Jour. Ind. Soc. Ag. Statistics
43 (2), 1991 : 172—185

Five Factor Central Composite Designs Robust to a Missing Observation

M. Akhtar,
Islamia University, Bahawalpur, Pakistan
(Received : March, 1990)

Summary

'Loss' due to a single missing observation is investigated in five factor central composite designs (c.c.d.) with different configurations of half and complete replicate of factorial part one or two replicate of axial part and some centre points. The variances of parameter estimates are also studied. Minimaxloss criterion is used to develop designs robust to a single missing observation. These designs are then compared with existing c.c.d. of the same configurations.

Key words : Central composite design; Missing observation; Robust designs; Minimaxloss criterion.

Introduction

Five factor central composite design (c.c.d.) consists of

i– half or complete replicate of a 2^5 factorial design. These n_f factorial points have co-ordinates (± 1, ± 1, ± 1, ± 1, ± 1) at some convenient scale,

ii– $n_a = 10$ axial points, two on each of the five axes at a distance $\pm\alpha$ from the centre of the design and with co-ordinates $(\pm\alpha, 0, 0, 0, 0)$, $(0, \pm\alpha, 0, 0, 0), \ldots, (0, 0, 0, 0, \pm\alpha)$ and

iii– $n_c = 1$ or more centre points with co-ordinates (o, o, o, o, o).

Total design points $n = n_f + n_a + n_c$

C.c.d. with different properties may be obtained by taking different values of α i.e. taking axial points at different distances from the centre of the design. Box [3] introduces orthogonal c.c.d. Box and Hunter [5] developed rotatable designs which have the property of equal response variances at points equi-distant from the centre, in any direction. Box and Draper [4] developed designs robust to outliers. These designs minimize the effects of the outlying observations at different design points. Herzberg and Andrews [6], [7] and Andrews and Herzberg [2] investigated the robustness of optimal designs to missing observations, each observation having certain probability of being missed. Mckee and Kshirsagar [9] studied the effects of missing observations on the parameter estimates and their variances for c.c.d. Akhtar and Prescott [1] introduced minimaxloss criterion for the selection of designs robust to missing observation. They also developed c.c.d. of different configurations robust to one or two missing observations.

Herzberg, Prescott and Akhtar [8] showed that it is not possible to obtain any design which retains equal information when any one and any two and also any three observations are missing.

In this paper the effect of a missing observation on $|X'X|$ is investigated for five factor c.c.d. of different configurations of factorial and axial parts. The designs for which the maximum reduction in $|X'X|$ due to a single missing observation is minimum, are developed. The variances of parameter estimates of these designs are also studied.

2. *Minimaxloss Criterion*

Akhtar and Prescott [1] developed minimaxloss criterion for the selection of designs robust to missing observations.

Consider a response surface model with p parameters

$$y = X\beta + \varepsilon$$

where $y = (y_1, \ldots, y_n)$ in an n x 1 vector of response at n design points, X is an nxp matrix constructed from the design matrix according to the response surface model, β is a p x 1 vector of coefficients, ε is an n x 1 vector of errors where $E(\varepsilon) = 0$ and $E(\varepsilon'\varepsilon) = \sigma^2$. The least square estimates of β and y are

$$\hat{\beta} = (X'X)^{-1}X'y$$

$$\hat{y} = X\hat{\beta} = X(X'X)^{-1}X'y = Ry$$

where $R = X(X'X)^{-1}X'$ is a matrix of order n.

For D-optimal design $|X'X|$ is maximum. In c.c.d. $|X'X|$ is an increasing function of α and is maximum when α reaches ∞. The ith observation missing reduces $|X'X|$ for the complete design to $_i|X'X|$ for $i = 1, \ldots, n$. We want $_i|X'X|$ to be as close to $|X'X|$ as possible.

