COMPROMISE DESIGNS IN HETEROSCEDASTIC LINEAR MODELS

by

Anirban DasGupta[1] Saurabh Mukhopadhyay[1] and William J. Studden[2]

Purdue University

Technical Report #91-48

Department of Statistics
Purdue University

August 1991

[1]Research supported by NSF grant DMS 89-230-71,
[2]Research supported by NSF grant DMS 91-017-30.

Abstract

We consider heteroscedastic linear models in which the variance of a response is an exponential or a power function of its mean. Such models have earlier been considered in Bickel (1978), Carroll and Ruppert (1982) etc. Classical as well as Bayes optimal experimental design is considered. We specifically address the problem of "compromise designs" where the experimenter is simultaneously interested in many estimation problems and wants to find a design that has an efficiency of at least $\frac{1}{1+\varepsilon}$ in each problem. For specific models we work out the smallest ε for which such a design exists. This is done for classical as well as Bayes problems. The effect of the variance function on the value of the smallest ε is examined. The maximin efficient design is then compared to the usual A-optimal design. Some general comments are made.

Key Words : compromise designs, optimal designs, heteroscedasticity, Chebyshev polynomials, canonical moments, linear and polynomial regressions.

AMS Subject Classification 62K05,62C10

1 Introduction

1.1 Overview.

One of the most widely used methodologies of contemporary statistics is linear regression. One commonly assumes that a n–dimensional vector $\underset{\sim}{Y}$ has mean $X\underset{\sim}{\theta}$ and variance $\sigma^2 I$ where X is the $n \times p$ matrix of design constants, $\underset{\sim}{\theta}$ is the p–dimensional vector of regression coefficients and $\sigma^2 > 0$ is the common unknown variance. This is the standard homoscedastic model. In practice, however, one commonly encounters the situation where the variance of a response depends on the corresponding values of the design variables. For example, if one had a simple linear regression with $E(y_i|x_i) = \theta_0 + \theta_1 x_i$, then one often encounters the situation where $\text{Var}(y|x_i)$ is a function of x_i, say $w(x_i)$. An even more complex scenario, not dealt with extensively in the literature, is the case when the variance of a response is a function of its mean, or in general when it is a function $V(x_i, \underset{\sim}{\theta})$ of both the design variable and the regression coefficients. Such models are of importance in multiple linear regression also. These are known as the heteroscedastic linear models.

Heteroscedastic linear models and the problem of finding efficient estimates in such models were considered in Box and Hill (1974), Bickel (1978), Jobson and Fuller (1980), Carroll and Ruppert (1982) etc. The models considered in these articles assume σ_i^2, the variance of y_i, to be functions of the form $(1 + |\tau_i|)^\lambda$ or $|\tau_i|^\lambda$ (Box and Hill), $e^{\lambda\tau_i}$ (Bickel), $1 + \lambda\tau_i^2$ (Jobson and Fuller) etc. where τ_i is the mean of y_i. These models are typically called Power or Exponential models. Carroll and Ruppert (1982) demonstrate efficient estimates of the regression vector $\underset{\sim}{\theta}$ in some of these models.

The emphasis in this article is on Bayesian estimation and design of experiments in heteroscedastic linear models. The classical case would often be regarded as a limiting Bayes case and the corresponding results will be presented. In comparison to the vast existing literature on classical optimal designs, the area of Bayes designs is only beginning to flourish. Part of the reason is the apparent difficulty in using convexity arguments; even when convexity arguments can be used, the associated numerical optimization frequently becomes difficult because of the introduction of a prior. For this and reasons of simplicity, much of the results here are presented for the case of simple linear regression. It is possible to prove some parallel results for, for example, polynomial regression but in many cases closed form Bayesian results are simply not feasible. For general results on Bayes designs, see Pilz (1979,1981,1984), Chaloner (1984), DasGupta and Studden (1991), Cook(1986) etc. For general references, see Cheng (1987), Hoel (1958), Karlin and Studden (1966), Kiefer and Studden (1976), Elfving (1952), Fedorov (1972), Kiefer (1974), Pukelsheim and Titterington (1983) etc.

1.2 Goals.

The two main goals of the present article are the following:

a. consideration of the design aspects in linear models with heteroscedastic variance functions; this includes the case where the response variable $\underset{\sim}{Y}$ has $E(\underset{\sim}{Y}) = X\underset{\sim}{\theta}$ and the variance-covariance matrix of $\underset{\sim}{Y}$ is diagonal but is possibly a function of both X and $\underset{\sim}{\theta}$, say, $\Sigma(X, \underset{\sim}{\theta})$; The covariance matrix can be generalized to the form $\sigma^2\Sigma(X, \underset{\sim}{\theta})$ where σ^2 is an unknown scalar. All classical results in the article and the Bayesian results (with a gamma prior for σ^2, independent of the prior on $\underset{\sim}{\theta}$) remain valid with the introduction of this unknown σ^2. We will thus assume $\sigma^2 = 1$ and make no reference to it.

b. finding designs with a guaranteed minimum efficiency simultaneously for estimating a number of coordinates of $\underset{\sim}{\theta}$: this is done in the context of specific heteroscedastic models of the type mentioned above.

1.3 Discussions of goals

More specifically, our goal is the following: suppose $\underset{\sim}{Y} \sim N(X\underset{\sim}{\theta}, \Sigma(X, \underset{\sim}{\theta}))$ where $\Sigma(X, \underset{\sim}{\theta})$ is diagonal and $\underset{\sim}{\theta}$ has a suitable specified prior distribution G. We are interested in simultaneously estimating k coordinates of $\underset{\sim}{\theta}$, say $\underset{\sim}{l}_i'\underset{\sim}{\theta}$, $1 \leq i \leq k$. The common approach is to let $L = (\underset{\sim}{l}_1', ..., \underset{\sim}{l}_k')'$ and estimate $L\underset{\sim}{\theta}$ by using a quadratic loss. Examples abound, however, which show that except in very special cases, an optimal design for the vector parameter $L\underset{\sim}{\theta}$ has rather poor efficiency for estimating the components of $L\underset{\sim}{\theta}$, or at least some components of $L\underset{\sim}{\theta}$. This phenomenon is well known in the context of, for example, Stein-estimation, where James-Stein type estimates can estimate individual estimate very poorly. See Rao(1976), Efron and Morris(1971,1972) etc. Our suggestion to instead look for designs which provide a minimum efficiency for esimating each component is tantamount to operating under vector risks, as for example in Cohen and Sackrowitz(1984).

1.4 Are the goals technically feasible?

In the generality that we have stated the problem, the answer seems to be 'no'. If Σ is allowed to depend on both X and $\underset{\sim}{\theta}$, even if we assume a normal prior G, the structure is simply too complicated for doing much more than isolated numerical calculations. However, the following can be achieved: suppose we let $\underset{\sim}{Y}$ have mean $X\underset{\sim}{\theta}$ and variance-covariance matrix C. We only allow linear estimate for $L\underset{\sim}{\theta}$, i.e., estimates of the form $A_0\underset{\sim}{Y}$, where A_0 is any fixed symmetric matrix. If we restrict to linear estimates, then we gain the generality that neither $\underset{\sim}{Y}$ nor $\underset{\sim}{\theta}$ needs to be normally distributed. We thus do not make any specific distributional assumptions on them. Theorem 2.1 says that in this case the optimal design for estimating the entire vector $L\underset{\sim}{\theta}$ minimizes $trL(X'B^{-1}X + C^{-1})^{-1}L'$ where $B = E_G(\Sigma(X, \underset{\sim}{\theta}))$. One immediately realizes that this is formally equivalent to finding an optimal design for estimating $L\underset{\sim}{\theta}$ when one has the more specific structure $\underset{\sim}{Y} \sim N(X\underset{\sim}{\theta}, B)$ and $\underset{\sim}{\theta} \sim N(\underset{\sim}{0}, C)$ but arbitrary estimates (not just the linear ones) for $L\underset{\sim}{\theta}$ are allowed. Notice that B depends only on X but not on $\underset{\sim}{\theta}$. Thus, if linear extimates do not seem adequate, one can turn arround and think of this as a normal problem with heteroscedastic variance function depending (only) on X. There have been some recent interest in adequacy of linear estimates in various estimation problems: see Donoho, Liu and MacGibbon(1990), Vidacovic and DasGupta(1991) etc.

1.5 Choice of variance function.

In view of the above discussion, we simply take the view that $\underset{\sim}{Y} \sim N(X\underset{\sim}{\theta}, B)$ where B depends on X, and $\underset{\sim}{\theta} \sim N(\underset{\sim}{0}, C)$. To derive specific optimal designs, specific choice of B has to be made. We must keep in mind, however, that B occurs as $E_G(\Sigma(X, \underset{\sim}{\theta}))$ in the more general variance structure. We therefore make specific choices for B. Recall Σ is always assumed to be diagonal.

