The Library of the Department of Statistics
North Carolina State University

Optimal Designs for Nonlinear Models with Correlated Error

BY Jacqueline M. Hughes-Oliver
Department of Statistics
North Carolina State University

Mimeo Series # 2503
September, 1997

Optimal Designs for Nonlinear Models with Correlated Error

BY JACQUELINE M. HUGHES-OLIVER

North Carolina State University

[12] Optimal designs for the efficient estimation of parameters in nonlinear regression models are usually local D-optimal designs, based on either a linear or quadratic approximation of the nonlinear surface, which assume that responses are independent. Very little has been done in the area of nonlinear models with correlated errors, and none of these works consider the effect of strong curvature when applying a linear approximation to the nonlinear surface in obtaining designs for small samples. This work compares the performance of three common design criteria in the presence of correlated errors. The results show that sometimes the more complicated design criteria (which account for correlation and/or curvature) can greatly improve the design, although not always. These ideas are illustrated through the intermediate product and exponential decay models.

1. Introduction. The advantages of a well-designed experiment are universally known; they include cost-effectiveness and the ability to quickly make valid inferences. When fitting nonlinear models, there is an additional benefit. A well-designed experiment for fitting a nonlinear model can allow the use of *simple* inferential procedures without sacrificing validity. Consider the model

$$y_i = f(\boldsymbol{x}_i; \boldsymbol{\theta}) + \epsilon_i, \qquad i = 1, \ldots, n \tag{1.1}$$

where y_i is the i^{th} response, $\boldsymbol{x}_i$ is the vector of covariates for the i^{th} response, $\boldsymbol{\theta}$ is a p-vector of unknown parameters, $f(\cdot, \cdot)$ is a known functional form, and ϵ_i is a random variable with mean 0 and variance σ^2. The vector representation of (1.1) is

$$\boldsymbol{y} = \boldsymbol{f}(\boldsymbol{\theta}) + \boldsymbol{\epsilon}, \tag{1.2}$$

Received ???; revised ???.

[1] *AMS* 1991 *subject classifications.* Primary 62K05; secondary 62J02, 62M10.

[2] *Key words and phrases.* Curvature, mean squared error criterion, quadratic volume criterion.

where $\boldsymbol{y} = (y_1, y_2, \ldots, y_n)'$, $\boldsymbol{f}(\boldsymbol{\theta}) = \big(f(x_1;\boldsymbol{\theta}), f(x_2;\boldsymbol{\theta}), \ldots, f(x_n;\boldsymbol{\theta})\big)'$, $\boldsymbol{\epsilon} = (\epsilon_1, \epsilon_2, \ldots, \epsilon_n)'$, and $\boldsymbol{\epsilon} \sim (\boldsymbol{0}, \sigma^2\boldsymbol{I})$. The least squares estimator $\widehat{\boldsymbol{\theta}}$ of $\boldsymbol{\theta}$ minimizes the sum of squares function

$$\begin{aligned} S(\boldsymbol{\theta}) &= [\boldsymbol{y} - \boldsymbol{f}(\boldsymbol{\theta})]'[\boldsymbol{y} - \boldsymbol{f}(\boldsymbol{\theta})] \\ &= \sum_{i=1}^{n} [y_i - f(x_i;\boldsymbol{\theta})]^2. \end{aligned}$$

The choice of joint inference region for $\boldsymbol{\theta}$ is usually limited either to one based on the asymptotic properties of $\widehat{\boldsymbol{\theta}}$, or one based on the likelihood function of (1.2). The $(1-\alpha)100\%$ joint inference region

$$\{\boldsymbol{\theta} : (\boldsymbol{\theta} - \widehat{\boldsymbol{\theta}})'\widehat{\boldsymbol{F}}_{\cdot}'\widehat{\boldsymbol{F}}_{\cdot}(\boldsymbol{\theta} - \widehat{\boldsymbol{\theta}}) \leq ps^2 F^{\alpha}_{p,n-p}\}, \tag{1.3}$$

where $\boldsymbol{F}_{\cdot} = \dfrac{\delta}{\delta\boldsymbol{\theta}'}\boldsymbol{f}(\boldsymbol{\theta})$, $\widehat{\boldsymbol{F}}_{\cdot}$ is $\boldsymbol{F}_{\cdot}$ evaluated at $\widehat{\boldsymbol{\theta}}$, and $s^2 = S(\widehat{\boldsymbol{\theta}})/(n-p)$, is easy to compute and is readily available from most computer packages. These regions are ellipsoids of the approximating multivariate normal distribution of $\widehat{\boldsymbol{\theta}}$. Unfortunately, they are also very dependent on the assumption that the expectation surface in the neighborhood of $\widehat{\boldsymbol{\theta}}$ is well approximated by the tangent plane in R^p, in other words, that

$$\boldsymbol{f}(\boldsymbol{\theta}) \approx \boldsymbol{f}(\widehat{\boldsymbol{\theta}}) + \widehat{\boldsymbol{F}}_{\cdot}(\boldsymbol{\theta} - \widehat{\boldsymbol{\theta}}) \tag{1.4}$$

is a good approximation. Unfortunately, this is not always the case, even for moderately large samples (Bates and Watts, 1988; Seber and Wild, 1989).

