Horwitz' Rule, Transforming Both Sides and the Design of Experiments for Mechanistic Models

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

December 11, 2001

Abstract

The paper develops methods for the design of experiments for mechanistic models when the response has to be transformed to achieve symmetry and constant variance. The power transformation used is partially justified by a rule in analytical chemistry. Because of the nature of the relationship between response and the mechanistic model, it is necessary to transform both sides of the model. Expressions are given for the parameter sensitivities in the transformed model and examples given of optimum designs, not only for single response models, but for experiments in which multivariate responses are measured and for experiments in which the model is defined by a set of differential equations which cannot be solved analytically.

Keywords: Analytical chemistry; Box-Cox transformation; Chemical kinetics; D-optimum design; Direct Method; Parameter sensitivities.

1 Introduction

This paper is concerned with the design of experiments for the mechanistic models which typically arise in pharmacokinetics and chemical kinetics. It is customary to convert such models into statistical models by including additive independent errors of constant variance. This paper presents evidence that this is often inappropriate, the variance increasing with the mean.

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

The variance can be stabilised, and the error distribution made symmetrical, by transforming the response. But, with the mechanistic models that are considered here, it is necessary to transform both sides of the model. The optimum design will then depend on the transformation that is appropriate as well as on the model.

The information matrix of the parameter estimates is a function of the parameter sensitivities, that is the derivatives of the response with respect to the parameters. A simple expression is obtained for the sensitivities for the transformation model. However, many mechanistic models are sets of differential equations which cannot be solved analytically. Differential equations for the sensitivities can however be derived, which are then solved numerically along with the equations for the model.

The paper starts with the simple example of designs for estimating the single parameter in exponential decay and shows how the design changes if, because of the error structure, it is appropriate to work with the logarithms of the observations. It is shown that the log transformation, which is sometimes used in pharmacokinetics, for example Mentré, Mallet, and Baccar (1997) or Stroud, Müller, and Rosner (2001, p.354), gives a silly design in this example. §3 presents evidence that constant variance is sometimes an inappropriate model for the errors. Part of this evidence is Horwitz' Rule in analytical chemistry, which establishes an empirical relationship between concentration and variance. A power transformation of the response is needed to stabilise variance. For linear regression models transformation of the response does not affect the design, unless, as in Atkinson and Cook (1996), it is required to estimate the transformation. This is not the case here, since Horwitz' rule gives a value for the transformation parameter. However, for nonlinear mechanistic models, both sides of the model have to be transformed, with consequences for the experimental design.

Optimum design for both single and multivariate responses with additive errors is reviewed in §4. The seminal statistical paper on design for parameter estimation in the mechanistic models which typically arise in pharmacokinetics and chemical kinetics under this error structure is Box and Lucas (1959). The addition of such errors has continued to be the standard statistical approach. Recent examples in the process modelling literature include Bauer, Bock, Körkel, and Schlöder (2000) and Asprey and Macchietto (2000). §5 develops the theory of designs for transformations. In particular, a simple form is found for the parameter sensitivities after transformation of both sides of the model. The simplest example, exponential decay, is investigated more fully in Section 6 and designs found for a range of transformations. In section 7 the model is that for two consecutive reactions. Designs are found for the transformed response when the concentrations of either one or two

chemical species are measured. In the example in Section 8 the differential equations defining the model do not have an analytical solution. The differential equations defining the sensitivities are derived and solved along with the equations defining the model. An optimum design for all four parameters in this model is found when only one response is measured. Comments in Section 9 conclude.

