D- and V-optimal population designs for the quadratic regression model with a random intercept term

Legesse Kassa Debusho[a],*, Linda M. Haines[b]

[a] *Department of Statistics, University of Pretoria, Information Technology Building, Pretoria 0002, South Africa* [b] *Department of Statistical Sciences, University of Cape Town, Rondebosch 7701, South Africa*

Abstract

In this paper D- and V-optimal population designs for the quadratic regression model with a random intercept term and with values of the explanatory variable taken from a set of equally spaced, non-repeated time points are considered. D-optimal population designs based on single-point individual designs were readily found but the derivation of explicit expressions for designs based on two-point individual designs was not straightforward and was complicated by the fact that the designs now depend on ratio of the variance components. Further algebraic results pertaining to d-point D-optimal population designs where dZ3 and to V-optimal population designs proved elusive. The requisite designs can be calculated by careful programming and this is illustrated by means of a simple example.

Introduction

This paper is concerned with optimal population designs for the quadratic regression model with a random intercept term and with observations constrained to be taken at equally spaced time points. The essential problem is that of choosing the numbers of individuals to be allocated to various groups or cohorts and of choosing the times for taking measurements on the individuals within each group. The article is a follow up to the paper by Debusho and Haines (2008) in which *V*- and D-optimal population designs for the simple linear regression model with a random intercept term were discussed.

The construction of optimal population designs for longitudinal models has been extensively studied in the design literature in various contexts. In particular Cheng (1995) and Atkins and Cheng (1999) presented results on optimal population designs for the quadratic regression model with a random intercept over the design space [-1,1]. Their results cannot, however, be directly translated to design spaces comprising a finite number of time points, especially when the number of such points is small. In addition Abt et al. (1998) studied optimal designs for the precise estimation of the linear and quadratic coefficients and for growth prediction in the quadratic regression model with a random intercept numerically. However, they only considered a limited number of individual designs on which to base the population designs and did not apply the Equivalence Theorem to check the global optimality or otherwise of the designs.

The aim of the present paper is to develop explicit expressions for optimal designs for the quadratic regression model with a random intercept term and with values of the explanatory variable taken from a set of equally spaced time points. Attention is restricted to two criteria, D-optimality for which the generalized variance of the estimates of the fixed effects

* Corresponding author. *E-mail addresses:* legesse.debusho@up.ac.za (L.K. Debusho),
linda.haines@uct.ac.za (L.M. Haines).

is minimized and *V*-optimality for which the average variance of estimates of the mean marginal responses over a given set of time points is minimized. The model, some basic ideas and notation, and an appropriate equivalence theorem are introduced in Section 2. The construction of one- and two-point *D*-optimal population designs for the model of interest, based on individual designs for which the time points are not repeated, is discussed in Section 3. In Section 4 some further results, in particular relating to *D*-optimal population designs over the space of all individual designs and to *V*-optimal population designs, are discussed. An illustrative example taken from an experiment on the foraging times of honeybees is presented in Section 5 and some broad conclusions are given in Section 6. The notation and terminology introduced in Debusho and Haines (2008) will be used throughout. The proofs of the theorems presented in the paper are intricate and algebraically tedious and are therefore relegated to a technical report (Debusho and Haines, 2010).

2. Preliminaries

Suppose that a study comprises K individuals and that d_i observations are taken on the ith individual, $i = 1, \ldots, K$. Then, within this context, the quadratic regression model with a random intercept term can be expressed as

$$y_{ij} = \beta_0 + \beta_1 t_{ij} + \beta_2 t_{ij}^2 + b_i + e_{ij}, \quad j = 1, \ldots, d_i, \ i = 1, \ldots, K, \tag{2.1}$$

where y_{ij} is the jth observation on the ith individual at time point t_{ij}, β_0, β_1 and β_2 are fixed effects, b_i is a random effect associated with the ith individual and e_{ij} represents a measurement error. The random terms are taken to be distributed as $b_i \sim \mathcal{N}(0, \sigma_b^2)$ and $e_{ij} \sim \mathcal{N}(0, \sigma_e^2)$ with b_i and e_{ij} independent both within and between individuals and with σ_b^2 and σ_e^2 comprising the variance components. Furthermore, in this study, the time points are assumed to be taken without replacement from a set of equally spaced values, represented for clarity and without loss of generality as $\{0, 1, \ldots, k\}$. Note that the variance structure of model (2.1) is invariant to linear transformations of the explanatory variables since only the intercept term of that model is random (Longford, 1993, Section 4.2.2). Note also that this invariance extends to design criteria based on the information matrix.

