection techniques and biased estimation techniques because the former requires the knowledge of the design matrix X.

Studies that belong to the second group start, naturally, with determining the design matrix X. In some situations, real data sets, which are known to be suffering from multicollinearity, are chosen to form the design matrix X in the simulation study. For example, this approach is adopted in the studies of, Hoerl and Kennard (1970), Lawless (1981) and Nebeb and Smith (1990). However, such studies are restricted to the observed state of nature and do not allow for measuring the effect of gradual changes in the degree of multicollinearity. Most of the studies within this group are based on the algorithm of McDonald and Galarneau (1975), denoted in this paper as the MG algorithm, which is used to generate the matrix X. For example, this algorithm is used in the studies of, McDonald and Galarneau (1975), Wichern and Churchill (1978) and Firinguetti (1989). According to this algorithm, the matrix X is generated such that, theoretically, the entries of the matrix $X'X=\{\omega_{ij}\}$ will be,

$$\omega_{ij}=\begin{cases}1 & i=j\\ \rho & i\neq j.\end{cases}$$

This algorithm has two main drawbacks. First, it assumes the same simple correlation coefficient among all the regressors. As will be shown, this limitation is relaxed by Wichern and Churchill (1978) who modified the algorithm to allow for a limited number of the entries in X'X to vary freely. Second, when X is generated, the entries of the matrix X'X will be different from the specified value of ρ and hence the results of any subsequent analysis will not reflect the true effect of this value. In the present work we introduce a new algorithm for generating the design matrix X. This algorithm eliminates the inconveniences associated with the MG algorithm. That is, it allows all the

ISSR, Cairo Univ, Vol.39 No.1, 1995
PP.106-119

entries of X'X to be different and it gives these entries exactly as they are prespecified. Hence, it can be said that, the new algorithm reflects the true effects of the specified correlation structure on the different aspects of the model through the simulation study.

In Section 2 we introduce the commonly used algorithms for generating the design matrix X . Section 3 introduces the new algorithm. Section 4 gives an example of a matrix generated using the new algorithm. Section 5 presents comparisons among some theoretical and simulation results obtained from applying the new and the MG algorithms under a linear regression model.

2- The McDonald and Galarneau Algorithm for Generating X

In this Section we introduce the algorithm of McDonald and Galarneau (1975) for generating the design matrix X. Also, we refer to the modification of this algorithm suggested by Wichern and Churchill (1978).

According to the MG algorithm, the regressors X_j , j=1,2,...,p, in the (n × p) design matrix X are generated using,

$$X_j = (1-\rho)^{1/2} Z_j + \rho^{1/2} Z_{p+1} \qquad j=1,2,...,p \tag{2.1}$$

where, Z_1 , Z_2 ,..., Z_{p+1} , are independent standard normal pseudo random (n × 1) vectors. It is easy to see that, the simple correlation coefficient between any two regressors generated using this algorithm is, theoretically ρ and hence the correlation matrix X'X should be of the form,

$$\begin{bmatrix} 1 & \rho & . & \rho \\ \rho & 1 & . & \rho \\ . & . & . & \rho \\ \rho & \rho & \rho & 1 \end{bmatrix}$$

To enrich the structure of X'X and make it more practical, Whichern and Churchill (1978) generated a design matrix with 5 regressors using,

$$\begin{aligned} X_j &= (1-\rho_1)^{1/2} Z_j + \rho_1{}^{1/2} Z_6 , \qquad j=1,2,3 \\ &= (1-\rho_2)^{1/2} Z_j + \rho_2{}^{1/2} Z_6 . \qquad j=4,5 \end{aligned} \tag{2.2}$$

Thus theoretically, the resulting correlation matrix has the form,

$$\begin{bmatrix} 1 & \rho_1 & \rho_1 & \rho_1\rho_2 & \rho_1\rho_2 \\ & 1 & \rho_1 & \rho_1\rho_2 & \rho_1\rho_2 \\ & & 1 & \rho_1\rho_2 & \rho_1\rho_2 \\ & & & 1 & \rho_2 \\ & & & & 1 \end{bmatrix}$$

which provides three different simple correlations two of which are free to be preassigned.