2 Transformations and Exponential Decay

A simple example of the effect of transformation of the response on experimental design and parameter estimation comes from the nonlinear response model resulting from first-order decay

$$A \xrightarrow{\theta} B$$

in which the concentration of chemical A at time t is given by the nonlinear function

$$[A] = \eta_A(t, \theta) = e^{-\theta t} \qquad (\theta, t \geq 0), \tag{1}$$

if it is assumed that the initial concentration of A is 1. If the ith experiment consists of measuring the concentration of A at time t_i, a simple statistical model of the observations is

$$y_i = \eta_A(t_i, \theta) + \epsilon_i, \tag{2}$$

where the errors ϵ_i are independently distributed with zero mean and constant variance. Unweighted least squares is the appropriate method of estimation. The variance of the least squares estimator $\hat{\theta}$ then depends on the parameter sensitivity

$$f(t_i, \theta) = \frac{d\eta_A(t_i, \theta)}{d\theta} = -t_i \exp(-\theta t_i). \tag{3}$$

Both Box and Lucas (1959) and Atkinson and Donev (1992) find the locally D-optimum designs minimising the variance of $\hat{\theta}$, which consist of taking all measurements where $f(t, \psi)$ is a maximum, that is at a time $t^* = 1/\theta$.

Now suppose that the model needs to be transformed to give constant variance. If the log transformation is appropriate and $[A]$ is measured, taking logarithms of both sides of (1), combined with additive errors, yields the statistical model

$$\log y_i = \log\{\eta_A(t_i, \theta)\} + \epsilon_i = -\theta t_i + \epsilon_i. \tag{4}$$

The log transformation thus results in a linear statistical model with response $\log y$, for which the parameter sensitivity is just the time t. The optimum design puts all observations at the maximum possible time, when the concentration is as small as possible, a clearly absurd answer. If there is an upper limit to the time of experimentation, all measurements will be taken at this time. The multiparameter designs found by Mentré et al. (1997) or Stroud et al. (2001), mentioned in §1 as incorporating a log transformation, do indeed contain some trials at the upper limit of time. As is shown in §6 for the model of exponential decay, less severe transformations give less extreme designs.

3 Empirical Evidence for Transformations

3.1 Power Transformations and Data

Power transformation of the response is helpful if the variance of Y increases with the expected value $E(Y)$ of Y. If

$$\text{var}Y \propto \{E(Y)\}^{2(1-\lambda)}, \tag{5}$$

Taylor series expansion shows that the variance is approximately stabilized by using as the response

$$\begin{array}{ll} y^{\lambda} & \lambda \neq 0 \\ \log y & \lambda = 0. \end{array}$$

So, for $\lambda = 1$, the variance is independent of the mean and no transformation is necessary. When $\lambda = 0.5$, the variance is proportional to the mean and the square root transformation is indicated, whereas, when $\lambda = 0$, the standard deviation is proportional to the mean and the logarithmic transformation provides approximately constant variance.

If the power law (5) holds, large observations will have larger standard deviations than small ones. Taking logarithms of the square root of both sides of this relationship yields

$$\log(\text{s.d.}Y) = \gamma_0 + (1-\lambda)\log\{E(Y)\}, \tag{6}$$

where s.d. is the standard deviation of Y. If replicate observations are available, a plot of log standard deviation against log mean will indicate whether the power law holds We check this for two examples.

Downing, Fedorov, and Leonov (2001) use an extensive data set to illustrate optimum design with a parameterised variance function. The data