Consider an individual design for model (2.1) which comprises non-repeated time points. Then the d-point design $t = (t_1, \ldots, t_d)$, with $t_j \in \{0, 1, \ldots, k\}$ and $0 \le t_1 < t_2 < \cdots < t_d \le k$, which puts equal weight on each point is termed a d-point individual design. The space of all such designs can thus be defined as the set

$$S_{d,k} = \{t : t = (t_1, t_2, \ldots, t_d), t_j \in \{0, 1, \ldots, k\}, j = 1, \ldots, d, \ 0 \le t_1 < t_2 < \cdots < t_d \le k\}.$$

and comprises $N_d = \binom{k+1}{d}$ designs. The information matrix for the fixed effects $\beta = (\beta_0, \beta_1, \beta_2)$ at the vector of time points $t = (t_1, \ldots, t_d)$ is readily derived as a special case of a more general result presented in Schwabe and Schmelter (2008) and is given by

$$I_\beta(t) = \frac{1}{\sigma_e^2(1 + d\gamma)} \begin{pmatrix} d & \sum_{j=1}^{d} t_j & \sum_{j=1}^{d} t_j^2 \\ \sum_{j=1}^{d} t_j & A_1 & A_2 \\ \sum_{j=1}^{d_i} t_j^2 & A_2 & A_3 \end{pmatrix}, \tag{2.2}$$

where

$$A_1 = (1 + d\gamma) \sum_{j=1}^{d} t_j^2 - \gamma \left(\sum_{j=1}^{d} t_j \right)^2,$$

$$A_2 = (1 + d\gamma) \sum_{j=1}^{d} t_j^3 - \gamma \left(\sum_{j=1}^{d} t_j \right) \left(\sum_{j=1}^{d} t_j^2 \right)$$

and

$$A_3 = (1 + d\gamma) \sum_{j=1}^{d} t_j^4 - \gamma \left(\sum_{j=1}^{d} t_j^2 \right)^2,$$

with $\gamma = \sigma_b^2/\sigma_e^2$. The information matrix for β at a population design comprising K individuals with associated sets of time points t_i, $i = 1, \ldots, K$, is thus given by $\sum_{i=1}^{K} I_\beta(t_i)$ and corresponds, at least approximately, to the inverse of the variance matrix of the ML and the REML estimates of β (Verbeke and Molenberghs, 2000, p. 64).

Consider now a population design comprising r distinct individual designs with n_i individuals allocated to the design with time points $t_i = (t_{i1}, \ldots, t_{id_i})$ for $i = 1, \ldots, r$. Suppose further that the cost incurred in taking a single observation is constant and that no extra costs are incurred on recruiting the $\sum_{i=1}^{r} n_i$ individuals to the study. Then the information matrix for β at this population design on a per observation basis is given by

$$\frac{1}{N} \sum_{i=1}^{r} n_i I_\beta(t_i) = \sum_{i=1}^{r} \frac{n_i d_i}{N} M_\beta(t_i),$$

where $N = \sum_{i=1}^{r} n_i d_i$ is the total number of observations taken and $M_\beta(t_i) = 1/d_i I_\beta(t_i)$ is the standardized information matrix at the individual design t_i, $i=1,\ldots,r$. Now consider relaxing the condition that n_i be an integer and introducing the approximate population design

$$\xi = \begin{cases} t_1, \ldots, t_r, \\ w_1, \ldots, w_r, \end{cases}$$

with w_i replacing $n_i d_i / N$ and thus with $0 < w_i < 1$ and $\sum_{i=1}^{r} w_i = 1$. Then the weight w_i represents the proportion of the total number of observations taken at the individual design t_i and the information matrix for β at the population design ξ is given by

$$M_\beta(\xi) = \sum_{i=1}^{r} w_i M_\beta(t_i).$$

Note that if the individual designs within the population design comprise the same number of time points, that is $d_i = d$, then the proportion of individuals allocated to the design t_i is equal to the weight w_i and that otherwise this proportion can immediately be recovered as $v_i = w_i/d_i / \sum_{i=1}^{r} w_i/d_i$ for $i=1,\ldots,r$.

Interest in the present study centres in particular on the construction of D-optimal population designs for the fixed effects β in model (2.1). Specifically the D-optimal criterion is defined in the usual way as

$$\Psi_D(\xi) = -\ln|M_\beta(\xi)| = -\ln\left|\sum_{i=1}^{r} w_i M_\beta(t_i)\right|$$

and clearly depends on the variance components σ_b^2 and σ_e^2 through their ratio $\gamma = \sigma_b^2/\sigma_e^2$. In order to accommodate this dependence, designs which are locally optimal in the sense of Chernoff (1953), that is designs which are based on a "best guess" for the ratio γ, are considered. The General Equivalence Theorem relating to approximate D-optimal population designs follows from the results presented in Debusho and Haines (2008) and, more fundamentally, is a special case of the Equivalence Theorem for multivariate design settings given in Fedorov (1972, p. 212). The theorem is based on the notion of individual designs t taken from a space of designs T and on the fact that the directional derivative of $\Psi_D(\xi) = -\ln|M_\beta(\xi)|$ at ξ in the direction of an individual design t is given by

$$\phi(t,\xi) = 3 - tr\{M_\beta(\xi)^{-1} M_\beta(t)\}$$

and is stated without proof as follows.