In the following we give a generalization of the algorithm in (2.2), which make it usable for all $p \geq 4$. Let $\mathcal{P}=\{1,2,...,p\}$ and let $J_1, J_2, ..., J_c$, $c \leq \frac{p}{2}$, be any set of mutually exclusive and exhaustive subsets of $\mathcal{P}$ with at least two elements in each subset, i.e.

$$\mathcal{P}=\bigcup_{i=1}^{c} J_i, \qquad J_i \bigcap J_k = \emptyset \quad i \neq k \qquad \text{and} \qquad \#(J_i) \geq 2 .$$

Also let, $\rho_1, \rho_2, ..., \rho_c$ be a set of different preassigned simple correlations then, the p regressors in the design matrix X can be generated using

$$X_j = \left(1-\rho_k\right)^{1/2} Z_j + \rho_k{}^{1/2} Z_{p+1}, \qquad k \in J_k \text{ , } k=1,2, ...,c \tag{2.3}$$

Two main disadvantages of the algorithms (2.1) and (2.3) are:

1- The structure of the resulting X'X matrix does not allow assigning different simple correlations for all different pairs of regressors. For example, (2.3) allows at most $c+\binom{c}{2}$ different simple correlations which is smaller than $\binom{p}{2}$ if all the simple correlations in X'X are different.

2- Although, $Z_1, Z_2, ..., Z_{p+1}$ are theoretically independent, their simple correlations will generally differ from zero when generated using any random number generator. As a result, the X'X matrix will not generally obey the preassigned simple correlations.

3- The New Algorithm

For a given $(p \times p)$ positive definite correlation matrix R, this algorithm is used to generate an $(n \times p)$ matrix X, whose columns may be used as the regressor variables in a linear regression model. The advantage of this algorithm is that, when the columns of X are standardized, the resulting X'X matrix will be identically R. The algorithm is based on the following theorem whose proof is obvious.

Theorem 3.1

Let H be any orthogonal $(n \times p)$ matrix and write R as, $R=G\Lambda G'$, where Λ is a $(p \times p)$ diagonal matrix with positive entries along its main diagonal and G is a $(p \times p)$ orthogonal matrix. Further,

ISSR, Cairo Univ. Vol.39 No.1, 1995
PP. 106-119

let $\Lambda^{1/2}$ denote the diagonal matrix whose elements are the square roots of the corresponding elements of Λ. If, $X=H\Lambda^{1/2}G'$ then, $X'X=R$.

Note that, the assumption of a positive definite correlation matrix R, suits all the regression situations in which the design matrix X is assumed to be full column rank, since then, X'X will be positive definite. Also, it may be noted that the above choice of H is essential to get the dimension of the matrix X as required.

To generate the matrix X we proceed as follows,

1- Generate a set of p, $(n \times 1)$ vectors, arbitrarily or using any random number generator.

2- Apply the Gram-Schmidt process to orthogonalize and normalize the p vectors generated in step 1. Form the matrix H from the resulting columns.

3- Calculate the eigenvalues and the normalized eigenvectors of the preassigned correlation matrix R.

4- Let $\Lambda^{1/2}$ be the diagonal matrix whose diagonal elements are the square roots of the calculated eigenvalues.

5- Form the columns of the matrix G from the normalized eigenvectors of step 3.

6- Set $X= H\Lambda^{1/2}G'$.

4-An Example of a Data Set Generated by the New Algorithm

Suppose that we want to run a simulation study including the four covariates, X_1, X_2, X_3 and X_4, with ten observations on each variable, such that X_1, X_2 and X_3 are independent of X_4, the correlation between X_1 and X_2 is 0.6, the correlation between X_1 and X_3 is 0.95 and that between X_2 and X_3 is 0.8. The correlation matrix under these relations among the four variables takes the form,

$$R=\begin{bmatrix} 1 & .6 & .95 & 0 \\ .6 & 1 & .8 & 0 \\ .95 & .8 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}.$$

We applied the above algorithm using the Gauss software version 2.0, to generate ten observations for each variable and the resulting data is given in Table 1. If the correlation matrix of this data is calculated, the result will be exactly the matrix R.

It may be noted that, using different random number generators or different seeds to start the generation will give different H matrices and hence different columns of the X matrix. However, the

X'X matrix will always have the same preassigned structure.