The likelihood-based $(1-\alpha)100\%$ joint inference region

$$\left\{\boldsymbol{\theta} : S(\boldsymbol{\theta}) \leq S(\widehat{\boldsymbol{\theta}})\right)\left(1 + \frac{p}{n-p}F^{\alpha}_{p,n-p}\right)\Bigg\} \tag{1.5}$$

is not ellipsoidal, is more difficult to compute and display, and is not readily available from popular computer packages. It is, however, preferable because it is not dependent on the assumption of linearity given in (1.4).

Is it possible to design an experiment for fitting a nonlinear model such that any combination of the following hold: $\widehat{\boldsymbol{\theta}}$ is both precise and accurate; asymptotic properties of the form (1.4) are not needed; the approximation (1.4) is reasonable, even for small samples? Optimal selection of design points for nonlinear estimation has been investigated by Box and Lucas (1959), Hill and Hunter (1974), Papakyriazis (1978), Hamilton and Watts (1985), Clarke and Haines (1995), to name a few,

but only a handful of these authors (for example, Hamilton and Watts 1985, Clarke and Haines 1995) have suggested design criteria which are not based on the asymptotic properties of $\widehat{\theta}$, and are thus applicable for the design of small experiments.

The goal of this article is to investigate and compare the performance of three design criteria applied to correlated responses in small experiments. Many processes, such as chemical reactions, occur over time and it is only reasonable to believe that the optimal designs are a function of the strength of correlation between responses. Section 2 contains a review of three optimality criteria. Section 3 describes the adjustment necessary for correlated errors. Section 4 contains applications to two nonlinear models, and Section 5 contains concluding remarks.

2. Optimality criteria. Unlike linear models, the optimal design for a nonlinear model can be heavily dependent on the value of θ. One approach is to consider θ_0-locally optimum designs, which are obtained by assuming the true value of θ is θ_0. Other approaches include sequential, minimax, and Bayesian designs (Atkinson and Donev, 1992; Chaudhuri and Mykland, 1993; Haines, 1995; Mukhopadhyay and Haines, 1995; Seber and Wild, 1989). In this article, interest is limited to locally optimum designs.

While there are a myriad of local optimality criteria, the following are pursued here: (linear) D-optimality, quadratic mean squared error D-optimality, and quadratic joint inference region volume optimality. They are described in the following subsections.

2.1. *D-optimality.* Asymptotic properties of the likelihood of (1.2) lead to $\boldsymbol{F}'.\boldsymbol{F}.$ as the asymptotic information matrix of $\widehat{\theta}$. The D-optimality criterion is equivalent to maximizing the determinant of this asymptotic information matrix (Pukelsheim, 1993). In other words, the D_{opt} criterion is to find $\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_n$ to maximize

$$|\boldsymbol{F}'.\boldsymbol{F}.|, \tag{2.6}$$

where calculations are done using $\theta = \theta_0$.

2.2. *Mean squared error D-optimality.* Under fairly general conditions, when the linear tangent-plane approximation (1.4) is valid, the least squares estimator $\widehat{\theta}$ is asymptotically unbiased and

normally distributed (Seber and Wild, 1989, pg 181). The D_{opt} criterion of Section 2.1 assumes this and so includes no bias term in the resulting information matrix. However, if the linear tangent-plane approximation (1.4) is inappropriate for a given model-design combination, then $\widehat{\boldsymbol{\theta}}$ can be very seriously biased (Box 1971, Clarke 1980, Hougaard 1985, Clarke and Haines 1995). Consequently, it is eminently sensible to consider the mean squared error (MSE) of $\widehat{\boldsymbol{\theta}}$, based on at least a quadratic approximation of $\boldsymbol{f}(\boldsymbol{\theta})$. This article considers the quadratic approximation only. The quadratic mean squared error D-optimality, $MSE_Q D_{opt}$, criterion minimizes $\det(MSE_Q(\widehat{\boldsymbol{\theta}}))$, the generalized mean squared error of $\widehat{\boldsymbol{\theta}}$, based on a quadratic approximation of the expectation surface.