It can be shown that

$$_i|X'X| = |X'X|(1 - r_{ii})$$

where r_{ii} is the diagonal element of the matrix R corresponding to the, ith design point at which the observation is missing. After some algebra it can be shown that loss due to ith observation missing

$$L_i = \frac{(\,|X'X| - |X'X|\,)}{|X'X|} = r_{11}$$

In c.c.d. L_i take three distinct values L_f , L_a and L_c for i corresponding to factorial, axial or centre point. Thus for c.c.d. Minimaxloss = Min[Max(L_i)] = min [max(L_f, L_a, L_c)]. For two or more centre points L_c L_f & L_a and max (L_i) is minimum when $L_f = L_a$. Thus α for minimaxloss design can be obtained by solving the equation $L_f = L_a$.

3. *Five Factor Designs with Half Factorial Replicate*

A five factor composite design with half replicate of a 2^5 factorial design consists of 16 factorial points, 10 axial points and one or more centre points. There are in total 27 or more design points. The half fractional factorial replicate corresponds to the highest factor interaction taken as defining contrast. Either of the two halves may be included in c.c.d. as both give identical results.

The explicit expressions for the losses L_f, L_a and L_c are

$$L_f = [(88+5n_c)\alpha^6 + 12(12+5n_c)\alpha^4 + 20(11n_c-402)\alpha^2 + 2560(10+n_c)]/B,$$

$$L_a = 8\,[(16+n_c)\alpha^6 + 4(n_c-24)\alpha^4 + 12(3n_c-20)\alpha^2 + 64(25+2n_c)]\,/\,B$$

and

$$L_c = [\,n_c + 16\,(5-\alpha^2)^2/(40+\alpha^4)^{-1}]$$

where

$$B = B[(16+n_c)\alpha^6 + 8(n_c-4)\alpha^4 + 40(n_c-22)\alpha^2 + 320(10+n_c)].$$

The plots of L_f ,L_a and L_c against α are shown in Fig. 1(a) and (b) for one or two centre points. It may be observed from these figures that L_c increases with the increase of α , reaches its maximum $1/n_c$ at $\alpha = \sqrt{k}$ and then decreases with further increases in α . Max(L_c) = $1/n_c$ means we can decrease max (L_c) by taking 2 or more centre points. L_f and L_a are decreasing and increasing functions of α and max (L_f, L_a) is minimum when $L_f = L_a$ at a particular value of α .

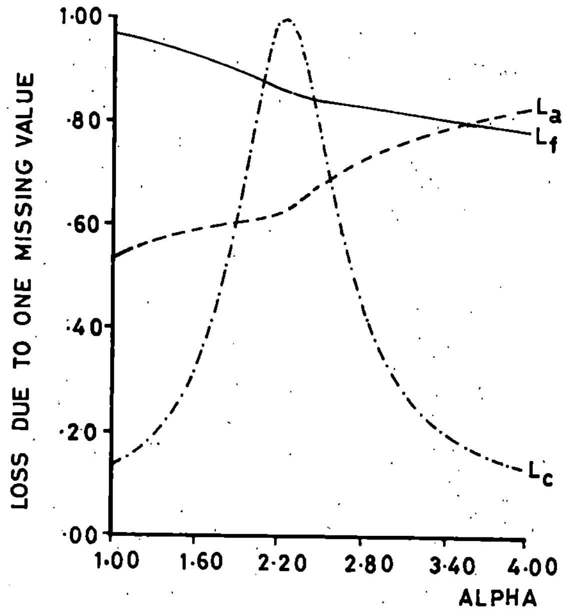
After some algebra the equation $L_f = L_a$ may be written as

$$(3n_c+40)\,\alpha^6 - 4(7n_c+228)\,\alpha^4 + 68(n_c+90)\,\alpha^2 - 512(3n_c+25) = 0$$