Table 1 Generated 10 observations on 4 variables using the new algorithm

X_1	X_2	X_3	X_4
-0.66801	-0.41099	-0.66614	0.21183
0.37828	-0.19903	0.16646	-0.15512
0.17111	0.22329	0.12461	0.03049
-0.27398	-0.27065	-0.25767	-0.23349
-0.24068	-0.17335	-0.22877	-0.64816
0.02719	0.58451	0.21612	0.01890
-0.03933	-0.19599	-0.08993	0.07760
0.01315	-0.20119	-0.04220	0.60828
0.16440	0.24823	0.25562	0.23861
0.46787	0.39517	0.52190	-0.14895

5- Comparisons Between the Results of Applying the New and the MG Algorithms

This Section starts with presenting X'X matrices calculated from five X matrices generated using the new and the MG algorithms with the same correlation coefficient between each pair of their columns. Also, it includes comparisons among the eigenvalues, index numbers and variance inflation factors calculated from matrices generated using each of the two algorithm. Finally, each of the five generated matrices was chosen to represent the design matrix of the regression model, $Y=X\beta+\sigma\epsilon$, which is then used to run a small simulation study as described below.

5-1 X'X matrices

In this part of the study, we used the random seeds 40619297, 12009857, 75113505 and 49531638 to generate, on the Gauss software version 2.0, four (20×4), X matrices via the MG algorithm, with simple correlation coefficient $\rho=0.9$ among each pair of the four covariates. The same seeds were used to generate four different II matrices which in turn, were used to generate different X matrices via the new algorithm, using the same value of ρ. Matrix 0 below, represent the X'X matrix calculated from each of the X matrices generated by the new algorithm. Matrices 1 through 4 represent the X'X matrices calculated from the matrices generated via the MG algorithm using the previous four seeds respectively. It is evident from the entries that, while the new algorithm gave X matrices which completely obey the preassigned correlation structure among the regressors, the correlations among the regressors generated using the MG algorithm are always different from the preassigned one.

Matrix 0

$$\begin{bmatrix} 1 & 0.9 & 0.9 & 0.9 \\ & 1 & 0.9 & 0.9 \\ & & 1 & 0.9 \\ & & & 1 \end{bmatrix}$$

Matrix 1

$$\begin{bmatrix} 1 & .812 & .760 & .893 \\ & 1 & .739 & .831 \\ & & 1 & .742 \\ & & & 1 \end{bmatrix}$$

Matrix 2

$$\begin{bmatrix} 1 & .977 & .952 & .952 \\ & 1 & .944 & .959 \\ & & 1 & .960 \\ & & & 1 \end{bmatrix}$$

Matrix 3

$$\begin{bmatrix} 1 & .908 & .883 & .865 \\ & 1 & .805 & .819 \\ & & 1 & .852 \\ & & & 1 \end{bmatrix}$$

Matrix 4

$$\begin{bmatrix} 1 & .933 & .944 & .953 \\ & 1 & .946 & .954 \\ & & 1 & .961 \\ & & & 1 \end{bmatrix}$$

5-2 Comparisons Among Non-Stochastic Measures

If we let p denote the number of regressors in the design matrix X generated using the new algorithm with the same correlation between each pair of the regressors then, the eigenvalues, the index number and the variance inflation factor of the j^{th} regressor $(VIF)_j$ are respectively given by the relations,

$$\text{Eigenvalues} = 1-\rho \qquad \text{with multiplicity (p-1),}$$

$$= 1+(p-1)\rho \qquad \text{with multiplicity one.}$$

$$\text{Index Number} = \left(1+(p-1)\rho\right)/\left(\rho-1\right)$$

$$(VIF)_j = \frac{1+(p-2)\rho}{(1-\rho)(1+(p-1)\rho)} \qquad j=1,2,\ldots,p.$$

On the other hand, if the corresponding measurements are calculated from matrices generated from the MG algorithm, they will differ from one seed to another and will differ from those calculated under the new algorithm. To illustrate the significance of such differences, we generated 30, (20×4) X matrices

using each algorithm with ρ=0.9, 0.95 and 0.99 . For each ρ, the index numbers, the sorted eigenvalues and the sorted variance inflation factors were calculated from the 30 design matrices generated using the MG algorithm. The results were then used to report, the averages, the standard