As in Bickel(1978), Box and Hill(1974) and Carroll and Ruppert(1982), we consider the following forms for $\Sigma = Diag(\sigma_1^2, \sigma_2^2, ..., \sigma_n^2)$, where

$$(i) \quad \sigma_i^2 = e^{\lambda E(y_i)}$$

$$(ii) \quad \sigma_i^2 = (E(y_i))^2.$$

Taking G to be a $N(\underset{\sim}{0}, I)$ prior, for instance, this results in $B = Diag(b_1, b_2, ..., b_n)$, where $b_i = e^{\frac{\lambda^2}{2}X_i'X_i}$ in case (i) and $b_i = X_i'X_i$ in case (ii), where X_i' is the i-th row of the design matrix X. Taking G to be $N(\underset{\sim}{0}, I)$ is just an artifact; the eventual derivations get only a bit more cumbersome for a general spherically symmetric prior (see the statement of Theorem 2.2 .

Example : Consider the simple linear regression model with $E(y|x) = \theta_0 + \theta_1 x$. Then the above choices are equivalent to $Var(y|x) = e^{\frac{\lambda^2}{2}(1+x^2)}$ and $Var(y|x) = 1 + x^2$ respectively. For polynomial regression with $E(y|x) = \theta_0 + \theta_1 x + ... + \theta_p x^p$, they are respectively equivalent to $Var(y|x) = e^{\frac{\lambda^2}{2}(1+x^2+...+x^{2p})}$ and $Var(y|x) = 1 + x^2 + ... + x^{2p}$.

1.6 Notation, outline.

To keep the article more focused, we derive our results for polynomial regressions only. In any case, polynomial models are clearly more used than other more general linear models. Thus we let the normally distributed observation y have mean $E(y|x) = \theta_0 + \theta_1 x + ... + \theta_p x^p$, where x is assumed to belong to the interval $[0,1]$ and we let $Var(y|x) = b(x)$ for specific choices of $b(x)$. The following two choices of $b(x)$ are considered in this article

$$\begin{aligned} (i)\quad b_1(x) &= e^{\frac{\lambda^2}{2}\sum_{i=0}^{p}|x|^{2i}} \text{ for some real } \lambda. \text{ and} \\ (ii)\quad b_2(x) &= 1 + c|x|^\lambda forc \geq 0, \lambda \geq 2, \end{aligned}$$

Motivation for using such variance functions was given above. The standard homoscedastic variance case is subsumed in our setup. The regression vector $\underset{\sim}{\theta}$ is also assumed to be normal and to have mean $\underset{\sim}{0}$ and covariance matrix C, where C is diagonal.The case of a nondiagonal C is certainly practically important, but results in substantially more complexity. We would, for this reason, usually recommend that if C was thought to be nondiagonal, one should bound it by a diagonal C_1(i.e., $C_1 - C$ is nonnegative definite) use C_1 instead. The results for the classical case can be formally obtained from our results by letting $C^{-1} \to 0$(the null matrix).

As mentioned in Section 1.2 and 1.3, the principal goal of this article is to derive compromise designs with with a minimal guaranteed efficiency for estimating a number of coefficients $\{\theta_i\}$. If we let ξ denote an arbitrary design and if $v_i(\xi)$ is the Bayes risk for estimating θ_i by using the design ξ, then the efficiency of ξ for estimating θ_i is defined as

$$e_i(\xi) = \frac{v_i}{v_i(\xi)},$$

where $v_i = \inf_\xi v_i(\xi)$.

Given $\varepsilon > 0$, we define

$$\Gamma_\varepsilon(C) = \left\{\xi : e_i(\xi) \geq \frac{1}{1+\varepsilon},\ 1 \leq i \leq p\right\}.$$

For small $\varepsilon > 0$, $\Gamma_\varepsilon(C)$ is usually empty. Of natural interest is the quantity

$$\varepsilon_0 \quad = \quad \inf\{\varepsilon > 0 : \Gamma_\varepsilon(C) \text{ is nonempty }\}.$$

$\frac{1}{1+\varepsilon_0}$ will be called the 'maximin efficiency' and if $\Gamma_{\varepsilon_0}(C)$ is itself nonempty (which it is in our examples), then any design in $\Gamma_{\varepsilon_0}(C)$ will be called a 'maximin efficient' design.

Notice that the quantities v_i are needed for the analysis here. This amounts to deriving the Bayes design for each θ_i separately. Thus the subsequent sections contain the following main type of results:

i derivation of optimal design for each θ_i,

ii derivation of the maximin efficiency and a maximin efficient design,

iii discussion of the effect of $b(x)$ on the maximin efficiency and maximin efficient designs.

Point iii is in particular important, because it is not easy to specify the parameters of the variance functions exactly. An alternative approach may be to treat them as unknowns and either use another prior for them or use a more classical approach such as minimaxing over them.

Section 2 treats the case $b(x)=e^{\frac{\lambda^2}{2}\sum_{i=0}^{p}|x|^{2i}}$, with illustrative emphasis on $p = 2$(quadratic regression); Section 3 treats $b(x) = 1 + c|x|^\lambda$ for $p = 1$(linear regression). The case $p > 1$ for this variance structure seems technically undoable. We assume x belong to $[0, 1]$. The case of compromise designs in the presence of many priors is discussed in Mukhopadhyay and DasGupta(1991). For some related work in general, see Lee(1988).

1.7 An illustrative example.

Here we give a brief example to illustrate the main idea of this article. For polynomial regression of arbitrary order on $[-1, 1]$, we describe a comprehensive method for deriving the maximin efficiency and a maximum efficient design assuming a homoscedastic classical setup.

Formally, consider the usual homoscedastic polynomial regression model

$$y = \theta_0 + \theta_1 x + ... + \theta_{p-1}x^{p-1} + \varepsilon \qquad -1 \leq x \leq 1, \tag{1}$$

where ε's are iid $N(0,\sigma^2)$, $\sigma^2(>0)$ is unknown.

Instead of using the usual moments, it is computationally convenient to get the solution in terms of canonical moments. We first briefly describe some relevant theory of canonical moments which will be useful to obtain the solutions. More detail and applications can be found in Skibinsky(1967,1968,1969,1986), Studden(1980,1982,1982a,1989), Lau(1983,1988) and Dette(1991). For an arbitrary design ξ on $[-1,1]$ let $c_k = \int_{-1}^{1} x^k d\xi(x)$ denote the k-th moment of ξ $(k = 0,1,...)$. Then the canonical moments are defined as

$$p_i = \frac{c_i - c_i^-}{c_i^+ - c_i^-} \quad i = 1,2,... \ ,$$

where c_i^+ and c_i^- denote respectively the maximum and the minimum values of the i-th moment for a given set of moments $c_0, c_1, ..., c_{i-1}$. Note that $0 \leq p_i \leq 1$ and the canonical moments are left undefined whenever $c_i^+ = c_i^-$. If i is the first index for which equality holds then $0 < p_k < 1$ $(k = 1,...,i-2)$ and p_{i-1} must be 0 or 1. The relationship between the two types of moments can be expressed by the use of certain recursive relationships(see Studden(1980) and Skibinsky(1968)). In the special cases when all the odd (ordinary) moments are zero (such as for symmetric designs), the even moments can be expressed recursively in terms of canonical moments. Note that for symmetric designs, all the odd canonical moments are equal to $\frac{1}{2}$.

Lemma 1.1 *Let ξ denote a symmetric probability measure on $[-1,1]$. Let $1-p_0 = S_{0,j} = 1$, $j = 1,2,....$ Take $\zeta_j = (1-p_{j-1})p_j$, $j = 1,2,...$ and recursively define*

$$S_{i,j} = \sum_{k=i}^{j} \zeta_{2(k-i+j)} S_{i-1,k}, \quad j = i, i+1, ..., i = 1,2,...;$$

then the following functional equality holds :

$$c_{2m} \quad = \quad S_{m,m}, \quad m = 1, 2, \ldots \ .$$

The support and the weights of probability measures corresponding to a terminating sequence of canonical moments can be found by standard techniques (see Studden(1982a) or Lau(1983,1988)). The following result is useful for the special sequences with all odd canonical moments equal to $\frac{1}{2}$ (see Lau(1983)).

Lemma 1.2 *Let ξ denote a probability measure on $[-1,1]$ with canonical moments $p_i \in (0,1)$ $(i \leq 2n-1)$, $p_{2i-1} = \frac{1}{2}$ $(i \leq n)$, $p_{2n} = 1$ and let the polynomials $Q_k(x,\xi)$ and $P_k(x,\xi)$ be defined recursively by ($Q_{-1}(x,\xi) = P_{-1}(x,\xi) = 0$, $Q_0(x,\xi) = P_0(x,\xi) = 1$)*

$$Q_{k+1}(x,\xi) \quad = \quad xQ_k(x,\xi) - p_{2k}(1-p_{2k+2})Q_{k-1}(x,\xi) \quad (k \geq 1)$$

$$P_{k+1}(x,\xi) \quad = \quad xP_k(x,\xi) - (1-p_{2k})p_{2k+2}P_{k-1}(x,\xi) \quad (k \geq 1).$$

The design ξ is supported at the zeros of the polynomial $(1-x^2)Q_{n-1}(x,\xi)$ and the weights at the support points $x_0, \ldots, x_n$ are given by

$$\xi(\{x_j\}) = \frac{P_n(x_j,\xi)}{\frac{d}{dx}(x^2-1)Q_{n-1}(x,\xi)|_{x=x_j}} \quad j = 0, 1, \ldots, n \ .$$

Suppose we want to estimate $\{\theta_{i_1}, \ldots, \theta_{i_t} : 0 \leq i_1 < \ldots < i_t \leq p-1\}$. We compute the maximin efficient designs in terms of canonical moments for different sets of θ_i for $p = 2, 3, 4$ and 5. It is not hard to see using standard invariance arguments that a symmetric maximin efficient design necessarily exists. Thus we can use Lemma 1.1 to express M in terms of the canonical moments. By using *Mathematica*, the exact solution can be obtained.

