Matematika, 2002, Jilid 18, bil. 2, hlm. 109–116


# Robustness of D-optimal Design to Prior Parameter Estimates

**Kassim Bin Haron**
Department of Mathematics
Universiti Putra Malaysia
43400 UPM Serdang, Selangor, Malaysia

**Abstract** This paper discusses the effect of errors in prior estimates on the efficiency of the D-optimal design used for estimating the parameters of a nonlinear regression model. The performance of the local D-optimal designs produced are evaluated by comparing their respective efficiencies of estimation , E (%). It was discovered that the D-optimal design is exceptionally robust to a wide range of initial parameter values and that over-estimating the parameters would consistently produce a more efficient design than under-estimating them by the same quantum.

**Keywords** Prior estimates, Criterion, Generalized Variance, D-optimal, Efficiency

***Abstrak*** *Kertas ini membincangkan kesan ralat dalam anggaran prior ke atas kecekapan rekabentuk D-optimum yang digunakan bagi menganggar parameter suatu model regresi tak linear. Pencapaian rekabentuk D-optimum tempatan yang terhasil dinilai dengan membandingkan kecekapan penganggaran, E(%) masing-masing. Kajian mendapati bahawa rekabentuk D-optimum amat teguh terhadap nilai awal parameter dalam satu julat yang lebar dan terlebih anggar parameter akan menghasilkan secara konsisten rekabentuk yang lebih cekap daripada terkurang anggar dengan 'quantum' yang sama.*

***Katakunci*** *Anggaran prior, Kriteria, Varians Teritlak, D-optimum , Kecekapan*

## 1 Introduction

Consider the univariate nonlinear response surface model

$$y_k = \eta(x_k, \theta) + \epsilon_k, \quad k = 1, 2, \ldots, n \tag{1}$$

where $x$ is an input variable; $\theta$ is a vector of unknown parameters and $\epsilon$ is a random error. It is well known that the design problem for such a model is to obtain an $n$–point design

$\xi$ to estimate some function of the $p$–dimensional parameter vector $\theta$ with high efficiency. In this paper we shall consider a continuous or approximate design $\xi$ , represented by the measure

$$\xi = \left\{ \begin{array}{cccc} x_1 & x_2 & \cdots & x_r \\ \omega_1 & \omega_2 & \cdots & \omega_r \end{array} \right\}$$

where the design points or vectors $x_1, x_2, \ldots, x_r$ are distinct elements of the design space $\chi$, for $r \geq p$ ( Atkinson [1]). The associated weights $\omega_1, \omega_2, \ldots, \omega_r$ are nonnegative real numbers which sum to unity and each $\omega_k = \frac{n_k}{n}$; $k = 1, 2, \ldots, r$, without the restriction of $n_k$ being an integer as proposed by Zocchi and Atkinson [9]. The optimal design problem is therefore consist of the choice of the pairs $(x_k, n_k)$, $k = 1, 2, \ldots, r$ in the design space $\chi$, subjected to a specific criterion.

When the errors associated with the model (1) are assumed to be uncorrelated Gaussian random variables with zero mean and constant variance $\sigma^2$, taken to be equal to one without loss of generality. The Fisher information matrix for the design $\xi$, given the parameter value $\theta$, is represented by

$$M(\xi, \theta) = \sum_{i=1}^{r} w_i \frac{\partial \eta(x_i, \theta)}{\partial \theta} \frac{\partial \eta(x_i, \theta)}{\partial \theta^T} = F^T \Omega F \tag{2}$$

where $F$ is the $r \times p$ Jacobian of $\eta$ with the $i^{th}$ row equal to the gradient $\frac{\partial \eta(x_i, \theta)}{\partial \theta}$ of the response function at the point $x_i$ and $\Omega$ is the diagonal matrix with diagonal elements $\omega_1, \omega_2, \ldots, \omega_r$.

Optimal designs typically minimize some convex function of the inverse Fisher information matrix. Box-Lucas [2] constructed optimal designs for precise estimation of the parameters of nonlinear models by using the D-optimality criterion. The D-optimality criterion seeks to minimize the generalized variance or the determinant of the first order variance-covariance matrix $M^{-1}(\xi, \theta^*)$ where $\theta^*$ is the vector of initial parameter estimates. The experimental designs that minimize the determinant of $M^{-1}(\xi, \theta^*)$ are referred to as continuous local D-optimal designs. Here, "D" stands for the determinant of $M^{-1}(\xi, \theta^*)$ associated with the model and the design is said to be local D-optimal because it is optimum only for the chosen set of initial values, $\theta^*$.

