TOWARDS A THEORY OF COMPROMISE DESIGNS: FREQUENTIST, BAYES, AND ROBUST BAYES

by

Anirban DasGupta * and William J. Studden **
Purdue University

Technical Report #89-20

Department of Statistics
Purdue University

June 1989

* Research supported by NSF Grant DMS 87–02620
** Research supported by NSF Grant DMS 88–02535

★

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 two 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. We also address the problem of Bayes sensitivity (to the prior) by varying the prior in a suitable family. We give a complete description of the two dimensional set of pairs of Bayes risks as the prior changes in the specified family. Sensitivity is measured in terms of the Lebesgue measure of this two dimensional set. We then address the problem of deriving a design that minimizes this Lebesgue measure among the designs which have an efficiency of at least $\frac{1}{1+\varepsilon}$ in each estimation problem.

1. Introduction. 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 $\mathrm{Var}(y|x_i)$ 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. We will consider the case of a single prior as well as the case when the prior can be specified only upto a family. This latter problem constitutes a part of what has come to be known as robustness in Bayesian inference with respect to the prior. Before we describe the projects undertaken in this article and outline the subsequent sections, we would like to point out that even though classical optimal designs have been studied in extremely great detail, in comparison the area of Bayes designs remains broadly unexplored. Part of the reason is the apparent difficulty in using convexity arguments; even when convexity arguments can be used, the optimization 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 not feasible and we have not attempted them. For general results on Bayes and robust Bayes designs, see Pilz (1979,1981), Chaloner (1984), DasGupta and Studden (1988) etc. For general references, see Cheng (1987), Hoel (1958), Karlin and Studden (1966), Kiefer and Studden (1976), Elfving (1952), Fedorov (1972), Kiefer (1976), Pukelsheim and Titterington (1983) etc.

In section 2, we first consider the heteroscedastic model where $E(Y) = X\underset{\sim}{\theta}$ and the variance–covariance matrix of Y is a diagonal matrix $\Sigma(\underset{\sim}{\theta})$. $\underset{\sim}{\theta}$ is assumed to have a prior distribution $G(\underset{\sim}{\theta})$. Normality is not assumed for either $\underset{\sim}{Y}$ or $\underset{\sim}{\theta}$. Best linear Bayes estimation of $\underset{\sim}{\theta}$ and the corresponding Bayes optimal designs are considered and it is pointed out that restricting to linear estimates in a general nonnormal model is formally equivalent to using the overall Bayes estimate in a normal model. This connection is actually quite well known but provides a motivation for some of the latter results.

Next in the same section we specialize to the canonical simple linear regression problem where $E(y|x) = \theta_0 + \theta_1 x$; it is assumed that $\text{Var}(y|x) = w(x)$ is of the form $1 + c|x|^\lambda$, $c > 0$, and $\lambda \geq 2$; the standard homoscedastic case corresponds to $c = 0$ and the results for this case can be obtained by letting $c \to 0$. The cases $0 \leq x \leq 1$, and $-1 \leq x \leq 1$ are considered. The case $\lambda < 2$ in general gives rise to different qualitative features and will not be considered in this article. Note that the present variance function resembles the power models described before. Exponential variance functions of the form $e^{\lambda x}$ or $e^{\lambda|x|}$ are also quite natural but have not been considered for reasons of space and compactness. We let $\underset{\sim}{\theta} = (\theta_0, \theta_1)'$ have a $N(\underset{\sim}{\mu}, C)$ distribution where C is diagonal. The case of a nondiagonal C is in fact possible to handle but gives rise to many more cases and generates more complexity than generality. Bayes designs for these priors are derived for both θ_0 and θ_1; the solution in the classical case can be obtained as a limiting Bayes solution by formally using $C^{-1} = 0$ (the null matrix). A surprising finding is that if $0 \leq x \leq 1$, then the Bayes optimal designs for the slope θ_1 are supported on 0 and 1 for sufficiently small λ, and on

0 and a suitable interior point (depending on λ) otherwise. However, the interior point starts moving back towards 1 for sufficiently large λ and eventually converges to 1 again.

A common feature of optimal design theory (classical as well as Bayes) is that the derived optimal designs are extremely problem–specific in addition to being model–specific. For instance, if a classical homoscedastic linear regression for $0 \leq x \leq 1$ is considered, the optimal design for the intercept θ_0 is supported on 0 and that for the slope θ_1 puts equal mass at 0 and 1. The optimal design for the slope results in an 100% increase in risk over the minimum value if used to estimate the intercept. The optimal design for the intercept in fact results in infinite risk if used for estimating the slope. Frequently, however, the experimenter is not interested in one specific problem but is probably interested in two or more problems. Towards this end, we let v_i, $i = 0, 1$, denote the minimum Bayes risks obtainable for estimating θ_i, $i = 0, 1$ respectively, by using the corresponding Bayes optimal designs. For any other arbitrary design ξ, let $v_i(\xi)$ denote the Bayes risks in the two problems by using the design ξ. Efficiency of ξ in problem i is defined as $\frac{v_i}{v_i(\xi)}$. We consider designs that maximize the minimum efficiency in the two problems. Formally, one can define $\Gamma_\varepsilon(C) = \{\xi : v_i(\xi) \leq (1+\varepsilon)v_i,\ i = 0, 1\}$. For small $\varepsilon > 0$, $\Gamma_\varepsilon(C)$ will usually be empty. Of natural interest is the value of the smallest $\varepsilon > 0$ such that $\Gamma_\varepsilon(C)$ is nonempty. We have a number of results on this smallest ε and how it depends on the variance function $w(X)$. Note that similar "compromise designs" are of interest in classical design theory also. Indeed, if one considers a canonical homoscedastic polynomial regression model $E(y_i) = \theta_0 + \theta_1 x_i + \theta_2 x_i^2 + \ldots + \theta_p x_i^p$, one can ask what is the smallest ε such that

$$\Gamma_\varepsilon = \{\xi : v_i(\xi) \leq (1+\varepsilon)v_i,\ i = 0, 1, \ldots, p\}$$

is nonempty. We in fact do touch on this problem as well. The indication seems to be that for this smallest ε, the first p inequalities $v_i(\xi) \leq (1+\varepsilon)v_i$ are equalities and the last inequality remains a strict inequality. These results are presented in section 3 for $0 \leq x \leq 1$ and in section 4 for $-1 \leq x \leq 1$. Notice that the importance of performing well on each of a few decision problems instead of collapsing into a single sum of risks has been emphasized by various authors in other contexts: see Rao (1976,1977), Efron and

Morris (1971,1972), Dey and Berger (1983), Stein (1981), etc. Also see Lee (1988) for some related work on constrained designs.

The question of robust Bayes designs and many priors is taken up in section 5. Some of the results we present in this section are of interest beyond the domain of designs. We consider the case when $\underset{\sim}{Y} \sim N(X\underset{\sim}{\theta}, \Sigma)$ (Σ does not depend on $\underset{\sim}{\theta}$, but may depend on X), and $\underset{\sim}{\theta}$ has a prior $N(\underset{\sim}{\mu}, C)$, $C_1 \leq C \leq C_2$, and μ belongs to some (convex) set S. This family of priors was first considered in Leamer (1978), and Polasek (1984); also see DasGupta and Studden (1988) for a variety of results. Suppose interest lies in estimating any two linear combinations $\underset{\sim}{\ell_1}'\underset{\sim}{\theta}$, $\underset{\sim}{\ell_2}'\underset{\sim}{\theta}$. The Bayes risk in each problem depends on the exact choice of the prior covariance matrix C; an experimenter interested simultaneously in both linear combinations would then want to know the two dimensional set of pairs of Bayes risks as the prior $N(\underset{\sim}{\mu}, C)$ changes. We characterize this set completely and derive an exact expression for its area (i.e., the Lebesgue measure). The robustness problem is to minimize sensitivity at a small cost in Bayes optimality with respect to a fixed prior. In the spirit of the results in the earlier sections, we then go back to a heteroscedastic simple linear regression problem and derive the design ξ_0 that minimizes the area of the two–dimensional set of Bayes risks in the class of designs $\Gamma_\varepsilon(C_0)$ where C_0 is a fixed matrix, $C_1 \leq C_0 \leq C_2$. This is done for $\underset{\sim}{\ell_1} = (1,0)'$ and $\underset{\sim}{\ell_2} = (0,1)'$, i.e., for estimating the intercept and the slope. The problem of model specificness of optimal designs was mentioned before in this section. We show, somewhat surprisingly, that the design ξ_0 is not unique and one can then choose an appropriate ξ_0 keeping model robustness in mind. Some alternative possible formulations of the robustness problem are also briefly discussed. Section 6 contains some concluding remarks. We hope that our results on "compromise designs" and Bayes sensitivity in heteroscedastic models will be useful for further research in this obviously important area of application.

2. Bayes and classical optimal designs in heteroscedastic models. 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 later sections. Through-

out we assume squared error loss.

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

Theorem 2.1. Let $\underset{\sim}{Y}_{n\times 1} \sim (X\underset{\sim}{\theta}, \Sigma(\underset{\sim}{\theta}))$ and let $\underset{\sim}{\theta}_{p\times 1} \sim (\underset{\sim}{\mu}, C)$. 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 $A_0(Y - X\mu) + \mu$ has $A_0 = LCX'(B + XCX')^{-1}$ and the corresponding Bayes risk equals $tr\ L(X'B^{-1}X + C^{-1})^{-1}L'$.