After we find the design in terms of canonical moments we can get the support and their weights by using Lemma 1.2. Table A provides the 'maximin efficiencies'. It is very interesting to note that in each case, the maximin efficiency for the set $\{\theta_0, ..., \theta_{p-2}\}$ coincides with that of the full set and that generally speaking, the maximin efficiencies are higher if even and odd coefficients are not grouped together.

2 Maximin efficient designs when $b(x) = b_1(x)$.

2.1 Two general theorems.

In this section we first consider the problem of finding optimal designs in general heteroscedastic linear models when attention is restricted to only linear estimates. Some of the results in this section should be considered known but they motivate other results in the latter sections. Throughout we assume squared error loss.

Notation: $\underset{\sim}{Z} \sim (\underset{\sim}{\mu}, \Sigma)$ will mean $E(\underset{\sim}{Z}) = \underset{\sim}{\mu}$ and the variance–covariance matrix of $\underset{\sim}{Z}$ is Σ.

Theorem 2.1 *Let* $\underset{\sim}{Y}_{n\times 1} \sim (X\underset{\sim}{\theta}, \Sigma(\underset{\sim}{\theta}))$ *and let* $\underset{\sim}{\theta}_{p\times 1} \sim (\underset{\sim}{0}, C)$. *Let* $B = E(\Sigma(\underset{\sim}{\theta}))$ *(under the prior; we assume* B *exists). For estimating* $L\underset{\sim}{\theta}$*, where* L *is a* $k \times p$ *matrix, the best linear estimate* $A_0\underset{\sim}{Y}$ *has* $A_0 = LCX'(B + XCX')^{-1}$ *and the corresponding Bayes risk equals* $tr\ L(X'B^{-1}X + C^{-1})^{-1}L'$.

Proof: Straightforward. ♣

Notice the formal equivalence of restricting to linear estimates and assuming that $\underset{\sim}{Y} \sim N(X\underset{\sim}{\theta}, B)$ and $\theta \sim N(\underset{\sim}{0}, C)$. If $\underset{\sim}{Y} \sim N(X\underset{\sim}{\theta}, B)$ and $\underset{\sim}{\theta}$ is normally distributed with a covariance matrix of C, then the posterior covariance matrix of $\underset{\sim}{\theta}$ is $(X'B^{-1}X + C^{-1})^{-1}$ and hence the Bayes risk for estimating $L\underset{\sim}{\theta}$ is $tr\ L(X'B^{-1}X + C^{-1})^{-1}L'$.

Suppose now $B = Diag(b_1, ..., b_n)$ with $b_i = e^{\frac{\lambda^2}{2}(1+x_i^2+...+x_i^{2p})}$, where λ is arbitrary. Optimal design theory for such B is simplfied for polynomial regression due to the following general theorem.

Theorem 2.2 *Let $E(y_i|x_i) = \theta_0 + \theta_1 x_i + \ldots + \theta_p x_i^p$, where $0 \leq x_i \leq 1$, and let $Var(y_i|x_i) = e^{\lambda E(y_i|x_i)}$. Let $\underset{\sim}{\theta} = (\theta_0, \theta_1, \ldots, \theta_p)'$ have a spherically symmetric prior G(i.e., $G(.)$ depends on $\underset{\sim}{\theta}$ only through $\underset{\sim}{\theta}'\underset{\sim}{\theta}$). Then the set D of designs with support on 0 and at most p other points forms a complete class in the sense that given any design ξ_1, there exists a design ξ_2 in D such that $M(\xi_2) \geq M(\xi_1)$ where for any design, M denotes $X'B^{-1}X$.*

Proof: First notice that using a standard argument, $b(x) = F(\underset{\sim}{x}'\underset{\sim}{x})$ for some suitable F, where $\underset{\sim}{x} = (1, x, ..., x^p)'$. Transforming to the new regression functions $f_i(x) = \frac{x^i}{\sqrt{b(x)}}$, $0 \leq i \leq p$, the Theorem follows on using a standard argument with Chebyshev-systems(see Karlin and Studden(1966)). ♣

2.2 Applications.

Example : The function $b(x) = e^{\frac{\lambda^2}{2}\sum_{i=0}^{p}|x|^{2i}}$ corresponds to taking special spherically symmetric $N(\underset{\sim}{0}, I)$ prior. For quadratic regression(i.e., p=2), Theorem 2.2 permits reduction of the design problem to a four parameter optimization. Table 1 gives the value of the maximin efficiency e and the maximin efficient design itself for various λ and n. Notice that unlike in classical design theory, Bayes optimal designs usually depend on n. In Table 1, the set $\{0, x_1, x_2\}$ consists the support of the maximin efficient design and p_i is the mass assigned to x_i; also, r denotes $\frac{1}{n}$.

Discussion of Table 1 : First of all, the maximin efficiencies are the lowest for the case $\lambda = 0$. Thus, in the usual homoscedastic case, one gets the worst maximin efficiencies. Secondly, in fact, the larger the value of λ, the better the maximin efficiency appears to be for any fixed value of n. On the other hand, for any fixed λ, the maximin efficiency decreases as n increases. The worst case is the situation of a homoscedastic model with a large sample.

Notice, however, that if λ is reasonably large, say $\lambda \geq 1.5$, then for moderate sample sizes like 25, the maximin efficient design guarantees a very respectable minimum efficiency of about 85% or more for estimating every coefficient θ_i.

As we stated in section 1.3, the common approach is to estimate the entire $\underset{\sim}{\theta}$ vector by using a collapsed quadratic loss. Therefore, a comparison of the minimum efficiency (over the different coefficients θ_i) achieved by using the optimal design for the entire $\underset{\sim}{\theta}$ vector with the maximin efficiencies of Table 1 seems in order. Table 2 gives the optimal designs for the whole $\underset{\sim}{\theta}$ vector and its minimum efficiency over the coordinates θ_i. Again, $\{0, x_1, x_2\}$ constitute the support set, p_i denotes the mass at x_i, $r = \frac{1}{n}$ and $e^\star$ is the minimum efficiency.

Discussion of Table 2 : In general, the phenomena evidenced in Table 1 seems present in Table 2 too. The gain in the minimum efficiency by using the maximin efficient design is substantial if $\lambda \leq 1.5$ or even if $\lambda \leq 2.5$ but only if n is large. For $\lambda = 3$, irrespective of the value of n, there seems to be practically nothing to gain by using the maximin efficient design.

Combining Table 1 and Table 2, generally speaking, deriving the maximin efficient design is worthwhile if λ is moderate or if λ is small and n not very large. In general, if λ is large, then deriving the maximin efficient design is not worthwhile because one may use the usual optimal design for the entire $\underset{\sim}{\theta}$ vector as well.

3 Maximin efficient design when $b(x) = b_2(x)$.

3.1 Bayes designs for each θ_i.

For the following analysis we will assume $y_i \overset{\text{indep}}{\sim} N(\theta_0 + \theta_1 x_i,\ 1 + c x_i^\lambda)$, for some $c > 0$, $\lambda \geq 2$, $0 \leq x_i \leq 1$, and $\underset{\sim}{\theta} = (\theta_0, \theta_1)' \sim N(\underset{\sim}{0}, C)$ where $\frac{C^{-1}}{n} = \begin{pmatrix} r_0 & 0 \\ 0 & r_2 \end{pmatrix}$ for $r_0,\ r_2 \geq 0$ (the matrix with r_0 or r_2 equal to zero is not invertible but formally the classical optimal designs can be found by substituting the null matrix for C^{-1}). We have the following Theorem.

