International Journal of Statistical Sciences *ISSN 1683–5603*
Vol. 6 (Special Issue), 2007, pp 259-268

Optimal Designs in Linear Regression Models With Heteroscedastic Errors

Premadhis Das

Department of Statistics
Kalyani University
Kalyani, Nadia
West Bengal, India

Nripes Kumar Mandal

Department of Statistics
Calcutta University
Kolkata, India

Bikas K. Sinha

Indian Statistical Institute
203, B.T. Road
Kolkata 700 108, India
E-mail : sinhabikas@yahoo.com

[Received October 7, 2005; Revised June 10, 2007; Accepted July 7, 2007]

Abstract

We examine standard linear regression models from the perspective of optimal spacings for the non-stochastic covariate, when the errors are distributed with heteroscedastic variances but zero covariances. We present results for both symmetric and asymmetric factor spaces and for certain special variance structures.

Keywords and Phrases: Linear Regression, Heteroscedastic Errors, Loewner Domination, Optimum Designs.

AMS Classification: 62K05, 62J05.

1 Introduction

Under standard linear regression model

$$Y_x = \beta_0 + \beta_1 x + e_x, \tag{1}$$

with uncorrelated and homoscedastic errors (mean 0 and variance σ^2) and with the factor space $\Xi = [a, b], a < b$, it is well-known that for most efficient estimation of the regression coefficient (β_1), *optimal* spacing corresponds to the *extreme* points viz., a and b. That means, one should collect observations only from the extreme points in the x-scale and that too, with 50 : 50 allocation. We will not pay much attention to the exact or approximate nature of such allocations. In other words, we will mostly deal with the *continuous* design allocation theory. Vide Pukelsheim (1993).
To keep the formulation in its most general form, we start with a continuous k-point design :

$$d_k = [(x_1, p_1), (x_2, p_2), ..., (x_k, p_k)], k \geq 2, a \leq x_1 < x_2 < ... < x_k \leq b. \tag{2}$$

The above result suggests that for estimation of β, any such d_k is dominated by $d_2^* = [(a, 0.50), (b, 0.50)]$ in the sense of providing increased *precision* i.e., smaller variance. Before proceeding further, we note that the Information Matrix for parameter estimation based on d_k has 4 elements given in

$$I(d_k) = ((I_{11}, I_{12} = I_{21}, I_{22})) \tag{3}$$

where

$$I_{11} = 1, I_{12} = \sum p_i x_i, I_{22} = \sum p_i x_i^2. \tag{4}$$

At this stage, it is pertinent to recall the celebrated **de la Garza Phenomenon** which states that in the case of pth degree polynomial regression with uncorrelated homoscedastic errors based on a given set of $k > p+1$ design points, one can provide a set of exactly $p+1$ design points such that *both* the designs provide *identical* information matrices. [de la Garza (1954)].
Applied to the present set-up, we note that corresponding to d_k, there is an "information-equivalent" design d_2 based on exactly 2 points, whenever $k > 2$. We set $d_2 = [(c, p), (d, q); q = 1 - p, 0 < p < 1]$ so that c and d are the support points of d_2. Naturally, the elements of d_2 are derived as functions of those of the given design d_k. It is known that $a \leq c < d \leq b$. Thus, de la Garza Phenomenon suggests that we can confine only to 2-point designs under homoscedastic linear regression. The design d_2^* is the *best* among all such designs for estimation of the β-coefficients.
There is some fact more into it. As an information matrix, apart from information equivalence, there is a possibility of "information dominance"! In the special case : $\Xi = [0, b]$, it is easily argued that corresponding to any 2-point design $d_2 = [(c, p), (d, q); c > 0, d < b, q = 1 - p]$ having information matrix I, there exists a 2-point design based on

