Robust Optimum Designs for Transformation of the Response in Nonlinear Models

Anthony C. Atkinson*, Department of Statistics,
London School of Economics, London WC2A 2AE, UK

November 22, 2004

Abstract

Often the responses from nonlinear models have to be transformed to achieve error distributions that are symmetric and have constant variance. This transformation can have a dramatic effect on optimum designs for the parameters. The paper develops methods for the design of experiments when the value of the power transformation λ is not precisely known. Examples are given of optimum designs for particular values of λ, together with the efficiency of these designs as λ varies. It is shown that, by a suitable choice of prior, Bayesian D-optimum designs can be found that are robust to the true value of λ. A typical equispaced design is introduced and the properties of designs studied under various departures from the assumptions.

Keywords: Bayesian D-optimality; Box-Cox transformation; Chemical kinetics; D-optimum design; Nonlinear model; Parameter sensitivities; Robust designs; Transform both sides.

1 Introduction

Large responses usually have a higher variance than small responses. Power transformation of the observations then provides a powerful method of obtaining a response with symmetrical errors of constant variance. Least squares is consequently efficient for fitting models and estimating parameters. The results given here show that such transformations can have a large effect on good designs for estimating the parameters of nonlinear models. The designs arise from transforming both sides of the model, so ensuring any constraints

*a.c.atkinson@lse.ac.uk

on the response imposed by the model still hold after transformation. As an example, the nonlinear models in this paper arise in pharmacokinetics and chemical kinetics; the concentrations of the chemicals vary over time, but the sum of the concentrations remains fixed, a property not shared by the power transformation of the concentrations.

The transformation is indexed by a parameter λ. Atkinson (2003) finds designs for known λ. In practice, λ will rarely be known. In this paper we find the efficiencies of designs as λ varies and then introduce Bayesian design for a prior distribution of values of λ. Suitable choice of prior yields designs that are robust to the value of λ, having high efficiency over a wide range of values of the parameter. Optimum design theory is then used to study the robustness of designs to mis-specification of the model.

The paper starts in §2 with a statement of the design problem when both sides of the model are to be transformed. §§3, 4 and 5 develop the relevant design theory and define design efficiencies as λ varies. The first example, of exponential decay, is in §6. Bayesian optimum designs are derived in §7 and an Equivalence Theorem formulated. This is applied to finding locally optimum designs for the model of exponential decay and, in §8, to a more complicated example involving a reversible reaction. For comparison a typical equispaced design is introduced in §9. The following five sections compare designs when the errors are correlated, when parameter uncertainty is expressed through a prior distribution and when the nonlinear model contains fewer or more parameters than were designed for. The paper concludes in §15 with a discussion of possible extensions of the results.

2 Transformations

The experimental design problem considered is that of selecting a set of times t_i at which readings y_i should be taken with weights w_i in order to obtain good estimates of the parameter vector ψ in the nonlinear mechanistic model

$$y = \eta(t, \psi). \tag{1}$$

This becomes a statistical problem when error is introduced.

The standard assumption is that the error variance is independent of the mean. If, however, the variance increases with the mean, Taylor expansion shows that power transformation of the response provides a random variable with variance approximately independent of the mean. For transformation of just the response y in a regression model, the parametric power transfor-

mation of Box and Cox (1964) is

$$y(\lambda) = \begin{cases} (y^\lambda - 1)/\lambda & \lambda \neq 0 \\ \log y & \lambda = 0. \end{cases} \tag{2}$$

When $\lambda = 1$, there is no transformation. The model to be fitted is (1) with response $y(\lambda)$ and additive errors of constant variance.

However, when, for example, $\eta(t, \psi)$ is a mechanistic model based on chemical kinetics, the relationship between the response and the concentrations of the other reactants needs to be preserved after transformation. This is achieved by transformation of both sides of the model, as described in Chapter 4 of Carroll and Ruppert (1988). For fixed $\lambda \neq 0$, estimation of the parameters ψ after transformation does not depend on whether the response is $y(\lambda)$ or the non-normalized y^λ. Simplification of the model and the introduction of observational error on this transformed scale leads to the statistical model

$$y^\lambda = \{\eta(t, \psi)\}^\lambda + \epsilon. \tag{3}$$

Analysis of two sets of pharmacokinetic data in Atkinson (2003) yields values of λ of 0.35 and 0.1. In comparison, Horwitz's rule for analytical chemistry based on the analysis of thousands of data sets (Lischer 1999) yields a value around 0.14 for λ. Further references to Horwitz's rule can be found, for example, in Zorn, Gibbons, and Sonzogni (1997) and in Rocke and Lorenzato (1995), to which we return in §15. Clearly transformation of such data is often required.

3 Optimum Design for a Multivariate Response

The experiments used as examples consist of measuring the concentration of one or more chemicals after a reaction has been running for a time t. As in Box and Lucas (1959), the measurements are taken at a single time point for each run, so that successive observations are independent. The design problem is to choose the times of measurement and the proportion of observations at these times. For mathematical convenience only continuous designs are discussed, in which the design ξ is a continuous measure specifying both a set of k distinct points in a design region $\mathcal{T}$ and the proportions, w_u, of observations taken at these points

$$\xi = \left\{ \begin{array}{c} t_1, \ldots, t_k \\ w_1, \ldots, w_k \end{array} \right\}, \qquad 0 < w_u \leq 1, \qquad \sum_{u=1}^{k} w_u = 1.$$

The times t_u are the points of support of the design ξ and w_u the design weights. Practical designs for n observations are found by rounding nw_u to give the integer number of trials at t_u.

Locally D-optimum designs maximize the logarithm of the determinant of the information matrix, $\log|M(\xi,\psi)|$, in which a best guess ψ^o is taken for the values of the parameters. Designs in which ψ has a prior distribution with more than one point of support are found in §10

If the concentration of more than one chemical is measured, there will be a model for each expected response giving a matrix F_i of parameter sensitivities for the ith response, $i = 1, \ldots, m$,

$$F_i = \{f_{ij}(t_u,\psi)\} = \left\{\frac{\partial\eta_i(t_u,\psi)}{\partial\psi_j}\right\},$$

where $j = 1, ..., p$ denotes the parameters. If no transformation is needed the observations follow the model

$$y_{iu} = \eta_i(t_u,\psi) + \varepsilon_{iu},$$

with

$$E(\varepsilon_{iu}) = 0, \qquad E(\varepsilon_{iu}\varepsilon_{lv}) = \begin{cases} 0 & if \quad u \neq v \\ \sigma_{il} & if \quad u = v \end{cases}, \tag{4}$$

when the variance-covariance matrix of the m responses at each time point is

$$\Sigma = \{\sigma_{il}\}_{i,l=1,\ldots,m}.$$

As (4) shows, observations at time points t_u and t_v are independent, even if $t_u = t_v$. Design for other error structures is mentioned briefly in §15.

For normally distributed errors the information matrix (Draper and Hunter 1966) is given by

$$M(\xi,\psi) = \sum_{i=1}^{m}\sum_{l=1}^{m}\sigma^{il}F_i^T W F_l, \tag{5}$$

where $\Sigma^{-1} = \{\sigma^{il}\}_{i,l=1,\ldots,m}$ and $W = \mathrm{diag}\{w_1,\ldots,w_k\}$. If ψ_j is absent from the model for response i, column j of F_i will be zero. In our examples, when all chemical components are measured $\sum_{i=1}^{m}\eta_i(t_u,\psi) = 1$, but with normally distributed errors the constraint does not apply to the observations $\sum_{i=1}^{m} y_{iu}$.

When one response is measured, optimum designs have at least p points of support. For nonlinear models in a single factor, similar to those in this paper, often $k = p$. However, (5) shows that each response contributes a rank one matrix, weighted by σ^{ii}, to the information matrix at t_u. Provided the rows of F_i and F_l do not lie in the same subspace, the m responses at each point of support therefore contribute a rank m matrix. The value of k may then be less than p. It usually is in the examples that follow.

4 Parameter Sensitivities and Transforming Both Sides

The scalar parameter sensitivities for the model in §2 with $\lambda = 1$ will be written

$$f_j^1(t,\psi) = \frac{\partial \eta(t,\psi)}{\partial \psi_j}. \tag{6}$$

The parameter sensitivities in the transformed model (3) are then

$$f_j^\lambda(t,\psi) = \frac{\partial \{\eta(t,\psi)\}^\lambda}{\partial \psi_j} = \lambda \{\eta(t,\psi)\}^{\lambda-1} \frac{\partial \eta(t,\psi)}{\partial \psi_j} = \lambda \{\eta(t,\psi)\}^{\lambda-1} f_j^1(t,\psi). \tag{7}$$

For fixed λ, multiplication by λ in (7) does not change the optimum design, so the sensitivities have the easily calculated form

$$f_j^\lambda(t,\psi) = \{\eta(t,\psi)\}^{\lambda-1} f_j^1(t,\psi) = f_j^1(t,\psi)/\{\eta(t,\psi)\}^{1-\lambda}. \tag{8}$$

If $\lambda < 1$, the variance of the observations increases with the value of $\eta(t,\psi)$. Thus transformation of both sides for such values of λ will increase the relative value of the sensitivities for times where the response is small. We can expect that designs for $\lambda < 1$ will include observations at lower concentrations than those when no transformation is needed.