Theorem 2.1. *For the random intercept model* (2.1) *and individual designs taken from a space of designs T, the following three conditions on the D-optimal population design ξ^* are equivalent:*

(1) *The design ξ^* minimizes $-\ln|M_\beta(\xi)|$.*
(2) *The design ξ^* minimizes $\max_{t\in T} tr\{M_\beta^{-1}(\xi)M_\beta(t)\}$.*
(3) *The directional derivative $\phi(t,\xi^*)$ attains its minimum at the support designs of ξ^* and $\max_{t\in T} tr\{M_\beta^{-1}(\xi^*)M_\beta(t)\} = 3$.*

The theorem is important in that it can be invoked to confirm the global optimality or otherwise of candidate D-optimal population designs and also in that it forms the basis for algorithmic design construction. Finally note that the D-efficiency of a design ξ_1 relative to a design ξ_2 is given by $\{|M_\beta(\xi_1)|/|M_\beta(\xi_2)|\}^{1/3}$ (Atkinson et al., 2007, p. 151).

3. D-optimal population designs based on individual designs with non-repeated time points

3.1. Individual one-point designs

Consider first population designs comprising a set of individual one-point designs. Then it follows immediately that in this particular case the random intercept model (2.1) reduces to the quadratic regression model

$$y_i = \beta_0 + \beta_1 t_i + \beta_2 t_i^2 + e_i, \quad i = 1, \ldots, K, \tag{3.3}$$

with $t_i \in \{0,\ldots,k\}$ and $e_i \sim N(0,\sigma_e^2 + \sigma_b^2)$ and with σ_e^2 and σ_b^2 therefore not identifiable. Thus the D-optimal population designs for model (2.1) coincide with the D-optimal designs for model (3.3). In particular it follows from Atkinson et al. (2007, p. 123) that for k even the D-optimal population design is given by

$$\xi_{D_e}^* = \begin{cases} (0) & \left(\dfrac{k}{2}\right) & (k) \\ \dfrac{1}{3} & \dfrac{1}{3} & \dfrac{1}{3} \end{cases}.$$

No analogous designs for the case of k odd appear to have been reported in the literature, however, and these designs are presented for completeness in the following theorem.

Theorem 3.1. *Consider the random intercept model* (2.1) *and the set of all one-point designs $S_{1,k}$ taken from $\{0,1,\ldots,k\}$ where k is an odd integer greater than or equal to 3 or, equivalently, the quadratic regression model* (3.3) *with time points taken from the set $\{0,1,\ldots,k\}$. Then the D-optimal population design for the fixed effects β over the space of all one-point designs and for all values of $\gamma \geq 0$ is given by*

$$
\xi^*_{D_o} = \left\{
\begin{array}{cccc}
(0) & \left(\dfrac{k-1}{2}\right) & \left(\dfrac{k+1}{2}\right) & (k) \\[2mm]
w & \dfrac{1}{2}-w & \dfrac{1}{2}-w & w
\end{array}
\right\},
$$

where $w = k^2-2+\sqrt{k^4-k^2+1}/6(k^2-1)$.

Certain points relating to the results of Theorem 3.1 are of interest. Specifically, it can be shown by some straightforward algebra that the weight w increases monotonically from 0.3238 at $k=3$ to $\frac{1}{3}$ as k approaches infinity and is thus, in effect, close to $\frac{1}{3}$ for all values of k. Furthermore, if the discrete time points $0,1,\ldots,k$ are scaled to lie between -1 and 1, then the D-optimal population design $\xi^*_{D_o}$ is given by

$$
\tilde{\xi}^*_{D_o} = \left\{
\begin{array}{cccc}
(-1) & \left(-\dfrac{1}{k}\right) & \left(\dfrac{1}{k}\right) & (1) \\[2mm]
w & \dfrac{1}{2}-w & \dfrac{1}{2}-w & w
\end{array}
\right\}
$$

and, as $k \to \infty$, approaches the design

$$
\tilde{\xi}^*_{A} = \left\{
\begin{array}{ccc}
(-1) & (0) & (1) \\[2mm]
\dfrac{1}{3} & \dfrac{1}{3} & \dfrac{1}{3}
\end{array}
\right\},
$$

which coincides with the approximate D-optimal design for model (3.3) based on the design space $[-1, 1]$. Note that the D-efficiency of $\tilde{\xi}^*_{D_o}$ relative to $\tilde{\xi}^*_{A}$ is given by the expression

$$
\left\{ \frac{(1+k^2+\sqrt{1-k^2+k^4})(1-4k^2+k^4+(1+k^2)\sqrt{1-k^2+k^4})}{4k^6} \right\}^{1/3}
$$

and is high, increasing monotonically as k increases from a value of 94.20% for $k=3$.

3.2. Individual two-point designs

Consider now D-optimal population designs based on the space of two-point designs $S_{2,k}$ for the precise estimation of the parameters β in the random intercept model (2.1). These designs can be constructed for specified ranges of the ratio γ and are presented in the following suite of theorems. Results presented here are for k even.

Theorem 3.2. *Consider the random intercept model* (2.1) *and the set of all individual two-point designs $S_{2,k}$ with k an even integer greater than or equal to 2. Then the D-optimal population design for the fixed effects β over the space of designs $S_{2,k}$ is given by*

$$
\xi^*_{D_0} = \left\{
\begin{array}{ccc}
\left(0,\dfrac{k}{2}\right) & (0,k) & \left(\dfrac{k}{2},k\right) \\[2mm]
\dfrac{1}{3} & \dfrac{1}{3} & \dfrac{1}{3}
\end{array}
\right\}
$$

provided

(i) $k=2$ *or* $k=4$ *and* $\gamma \geq 0$ *and*
(ii) $k \geq 6$ *and* $0 \leq \gamma \leq \gamma_c$ *where* $\gamma_c = 3(k+2)/(k^2-3k-6)$.