$MSE_Q(\widehat{\boldsymbol{\theta}})$ has been derived by many authors. (There are also formulas based on higher-order approximations of $\boldsymbol{f}(\boldsymbol{\theta})$. See Clarke 1980, Hougaard 1985). Only the basic required formulas will be presented here. (See Seber and Wild (1989) for a more complete treatment, as well as a review of multiplication of three dimensional arrays.) The notation in this article follows that of Seber and Wild (1989). First, apply the QR decomposition to $\boldsymbol{F}.$ to get

$$\boldsymbol{F}. = \boldsymbol{Q}\boldsymbol{R}_1 = \left(\underbrace{\boldsymbol{Q}_p}_{n\times p} \middle| \underbrace{\boldsymbol{Q}_{n-p}}_{n\times(n-p)}\right)\begin{pmatrix}\boldsymbol{R}_{11}\\ 0\end{pmatrix} = \boldsymbol{Q}_p\boldsymbol{R}_{11}$$

Let $\boldsymbol{F}..$ be the $n \times p \times p$ three-dimensional array of second derivatives of $\boldsymbol{f}(\boldsymbol{\theta})$, where $\boldsymbol{F}_{i}.. = \left[\left(\frac{\delta^2}{\delta\theta_r \delta\theta_s} f(\boldsymbol{x}_i;\boldsymbol{\theta})\right)\right]$, $r,s = 1,2,\ldots,p$ is the i^{th} face of $\boldsymbol{F}..$ for $i = 1,2,\ldots,n$. Now define the $n \times p \times p$ array $\boldsymbol{G}.. = (\boldsymbol{R}_{11}^{-1})'\boldsymbol{F}..(\boldsymbol{R}_{11}^{-1})$. The intrinsic curvature array is the $(n-p) \times p \times p$ array $\boldsymbol{A}_{..}^{N} = [\boldsymbol{Q}'_{n-p}][\boldsymbol{G}..]$ which measures the "inherent" curvature in the model itself, irrespective of parameterization. The parameter-effects curvature array is the $p \times p \times p$ array $\boldsymbol{A}_{..}^{T} = [\boldsymbol{Q}'_p][\boldsymbol{G}..]$ and can be eliminated with a clever choice of parameterization. Also let $\boldsymbol{M}^T = [(\text{trace}\{\boldsymbol{A}_{i}^{T}..\boldsymbol{A}_{j}^{T}..\})]$, $i,j = 1,2,\ldots,p$ and $\boldsymbol{J}^T = [(\text{trace}\{\boldsymbol{A}_{i}^{T}..\}\text{trace}\{\boldsymbol{A}_{j}^{T}..\})]$, $i,j = 1,2,\ldots,p$ be $p \times p$ matrices which are functions of $\boldsymbol{A}_{..}^{T}$, but not $\boldsymbol{A}_{..}^{N}$. Then

$$MSE_Q(\widehat{\boldsymbol{\theta}}) = \sigma^2\boldsymbol{R}_{11}^{-1}\left\{\boldsymbol{I} + \sigma^2\left[\frac{1}{4}\boldsymbol{J}^T + \sum_{i=1}^{n-p}(\boldsymbol{A}_{..}^{N})^2 + \frac{1}{2}\boldsymbol{M}^T\right]\right\}\left(\boldsymbol{R}_{11}^{-1}\right)'. \tag{2.7}$$

The $MSE_Q D_{opt}$ criterion is to find $\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_n$ to minimize

$$|(\boldsymbol{F}'.\boldsymbol{F}.)^{-1}|\left|\boldsymbol{I} + \sigma^2\left[\frac{1}{4}\boldsymbol{J}^T + \sum_{i=1}^{n-p}(\boldsymbol{A}_{..}^{N})^2 + \frac{1}{2}\boldsymbol{M}^T\right]\right|, \tag{2.8}$$

where all calculations are done using $\theta = \theta_0$ and $\sigma^2 = \sigma_0^2$. Note that if there is no curvature, that is, $A^N_{\cdot\cdot} = A^T_{\cdot\cdot} = 0$, then the $MSE_Q D_{opt}$ criterion reduces to the D_{opt} criterion.

For $n = p$, Equation (2.8) is simplified. When $F_{\cdot}$ is square and nonsingular, this causes the expectation surface to be flat and the same as the tangent plane. This results in no intrinsic curvature ($A^N_{\cdot\cdot} = 0$), but there may still be parameter-effects curvature (Seber and Wild, 1989, pg. 262–264). In addition, we get $|(F'_{\cdot}F_{\cdot})^{-1}| = 1/(|F_{\cdot}|)^2$.

2.3. *Joint inference region volume optimality.* If the linear approximation (1.4) is invalid, then the joint inference region (1.3) is inappropriate but may be modified so that it is based on a better, for example, quadratic, approximation of the expectation surface. Hamilton and Watts (1985) provide such an approximation and suggest that designs be selected to minimize the approximate volume of the resulting joint inference region. The Vol_Q criterion is to find $x_1, x_2, \ldots, x_n$ to minimize

$$|(F'_{\cdot}F_{\cdot})^{-1/2}|\left[1 + \frac{\sigma^2\chi^2_{p,\alpha}}{2(p+2)}\sum_{s=1}^{p}\{[\text{trace}(A^T_{(s)})]^2 + \text{trace}[(A^T_{(s)})^2] + c'a_{ss}\}\right], \tag{2.9}$$