Table 2 Summary Statistics of Index Numbers of 30 X Matrices Generated Using the MG Algorithm

ρ	True	Mean	Standar Error	t	ConfidenceLimits Lower	Upper
0.9	37	74.891	4.767	7.948	65.135	84.647
0.95	77	155.885	10.099	7.811	135.232	176.539
0.99	397	802.925	53.051	7.652	694.436	911.413

Table 3 Summary Statistics of Eigenvalues of 30 X Matrices Generated Using the MG Algorithm

ρ	True	Mean	Standard Error	t	ConfidenceLimits Lower	Upper
0.9	0.1	0.056	0.0035	-12.795	0.048	0.063
	0.1	0.094	0.0051	-1.174	0.083	0.104
	0.1	0.153	0.0086	6.177	0.135	0.170
	3.7	3.698	0.0142	-0.141	3.669	3.727
0.95	0.05	0.028	0.0020	-10.945	0.024	0.032
	0.05	0.047	0.0026	-1.1720	0.047	0.053
	0.05	0.077	0.0048	5.684	0.067	0.087
	3.85	3.848	0.0078	-0.255	3.832	3.864
0.99	0.01	0.0037	0.0004	-10.536	0.0048	0.0056
	0.01	0.0095	0.0005	-1	0.0083	0.0106
	0.01	0.0155	0.0011	5	0.0133	0.0177
	3.97	3.9693	0.0018	-.3895	3.9657	3.9728

Table 4 Summary Statistics of Variance inflation Factors of 30 X Matrices Generated Using the MG Algorithm

ρ	True	Mean	Standar Error	t	ConfidenceLimits Lower	Upper
0.9	7.567	6.572	0.276	-3.607	6.007	7.136
	7.567	8.275	0.377	1.877	7.504	9.045
	7.567	10.693	0.509	6.137	9.652	11.735
	7.567	13.484	0.841	7.036	11.764	15.204
0.95	15.065	13.111	0.615	-3..177	11.853	14.369
	15.065	16.443	0.7825	1.761	14.843	18.043
	15.065	21.359	1.083	5.810	19.144	27.574
	15.065	27.073	1.747	6.876	23.502	30.645
0.99	75.063	65.873	3.431	-2.678	58.856	72.890
	75.063	81.780	4.092	1.642	73.413	90.148
	75.063	106.298	5.885	5.308	94.264	118.332
	75.063	135.621	8.964	6.756	117.290	153.952

errors, the t-statistics for testing the significance of the deviation from the correct values of the new algorithm and the limits of a 95% confidence intervals. Tables 2,3 and 4 respectively give the results for, the index numbers, the eigenvalues and the variance inflation factors.

The averages and t-statistics in Table 2 show that, the index numbers of the matrices generated using the MG algorithm are always much greater than the correct values and vary internally over a wide range specially for ρ=0.99. Table 3 indicates that, the smallest and the third largest eigenvalues are significantly different from the correct values, while Table 4 shows that, variance inflation factors are significantly different from the correct values, except for the second largest VIF.

5-3 Simulation Results

It is well known that, the existence of high collinearity among the regressors in the linear regression model $Y=X\beta+\sigma\epsilon$, highly inflates the variance of the least squares estimator of β. In such situations, the ridge regression technique may be used to get biased estimators with smaller mean square errors.