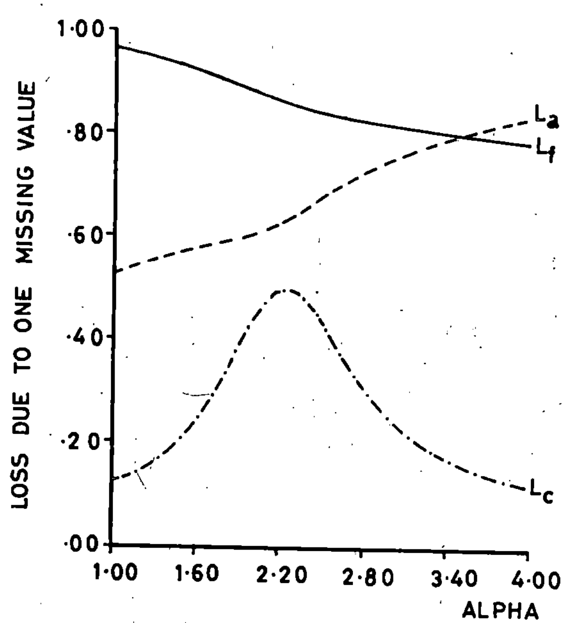
For particular value of n_c this equation may be solved for α . For example for n_c=2

Figure-1 Losses due to a factorial, an axial or a centre observation missing for design with k=5, n_f =32, n_a =10 and n_c =1 or 2, plotted against α .

(a) $n_c = 1$



(b) $n_c = 2$



this equation reduces to

$$23\alpha^6 - 48\,\alpha^4 + 3128\,\alpha^2 - 7936 = 0$$

and the only real root of this cubic equation is $\alpha^2=12.2304$ which gives $\alpha=3.4972$. The α solutions of equation $L_f = L_a$ for $n_c=1, 2, \ldots, 10$ and corresponding L_f, L_a and L_c are shown in Table 1. For all the designs in the Table $L_c < L_f = L_a$ which implies that they all are minimaxloss designs.

The variances of the parameter estimates for five factor design of this configuration may be expressed as

$$\mathrm{Var}(\hat{\beta}_0) = [\, n_c + 16(5-\alpha^2)^2/(40+\alpha^4)\,]^{-1},$$

$$\mathrm{Var}(\hat{\beta}_i) = (16 + 2\alpha^2)^{-1},$$

$$\mathrm{Var}(\hat{\beta}_{ii}) = \frac{1}{2\alpha^4}\left[1 + \frac{2\alpha^4 + 32\alpha^2 - 8(10+n_c)}{(16+n_c)\,\alpha^4 - 160\alpha^2 + 40(10+n_c)}\right]$$

and $$\mathrm{Var}(\hat{\beta}_{ij}) = \frac{1}{16}$$

The plots of these variances against α for one or two centre points are shown in Figures-2 (a) and (b) respectively. These variances are very small for minimaxloss design.

In the following we compare minimaxloss design with two centre points with the following five factor designs of the same configuration:

i. design with α=1.0 ,

ii. orthogonal design with α =1.6072 ,

iii. rotatable design with a =2.0 ,

iv. design with minimum variance of losses and two centre points with α = 2.4537

v. Box and Draper outlier robust design with α =2.6 and

vi. minimaxloss design with one centre point and α =3.5293.

The losses L_f, L_a and L_c alongwith their maximum and variance for the above mentioned designs with two or one centre point are shown in Table 2. From this table it may be seen that the maximum loss for minimaxloss design is 0.7957 which is minimum.

Table 1. Alpha solution of the equations $L_f = L_a$ and different losses for five factor designs with 1 to 10 centre points and axial part replicated twice.

n	n_c	Alpha	L_f	L_a	L_c
53	1	3.5509	0.4020	0.4020	0.0970
54	2	3.5293	0.4004	0.4004	0.0900
55	3	3.5117	0.3990	0.3990	0.0836
56	4	3.4972	0.3978	0.3978	0.0780
57	5	3.4851	0.3968	0.3968	0.0729
58	6	3.4749	0.3960	0.3960	0.0684
59	7	3.4661	0.3952	0.3952	0.0643
60	8	3.4587	0.3945	0.3945	0.0607
61	9	3.4522	0.3939	0.3939	0.0574
62	10	3.4465	0.3934	0.3934	0.0545

4. *Five Factor Designs with Axial Part Replicated Twice*

A five factor c.c.d. with full replicate of a 2^5 factorial design and axial points replicated twice consists of $n_f=32$, $n_a=20$ and one or more centre point. There are n=53 or more design points.