It is common knowledge that the procedures for estimating the parameter of any nonlinear models is not straightforward as for its linear counterpart. In order to progress, one has to provide a set of good and reliable prior estimates $\theta^*$ for the parameters $\theta$ that one is supposed to estimate. As a result, experimental designs produced by optimizing an appropriate criterion will undoubtedly vary with different choices of values for the initial estimates and this, in turns, may affect the precision of the resulting parameter estimates. In this paper, we present both the theory and numerical examples to highlight the effect of uncertainty in prior estimates on the performance of local D-optimal designs for estimating the parameters of a selected nonlinear model.

## 2 The Underlying Model

Throughout our investigation, we shall consider generating local D-optimal designs in which the response is to be modelled by the one-factor inverse quadratic polynomial

$$y_k = \frac{x_k}{\theta_0 + \theta_1 x_k + \theta_2 x_k^2} + \epsilon_k \qquad \begin{array}{l} x \geq 0 \\ \theta_0, \theta_2 > 0 \\ \epsilon \sim N(0, \sigma^2) \end{array} \tag{3}$$

which was first introduced by Nelder [8].

The family of inverse polynomials has been known to be extensively used in biological and agricultural investigations to describe the relationship between some responses to various stimuli (see Cousens [4], Mead [7], Large et al. [6], and Cobby et al. [3] ). The model (1) has a non-symmetric optimum, allowing the response to rise to a maximum and then falls steadily to a zero asymptote. Such an appealing feature makes the quadratic inverse polynomial a more practical representation than the ordinary linear quadratic model.

## 3 Design Efficiency

By taking the first order information matrix $M(\xi, \theta)$, Kassim [5] has shown that continuous local D-optimal designs for estimating the parameters of the quadratic inverse polynomial (1) is given by the measure

$$\xi = \begin{pmatrix} x_1 & x_2 & x_3 \\ \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \end{pmatrix}$$

where

$$x_2 = \sqrt{\frac{\theta_0}{\theta_2}} = r x_1, \quad x_3 = r x_2$$

and the geometric ratio

$$r = f(\beta), \quad \beta = \frac{\theta_1}{\sqrt{\theta_0 \theta_2}}.$$

In other words, the D-optimality criterion gave rise to designs consisting of three distinct geometrically-spaced points each with equal number of replication and the number of distinct design points were found to be equal to the number of parameters $p$ in the model. So using these results, we have for the assumed model, $r = p = 3$.

Let $M(\xi, \theta) = (m_{ij})$, $i = 1, 2, 3$; $j = 1, 2, 3$ then

$$m_{ij} = \sum_{k=1}^{3} \frac{\omega_k x_k^{i+j}}{(\theta_0 + \theta_1 x_k + \theta_2 x_k^2)^4} \quad \text{where} \quad \sum_{k=1}^{3} \omega_k = 1$$

Hence,

$$|M(\xi, \theta)| = (m_{11} m_{22} m_{33} + 2 m_{12} m_{22} m_{23}) - (m_{22}^3 + m_{12}^2 m_{33} + m_{23}^2 m_{11}) \tag{4}$$

Now, by letting $\alpha_k = \theta_0 + \theta_1 x_k + \theta_2 x_k^2$ we have

$$\begin{aligned}
|M(\xi,\theta)| &= \left(\frac{n}{3}\right)^3 \sum_{\substack{i \\ i\neq j\neq k}}^{3}\sum_{j}^{3}\sum_{k}^{3} \frac{x_i^2 x_j^2 x_k^2}{(\alpha_i\alpha_j\alpha_k)^4}\left(\begin{array}{l} x_j^2 x_k^4 + 2x_i x_j^2 x_k^3 - x_i^2 x_i^2 x_k^2 - \\ x_i x_j x_k^4 - x_i^3 x_j^3 \end{array}\right) \\
&= \frac{n^3 x_2^{12} r^4 (r-1)^6 (r+1)^2}{27(\theta_0 r^2 + \theta_1 x_2 r + \theta_2 x_2^2)^4 \alpha_2^4 (\theta_0 + \theta_1 x_2 r + \theta_2 x_2^2 r^2)^4} \\
&= \frac{n^3 x_2^{12} r^4 (r-1)^6 (r+1)^2}{27[(1+r^2)\theta_0 + \theta_1 x_2 r]^8 (2\theta_0 + \theta_1 x_2)^4} \qquad (5)
\end{aligned}$$