5 Design Efficiencies

The parameter sensitivities for the multivariate model of §3 when both sides of the model are transformed depend on both the response and on λ. For response i let the matrix of sensitivities now be $F_i(\lambda)$ with typical row $f_i^T(t,\psi,\lambda)$. The information matrix (5) is then written

$$M(\xi,\psi,\lambda) = \sum_{i=1}^{m} \sum_{l=1}^{m} \sigma^{il} F_i(\lambda)^T W F_l(\lambda). \tag{9}$$

Let the optimum design for a specified λ_0 be ξ_0^* and for some other λ be ξ_λ^*. When the value of the transformation parameter is λ the information matrix for the design ξ_0^* is $M(\xi_0^*,\psi,\lambda)$. Then the efficiency of the design ξ_0^* for some λ, compared to the most efficient design for that λ, is the ratio of determinants

$$\mathit{Eff}(\xi_0^*,\lambda) = \{|M(\xi_0^*,\psi,\lambda)|/|M(\xi_\lambda^*,\psi,\lambda)|\}^{1/p}, \tag{10}$$

where p is the number of parameters in the model.

6 Exponential Decay

The first example of the effect of assumptions about the mean-variance relationship on experimental design comes from the model for exponential decay in which the concentration of chemical A at time t is given by the nonlinear function

$$[A] = \eta_A(t, \theta) = e^{-\theta t} \qquad (\theta, t \geq 0), \tag{11}$$

if it is assumed that the initial concentration of A is 1.

Since no material is lost during the reaction, $[A] + [B] = 1$, although the observed concentrations will not obey this relationship. From (11) the concentration of B at time t is therefore

$$[B] = \eta_B(t, \theta) = 1 - e^{-\theta t} \qquad (\theta, t \geq 0).$$

If $[A]$ is measured in the absence of transformation the sensitivity is

$$f_A^1(t, \theta) = -t \exp(-\theta t), \tag{12}$$

whereas if $[B]$ is measured

$$f_B^1(t, \theta) = t \exp(-\theta t).$$

Now suppose that the model needs to be transformed to give constant variance. From (7) the parameter sensitivity for the power transformation λ when $[A]$ is measured is

$$f_A^\lambda(t, \theta) = \{\eta_A(t, \theta)\}^{\lambda - 1} f_A^1(t, \theta) = -t \exp(-\lambda \theta t). \tag{13}$$

Differentiation of (13) with respect to t shows that the optimum design is at a time of $1/(\lambda\theta)$. As λ decreases, the time for the optimum design increases.

The analysis when $[B]$ is measured is similar, but does not yield an explicit value for the optimum time. The sensitivity is now

$$f_B^\lambda(t, \theta) = \{\eta_B(t, \theta)\}^{\lambda - 1} f_B^1(t, \theta) = t \exp(-\theta t)\{1 - \exp(-\theta t)\}^{\lambda - 1}, \tag{14}$$

which is maximized by the optimum time. As $\lambda \to 0$, the optimum time does likewise.

Figure 1 is a plot for the D-optimum designs of the time at which the reading of the concentration of A or B or of both should be taken as a function of λ, when $\theta = 0.2$. The maximum time for which the experiment can be run is 20. The optimum time of measurement, in the absence of a transformation, is 5. When both $[A]$ and $[B]$ are measured, it is assumed that, after transformation, both components are measured with the same variance

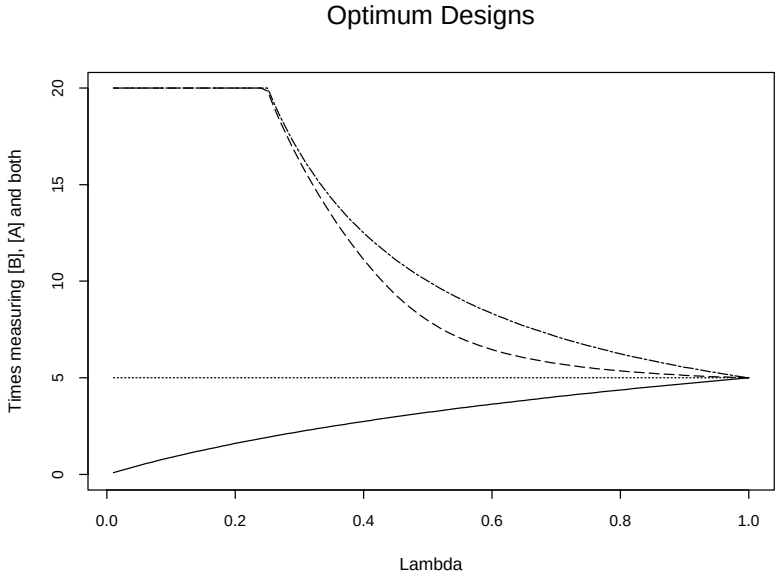

Figure 1: Exponential decay: time of optimum reading. Continuous line, measure $[B]$; dotted and dashed line, measure $[A]$; dashed line, both $[A]$ and $[B]$ measured

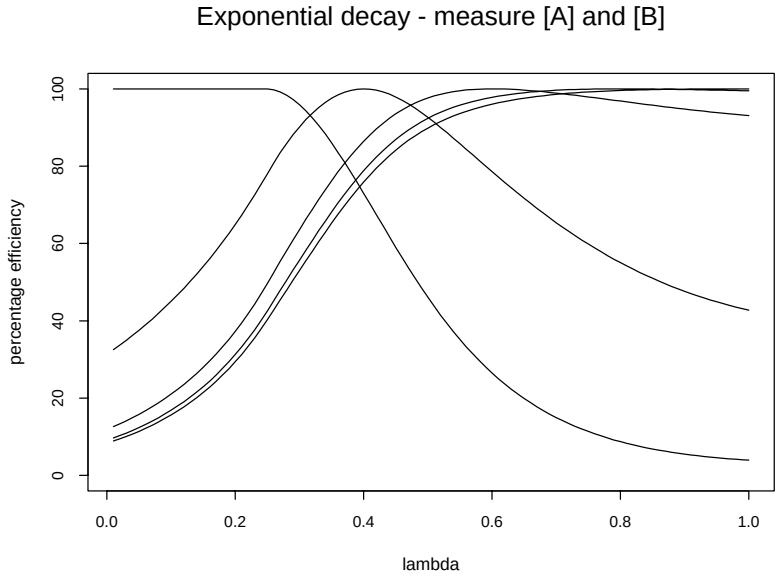

Figure 2: Exponential decay: efficiencies of optimum designs when both [A] and [B] are measured, for five values of λ_0: 0.2, 0.4, 0.6, 0.8 and 1.0. Each efficiency is 100% when $\lambda = \lambda_0$

and that the covariance is zero. The figure shows this design is similar to that when only $[A]$ is measured. For $\lambda \leq 0.25$ the single measurement is taken at the maximum time, 20. The concentration of B is small at the beginning of the experiment and that of A is small for large times. The figure shows that, as λ decreases and a stronger transformation is needed, the design points for single responses move to regions of lower concentration. Whether one or both components are measured, the optimum design requires measurement at a single time.

The plots of support points in Figure 1 show that, as $\lambda \to 0$, the support points of the designs tend to either 0 or to the maximum available time. For values of $\lambda < 0$ the effect of transformation is, where possible, to make the designs even more extreme. In this example this is not possible, the support points remaining at the minimum or maximum time available. Negative values of λ are accordingly not considered in the remaining examples. The extreme nature of these designs is considered in the discussion of Rocke and Lorenzato (1995) in §15.

Figure 2 shows the efficiencies defined in §5 of the D-optimum designs for five values of λ_0: 0.2, 0.4, 0.6, 0.8 and 1 when both $[A]$ and $[B]$ are measured. These are plotted over a range of values of λ between 0.01 and one Each efficiency is, of course, 100% when $\lambda = \lambda_0$. What is particularly noticeable is that all designs except that for $\lambda_0 = 0.2$ are similarly inefficient for low values of λ. The design for $\lambda_0 = 0.2$ has an efficiency of 100% for low values of λ because of the range of values for which the optimum design consists of a single measurement at $t = 20$. It is, however, inefficient for high values of λ. Thus no one of these designs will be efficient for estimating θ if there is little information about the true value of λ.

7 Bayesian Optimum Designs

To find optimum designs which have good properties over a range of values of λ we can maximise a weighted product of the efficiencies for a vector of $n(q)$ values λ_0 with a weight q_i for the efficiency for λ_i. From (10) this weighted product is

$$G(\xi, \psi, \lambda_0) = \prod_{i=1}^{n(q)} \{|M(\xi, \psi, \lambda_i)|/|M(\xi^*_{\lambda i}, \psi, \lambda_i)|\}^{q_i/p}, \tag{15}$$

where $\xi^*_{\lambda i}$ is the optimum design for λ_i. The weighted product of efficiencies (15) is to be maximised by choice of ξ, which is equivalent to maximising its logarithm. Since the terms $|M(\xi^*_{\lambda i}, \psi, \lambda_i)|$ are not functions of ξ the design

criterion reduces to maximization of

$$\Phi(\xi, \psi, \lambda_0) = \sum_{i=1}^{n(q)} q_i \log |M(\xi, \psi, \lambda_i)|. \tag{16}$$

We thus obtain a Bayesian design in which the weights q_i form a discrete prior for the values λ_0. This particular criterion is also a form of generalized D-optimality, for which an Equivalence Theorem can be formulated. Atkinson and Cox (1974) and Läuter (1974) give examples of the use of this Equivalence Theorem for linear models.