where all calculations are done using $\theta = \theta_0$, $\sigma^2 = \sigma_0^2$, and α. The notation is similar to that of Section 2.2, with the following additions: $\chi^2_{p,\alpha}$ is the $(1-\alpha)100^{th}$ percentile of the chi-square distribution with p degrees of freedom; $A^T_{(s)}$ is the s^{th} column slice of $A^T_{\cdot\cdot}$, meaning it is a $p \times p$ matrix with $(i,r)^{th}$ entry a^T_{irs} which is the $(r,s)^{th}$ entry in the i^{th} face of $A^T_{\cdot\cdot}$; $c' = (\text{trace}(A^T_{(1)}), \text{trace}(A^T_{(2)}), \ldots, \text{trace}(A^T_{(p)}))$; and a_{ss} is the $n \times 1$ vector with i^{th} entry a_{iss}. However, Hamilton and Watts comment that this approximation can sometimes be bad and that (2.9) may not even be calculable. This usually happens only in cases where the curvature of a model-design combination is really high.

Finally, if there is no curvature, that is, $A^N_{\cdot\cdot} = A^T_{\cdot\cdot} = 0$, then the Vol_Q criterion reduces to the D_{opt} criterion.

3. Correlated errors. Correlated errors occur very often in practice, for example, when the independent variable is time, and this correlation should not be ignored as it may affect the choice of design (Atkinson and Donev, 1992, pg 197). Consider the model

$$y = f(\theta) + \epsilon, \epsilon \sim (0, \sigma^2 V), \tag{3.10}$$

where $\boldsymbol{V} = \boldsymbol{V}(\boldsymbol{\alpha})$ is a function of $\boldsymbol{\alpha}$ and does not involve $\boldsymbol{\theta}$. If $\boldsymbol{\alpha}$ is known, we can apply a transformation to remove the dependence:

$$\begin{aligned} &\boldsymbol{L}^{-1}\boldsymbol{y} = \boldsymbol{L}^{-1}\boldsymbol{f}(\boldsymbol{\theta}) + \boldsymbol{L}^{-1}\boldsymbol{\epsilon}, \\ &\hookrightarrow \boldsymbol{y}^* = \boldsymbol{f}^*(\boldsymbol{\theta}) + \boldsymbol{\epsilon}^*, \qquad \boldsymbol{\epsilon}^* \sim (0, \sigma^2\boldsymbol{I}), \end{aligned} \tag{3.11}$$

where $\boldsymbol{V} = \boldsymbol{L}\boldsymbol{L}'$ is the Cholesky decomposition of $\boldsymbol{V}$. Using model (3.11), we may now proceed as in the case of independent errors. It is easily seen that

$$\boldsymbol{F}^*_{\cdot\cdot} = \boldsymbol{L}^{-1}\boldsymbol{F}_{\cdot\cdot}$$

$$\boldsymbol{F}^*_{\cdot\cdot\cdot} = [\boldsymbol{L}^{-1}][\boldsymbol{F}_{\cdot\cdot\cdot}]$$

to allow computation of all formulas in Section 2.

4. Examples. Two-point ($n = 2$) optimal designs are obtained for two commonly used models. All three criteria from Section 2 are applied, using varying degrees of correlation. Suppose the selected design is $\{x_1, x_2\}$. The correlation assumed here is the exponential decay model $Corr(\epsilon_1, \epsilon_2) = \rho^{|x_1 - x_2|}$. Results are presented for $\rho = 0, .1, \ldots, .9$.

To the extent possible, symbolic computation in Maple V was used to derive formulas and to perform the optimization of the D_{opt} objective function. The objective functions for the $MSE_Q D_{opt}$ and Vol_Q criteria were too complicated to allow optimization in Maple. For these cases, the `fortran` export function in Maple was used to allow numerical optimization to be done using double precision IMSL routines `duminf` and `dumpol`. Without loss of generality, the restriction $0 < x_1 < x_2$ was enforced by optimizing over the transformed (unrestricted) space of (δ_1, δ_2) where $x_1 = \exp(\delta_1)$ and $x_2 = \exp(\delta_1) + \exp(\delta_2)$.

4.1. *Intermediate product model.* The intermediate product model has been studied by many authors (see, for example, Atkinson and Donev 1992, Bates and Watts 1988, Box and Lucas 1959, Hamilton and Watts 1985). It is useful in modeling the concentration of the intermediate substance

created from two consecutive irreversible first-order chemical reaction. The expected response is

$$f(x;\theta) = \frac{\theta_1}{\theta_1 - \theta_2}[\exp(-\theta_2 x) - \exp(-\theta_1 x)] \qquad \theta_1, \theta_2, x > 0. \tag{4.12}$$

The resulting concentration increases from 0 at time $x = 0$ to its maximum at time $x = [\ln\theta_2 - \ln\theta_1]/[\theta_2 - \theta_1]$, then decreases to 0 as x increase. To maintain consistency with Box and Lucas (1959) and Hamilton and Watts (1985), locally optimum two-point designs are obtained using $\theta_0 = (.7, .2)$, $\sigma^2 = .01$, and $\alpha = .05$.