Proof: By straightforward calculations, the Bayes risk of any linear estimate $A\underset{\sim}{Y}$ under the prior $(\underset{\sim}{\mu}, C)$ equals

$$r(A, \underset{\sim}{\mu}, C) = tr\ APA' - 2tr\ AQ + tr\ LCL', \tag{2.1}$$

where $P = B + XCX'$, and $Q = XCL'$. This is minimized by $A_0 = Q'P^{-1} = LCX'(B + XCX')^{-1}$. Hence the corresponding Bayes risk equals

$$r(A_0, \underset{\sim}{\mu}, C) = tr\ LCL' - tr\ LCX'(B + XCX')^{-1}XCL'. \tag{2.2}$$

Using the identity $(B + XCX')^{-1} = B^{-1} - B^{-1}X(X'B^{-1}X + C^{-1})^{-1}X'B^{-1}$ (see, for example, Rao (1973), page 33), one has

$$\begin{aligned} r(A_0, \underset{\sim}{\mu}, C) &= tr\ LCL' - tr\ LCX'B^{-1}XCL' + tr\ LCX'B^{-1}X(X'B^{-1}X + C^{-1})^{-1} \\ &\qquad X'B^{-1}XCL' \\ &= tr\ LCL' - tr\ LCX'B^{-1}XCL' + tr\ LC(X'B^{-1}X + C^{-1} - C^{-1})(X'B^{-1}X + C^{-1})^{-1} \\ &\qquad X'B^{-1}XCL' \\ &= tr\ LCL' - tr\ L(X'B^{-1}X + C^{-1})^{-1}X'B^{-1}XCL' \\ &= tr\ LCL' - tr\ L(X'B^{-1}X + C^{-1})^{-1}(X'B^{-1}X + C^{-1} - C^{-1})CL' \\ &= tr\ L(X'B^{-1}X + C^{-1})^{-1}L'. \end{aligned} \tag{2.3}$$

In fact, the estimate in Theorem 2.1 is the best Bayes estimate of $\underset{\sim}{\theta}$ of the form $A_0\underset{\sim}{Y} + A_1\underset{\sim}{\mu}$. Notice the formal equivalence of restricting to linear estimates and assuming that $Y \sim N(X\theta, B)$ and $\theta \sim N(\mu, C)$. If $Y \sim N(X\theta, B)$ and θ 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 $tr\ L(X'B^{-1}X+C^{-1})^{-1}L'$. Thus if restriction to linear estimates is considered undesirable, one can alternatively take the viewpoint of adopting a normal model with a normal prior.

Suppose now $y_1, \ldots, y_n$ are independent observations with $\text{Var}(y_i) = e^{\lambda \underset{\sim}{X}_i' \underset{\sim}{\theta}}$ where $\underset{\sim}{X}_i'\underset{\sim}{\theta}$ is the mean of y_i. If one, for example, assumes that $\underset{\sim}{\theta}$ has a $N(\underset{\sim}{0}, I)$ prior, then the matrix B can be easily seen to be

$$B = \text{diag}(b_1, \ldots, b_n)$$

where $b_i = e^{\frac{\lambda^2}{2} \underset{\sim}{X}_i' \underset{\sim}{X}_i}$. In particular, for polynomial regression of degree p on the interval [0,1], $b_i = e^{\frac{\lambda^2}{2}}(1 + x_i^2 + x_i^4 + \ldots + x_i^{2p})$. Optimal design theory for such a matrix B is simplified due to the fact that it is possible to prove that the set of designs supported on 0 and at most p other points forms a complete class. Indeed, a general theorem in this direction is the following.

<u>Theorem 2.3</u>. Let $E(y_i|x_i) = \theta_0 + \theta_1 x_i + \ldots + \theta_p x_i^p$, where $0 \le x_i \le 1$, and let $\text{Var}(y_i|x_i) = e^{\lambda E(y_i|x_i)}$. Let $\underset{\sim}{\theta} = (\theta_0, \theta_1, \ldots, \theta_p)'$ have a spherically symmetric prior G. 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 $M(\xi_2) \ge M(\xi_1)$ where for any design, M denotes $X'B^{-1}X$.

Thus for quadratic regression, this result reduces the problem of deriving an optimal design to a four parameter minimization. As opposed to the exponential model $e^{\lambda E(y_i)}$, if one considers the power model $(E(y_i))^2$ for variance, then for a simple linear regression model and a $N(\underset{\sim}{0}, I)$ prior, B works out to

$$B = \text{diag}(1 + x_1^2, \ldots, 1 + x_n^2).$$

In view of the remark following (2.3), restricting to linear estimates in Theorem 2.1 is equivalent to having $\underset{\sim}{Y} \sim N(X\underset{\sim}{\theta}, B)$ and $\underset{\sim}{\theta} \sim N(\underset{\sim}{\mu}, C)$. A more general variance function to consider is $w(x) = 1 + cx^\lambda$, $c > 0$, $\lambda \ge 2$. This will imply

$$B = \text{diag}(1 + cx_1^\lambda, \ldots, 1 + cx_n^\lambda).$$

These are the variance functions we now consider. For the following analysis we will assume $y_i \overset{\text{indep}}{\sim} N(\theta_0 + \theta_1 x_i,\ 1 + c x_i^\lambda)$, $0 \le x_i \le 1$, and $\underset{\sim}{\theta} = (\theta_0, \theta_1)' \sim N(\underset{\sim}{\mu}, C)$ where $\frac{C^{-1}}{n} = \begin{pmatrix} r_0 & 0 \\ 0 & r_2 \end{pmatrix}$ for $r_0,\ r_2 \ge 0$ (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 2.3. Let $E(y_i) = \theta_0 + \theta_1 x_i$, $0 \le x_i \le 1$ and let $\text{Var}(y_i) = 1 + c x_i^\lambda$, $\lambda \ge 2$. Assume $\underset{\sim}{\theta}$ has a $N(\underset{\sim}{\mu}, C)$ 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:

(i) If $c(\lambda - 2) \le 2$ or $c(\lambda - 2) > 2$ and $c \le \frac{4(\lambda-1)}{(\lambda-2)^2}$, then the design is supported on 0 and 1 with mass $p = \frac{(1+r_0)\sqrt{1+c}(\sqrt{1+c}-1)}{c}$ at 1 if $\frac{c}{1+c} \le 1 - r_0^2$, on only 1 if $\frac{c}{1+c} > 1 - r_0^2$ and $c(\lambda - 2) \le \frac{2(1+r_0)}{r_0}$, and on only $X_1 = \left(\frac{2(1+r_0)}{c r_0 (\lambda-2)}\right)^{\frac{1}{\lambda}}$ if $\frac{c}{1+c} > 1 - r_0^2$ and $c(\lambda - 2) > \frac{2(1+r_0)}{r_0}$.

(ii) If $c > \frac{4(\lambda-1)}{(\lambda-2)^2}$ (which implies $c(\lambda-2) > 2$), then the design is supported on only X_1 (defined above) if $\lambda(1 - r_0) \le 2$ and on 0 and $X_0 = \left(\frac{4(\lambda-1)}{c(\lambda-2)^2}\right)^{\frac{1}{\lambda}}$ if $\lambda(1 - r_0) > 2$, with mass $p = \frac{\lambda(1+r_0)}{2(\lambda-1)}$ at X_0.

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 $0 \le x \le 1$) the inequality $((c_0 + r_0)x - c_1)^2 \le (1 + c x^\lambda) \cdot Q$, 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 $E\frac{1}{w(X)}$ and $E\frac{X}{w(X)}$ as given by the designs in the statements of the theorem.

Remark: For a nondiagonal prior covariance matrix C, the optimal design for θ_1 cannot in general be written down in a closed form.

Discussion of Theorem 2.3. 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 2.3 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. Maximin efficient designs: $0 \leq x \leq 1$. 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 $\lambda \geq 2$, 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. For the purpose of the following analysis we will assume that we have a simple linear regression with the independent variable varying in [0,1] and a variance function $w(x) = 1 + cx^\lambda$, $\lambda \geq 2$, $c > 0$. Also, for the classical case, $\Gamma_\varepsilon(C)$ will be denoted as simply Γ_ε.

Theorem 3.1. Let $r_0 = r_2 = 0$. Then

(i) $v_0 = 1$

(ii)

$$v_1 = (\sqrt{1+c}+1)^2 \text{ if } \lambda \le 2\sqrt{\frac{1+c}{c}}\left(\sqrt{\frac{1+c}{c}}+1\right)$$
$$= c^{\frac{2}{\lambda}} \cdot 4^{1-\frac{2}{\lambda}} \cdot (\lambda-1)^{2-\frac{2}{\lambda}} \cdot (\lambda-2)^{\frac{4}{\lambda}-2}$$
$$\text{if } \lambda > 2\sqrt{\frac{1+c}{c}}\left(\sqrt{\frac{1+c}{c}}+1\right).$$

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

Part (i) follows from the fact that $v_0 = \frac{c_2}{c_0c_2-c_1^2}$ when the design is a one point design at zero (that this is the optimal design for θ_0 follows from Theorem 2.2). To prove part (ii), conclude using part (b) in Theorem 2.3 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 $v_1 = \frac{c_0}{c_0c_2-c_1^2}$ where in c_i expectation is taken with respect to the designs described above.