Theorem 3.1 *Let $E(y_i) = \theta_0 + \theta_1 x_i$, $0 \leq x_i \leq 1$ and let $Var(y_i) = 1 + cx_i^\lambda$, $c \geq 0$, $\lambda \geq 2$. Assume $\underset{\sim}{\theta}$ has a $N(\underset{\sim}{0}, C)$ prior where $\frac{C^{-1}}{n} = \begin{pmatrix} r_0 & 0 \\ 0 & r_2 \end{pmatrix}$. Then*

a. *The optimal Bayes design for estimating θ_0 is supported on $x = 0$.*

b. *The optimal design for estimating θ_1 is given as follows :*

Define quantities x_0, x_1, p_0, p_1 whenever they are well defined as follows :

$$\begin{aligned} x_0 &= \left(\frac{4(\lambda-1)}{c(\lambda-2)^2}\right)^{\frac{1}{\lambda}}, \\ x_1 &= \left(\frac{2(1+r_0)}{cr_0(\lambda-2)}\right)^{\frac{1}{\lambda}}, \\ p_0 &= \frac{\lambda(1+r_0)}{2(\lambda-1)} \text{ and} \\ p_1 &= \frac{(1+r_0)\sqrt{1+c}(\sqrt{1+c}-1)}{c}. \end{aligned}$$

(i) For $c(\lambda-2) \leq 2$, the design is supported on $\{0,1\}$ with mass p_1 at 1;

(ii) For $2 < c(\lambda-2) \leq \frac{2(1+r_0)}{r_0}$ and $c \leq \frac{4(\lambda-1)}{(\lambda-2)^2}$, the design is supported on

(iia) $\{0,1\}$ with mass p_1 at 1 if $\frac{c}{1+c} \leq 1 - r_0^2$;

(iib) only $\{1\}$ if $\frac{c}{1+c} > 1 - r_0^2$;

(iii) For $2 < c(\lambda-2) \leq \frac{2(1+r_0)}{r_0}$ and $c > \frac{4(\lambda-1)}{(\lambda-2)^2}$, the design is supported on

(iiia) only $\{x_1\}$ if $\lambda(1-r_0) \leq 2$;

(iiib) $\{0, x_0\}$ with mass p_0 at x_0 if $\lambda(1-r_0) > 2$;

(iv) For $c(\lambda-2) > \frac{2(1+r_0)}{r_0}$ and $c \leq \frac{4(\lambda-1)}{(\lambda-2)^2}$, the design is supported on

(iva) $\{0,1\}$ with mass p_1 at 1 if $\frac{c}{1+c} \leq 1 - r_0^2$;

(ivb) only $\{x_1\}$ if $\frac{c}{1+c} > 1 - r_0^2$;

(v) For $c(\lambda-2) > \frac{2(1+r_0)}{r_0}$ *and* $c > \frac{4(\lambda-1)}{(\lambda-2)^2}$*, the design is supported on*

(va) only $\{x_1\}$ *if* $\lambda(1-r_0) \leq 2$*;*

(vb) $\{0, x_0\}$ *with mass* p_0 *at* x_0 *if* $\lambda(1-r_0) > 2$.

Proof: The proof of part (a) is trivial.

Part (b) follows on using the usual equivalence theorem arguments. In each case one has to check (for $0 \leq x \leq 1$) the inequality $((c_0+r_0)x - c_1)^2 \leq (1+cx^\lambda)\cdot Q$, where $Q = (-c_1,\ c_0+r_0)\begin{pmatrix} c_0 & c_1 \\ c_1 & \frac{1-c_0}{c} \end{pmatrix}\begin{pmatrix} -c_1 \\ c_0 \end{pmatrix}$, and c_0, c_1 are the values of $E\frac{1}{b(X)}$ and $E\frac{X}{b(X)}$ as given by the designs in the statements of the theorem. ♣

<u>Remark</u> : For a nondiagonal prior covariance matrix C, the optimal design for θ_1 cannot in general be written down in a closed form. Also, as in Section 2, the the optimal Bayes design (for θ_1) depends on the sample size n.

<u>Discussion of Theorem 3.1</u>. Notice that the design for θ_0 is in general supported on 0. If however, C is not diagonal, this may not be the case in general. The optimal design for estimating θ_1 can be one of three different kinds: it may be supported on 0 and 1, or on 0 and another interior point, or it may even be a one–point design. Notice also that whatever be c, if λ gets very large then the design will be supported on 0 and another interior point x_0. However, surprisingly, as $\lambda \to \infty$, x_0 converges back to 1 as can be easily seen by checking that $\log x_0$ converges to zero. Note that this interior point is independent of the prior. Consider now the classical case with a large λ. Since the classical case can be thought of as the case with $r_0 = r_2 = 0$, for large enough λ the support of the Bayes design coincides with that of the classical design as long as $r_0 < 1$. If the constant c in the weight function $= 1$, and if one considers the classical case $r_0 = 0$, then it follows from Theorem 3.1 that the optimal design for θ_1 is supported on 0 and 1 with mass $p = 2-\sqrt{2}$ at 1 if $\lambda \leq 4+2\sqrt{2}$, and is supported on 0 and $\left(\frac{4(\lambda-1)}{(\lambda-2)^2}\right)^{\frac{1}{\lambda}}$ with mass $\frac{\lambda-2}{2\lambda-2}$ at zero. This interior point goes down

to approximately .92866 at $\lambda = 16.4245$ and then starts moving back to 1.

3.2 Maximin efficient designs, $0 \le x \le 1$.

In this section, we derive the value of the smallest ε for which $\Gamma_\varepsilon(C)$ is nonempty and also give geometric descriptions of the set $\Gamma_\varepsilon(C)$ in terms of the moments c_0, c_1, etc. (each design corresponds to a moment sequence). For ease of representation and understanding, we will present most of the analysis for the classical case while keeping in mind that the analysis is similar for the Bayes case although the algebra is of necessity more complicated. We first need the following notations and a theorem.

Given $\lambda \ge 2$, denote $r = \frac{1}{\lambda-1}$ and $p = \frac{\lambda}{\lambda-2}$. Note $p = \infty$ if $\lambda = 2$. Also let $v_i = \inf_\xi v_i(\xi)$, where $v_i(\xi)$ denotes the risk for estimating θ_i using the design ξ. v_i thus simply represents the risk obtained by using the corresponding optimal design. As stated before, we will assume that we have a simple linear regression with the independent variable varying in [0,1] and a variance function $w(x) = 1 + cx^\lambda,\ \lambda \ge 2,\ c > 0$. Also, for the classical case, $\Gamma_\varepsilon(C)$ will be denoted as simply Γ_ε.

Theorem 3.2 *Let $r_0 = r_2 = 0$. Then*

(i) $v_0 = 1$

(ii)

$$\begin{aligned} v_1 &= (\sqrt{1+c}+1)^2 \ \textit{if}\ \lambda \le 2\sqrt{\frac{1+c}{c}}\left(\sqrt{\frac{1+c}{c}}+1\right) \\ &= c^{\frac{2}{\lambda}} \cdot 4^{1-\frac{2}{\lambda}} \cdot (\lambda-1)^{2-\frac{2}{\lambda}} \cdot (\lambda-2)^{\frac{4}{\lambda}-2} \\ &\quad \textit{if}\ \lambda > 2\sqrt{\frac{1+c}{c}}\left(\sqrt{\frac{1+c}{c}}+1\right). \end{aligned}$$

Proof: Recall the definitions $c_i = E\frac{X^i}{b(X)}$, $i = 0,1,2$ where $E(\cdot)$ denotes expectation with respect to the relevant design (measure).

Part (i) follows from Theorem 3.1. To prove part (ii), conclude using part (b) in Theorem 3.1 that the optimal design for θ_1 is supported on 0 and 1 with mass $\frac{\sqrt{1+c}(\sqrt{1+c}-1)}{c}$ at 1 if $\lambda \le 2\sqrt{\frac{1+c}{c}}\left(\sqrt{\frac{1+c}{c}}+1\right)$ and otherwise it is supported on 0 and $\left(\frac{4(\lambda-1)}{c(\lambda-2)^2}\right)^{\frac{1}{\lambda}}$ with mass $\frac{\lambda-2}{2(\lambda-1)}$ at zero. The second assertion in part (ii) of the current theorem now follows on algebra by using the fact that $v_1 = \frac{c_0}{c_0 c_2 - c_1^2}$ where in c_i expectation is taken with respect to the designs described above. ♣

Lemma 3.1 *For any $\lambda > 2$, if Γ_ε is nonempty for some $\varepsilon > 0$, then Γ_ε also contains a two point design supported at 0 and some other point in the interval [0,1].*

Proof: The proof uses a standard complete class argument by arguing that

$$\left(1, \frac{1}{1+cx^\lambda}, \frac{x}{1+cx^\lambda}, \frac{-x^2}{1+cx^\lambda}\right)$$

form a Chebyshev system; again, see Karlin and Studden(1966). ♣

In view of the above lemma, it is enough to consider designs supported on 0 and x_0 (where $0 \le x_0 \le 1$) with mass p at x_0. Here p and x_0 are kept arbitrary. For such designs, there is a convenient representation of c_2 in terms of c_0 and c_1. This is the assertion of the following theorem.

Theorem 3.3 *For $\lambda \ge 2$, $c_2 = \frac{(1-c_0)^r c_1^{1-r}}{c^r}$ for all two point designs supported on 0 and some other x_0 in [0,1].*

Proof: For $\lambda = 2$, r equals 1 and the above representation is trivial (in fact it is valid for all designs). For $\lambda > 2$, note that

$$c_0 = 1 - p + \frac{p}{1+cx_0^\lambda}, \tag{2}$$

$$c_1 = \frac{px_0}{1+cx_0^\lambda}, \tag{3}$$

$$\text{and } c_2 = \frac{px_0^2}{1+cx_0^\lambda}. \tag{4}$$

Solving the first two equations for p and x_0 one obtains

$$x_0 = \left(\frac{1-c_0}{cc_1}\right)^r \tag{5}$$

$$\text{and } p = \frac{c^r c_1^{1+r} + (1-c_0)^{1+r}}{(1-c_0)^r}. \tag{6}$$

Substituting (6) into c_2 in (4) one gets the required result. ♣

We now go onto deriving the value of the smallest ε such that the set of designs Γ_ε is not empty. Note that if ε_0 is the smallest such value then $\frac{1}{1+\varepsilon_0}$ is the maximin efficiency and the corresponding design is a maximin efficient design.