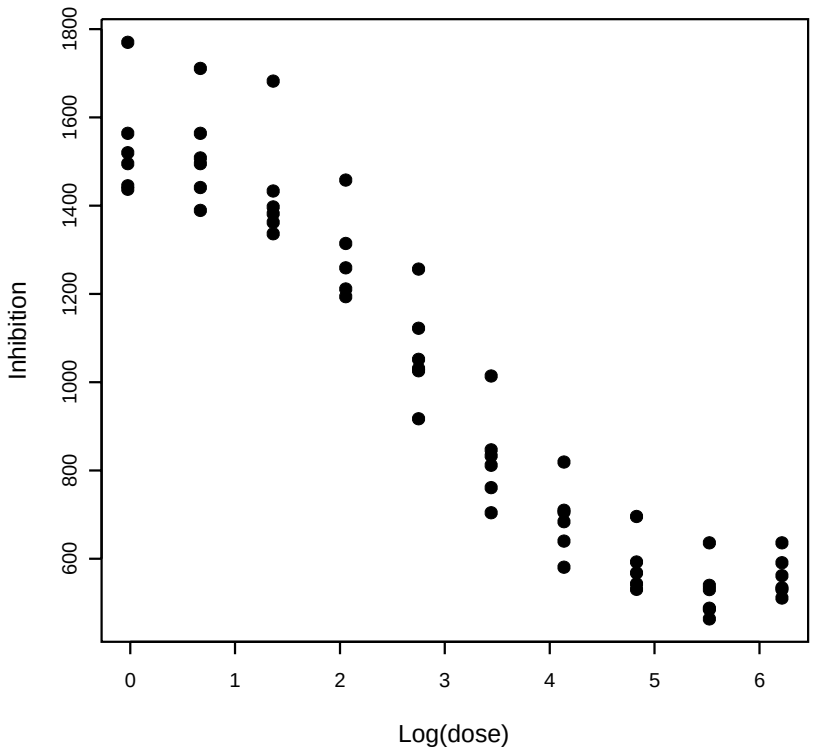

Figure 1: Inhibition data, preparation five: inhibition against log dose

come from a study on the ability of a compound to inhibit the proliferation of cells in a cell-based assay. The design used was a two-fold dilution series, starting with a concentration of 500ng./ml., which gave ten equally-spaced doses on the log scale. Twelve preparations were used, with readings on six samples at each dose level.

Figure 1 shows a plot of activity against log dose for preparation five. It is clear that the variance depends on concentration and that the error distribution is skew. Figure 2 is the plot of log standard deviation against log mean for the 120 sets of replicate observations. The slope of the least squares line is 0.9034 with a t value of 16.3, indicating a strong relationship between standard deviation and mean, even if there is appreciable scatter in

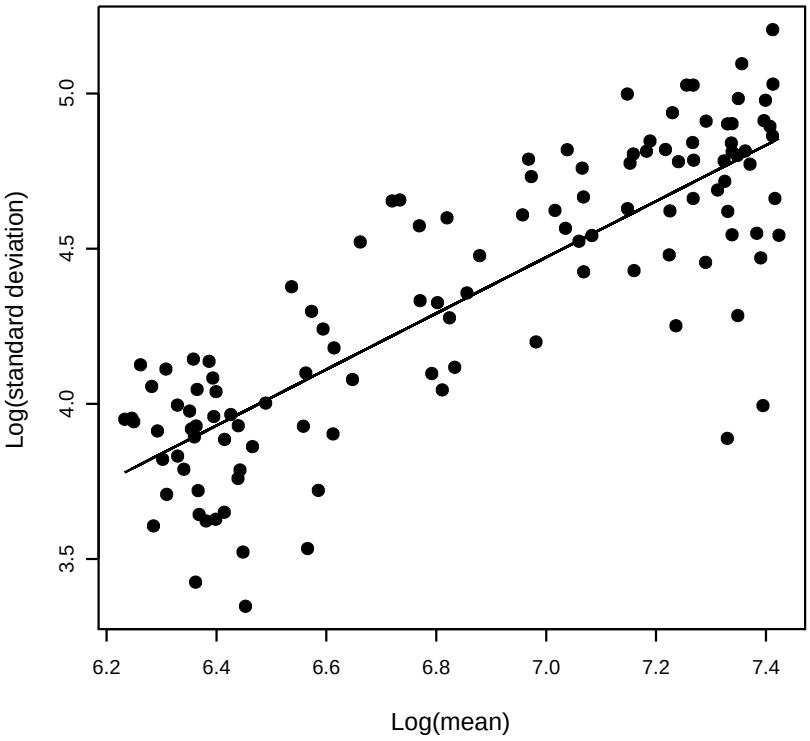

Figure 2: Inhibition data: log standard deviation against log mean for the 120 groups of six observations. The estimated slope is 0.90

the plot.