Theorem 3.2. 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 if

$$M(\xi) = M = \begin{pmatrix} c_0 & c_1 \\ c_1 & c_2 \end{pmatrix} \text{ where } c_i = E\frac{X^i}{1+cX^\lambda},$$

then there is a two point design ξ^* supported on 0 and some other point in the interval [0,1] such that $M(\xi) \le M(\xi^*)$ in the sense of nonnegativeness. The condition $\lambda > 2$ is required to assert that $\left(1, \frac{1}{1+cX^\lambda}, \frac{X}{1+cX^\lambda}, \frac{-X^2}{1+cX^\lambda}\right)$ form a T–system. Since $v_i(\xi^*) \le v_i(\xi)$, $i = 0,1$, the theorem follows.

In view of the above theorem, 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 $\lambda \geq 2$, $c_2 = \frac{(1-c_0)^r c_1^{1-r}}{c^r}$ 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

$$\begin{aligned} c_0 &= 1 - p + \frac{p}{1 + cx_0^\lambda}, \\ c_1 &= \frac{px_0}{1 + cx_0^\lambda}, \\ \text{and } c_2 &= \frac{px_0^2}{1 + cx_0^\lambda}. \end{aligned} \tag{3.1}$$

Solving the first two equations for p and x_0 one obtains

$$\begin{aligned} x_0 &= \left(\frac{1 - c_0}{cc_1}\right)^r \\ \text{and } p &= \frac{c^r c_1^{1+r} + (1 - c_0)^{1+r}}{(1 - c_0)^r}. \end{aligned} \tag{3.2}$$

Substituting (3.2) into c_2 in (3.1) 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 Theorem 3.2 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 $\frac{1}{1+c} \leq c_0 \leq 1$, $c_1 \geq \frac{1-c_0}{c}$, and $c_1 \leq c_0^{1-\frac{1}{\lambda}} \left(\frac{1-c_0}{c}\right)^{\frac{1}{\lambda}}$ (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 $c_1 \leq c_0^{1-\frac{1}{\lambda}} \left(\frac{1-c_0}{c}\right)^{\frac{1}{\lambda}}$ holds follows from the Cauchy–Schwartz inequality). We will call

$$\mathcal{M} = \left\{ (c_0, c_1) \colon \frac{1}{1+c} \leq c_0 \leq 1,\ c_1 \geq \frac{1-c_0}{c},\ c_1 \leq c_0^{1-\frac{1}{\lambda}} \left(\frac{1-c_0}{c}\right)^{\frac{1}{\lambda}} \right\} \tag{3.3}$$

the moment space of the problem.

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

$$\mathcal{M} = \left\{ (c_0, c_1) \colon \frac{1}{1+c} \le c_0 \le 1,\ c_1 \ge \frac{1-c_0}{c},\ c_1 \le c_0^{\frac{1}{1+r}} \left(\frac{1-c_0}{c}\right)^{\frac{r}{1+r}} \right\}. \tag{3.4}$$

Now for a two point design ξ described above,

$$e_i(\xi) = \frac{v_i}{v_i(\xi)} = \left(\frac{c_{2-i}}{c_0c_2 - c_1^2}\right)^{-1} \cdot v_i \ge \frac{1}{1+\varepsilon}$$
$$\Leftrightarrow \frac{c_{2-i}}{c_0c_2 - c_1^2} \le (1+\varepsilon)v_i,\ i = 0, 1. \tag{3.5}$$

Using Theorem (3.3) and part (i) of Theorem 3.1, the first inequality in (3.5) reduces to

$$(1-c_0)^r c_1^{1-r} - (1+\varepsilon)(1-c_0)^r c_1^{1-r} c_0 + (1+\varepsilon)c^r c_1^2 \le 0. \tag{3.6}$$

Similarly the second inequality in (3.5) reduces to

$$c^r c_0 - (1+\varepsilon)v_1 c_0 (1-c_0)^r c_1^{1-r} + (1+\varepsilon)v_1 c^r c_1^2 \le 0. \tag{3.7}$$

Motivated by these, we will define

$$S_0 = \{(c_0, c_1) \colon (3.6) \text{ holds}\},$$
$$S_1 = \{(c_0, c_1) \colon (3.7) \text{ holds}\}. \tag{3.8}$$

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 $S_0 \cap S_1 \cap \mathcal{M} \ne \phi$, 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 3.4. For $\varepsilon \ge 0$, 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 $\varepsilon_0 = \inf\{\varepsilon > 0 \colon A_\varepsilon \cap B_\varepsilon \cap C \ne \phi\}$. Then $A_{\varepsilon_0} \cap B_{\varepsilon_0} \cap C = \partial A_{\varepsilon_0} \cap \partial B_{\varepsilon_0} \cap C$.

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.

$$\therefore A_{\varepsilon_0} \cap B_{\varepsilon_0} \cap C = (\partial A_{\varepsilon_0} \cap \partial B_{\varepsilon_0} \cap C) \cup (\partial A_{\varepsilon_0} \cap B^0_{\varepsilon_0} \cap C)$$
$$\cup (A^0_{\varepsilon_0} \cap \partial B_{\varepsilon_0} \cap C) \cup (A^0_{\varepsilon_0} \cap B^0_{\varepsilon_0} \cap C).$$

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

$$\frac{c_2}{c_0} = \frac{1}{v_1}$$
$$\Rightarrow c_1^{1-r} = \frac{1}{v_1} c^r c_0 (1-c_0)^{-r}. \tag{3.9}$$

(Set both inequalities in (3.5) as equalities, divide, and then use Theorem 3.3).

Substituting (3.9) into (3.6) (with an equality in (3.6)), one gets

$$\frac{1}{1+\varepsilon_0} = c_0 - c^{\frac{2r}{1-r}} v_1^{-\frac{1+r}{1-r}} c_0^{\frac{1+r}{1-r}} (1-c_0)^{-\frac{2r}{1-r}}. \tag{3.10}$$

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 (3.9) and (3.4),

$$\frac{1}{v_1} c^r c_0 (1-c_0)^{-r} \geq \left(\frac{1-c_0}{c}\right)^{1-r}, \tag{3.11}$$

$$\text{and } \frac{1}{v_1} c^r c_0 (1-c_0)^{-r} \leq c_0^{\frac{1-r}{1+r}} \left(\frac{1-c_0}{c}\right)^{\frac{r(1-r)}{1+r}}. \tag{3.12}$$

On algebra, (3.11) and (3.12) reduce to

$$\frac{v_1}{c+v_1} \leq c_0 \leq \frac{v_1^{\frac{1+r}{2r}}}{c + v_1^{\frac{1+r}{2r}}}. \tag{3.13}$$

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 (3.13) such that (3.10) holds. Conversely, if there is a c_0 satisfying (3.13) such that (3.10) 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 (3.10), i.e.,

$$f(c_0) = c_0 - c^{\frac{2r}{1-r}} v_1^{-\frac{1+r}{1-r}} c_0^{\frac{1+r}{1-r}} (1-c_0)^{\frac{-2r}{1-r}}, \tag{3.14}$$

for c_0 satisfying (3.13). 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

$$h(z) = \frac{z - \frac{1}{c} \cdot \left(\frac{c}{v_1}\right)^p z^p}{1+z}, \tag{3.15}$$

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

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

$$N(z) = c \cdot \left(\frac{v_1}{c}\right)^p - pz^{p-1} - (p-1)z^p, \tag{3.16}$$

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 $N(z) \le 0$ 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 $N(z) \le 0$ at $z = \frac{v_1}{c}$ iff $p \ge \frac{v_1(1+c)}{v_1+c}$. 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

$$pz^{p-1} + (p-1)z^p = c \cdot \left(\frac{v_1}{c}\right)^p. \tag{3.17}$$

We thus have the following theorem:

Theorem 3.5. Let $\varepsilon_0 = \inf\{\varepsilon > 0 : \Gamma_\varepsilon \neq \phi\}$. Then

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

(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 (3.17).

Using the formulas for v_1 derived in Theorem 3.1 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 3.5 can be much improved. In fact, once one finds the unique root z_0 of (3.17), 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 3.6. Let $p = \frac{\lambda}{\lambda-2}$.

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

$$\varepsilon_0 = \frac{1+c}{v_1-1}, \quad c_0 = \frac{v_1}{v_1+c}, \quad c_1 = \frac{1}{v_1+c}, \quad X_0 = 1, \text{ and } p_0 = \frac{1+c}{v_1+c},$$

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

$$\varepsilon_0 = \frac{p}{(p-1)z_0}, \quad c_0 = \frac{z_0}{1+z_0},$$

$$c_1 = c^{\frac{p-1}{2}} \cdot \frac{1}{1+z_0}\left(\frac{z_0}{v_1}\right)^{\frac{p+1}{2}}, \quad X_0 = \left(\frac{v_1}{cz_0}\right)^{\frac{p-1}{2}}, \text{ and}$$

$$p_0 = \frac{p}{v_1}\left(\frac{cz_0}{v_1}\right)^{p-1},$$

where z_0 is the root of (3.17).