In this subsection a simulation study is conducted to examine the effect of the algorithm used in generating X on the simulated mean square errors (MSE(0)) of the ordinary least squares estimator and the mean square errors (MSE($\hat{k}$)) of the ridge regression estimator , based on the Hoerl, Kennard and Baldwin (1975) selection rule of the ridge parameter. The simulation study was executed as a six-factor cross-classified factorial experiment with 4 replicates. The first factor in the study was the method of generation (Meth.) with two levels, the new and the MG algorithms. The second factor was the sample size (n) with two levels, n=20 and 50. The third factor was the number of regressors (p) in the design matrix X, with two levels, p=5 and 10. the fourth factor was the correlation coefficient among the regressors (ρ) with two levels, ρ=0.90 and 0.99 . The fifth factor was the variance of the error term (σ^2) with two levels, σ^2= .1 and 10 . The last factor (or.) represented the choice of the vector β where it was one time chosen to be the eigenvector corresponding to the largest eigenvalue of X'X and the other time was chosen to be the eigenvector corresponding to the smallest eigenvalue. These two choices are known as the favorable and the least favorable orientations respectively. The replicates are determined by the four seeds given in §5-2. For each replicate and under each combination of the factors levels, an (n × p) design matrix X was generated. The design matrices, thus obtained, were used to generate 500 pseudo (n × 1), $N(X\beta,\sigma^2 I_n)$ random vectors which simulate the response vectors Y. For each of the 500 simulation runs, the point estimates of β are given by,

$$\hat{\beta}(\hat{k}) = (X'X+\hat{k}I_p)^{-1}X'Y$$

where, $\hat{k} = 0$ for the least square estimator and $\hat{k} = \frac{p\hat{\sigma}^2}{\hat{\beta}(0)'\hat{\beta}(0)}$ for the ridge estimator, and

$$\hat{\sigma}^2 = (Y - X\hat{\beta}(0))'(Y - X\hat{\beta}(0)) \Big/ (n - p) .$$

The mean square errors for both types of estimators were calculated as,

$$MSE(.)= \sum_{i=1}^{500} (\hat{\beta}(.) - \beta)'(\hat{\beta}(.) - \beta) .$$

To choose the scale on which we analyze the simulation experiment, we examined several choices and it was found that, the percent change (PCh) in the mean square errors between the two estimators defined as,

Table 5 Partial analysis of variance results

Source	Degrees of Freedom	Mean Square	F-ratio	P-value
Meth.	1	.04813	440	0.00
n	1	.00309	29	0.00
p	1	.04478	410	0.00
ρ	1	.04397	402	0.00
σ^2	1	.05429	479	0.00
or.	1	.01284	117	0.00
Meth. & n	1	.02056	188	0.00
Meth. &p	1	.00970	89	0.00
n & ρ	1	.00067	6.15	0.014
n & σ^2	1	.00069	6.31	0.013
p & ρ	1	.00679	62	0.00
p & σ^2	1	.00806	75	0.00
p & or.	1	.00219	20	0.00
ρ & σ^2	1	.03554	325	0.00
ρ & oρ.	1	.00832	76	0.00
σ^2 & or.	1	.01230	113	0.00
Meth. & n & p	1	.00136	12.41	0.001
Meth. & n & ρ	1	.00048	4.39	0.037
Meth. & n & σ^2	1	.00048	4.39	0.037
n & ρ & σ^2	1	.00045	4.11	0.044
p & ρ & σ^2	1	.00522	48	0.00
0.00	1	.00139	13	0.00
p & σ^2 & or.	1	.00210	19	0.00
ρ & σ^2 & or.	1	.00797	73	0.00
p & ρ & σ^2& or.	1	.00134	12	0.001
Error	190	.0001093		

ISSR, Cairo univ. Vol.39 No.1, 1995
PP. 106-119

$$PCh = (MSE(0) - MSE(\hat{k})) \Big/ MSE(0) . \qquad (5.1)$$

was a sensible scale and had a distribution which did not depart markedly from normality. In a preliminary analysis two observations had been detected as outliers (having standardized residuals exceeded 2 in magnitude). These two observations were deleted and we analyzed the data as a general linear model since it became no more balanced.

In reporting the results of the analysis, our comments will mainly be restricted to those concerning the effects of the methods of generating X and how they interact with the other factors involved in the experiment. The main and interaction effects of the remaining factors will not be emphasized since they are well known in the literature of ridge regression, and they are merely confirmed here in the presence of the new method of generation.