Towards this end, recall that in view of Lemma 3.1 it is enough to consider designs supported on 0 and some other point in the interval [0,1]. Also recall that for such designs c_2 is completely determined from c_0 and c_1. Finally note that a pair (c_0, c_1) arises from a valid design (for the variance function $1 + cx^\lambda$) if an only if $\frac{1}{1+c} \le c_0 \le 1$, $c_1 \ge \frac{1-c_0}{c}$, and $c_1 \le c_0^{1-\frac{1}{\lambda}} \left(\frac{1-c_0}{c}\right)^{\frac{1}{\lambda}}$ (the designs for which $c_1 = \frac{1-c_0}{c}$ are those supported on $\{0,1\}$ and the designs for which $c_1 = c_0^{1-\frac{1}{\lambda}} \left(\frac{1-c_0}{c}\right)^{\frac{1}{\lambda}}$ are the one point designs; that for every other design the third inequality $c_1 \le c_0^{1-\frac{1}{\lambda}} \left(\frac{1-c_0}{c}\right)^{\frac{1}{\lambda}}$ holds follows from the Cauchy–Schwartz inequality). We will call

$$\mathcal{M} = \left\{(c_0, c_1)\colon \frac{1}{1+c} \le c_0 \le 1,\ c_1 \ge \frac{1-c_0}{c},\ c_1 \le c_0^{1-\frac{1}{\lambda}} \left(\frac{1-c_0}{c}\right)^{\frac{1}{\lambda}}\right\} \tag{7}$$

the moment space of the problem.

Notice that $\mathcal{M}$ can also be written as

$$\mathcal{M} = \left\{(c_0, c_1)\colon \frac{1}{1+c} \le c_0 \le 1,\ c_1 \ge \frac{1-c_0}{c},\ c_1 \le c_0^{\frac{1}{1+r}} \left(\frac{1-c_0}{c}\right)^{\frac{r}{1+r}}\right\}. \tag{8}$$

Now for a two point design ξ described above,

$$e_i(\xi) = \frac{v_i}{v_i(\xi)} = \left(\frac{c_{2-i}}{c_0c_2 - c_1^2}\right)^{-1} \cdot v_i \ge \frac{1}{1+\varepsilon} \tag{9}$$

$$\Leftrightarrow \frac{c_{2-i}}{c_0c_2 - c_1^2} \le (1+\varepsilon)v_i,\ i = 0, 1. \tag{10}$$

Using Theorem 3.3 and part (i) of Theorem 3.1, the first inequality in (10) reduces to

$$(1-c_0)^r c_1^{1-r} - (1+\varepsilon)(1-c_0)^r c_1^{1-r} c_0 + (1+\varepsilon)c^r c_1^2 \leq 0. \tag{11}$$

Similarly the second inequality in (10) reduces to

$$c^r c_0 - (1+\varepsilon)v_1 c_0 (1-c_0)^r c_1^{1-r} + (1+\varepsilon)v_1 c^r c_1^2 \leq 0. \tag{12}$$

Motivated by these, we will define

$$S_0 = \{(c_0, c_1) : (11) \text{ holds}\}, \tag{13}$$

$$S_1 = \{(c_0, c_1) : (12) \text{ holds}\}. \tag{14}$$

Notice that elements of S_0 or S_1 need not be within the moment space $\mathcal{M}$ but Γ_ε is nonempty for a specific $\varepsilon > 0$ if and only if $S_0 \cap S_1 \cap \mathcal{M}$ is nonempty for this $\varepsilon > 0$.

We now claim that if ε_0 is the smallest ε with the property that $S_0 \cap S_1 \cap \mathcal{M} \neq \phi$, then there exists a point in $\partial S_0 \cap \partial S_1 \cap \mathcal{M}$ where ∂S_i denotes the boundary of S_i (in fact we will prove a stronger assertion).

Lemma 3.2 *For $\varepsilon \geq 0$, let $A_\varepsilon, B_\varepsilon$ be closed sets and let C be another fixed closed set. Suppose $A_\varepsilon \cap C$, $B_\varepsilon \cap C$ are (closed) convex sets with nonempty interiors for each $\varepsilon > 0$ but $A_0 \cap B_0 \cap C = \phi$. Let $\varepsilon_0 = \inf\{\varepsilon > 0\colon A_\varepsilon \cap B_\varepsilon \cap C \neq \phi\}$. Then $A_{\varepsilon_0} \cap B_{\varepsilon_0} \cap C = \partial A_{\varepsilon_0} \cap \partial B_{\varepsilon_0} \cap C$, provoded A_ε, B_ε are continuos in ε.*

Proof: First note that $\varepsilon_0 > 0$ and also that $A_{\varepsilon_0} \cap B_{\varepsilon_0} \cap C = (\partial A_{\varepsilon_0} \cup A^0_{\varepsilon_0}) \cap (\partial B_{\varepsilon_0} \cup B^0_{\varepsilon_0}) \cap C$ where D^0 denotes the interior of D.

$$\begin{aligned} A_{\varepsilon_0} \cap B_{\varepsilon_0} \cap C &= (\partial A_{\varepsilon_0} \cap \partial B_{\varepsilon_0} \cap C) \cup (\partial A_{\varepsilon_0} \cap B^0_{\varepsilon_0} \cap C) \\ &\quad \cup (A^0_{\varepsilon_0} \cap \partial B_{\varepsilon_0} \cap C) \cup (A^0_{\varepsilon_0} \cap B^0_{\varepsilon_0} \cap C). \end{aligned}$$

By definition of ε_0, we have that $A^0_{\varepsilon_0} \cap B^0_{\varepsilon_0} \cap C = \phi$. Now observe that $B_{\varepsilon_0} \cap C$ is a closed convex set with a nonempty interior and is therefore regular, i.e., $\overline{(B_{\varepsilon_0} \cap C)^0} = B_{\varepsilon_0} \cap C$.

Suppose now $A^0_{\varepsilon_0} \cap B_{\varepsilon_0} \cap C \neq \phi$. Then $\exists x \varepsilon A^0_{\varepsilon_0}$ which is also in $B_{\varepsilon_0} \cap C$. Therefore, there is a sphere $S(x,r) \subseteq A^0_{\varepsilon_0}$; also by the property that $\overline{(B_{\varepsilon_0} \cap C)^0} = B_{\varepsilon_0} \cap C$, we have that there is $y \varepsilon S(x,r)$ such that $y \varepsilon (B_{\varepsilon_0} \cap C)^0$. Thus we now have $y \varepsilon (B_{\varepsilon_0} \cap C)^0 = B^0_{\varepsilon_0} \cap C^0$ and this y is also in $A^0_{\varepsilon_0}$ implying $y \varepsilon A^0_{\varepsilon_0} \cap B^0_{\varepsilon_0} \cap C^0$ which is a contradiction to the fact that $A^0_{\varepsilon_0} \cap B^0_{\varepsilon_0} \cap C = \phi$. Hence $A^0_{\varepsilon_0} \cap B_{\varepsilon_0} \cap C = \phi$, implying $A^0_{\varepsilon_0} \cap \partial B_{\varepsilon_0} \cap C = \phi$ since $\partial B_{\varepsilon_0} \subseteq B_{\varepsilon_0}$. Similarly, $\partial A_{\varepsilon_0} \cap B^0_{\varepsilon_0} \cap C = \phi$. This proves the lemma. ♣

In view of the above lemma, if ε_0 is the smallest ε such that $S_0 \cap S_1 \cap \mathcal{M}$ is nonempty, then we can find a point (c_0, c_1) in the common boundary of S_0 as well as S_1 which is also in the moment space $\mathcal{M}$. For this point (c_0, c_1), we then must have

$$\frac{c_2}{c_0} = \frac{1}{v_1} \tag{15}$$

$$\Rightarrow c_1^{1-r} = \frac{1}{v_1} c^r c_0 (1-c_0)^{-r}. \tag{16}$$

(Set both inequalities in (10) as equalities, divide, and then use Theorem 3.3).