Similar relationships are obtained from other examples. Lindsey (2001, p.113) presents data on flosequinan, a now withdrawn drug used in the management of patients with chronic heart failure. The results for several doses are recorded. A dose of 150mg. was given to 18 healthy volunteers and readings taken at twelve times, at the first and last of which all readings were zero. The highest reading at each time was for volunteer 12, who Lindsey states was one of the two lightest and youngest of the volunteers. If this subject is removed from the calculations there are 17 non-zero observations at ten times. The plot of log standard deviation against log mean again shows a linear relationship. The least squares regression line in this example has a slope of 0.6494 and a t value of 22.84.

3.2 Horwitz' Rule

The two preceding examples show a strong relationship between mean and standard deviation. From (6) the slope of the plots is an estimate of $1-\lambda$, so that quite strong transformations are indicated for both data sets: 0.35 for flosequin and 0.1, not far from the log, for the inhibition data. An estimate of λ with better statistical properties, which is not needed here, can be found using the transformation methods of §5.

The linear relationship between log standard deviation and log mean is well established in analytical chemistry, where it is known as Horwitz' rule, an empirical relationship between the variability of chemical measurements and the concentration of the analyte. Lischer (1999) states that it is supported by the results of studies involving almost 10,000 individual data sets with analytes varying in concentration in mass/mass units from 1 (pure substances) to ultratrace contaminants, with concentrations of one part in 10^{-12}.

If the slope of the relationship between log standard deviation and log concentration were one, the log transformation would be appropriate. The increase is slightly less than this, depending on the laboratory, averaging around 0.86, so that the transformation is to the power 0.14, which will be the value used in this paper. No interpretation is given by Lischer for this value, which goes against standard statistical advice of using values with simple physical interpretations, such as the square root or the one third power for volumes. However the evidence for this rule is overwhelming.

It is surprising that both sets of data in §3.1 seem to support a Horwitz' style relationship. Rocke and Lorenzato (1995) argue that a mixed variance function is more appropriate for data taken on one instrument: the variance should be constant for small measurements, reflecting instrument error, but will increase for larger measurements. If so the plot of log standard deviation

against log mean should be hyperbolic, but there is no evidence of this for these two data sets.

4 Optimum Design

4.1 A Single Response

The focus of this paper is on the effect of the need to transform the response on the experimental design. This section summarises the theory for optimum design in the absence of transformation. More details are given in, for example, Atkinson and Donev (1992). The extension to transformation follows in the next section.

The experiments used as examples consist of measuring the concentration of one or more chemicals after the reaction has been running for a time t. In this section the theory is given when the concentration of only one chemical is measured. One experimental run yields one observation y_i and the experimental design is a list of the n times, $t_i, i = 1, \ldots, n$, not necessarily distinct, at which measurements are to be made.

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_i, of observations taken at these points

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

The times t_i are the points of support of the design ξ and w_i the design weights. In practice, when n observations must be taken, the design has the number of trials at t_i the integer closest to nw_i.

The nonlinear regression model is

$$y = \eta(t, \psi) + \epsilon \tag{7}$$

where the random errors ϵ are additive and independently, identically, normally distributed with zero mean and constant variance σ^2.

The information matrix of a design ξ for the p parameters ψ is

$$M(\xi, \psi) = F^T W F,$$

where F is a $k \times p$ matrix and the i-th row vector $f^T(t_i, \psi)$ has j-th element

$$f_j(t_i, \psi) = \frac{\partial \eta(t_i, \psi)}{\partial \psi_j}, \quad \text{for} \quad j = 1, \ldots, p,$$

called the sensitivity for parameter j. The matrix of weights $W = \text{diag}\{w_1, \ldots, w_p\}$. The extension of the parameter sensitivities to designs for transformation is described in §5. The information matrix thus depends on the unknown parameters ψ. Here only locally optimum designs will be considered in which a best guess ψ^o is taken for the parameters.