The explicit expressions of the losses for designs of this configuration are :

$$L_f = [(176+5n_c)\alpha^6 + 4(72+15n_c)\alpha^4 + 20(11n_c-804)\alpha^2 + 2560(20+n_c)]/B$$

$$L_a = 8[(32+n_c)\alpha^6 + 4(n_c-48)\alpha^4 + 12(3n_c-40)\alpha^2 + 128(n_c+25)]/B$$

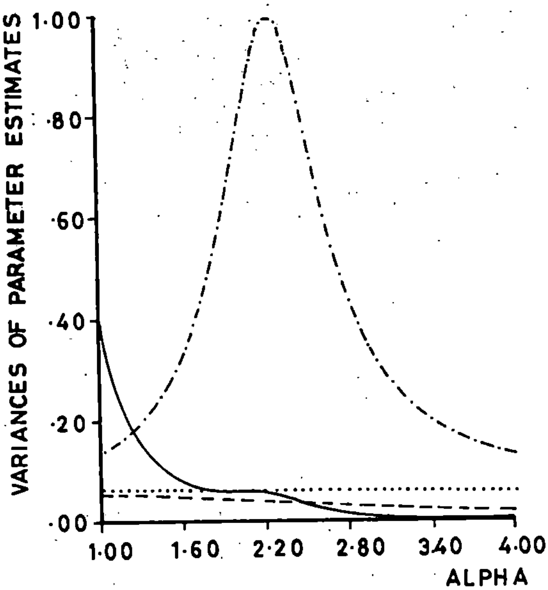
and $L_c = [n_c + 32(5-\alpha^2)^2/(40+\alpha^4)]^{-1}$

where $B = [(32+n_c)\alpha^6 + 8(n_c-8)\alpha^4 + 40(n_c-44)\alpha^2 + 320(20+n_c)]$

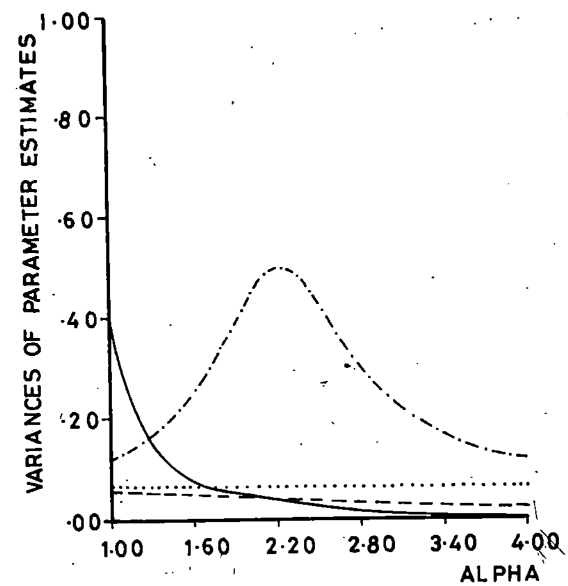
The losses L_f, L_a and L_c for one or two centre points are plotted against α in Figure-3 (a) and (b) respectively. It may be seen from these figures that losses of missing factorial and axial points are small for the whole range of α as compared to the five factor design discussed in the previous section. For one or two centre points the maximum loss corresponds to either L_f, L_c or L_a with the increasing α. The maximum loss is minimum when $L_f=L_a$ which occurs near $\alpha=3.5$ but in both

Figure-2 Variances of parameter estimate for design with k=5, n_f=16, n_a=10 and n_c=1 or 2, plotted against α.

(a) n_c=1



(b) n_c=2



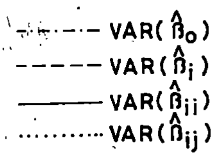


Table 2. Loss due to a single missing observation at factorial, axial or centre point together with the maximum loss and variance of losses.