Substituting (16) into (11) (with an equality in (11), one gets

$$\frac{1}{1+\varepsilon_0} = c_0 - c^{\frac{2r}{1-r}} v_1^{-\frac{1+r}{1-r}} c_0^{\frac{1+r}{1-r}} (1-c_0)^{-\frac{2r}{1-r}}. \tag{17}$$

Note that for this point $c_0 \neq 1$ (in fact it is also $\geq \frac{1}{1+\varepsilon_0}$). Also since this point (c_0, c_1) is in the moment space $\mathcal{M}$, we must have, by (16) and (8),

$$\frac{1}{v_1} c^r c_0 (1-c_0)^{-r} \geq \left(\frac{1-c_0}{c} \right)^{1-r}, \tag{18}$$

$$\text{and } \frac{1}{v_1} c^r c_0 (1-c_0)^{-r} \leq c_0^{\frac{1-r}{1+r}} \left(\frac{1-c_0}{c} \right)^{\frac{r(1-r)}{1+r}}. \tag{19}$$

On algebra, (18) and (19) reduce to

$$\frac{v_1}{c+v_1} \leq c_0 \leq \frac{v_1^{\frac{1+r}{2r}}}{c + v_1^{\frac{1+r}{2r}}}. \tag{20}$$

We have thus in effect proved that if ε_0 is the smallest ε for which $S_0 \cap S_1 \cap \mathcal{M}$ is nonempty, then there exists a c_0 satisfying (20) such that (17) holds. Conversely, if there is a c_0 satisfying (20) such that (17) holds for some given ε, then $S_0 \cap S_1 \cap \mathcal{M}$ is nonempty (in fact $\partial S_0 \cap \partial S_1 \cap \mathcal{M}$ is nonempty) for that ε. Therefore the smallest ε_0 can be found by maximizing the right side of (17), i.e.,

$$f(c_0) = c_0 - c^{\frac{2r}{1-r}} v_1^{-\frac{1+r}{1-r}} c_0^{\frac{1+r}{1-r}} (1-c_0)^{\frac{-2r}{1-r}}, \tag{21}$$

for c_0 satisfying (20). At this stage it is convenient to reparametrize to $z = \frac{c_0}{1-c_0}$ and $p = \frac{\lambda}{\lambda-2} = \frac{1+r}{1-r}$. We then have to maximize

$$h(z) = \frac{z - \frac{1}{c} \cdot \left(\frac{c}{v_1}\right)^p z^p}{1+z}, \tag{22}$$

in the interval $\frac{v_1}{c} \leq z \leq \frac{v_1^{\frac{1+r}{2r}}}{c}$.

To maximize h in the above interval, we take the derivative of $\log h$; algebra gives that the numerator in $(\log h)'$ is proportional to

$$N(z) = c \cdot \left(\frac{v_1}{c}\right)^p - pz^{p-1} - (p-1)z^p, \tag{23}$$

and the denominator in $(\log h)'$ is positive for z in the above interval. Clearly now, $N(z)$ is decreasing in z so that if $N(z) \leq 0$ at $z = \frac{v_1}{c}$ then it is ≤ 0 for all z in the above interval, implying that $\log h$ and hence h is decreasing and is therefore maximized at $z = \frac{v_1}{c}$. It is easy to check that $N(z) \leq 0$ at $z = \frac{v_1}{c}$ iff $p \geq \frac{v_1(1+c)}{v_1+c}$. Otherwise, there is a unique zero of $N(z)$ in the above interval and this is the unique maxima of $\log h$ and hence of h. Clearly, then, if $p < \frac{v_1(1+c)}{v_1+c}$, then the unique maxima of h is at the root of

$$pz^{p-1} + (p-1)z^p = c \cdot \left(\frac{v_1}{c}\right)^p. \tag{24}$$

Theorem 3.4 *Let* $\varepsilon_0 = \inf\{\varepsilon > 0 \colon \Gamma_\varepsilon \neq \phi\}$. *Then*

(i) $\varepsilon_0 = \frac{1+c}{v_1-1}$ *if* $p = \frac{\lambda}{\lambda-2} \geq \frac{v_1(1+c)}{v_1+c}$

(ii) $\varepsilon_0 = \frac{1-h(z_0)}{h(z_0)}$ *if* $p = \frac{\lambda}{\lambda-2} < \frac{v_1(1+c)}{v_1+c}$,

where z_0 *is the unique root of (24).*

Using the formulas for v_1 derived in Theorem 3.2 it is actually possible to get a better idea of which values of the pair (λ, c) imply $p \geq \frac{v_1(1+c)}{v_1+c}$. We omit these details.

It is interesting, however, that part (ii) of Theorem 3.4 can be much improved. In fact, once one finds the unique root z_0 of (24), it is possible to write down very convenient expressions for the maximin efficiency ε_0, the values of the moments c_0, c_1, and the two point design these c_0, c_1 correspond to. In effect, thus, it is possible to exactly write down what the two point maximin efficient design is. This is the assertion of the next theorem.

Theorem 3.5 *Let* $p = \frac{\lambda}{\lambda-2}$.

(i) Suppose $p \geq \frac{v_1(1+c)}{v_1+c}$. *Then*

$$\varepsilon_0 = \frac{1+c}{v_1-1}, \quad c_0 = \frac{v_1}{v_1+c}, \quad c_1 = \frac{1}{v_1+c}, \quad x_0 = 1, \quad \textit{and } p_0 = \frac{1+c}{v_1+c},$$

where x_0 *and* p_0 *denote the two point maximin efficient design.*

(ii) Suppose $p < \frac{v_1(1+c)}{v_1+c}$. *Then*

$$\begin{aligned} \varepsilon_0 &= \frac{p}{(p-1)z_0}, \quad c_0 = \frac{z_0}{1+z_0}, \\ c_1 &= c^{\frac{p-1}{2}} \cdot \frac{1}{1+z_0}\left(\frac{z_0}{v_1}\right)^{\frac{p+1}{2}}, \quad x_0 = \left(\frac{v_1}{cz_0}\right)^{\frac{p-1}{2}}, \quad \textit{and} \\ p_0 &= \frac{p}{v_1}\left(\frac{cz_0}{v_1}\right)^{p-1}, \end{aligned}$$

where z_0 *is the root of (24).*

Proof: Proof of part (ii): To get ε_0, just use $\frac{1}{1+\varepsilon_0} = h(z_0)$ and then use that z_0 solves (24) for part (ii). To get the expression for c_0, simply use the fact that $z_0 = \frac{c_0}{1-c_0}$. For c_1, use (16). For x_0, use (6), and for p_0 use $p_0 = 1 - c_0 + \frac{c_1}{x_0}$ (which is an implication of (4). ♣

Discussion : Of course, given λ and c, it will be easy to see using Theorem 3.2 whether case (i) or (ii) applies in Theorem 3.5. Then Theorem 3.5 provides a very convenient vehicle for finding the maximin efficiency and the required design. The case $\lambda = 2$ is of some special interest because this corresponds to regression on an ellipse. On the other hand, the case $\lambda \to \infty$ is of interest as the other limiting case. We briefly describe the nature of various things such as the maximin design, the maximin efficiency, etc. in these two cases.

Theorem 3.6 *(i) Let $w(x) = 1 + cx^2$, $c > 0$. Then (the) maximin efficient design is supported on 0 and 1 with mass $p_0 = \frac{\sqrt{1+c}}{2(\sqrt{1+c}+1)}$ at 1 and the maximin efficiency equals $\frac{\sqrt{1+c}+2}{2(\sqrt{1+c}+1)}$, which is monotone decreasing in c with a maximum possible value of $\frac{3}{4}$.*

(ii) Let $w(x) = 1 + cx^\lambda$, $c > 0$. As $\lambda \to \infty$, for every c (the) maximin efficient design converges to a design supported on 0 and 1 with mass $\frac{3}{4}$ and $\frac{1}{4}$ respectively. Also, the maximin efficiency converges to $\frac{3}{4}$ for every c.

Proof: Part (i) is a direct consequence of part (i) of Theorem 3.5 To prove part (ii), note that $p \to 1$ as $\lambda \to \infty$. Also, using part (ii) of Theorem 3.2 and the definition of p, it is easy to check that given any c, for large λ,

$$v_1 = c^{\frac{p-1}{p}} \cdot \frac{(p+1)^{\frac{p+1}{p}}}{(p-1)^{\frac{p-1}{p}}}. \tag{25}$$

Note that (25) converges to 4 as $p \to 1$. Thus, given any c, for large λ, case (ii) in Theorem 3.5 applies. Use now the fact that z_0 solves (24), or equivalently,

$$pz_0^{p-1} + (p-1)z_0^p = \frac{(p+1)^{p+1}}{(p-1)^{p-1}} \tag{26}$$

$$\Leftrightarrow \; pa_p^{p-1} + a_p^p = (p+1)^{p+1}, \tag{27}$$

where $a_p = (p-1)z_0$.

It is easy to show that $\{a_p\}$ is a bounded sequence (and is bounded away from zero) and therefore has a convergent subsequence. From (27) it now follows immediately that every convergent subsequence of a_p converges to 3. Hence, $\lim_{p \to 1}\{(p-1)z_0\} = 3$.

Now from part (ii) of Theorem 3.5, we have that x_0 converges to 1 as $p \to 1$, the mass (at 1) p_0 converges to $\frac{1}{4}$ as $p \to 1$ and $\varepsilon_0 \to \frac{1}{3}$ as $p \to 1$ and hence the maximin efficiency converges to $\frac{3}{4}$. This proves the theorem. ♣

Discussion : It is interesting to note that as $\lambda \to 2$ or ∞, the maximin efficient design and the maximin efficiency behave similarly. Indeed, for $c \to 0$, the designs in parts (i) and (ii) of the above theorem are exactly the same and so are the maximin efficiencies.

3.3 Example.

Proceeding as in the preceding section, it is possible to derive a maximin efficient Bayes design and the corresponding maximin efficiency provided one has a diagonal prior covariance matrix for the regression vector $\underset{\sim}{\theta} = (\theta_0, \theta_1)'$. We do not give this details. Table 3 below gives these quantities assuming a $N(\underset{\sim}{0}, I)$ prior. Table 4 gives the ordinary A-optimal design assuming the same prior and its minimum efficiency over the coordinates. In Table 3, $\{x_1, x_2\}$ constitutes the support of the design, p denotes the mass at x_2, r denotes $\frac{1}{n}$ and e stands for the maximin efficiency. In Table 4, the same notation is used for the A-optimal design and e^* denotes its minimum efficiency.