For a single-response linear model the Equivalence Theorem (Kiefer and Wolfowitz 1960) states the equivalence of D-optimum designs to those minimising the maximum of the variance $d(t, \xi)$ of the predicted response over the design region. For multivariate designs maximising $\log |M(\xi, \psi, \lambda)|$ the variance, from (5), is

$$d(t, \xi, \psi, \lambda) = \sum_{i=1}^{m} \sum_{l=1}^{m} \sigma^{il} f_i^T(t, \psi, \lambda) M^{-1}(\xi, \psi, \lambda) f_l(t, \psi, \lambda), \tag{17}$$

where $f_i^T(t, \psi, \lambda)$ is the vector of parameter sensitivities introduced in §5. For maximisation of the Bayesian criterion (16) the required function is

$$d(t, \xi, \psi, \lambda_0) = \sum_{i=1}^{n(q)} q_i d(t, \xi, \psi, \lambda_i). \tag{18}$$

Let the design maximising (16) be $\xi^*_{\lambda 0}$. Then the Equivalence Theorem states that $\xi^*_{\lambda 0}$ also minimises the maximum over the design region of $d(t, \xi, \psi, \lambda_0)$. Furthermore, the maximum value of $d(t, \xi^*_{\lambda 0}, \psi, \lambda_0)$ is p and all points of support of $\xi^*_{\lambda 0}$ are at a maximum. This equivalence provides algorithms for the construction of designs and, here, for checking that a design found by numerical optimization is indeed optimum. Discussion of Figure 9 in §8 shows how the plot of $d(t, \xi^*_{\lambda 0}, \psi, \lambda_0)$ can provide information about the sensitivity of the optimum to small changes in $\xi^*_{\lambda 0}$.

This derivation makes it clear that the motivation is not primarily to find designs which reflect prior knowledge about λ but rather to find designs which are robust to a wide range of values of λ. For illustration we take $n(q) = 5$; a different number of prior values or integration of a continuous prior would not give substantially different robust designs.

To see the effect of the prior distribution on the design and efficiencies, optimum designs were found for the exponential decay model for three priors for λ. The resulting designs are in Table 1. All designs are for a five-point prior with weights of 0.2 at each point. The first design, for a uniform prior

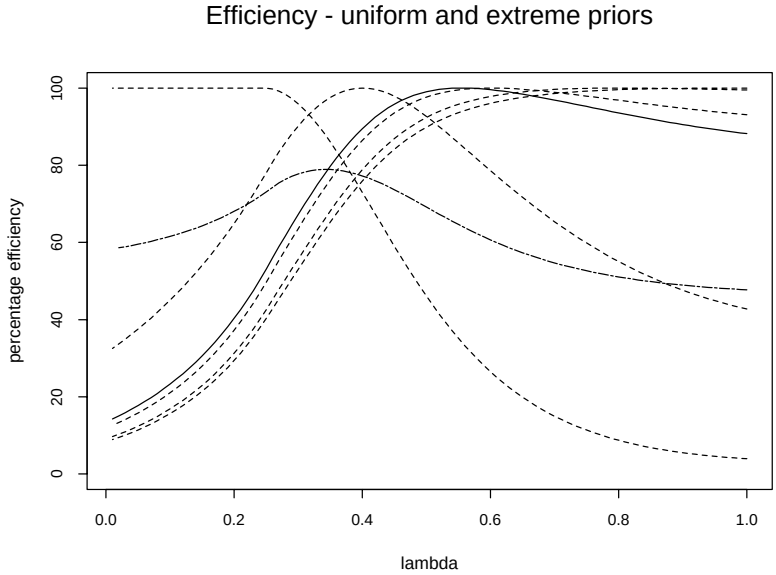

Figure 3: Exponential decay when both [A] and [B] are measured. Continuous line, efficiency of Bayesian optimum design for $\lambda_0 = (0.2, 0.4, 0.6, 0.8, 1.0)^T$; dotted and dashed line, efficiency of Bayesian optimum design for $\lambda_0 = (0.05, 0.1, 0.15, 0.95, 1.0)^T$; dashed lines, efficiencies of the optimum designs for individual members of λ_0

on the five equally spaced points 0.2, 0.4, 0.6, 0.8 and 1.0, consists of a single measurement at $t = 6.989$. The optimum design for $\lambda = 0.56$ takes one measurement at $t = 6.925$ and the continuous line in Figure 3 shows that the trace of the efficiency of this Bayesian design is close to that of the design for $\lambda = 0.6$. In particular, it drops to below 20% as $\lambda \to 0$.

To improve efficiency for low values of λ the priors for the other two designs in Table 1 put more weight on low values of λ. The second prior is supported on points 0.1, 0.2, 0.3, 0.9 and 1. The design has two support points, with weight 0.382 at $t = 20$, the optimum design point for low values of λ. The efficiency for this non-uniformly spaced prior is increased for lower values of λ at the cost of lower efficiency for values of λ near one.

The third design in Table 1 is for an extreme prior on the points 0.05, 0.1, 0.15., 0.95 and 1.0. As the plot in Figure 3 shows, this has efficiencies between 50 and 60% for values of λ close to zero and close to one with a higher efficiency in the centre of the range. In the absence of any firm information about possible values of λ, this design with just over half its weight at $t = 20$ and the remainder near the value of $t = 5$, optimum for no transformation, provides a good design for estimation of θ over a wide range

Table 1: Exponential decay when both [A] and [B] are measured. Bayesian optimum designs for three equiweighted five-point priors on λ

Prior λ_0	t, w	Design	
0.2, 0.4, 0.6, 0.8, 1.0	t_i	6.989	
	w_i	1.000	
0.1, 0.2, 0.3, 0.9, 1.0	t_i	6.490	20.000
	w_i	0.618	0.382
0.05, 0.1, 0.15, 0.95, 1.0	t_i	5.641	20.000
	w_i	0.462	0.538

of possible transformations.

8 Reversible Reaction

The behaviour of the designs in Figure 1 for experiments in which only one component was measured changed in a predictable way: as λ decreased, the design points moved towards regions of lower concentration. The multiresponse experiment also showed this general behaviour. This section considers a reversible experiment in which the behaviour is not so easily characterized.

The model is for two consecutive first-order reactions with the second reaction reversible which can be written

$$A \overset{\theta_1}{\rightarrow} B \underset{\theta_3}{\overset{\theta_2}{\rightleftharpoons}} C, \tag{19}$$

where θ_3 the rate of the reverse reaction and all $\theta_j > 0$. The kinetic differential equations for $[A], [B]$ and $[C]$ are

$$\begin{aligned} \frac{d[A]}{dt} &= -\theta_1[A] \\ \frac{d[B]}{dt} &= \theta_1[A] - \theta_2[B] + \theta_3[C] \\ \frac{d[C]}{dt} &= \theta_2[B] - \theta_3[C]. \end{aligned} \tag{20}$$

If the initial concentration of A is one and those of B and C are zero, $\eta_A(t, \theta)$ again follows exponential decay (11) with $\theta = \theta_1$. The other concen-

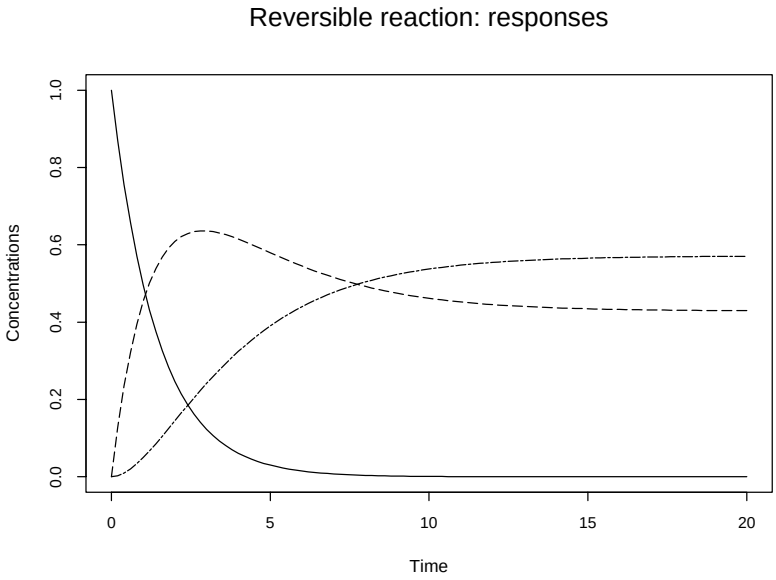

Figure 4: Reversible reaction: concentrations of reactants over time. Reading upwards for large t: [A], [B] and [C]

trations are given by

$$\begin{aligned} \eta_B(t,\theta) &= \frac{\theta_3}{\theta_2+\theta_3}\left\{1-e^{-(\theta_2+\theta_3)t}\right\} - \frac{\theta_1-\theta_3}{\theta_1-\theta_2-\theta_3}\left\{e^{-\theta_1 t}-e^{-(\theta_2+\theta_3)t}\right\} \\ \eta_C(t,\theta) &= 1-\eta_A(t,\theta)-\eta_B(t,\theta). \end{aligned} \tag{21}$$

An important difference from the earlier example is that $[B] \to \theta_3/(\theta_2+\theta_3)$ as $t \to \infty$, so that $[B]$ and $[C]$ have informative values for large t.

Figure 4 shows the responses as a function of time when $\theta_1 = 0.7$, $\theta_2 = 0.2$ and $\theta_3 = 0.15$: the asymptotic value of $[B]$ is therefore $3/7$. Since there are three parameters, single-response designs will have at least three points of support, the third being at the maximum value of t. We can also expect, since the value of $[B]$ no longer decreases to zero with time, that there will be only a slight effect of λ on any design points in the middle of the region when only $[B]$ is measured.