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 (3.17) 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 (3.9). For X_0, use (3.2), and for p_0 use $p_0 = 1 - c_0 + \frac{c_1}{X_0}$ (which is an implication of (3.1)).

Discussion: Of course, given λ and c, it will be easy to see using Theorem 3.1 whether case (i) or (ii) applies in Theorem 3.6. Then Theorem 3.6 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 3.7.

(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 3.6. To prove part (ii), note that $p \to 1$ as $\lambda \to \infty$. Also, using part (ii) of Theorem 3.1 and the definition of p, it is easy to check that given any c, for large λ,

$$v_1 = c^{\frac{p-1}{p}} \cdot \frac{(p+1)^{\frac{p+1}{p}}}{(p-1)^{\frac{p-1}{p}}}. \tag{3.18}$$

Note that (3.18) converges to 4 as $p \to 1$. Thus, given any c, for large λ, case (ii) in Theorem 3.6 applies. Use now the fact that z_0 solves (3.17), or equivalently,

$$\begin{aligned} pz_0^{p-1} + (p-1)z_0^p &= \frac{(p+1)^{p+1}}{(p-1)^{p-1}} \\ \Leftrightarrow pa_p^{p-1} + a_p^p &= (p+1)^{p+1}, \end{aligned} \tag{3.19}$$

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 (3.19) 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 3.6, 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.

4. Maximin efficient designs: $-1 \le X \le 1$.

The case when the independent variable X belongs to the interval [-1,1] is technically very similar to the case when $0 \le X \le 1$ and indeed in some respects is considerably easier. We will first give a simple example to illustrate the theory of maximin efficient designs when X is in [-1,1]. This example will also provide the basis for another example in section 4.

Theorem 4.1. Consider the simple linear regression model with a variance function $w(x) = 1 + cx^2$, $c > 0$, $-1 \le x \le 1$. Suppose $\underset{\sim}{\theta}$ has a $N(\underset{\sim}{\mu}, C)$ prior where $\frac{C^{-1}}{n} = \begin{pmatrix} r_0 & 0 \\ 0 & r_2 \end{pmatrix}$. Then

(i) The optimal Bayes design for estimating θ_0 is supported on 0.

(ii) The optimal Bayes design for estimating θ_1 is supported on $x = \pm 1$ with mass $\frac{1}{2}$ at each point.

(iii) $v_0 = \frac{1}{1+r_0}$

(iv) $v_1 = \frac{1+c}{1+(1+c)r_2}$.

Proof: First note that for every design, $c_2 = E\frac{X^2}{w(X)} = \frac{1-c_0}{c}$. The moment space $\mathcal{M}$ can then be described as

$$\mathcal{M} = \left\{ (c_0, c_1): \frac{1}{1+c} \le c_0 \le 1,\ -\sqrt{\frac{c_0(1-c_0)}{c}} \le c_1 \le \sqrt{\frac{c_0(1-c_0)}{c}} \right\} \tag{4.1}$$

The Bayes design for estimating θ_0 then minimizes $\frac{\frac{1-c_0}{c}+r_2}{(c_0+r_0)(\frac{1-c_0}{c}+r_2)-c_1^2}$ which is plainly the design in (i). This also gives (iii) immediately. The Bayes design for θ_1 minimizes $\frac{c_0+r_0}{(c_0+r_0)(\frac{1-c_0}{c}+r_2)-c_1^2}$, and is thus the design in (ii); (iv) then follows immediately.

Remark: Notice that the constants r_0, r_2, c have no effect on the Bayes designs. This is really because C was chosen as diagonal. Also note that the Bayes designs in (i) and (ii) are also the classical optimal designs in these problems.

Theorem 4.2. Let $\varepsilon_0 = \inf\{\varepsilon > 0: \Gamma_\varepsilon(C) \neq \phi\}$. Then $\varepsilon_0 = \frac{c}{(1+c)(1+cr_2+r_0)}$.

Proof: Define again

$$S_i = \{(c_0, c_1): v_i(\xi) \leq (1+\varepsilon)v_i,\ i = 0, 1\}. \tag{4.2}$$

Using Theorem 4.1 now,

$$S_0 = \left\{(c_0, c_1): \frac{\frac{1-c_0}{c} + r_2}{(c_0 + r_0)\left(\frac{1-c_0}{c} + r_2\right) - c_1^2} \leq \frac{1+\varepsilon}{1+r_0}\right\},$$

which, on straightforward but rather lengthy algebra, reduces to

$$S_0 = \left\{(c_0, c_1): \left(c_0 - \frac{1 + cr_2 + \frac{1-\varepsilon r_0}{1+\varepsilon}}{2}\right)^2 + cc_1^2 \leq \left(\frac{\varepsilon(1 + r_0 + cr_2) + cr_2}{2(1+\varepsilon)}\right)^2\right\}. \tag{4.3}$$

Notice that S_0 is thus just an ellipsoid with axes parallel to the coordinate axes and symmetric along the c_1–axis.

Similarly,

$$S_1 = \left\{(c_0, c_1): \frac{c_0 + r_0}{(c_0 + r_0)\left(\frac{1-c_0}{c} + r_2\right) - c_1^2} \leq \frac{(1+\varepsilon)(1+c)}{1 + (1+c)r_2}\right\},$$

which reduces to

$$S_1 = \left\{(c_0, c_1): \left(c_0 - \frac{1 + \varepsilon(1+c) + \varepsilon c(1+c)r_2 - (1+\varepsilon)(1+c)r_0}{2(1+\varepsilon)(1+c)}\right)^2 + cc_1^2 \leq \left(\frac{1 + \varepsilon(1+c) + \varepsilon c(1+c)r_2 + (1+\varepsilon)(1+c)r_0}{2(1+\varepsilon)(1+c)}\right)^2\right\}. \tag{4.4}$$

Note that this is another ellipsoid with axes parallel to the coordinate axes and again symmetric along the c_1–axis.

By using routine calculus, it is very easy to prove that the center of the first ellipsoid (4.3) moves left and the radius along the c_0–axis increases as ε increases and the center of the second ellipsoid (4.4) moves right and the radius along the c_0–axis increases as ε increases. Clearly, therefore, the smallest ε for which S_0 and S_1 will intersect is such that the left boundary point of S_0 on the c_0–axis merges with the right boundary point of S_1, i.e.,

$$\frac{1+cr_2+\frac{1-\varepsilon r_0}{1+\varepsilon}}{2}-\frac{\varepsilon(1+r_0+cr_2)+cr_2}{2(1+\varepsilon)}$$
$$=\frac{1+\varepsilon(1+c)+\varepsilon c(1+c)r_2-(1+\varepsilon)(1+c)r_0}{2(1+\varepsilon)(1+c)}$$
$$+\frac{1+\varepsilon(1+c)+\varepsilon c(1+c)r_2+(1+\varepsilon)(1+c)r_0}{2(1+\varepsilon)(1+c)},$$

which gives

$$\varepsilon=\frac{c}{(1+c)(1+cr_2+r_0)}.$$

This proves the theorem.

Corollary 4.3. Consider the set up of Theorem 4.1. The maximin efficient design has

$$c_0=\frac{1+cr_2+r_0+c(1+cr_2)}{1+cr_2+r_0+c(2+cr_2+r_0)},$$

and $c_1=0$ (this point is in the moment space $\mathcal{M}$); also, the maximin efficiency equals

$$\frac{1}{1+\varepsilon_0}=\frac{(1+c)(1+cr_2+r_0)}{1+cr_2+r_0+c(2+cr_2+r_0)}.$$

Proof: Evaluating the values of c_0 and c_1 amounts to finding the point where S_0 and S_1 first touch. This is immediate from Theorem 4.2. The maximin efficiency equals $\frac{1}{1+\varepsilon_0}$ and thus is immediate again.

Discussion: Clearly the maximin efficient design is not unique because the given values of c_0 and c_1 in corollary 4.3 can be attained by many measures on the set [-1,1]. In particular, there is a symmetric 3 point design with support on $0, \pm 1$ that is maximin efficient. An attractive possibility is to choose among these measures by using model robustness as a criterion. We do not go into the details of this here.

Also, for any given C, the maximin efficiency is greater if prior information is more precise. That is, if $C_1 \leq C_2$, then $\frac{C_1^{-1}}{n} \geq \frac{C_2^{-1}}{n}$ and it follows that ε_0 is smaller under C_1

than under C_2. Thus if prior information is very precise, then it is easier to get designs which are near optimum for both θ_0 and θ_1 simultaneously.

Using essentially the same argument as above, the value of the smallest ε for which $\Gamma_\varepsilon(C)$ is nonempty can be worked out for other variance functions. We have a theorem below for the case $w(x) = 1 + c|x|^\lambda,\ c > 0,\ \lambda \geq 2$.