Equations (2.6), (2.8), (2.9), sufficiently modified as per Section 3, are used to obtain the objective functions for all three criteria, for different values of ρ. The resulting optima are shown in Table 1. For example, $\rho = 0$ results in the usual designs obtained from assuming independent responses. The $\rho = 0, D_{opt}$ design $(1.23, 6.86)$ is the same obtained by Box and Lucas (1959) and Hamilton and Watts (1985). The $\rho = 0, Vol_Q$ design $(1.04, 5.59)$ is slightly different from the $(1.04, 5.56)$ reported by Hamilton and Watts (1985). As correlation (ρ) increases, the optimum values of x_1 and x_2 become closer to each other.

Figure 1 displays the optimum designs of Table 1 relative to the corresponding $\rho = 0, D_{opt}$ design, as a function of ρ. All three design criteria appear to be approaching the same limiting design as $\rho \to 1$.

While Figure 1 helps the reader see the pattern in the optimum designs, it gives no information on the relevant difference between these designs — their adjusted relative efficiencies with respect to the criterion objective function. The adjusted relative efficiency (AdRE) used here for a criterion whose objective function $O(x_1, \ldots, x_n)$ needs to be maximized is

$$AdRE = \left[\frac{O(\rho = 0, D_{opt} \text{ design})}{\max_{x_1,\ldots,x_n} O(x_1, \ldots, x_n)}\right]^{1/p}, \tag{4.13}$$

where p is the number of parameters in the model. For a criterion whose objective function $O(x_1, \ldots, x_n)$ needs to be minimized,

$$AdRE = \left[\frac{\min_{x_1,\ldots,x_n} O(x_1, \ldots, x_n)}{O(\rho = 0, D_{opt} \text{ design})}\right]^{1/p}. \tag{4.14}$$

Efficiency relative to the $\rho = 0, D_{opt}$ design seems most appropriate here because this is the default design (i.e., all other criteria simplify to give this design) when both correlation and curvature are ignored or do not exist. An adjusted relative efficiency value near 1 means little information is lost

TABLE 1

Two-point optimum designs for the intermediate product model.

	Optimality Criteria		
ρ	D_{opt}	$Vol_Q D_{opt}$	$MSE_Q D_{opt}$
0	(1.23,6.86)	(1.04,5.59)	(1.18,6.48)
.1	(1.23,6.86)	(1.04,5.59)	(1.18,6.48)
.2	(1.23,6.86)	(1.04,5.59)	(1.18,6.48)
.3	(1.23,6.86)	(1.04,5.59)	(1.18,6.48)
.4	(1.23,6.86)	(1.04,5.59)	(1.18,6.48)
.5	(1.23,6.85)	(1.05,5.58)	(1.18,6.48)
.6	(1.23,6.82)	(1.06,5.53)	(1.18,6.44)
.7	(1.24,6.68)	(1.08,5.40)	(1.20,6.29)
.8	(1.26,6.29)	(1.13,5.10)	(1.23,5.86)
.9	(1.32,5.55)	(1.21,4.64)	(1.29,4.98)

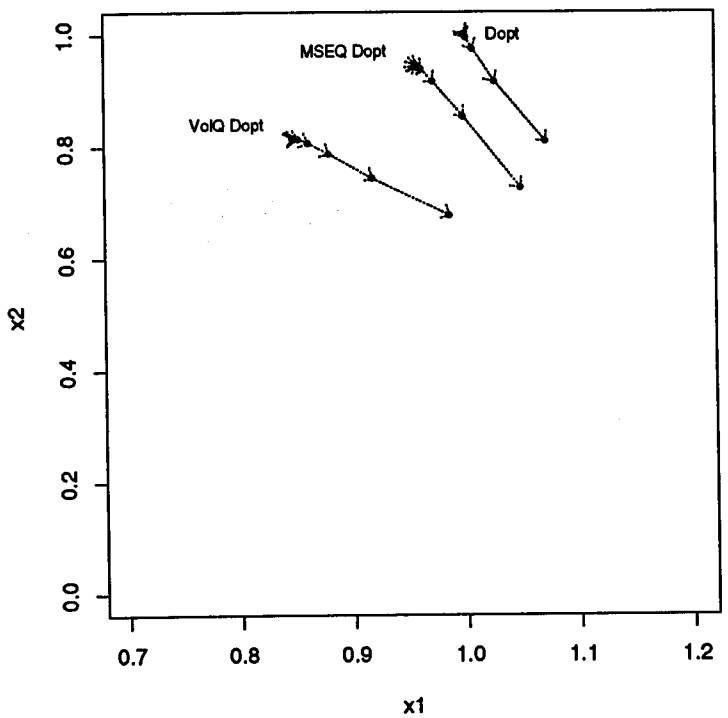

FIG. 1. *Intermediate product model. Optimal D_{opt}, $MSE_Q D_{opt}$, and Vol_Q designs for $\rho = 0, .1, \ldots, .9$, relative to the $\rho = 0, D_{opt}$ design. The arrows show the direction of change for increasing ρ.*