The parameter sensitivities are found by differentiation of (21) and, as before, use of (8). The resulting designs when either $[B]$ or $[C]$ are measured are given in Figure 5. Both are three-point designs with weights one third at each time and with one support point at the maximum time, here taken to be 20. The two upper design points of the design for $[B]$ in the left-hand panel are indeed virtually unaffected by λ, t_2 increasing very slightly from 4.96 to 5.30. The lowest time point however decreases from 1.17 to

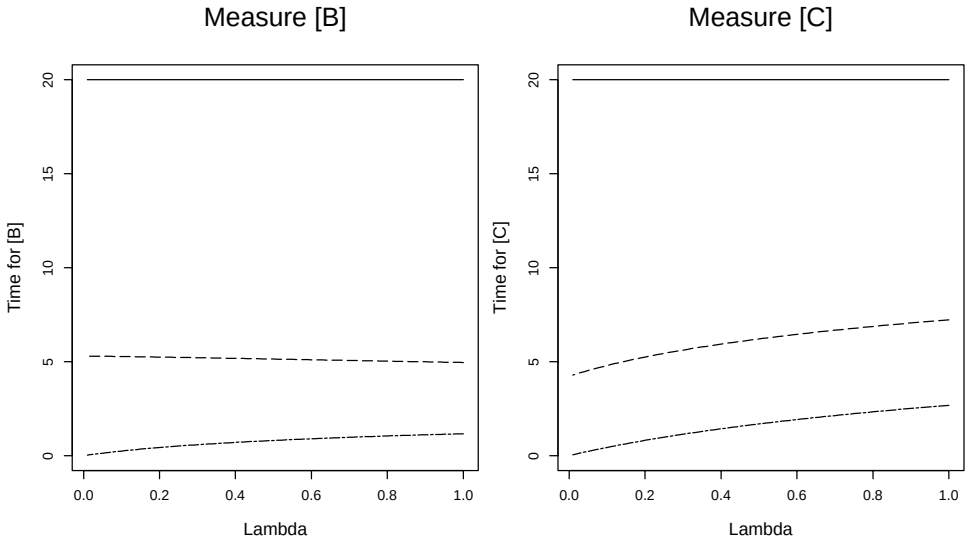

Figure 5: Reversible reaction: times of measurement for optimum designs when one response is measured. Left-hand panel, measure [B]; right-hand panel, measure [C]

zero as λ decreases, behaviour which is similar to that shown in Figure 1 for exponential decay when only $[B]$ is measured. The design when just $[C]$ is measured has one observation at $t = 20$. As λ decreases, the other two support points move towards regions of lower concentration, that is lower values of t.

The design however changes in a less predictable way when more than one response is measured. The left-hand panel of Figure 6 shows the design points when $[A]$, $[B]$ and $[C]$ are all measured. The three measurements are assumed independent with the same variance, so that Σ is again a multiple of the identity matrix. For $\lambda = 1$, there is a three-point design, similar to that when either response is measured separately. The weights, in the right-hand panel of Figure 6, are very close to one third when $\lambda = 1$. However, by the time λ has decreased to 0.89, a two-point design is optimal, with weight $2/3$ on the lower of the two design points. Two point designs with these weights are optimum until $\lambda = 0.25$. The design changes at lower values of λ, with the upper support point of the design, for a while, less than 20. For the smallest values of λ we again obtain a two-point design, with the upper time point again at 20. But now the maximum weight is on the upper design point. In Figure 6, the same type of line is used for each time and its associated weight.

The efficiencies for some of these designs as λ varies, defined in §5, are

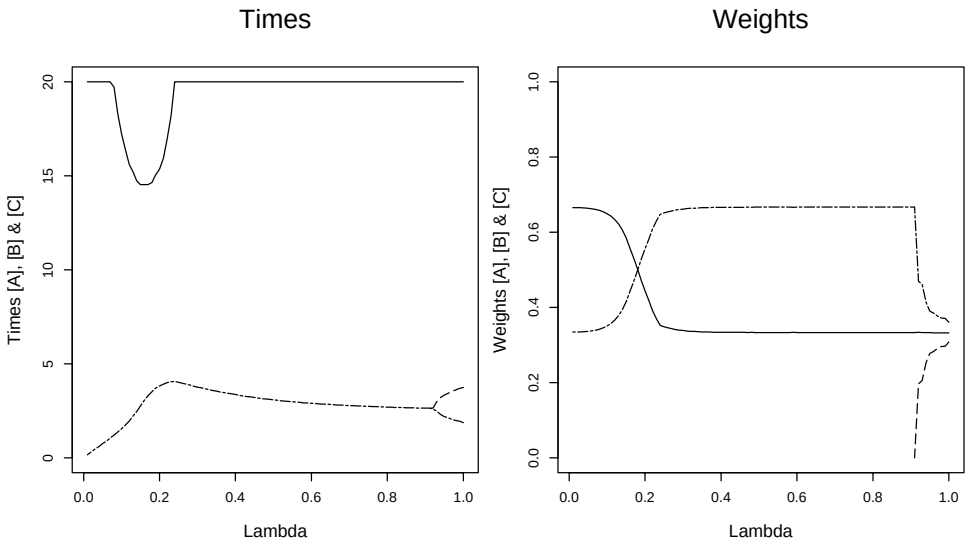

Figure 6: Reversible reaction: optimum designs when [A], [B] and [C] are all measured. Left-hand panel, times; right-hand panel, design weights

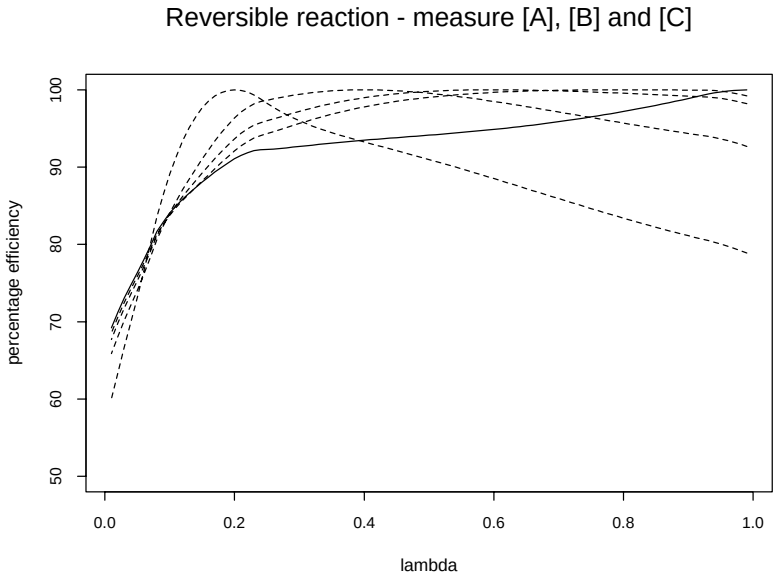

Figure 7: Reversible reaction: efficiencies of optimum designs when [A], [B] and [C] are measured, for five values of λ_0: 0.2, 0.4, 0.6, 0.8 and (continuous line) 1.0. Each efficiency is 100% when $\lambda = \lambda_0$

Table 2: Reversible reaction when [A], [B] and [C] are measured. Bayesian optimum designs for three five-point priors on λ

Prior λ_0	t, w	Design		
0.2, 0.4, 0.6, 0.8, 1.0	t_i	2.902	20.000	
	w_i	0.656	0.344	
0.1, 0.2, 0.3, 0.9, 1.0	t_i	2.882	19.085	
	w_i	0.601	0.399	
0.05, 0.1, 0.15, 0.95, 1.0	t_i	2.037	3.493	19.584
	w_i	0.297	0.216	0.487

shown in Figure 7 for the five values of λ_0: 0.2, 0.4, 0.6, 0.8 and 1. By comparison with the efficiencies of designs for exponential decay plotted in Figure 2, these efficiencies are all relatively high, although all are below 70% for small λ. The efficiency for $\lambda_0 = 1$ is shown in Figure 7 by a continuous line. This is the only three-point design amongst those considered, and is the most efficient not only for $\lambda = 1$, but also for the lowest values of λ. It is therefore likely that any Bayesian design that has good properties over a wide range of values for λ will also have three design points.

Table 2 gives Bayesian designs for the reversible reaction for the same three priors for λ as were used in Table 1. The uniform prior on 0.2, 0.4, 0.6, 0.8 and 1.0 leads to a two-point design with weight almost one third at $t = 20$. It is similar to designs for λ around 0.6, the average value of λ for this prior. The second design, for a non-uniform prior gives a similar design, although the maximum of the two times of measurement, at 19.085, is just within the boundary of the design region.

The third, and most extreme, prior in Table 2 does give a three-point design, with the third time again just within the boundary of the region. The design has more weight, 0.487, at this high value than does the design for $\lambda_0 = 1$, which puts weight 0.332 at $t = 20$. The efficiencies for this Bayesian design are plotted in Figure 8. Comparison with Figure 7 shows that the Bayesian design has an efficiency more than 10% higher than any other design for λ close to zero. Over most of the range of λ the efficiency is around 90%.