Note that the optimal design $\xi^*_{D_0}$ presented in this theorem has an intuitively appealing form and provides a valuable benchmark in that it is optimal for $\gamma = 0$, that is for the case of no correlation between observations within individuals. Note also that the upper bound γ_c decreases to 0 monotonically as k increases. This result can be explained by observing that as k increases so there are many more designs $t=(t_1,t_2)$ for which the condition $\phi(t,\xi^*_{D_0}) \geq 0$ must be satisfied and thus that the degree of correlation within individuals, as reflected in γ, becomes increasingly important.

Theorem 3.3. *Consider the random intercept model (2.1) and the set of all individual two-point designs in $S_{2,k}$ with k an even integer greater than or equal to 6. Let γ_{0k} and γ_{1k} denote the only positive roots of the cubic equations*

$$(k^2-3k-6)(k^2+9k+2)\gamma^3+(k^4-61k^2-116k-52)\gamma^2-(k+2)(2k^2+21k+26)\gamma-3(k+2)^2=0 \tag{3.4}$$

and

$$(k^2-8k-4)(k^2+16k+12)\gamma^3+(k^4-4k^3-192k^2-368k-144)\gamma^2 \quad 4(k^3+21k^2+64k+36)\gamma-12(k+2)^2=0, \tag{3.5}$$

respectively. Then the D-optimal population design for the fixed effects β over the space of designs $S_{2,k}$ is given by

$$\xi^*_{D_1}=\left\{\begin{matrix}\left(0,\dfrac{k}{2}+1\right) & (0,k) & \left(\dfrac{k}{2}-1,k\right)\\ w & 1-2w & w\end{matrix}\right\},$$

where $w=B-\sqrt{A}/3(k-2)\{(k+2)+\gamma(3k+2)\}$ with

$$A=(k-2)^2(3k+2)^2\gamma^4+2(k-2)(3k+2)(3k^2-4k-8)\gamma^3+(15k^4-28k^3-60k^2+96k+96)\gamma^2$$
$$+2(k+2)(3k^3-8k^2+16)\gamma+(k^4-4k^2+16)$$

and

$$B=(k-2)(3k+2)\gamma^2+2(3k^2-2k-4)\gamma+2(k^2-2)$$

provided

 (i) *k=6, 8 and $\gamma\geq\gamma_{0k}$ and*
(ii) *$k\geq 10$ and $\gamma_{0k}\leq\gamma\leq\gamma_{1k}$.*

A more stringent result than that of Theorem 3.3 was in fact obtained for $k=10, 12, 14, 16, 18$ and 20 and is summarized as follows.

Theorem 3.4. *Consider the random intercept model (2.1) and the set of all individual two-point designs in $S_{2,k}$ with k an even integer. Let γ_{2k} denote the only positive root of the cubic equation*

$$(k^3-9k^2-46k-24)(k^3+27k^2+122k+72)\gamma^3+(k^6-437k^4-4188k^3-13316k^2-15072k-5184)\gamma^2$$
$$-3(k+2)(k+4)(2k^3+63k^2+314k+216)\gamma-27(k+2)^2(k+4)^2=0. \tag{3.6}$$

Then the D-optimal population design for the fixed effects β of model (2.1) over the space of designs $S_{2,k}$ is given by $\xi^*_{D_1}$ provided

 (i) *k=10, 12 and $\gamma\geq\gamma_{0k}$ and*
(ii) *k=14, 16, 18, 20 and $\gamma_{0k}<\gamma\leq\gamma_{2k}$.*

The proof of this result developed in the present study was particularly intricate in that it required, *inter alia*, demonstrating that an expression comprising terms of degree 12 in k and of degree 9 in γ is less than or equal to zero for each specific value of k. A more general proof encompassing all even values of $k\geq 10$ could not be found, however, although numerical studies suggest that such a result does indeed hold.

A number of issues arise immediately from the results presented in Theorems 3.2–3.4. In particular the form of the D-optimal population designs is governed by the bounds $\gamma_c,\gamma_{0k},\gamma_{1k}$ and γ_{2k} and selected values for these limits for $k=6,8,\ldots,20$ are summarized in Table 1. Furthermore it can be shown that γ_c is strictly less than γ_{0k} and it thus follows that neither $\xi^*_{D_0}$ nor $\xi^*_{D_1}$ separately are optimal within these bounds. In fact for $k\geq 14$ and $\gamma_c<\gamma<\gamma_{0k}$ the D-optimal population design based on the space of designs $S_{2,k}$ is a mixture of the two designs $\xi^*_{D_0}$ and $\xi^*_{D_1}$. This result is summarized formally in the following theorem.