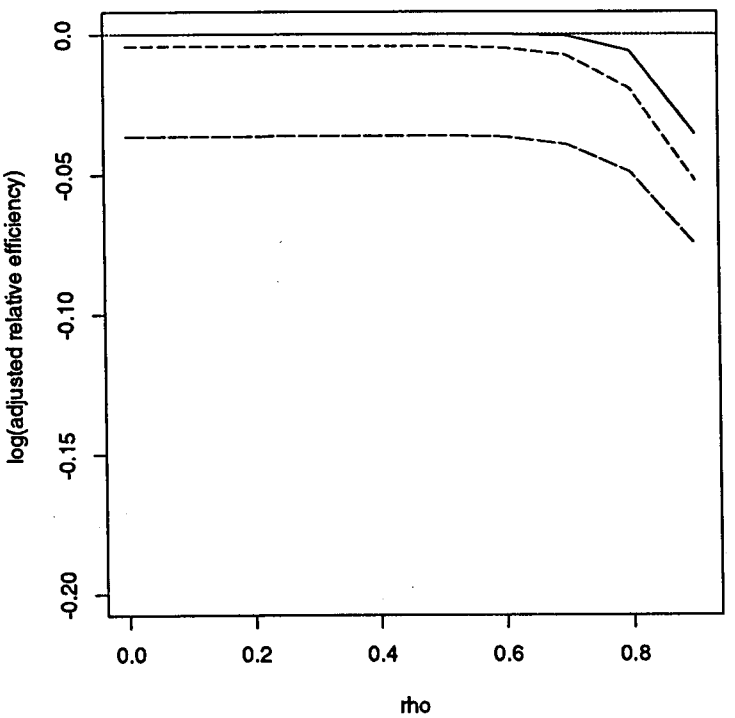

FIG. 2. *Intermediate product model. Adjusted efficiencies of the "best" designs relative to the* $\rho = 0, D_{opt}$ *design, for the* D_{opt}, MSE_QD_{opt}, *and* Vol_Q *criteria, and for* $\rho = 0, .1, \ldots, .9$. *The curves are:* D_{opt} ———; MSE_QD_{opt} - - - -; Vol_Q — — — —.

by ignoring correlation and/or curvature. Figure 2 contains the logarithm of this adjusted relative efficiency for the designs in Table 1, as a function of ρ, with separate curves for the different criteria. The D_{opt} criterion, which accounts for correlation but not curvature, gives the least improvement over the $\rho = 0, D_{opt}$ design. The MSE_QD_{opt} criterion is marginally better than the D_{opt} criterion, while the Vol_Q criterion gives the greatest improvement. Unfortunately, the maximum improvement corresponds to an $AdRE$ value of 0.928, which is practically no improvement at all. The conclusion is that for the intermediate product model with $\theta_0 = (.7, .2)$, very little is gained from the more complicated procedures that account for correlation and/or curvature.

4.2. *Exponential decay.* The exponential decay model has also been studied by many (see, for example, Atkinson and Donev 1992, Bates and Watts 1988, Seber and Wild 1989). It is useful in modeling concentration of the substance created in a single irreversible first-order chemical reaction. The expected response is

$$f(x;\theta) = \theta_1[1 - \exp(-\theta_2 x)] \qquad 0 < \theta_1, \theta_2, 0 < x < B. \tag{4.15}$$

The resulting concentration increases from 0 at time $x = 0$, then asymptotes to θ_1, the initial concentration of the original substance, as time x increases. The rate of increase is governed by θ_2, with larger values of θ_2 implying a rapid rate of increase of the new substance.

The exponential decay model has two distinguishing features when $\rho = 0$. The first is that because θ_1 enters the model (4.15) linearly, it has no effect on the optimal design. The second is that the optimum two-point D_{opt} design is analytically very easy to obtain as

$$x_2 = B, \qquad x_1 = \frac{1}{\theta_2} - \frac{B\exp(-\theta_2 B)}{1 - \exp(-\theta_2 B)}.$$

However, the second feature disappears for $\rho > 0$, as will be demonstrated below. Locally optimum two-point designs are obtained using $\theta_0 = (1, 10)$, $B = 5$, $\sigma^2 = .01$, and $\alpha = .05$.