No. of variables k= 5; No. of parameters p=21			Total design points n=28; No. of centre points =2					
Alpha	n	$\|X'X\|$ (complete design)	Loss due to a single missing observation				Minimum $\|X'X\|$ (Reduced design)	Variance of losses
			Factorial obs.	Axial obs.	Centre obs.	Maximum loss		
1.0000	28	0.2247E+23	0.9649	0.5319	0.1213	0.9649	0.7886E+21	0.7427E–01
	27	0.1975E+23	0.9651	0.5421	0.1380	0.9651	0.6898E+21	0.5869E–01
1.6072	28	0.5534E+26	0.9193	0.5792	0.2498	0.9193	0.4468E+25	0.4631E–01
	27	0.4152E+26	0.9208	0.5935	0.3331	0.9208	0.3290E+25	0.3325E–01
2.0000	28	0.2350E+28	0.8802	0.6042	0.4375	0.8802	0.2815E+27	0.2516E–01
	27	0.1322E+28	0.8819	0.6111	0.7778	0.8819	0.1561E+27	0.1736E–01
2.2361	28	0.2124E+29	0.8558	0.6308	0.5000	0.8558	0.3063E+28	0.1652E–01
	27	0.1062E+29	0.8558	0.6308	1.0000	1.0000	0.0000E+00	0.1395E–01
2.6000	28	0.7600E+30	0.8310	0.6928	0.3878	0.8310	0.1284E+30	0.1482E–01
	27	0.4652E+30	0.8358	0.6994	0.6336	0.8358	0.7641E+29	0.5230E–02
3.4972	28	0.2322E+34	0.7957	0.7957	0.1560	0.7957*	0.4744E+33	0.2815E–01
	27	0.1960E+34	0.8020	0.7984	0.1848	0.8020	0.3881E+33	0.1405E–01
3.5293	28	0.2991E+34	0.7946	0.7982	0.1525	0.7982	0.6035E+33	0.2848E–01
	27	0.2534E+34	0.8008	0.8008	0.1799	0.8008	0.5049E+33	0.1428E–01

* Minimaxloss due to one missing observation.

figures it has local minima. Also note that for one centre point maximum loss reaches 1 near $\alpha = 2.3$ which causes the breakdown of the design. Actually this is the point where $L_c=1/n_c$ when $\alpha = \sqrt{k}$ and for one point maximumloss=L_c =1.

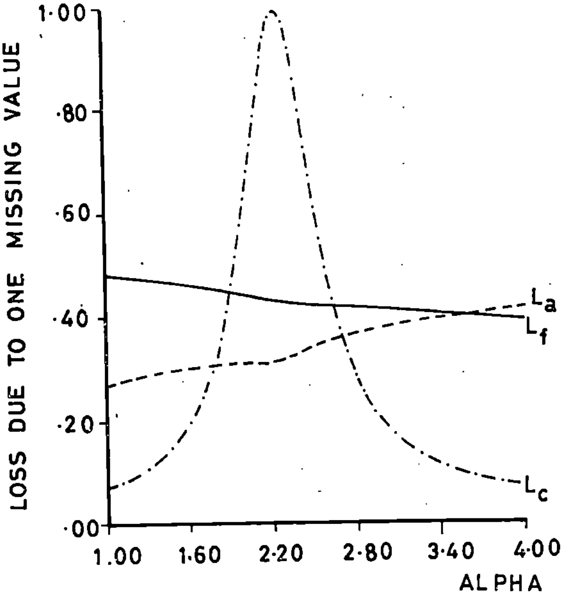
The equation $L_f=L_a$ for this configuration is

$$(80+3n_c)\alpha^6 - 4(456+7n_c)\alpha^4 + 4(3060+17n_c)\alpha^2 - 512(50+3n_c) = 0.$$