the allocation of mass at 0 and at another point r $(d < r \leq b)$ such that the latter information matrix dominates the former in the sense of "Loewner Domination"! In other words, the difference of the matrices [latter minus former] is an nnd matrix. This is the best result one can hope for. The interpretation is that when the factor space is $\Xi = [0, b]$, only designs with some allocation of mass at 0 and the rest at just one other point away from 0 form a "Complete Class" in the sense that any design outside this class can be improved by one such design in terms of Loewner Domination.
We will extend the validity of the above result for an arbitrary finite interval $\Xi = [a, b]$ with $0 < a < b < \infty$.
Such domination results for quadratic regression are established only recently. Vide Pukelsheim (1993) and Liski et al (2002).
Our purpose in this paper is to discuss such domination results only in the case of linear regression with heteroscedastic errors. A preliminary version of this article was presented at Taiwan Design Conference (December 22-24, 2003).
We organize the paper as follows. In Section 2, we establish the general domination result mentioned above. Next in Section 3, we examine the prospect of Loewner Domination in the context of linear regression involving heteroscedastic but uncorrelated errors. Specific optimal designs are characterized in Section 4.

2 Loewner Domination in Homoscedastic Linear Regression

We use the same set-up and notations as above. We establish the following result.
Theorem 1. Given a 2-point design $d_2 = [(c, p), (d, q); c > a, d < b, 0 < p < 1, q = 1 - p]$ over $\Xi = [a, b], 0 < a < b$, there exist $P, 0 < P < 1$ and e , $d < e < b$ such that the 2-point design $d_2^* = [(a, P), (e, Q); Q = 1 - P]$ dominates over the given design in the Loewner sense under homoscedastic linear regression.
Proof. We will follow closely the arguments as in Pukelsheim (1993) and Liski et al (2002). It is readily seen that the $(1, 1)^{th}$ term of I_{d_2} and $I_{d_2}^*$ are each equal to 1. We equate $(1, 2)^{th}$ term of I_{d_2} and $I_{d_2^*}$ and then establish that the $(2, 2)^{th}$ term of I_{d_2} is less than the $(2, 2)^{th}$ term of $I_{d_2^*}$. Set $w = a/b, u = c/b, v = d/b, e/b = s$. Given $0 < w < u < v < 1; p, 0 < p < 1$, we will show existence of $s, v < s < 1$ and $P, 0 < P < 1$ such that both (5) and (6) stated below are simultaneously satisfied [(5) corresponds to $(1, 2)^{th}$ term and (6) corresponds to the $(2, 2)^{th}$ term respectively] :

$$up + vq = wP + sQ, \tag{5}$$

$$u^2p + v^2q < w^2P + s^2Q. \tag{6}$$

Set again $wP = \alpha(up + vq)$ so that $sQ = (1 - \alpha)(up + vq)$. Note that $0 < \alpha < 1$. These yield :

$$P = \alpha(up + vq)/w \tag{7}$$

and, hence,

$$s = [w(1-\alpha)(up+vq)]/[w-\alpha(up+vq)]. \tag{8}$$

Therefore, (6) can be expressed as :
$w^2P + s^2Q = w\alpha(up+vq) + s(1-\alpha)(up+vq) = (up+vq)[w\alpha + s(1-\alpha)] > u^2p + v^2q$
and this is satisfied whenever (9) stated below holds :

$$w\alpha + s(1-\alpha) > B = [u^2p + v^2q]/[up+vq]. \tag{9}$$

Set $A = up + vq$.
Substituting the expression for s from (8) in (9), we obtain
$w\alpha + [w(1-\alpha)^2 A]/[w - A\alpha] > B$
i.e., $[w^2\alpha - w\alpha^2 A + w(1-\alpha)^2 A]/[w - A\alpha] > B$
i.e., $w^2\alpha + A(B-2w)\alpha > w(B-A)$
i.e.,

$$\alpha > [w(B-A)]/[w^2 + A(B-2w)]. \tag{10}$$

Again from the requirement $s < 1$, we obtain, from (8),
$wA(1-\alpha) < w - A\alpha$
i.e.,