Table 5 gives partial results of the analysis of variance of the simulation experiment. We

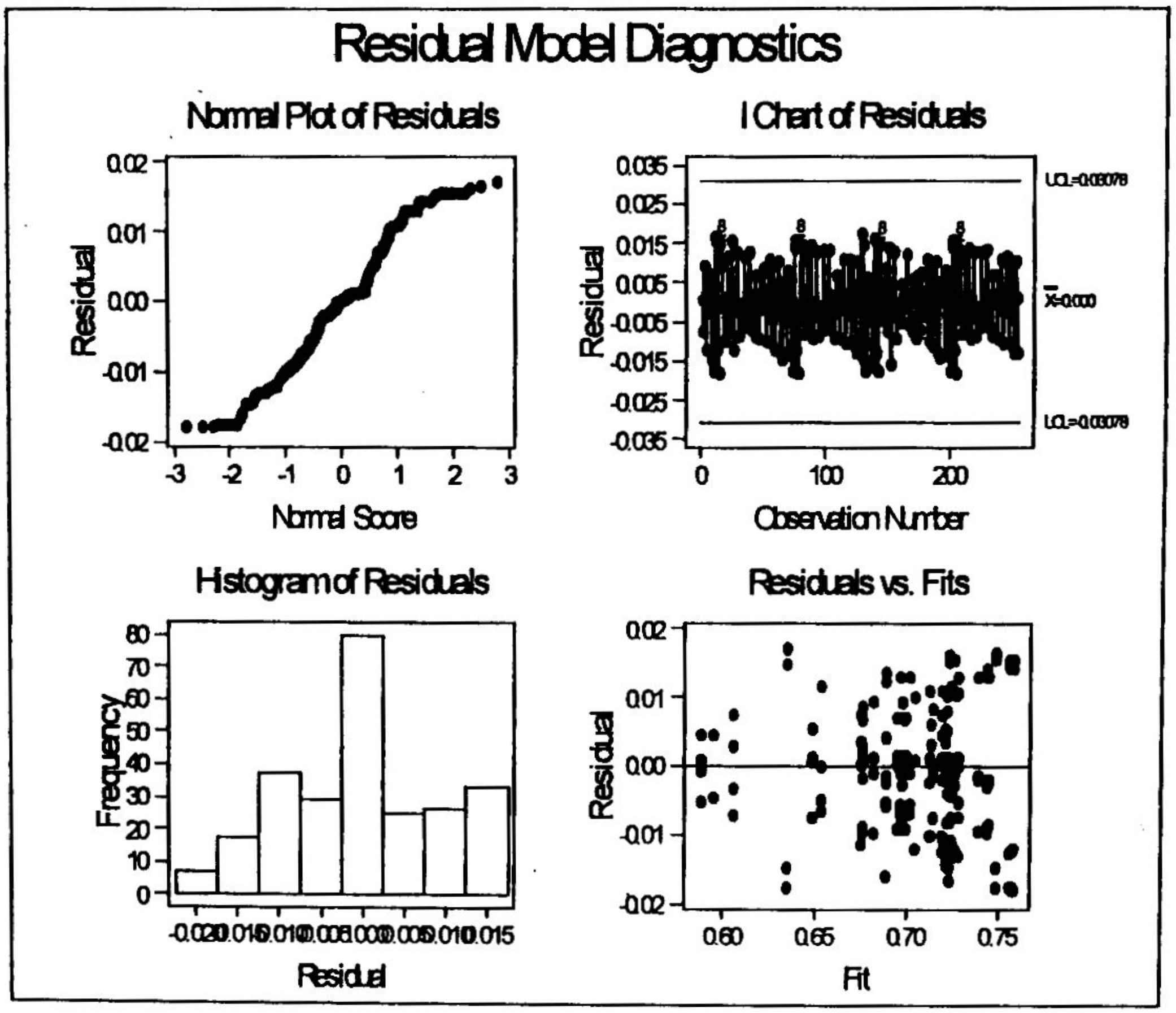

Figure (1)

excluded from the table all the nonsignificant effects at level 0.05 . As the table shows, the main effects of the method of generation factor, as well as the other five factors, are significant. For the 2-factor interactions, the method of generation is found to interact significantly only with n and with p . The significant 3-factor interactions, involving the method of generation are, (Meth.&n&p), (Meth.&n&ρ) and (Meth.&n&σ^2). No other higher order interactions, including the method of generation, are found to be significant.