Theorem 4.4. Consider the simple linear regression model with a variance function $w(x) = 1 + c|x|^\lambda,\ c > 0,\ -1 \leq x \leq 1$. Suppose $\underset{\sim}{\theta}$ has a $N(\underset{\sim}{\mu}, C)$ prior where $\frac{C^{-1}}{n} = \begin{pmatrix} r_0 & 0 \\ 0 & r_2 \end{pmatrix}$. Then

(i) The optimal Bayes design for estimating θ_0 is supported on 0; also $v_0 = \frac{1}{1+r_0}$.

(ii) If $c(\lambda - 2) \leq 2$, the optimal Bayes design for estimating θ_1 is supported on ± 1 with mass .5 at each point and $v_1 = \frac{1+c}{1+(1+c)r_2}$; if $c(\lambda - 2) > 2$, the optimal Bayes design for estimating θ_1 is supported on $\pm x_0$ with mass .5 at each point where

$$x_0 = \left(\frac{2}{c(\lambda - 2)}\right)^{\frac{1}{\lambda}}$$

and $v_1 = \lambda / \left(4^{\frac{1}{\lambda}} \cdot c^{-\frac{2}{\lambda}}(\lambda - 2)^{1-\frac{2}{\lambda}} + \lambda r_2\right)$.

(iii) If $\varepsilon_0 = \inf\{\varepsilon > 0 : \Gamma_\varepsilon(C) \neq \phi\}$, then

$$\varepsilon_0 = \frac{1}{v_0(\bar{c}_0 + r_0)} - 1,$$

where $\bar{c}_0$ is the largest root of the equation

$$c_0 = \frac{v_1}{v_0}\left\{\frac{(1-c_0)^{\frac{2}{\lambda}} c_0^{1-\frac{2}{\lambda}}}{c^{2/\lambda}} + r_2\right\} - r_0 \tag{4.5}$$

Proof: We will only give a short sketch of the proof of (iii). Using a symmetry and convexity argument, it follows that if $\Gamma_\varepsilon(C)$ is nonempty for some ε, then there exists a symmetric design in $\Gamma_\varepsilon(C)$. We can therefore restrict attention to designs having $c_1 = E\frac{X}{w(X)} = 0$.

The moment space $\mathcal{M}$ now can be written as

$$\mathcal{M} = \left\{(c_0, c_2) : \frac{1}{1+c} \leq c_0 \leq 1,\ c_2 \geq \frac{1-c_0}{c},\ c_2 \leq \frac{(1-c_0)^{\frac{2}{\lambda}} \cdot c_0^{1-\frac{2}{\lambda}}}{c^{\frac{2}{\lambda}}}\right\}. \tag{4.6}$$

Thus $\Gamma_\varepsilon(C)$ is nonempty iff for some point $(c_0, c_2)\varepsilon\mathcal{M}$,

$$\frac{1}{c_0 + r_0} \le (1+\varepsilon)v_0, \tag{4.7}$$

$$\text{and } \frac{1}{c_2 + r_2} \le (1+\varepsilon)v_1. \tag{4.8}$$

Using the argument of Theorem 3.5, one can then show that ε_0 satisfies

$$1 + \varepsilon_0 = \frac{1}{v_0(\bar{c}_0 + r_0)},$$

$$\text{where } \bar{c}_0 = \sup\left\{c_0 : c_0 \ge \frac{cv_0 + v_1 - c}{cv_0 + v_1},\ c_0 \le \frac{v_1}{v_0}\left(\frac{(1-c_0)^{\frac{2}{\lambda}} c_0^{1-\frac{2}{\lambda}}}{c^{\frac{2}{\lambda}}} + r_2\right) - r_0\right\}$$

(essentially, one uses equalities in (4.7) and (4.8), solves for c_2 in terms of c_0, and then uses (4.8) to write $1 + \varepsilon_0 = \frac{1}{v_0(c_0 + r_0)}$. Then one forces the constraint that the point must belong to the moment space $\mathcal{M}$).

Using now the fact that $\frac{v_1}{v_0}\left(\frac{(1-c_0)^{\frac{2}{\lambda}} c_0^{1-\frac{2}{\lambda}}}{c^{\frac{2}{\lambda}}} + r_2\right) - r_0$ is a concave function of c_0 if $\lambda \ge 2$, one gets that $\bar{c}_0$ is in fact the largest root of (4.5).

Using the result of Theorem 4.4, we have numerically calculated the value of ε_0 and the maximin efficiency $\frac{1}{1+\varepsilon_0}$ for some values of c and λ. The entries are the values of the maximin efficiencies. We take $r_0 = r_2 = .1$.

Table of maximin efficiency

λ / c	2	3	4	5	10
.5	.7752	.8092	.8374	.8625	.9327
1	.7059	.7699	.8227	.8604	.9326
2	.6610	.7631	.8294	.8673	.9336
5	.6575	.7820	.8389	.8720	.9337

For any c, the maximin efficiency seems to increase with λ, while for given λ, it seems to first decrease and then increase as c increases. Of course, more computation is required before a general statement can be made with more confidence.

As stated before, the problem of finding maximin efficient designs is of interest in polynomial regression too. For the classical homoscedastic model $E(y|x) = \theta_0 + \theta_1 x +$

$\ldots + \theta_p x^p$ with $\text{Var}(y|x) = 1$, the value of ε_0 (i.e., the smallest ε) equals 0, 1, .715, and 1.072 respectively for $p = 1, 2, 3, 4$ when $-1 \leq x \leq 1$. This ε_0 is thus the smallest ε for which the set of designs $\Gamma_\varepsilon = \{\xi: v_i(\xi) \leq (1+\varepsilon)v_i,\ 0 \leq i \leq p\}$ is nonempty. Interestingly, we found that for each $p = 2, 3, 4$, one actually has $v_i(\xi_0) = (1+\varepsilon_0)v_i$ for $0 \leq i \leq p-1$ and $v_p(\xi_0) < (1+\varepsilon_0)v_p$ where ξ_0 is the maximin efficient design. It may be of some interest to see whether this holds for general p and also to see the behavior of ε_0 as p increases.

5. Robust Bayes designs. In this section, we address the problem of minimizing sensitivity with respect to the prior at a small cost in Bayes optimality for a fixed prior. We will do this analytically only in the case of a simple linear regression with variance function $1+x^2$. We will consider a family of normal priors $N(\underset{\sim}{\mu}, C)$ for $\underset{\sim}{\theta} = (\theta_0, \theta_1)'$ where $\underset{\sim}{\mu}$ belongs to any arbitrary set in $\mathbb{R}^2$ and $C_1 \leq C \leq C_2$ for fixed C_1 and C_2. The cases $C \geq C_1$ and $C \leq C_2$ are included in this. As discussed in section 1, we will give a complete description of the set of vectors of Bayes risks for estimating θ_0, θ_1 as the prior changes in the above manner. Sensitivity to the prior will be measured in terms of the area of the above set, say S, (other sensitivity measures such as the Euclidean diameter are possible. In fact, the diameter always works out to $tr(X'X + C_2^{-1})^{-1} - tr(X'X + C_1^{-1})^{-1}$ and may be easier to handle because of some known monotonicity and convexity results for this functional of $X'X$. See DasGupta and Studden (1988)). In the spirit of the results of the previous section, we will fix one prior $\pi_0 \equiv N(\mu_0, C_0)$ and minimize sensitivity at a small cost in Bayes optimality with respect to π_0 by minimizing the area of the set S (this area is a functional of the design ξ) subject to the restriction that the design ξ is in $\Gamma_\varepsilon(C_0)$ where ε is such that $\Gamma_\varepsilon(C_0)$ is not empty. We start with a few general results on description of the set S. Recall that if the prior covariance matrix is C then the posterior covariance matrix is $(X'X + C^{-1})^{-1}$ which varies in the range $(X'X + C_1^{-1})^{-1} \leq (X'X + C^{-1})^{-1} \leq (X'X + C_2^{-1})^{-1}$ (and every p.d. matrix in this range is a possible posterior covariance matrix). For the purpose of describing the set of vectors of Bayes risks, we can therefore simply address the problem of describing the set of vectors of diagonal elements of Σ when Σ varies in a fixed range $\Sigma_1 \leq \Sigma \leq \Sigma_2$. Also by reparametrizing to $\Sigma - \frac{\Sigma_1 + \Sigma_2}{2}$, we can without loss of generality assume that Σ lies in the range $-\Sigma_1 \leq \Sigma \leq \Sigma_1$ where Σ_1 is a p.d. matrix. Note that this

reparametrization would not change the area of the set of diagonal elements.

Theorem 5.1. Let $\Sigma = \begin{pmatrix} \lambda & \rho \\ \rho & \mu \end{pmatrix}$ be any symmetric matrix in the range $-\Sigma_1 = -\begin{pmatrix} a & b \\ b & c \end{pmatrix} \leq \Sigma \leq \Sigma_1 = \begin{pmatrix} a & b \\ b & c \end{pmatrix}$, where Σ_1 is a fixed symmetric p.d. matrix. Then the set S of all (λ, μ) when $-\Sigma_1 \leq \Sigma \leq \Sigma_1$ is the set of all (λ, μ) in the rectangle $[-a, a] \times [-c, c]$ satisfying $\sqrt{(a+\lambda)(c+\mu)} + \sqrt{(a-\lambda)(c-\mu)} \geq 2|b|$. Geometrically, it is the rectangle $[-a, a] \times [-c, c]$ with its left upper and right lower corners chopped off by the ellipse (5.4).