$$\alpha < w(1-A)/A(1-w). \tag{11}$$

Note that $A > u > w$ and so $1 - A < 1 - w$, implying thereby that $\alpha < w/A < 1$ and hence, from (7), $P = A\alpha/w < 1$.
It remains to show that the inequalities (10) and (11) are consistent i.e.,
$w(1-A)/A(1-w) > w(B-A)/[w^2 + A(B-2w)]$
i.e., $(1-A)[w^2 + (B-2w)A] > A(1-w)(B-A)$
i.e., $(A-w)^2 + wA^2 - w^2A > AB(A-w)$
i.e., $(A-w) + wA > AB$
i.e.,

$$A(1 + w - B) > w. \tag{12}$$

Substituting the expressions for A and B i.e., $A = up + vq$ and $B = \frac{(u^2p+v^2q)}{up+vq}$, we simplify (12) as
$(1+w)(up+vq) - (u^2p + v^2q) > w$
i.e.,

$$up(1 + w - u) + vq(1 + w - v) > w. \tag{13}$$

Clearly, (13) holds trivially since $u(1+w-u) - w = (u-w)(1-u) > 0$ and $v(1 + w - v) - w = (v-w)(1-v) > 0$.
Thus we have established that there are choices of P and s for which (5) and (6) hold i.e., there is Loewner domination of the given design located at the points c and d by the one located at a and e. There are plenty of choices of e since this is true of α i.e., of P.

Example 1. Take $w = 0.2 < u = 0.3 < v = 0.7; p = 0.3$. Then $A = 0.58, B = 0.64$ (approx.). From (10) and (11), bounds for α are [0.067, 0.181]. For $\alpha = 0.10$, we obtain $P = A\alpha/w = 0.29, s = (1-\alpha)wA/[w - A\alpha] = 0.735$. Again, for $\alpha = 0.15$, we obtain $P = 0.435, s = 0.873$.
The information matrices are given below for the case $b = 1$.
$I_d = ((1,\ 0.58,\ 0.37)); I_{d^*} = ((1,\ 0.58,\ 0.395)); I_{d^{**}} = ((1,\ 0.58,\ 0.4484))$.

3 Loewner Domination in Heteroscedastic Linear Regression

We refer to the linear regression model in (1) but this time we assume that the errors are uncorrelated with heteroscedastic variances. In particular, we assume $V(e_x) = \sigma^2 v(x)$ where $v(x)$ is strictly positive non-constant. Since we are interested in the estimation of the β-coefficients, without any loss of generality, we take $\sigma = 1$. Taking the factor space to be the positive half of the real line including the point 0, Minkin (1993) obtained an explicit characterization of the optimum design under the above set-up for estimation of $1/\beta$, assuming $v(x) = e^x$. See Liski et al (2002) for an alternative derivation of Minkin's result, by exploiting the technique of Loewner domination in the specific heteroscedastic situation.
The following extensions will be made in this article. Continuing as in Liski et al (2002), a complete class result is established for the estimation of the β-coefficients with the following error functions:
(i) $v(x) = k^x, x \geq 0, k \geq 1$; *(ii)* $v(x) = (1+x)^{\gamma+1}, x \geq 0, \gamma \geq 0$.
Note that the elements of the information matrix are now defined in terms of the weighted sums involving the reciprocals of the variances i.e., $I_{11} = \sum p_i/v(x_i); I_{12} = \sum x_i p_i/v(x_i); I_{22} = \sum x_i^2 p_i/v(x_i)$.
We now consider the main result of this section which ensures that, in finding optimum designs, it is sufficient to confine our search in the class of 2-point designs **including** the point 0.
Towards establishing the result stated above, first we start with a 2-point design:

$$d_2 = [(a,p),(b,q); 0 < a < b < \infty; 0 < p, q < 1, p + q = 1]. \tag{14}$$

For this d_2

$$I_{11} = p/v(a) + q/v(b), I_{12} = pa/v(a) + qb/v(b), I_{22} = pa^2/v(a) + qb^2/v(b). \tag{15}$$