Theorem 3.5. *Consider the random intercept model (2.1) and the set of all two-point individual designs $S_{2,k}$ for k an even integer greater than or equal to 6. Then the D-optimal population design for the fixed effects β over the space of designs $S_{2,k}$ is given by*

$$\xi^*_{D_M}=\left\{\begin{matrix}\left(0,\dfrac{k}{2}\right) & \left(0,\dfrac{k}{2}+1\right) & (0,k) & \left(\dfrac{k}{2}-1,k\right) & \left(\dfrac{k}{2},k\right)\\ w_1 & w_2 & 1-2w_1-2w_2 & w_2 & w_1\end{matrix}\right\},$$

where

$$w_1=\dfrac{1}{32k^2\gamma^2}\{3(k+2)^2+(k+2)(2k^2+21k+26)\gamma-(k^4-61k^2-116k-52)\gamma^2-(k^2-3k-6)(k^2+9k+2)\gamma^3\}$$

Table 1
Values of $\gamma_c, \gamma_{0k}, \gamma_{1k}$ and γ_{2k} for even values of k from 6 to 20.

k	γ_c	γ_{0k}	γ_{1k}	γ_{2k}
6	2.000	2.249	n	n
8	0.882	0.956	n	n
10	0.563	0.595	4.673	n
12	0.412	0.429	1.975	n
14	0.324	0.335	1.236	8.457
16	0.267	0.274	0.893	3.197
18	0.227	0.232	0.697	1.953
20	0.198	0.201	0.570	1.398

The symbol n indicates that no positive root of the associated cubic equation exists.

and

$$w_2 = \frac{1}{32(k-2)(k+2)\gamma^2}\{-3(k+2)^2-(k+2)(2k^2+21k+42)\gamma+(k^4-45k^2-180k-180)\gamma^2+(k+3)(k+6)(k^2-3k-6)\gamma^3\}$$

provided that $\gamma_c < \gamma < \gamma_{0k}$, where γ_c and γ_{0k} are defined in Theorems 3.2 and 3.3, respectively.

A further issue emanating from Theorem 3.4 is the construction of *D*-optimal population designs based on the space of two- point designs $S_{2,k}$ with $k \geq 14$ and $\gamma \geq \gamma_{2k}$. From the results presented above it is not unreasonable to conjecture that designs of the form

$$\xi_{D_a}^* = \left\{ \begin{array}{ccc} \left(0,\frac{k}{2}+a\right) & (0,k) & \left(\frac{k}{2}-a,k\right) \\ w & 1-2w & w \end{array} \right\} \tag{3.7}$$

for some positive integer a and weight w are appropriate. However, as already indicated, the proof of Theorem 3.4 is algebraically tedious and essentially incomplete. For this reason attempts to demonstrate that candidate *D*-optimal population designs of the form (3.7) with $a \geq 2$ are globally optimal for $k \geq 14$ and ranges of γ with $\gamma \geq \gamma_{2k}$ were not made. From a practical point of view the required *D*-optimal population designs can always be constructed numerically.

The results presented thus far hold for k even. Since the scope of these results is limited corresponding theorems for k odd were not fully explored. In fact numeric studies indicate that the *D*-optimal population designs for k odd mirror those for k even and are based on support points of the form $(0,(k+1)/2+a),(0,k)$ and $((k-1)/2-a,k)$ for integers $a=0, 1, 2,\dots$. The weights on the points of these designs necessarily coincide with those for population designs based on the points $(0, k+1+2a)$, $(0, 2k)$ and $(k-1-2a, 2k)$ and, in the case of $a=0$, can be obtained immediately from the formulae given in Theorem 3.3. However, values of γ which define bounds between *D*-optimal population designs with differing values of a for k odd cannot be deduced directly from those for $2k$ even.

It is interesting to relate the results developed here to those reported by Atkins and Cheng (1999) for the corresponding *D*-optimal population designs with time points taken from the interval $[-1,1]$. Their designs can be readily translated into designs with points taken from the interval $[0, k]$. It then follows that *D*-optimal population designs on the continuous design space $[0, k] \times [0, k]$ are of the same form as those based on the set of integer time points $\{0, 1, \dots, k\}$ and correspond to the design (3.7) but with a positive, not necessarily integer, and given for specified k and γ by $a_{k,\gamma} = ka_\rho/2$ where a_ρ is the unique solution to the quartic equation

$$3x^4+2(3+7\rho)x^3+3(1+4\rho+5\rho^2)x^2+2\rho(1-\rho-2\rho^2)x-\rho^2 = 0 \tag{3.8}$$

with $x \in (4\rho/17, \rho/3)$ and $\rho = \gamma/(1+\gamma)$ (Atkins and Cheng, 1999). The population designs on $[0,k] \times [0,k]$ can be "rounded" to give designs based on the time points $\{0,1,\dots,k\}$ which are expected to be near-*D*-optimal. Specifically, for given values of k and γ, the scalar $a_{k,\gamma}$ calculated from (3.8) can be replaced by the integer a satisfying $a-\frac{1}{2} \leq a_{k,\gamma} < a+\frac{1}{2}$, that is a can be taken to be $\lfloor a_{k,\gamma}+\frac{1}{2} \rfloor$, and the weights on the support points can be taken as unchanged. The calculation of near-*D*-optimal population designs in this way is particularly useful for values of γ which exceed the upper limit of γ_{2k} specified in Theorem 3.4.