Proof: Notice that

$$
\begin{aligned}
S = \Bigg\{ & (\lambda, \mu) \colon -a \leq \lambda \leq a, \ -c \leq \mu \leq c, \\
& (\rho + b)^2 \leq (a+\lambda)(c+\mu), \ (\rho - b)^2 \leq (a-\lambda)(c-\mu) \text{ for some } \rho \Bigg\} \\
= \Bigg\{ & (\lambda, \mu) \colon -a \leq \lambda \leq a, \ -c \leq \mu \leq c, \text{ for some } \rho \\
& \max\left(b - \sqrt{(a-\lambda)(c-\mu)}, \ -b - \sqrt{(a+\lambda)(c+\mu)}\right) \\
& \leq \rho \\
& \leq \min\left(b + \sqrt{(a-\lambda)(c-\mu)}, \ \sqrt{(a+\lambda)(c+\mu)} - b\right) \Bigg\}. \qquad (5.1)
\end{aligned}
$$

Therefore the set S can simply be described as

$$
\begin{aligned}
S = \Bigg\{ & (\lambda, \mu) \colon -a \leq \lambda \leq a, \ -c \leq \mu \leq c, \ b + \sqrt{(a-\lambda)(c-\mu)} \geq -b - \sqrt{(a+\lambda)(c+\mu)}, \\
& -b + \sqrt{(a+\lambda)(c+\mu)} \geq b - \sqrt{(a-\lambda)(c-\mu)} \Bigg\} \\
= \Big\{ & (\lambda, \mu) \colon -a \leq \lambda \leq a, \ -c \leq \mu \leq c, \ \sqrt{(a+\lambda)(c+\mu)} + \sqrt{(a-\lambda)(c-\mu)} \geq 2|b| \Big\} \qquad (5.2)
\end{aligned}
$$

$$
\begin{aligned}
\text{Now } & \sqrt{(a+\lambda)(c+\mu)} + \sqrt{(a-\lambda)(c-\mu)} \geq 2|b| \\
\Leftrightarrow & \sqrt{(a^2-\lambda^2)(c^2-\mu^2)} \geq 2b^2 - (ac + \lambda\mu)
\end{aligned}
$$

(square both sides of the given inequality and regroup terms).

$$\text{Now } \Big\{(\lambda,\mu)\varepsilon[-a,a]\times[-c,c]\colon \sqrt{(a^2-\lambda^2)(c^2-\mu^2)} \geq 2b^2-(ac+\lambda\mu)\Big\}$$
$$=\Big\{(\lambda,\mu)\varepsilon[-a,a]\times[-c,c]\colon \sqrt{(a^2-\lambda^2)(c^2-\mu^2)} \geq 2b^2-(ac+\lambda\mu),\ ac+\lambda\mu\leq 2b^2\Big\}$$
$$\cup\Big\{(\lambda,\mu)\varepsilon[-a,a]\times[-c,c]\colon \sqrt{(a^2-\lambda^2)(c^2-\mu^2)} \geq 2b^2-(ac+\lambda\mu),\ ac+\lambda\mu> 2b^2\Big\}. \tag{5.3}$$

Now, if $ac+\lambda\mu\leq 2b^2$, then

$$\sqrt{(a^2-\lambda^2)(c^2-\mu^2)} \geq 2b^2-(ac+\lambda\mu)$$
$$\Leftrightarrow a^2\mu^2+\lambda^2c^2+2\lambda\mu(ac-2b^2)\leq 4b^2(ac-b^2), \tag{5.4}$$

which is an ellipse (say E) contained in the rectangle $[-a,a]\times[-c,c]$.

On the other hand, if $ac+\lambda\mu > 2b^2$, then the inequality $\sqrt{(a^2-\lambda^2)(c^2-\mu^2)} \geq 2b^2-(ac+\lambda\mu)$ automatically holds. Therefore, by using (5.3) and (5.4),

$$S=\{(\lambda,\mu)\varepsilon[-a,a]\times[-c,c]\colon (\lambda,\mu)\varepsilon E,\ ac+\lambda\mu\leq 2b^2\}$$
$$\cup\{(\lambda,\mu)\varepsilon[-a,a]\times[-c,c]\colon ac+\lambda\mu> 2b^2\}. \tag{5.5}$$

Geometrically, the set S is the rectangle $[-a,a]\times[-c,c]$ with its left upper and right lower corners chopped off by the boundaries of the ellipse E. This can be seen by considering the cases $2b^2-ac>0$ and $2b^2-ac\leq 0$ separately and by examining the orientation of the ellipse E and the hyperbola $\lambda\mu=2b^2-ac$. The ellipse E and the set S are plotted in figure 1. It turns out that while the pairs (a,c) and $(-a,-c)$ are always in the set S, the pairs $(a,-c)$ and $(-a,c)$ are not unless $b=0$, (i.e., Σ_1 is a diagonal matrix) in which case (and only in that case) S is the full rectangle $[-a,a]\times[-c,c]$. Indeed, for $\mu=-c$, the maximum possible value of λ is $a-\frac{2b^2}{c}$, and for $\mu=c$, the minimum possible value of λ is $-a+\frac{2b^2}{c}$. These are precisely the points at which the ellipse E touches the lines $\mu=-c$ and $\mu=c$.

The following theorem is useful in describing how the set of any two diagonal elements of Σ can be obtained from the above theorem if Σ is a $p\times p$ symmetric p.d. matrix in the range $\Sigma_1\leq\Sigma\leq\Sigma_2$, where Σ_1,Σ_2 are also p.d. and $\Sigma_2>\Sigma_1$.

Theorem 5.2. Let $\Sigma_1 = \begin{pmatrix} A & \underset{\sim}{u}_1 \\ \underset{\sim}{u}_1' & a_{pp} \end{pmatrix} \leq \Sigma = \begin{pmatrix} \Sigma_{11} & \underset{\sim}{u} \\ \underset{\sim}{u}' & \sigma_{pp} \end{pmatrix} \leq \Sigma_2 = \begin{pmatrix} B & \underset{\sim}{u}_2 \\ \underset{\sim}{u}_2' & b_{pp} \end{pmatrix}$. Then $A \leq \Sigma_{11} \leq B$, and given any Σ_{11} such that $A \leq \Sigma_{11} \leq B$, there exist $\underset{\sim}{u}$ and σ_{pp} such that

$$\Sigma_1 \leq \begin{pmatrix} \Sigma_{11} & \underset{\sim}{u} \\ \underset{\sim}{u}' & \sigma_{pp} \end{pmatrix} \leq \Sigma_2.$$

Remark: This theorem will imply that in order to get a characterization of the set of the first $(p-1)$ diagonal elements of Σ, one only need consider the constraint $A \leq \Sigma_{11} \leq B$. By a repeated application of this theorem it in fact follows that all projections of the set of all p diagonal elements of Σ are just the sets one would obtain by considering the lower dimensional inequalities on the corresponding block of Σ. In particular, to get the set of any two diagonal elements of Σ, one can simply restrict attention to the corresponding 2×2 block in Σ, and then use Theorem 5.1. Indeed, it also follows from this theorem that in order to get the set of all vectors $(\ell_1'\Sigma\ell_1, \ell_2'\Sigma\ell_2)$ where $\underset{\sim}{\ell}_1, \underset{\sim}{\ell}_2$ are two independent vectors, one can extend $\begin{pmatrix} \ell_1' \\ \ell_2' \end{pmatrix}$ to a full rank matrix L, use the fact that $L\Sigma L'$ varies in the entire range $L\Sigma_1 L' \leq L\Sigma L' \leq L\Sigma_2 L'$ and then get the set of vectors $(\ell_1'\Sigma\ell_1, \ell_2'\Sigma\ell_2)$ by using the fact that these are just the first two diagonal elements of $L\Sigma L'$.

Proof of Theorem 5.2: That $A \leq \Sigma_{11} \leq B$ is well known. To prove the second part we will prove that if $A < \Sigma_{11} < B$, then there exist $\underset{\sim}{u}$ and σ_{pp} such that $\Sigma_1 < \begin{pmatrix} \Sigma_{11} & \underset{\sim}{u} \\ \underset{\sim}{u}' & \sigma_{pp} \end{pmatrix} < \Sigma_2$. The assertion in the theorem will then follow by a closure argument.