Discussion of Table 3 : The most interesting feature of Table 3 is the remarkable robustness of the maximin efficiency e to the choice of λ for any fixed value of c. Indeed, in some cases, the design and the maximin efficiency e remain the same for a wide range of λ for fixed c. Also notice that similar robustness seems to hold over the choice of c for fixed λ as well, with the exception of the case $c = 0$ which corresponds to the homoscedastic case.

Notice that in this case the value of e generally seems to be higher, indicating a penalty for lack of homoscedasticity. As in Table 1, if one fixes c and λ, then e seems to decrease as n increases. In general, the value of e is moderate for small to moderate sample sizes.

Discussion of Table 4 : Again, the general features of Table 3 are present in Table 4 too. As before, the interesting question is whether it is worthwhile to derive the maximin efficient design. Even a cursory glance of Table 3 and 4 evidently indicate that virtually for any c, λ and n, the maximin efficient design provides quite major gain in the minimum efficiency over the standard A-optimal design. This was not the case for the variance function considered in section 2.2.

4 Conclusions.

The results in this article pertain to the question of simultaneous optimization of the design when interest lies in more than one specific problem and one wants to work with a vector loss instead of collapsing the different problems into a single problem by taking a sum of the coordinate wise losses. The analysis seems to indicate that for general polynomial regression it may be hard to derive the maximin efficient designs for heteroscedastic models. A prior can further complicate the situation. Bayesian results have been emphasized in this article while keeping in mind that their frequentist analogs are often easier to derive. We consider these important from a practical viewpoint and hope that compromise designs will be emphasized in other contexts as well.

Acknowledgement. We are grateful to Mr. Jaime San Martin for giving a very neat proof of Lemma 3.2 and to Mr. F.C. Chang for help with the numerical work.

Table A

p	θ_i's of interest	Maximin efficiecy	p	θ_i's of interest	Maximin efficiecy
2	$\{\theta_0, \theta_1\}$	1.0	5	$\{\theta_1, \theta_3\}$	0.933012
3	$\{\theta_0, \theta_1\}$	0.5		$\{\theta_1, \theta_4\}$	0.735753
	$\{\theta_0, \theta_2\}$	0.75		$\{\theta_2, \theta_3\}$	0.681868
	$\{\theta_1, \theta_2\}$	0.75		$\{\theta_2, \theta_4\}$	0.968019
	$\{\theta_0, \theta_1, \theta_2\}$	0.5		$\{\theta_3, \theta_4\}$	0.75
4	$\{\theta_0, \theta_1\}$	0.666472		$\{\theta_0, \theta_1, \theta_2\}$	0.489838
	$\{\theta_0, \theta_2\}$	0.75		$\{\theta_0, \theta_1, \theta_3\}$	0.482620
	$\{\theta_0, \theta_3\}$	0.671464		$\{\theta_0, \theta_1, \theta_4\}$	0.498835
	$\{\theta_1, \theta_2\}$	0.648998		$\{\theta_0, \theta_2, \theta_3\}$	0.497546
	$\{\theta_1, \theta_3\}$	0.934186		$\{\theta_0, \theta_2, \theta_4\}$	0.632240
	$\{\theta 2, \theta_3\}$	0.75		$\{\theta_0, \theta_3, \theta_4\}$	0.497850
	$\{\theta_0, \theta_1, \theta_2\}$	0.582961		$\{\theta_1, \theta_2, \theta_3\}$	0.664791
	$\{\theta_0, \theta_1, \theta_3\}$	0.656609		$\{\theta_1, \theta_2, \theta_4\}$	0.668603
	$\{\theta_0, \theta_2, \theta_3\}$	0.632688		$\{\theta_1, \theta_3, \theta_4\}$	0.722619
	$\{\theta_1, \theta_2, \theta_3\}$	0.648998		$\{\theta_2, \theta_3, \theta_4\}$	0.681868
	$\{\theta_0, \theta_1, \theta_2, \theta_3\}$	0.582961		$\{\theta_0, \theta_1, \theta_2, \theta_3\}$	0.482620
5	$\{\theta_0, \theta_1\}$	0.489838		$\{\theta_0, \theta_1, \theta_2, \theta_4\}$	0.498835
	$\{\theta_0, \theta_2\}$	0.692790		$\{\theta_0, \theta_1, \theta_3, \theta_4\}$	0.482620
	$\{\theta_0, \theta_3\}$	0.499669		$\{\theta_0, \theta_2, \theta_3, \theta_4\}$	0.497546
	$\{\theta_0, \theta_4\}$	0.632240		$\{\theta_1, \theta_2, \theta_3, \theta_4\}$	0.664791
	$\{\theta_1, \theta_2\}$	0.668603		$\{\theta_0, \theta_1, \theta_2, \theta_3, \theta_4\}$	0.482620

Table 1

λ	r	x_1	x_2	p_1	p_2	e
0	0.2	0	1	0	0.217	0.876
	0.1	0	1	0	0.215	0.842
	0.04	0	1	0	0.259	0.769
	0.01	0	1	0	0.461	0.548
1	0.2	0	0.979	0	0.234	0.869
	0.1	0	0.988	0	0.220	0.847
	0.04	0	0.975	0	0.225	0.813
	0.01	0	0.931	0	0.329	0.686
1.5	0.2	0	0.628	0	0.276	0.910
	0.1	0	0.709	0	0.232	0.884
	0.04	0	0.768	0	0.226	0.843
	0.01	0	0.995	0	0.257	0.770
2	0.2	0	0.354	0	0.415	0.955
	0.1	0	0.390	0	0.346	0.933
	0.04	0	0.436	0	0.281	0.904
	0.01	0	0.498	0	0.249	0.848
2.5	0.2	0.097	0.246	0.00007	0.430	0.986
	0.1	0.043	0.252	0.0008	0.424	0.974
	0.04	0.275	0.284	0.00003	0.357	0.952
	0.01	0	0.361	0	0.263	0.908
3	0.2	0.099	0.127	0.078	0.769	0.997
	0.1	0	0.141	0	0.680	0.994
	0.04	0	0.177	0	0.464	0.986
	0.01	0	0.237	0	0.304	0.961

Table 2

λ	r	x_1	x_2	p_1	p_2	e^*
0	0.2	0	1	0	0.4	0.733
	0.1	0.423	1	0.163	0.345	0.652
	0.04	0.478	1	0.344	0.269	0.534
	0.01	0.498	1	0.455	0.210	0.413
1	0.2	0	1	0	0.497	0.692
	0.1	0	1	0	0.492	0.626
	0.04	0.436	1	0.175	0.416	0.541
	0.01	0.454	1	0.375	0.300	0.424
1.5	0.2	0.551	0.683	0	0.451	0.826
	0.1	0.748	1	0.496	0	0.714
	0.04	0.798	1	0.535	0	0.585
	0.01	0.425	0.920	0.219	0.442	0.439
2	0.2	0.306	1	0.163	0	0.941
	0.1	0.462	1	0.327	0	0.917
	0.04	0	0.579	0	0.467	0.754
	0.01	0.625	0.657	0	0.567	0.536
2.5	0.2	0.00008	1	0.2	0	0.981
	0.1	0	1	0	0	0.966
	0.04	0.25	1	0.137	0	0.939
	0.01	0.482	1	0.463	0	0.747
3	0.2	0	1	0	0	0.996
	0.1	0	1	0	0	0.993
	0.04	0	1	0	0	0.983
	0.01	0	1	0	0	0.949

Table 3

c	λ	r	x_1	x_2	p	e	c	λ	r	x_1	x_2	p	e
0	-	0.2	0	1	0.16667	0.93687	1	2.5	0.2	0	1	0.30770	0.81605
		0.1	0	1	0.09091	0.96065			0.1	0	1	0.29898	0.78267
		0.04	0	1	0.03846	0.98187			0.04	0	1	0.29493	0.74667
		0.01	0	1	0.25250	0.75952			0.01	0	1	0.29334	0.71883
0.5	2	0.2	0	1	0.30707	0.82842		5	0.2	0	0.88338	0.32904	0.82297
		0.1	0	1	0.29078	0.79563			0.1	0	0.92154	0.30689	0.78633
		0.04	0	1	0.28135	0.76117			0.04	0	0.95614	0.29668	0.74798
		0.01	0	1	0.27676	0.73537			0.01	0	0.97929	0.29343	0.71915
	2.5	0.2	0	1	0.30707	0.82842	2	2	0.2	0	1	0.30120	0.80470
		0.1	0	1	0.29078	0.79563			0.1	0	1	0.30496	0.76859
		0.04	0	1	0.28135	0.76117			0.04	0	1	0.31081	0.72889
		0.01	0	1	0.27676	0.73537			0.01	0	1	0.31523	0.69693
	5	0.2	0	1	0.30707	0.82842		2.5	0.2	0	1	0.30120	0.80470
		0.1	0	1	0.29078	0.79563			0.1	0	1	0.30496	0.76859
		0.04	0	1	0.28135	0.76117			0.04	0	1	0.31081	0.72889
		0.01	0	1	0.27676	0.73537			0.01	0	1	0.31523	0.69693
1	2	0.2	0	1	0.30770	0.81605		5	0.2	0	0.75471	0.33452	0.83537
		0.1	0	1	0.29898	0.78267			0.1	0	0.79212	0.31206	0.79551
		0.04	0	1	0.29493	0.74667			0.04	0	0.82762	0.30264	0.75081
		0.01	0	1	0.29334	0.71883			0.01	0	0.85241	0.30067	0.71485