4. Further results

4.1. D-optimal population designs

Consider now the construction of *D*-optimal population designs for the precise estimation of the fixed effects β in the quadratic model (2.1) which are based on *d*-point individual designs with $d \geq 3$. It follows immediately from the results of Atkins and Cheng (1999) that for $d=3$ and k even, the requisite population design comprises the single design $(0,k/2,k)$. More generally, however, since the proofs of theorems relating to one- and two-point designs given in the previous section

are intricate and in essence incomplete, no attempt was made to extend these proofs to cases where $d \geq 3$. Rather such designs can be readily obtained numerically for given values of k, d and γ.

4.2. V-optimal population designs

Suppose now that interest centres on the estimation of the mean marginal responses at a given vector of time points t_g, where the elements of t_g are taken without repetition from the set $\{0,1,\ldots,k\}$. Suppose also that the elements of t_g, together with the squares of those elements, are appropriately assembled in the design matrix X_g and thus that the population mean response at t_g is given by $\mu_g = X_g \beta$. The V-optimality criterion, which is proportional to the average of the variances of the estimators of the population mean responses at t_g, can then be formulated as the concave function

$$\Psi_V(\xi) = tr\{X_g M_\beta^{-1}(\xi) X_g^T\} = tr\{M_\beta^{-1}(\xi) X_g^T X_g\}$$

and a V-optimal design can in turn be defined as a design for which $\Psi_V(\xi)$ is minimized over the space of possible population designs. The Equivalence Theorem for this V-criterion is given in the paper by Debusho and Haines (2008) and in essence states that a population design ξ_V^* is V-optimal if and only if the directional derivative of $\Psi_V(\xi)$ at ξ_V^* in the direction of an individual design t is such that

$$\phi_V(t,\xi_V^*) = tr\{M_\beta^{-1}(\xi_V^*) X_g' X_g\} - tr\{M_\beta^{-1}(\xi_V^*) X_g' X_g M_\beta^{-1}(\xi_V^*) M_\beta(t)\} \leq 0 \, boo$$

with equality holding at the support designs of ξ_V^*. Furthermore the V-optimality criterion of interest here, namely $\Psi_V(\xi)$, is invariant to a linear transformation of the explanatory variables. This invariance follows directly from the fact that both the population mean responses $X_g \beta$ and the variance structure of the random components are themselves invariant to such a transformation. Finally note that the V-efficiency of the design ξ_1 relative to the design ξ_2 can be defined straightforwardly as the ratio $\Psi_V(\xi_2)/\Psi_V(\xi_1)$.

4.2.1. One-point individual designs

Consider again population designs for model (2.1) which comprise individual one-point designs. Then it follows immediately from the arguments presented in the case of D-optimality in Section 3.1 that the V-optimal population designs do not depend on the parameters of the model, and specifically on γ, and indeed coincide with those for the quadratic regression model (3.3). For k even, the requisite V-optimal design is readily derived and is presented in the following theorem.

Theorem 4.1. *Consider the random intercept model* (2.1) *and the set of all one-point individual designs* $S_{1,k}$ *taken from* $\{0,1,\ldots,k\}$ *where k is an even integer greater than or equal to 2. Then the V-optimal population design for the mean responses μ_g at $t_g=(0,1,\ldots,k)$ over the space of all one-point designs and for all values of $\gamma \geq 0$ is given by*

$$\xi_{V_e}^* = \left\{ \begin{matrix} (0) & \left(\dfrac{k}{2}\right) & (k) \\ w & 1-2w & w \end{matrix} \right\},$$

where $w = (k+2)(4k^2+3k-2)-2\sqrt{B}/30k^2$ *with* $B = (k-1)(2+k)(1+k^2)(4k^2+3k-2)$.

The weight w specified equals $\frac{1}{3}$ at $k=2$ and decreases monotonically with increasing k, approaching $\frac{1}{4}$ as k approaches infinity. In other words, as k increases, the V-optimal population design $\xi_{V_e}^*$ converges to a design which puts 25% of the weight at the designs with extreme time points 0 and k, and the remaining 50% at the design with time point $k/2$.

For k odd, numerical studies suggest that the V-optimal population design for the mean responses μ_g at $t_g=(0,1,\ldots,k)$ over the set of all one-point individual designs $S_{1,k}$ and for all $\gamma \geq 0$ is given by

$$\xi_{V_o}^* = \left\{ \begin{matrix} (0) & \left(\dfrac{k-1}{2}\right) & \left(\dfrac{k+1}{2}\right) & (k) \\ w & \dfrac{1}{2}-w & \dfrac{1}{2}-w & w \end{matrix} \right\},$$

where $0 < w < 1/2$. However, a proof of this observation did not appear feasible, specifically since the optimal weight w for the design $\xi_{V_o}^*$ is the root of a quartic equation. It would seem from numerical investigations that w increases monotonically with increasing k and that as k approaches infinity so $\xi_{V_o}^*$ converges to a design which puts equal weights of $\frac{1}{4}$ on the individual one-point designs.