So given $A < \Sigma_{11} < B$, we need to prove that there exist $\underset{\sim}{u}$ and σ_{pp} such that

$$\begin{aligned} (\sigma_{pp} - a_{pp}) &\geq (\underset{\sim}{u} - \underset{\sim}{u}_1)'(\Sigma_{11} - A)^{-1}(\underset{\sim}{u} - \underset{\sim}{u}_1) \\ \text{and } (b_{pp} - \sigma_{pp}) &\geq (\underset{\sim}{u} - \underset{\sim}{u}_2)'(B - \Sigma_{11})^{-1}(\underset{\sim}{u} - \underset{\sim}{u}_2), \end{aligned} \tag{5.6}$$

or equivalently, there exist $\underset{\sim}{u}$ and σ_{pp} such that

$$a_{pp} + (\underset{\sim}{u} - \underset{\sim}{u}_1)'(\Sigma_{11} - A)^{-1}(\underset{\sim}{u} - \underset{\sim}{u}_1) \leq \sigma_{pp} \leq b_{pp} - (\underset{\sim}{u} - \underset{\sim}{u}_2)'(B - \Sigma_{11})^{-1}(\underset{\sim}{u} - \underset{\sim}{u}_2).$$

Therefore it is enough to prove that there exists $\underset{\sim}{u}$ such that

$$a_{pp} + (\underset{\sim}{u} - \underset{\sim}{u}_1)'(\Sigma_{11} - A)^{-1}(\underset{\sim}{u} - \underset{\sim}{u}_1) \leq b_{pp} - (\underset{\sim}{u} - \underset{\sim}{u}_2)'(B - \Sigma_{11})^{-1}(\underset{\sim}{u} - \underset{\sim}{u}_2), \tag{5.7}$$

$$\Leftrightarrow (\underset{\sim}{u} - \underset{\sim}{u_1})'(\Sigma_{11} - A)^{-1}(\underset{\sim}{u} - \underset{\sim}{u_1}) + (\underset{\sim}{u} - \underset{\sim}{u_2})'(B - \Sigma_{11})^{-1}(\underset{\sim}{u} - \underset{\sim}{u_2}) \le b_{pp} - a_{pp}; \qquad (5.8)$$

this is because if (5.7) holds for some $\underset{\sim}{u}$, we can then take σ_{pp} to be any number in the interval (5.7).

That (5.8) holds for some $\underset{\sim}{u}$ is easily seen by noting that the minimum value of $(\underset{\sim}{u} - \underset{\sim}{u_1})'(\Sigma_{11} - A)^{-1}(\underset{\sim}{u} - \underset{\sim}{u_1}) + (\underset{\sim}{u} - \underset{\sim}{u_2})'(B - \Sigma_{11})^{-1}(\underset{\sim}{u} - \underset{\sim}{u_2})$ is $(\underset{\sim}{u_2} - \underset{\sim}{u_1})'(B - A)^{-1}(\underset{\sim}{u_2} - \underset{\sim}{u_1})$ (whatever be Σ_{11}) and this is $\le b_{pp} - a_{pp}$ since $\Sigma_1 < \Sigma_2$. This proves the theorem.

We now go back to Theorem 5.1 and get an expression for the area of the set S.

Theorem 5.3. The area of S equals $4ac \cdot g(\frac{b^2}{ac})$ where

$$g(z) = 2(1-z) + \sqrt{z(1-z)}\left(\frac{\pi}{2} - \sin^{-1}(1-2z)\right),\ 0 \le z \le 1.$$

Here $\sin^{-1}(\cdot)$ is defined such that $\sin^{-1}(-1) = -\frac{\pi}{2}$ and $\sin^{-1}(1) = \frac{\pi}{2}$.

Proof: The area of the whole rectangle is $4ac$. Because of symmetry, we need to subtract from $4ac$ twice the area of the right lower corner of the rectangle chopped off by the ellipse (5.4). The required integration is a routine exercise and is omitted.

We will now work out analytically the design minimizing the area of S among the set of designs that belong to $\Gamma_\varepsilon(C)$. We will do this in the case of a simple linear regression with variance function $w(x) = 1 + x^2$, $-1 \le x \le 1$. This should be considered more as an artifact rather than anything else. Some comments on the case $1 + cx^2$ are made afterwards.

Theorem 5.4. Let $E(y|x) = \theta_0 + \theta_1 x$, $-1 \le x \le 1$, and $\text{Var}(y|x) = 1 + x^2$. Suppose $\underset{\sim}{\theta} \sim N(\underset{\sim}{\mu}, C)$ where $\underset{\sim}{\mu}$ is in some set T in $\mathbb{R}^2$ and $K_1 I \le \frac{C^{-1}}{n} \le K_2 I$ for some $K_2 \ge K_1$. Let C_0 be a fixed prior covariance matrix and let $\frac{C_0^{-1}}{n} = \begin{pmatrix} r_0 & 0 \\ 0 & r_2 \end{pmatrix}$. The design minimizing the area of the set of (vectors of) Bayes risks subject to the restriction that it belongs to $\Gamma_\varepsilon(C_0)$ has

$$c_0 = \frac{1 - \varepsilon r_0}{1 + \varepsilon},\ c_1 = 0$$

unless $\varepsilon \ge \frac{1}{1+2r_0}$, in which case the unconstrained minima $c_0 = \frac{1}{2}$, $c_1 = 0$ belongs to $\Gamma_\varepsilon(C_0)$.

Proof: Note that the posterior covariance matrix Λ varies in the range

$$\Lambda_1 = \frac{1}{n}(M + K_2 I)^{-1} \leq \Lambda \leq \frac{1}{n}(M + K_1 I)^{-1} = \Lambda_2, \text{ where}$$
$$M = \begin{pmatrix} c_0 & c_1 \\ c_1 & 1 - c_0 \end{pmatrix}.$$

In the notation of Theorems 5.1 and 5.3, then, the area of the set of Bayes risks equals $4ac \cdot g(\frac{b^2}{ac})$,

$$\text{where } \begin{pmatrix} a & b \\ b & c \end{pmatrix} = \frac{\Lambda_2 - \Lambda_1}{2}. \tag{5.9}$$

$$\begin{aligned} \text{Now, } \frac{\Lambda_2 - \Lambda_1}{2} &= \frac{1}{2n}\left[(M + K_1 I)^{-1} - (M + K_2 I)^{-1}\right] \\ &= \frac{K_2 - K_1}{2n} \cdot (M + K_1 I)^{-1}(M + K_2 I)^{-1} \\ &= \frac{K_2 - K_1}{2n}(M^2 + (K_2 + K_1)M + K_1 K_2 I)^{-1}. \end{aligned} \tag{5.10}$$

We can and will ignore the multiplicative factor $\frac{K_2 - K_1}{2n}$ as it will have no effect on the design problem.

Let $\begin{pmatrix} \alpha & \beta \\ \beta & \delta \end{pmatrix} = M^2 + (K_1 + K_2)M + K_1 K_2 I$. Therefore,

$$\alpha = c_0^2 + c_1^2 + (K_1 + K_2)c_0 + K_1 K_2,$$
$$\beta = (K_1 + K_2 + 1)c_1,$$
$$\delta = (1 - c_0)^2 + c_1^2 + (K_1 + K_2)(1 - c_0) + K_1 K_2.$$

Thus the area of the set of Bayes risks is proportional to (because we have ignored the factor $\frac{K_2 - K_1}{2n}$)

$$\begin{aligned} &\frac{\alpha\delta}{(\alpha\delta - \beta^2)^2} \cdot g\left(\frac{\beta^2}{\alpha\delta}\right) \\ &= \frac{1}{\alpha\delta} \cdot h\left(\frac{\beta^2}{\alpha\delta}\right), \end{aligned} \tag{5.11}$$

where $h(z) = \frac{g(z)}{(1-z)^2}$.

Let us first indicate how the proof will go and then actually prove the theorem. We will prove that for any given $c_1, \alpha\delta$ is decreasing in c_0 and $h(z)$ is an increasing function of

z. Since β involves only c_1 but not c_0, this will prove that the area in (5.11) is increasing in c_0 for given c_1. It was described in Theorem 4.2 that the set $\Gamma_\varepsilon(C_0)$ is an intersection of two circles (circles because $c = 1$ here). Since the area is increasing in c_0 for any given c_1, it then will follow that the minimizing pair (c_0, c_1) must be on the left boundary of the first circle S_0. This makes the design problem a one dimensional minimization problem. A simple differentiation argument then gives that the minimizing pair is in fact on the c_0–axis, i.e., has $c_1 = 0$. The required point then is just the left boundary point of S_0 along the c_0–axis, which is what the theorem asserts. We will now prove that $\alpha\delta$ is decreasing in c_0 for fixed c_1 and $h(z)$ is increasing in z.

Note that $\alpha\delta = f(c_0) \cdot f(1 - c_0)$,

$$\text{where } f(c_0) = c_0^2 + (K_1 + K_2)c_0 + c_1^2 + K_1K_2.$$

We will therefore need to show that given c_1,

$$f'(c_0)f(1 - c_0) \leq f(c_0)f'(1 - c_0)$$

$$\begin{aligned}
&\Leftrightarrow (2c_0 + (K_1 + K_2))((1 - c_0)^2 + (K_1 + K_2)(1 - c_0) + c_1^2 + K_1K_2) \\
&\leq (c_0^2 + (K_1 + K_2)c_0 + c_1^2 + K_1K_2)(2(1 - c_0) + K_1 + K_2) \\
&\Leftrightarrow 4c_0^3 - 6c_0^2 + 2[1 - K_1 - K_2 - K_1^2 - K_2^2 + 2c_1^2]c_0 \\
&\quad + K_1 + K_2 + K_1^2 + K_2^2 - 2c_1^2 \leq 0. \qquad (5.12)
\end{aligned}$$

The cubic in (5.12) can be factorized as $(2c_0 - 1)(2c_0^2 - 2c_0 - K_1 - K_2 - K_1^2 - K_2^2 + 2c_1^2)$. Thus its roots are $\frac{1}{2}$, $\frac{1 \pm \sqrt{\sigma_1}}{2}$, where

$$\sigma_1 = 1 + 2(K_1 + K_2 + K_1^2 + K_2^2 - 2c_1^2).$$

Now given c_1, c_0 satisfies $\frac{1}{2} \leq c_0 \leq 1$, and $c_1^2 \leq c_0(1 - c_0)$. Thus for fixed c_1, $\frac{1}{2} \leq c_0 \leq \frac{1+\sqrt{\sigma_2}}{2}$, where $\sigma_2 = 1 - 4c_1^2$. In order to prove (5.12) it is then sufficient to prove that the root $\frac{1+\sqrt{\sigma_1}}{2}$ is outside of the interval to which c_0 belongs which is clearly true since $\sigma_1 > \sigma_2$. This proves that for fixed c_1, $\alpha\delta$ is decreasing in c_0.