Therefore,

$$|M^{-1}(\xi,\theta)| = \frac{27[(1+r^2)\theta_0 + \theta_1 x_2 r]^8 (2\theta_0 + \theta_1 x_2)^4}{n^3 x_2^{12} r^4 (r-1)^6 (r+1)^2} \qquad (6)$$

and

$$|M^{-1}(\xi^*,\theta^*)| = \frac{27[(1+{r^*}^2)\theta_0^* + \theta_1^* x_2 r^*]^8 (2\theta_0^* + \theta_1^* x_2)^4}{n^3 x_2^{12} r^4 (r^*-1)^6 (r^*+1)^2} \qquad (7)$$

where

$$r^* = f(\beta^*), \quad \beta^* = \frac{\theta_1^*}{\sqrt{\theta_0^*\theta_2^*}}.$$

Following Silvey [9], we can compute the efficiency of design $\xi^*$ relative to design $\xi$ for estimating the model parameters by calculating the cubic root of the ratio of the D-optimality criterion for the two designs applied to the true model respectively. Therefore, the D-efficiency of design $\xi^*$ relative to $\xi$ is given by

$$E = \left(\frac{|M^{-1}(\xi,\theta)|}{|M^{-1}(\xi^*,\theta^*)|}\right)^{\frac{1}{3}} \times 100\% \qquad (8)$$

Also if we let let $\delta_i, i = 0,1,2$ such that $|\delta_i| \leq 1$ represents an error in $\theta_i$, then the prior parameter estimates can be written as

$$\theta_i^* = (1 \pm \delta_i)\theta_i.$$

Using these notation, we have the following theorem,

## Theorem

For the one factor quadratic inverse polynomial model given in (1), the effect of varying $\theta_0^*$ and $\theta_2^*$ on the efficiency of estimation is the same regardless of $\theta_1^*$.

## Proof:

The proof is of two parts.

(i) Let $\delta_0 = \delta$ and without loss of generality, $\delta_2 = 0$. Then, from (7) we have

$$E_{\delta_0} = \left|M^{-1}(\xi^*, \theta^*)\right|_{\delta_0} \frac{27 A_0 B_0 C_0}{n^3 D_0} \tag{9}$$

where

$A_0 = \{(1 + r_1^{*2} + \delta)\theta_0 + \theta_1 x_2 r_1^* \sqrt{1+\delta}\}^4$, $B_0 = \{(\delta + 2)\theta_0 + \theta_1 x_2 \sqrt{1+\delta}\}^4$

$C_0 = \{(1 + r_1^{*2} + \delta r_1^{*2})\theta_0 + \theta_1 x_2 r_1^* \sqrt{1+\delta}\}^4$, $D_0 = \{(1+\delta_0)^6 x_2^{12} r_1^{*4} (r_1^* - 1)^6 (r_1^* + 1)^2$

and

$$r_1^* = f(\beta_1^*) \ \text{ and } \ \beta_1^* = \left(\frac{1+\delta_1}{\sqrt{1+\delta}}\right)\beta$$

(ii) Let $\delta_2 = \delta$ and without loss of generality, $\delta_0 = 0$. Again, from (7), we have

$$E_{\delta_2} = \left|M^{-1}(\xi^*, \theta^*)\right|_{\delta_2} \frac{27 A_2 B_2 C_2}{n^3 D_2} \tag{10}$$

where

$A_2 = \{(r_1^{*2} + (1+\delta)^{-1})\theta_0 + (1+\delta)^{-\frac{1}{2}}\theta_1 x_2 r_1^*\}^4 = (1+\delta)^{-4} C_0$

$B_2 = \{(1 + (1+\delta)^{-1})\theta_0 + \theta_1 x_2 \sqrt{1+\delta}\}^4 = (1+\delta)^{-4} B_0$

$C_2 = \{(1 + (1+\delta)^{-1} r_1^{*2})\theta_0 + (1+\delta)^{-\frac{1}{2}}\theta_1 x_2 r_1^*\}^4 = (1+\delta)^{-4} A_0$