4.2.2. Individual d-point designs with $d \geq 2$

Algebraic results for V-optimal population designs for model (2.1) based on the set of d-point individual designs with $d \geq 2$ proved elusive. The requisite designs can, however, be readily obtained numerically and, within this context, it is interesting to examine some related results that have been reported in the literature. Specifically, Abt et al. (1998) considered model (2.1) with explanatory variables taken to be the integers which, in the present setting, fall in the interval $[0,k]$ with k even, and derived V-optimal population designs based on the single-point designs (0), $(k/2)$ and (k) and the

$(k+1)$-point design $(0, \ldots k/2 \ldots k)$. In particular, they demonstrated that for γ values below a cut-off point of 1.5, the V-optimal population design comprises only the single point designs but that for $\gamma > 1.5$ the requisite optimal design also includes the $(k+1)$-point design. In the present study designs which are V-optimal over all individual d-point designs where $1 \le d \le k+1$ were constructed for selected k and γ and were found to be based on individual one- and two-point designs, with the form of the two-point designs depending on γ. For example, it is interesting to note that for $k=10$ and $\gamma = 3$ the best V-optimal population design comprises the single point designs (0), (5) and (10) with weights 0.085, 0.349 and 0.085 and the two-point designs $(0, 10)$, $(0, 6)$ and $(4, 10)$ with weights 0.165, 0.158 and 0.158, respectively, and that the corresponding design given in Abt et al. (1998) is approximately 84% efficient relative to this design.

5. An example

A survey on the foraging behaviour of honeybees was undertaken on a commercial farm near Bela–Bela in Limpopo, South Africa (Bezabih, 2009). Specifically, honeybees entering a hive were counted over a five minute period at a series of 12 different equally spaced time points during day time, that is at 7:00, 8:00, 9:00, 10:00, 11:00, 12:00, 13:00, 14:00, 15:00, 16:00, 17:00 and 18:00 hours. The experiment was run over nine hives giving a total of 108 observations. The counts were high, were therefore assumed to be normally distributed and were modelled against time with hives as a random effect. The quadratic regression model (2.1) provided a good fit to the data, with the REML estimate of the ratio of variance components given by $\hat{\gamma} = 0.115$.

Consider now redesigning the experiment in order to estimate the model parameters β, or some function of them, as precisely as possible. Assume, as in the original experiment, that only 108 observations can be taken in total, that at most one measurement can be taken on each hive at each time of day and that the ratio γ is given by its REML estimate of 0.115. Assume also that the cost of each observation is constant and that no extra costs are incurred on introducing the hives. Then, more specifically, the problem posed is that of establishing how many hives should be used in the experiment and at which of the equally spaced times the number of honeybees at each hive should be counted in order to achieve the required optimal precision. For example, should the researcher take two counts on each of 54 hives, with an equal allocation of hives to the designs based on time points 7:00 and 13:00, on time points 12:00 and 18:00 and on time points 7:00 and 18:00, or should the researcher rather take some other allocation based on say 18, 27 or 36 hives?

D- and V-optimal population designs based on d-point individual designs and on the set of all possible individual designs were obtained using the results of Theorems 3.1 with $k=11$ for $d=1$ and otherwise numerically. Exact designs were constructed directly from the resultant approximate designs by rounding and are summarized, together with values of the appropriate optimality criteria, for selected values of d in Table 2. It is interesting to note that the D-optimal criterion

Table 2
Exact population designs. The time points 7:00, …, 18:00 hours are denoted by 0, …, 11, the numbers associated with each individual design refer to the numbers of hives allocated to that design and B indicates the best design over the set of all individual designs.

d	Exact design (ξ_n)	$\Psi_D(\xi_n)$	$\Psi_V(\xi_n)$
(a) D-optimal			
1	(0) (5) (6) (11) 36 18 18 36	-2921.67	33.3404
2	(0,11) (0,6) (5,11) 18 18 18	-3017.99	33.4759
3	(0,5,11) (0,6,11) 18 18	-3010.09	34.0442
4	(0,5,6,11) (0,5,10,11) (0,1,6,11) 23 2 2	-2359.39	33.5883
12	(0,1,2,3,4,5,6,7,8,9,10,11) 9	-556.89	44.6196
B	(0,6) (5,11) (0,11) (0,5,11) (0,6,11) 13 13 13 5 5	-3019.24	33.5664

d	Exact design, ξ_n	$\Psi_V(\xi_n)$	$\Psi_D(\xi_n)$
(b) V-optimal			
1	(0) (5) (6) (11) 29 25 25 29	30.9570	-2641.08
2	(0,11) (0,6) (5,11) 4 25 25	31.2349	-2743.75
3	(0,5,11) (0,6,11) (0,5,6) (5,6,11) 11 11 7 7	32.2861	-2592.85
4	(0,5,6,11) 27	33.4215	-2350.17
B	(5) (6) (0,11) 25 25 29	30.6644	-2667.53

values for the designs in Table 2(a) increase with increasing d for $d \geq 2$ and that the V-optimal values for the designs in Table 2(b) increase with increasing d for all d. The D- and V-optimal designs are, however, very different in form. Thus, if the number of observations taken at each hive is fixed, then the best D-optimal population design comprises two-point individual designs, while the best V-optimal design comprises single-point designs. Also, the best D-optimal population design over the set of all individual designs is a mixture of 2- and 3-point designs, while the corresponding V-optimal design is based on two single-point designs and one 2-point design. More generally, it can be argued that the best D- and V-optimal population designs of those presented in Table 2 are disappointing in a practical sense in that they require experiments with large numbers of hives and very few observations taken on each hive. As a counter to this, it can be argued that the original design comprising observations at 12 distinct times on each of nine hives has a D-efficiency of 56.92% and a V-efficiency of 58.34%, both of which are alarmingly low.