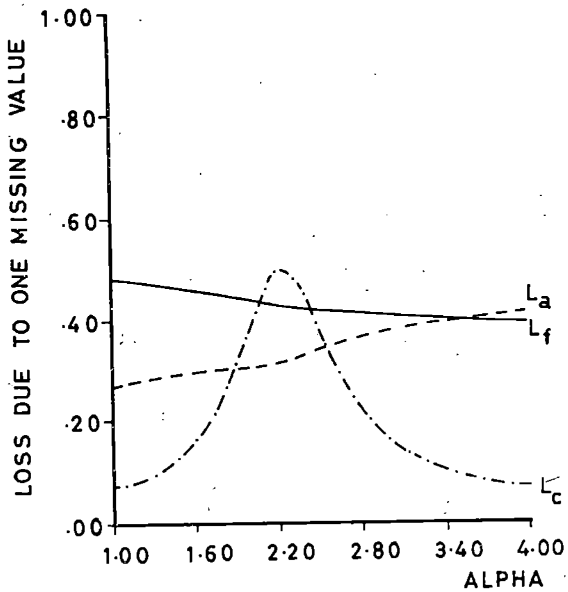
For n_c=2 this equation reduces to

Figure-3 Losses due to a factorial, an axial or a centre observation missing for design with k=5, n_f=32, n_a=20 and n_c=1 or 2, plotted against α.

(a) $n_c = 1$



(b) $n_c = 2$



$$43\,\alpha^6 - 940\,\alpha^4 + 6188\,\alpha^2 - 14336 = 0$$

The only real root for this equation is α^2=12.456 which gives α=3.5293.

The α solution for the equation L_f=L_a for n_c=1, . . . , 10 is shown in Table 3. The losses L_f, L_a and L_c for these pairs of α and n_c are also given. It may be noted from the table that L_c is less than L_f and L_a for all the designs. So the designs with the pair of α and n_c in this table are all minimaxloss designs robust to a single missing observation. Also note that L_f(=L_a) in Table-3 are almost half of the corresponding values in Table 1, in the previous section. If the size of each part of a design in Table 3 is exactly double the size of corresponding parts in a design in Table 1 then the α values for both designs are the same and L_f, L_a and L_c for the design in Table 3 are exactly half of the corresponding losses for the design in Table 1. For example designs with n=54 (n_f = 32, n_a = 20, n_c = 2) and n=27(n_f=16,n_a=10, n_c=1) both have α = 3.52293 and losses for the first design are half the corresponding losses for the second design.

The variances of the parameter estimates for k=5, n_f=32 and n_a=20 may be expressed as

$$\mathrm{Var}(\hat{\beta}_0) = [\, n_c + 32(5-\alpha^2)^2/(40-\alpha^4)\,]^{-1} ,$$

$$\mathrm{Var}(\hat{\beta}_i) = (32 + 4\,\alpha^2)^{-1} ,$$

Table 3. Alpha solution of the equations L_f=L_a and different losses for five factor designs with 1 to 10 centre points and half factorial replicate.

n	n_c	Alpha	L_f	L_a	L_c
27	1	3.5293	0.8008	0.8008	0.1799
28	2	3.4972	0.7957	0.7957	0.1560
29	3	3.4749	0.7919	0.7919	0.1368
30	4	3.4587	0.7890	0.7890	0.1214
31	5	3.4465	0.7867	0.7867	0.1089
32	6	3.4370	0.7849	0.7849	0.0987
33	7	3.4295	0.7834	0.7834	0.0901
34	8	3.4233	0.7822	0.7822	0.0828
35	9	3.4183	0.7812	0.7812	0.0766
36	10	3.4140	0.7803	0.7803	0.0713

$$\text{Var}(\hat{\beta}_{ii}) = \frac{1}{4\alpha^4}\left[1 + \frac{4\alpha^4 + 64\alpha^2 - 160 - 8n_c}{(32+n_c)\alpha^4 - 320\alpha^2 + 800 + 40n_c}\right]$$

and $\text{Var}(\hat{\beta}_{ij}) = \frac{1}{32}$

These variances for one or two centre points are plotted against α in Figure-4 (a) and (b) respectively. These variances are small for minimaxloss design.

Now we compare minimaxloss design with two centre points, with the following five factor designs of the same configuration

i. design with α=1.0,

ii. orthogonal design with α = 1.5467,

iii. rotatable design with α = 2.0,

iv. design with $\alpha = \sqrt{k} = 2.2361$,

v. design with two centre points and minimum variance of losses with α = 2.5998 and

vi. design with one centre point and minimum variance of losses with α = 2.74.