The fact that $h(z)$ is increasing in z follows quite easily on observing that

$$h(z) = \frac{2}{1-z} + \frac{\sqrt{z}}{(1-z)^{\frac{3}{2}}} \cdot \left(\frac{\pi}{2} - \sin^{-1}(1-2z)\right)$$

and each of $\frac{2}{1-z}$, $\frac{\sqrt{z}}{(1-z)^{\frac{3}{2}}}$ and $\frac{\pi}{2} - \sin^{-1}(1-2z)$ is a nonnegative increasing function for $0 \leq z \leq 1$. This now proves the theorem.

The area of the set S was proved above to be increasing in c_0 for fixed c_1. This implies that the global minima of the area is on the line $c_0 = \frac{1}{2}$. This reduces the dimension of the problem again to one and one can then show that the global minima is attained at the point $c_0 = \frac{1}{2}$, $c_1 = 0$ (notice this corresponds to the design assigning mass .5 at each of ± 1). However, for the variance function $w(x) = 1 + cx^2$, it is not in general true that the area is increasing in c_0 for any given c_1. It is also not necessarily the case that the global minima of the area is at the point $c_0 = \frac{1}{1+c}$, $c_1 = 0$. The global minima still occurs at this point if $c \leq 1$, but moves into a point $(c_0, 0)$ for $c_0 > \frac{1}{1+c}$ if $c > 1$. Using $r_0 = r_2 = .1$ and $K_1 = .1$, $K_2 = .5$ (in Theorem 5.4), we also numerically found the point (c_0, c_1) in $\Gamma_\varepsilon(C)$ at which the area of S is minimized. In each case we found that this point is simply that point in $\Gamma_\varepsilon(C)$ that is closest to the point of global minima of the area. As an example, if $c = 5$, then the smallest ε_0 equals .52 if $r_0 = r_2 = .1$. If one now takes an $\varepsilon \geq \varepsilon_0$, say $\varepsilon = .6$, then for points (c_0, c_1) in $\Gamma_\varepsilon(C)$, c_0 varies in the range $.5875 \leq c_0 \leq .6667$. The global minima of the area is attained at $c_0 = .8213$, $c_1 = 0$. The minimum of the area in $\Gamma_\varepsilon(C)$ is attained at $c_0 = .6667$, $c_1 = 0$, which is the point in $\Gamma_\varepsilon(C)$ closest to the point of the global minima.

6. 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 have also addressed the issue of sensitivity of the designs to

the uncertainty in the prior. 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 3.4 and to Mr. F.C. Chang for help with the numerical work.

References

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

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

R.J. Carroll and David Ruppert (1982). Robust estimation in heteroscedastic linear models. *Ann. Stat.* **10**, 429–441.

Chaloner, Kathryn (1984). Optimal Bayesian experimental design for linear models. *Ann. Stat.* **12**, 283–300.

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

DasGupta, A. and Studden, W.J. (1988). Robust Bayesian analysis and optimal experimental designs in normal linear models with many parameters – II. Technical report, Department of Statistics, Purdue University (submitted for publication).

Dey, D.K. and Berger, J. (1983). Combining coordinates in simultaneous estimation of normal means. *J. Stat. Planning and Inference* **8**, 143–160.

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**, 807–815.

Efron, B. and Morris, C. (1972). Limiting the risk of Bayes and empirical Bayes estimators – Part II: The empirical Bayes case. *J. Amer. Stat. Assoc.* **67**, 130–139.

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

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

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

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**, 176–181.

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

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

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

Leamer, E.E. (1978). Specification searches. Wiley, New York.

Lee, C–M.S. (1988). Constrained optimal designs, *Jour. Stat. Planning and Inf.* **18**, 377–389.

Pilz, J. (1979). Das bayes she Schätzproblem im linearen Regressionmodell. *Freiberger Forschungshefte* **117**, 21-55.

Pilz, J. (1981). Robust Bayes and minimax–Bayes estimation and design in linear regression. *Math. Operations forsch. Stat., Ser. Statistics* **12**, 163–177.

Polasek, W. (1984). Multivariate regression systems: estimation and sensitivity analysis of two–dimensional data. Robustness of Bayesian Analyses, J. Kadane (Ed.), North–Holland, Amsterdam.

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

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

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

Stein, C. (1981). Estimation of the mean of a multivariate normal distribution. *Ann. Stat.* **9**, 1135–1151.

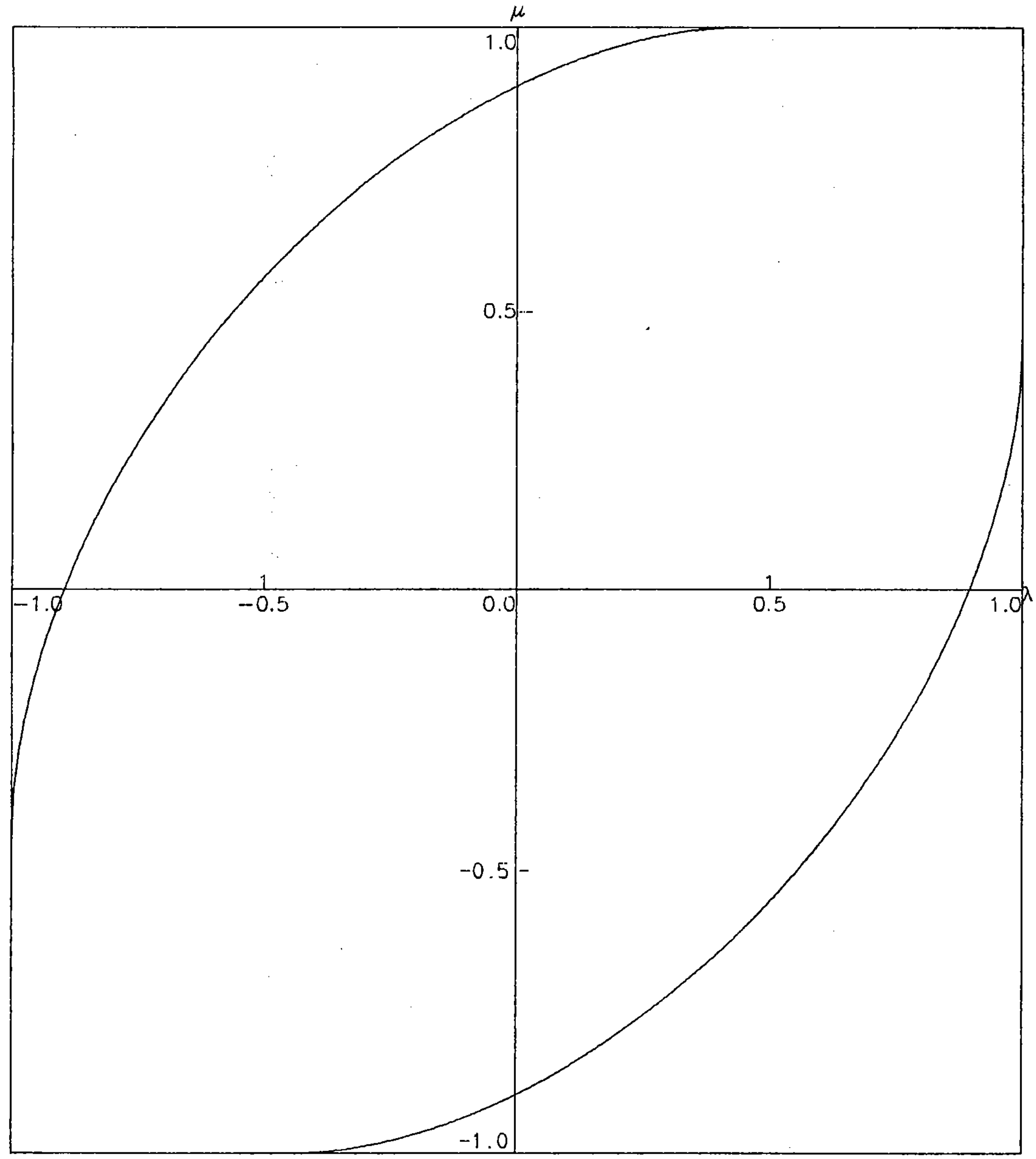

Figure 1: plot of the set S, a=c=1, b=.85