The optimum designs are presented in the same format used for the intermediate product model. Table 2 lists the designs obtained from the various criteria and values of ρ; Figure 3 plots the designs relative to the $\rho = 0, D_{opt}$ design; and Figure 4 plots the logarithm of the adjusted efficiencies of the "best" designs relative to the $\rho = 0, D_{opt}$ design. Once again, from Table 2 we see that the optimum values of x_1 and x_2 become closer to each other as ρ increases. We also see that a very small amount of correlation can cause the design to change rather dramatically. From Figure 3 we see that all three design criteria appear to be approaching the same limiting design as $\rho \longrightarrow 1$, but at a much faster rate than for the intermediate product model. Figure 4 shows the huge improvements possible by accounting either for correlation or curvature or both. (The scale of Figure 4 is 100-times the scale of Figure 2!) For this exponential decay model, the $MSE_Q D_{opt}$ criterion gives the greatest improvement. The conclusion is that, unlike with the intermediate product model, the designs for the exponential decay model change a lot as ρ changes and/or different curvature adjustments are made. For example, even if no curvature adjustment is made but correlation is accounted for, the

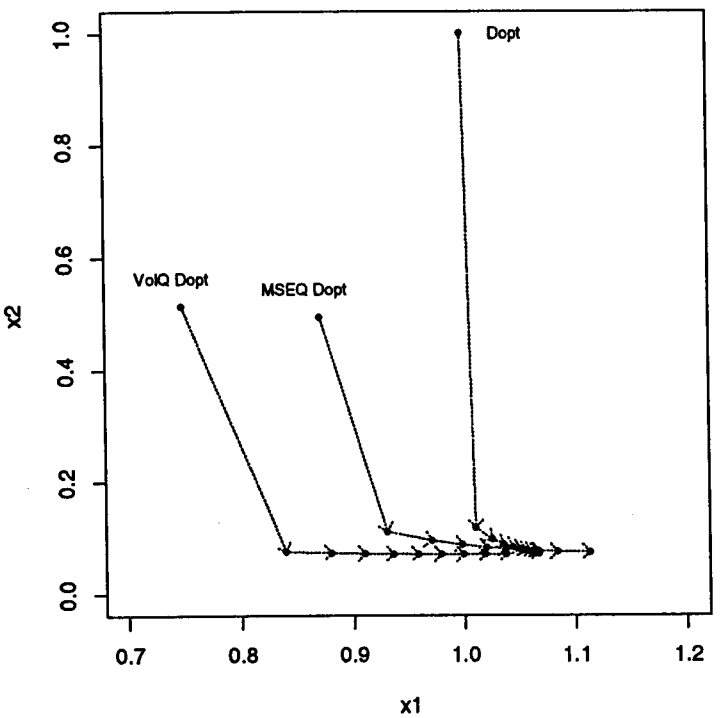

FIG. 3. *Exponential decay model. Optimal D_{opt}, $MSE_Q D_{opt}$, and Vol_Q designs for $\rho = 0, .1, \ldots, .9$, relative to the $\rho = 0, D_{opt}$ design. The arrows show the direction of change for increasing ρ.*

$\rho = .9, D_{opt}$ design is almost 3 times more informative than the $\rho = 0, D_{opt}$ design. The $MSE_Q D_{opt}$ designs offer such an improvement over the $\rho = 0, D_{opt}$ design that the $AdRE$ is practically zero.

5. Discussion. This article illustrates some of the effects of correlation and inappropriate use of the tangent-plane linear approximation to the expectation surface on the selection of designs for estimation of $\boldsymbol{\theta}$. The result depends on the model and on $\boldsymbol{\theta}_0$. There are no general "rules" to follow. For some models, the improvement due to correlation and/or curvature adjustments can be large; for other models, the adjustments are not worth the effort. If one suspects correlated errors then optimal designs based on this correlation structure should be compared (in terms of adjusted relative efficiency) to the optimal design for independent errors. The same comment applies to checking the

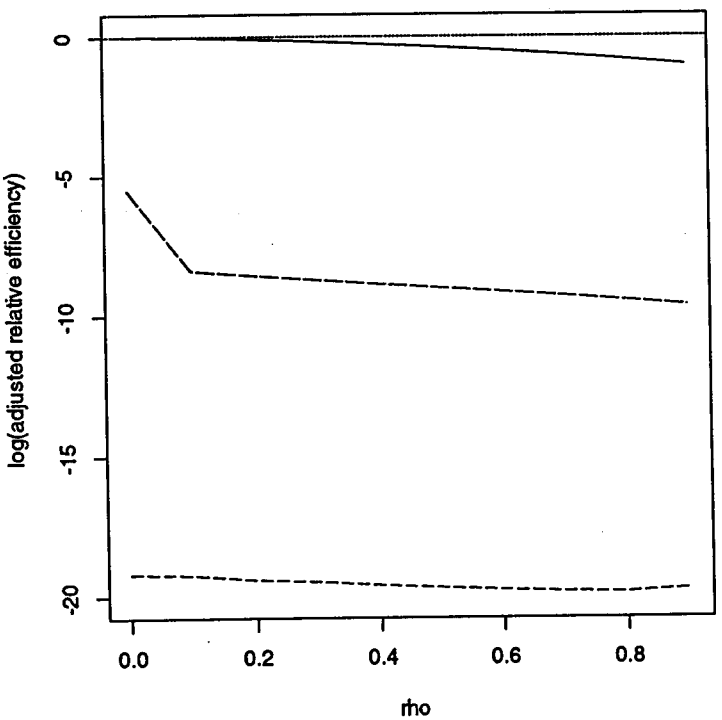

FIG. 4. *Exponential decay model. Adjusted efficiencies of the "best" designs relative to the* $\rho = 0, D_{opt}$ *design, for the* D_{opt}, $MSE_Q D_{opt}$, *and* Vol_Q *criteria, and for* $\rho = 0, .1, \ldots, .9$. *The curves are:* D_{opt} ———; $MSE_Q D_{opt}$ - - - -; Vol_Q — — — —.