The losses L_f, L_a and L_c, the maximum loss and variance of losses for designs listed above are shown in Table 4 for one or two centre points. The maximum loss for minimaxloss design is 0.4 as compared to 0.4154 for outlier robust design, 0.5 for design with $\alpha = \sqrt{k}$, 0.44 for rotatable design, 0.463 for orthogonal design and 0.4825 for design with α = 1.0.

5. *Five Factor Designs with One Replication of Factorial and Axial Parts.*

Five factor c.c.d. of this configuration consists of n_f=32, n_a=10 and one or two centre points. Total design points are 43 or more.

The α solution of equation L_f=L_a for this design with n_c=2 is α =0.705. This implies that the axial points (or star points) are inside the five dimensional 'cube' formed by the 2^5 factorial points. Because of this these designs are not discussed any more.

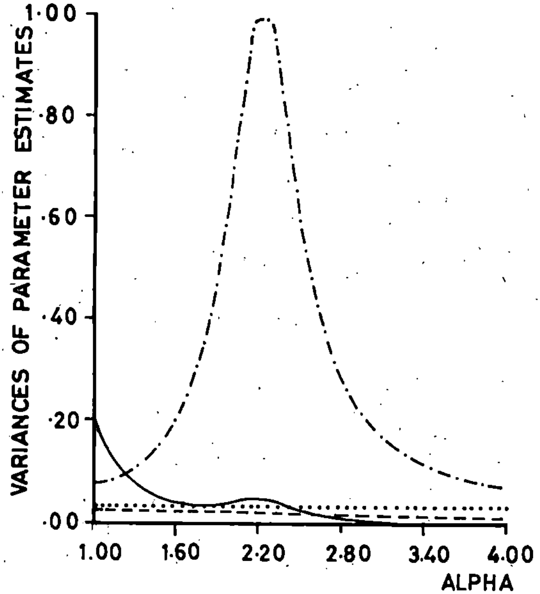
6. *Conclusions*

The minimaxloss five factor central composite designs developed in the above sections are robust to a single missing observations. The variances of the parameter estimates for these designs are also small.

The use of $(|X'X|)^{1/p}$ or $|X'X|(n)^{-p}$ instead of $|X'X|$ also leads to

Figure–4 Variances of parameter estimates for design with k=5, n_f=32, n_a=20 and n_c=1 or 2, plotted against α.

(a) n_c=1



(b) n_c=2

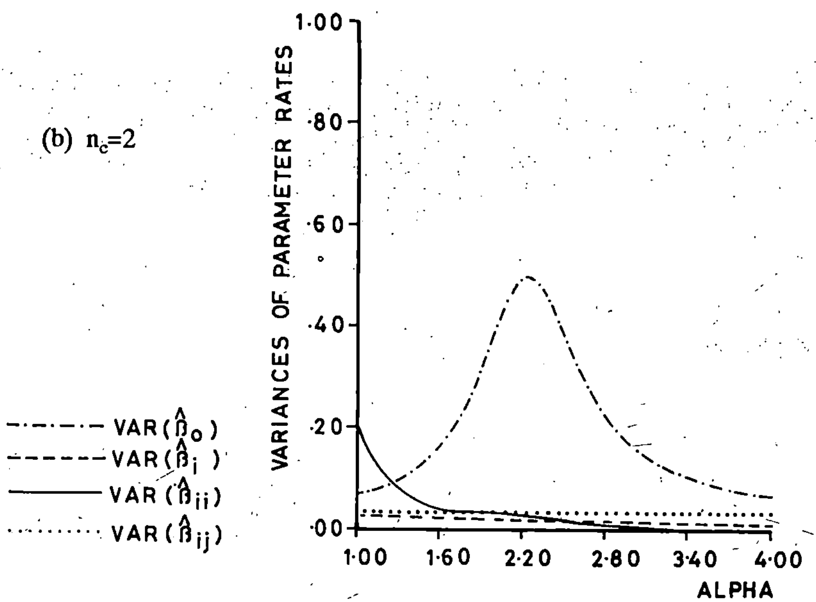


Table 4. Loss due to a single missing observation at factorial, axial or centre point together with the maximum loss and variance of losses.