Finally, Figure 9 is a plot of the variance $d(t, \xi^*_{\lambda 0}, \psi, \lambda_0)$ for the extreme prior. The three points of support of the design are shown by dots. At these points the variance is indeed equal to three, the number of parameters in the model, so that an optimum design has been found. The flatness of the curve for high values of t shows that the design is not sensitive to the exact value

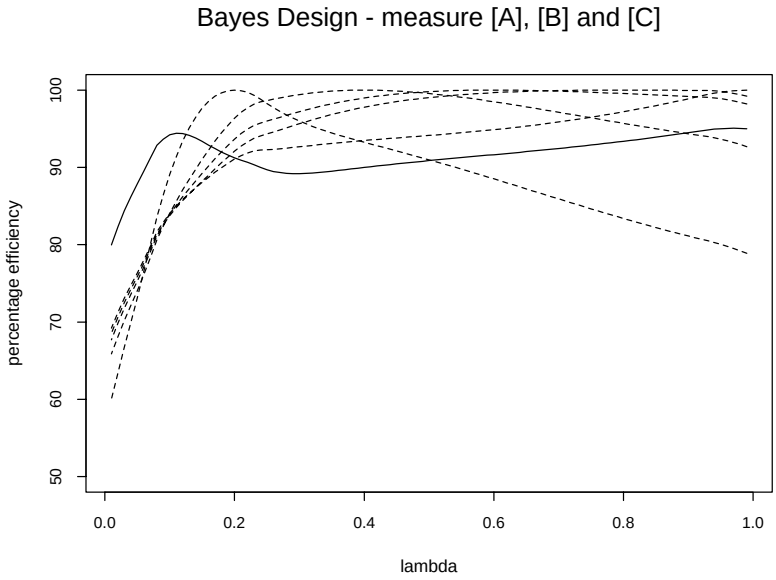

Figure 8: Reversible reaction: efficiencies of optimum designs when [A], [B] and [C] are measured. Continuous line, efficiency of Bayesian optimum design for $\lambda_0 = (0.05, 0.1, 0.15, 0.95, 1.0)^T$; dashed lines, efficiencies of the optimum designs for individual values 0,2, 0.4, 0.6, 0.8, 1.0 for λ_0

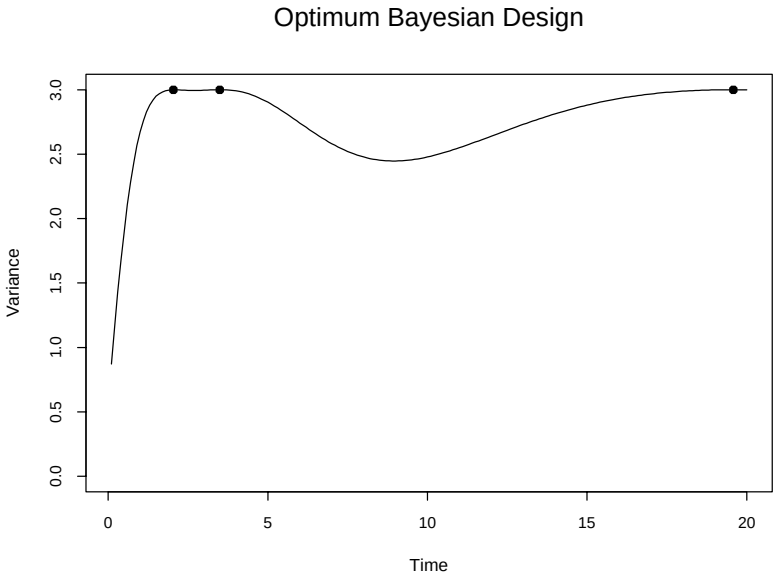

Figure 9: Reversible reaction when [A], [B] and [C] are measured. Variance $d(t, \xi^*_{\lambda 0}, \psi, \lambda_0)$ for the extreme prior $\lambda_0 = (0.05, 0.1, 0.15, 0.95, 1.0)^T$

of the upper time point. However, observations at very low times are not informative.

9 An Equispaced Design

The Bayesian procedure of the preceding section provides a design that is robust to the value of λ. We now consider robustness to other departures from the conditions under which the locally-optimum designs were found.

One frequent form of design has points of support approximately equispaced in log dose or log time. It is often claimed in teh design literature that such designs, while not optimum for any specific formulation, do provide good behaviour for a variety of criteria. For example, Downing et al. (2001) describe an experiment with ten levels of concentration from a two-fold dilution series, with a highest concentration of 500 ng/ml. Similarly, in the experiment on measurements leading to compartmental model for the concentration of aminophylline in horses quoted by Atkinson et al. (1993), measurements are taken at 18 different levels found by approximate doubling of times of measurement, with some rounding to give experimentally simpler conditions. The equispaced design in this paper consists of measurements at ten time points, the maximum being 20, with lower values found by division by $\sqrt{2}$, so that the minimum value is 0.8839.

As Figure 6 shows, except for very low values of λ, the two or three support points of the D-optimum designs lie within the range 0.8839 to 20. However, the dispersed nature of a design with ten support points means that the efficiency of the design will not be especially high for any λ. Figure 10 shows the efficiency of the equispaced design together with the efficiencies from the designs for five values of λ_0, as in Figures 7 and 8. The efficiency of this design is around or slightly above 80% for λ above 0.2. For smaller values of λ the efficiency decreases rapidly, as the design does not have any support at the very low values of t that are optimum as λ approaches zero. Comparison with Figure 8 shows that the Bayesian optimum design has higher efficiency for all λ than does the equispaced design. However, the equispaced design may be reasonably efficient for some other departure and we continue to investigate its properties.

10 Imprecise Parameter Values

In addition to λ the parameters ψ of the model (20) include the rate constants θ and, implicity, the orders of reaction, assumed 1 in (20). Instead of

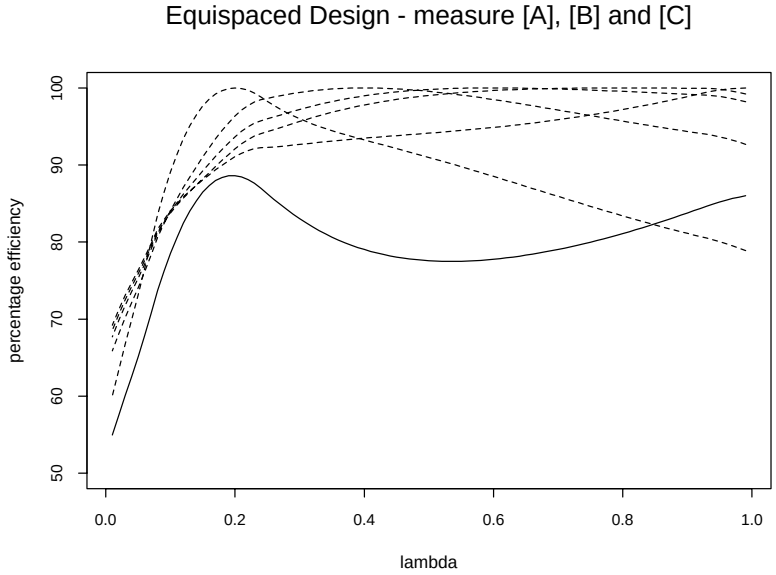

Figure 10: Reversible reaction when [A], [B] and [C] are measured: continuous line, efficiency of ten-point equispaced design; dashed lines, efficiencies of the optimum designs for individual values 0,2, 0.4, 0.6, 0.8, 1.0 for λ_0

locally optimum designs for known parameter value ψ^0, we explore Bayesian optimum designs when there is a prior distribution of parameter values.

When the prior distribution is $\pi(\psi)$ the Bayesian design criterion (16) is replaced by maximization of

$$\Phi(\xi, \lambda_0) = \mathrm{E}_\psi \log |M(\xi, \psi, \lambda_0)| = \int \log |M(\xi, \psi, \lambda_0)| \pi(\psi) d\psi. \qquad (22)$$

The integral in (22) can be evaluated numerically, as was done for a different model in Atkinson et al. (1993). An alternative (Atkinson et al. 1995) is to sample parameter values from the prior and then to replace the integral in (22) with a summation. If only values of θ are varying (22) becomes maximization of

$$\Phi(\xi, \lambda_0) = \sum_{i=1}^{n(\theta)} \log |M(\xi, \theta_i, \lambda_0|, \qquad (23)$$

where $n(\theta)$ is the number of values sampled from $\pi(\psi)$. Formally, this is similar to the summation in (16) and an equivalence theorem applies, analogous to that in (18).