TABLE 2

Two-point optimum designs for the exponential demand model.

	Optimality Criteria		
ρ	D_{opt}	$Vol_Q D_{opt}$	$MSE_Q D_{opt}$
0	(0.100,5.000)	(0.087,2.470)	(0.075,2.563)
.1	(0.101,0.587)	(0.093,0.553)	(0.084,0.375)
.2	(0.102,0.480)	(0.097,0.471)	(0.088,0.363)
.3	(0.104,0.438)	(0.100,0.435)	(0.091,0.358)
.4	(0.104,0.414)	(0.102,0.413)	(0.094,0.355)
.5	(0.105,0.397)	(0.104,0.398)	(0.096,0.353)
.6	(0.106,0.385)	(0.105,0.387)	(0.098,0.352)
.7	(0.106,0.376)	(0.107,0.379)	(0.100,0.352)
.8	(0.106,0.368)	(0.108,0.374)	(0.102,0.352)
.9	(0.107,0.362)	(0.111,0.373)	(0.104,0.352)

effect of curvature adjustments. The effect of correlation and curvature appear to interact in that when there is very little curvature (as in the intermediate product model), correlation has a small effect, but when there is a lot a curvature (as in the exponential decay model), correlation has a big effect on changing the optimal design.

Much work remains to be done. For example, do the optimal designs from the different criteria approach a limiting design as $\rho \to 1$. If so, do they approach the same limit, as suggested by Figures 1 and 3? Can a formal relationship between correlation and curvature be derived? For any given criterion, the designs corresponding to different ρ seem to fall on a curve, and this curve is almost a straight line for the intermediate product model.

The results presented here have been limited to $n = p = 2$. What can we expect for $n > p$? Are the designs replicated p-point designs? In general, how does one confirm that a continuous ($n = \infty$) design has been found? The equivalence theorem cannot, in general, be applied to the combination of criteria (including correlation) presented here.

REFERENCES

ATKINSON, A. C. and DONEV, A. N. (1992). *Optimum Experimental Design.* Oxford University Press. New York.

BATES, D. M. and WATTS, D. G. (1988). it Nonlinear Regression Analysis & its Applications. Wiley.

BOX, G. E. P. and LUCAS, H. L. (1959). Design of experiments in non-linear situations. *Biometrika* **47** 77–99.

BOX, M. J. (1971). Bias in non-linear estimation. *J. R. Statist. Soc.* **33** 171–201.

CHAUDHURI, P. and MYKLAND, P. A. (1993). Nonlinear experiments: Optimal design and inference based on likelihood. *J. Amer. Statist. Assoc.* **88** 538–546.

CLARKE, G. P. Y. (1980). Moments of the least squares estimators in a non-linear regression model. *J. R. Statist. Soc.* B **42** 227–237.

CLARKE, G. P. Y. and HAINES, L. M. (1995). Optimal design for models incorporating the Richards function. *Proceedings of the 10th International Workshop on Statistical Modelling.* 61–66. Springer-Verlag.

HAINES, L. M. (1995). A geometric approach to optimal design for one-parameter non-linear models. *J. R. Statist. Soc.* B **57** 575–598.

HAMILTON, D. C. and WATTS, D. G. (1985). Accounting for intrinsic nonlinearity in nonlinear regression parameter inference regions. *Technometrics* **27** 241–250.

HILL, W. J. and HUNTER, W. G. (1974). Design of experiments for subsets of parameters. *Technometrics* **16** 425–434.

HOUGAARD, P. (1985). The appropriateness of the asymptotic distribution in a nonlinear regression. *J. R. Statist. Soc.* B **47** 103–114.

MUKHOPADHYAY, S. and HAINES, L. M. (1995). Bayesian D-optimal designs for the exponential growth model. *J. Statist. Plann. Inference* **44** 385–397.

PAPAKYRIAZIS, P. A. (1978). Optimal experimental design in econometrics. *J. Econometrics* **7** 351–372.

PUKELSHEIM, F.. (1993). *Optimal Design of Experiments.* Wiley.

SEBER, G. A. F. and WILD, C. J. (1989). *Nonlinear Regression.* Wiley.

DEPARTMENT OF STATISTICS
NORTH CAROLINA STATE UNIVERSITY
RALEIGH, NC 27695-8203