$D_2 = (1+\delta_0)^{-6} x_2^{12} r_1^{*4} (r_1^* - 1)^6 (r_1^* + 1)^2 = (1+\delta)^{-12} D_0$

Hence, from (9) and (10), it follows that

$$E_{\delta_2} = |M^{-1}(\xi^*, \theta^*)|_{\delta_2} = \frac{27 A_2 B_2 C_2}{n^3 D_2} = \frac{27 A_0 B_0 C_0}{n^3 D_0} = |M^{-1}(\xi^*, \theta^*)|_{\delta_0} = E_{\delta_0}$$

Thus, by varying $\delta_i$ accordingly, we shall be able to have a set of prior estimates for parameter $\theta_i$ and whence generate the required local D-optimal designs. The D-efficiencies of the resulting designs will be investigated and compared.

# 4 Results and Discussion

By carrying out successive numerical minimizations of the D-optimality criterion on the computer, we obtained the required local optimal designs. Table 1(a,b) show the corresponding local D-optimal designs generated for various values of $\theta^* = (\theta_0^*, \theta_1^*, \theta_2^*)$ while Table 2 shows the corresponding D-efficiencies, E(%) calculated by varying $\theta^*$ when the true parameter values are assumed to be $\theta = (4, 1, 1)$.

It is clear from Table (1) and Table (2) that the D-optimal designs obtained are quite robust to errors in the initial choices of the parameter values and this is especially true for $\theta_1$. Intuitively, this phenomenon can be explained by the fact that $\theta_1$ is known to be the least important parameter in the model since its main function is only to jointly govern the height of the response surface.

Table 1: Local D-optimal designs constructed by varying $\theta_0^*$ and $\theta_2^*$ when the true values are $\theta = (4, 1, 1)$

| $\theta^*$ | $\xi$ | $\theta^*$ | $\xi$ |
|---|---|---|---|
| (2,1,1) | $\left\{\begin{array}{ccc} 0.387 & 1.414 & 5.164 \\ 0.333 & 0.333 & 0.333 \end{array}\right\}$ | (4,1,0.5) | $\left\{\begin{array}{ccc} 0.775 & 2.282 & 10.328 \\ 0.333 & 0.333 & 0.333 \end{array}\right\}$ |
| (2.8,1,1) | $\left\{\begin{array}{ccc} 0.470 & 1.673 & 5.951 \\ 0.333 & 0.333 & 0.333 \end{array}\right\}$ | (4,1,0.7) | $\left\{\begin{array}{ccc} 0.672 & 2.390 & 8.505 \\ 0.333 & 0.333 & 0.333 \end{array}\right\}$ |
| (3.2,1,1) | $\left\{\begin{array}{ccc} 0.508 & 1.789 & 6.303 \\ 0.333 & 0.333 & 0.333 \end{array}\right\}$ | (4,1,0.8) | $\left\{\begin{array}{ccc} 0.635 & 2.236 & 7.878 \\ 0.333 & 0.333 & 0.333 \end{array}\right\}$ |
| (3.6,1,1) | $\left\{\begin{array}{ccc} 0.543 & 1.897 & 6.633 \\ 0.333 & 0.333 & 0.333 \end{array}\right\}$ | (4,1,0.9) | $\left\{\begin{array}{ccc} 0.603 & 2.108 & 7.370 \\ 0.333 & 0.333 & 0.333 \end{array}\right\}$ |
| (4.0,1,1) | $\left\{\begin{array}{ccc} 0.576 & 2.000 & 6.945 \\ 0.333 & 0.333 & 0.333 \end{array}\right\}$ | (4,1,1.0) | $\left\{\begin{array}{ccc} 0.576 & 2.000 & 6.945 \\ 0.333 & 0.333 & 0.333 \end{array}\right\}$ |
| (4.4,1,1) | $\left\{\begin{array}{ccc} 0.608 & 2.098 & 7.242 \\ 0.333 & 0.333 & 0.333 \end{array}\right\}$ | (4,1,1.1) | $\left\{\begin{array}{ccc} 0.552 & 1.907 & 6.853 \\ 0.333 & 0.333 & 0.333 \end{array}\right\}$ |
| (4.8,1,1) | $\left\{\begin{array}{ccc} 0.638 & 2.191 & 7.525 \\ 0.333 & 0.333 & 0.333 \end{array}\right\}$ | (4,1,1.2) | $\left\{\begin{array}{ccc} 0.532 & 1.826 & 6.271 \\ 0.333 & 0.333 & 0.333 \end{array}\right\}$ |
| (5.2,1,1) | $\left\{\begin{array}{ccc} 0.667 & 2.280 & 7.798 \\ 0.333 & 0.333 & 0.333 \end{array}\right\}$ | (4,1,1.3) | $\left\{\begin{array}{ccc} 0.513 & 1.754 & 5.998 \\ 0.333 & 0.333 & 0.333 \end{array}\right\}$ |
| (6.0,1,1) | $\left\{\begin{array}{ccc} 0.722 & 2.449 & 8.312 \\ 0.333 & 0.333 & 0.333 \end{array}\right\}$ | (4,1,1.5) | $\left\{\begin{array}{ccc} 0.481 & 1.633 & 5.541 \\ 0.333 & 0.333 & 0.333 \end{array}\right\}$ |