As a numerical example the θ_j in (21) are taken to have log normal distributions generated by the exponentiation of normal distributions with mean $\log \theta_j$ and standard deviation $\log 2/2$. Thus, for each θ_j, just over 95%

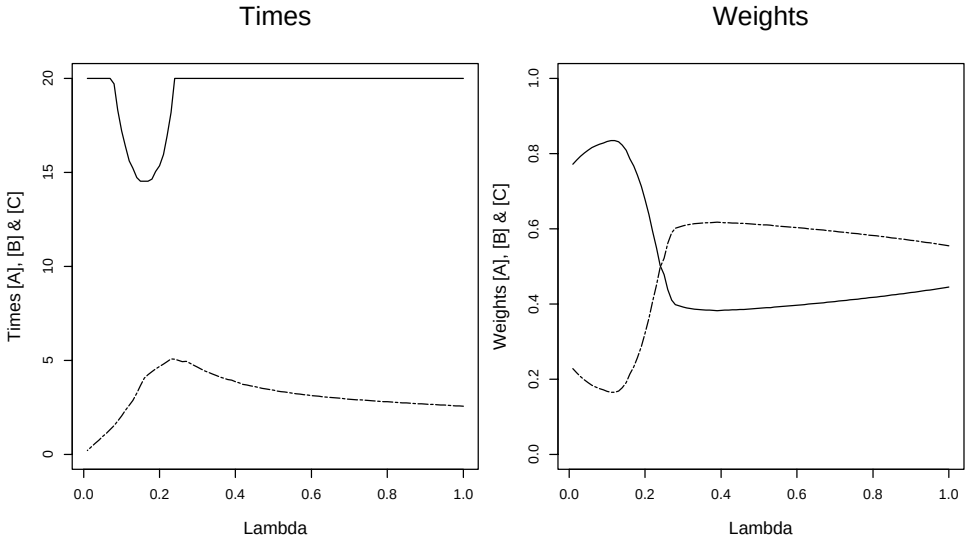

Figure 11: Reversible reaction: Bayesian optimum designs with $n(\theta) = 50$ when [A], [B] and [C] are all measured. Left-hand panel, times; right-hand panel, design weights

of the prior distribution lies between $\theta_j/2$ and $2\theta_j$. Some values of θ however have to be excluded. As $\theta_1 - (\theta_2 + \theta_3) \to 0$, the expression for $\eta_B(t, \theta)$ (21) becomes indeterminate and needs to be replaced by

$$\eta_B(t, \theta) = \frac{\theta_3}{\theta_2 + \theta_3} \left\{1 - e^{-(\theta_2+\theta_3)t}\right\} - (\theta_1 - \theta_3)te^{-\theta_1 t}, \tag{24}$$

with a consequent effect on the parameter sensitivities. Such complications were avoided by rejecting any set of simulated values for which $1.1(\theta_2 + \theta_3) > \theta_1$. The value of $n(\theta)$ was taken as 50. In the generation of 50 satisfactory sets of parameter values, only four sets had to be rejected.

The Bayesian D-optimum design based on the sample of 50 sets of parameters is given in Figure 11. Surprisingly all optimum designs have only two points of support. Otherwise, the support points of the designs are close to those for the locally optimum design plotted in Figure 6. The design weights are however rather different. Instead of weights of around 1/3 and 2/3 switching at around $\lambda = 0.2$, the weights are closer to one half for λ near one and more extreme for small values of λ.

It is not immediately obvious how to compare the Bayesian and locally optimum designs. By definition, the integration of the Bayesian design over the parameter space Θ gives a higher value of the criterion (22) than would integration of the locally optimum design. However, neither design is the

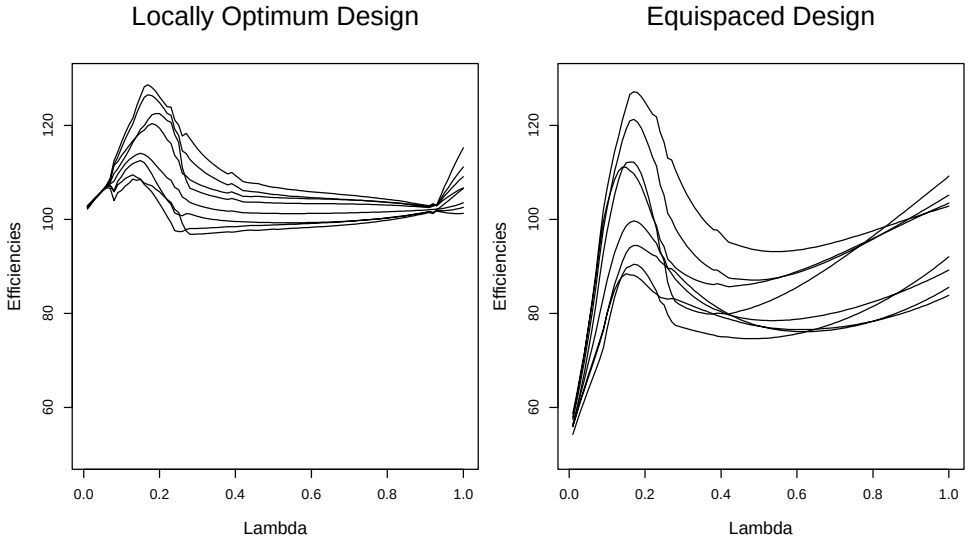

Figure 12: Reversible reaction: Efficiency of designs relative to the Bayesian D-optimum designs with $n(\theta) = 50$ for eight parameter values. Left-hand panel, locally optimum design; right-hand panel, equispaced design

most efficient for any specific element $\theta \in \Theta$, apart from the local design at θ^0. Accordingly designs were compared by calculating efficiency relative to the Bayesian design. Such "efficiencies" may be less than, or greater than, 100%.

The efficiencies were calculated for eight parameter values. One was the set of values used when finding the locally optimum designs of §8, that is $\theta_1^0 = 0.7$, $\theta_2^0 = 0.2$ and $\theta_3^0 = 0.15$. The others, on the log scale, were found by adding $\pm \log 2/2$ to these values. For one of these eight factorial values $\theta_1 = \theta_2 + \theta_3$, so that this set of values was excluded, making eight parameter sets in all.

The efficiencies of the locally optimum design are shown in the left-hand panel of Figure 12. These efficiences are surprisingly high. For λ above about 0.3 they range upwards from 95%. For lower values of λ, where the Bayesian design is less balanced than the locally optimum design, the efficiences are all above 100%. The efficiencies for the equispaced design in the right-hand panel of the figure are lower, showing appreciably more variation with θ than do the locally optimum designs. For λ above 0.3 the efficiencies are around 80% or more, with a peak of higher values near $\lambda = 0.2$. For very small values of λ the design has small efficiency, as would be expected from the discussion of Figure 10. An interpretation of these results is that the locally optimum design is efficient over the central region of parameter space. The Bayesian

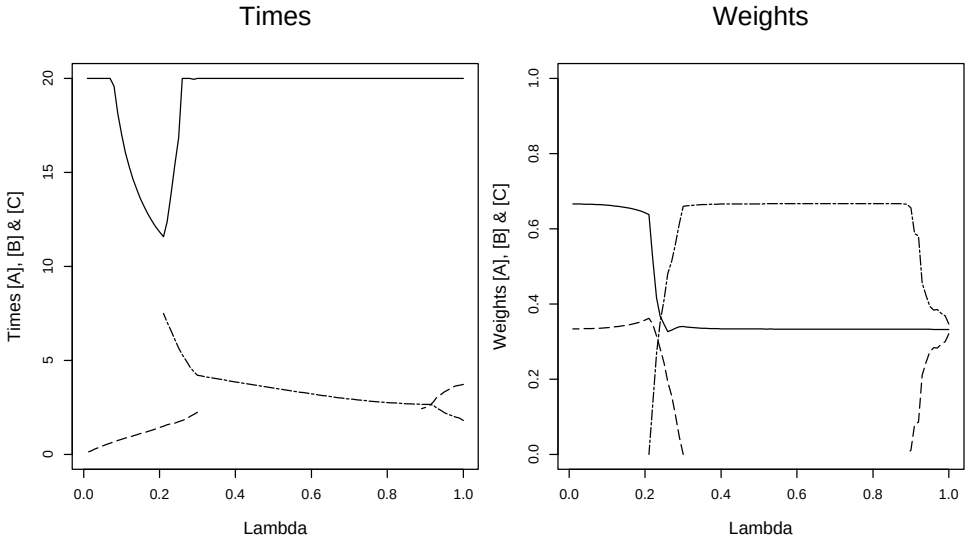

Figure 13: Reversible reaction: Locally optimum designs for correlated responses. Left-hand panel, times; right-hand panel, design weights

design is less efficient there because it has been constructed to provide some information over a wider distribution of parameter values. Some of these values have caused a relative distortion of the design for low values of λ.

11 Correlated Responses

The design criterion given in §3 allowed for covariances σ_{il} between responses i and l. Here we find optimum designs when such covariances exist and compare the designs with the locally optimum design of §8 found in the absence of correlation.

Let the variances of the individual responses be 1 and the covariances ρ. Then the 3×3 information matrix Σ^{-1} used in the design criterion (5) has diagonal elements

$$\sigma^{ii} \;=\; (1+\rho)/d(\rho) \qquad \text{and} \qquad \sigma^{il} \;=\; -\rho/d(\rho), \tag{25}$$

where $d(\rho) = 1 + \rho + 2\rho^2$.

The resulting optimum design for $\rho = -0.4$ is shown in Figure 13. It seems as if the effect of correlation between the observations is slight, as far as the design is concerned. Figure 13 is similar to Figure 6 except for a region around $\lambda = 0.25$ where there are some three-point designs.