Theorem 2. Given d_2 as above, there exists another 2-point design $d_2^* = [(0,s),(c,1-s); c > 0, 0 < s < 1]$ which dominates d_2 in the Loewner Domination sense.
Proof. To prove this theorem, we will proceed as in Liski et al (2002). As before, we equate the terms in the $(1,1)^{th}$ and $(1,2)^{th}$ positions in the two information matrices

corresponding to two designs d_2 and d_2^* and solve for c and s. Next we show that for this choice of c and s, $I_{22}(d_2^*) > I_{22}(d_2)$.
The equations in terms of c and s are given by

$$s/v(0) + (1-s)/v(c) = p/v(a) + q/v(b), \tag{16}$$

$$(1-s)c/v(c) = pa/v(a) + qb/v(b). \tag{17}$$

Eliminating s, we obtain the following equation involving c:

$$\phi(c) = w\phi(a) + (1-w)\phi(b) \tag{18}$$

where

$$\phi(x) = [v(x) - 1]/x; w = pa/v(a)/[pa/v(a) + qb/v(b)]. \tag{19}$$

It is readily seen that the function $\phi(x)$ satisfies the following conditions:

$$\phi(x) \textit{ is convex and increasing in } x \textit{ over } [0, \infty) \tag{20}$$

for both the choices of $v(x)$ given in the statement of the Theorem. Hence we have a unique solution for c in (a, b) satisfying the requirement above. Next, s can now be deter mined from (16) i.e., from

$$1 - s = [p(1 - 1/v(a)) + q(1 - 1/v(b))]/[1 - 1/v(c)].$$

It is easy to see that $s < 1$. Moreover, $s > 0$ if and only if

$$1/v(c) < p/v(a) + q/v(b). \tag{21}$$

We will establish (21) in the Appendix for both the error functions considered above. We are yet to establish $I_{22}(d_2^*) > I_{22}(d_2)$. This, after a little simplification, reduces to establishing the inequality : $c > wa + (1-w)b$. This follows readily from the strict convexity of the function $\phi(x)$ and the defining equation for c. This establishes the Theorem.
Remark 1. In view of the Remark 7.6.1 made in Liski et al (2002), we have the following Corollary.
Corollary 1. Given d_n, an n-point design in Ξ, there exists a 2-point design $d_2 = [(0, s), (c, 1-s); c > 0, 0 < s < 1]$ that dominates d_n in the sense of Loewner Order Domination.
In view of Theorem 2 and Corollary 1 cited above, it follows that in order to find optimum designs we can restrict to 2-point designs with one point as 0. In the following optimum designs are given for different optimality criteria.

4 Specific Optimal Designs

4.1 Variance Structure : $v(x) = k^x, x \geq 0$

For the estimation of slope parameter, we have to maximize $I_{22.1}$, which after a little algebra, reduces in terms of its reciprocal, to

$$I_{22.1}^{-1} = [c^2 k^{-c}]^{-1}[1/(1-s) + k^{-c}/s]. \tag{22}$$

We first minimize this expression in (22) wrt s which yields

$$[c^2 k^{-c}]^{-1}[1 + k^{-c/2}]^2 \tag{23}$$

and the minimum is attained when $s = 1/[1 + k^{c/2}] = s(c)$ (say).
Next we minimize the resulting expression wrt c which results in an implicit equation in c given by : $c^* = 2[k^{-c^*/2} + 1]/log_e k$. This in its turn produces optimum value of s.