Table 4

c	λ	r	x_1	x_2	p	e	c	λ	r	x_1	x_2	p	$e^{\star}$
0	-	0.2	0	1	0.44391	0.74497	1	2.5	0.2	0	1	0.5	0.67593
		0.1	0	1	0.42788	0.68470			0.1	0	1	0.5	0.61039
		0.04	0	1	0.41933	0.63191			0.04	0	1	0.5	0.55239
		0.01	0	1	0.41544	0.59834			0.01	0	1	0.5	0.51447
0.5	2	0.2	0	1	0.47974	0.70276		5	0.2	0	0.97464	0.49774	0.68296
		0.1	0	1	0.47124	0.64056			0.1	0	1	0.5	0.61039
		0.04	0	1	0.46675	0.58528			0.04	0	1	0.5	0.55239
		0.01	0	1	0.46473	0.54946			0.01	0	1	0.5	0.51447
	2.5	0.2	0	1	0.47974	0.70276	2	2	0.2	0	1	0.51761	0.64585
		0.1	0	1	0.47124	0.64056			0.1	0	1	0.53569	0.57128
		0.04	0	1	0.46675	0.58528			0.04	0	1	0.54512	0.50737
		0.01	0	1	0.46473	0.54946			0.01	0	1	0.54924	0.46558
	5	0.2	0	1	0.47974	0.70276		2.5	0.2	0	1	0.51761	0.64585
		0.1	0	1	0.47124	0.64056			0.1	0	1	0.53569	0.57128
		0.04	0	1	0.46675	0.58528			0.04	0	1	0.54512	0.50737
		0.01	0	1	0.46473	0.54946			0.01	0	1	0.54924	0.46558
1	2	0.2	0	1	0.5	0.67593		5	0.2	0	0.84199	0.50398	0.69562
		0.1	0	1	0.5	0.61039			0.1	0	0.88485	0.51436	0.60891
		0.04	0	1	0.5	0.55239			0.04	0	0.91410	0.52425	0.53402
		0.01	0	1	0.5	0.51447			0.01	0	0.92938	0.53016	0.48598

References

[1] Bickel, P.J. (1978), Using residuals robustly I: tests for heteroscedasticity, nonlinearity, *Ann. Stat.*, **6**, pp. 266–291.

[2] Box, G.E.P. and Hill, W.J. (1974), Correcting inhomogeneity of variance with power transformation weighting, *Technometrics*, **16**, pp. 385–389.

[3] Carroll, R.J. and Ruppert, D. (1982), Robust estimation in heteroscedastic linear models, *Ann. Stat.*, **10**, pp. 429–441.

[4] Chaloner, Kathryn (1987), Optimal Bayesian experimental design for linear models, *Ann. Stat.*, **12**, pp. 283–300.

[5] Cheng, C.S. (1987), An optimization problem with applications to optimal design theory, *Ann. Stat.*, **15**, pp. 712–723.

[6] Cohen, A. and Sackrowitz, H.D. (1984) Decision theory results for vector risks with applications, in *Stat. and Decisions Suppl. no. 1: Recent results in estimation theory and related topics*, pp. 150–176 Oldenbourg,Munich.

[7] Cook, R.D. (1986), Assessment of locla influence (with discussion)., *J.R.S.S.(ser-B)*, **48**, pp. 133-169.

[8] DasGupta, A. and Studden, W.J. (1991), Robust Bayesian Analysis and optimal experimental designs in normal linear models with many parameters - I, To appear in *Ann. Stat.*

[9] Dette, H. (1991), Robust designs for multivariate polynomial regression on the d-cube, Submitted to *J. Mult. Annals.*

[10] Donoho, D., Liu, R., and MacGibbon, B. (1990), Minimax risk and hyperrectangles, *Ann. Stat.*, **18**, pp. 1416-1437.

[11] Efron, B. and Morris, C. (1971), Limiting the risk of Bayes and empirical Bayes estimators – Part I: The Bayes Case, *J. Amer. Stat. Assoc.*, **66**, pp. 807–815.

[12] Efron, B. and Morris, C. (1971), Limiting the risk of Bayes and empirical Bayes estimators – Part II: The empirical Bayes case, *J. Amer. Stat. Assoc.*, **67**, pp. 130–139.

[13] Elfving, G. (1952), Optimum allocation in linear regression theory, *Ann. Math. Stat.*, **23**, pp. 255–262.

[14] Fedorov, V.V. (1972), *Theory of optimal experiments*,Trans. and ed. W.J. Studden and E.M. Klimko: Academic Press, New York.

[15] Hoel, P.G. (1958), Efficiency problems in polynomial estimation, *Ann. Math. Stat.*, **29**, pp. 1134–1145.

[16] Jobson, J.D. and Fuller, W.A. (1980), Least squares estimation when the covariance matrix and parameter vector are functionally related, *J. Amer. Stat. Assoc.*, **75**, pp. 176–181.

[17] Karlin, S. and Studden, W.J. (1966), Optimal experimental designs, *Ann. Math. Stat.*, **37**, pp. 783–815.

[18] Kiefer, J. (1974), General equivalence theory for optimum designs, *Ann. Stat.*, **2**, pp. 849–879.

[19] Kiefer, J. and Studden, W.J. (1976), Optimal designs for large degree polynomial regression, *Ann. Stat.*, **4**, pp. 1113–1123.

[20] Lau, T.S. (1983), Theory of canonical moments and its application in polynomial regression, I and II, , Technical Report, **83-23 and 83-24**, Department of Statistics, Purdue University.

[21] Lau, T.S. (1988), D-optimal designs on the unit q-ball, *J. Stat. Plan. Inf.*, **19**, pp. 299-315.

[22] Leamer, E.E. (1978), in *Specification searches*, Wiley, New York

[23] Lee, C–M.S. (1988), Constrained optimal designs, *J. Planning and Inf.*, **18**, pp. 377–389.

[24] Mukhopadhyay, S. and DasGupta, A. (1991), Compromise designs and robustness, Technical Report, Dept. of Statistics, Purdue University.

[25] Pilz, J (1979), Das bayes she Schätzproblem im linearen Regressionmodell *Freiberger Forschungshefte*, **117**, pp. 21–55.

[26] Pilz, J (1981), Robust Bayes and minimax–Bayes estimation and design in linear regression, *Math. Operations forsch. Stat., ser Statistics*, **12**, pp. 163-177.

[27] Pilz, J. (1984), Robust Bayes regression estimation under weak prior knowledge, in *Robustness of Stat. Methods and Nonpar. Stat.*,Rasch, D. and Tiku, M.L. (Eds.): Deutscher Verlag, Berlin.

[28] Polasek, W. (1984), Multivariate regression systems: estimation andsensitivity analysis of two–dimensional data, in *Robustness of Bayesian Analyses*,J. Kande(Ed.): North–Holland, Amsterdam.

[29] Pukelsheim, F. and Titterington, D.M. (1983), General differential and Lagrangian theory for optimal experimental designs, *Ann. Stat.*, **11**, pp. 1060–1068.

[30] Rao, C.R. (1976), Estimation of parametes in a linear model, *Ann. Stat.*, **4**, pp. 1023–1037.

[31] Rao, C.R. (1977), Simultaneous estimation of parameters – a compound ecision problem, in *Stat. Dec. Theory and Related Topics II*,S.S. Gupta and D.S. Moore (Eds.): Academic Press, New York.

[32] Skibinsky, M. (1967), The range of the (n+1)-th moment for distributions on [0,1], *J. Appl. Probability*, **4**, pp. 543-552.

[33] Skibinsky, M. (1968), Extreme n-th moments for distributions on [0,1] and the inverse of a moment space map, *J. Appl. Probability*, **5**, pp. 693-701.

[34] Skibinsky, M. (1969), Some striking properties of binomial and beta moments, *Ann. Math. Stat.*, **40**, pp. 1753-1764.

[35] Skibinsky, M. (1986), Principal representations and canonical moment sequences for distribution on an interval, *J. Math. Anal. and Applications*, **120**, pp. 95-120.

[36] Studden, W.J. (1980), D_s-optimal design for polynomial regression using continued fractions, *Ann. Stat.*, **8**, pp. 1132-1141.

[37] Studden, W.J. (1982), Some robust type D-optimal designs in polynomial regression, *J. American Stat. Assoc.*, **77**, pp. 916-921.

[38] Studden, W.J. (1982a), Optimal designs for weighted polynomial regression using canonical moments, in *Statistical Decision Theory and Related Topics III*,, pp. 335-350, Academoc Press, New York.

[39] Studden, W.J. (1989), Note on some Φ_p optimal design for polynomial regression, *Ann. Stat.*, **17**, pp. 618-623.

[40] Vidacovic, B. and DasGupta, A. (1991), Linear vs. nonlinear rules: estimation of multivariate normal mean, Technical Report # 91–45, Departmen of Statistics, Purdue University.