6. Conclusions

The aim of this paper has been to investigate the construction of D- and V-optimal population designs for the quadratic regression model with a random intercept term and with values of the explanatory variable taken from a set of equally spaced, non-repeated time points, $0,1,\ldots,k$. Algebraic results proved somewhat elusive. D-optimal population designs based on single-point individual designs were readily found and are independent of the choice of the ratio $\gamma = \sigma_e^2 / \sigma_b^2$. However, the derivation of explicit expressions for D-optimal population designs based on two-point individual designs was not straightforward and was complicated by the fact that the designs now depend on γ. The theorems presented in this study are intricate, involve tedious algebra and are limited in that they only pertain to certain values of k and γ. The approach to the proofs of the theorems draws heavily on the work of Cheng (1995) and Atkins and Cheng (1999) but is complicated by the fact that the candidate support designs form a lattice rather than a continuum. From the limited results obtained it is tempting to conjecture that D-optimal population designs for k even based on two-point individual designs are of the form

$$\left\{ \begin{matrix} \left(0, \dfrac{k}{2}+a\right) & (0,k) & \left(\dfrac{k}{2}-a,k\right) \\ w & 1-2w & w \end{matrix} \right\},$$

for $a=0,1,\ldots$ and weight w or comprise a mixture of such designs. But this is only a conjecture. No attempt was therefore made to extend the theory to cases of d-point D-optimal population designs where $d \geq 3$ and, more generally to population designs which are D-optimal over the set of all possible individual designs. Furthermore V-optimal population designs proved very much more difficult to derive than their D-optimal counterparts, even for the case of $d=1$. These findings and limitations for the quadratic regression model with a random intercept term are in sharp contrast to the compact algebraic results obtained by Debusho and Haines (2008) for the corresponding simple linear model. It should of course be emphasized that D- and V-optimal population designs for the quadratic model with random intercept can be constructed numerically by some careful programming, at least for moderate values of d and k, and a suite of gauss routines to do this is available from the authors on request. Finally the example of honeybee foraging presented here involves $k=11$ and $d=1,\ldots,12$ and was introduced in order to illustrate the nature of the D- and V-optimal population designs that can be obtained, albeit numerically. So the challenge of obtaining explicit general formulations for D- and V-optimal population designs for polynomial regression models with a random intercept remains and innovative approaches other than the one introduced here could well be sought.

References

Abt, M., Gaffke, N., Liski, E.P., Sinha, B.K., 1998. Optimal designs in growth curve models—II correlated model for quadratic growth: optimal designs for parameter estimation and growth prediction. Journal of Statistical Planning and Inference 67, 287–296.

Atkins, J.E., Cheng, C.S., 1999. Optimal regression designs in the presence of random block effects. Journal of Statistical Planning and Inference 77, 321–335.

Atkinson, A.C., Donev, A.N., Tobias, R.D., 2007. Optimum Experimental Designs, with SAS. Oxford University Press, Oxford.

Bezabih, G., 2009. Unpublished results.

Cheng, C.S., 1995. Optimal regression designs under random block-effects models. Statistica Sinica 5, 485–497.

Chernoff, H., 1953. Locally optimal designs for estimating parameters. Annals of Mathematical Statistics 24, 586–602.

Debusho, L.K., Haines, L.M., 2008. V- and D-optimal population designs for the simple linear regression model with a random intercept term. Journal of Statistical Planning and Inference 138, 1116–1130.

Debusho, L.K., Haines, L.M., 2010. Proofs of the theorems presented in the paper "*D*- and *V*-optimal population designs for the quadratic regression model with a random intercept term" by L.K. Debusho and L.M. Haines. Technical Report ⟨http://web.uct.ac.za/depts/stats/lhaines.html⟩.

Fedorov, V.V., 1972. Theory of Optimal Experiments. Academic Press, New York.

Longford, N.T., 1993. Random Coefficient Models. Oxford University Press, Oxford.

Schwabe, R., Schmelter, T., 2008. On optimal designs in random intercept models. In: Pázman, A., Volaufová, J., Witkovský, V. (Eds.), PROBASTAT 2006. Proceedings of the Fifth International Conference on Mathematical Statistics, Smolenice 2006, vol. 39. Tatra Mountains Mathematical Publications, pp. 145–153.

Verbeke, G., Molenberghs, G., 2000. Linear Mixed Models for Longitudinal Data. Springer-Verlag, New York.