Figure (1) includes, the normal probability plot, a control chart, a histogram and a residual plot for the residuals of the fitted model. The control chart and the residual plot show that, all the residuals are within control limits and their variances are homogeneous, which fulfill the basic requirement for the validity of the given analysis. On the other hand, the shapes of the histogram and the normal probability plot indicate that the distribution of the residual is symmetric with heavy tails. However, since the analysis of variance models are robust under moderate departures from normality, we proceed with the analysis using the data as is.

Table (6) and Figure (2), present the means of the PCh corresponding to the significant 3-factor interactions, including the method of generation. The numbers between parentheses beneath each PCh mean, give the corresponding MSE(0). Substitution for the mean PCh and MSE(0) in (5.1), gives the simulated MSE($\hat{k}$). Examination of the table and the figure indicates the following trends,

Table 6 Mean PCh and MSE(0)

n	20		50	
	p=5	p=10	p=5	p=10
MG	.7066 (1327)	.7522 (5613)	.7093 (1020)	.7346 (3273)
New	.6779 (845)	.6895 (2542)	.7003 (850)	.7173 (2499)
	ρ=.9	ρ=.99	ρ=.9	ρ=.99
MG	.7203 (627)	.7385 (6313)	.7104 (391)	.7335 (3902)
New	.6697 (308)	.6977 (3078)	.6942 (305)	.7233 (3044)
	σ^2=.1	σ^2=10	σ^2=.1	σ^2=10
MG	.7189 (69)	.7399 (6871)	.7089 (43)	.7349 (4251)
New	.6682 (34)	.6992 (3354)	.6927 (33)	.7249 (3316)