No. of variables k= 5 Total design points n=54
No. of parameters p =21 No. of centre points =2

Axial observations replicated twice

Alpha	n	|X' X| (complete design)	Loss due to a single missing observation				Minimum |X' X| (Reduced design)	Variance of losses
			Facto rial obs.	Axial obs.	Centre obs.	Maxi mum loss		
1.0000	54	0.4141E+29	0.4825	0.2710	0.0690	0.4825	0.2143E+29	0.1440E–01
	53	0.3855E+29	0.4826	0.2742	0.0741	0.4826	0.1995E+29	0.1231E–01
1.5467	54	0.4744E+32	0.4629	0.2946	0.1479	0.4629	0.2548E+32	0.8850E–02
	53	0.4043E+32	0.4634	0.3000	0.1736	0.4634	0.2169E+32	0.7290E–02
2.0000	54	0.2772E+34	0.4410	0.3056	0.3889	0.4410	0.1550E+34	0.4258E–02
	53	0.1694E+34	0.4422	0.3106	0.6364	0.6364	0.6161E+33	0.5231E–02
2.2361	54	0.2227E+35	0.4279	0.3154	0.5000	0.5000	0.1113E+35	0.3423E–02
	53	0.1113E+35	0.4279	0.3154	1.0000	1.0000	0.0000E+00	0.1014E–01
2.5998	54	0.9736E+36	0.4179	0.3497	0.3169	0.4179	0.5667E+36	0.1283E–02
	53	0.6651E+36	0.4207	0.3537	0.4639	0.4639	0.3565E+36	0.1154E–02
2.7400	54	0.4210E+37	0.4154	0.3609	0.2446	0.4154	0.2461E+37	0.1506E–02
	53	0.3180E+37	0.4184	0.3643	0.3238	0.4184	0.1849E+37	0.7962E–03
3.5293	54	0.5315E+40	0.4004	0.4004	0.0900	0.4004*	0.3187E+40	0.3502E–02
	53	0.4837E+40	0.4024	0.4012	0.0989	0.4024	0.2891E+40	0.1734E–02

* Minimaxloss due to one missing observation.

the same minimaxloss design. These five factor minimaxloss design also minimize the maximum discrepancy caused by any outlying observation. In c.c.d. with some centre points it is not possible to have $L_f = L_a = L_c$ i.e. equal loss due to any single missing observation.

REFERENCES

[1] Akhtar, M., and Prescott, P., 1986. Response surface designs robust to missing observations, *Commun. Statist.-Simula., Comput.*, **15**(2) 345-363.

[2] Andrews, D.F. and Herzberg, A.M., 1979. The robustness and optimality of response surface designs. *J. Statist. Plan. Inf.* **3** , 249-257.

[3] Box, G.E.P., 1952. Multifactor Designs of First Order. *Biometrika* **39**, 49-57.

[4] Box, G.E.P., and Draper, N.R., 1975. Robust Designs. *Biometrika*, **62**, 347-352.

[5] Box, G.E.P. and Hunter, J.S., 1957. Multifactor Response Surface Design for Exploring Response Surfaces. *Ann. Math. Statist.* **28**, 195-241.

[6] Herzberg, A.M. and Andrews, D.F., 1976. Some considerations in the Optimal Design of Experiments in Non-optimal Situations. *J. Roy. Statist. Soc. B*, **38**, 284-289.

[7] Herzberg, A.M. and Andrews, D.F., 1978. The robustness of Chain Block Designs and Coat-of-Mail Designs. *Commun. Statist.- Theory Meth.* **7**, 479-485.

[8] Herzberg, A.M., Prescott, P. and Akhtar, M., 1987. Equi- information Designs : Which Designs are Possible? *Canad. J. Statist.* **15**(1), 71-76.

[9] McKee, B. and Kshirsagar, A.M., 1982. Effect of Missing Plots in some Response Surface Designs. *Commun, Statist.-Theor. Meth.* **14**, 1525-1549.