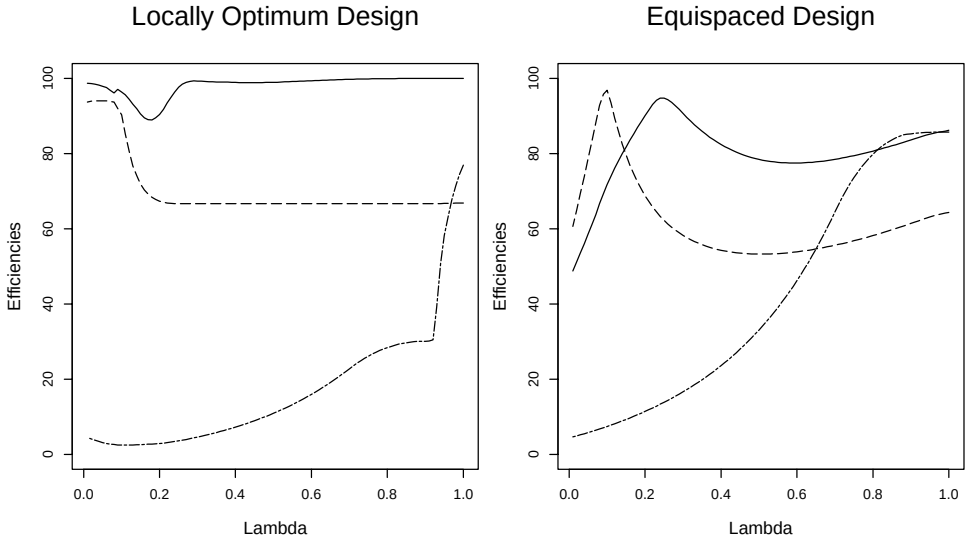

Figure 14: Reversible reaction: Efficiency of designs for departures from assumptions. Left-hand panel, locally optimum designs; right-hand panel, equispaced design. Continuous line, correlated responses; dashed line, two-parameter model; dotted and dashed line, model with four parameters

The efficiency of the locally D-optimum design and of the equispaced design are shown by continuous lines in Figure 14. In line with the similarity between the locally optimum design for independent observations and the design of Figure 13, the efficiency of the locally optimum design is virtually 100%, except for values of λ around 0.2 where the three-point design is optimum. The efficiency of the equispaced design in the right-hand panel of the figure is similar to that shown in Figure 10 - since the locally optimum design mostly has an efficiency near 100%, the efficiency of the equispaced design relative to the correlated design is close to its efficiency relative to the locally optimum design.

12 A Two-Parameter Model

So far in the analysis of design robustness, it has been assumed that the correct form of the model is known, even if, as in §10, the parameter values are uncertain or the errors are not necessarily independent. In this section and the next we explore how uncertainties about the mechanistic model affect the optimum design.

The terms in the expression for $\eta_B(t, \theta)$ in (21) contain fixed terms and

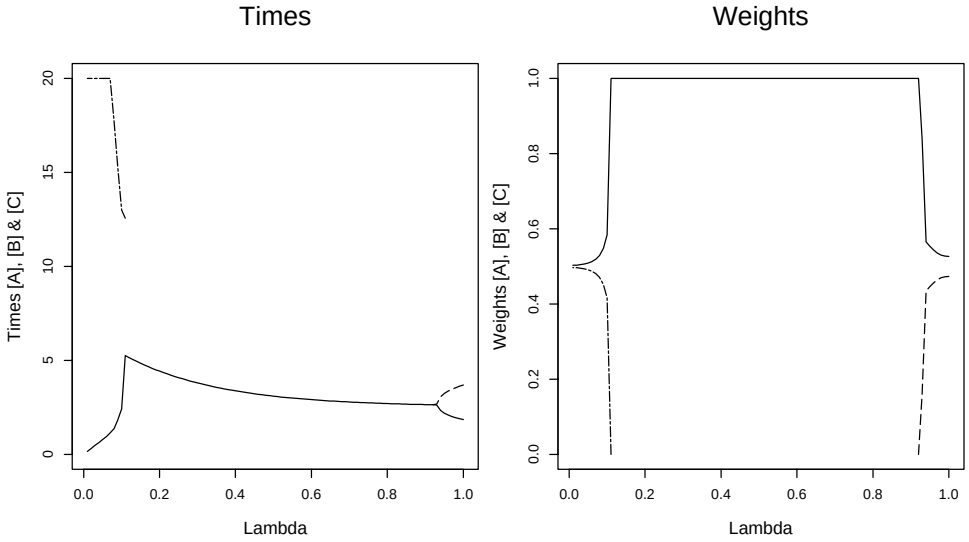

Figure 15: Two-parameter reversible reaction (γ known): Locally optimum designs when three responses are measured. Left-hand panel, times; right-hand panel, design weights

transient terms. As $t \to \infty$, the concentration of B tends to

$$\eta_B(\infty, \theta) = \frac{\theta_3}{\theta_2 + \theta_3}. \tag{26}$$

The concentration of A will become zero with increasing time. The equilibrium concentration of C is therefore 1 - $\eta_B(\infty, \theta)$. The ratio of such equilibrium concentrations can be known very precisely, even when the dynamics of the system are not so well estimated. Accordingly we take $\theta_3 = \gamma\theta_2$ when $\eta_B(\infty, \theta) = \gamma/(1+\gamma)$.

With γ known there are only two unknown parameters in the model, that is $p = 2$. The locally D-optimum design for this reduced model as λ varies and all three responses are measured is shown in Figure 15. For most values of λ the design has one point of support, except for low and high values of λ when two are required.

The design for the reduced model in Figure 15 has a strong resemblance to the design for the three-parameter model in Figure 6. The times at which the lower measurement is taken follow a very similar trajectory in the two figures as λ varies. However, for the two parameter model, observations at a high value of time are only taken for $\lambda \leq 0.1$. Below this value the two designs are similar, with the lower time of measurement decreasing and the upper increasing to $t = 20$.

The similarity in the designs for the two- and three-parameter models is reflected in the left-hand panel of Figure 14, where the efficiency for the two-parameter model is plotted using a dashed line. For most values of λ this value is 2/3, rising to over 90% for small values of λ. The efficiency for the ten-point design is lower for the larger values of λ, around 50 to 60%, but is surprisingly high, over 90%, around $\lambda = 0.1$.

13 A Four-Parameter Model

Figure 14 shows that optimum designs for a too large model can be surprisingly efficient if a smaller model is fitted. This section considers the reverse situation where the true model has four parameters. Of course, for linear models when a single response is measured, the D-optimum design for a quadratic, with three design points, would be totally non-informative about the four parameters of a cubic model. As we shall see, with three responses measured, the situation in our example is not quite so bad.

Seemingly simple chemical reactions may have a more complicated structure. For example, the first reaction in (19) may consist not of A going directly to B, but through some intermediate species D. Then the reaction scheme becomes

$$A \xrightarrow{\theta_4} D \xrightarrow{\theta_1} B \underset{\theta_3}{\overset{\theta_2}{\rightleftharpoons}} C. \tag{27}$$

If $\theta_1 >> \theta_4$, the overall rate of reaction will be determined by θ_4 (the rate determining step) and the reaction reduces to (19), as it does if, conversely, $\theta_4 >> \theta_1$.

Solutions of the first-order differential equations for the mechanism given in (19) are in the Appendix. These equations are similar in form, although more complicated in detail, than those for the concentrations given in (21) for the three-parameter model. We assume that, as before, $[A]$, $[B]$ and $[C]$ are measured, but not $[D]$. Differentiation of the three measured responses with respect to the four parameters gives, as before, a set of parameter sensitivities that can be used in calculating optimum designs. For numerical work the prior values of θ_2 and θ_3 remain as before at 0.2 and 0.15, but now $\theta_1 = 0.8$ and $\theta_4 = 1.4$

As is to be expected, the designs are more complicated than any we have so far seen. Whatever the value of λ, one quarter of the design weight is at $t = 20$. For λ between 0.87 and 0.64 the design has four points of support: otherwise there are three. As λ decreases so does the minimum time of measurement. Here the minimum value has been taken as 0.1, which

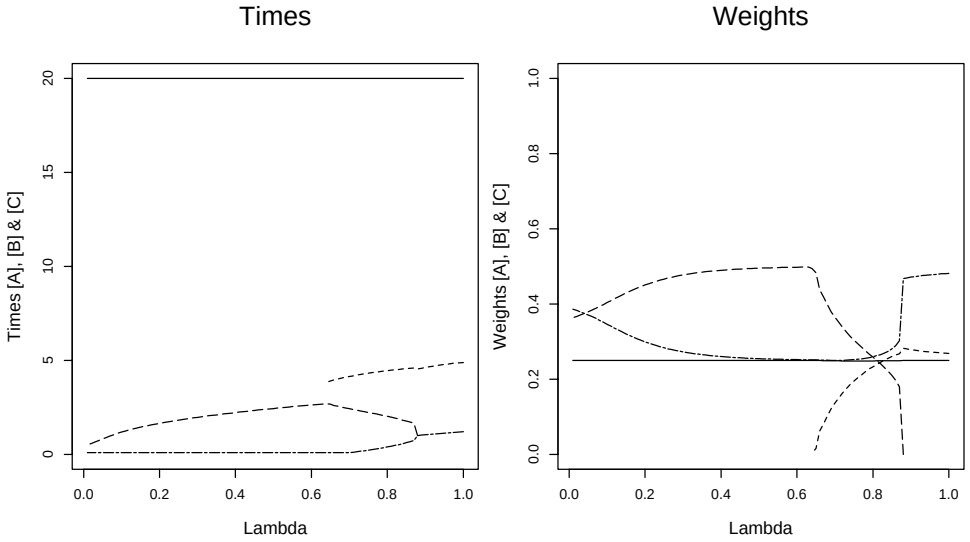

Figure 16: Four-parameter reversible reaction, including D: Locally optimum designs when three responses are measured. Left-hand panel, times; right-hand panel, design weights

applies at and below $\lambda = 0.7$. As λ decreases below this value, the second time of measurement also becomes smaller. One effect of transformations, as has been seen, is to place design points at regions of increasingly low concentration as λ decreases. Here information at low times is necessary to enable estimation of θ_1 and θ_4, which both regulate the early stages of the reaction.

The efficiencies of the locally D-optimum design for the three parameter model for estimation of the four parameters are plotted in the left-hand panel of Figure 14, using a dotted and dashed line. For $\lambda = 1$ the efficiency is 76%, a surprisingly high value. However the efficiency decreases rapidly as λ decreases, being only 30% when $\lambda = 0.92$. Thereafter the decrease is more gradual, but, even so, the minimum value is a useless 2.5%. The ten-point design in the right-hand panel has higher efficiencies throughout, which, since the design does not change with λ, decrease in a smooth way. However, for small values of λ, the efficiency is unacceptably low, due to the absence of measurements at small values of t.