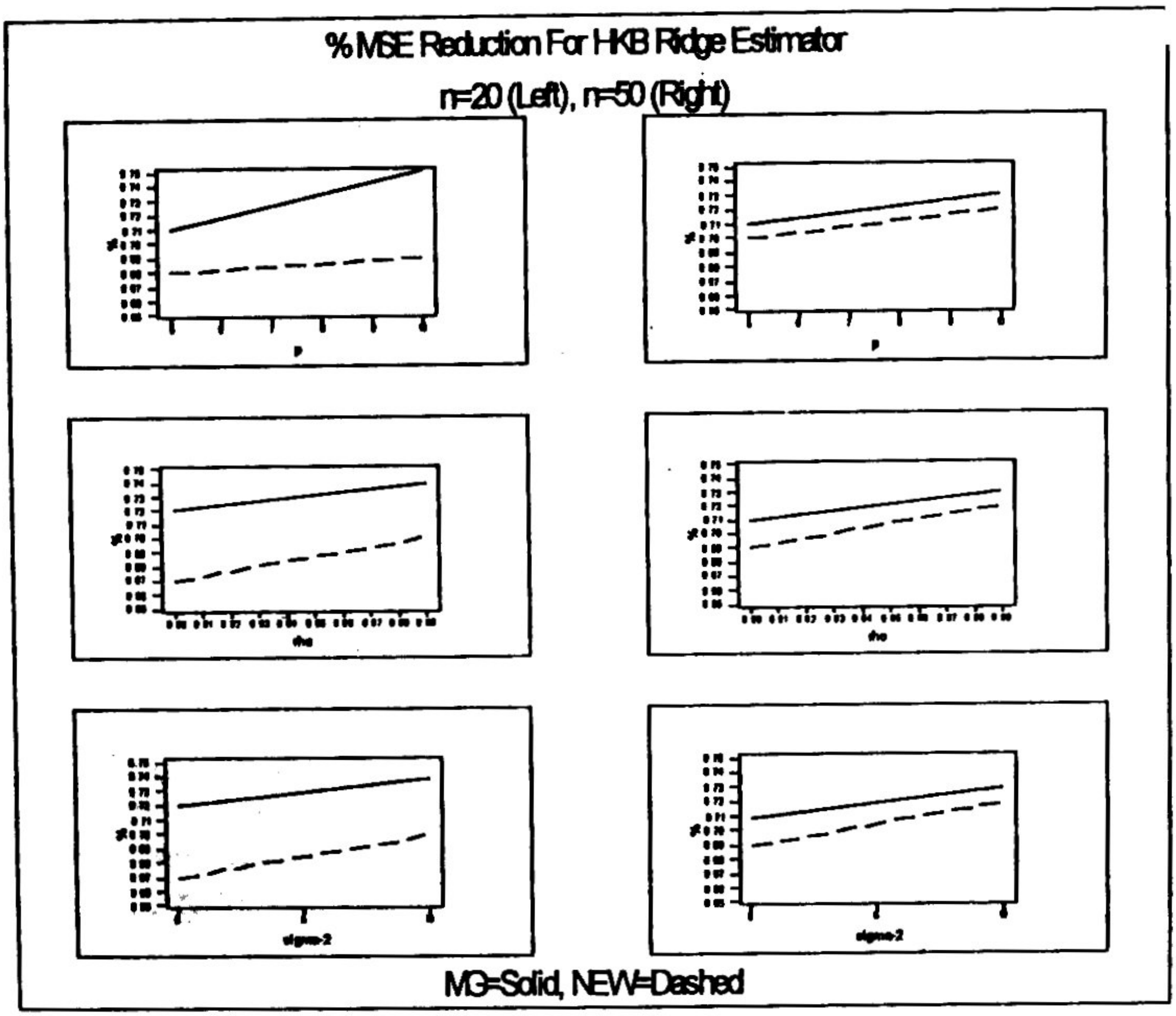

Figure (2)

1- The positive signs of the PCh means imply that under the present choices of the simulation parameters, MSE($\hat{k}$) is always smaller than MSE(0). However, this may not be the case under other choices.

2- The solid lines in the figure are always higher than the dashed lines. This implies that the MG method tends to exaggerate the percent reduction of MSE($\hat{k}$) relative to the new method.

3- The differences in the means of PCh between the two methods are smaller under n=50 than what they are under n=20. This may be the case because, under larger sample sizes, we expect the covariance structure of the design matrix generated via the MG algorithm to be close to the specified structure which is satisfied by the matrix generated via the new algorithm.

4- The largest gaps between the means of the PCh of the two methods of generation are observed under, the small sample size with large p, the small sample size with small ρ and under the small sample size with small σ^2.

5- The mean square errors of the least squares estimator under the MG method are always greater than those observed under the new method (doubled when n=20).

6- Despite result 2, the mean square errors of the ridge estimator, if calculated from (5.1), under the MG method will be found, in most cases, much greater than those calculated under the new algorithm.

Acknowledgement

The authors would like to express their thanks to professor Yehia El-Bassiouni for his help in reading this article and for his very fruitful suggestions about the analysis of Section 5.

Bibliography

1- Dempster, A. , Schatzoff, M. and Wermuth, N. (1977). " A Simulation Study of Alternatives to Ordinary Least Squares" J.A.S.A., 72, 77-106.

2- El Bassiouni, M. Y. and El Sayed, S.M. (1986). "On the Selection of the Ridge Parameter" The 21st. Annual Conference in Statistics, Computer Science and Operation Research Cairo University. Vol. 1, 1-18

3- Firinguetti, L. (1989). "A simulation Study of Ridge Regression Estimators with Autocorrelated Error" Communi. Statist. – Simula. Computa., 18, 673-702.

4- Hoerl, A.E. and Kennard, R.W. (1970-b). "Ridge Regression: Application to Non-Orthogonal Problem" Techno., 12, 69-82.

5- Hoerl, A.E., Kennard, R.W. and Baldwin, K.F. (1975). "Ridge Regression: Some Simulations" Communi. Statist., 4, 105-123.

6- Lawless, J.F. (1981). "Mean Squared Properties of Generalized Ridge Estimators" J.A.S.A., 76, 462-466.

7- Lawless, J.F. and Wang, P. (1976). "A Simulation Study of Ridge and other Regression Estimators" Communi. Statist., 5, 307-323.

8- McDonald, D.C. and Galarneau, D. (1975). "A Monte-Carlo Evaluation of some Ridge Type Estimators" J.A.S.A., 70, 407-416.

9- Nebebe, F. and Smith, A.B. (1990). " The Relative Performance of Improved Ridge Estimators and an Empirical Bayes Estimator" Communi. Statist., 19, 3469-3495.

10- Wichern, D.W. and Churchill, G.A. (1978). "A Comparison of Ridge Estimators" Techno., 20, 301-311.