Table 2: Local D-optimal designs constructed by varying $\theta_1^*$ when the true values are $\theta = (4, 1, 1)$

| $\theta^*$ | $\xi$ | $\theta^*$ | $\xi$ |
|---|---|---|---|
| (4,0.5,1) | $\left\{\begin{array}{ccc} 0.775 & 2.828 & 10.328 \\ 0.333 & 0.333 & 0.333 \end{array}\right\}$ | (4,1.1,1) | $\left\{\begin{array}{ccc} 0.552 & 1.907 & 6.853 \\ 0.333 & 0.333 & 0.333 \end{array}\right\}$ |
| (4,0.7,1) | $\left\{\begin{array}{ccc} 0.672 & 2.390 & 8.505 \\ 0.333 & 0.333 & 0.333 \end{array}\right\}$ | (4,1.2,1) | $\left\{\begin{array}{ccc} 0.552 & 1.826 & 6.271 \\ 0.333 & 0.333 & 0.333 \end{array}\right\}$ |
| (4,0.8,1) | $\left\{\begin{array}{ccc} 0.635 & 2.236 & 7.878 \\ 0.333 & 0.333 & 0.333 \end{array}\right\}$ | (4,1.3,1) | $\left\{\begin{array}{ccc} 0.513 & 1.754 & 5.998 \\ 0.333 & 0.333 & 0.333 \end{array}\right\}$ |
| (4,0.9,1) | $\left\{\begin{array}{ccc} 0.603 & 2.108 & 7.370 \\ 0.333 & 0.333 & 0.333 \end{array}\right\}$ | (4,1.5,1) | $\left\{\begin{array}{ccc} 0.481 & 1.633 & 5.541 \\ 0.333 & 0.333 & 0.333 \end{array}\right\}$ |
| (4,1.0,1) | $\left\{\begin{array}{ccc} 0.576 & 2.000 & 6.945 \\ 0.333 & 0.333 & 0.333 \end{array}\right\}$ | | |

Table 3: Efficiencies (%) of the local D-optimal designs constructed by varying $\theta^*$ when the true values are $\theta = (4, 1, 1)$

| $\delta_i(\%)$ | $\theta^*$ | $E_{\delta_0}$ | $\theta^*$ | $E_{\delta_1}$ | $\theta^*$ | $E_{\delta_2}$ |
|---|---|---|---|---|---|---|
| $-50$ | $(2.0, 1, 1)$ | 70.86 | $(4, 0.5, 1)$ | 98.84 | $(4, 1, 0.5)$ | 70.86 |
| $-30$ | $(2.8, 1, 1)$ | 91.17 | $(4, 0.7, 1)$ | 99.60 | $(4, 1, 0.7)$ | 91.17 |
| $-20$ | $(3.2, 1, 1)$ | 96.43 | $(4, 0.8, 1)$ | 99.83 | $(4, 1, 0.8)$ | 96.43 |
| $-10$ | $(3.6, 1, 1)$ | 99.19 | $(4, 0.9, 1)$ | 99.96 | $(4, 1, 0.9)$ | 99.19 |
| 0 | $(4.0, 1, 1)$ | 100.00 | $(4, 1.0, 1)$ | 100.00 | $(4, 1, 1.0)$ | 100.00 |
| 10 | $(4.4, 1, 1)$ | 99.34 | $(4, 1.1, 1)$ | 99.96 | $(4, 1, 1.1)$ | 99.34 |
| 20 | $(4.8, 1, 1)$ | 97.58 | $(4, 1.2, 1)$ | 99.83 | $(4, 1, 1.2)$ | 97.58 |
| 30 | $(5.2, 1.1)$ | 95.04 | $(4, 1.3, 1)$ | 99.62 | $(4, 1, 1.3)$ | 95.04 |
| 50 | $(6.0, 1, 1)$ | 88.53 | $(4, 1.5, 1)$ | 99.00 | $(4, 1, 1.5)$ | 88.53 |