The optimum values of (c, s) for different values of k are given below: $(k, c, s) =$ $(1,\ 16.192,\ 0.129); (2,\ 4.9887,\ 0.474) : (e,\ 2.9887,\ 0.4137);$
$(3,\ 2.62731,\ 0.399); (4,\ 1.92952,\ 0.3633); (5,\ 1.59893,\ 0.2454).$

For the D-optimum design we have to maximize the determinant of the information matrix or equivalently $s(1-s)c^2k^{-c}$. Optimum values of s and c are given by $s = 0.5; c = 2/log_e k$.
For A-optimality we have to minimize trace I^{-1} which, after a little algebra, is equivalent to minimizing

$$[c^2 k^{-c}]^{-1}[1/(1-s) + (c^2+1)k^{-c}/s]. \tag{24}$$

Minimization wrt s yields the lower bound $[c^2k^{-c}]^{-1}[1 + (c^2+1)^{1/2}k^{-c/2}]^2$, equality holding at $s = 1/[1 + (c^2+1)^{-1/2}k^{c/2}]$. The resulting bound is further minimized wrt choice of c numerically.
The optimum values of (c, s) for different values of k are given below: $(k, c, s) =$ $(1,\ \infty,\ 1.00); (2,\ 3.17,\ 0.5256); (3,\ 2.07,\ 0.4340);$
$(4,\ 1.67,\ 0.3795); (5,\ 1.46,\ 0.3534); (6,\ 1.32,\ 0.3367);$
$(7,\ 1.23,\ 0.3239); (8,\ 1.15,\ 0.3155); (9,\ 1.09,\ 0.3087); (10,\ 1.05,\ 0.3021).$

4.2 Variance Structure : $v(x) = (1 + x)^{1+\gamma}$

This time again we have to maximize $I_{22.1}$ i.e., minimize $I_{22.1}^{-1}$ which, after a little algebra, reduces to

$$I_{22.1}^{-1} = [c^2(1+c)^{-(1+\gamma)}]^{-1}[1/(1-s) + (1+c)^{-(1+\gamma)}/s] \tag{25}$$

which, for fixed c, has an attainable lower bound given by

$$[c^2(1+c)^{-(1+\gamma)}]^{-1}[1+(1+c)^{-(1+\gamma)}/2]^2, \tag{26}$$

with equality holding at $s = 1/[1+(1+c)^{-(1+\gamma)/2}]$.
The resulting expression is further minimized wrt c numerically. Let us write $\alpha = (1+\gamma)/2$.
The optimum values of c and s for different values of α are given below.

$(\alpha, c, s) = (0.5,\ \infty, 1.00); (1.0,\ 45918.3, 0.99998); (1.5,\ 3.0, 0.8888);$
$(2.0,\ 1.41421 = e, 0.85355); (2.5,\ 0.91703, 0.83587), (3.0,\ 0.677651, 0.82523).$
Proceeding as before, for the D-optimum design, the optimum value of s is again and the maximizing function for the determination of c is $c^2(1+c)^{-(1+\gamma)}$. The optimum values of c for different values of γ are derived as $c_{opt} = 2/(\gamma - 1),\ \gamma > 1; = \infty,\ -1 \leq \gamma \leq 1$.
MV-Optimality : Here we find a design which minimizes the maximum of the two variances. It is known that this occurs at

$$V(b_0) = V(b_1)$$

where b_0 and b_1 denote least squares estimates of β_0 and β_1 respectively and the above variance equality yields

$$1/s_{opt} = 1 + [(1+c)^{(1+\gamma)}]/[(c^2-1)]$$

and c is to be determined by minimizing

$$1 + [(1+c)^{(1+\gamma)}]/[(c^2-1)].$$

Finally, we have

$$c_{opt} = (\gamma+1)/(\gamma-1), \gamma > 1; = \infty, -1 \leq \gamma \leq 1.$$