14 Orders of Reaction

It was assumed in (20) that all reactions for the three-parameter reversible model were first order. If the orders are unknown parameters ν_j the differential equations become

$$\begin{aligned} \frac{d[A]}{dt} &= -\theta_1[A]^{\nu_1} \\ \frac{d[B]}{dt} &= \theta_1[A]^{\nu_1} - \theta_2[B]^{\nu_2} + \theta_3[C]^{\nu_3} \\ \frac{d[C]}{dt} &= \theta_2[B]^{\nu_2} - \theta_3[C]^{\nu_3}. \end{aligned} \tag{28}$$

Optimum designs can then be found for the vector of parameters $\psi^T = (\theta^T \nu^T)$ of dimension $2p$. Except for special values of ν, in particular one, it is not possible to find analytical solutions to such sets of differential equations. The sensitivities then have to be found by the numerical solution of further sets of differential equations. An outline is given in §8 of Atkinson (2003) with further details and examples in Atkinson and Bogacka (2002)

It is clear that the designs compared for robustness in this paper will not be highly efficient in this situation with twice as many parameters. The analogous analysis to that given here is to compare locally optimum designs intended for the estimation of θ and ν as λ varies and all ν_j^0 are taken equal to one.

15 Discussion

The results of this paper show how, by suitable choice of prior distribution, Bayesian D-optimum designs can be found that are efficient for estimation of the parameters in a nonlinear model over a wide range of response transformations. The method applies to nonlinear models. For linear regression models, transformation of the response does not affect the range of the response; transformation of both sides is not required and the design is unaffected, unless, as in Atkinson and Cook (1997), it is required to estimate the transformation.

The comparisons of designs in Figures 12 and 14 have shown that the locally optimum design for the correct value of λ has good properties against other departures. The equispaced design, which does not vary with λ, also is reasonably efficient, unless λ is low. In this case the design lacks support at low values of t. It is tempting to explore the efficiency of designs generated by adding more support points in the sequence created by division by $\sqrt{2}$

used to generate the design. Perhaps also some support points in the region between $t = 5$ and 15, that do not appear in Figure 6, could be removed. A further exploration would be to include the most extreme Bayesian design in Table 2 in the comparisons. However, the purpose of this paper is to show how such comparisons can be used to assess the robustness of designs. Generation of designs can be done from an extension of (16) to generalized D-optimality with criterion

$$\Phi(\xi, \psi, \lambda) = \sum_{\iota} q_{\iota} \log |M_{\iota}(\xi, \psi, \lambda)|,$$

where $M_{\iota}(\xi, \psi, \lambda)$ is the information matrix for models such as those in the preceding sections and q_{ι} is a weight. If interest is primarily in robustness against departures, D-optimality can be replaced by minimax designs (Sitter 1992, Dette and Biedermann 2003).

A feature of the designs found here as $\lambda \to 0$ is that design conditions become extreme. This is because, as the concentrations go to zero, so do the errors of measurement. Rocke and Lorenzato (1995) query this aspect of Horwitz's rule and suggest an alternative with two error components for which, although standard deviation decreases with concentration, it does not go to zero. Such an error model would give less extreme designs than those found here for very small positive λ and indicated for negative λ.

It has been assumed that all responses are subject to the same transformation. However, as the examples of multivariate data analysis in Chapter 4 of Atkinson, Riani, and Cerioli (2004) show, each response may have an individual transformation parameter λ_i. The change to the design criterion is slight, the common value λ in the sensitivities (8) being replaced by λ_i in the calculation of the sensitivity for η_i. However, if each response has a different λ_i, the optimum design will vary over the m-dimensional space Λ, rather than over the single dimension studied here.

Several other extensions of this work are possible. One important direction is to the design of experiments for industrial processes where the dynamics of the model are described by sets of differential equations. Any correlation of errors is ignored, for example as in Bauer, Bock, Körkel, and Schlöder (2000). So a second extension is to design when the errors are correlated. Recent papers describing algorithms for optimum designs in the presence of correlated errors include Müller and Pázman (2003) and Uciński and Atkinson (2004).

References

Atkinson, A. C. (2003). Horwitz's rule, transforming both sides and the design of experiments for mechanistic models. *Applied Statistics 52*, 261–278.

Atkinson, A. C. and B. Bogacka (2002). Compound and other optimum designs for systems of nonlinear differential equations arising in chemical kinetics. *Chemometrics and Intelligent Laboratory Systems 61*, 17–33.

Atkinson, A. C., K. Chaloner, A. M. Herzberg, and J. Juritz (1993). Optimum experimental designs for properties of a compartmental model. *Biometrics 49*, 325–337.

Atkinson, A. C. and R. D. Cook (1997). Designing for a response transformation parameter. *Journal of the Royal Statistical Society B 59*, 111–124.

Atkinson, A. C. and D. R. Cox (1974). Planning experiments for discriminating between models (with discussion). *Journal of the Royal Statistical Society, Series B 36*, 321–348.

Atkinson, A. C., C. G. B. Demetrio, and S. Zocchi (1995). Optimum dose levels when males and females differ in response. *Applied Statistics 44*, 213–226.

Atkinson, A. C., M. Riani, and A. Cerioli (2004). *Exploring Multivariate Data with the Forward Search*. New York: Springer–Verlag.

Bauer, I., H. G. Bock, S. Körkel, and J. P. Schlöder (2000). Numerical methods for optimum experimental design in DAE systems. *Journal of Computational and Applied Mathematics 120*, 1–25.

Box, G. E. P. and D. R. Cox (1964). An analysis of transformations (with discussion). *Journal of the Royal Statistical Society, Series B 26*, 211–246.

Box, G. E. P. and H. L. Lucas (1959). Design of experiments in nonlinear situations. *Biometrika 46*, 77–90.

Carroll, R. J. and D. Ruppert (1988). *Transformation and Weighting in Regression*. London: Chapman and Hall.

Dette, H. and S. Biedermann (2003). Robust and efficient designs for the Michaelis-Menten model. *Journal of the American Statistical Association 98*, 679–686.

Downing, D., V. V. Fedorov, and S. Leonov (2001). Extracting information from the variance function: optimal design. In A. C. Atkinson, P. Hackl,

and W. G. Müller (Eds.), *MODA 6 – Advances in Model-Oriented Design and Analysis*, pp. 45–52. Heidelberg: Physica-Verlag.

Draper, N. R. and W. G. Hunter (1966). Design of experiments for parameter estimation in multiresponse situations. *Biometrika 53*, 596 – 599.

Kiefer, J. and J. Wolfowitz (1960). The equivalence of two extremum problems. *Canadian Journal of Mathematics 12*, 363–366.

Läuter, E. (1974). Experimental design in a class of models. *Mathematische Operationsforschung und Statistik, Serie Statistik 5*, 379–398.

Lischer, P. (1999). Good statistical practice in analytical chemistry. In B. Grigelionis (Ed.), *Probability Theory and Mathematical Statistics*, pp. 1–12. Dordrecht: VSP.

Müller, W. G. and A. Pázman (2003). Measures for designs in experiments with correlated errors. *Biometrika 90*, 423–434.

Rocke, D. M. and S. Lorenzato (1995). A two-component model for measurement error in analytical chemistry. *Technometrics 37*, 176–184.

Sitter, R. R. (1992). Robust designs for binary data. *Biometrics 48*, 1145–1155.

Uciński, D. and A. C. Atkinson (2004). Experimental design for processes over time. *Studies in Nonlinear Dynamics and Econometrics 2*(2). (Article 13).

Zorn, M. E., R. D. Gibbons, and W. C. Sonzogni (1997). Weighted least-squares approach to calculating limits of detection and quantification by modeling variability as a function of concentration. *Analytical Chemistry 69*, 3069–3075.

A Appendix: Four-Parameter Model

The reaction scheme for the four-parameter model is given in (27). If all reactions are first order and the initial concentrations of A is one with those of B C and D all zero, $\eta_A(t,\theta)$ again follows exponential decay (11) with $\theta = \theta_4$. The other concentrations are given by

$$\begin{aligned}
\eta_D(t,\theta) &= \frac{\theta_4}{\theta_4-\theta_1}\left\{e^{-\theta_1 t} - e^{-\theta_4 t}\right\} \\
\eta_B(t,\theta) &= \frac{\theta_3}{\theta_2+\theta_3}\left\{1 - e^{-(\theta_2+\theta_3)t}\right\} - \frac{(\theta_1-\theta_3)\theta_4}{(\theta_1-\theta_2-\theta_3)(\theta_4-\theta_1)}\left\{e^{-\theta_1 t} - e^{-(\theta_2+\theta_3)t}\right\}
\end{aligned}$$

$$
\begin{aligned}
& -\frac{(\theta_4-\theta_3)\theta_1}{(\theta_2+\theta_3-\theta_4)(\theta_4-\theta_1)}\left\{e^{-\theta_4 t}-e^{-(\theta_2+\theta_3)t}\right\} \\
\eta_C(t,\theta) &= 1-\eta_A(t,\theta)-\eta_B(t,\theta)-\eta_D(t,\theta). \qquad \text{(A1)}
\end{aligned}
$$

The parameter sensitivities are found by straightforward, albeit tedious, differentiation of these expressions with respect to the four parameters.