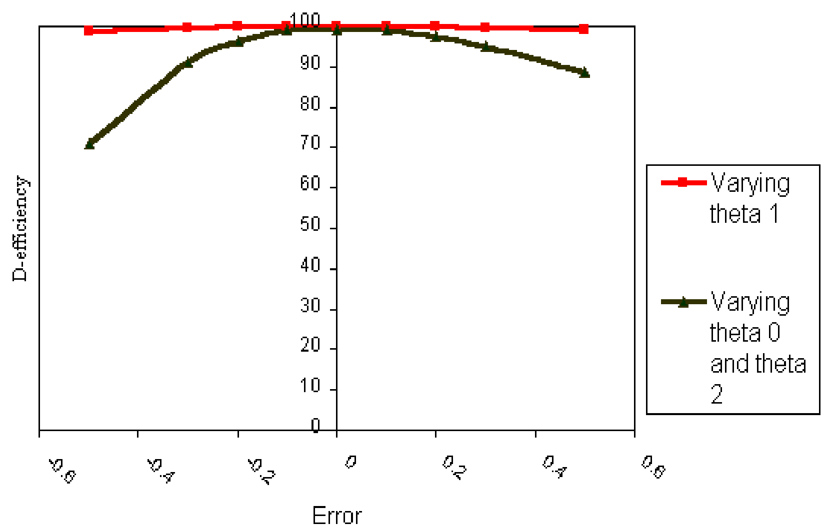


Figure 1: Plot of Efficiency (%) Against Error.

Table 3, on the other hand, reveals the fact that errors in prior estimates of both $\theta_0$ and $\theta_2$ for the model have equal effect on the resulting design efficiencies and this is in total agreement with the theorem proven above. A close examination of the entries in Table 3 and also the plot in Fig.1 clearly show that over-estimating the parameter values would consistently produce designs with higher D-efficiencies than under-estimating them by the same quantum.

## 5 Conclusion

Although D-optimal designs were found to be exceptionally robust to a wide range of initial parameter estimates, one however should pursue a good and reliable set of initial estimates as one possibly can. As far as the quadratic inverse polynomial model is concerned, acquiring good prior estimates is possible if one knows at the outset the stimulus level at which the maximum response lies as well as the maximum value and the slope of the response curve at zero stimulus. Undoubtedly, such prior information can be utilized to obtain good prior estimates for the unknown model parameters.

## References

[ **1** ] A.C. Atkinson & A.N. Donev, *Optimum experimental designs*, Oxford University Press, Oxford, 1992

[ **2** ] G. E.P. Box & H.L. Lucas, *Design of experiments in nonlinear situations.* Biometrika, **46** (1959), 77-90

[ **3** ] J.M. Cobby et al., *Design of experiments for estimating inverse quadratic polynomial responses*, Biometrics, **42** (1986), 659-664

[ **4** ] R. Cousens, *A simple model relating yield loss to weed density*, Annals of Applied Biology, **107** (1985), 239-252

[ **5** ] H. Kassim, *Optimal design spacing for the estimation of the quadratic inverse polynomial parameters*, Pertanika, **12(2)** (1989), 239-244

[ **6** ] R.V. Large et al., *The design and conduct of experiments to measure animal and herbage production responses to fertilizer nitrogen under cutting and grazing managements*, Journal of Agricultural Science ( Cambridge), **104** (1985), 85-94

[ **7** ] R. Mead, *Plant density and crop yield*, Applied Statistics, **19** (1970), 64-81

[ **8** ] J. A. Nelder , *Inverse Polynomials, a useful group of multifactor response functions*, Biometrics, **22** (1966), 128-141

[ **9** ] S.D. Silvey, *Optimal design - an introduction to the theory for parameter estimation*, Chapman Hall, London, 1980

[**10** ] S.S. Zocchi and A.C.Atkinson, *Optimum Experimental Designs for Multinomial Logistic Models*, Biometrics, **55** (1999), 437-444