For A-optimality we have to minimize trace I^{-1} which, after a little algebra, is equivalent to minimizing $A/(1-s) + B/s$ with $A = c^{-2}(1+c)^{(1+\gamma)}$ and $B = c^{-2}(1+c^2)$. The opt. choice of s comes out as $B^{1/2}/[B^{1/2} + A^{1/2}]$. Finally, c is to be determined by minimizing $[(c^2+1)(1+c)^{(1+\gamma)/2}]/c$. The opt. values of c and s for different values of α are given below.
$(\alpha, c, s) = (0.5,\ 0.798,\ 0.4883); (1.0,\ 0.657,\ 0.4193); (1.5,\ 0.552,\ 0.3714);$
$(2.0,\ 0.478,\ 0.3366); (2.5,\ 0.405,\ 0.3156); (3.0,\ 0.352,\ 0.3002).$

Appendix

We will establish (21) for both the error functions (3) one by one.
(i) Error function $v(x) = k^x$.
Note that equation (21) is equivalent to
$pk^{-a} + qk^{-b} \leq k^{-[(apk^{-a}+bqk^{-b}]/[pk^{-a}+qk^{-b}]}$
i.e., $log_k[pk^{-a} + qk^{-b}] \leq -[pak^{-a} + qbk^{-b}]/[pk^{-a} + qk^{-b}]$
which, after a little simplification, can be written as
$log_k[p + qk^t] \leq tqk^t/[p + qk^t]$
i.e., $tqk^t - [p + qk^t]log_k[p + qk^t] \geq 0$.
Write $\delta(t) = tqk^t - [p + qk^t]log_k[p + qk^t]$.
It is easy to see that $\delta(0) = 0$ and
$\delta'(t)$ is proportional to $tlog_e klog_e[p + qk^t] > 0$ for $k > 1$.
And this establishes (21).

(ii) Error function $v(x) = (1 + x)^{1+\gamma}$.
Equation (21) is equivalent to

$$p/(1 + a)^{1+\gamma} + q/(1 + b)^{1+\gamma} \leq 1+$$
$$[pa(1 + a)^{-(1+\gamma)} + qb(1 + b)^{-(1+\gamma)}]/$$
$$[p(1 + a)^{-(1+\gamma)} + q(1 + b)^{-(1+\gamma)}]^{-(1+\gamma)}$$

i.e., $[p(1 + a)^{-\gamma} + q(1 + b)^{-\gamma}]^{(1+\gamma)} \leq [p(1 + a)^{-(1+\gamma)} + q(1 + b)^{-(1+\gamma)}]^{\gamma}$
i.e., $[p(1 + a)^{-\gamma} + q(1 + b)^{-\gamma}]^{1/\gamma} \leq [p(1 + a)^{-(1+\gamma)} + q(1 + b)^{-(1+\gamma)}]^{1/(1+\gamma)}$
which is the so called Liapunovs inequality. Hence the result is established.

References

1. de la Garza (1954). Spacing of information in polynomial regression. Ann. Math. Stat. 25, 123-130.

2. Ford,I., Torsney, B. and Wu, C.F.J. (1992). The use of a canonical form in the construction of locally optimal designs for non-linear problems. Jour. Royal Stat. Soc. Ser. B 54, 569-583.

3. Hedayat, A.S., Yan, B. and Pezzuto, J.M.(1997). Modelling and identifying optimum designs for fitting dose-response curves based on raw optical density data.Jour. Amer. Stat. Assoc. 92, 1132-1140.

4. Liski, E.P., Mandal, N.K., Shah, K.R. and Sinha, B.K. (2002). Topics in optimal design. Springer-Verlag. Lecture Notes Series No. 163.

5. Mathew, T. and Sinha, Bikas K. (2001). Optimal designs for binary data under logistic regression. Jour. Statist. Plann. Inference. 93, 295-307.

6. Minkin, S. (1993). Experimental design for clonogenic assessment in chemotherapy. Jour. Amer. Stat. Assoc. 88, 410-420.

7. Pukelsheim, F. (1993). Optimal design of experiments. Wiley, New York.