In the examples the calculations are of D-optimum designs maximizing the logarithm of the determinant of the information matrix, $\log|M(\xi, \psi)|$. The use of other design criteria, as well as Bayesian designs to reflect parameter uncertainty, in experiments for chemical kinetics is described by Atkinson and Bogacka (1997). The well-known Equivalence Theorem of Kiefer and Wolfowitz (1960) relates maximization of the determinant of the information matrix to minimization of the maximum variance of the predicted response over $\mathcal{T}$. With the standardized variance of the prediction at t defined by

$$d(t, \xi, \psi) = f^T(t, \psi) M^{-1}(\xi, \psi) f(t, \psi), \tag{8}$$

the Equivalence Theorem states that, for the optimum design, ξ^*, the maximum value of $d(t, \xi^*, \psi)$ over the design region, $\mathcal{T}$, is p, the number of parameters in the model, and further that this maximum value is attained at the support points t_i^* of ξ^*. The theorem provides a basis for the construction and checking of D-optimum designs. All the locally D-optimum designs for nonlinear models found in this paper have exactly p points of support. For univariate responses the weights at the support points are equal to $1/p$. This is not the case for multivariate designs.

4.2 Optimum Design for Multivariate Response

Now suppose that the concentration of more than one chemical is measured. There will then 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 $u = 1, ..., k$ denote the design points and $j = 1, ..., p$ denote the parameters. The generalization of the single-response case is that now the observations follow the model

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

with

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

when the variance-covariance matrix of the responses is

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

Draper and Hunter (1966), following arguments similar to those of Box and Lucas (1959) for the single-response case, show that for normally distributed errors the information matrix is given by

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

where $\Sigma^{-1} = \{\sigma^{il}\}_{i,l=1,\ldots,m}$.

The results of Fedorov (1972, p.212) show that a form of the usual equivalence theorem applies for D-optimality. If the standardized variance of prediction in (8) is extended to

$$d_{il}(t,\xi,\psi) = f_i^T(t,\psi) M^{-1}(\xi,\psi) f_l(t,\psi), \tag{10}$$

with $M(\xi,\psi)$ given by (9), the Equivalence Theorem of §2.2 applies to

$$d(t,\xi,\psi) = \sum_{i=1}^{m}\sum_{l=1}^{m} \sigma^{il} d_{il}(t,\xi,\psi).$$

5 Parameter Sensitivities and Transforming Both Sides

Horwitz' rule is expressed by a linear relationship between log standard deviation and log concentration reflecting the increase in standard deviation at higher concentrations. Since concentrations are non-negative, the distributions of measurements are skewed to the right and stable variance, together with symmetry, can be achieved by use of the Box-Cox power transformation.

For transformation of just the response y in a regression model, Box and Cox (1964) analyze the normalized power transformation

$$z(\lambda) = \begin{cases} (y^\lambda - 1)/(\lambda \dot{y}^{\lambda-1}) & \lambda \neq 0 \\ \dot{y} \log y & \lambda = 0, \end{cases} \tag{11}$$

where the geometric mean of the observations is written as $\dot{y} = \exp(\Sigma \log y_i / n)$. When $\lambda = 1$, there is no transformation. The model to be fitted is (7) with response $z(\lambda)$. The parameter sensitivities for this model with $\lambda = 1$ will be written

$$f_j^1(t,\psi) = \frac{\partial \eta(t,\psi)}{\partial \psi_j}. \tag{12}$$

If the model $\eta(t, \psi)$ is a polynomial, design for known λ does not depend on the need to transform the data. The data are transformed before fitting the model to give homogeneity of variance so that the optimum design does not depend on the particular transformation employed. However, if $\eta(t, \psi)$ is a mechanistic model based on chemical kinetics, the relationship between the observed response and the concentrations of the other reactants needs to be preserved after transformation.

An example of this need to transform both sides is transformation of the concentration of A in §2. For the exponential decay model (1) the response varies between one and zero. But if the log transformation is needed to give a symmetrical error distribution with constant variance, the response in (4) goes from zero to minus infinity. The right-hand side of the model needs to have the same range as indeed it does. This is achieved by transformation of both sides of the model, as described in Chapter 4 of Carroll and Ruppert (1988) and §4.11 of Atkinson and Riani (2000). We transform both sides of the expectation of the nonlinear model (7)

$$E(Y) = \eta(t, \psi)$$

with the normalized Box–Cox transformation (11), to obtain

$$z(\lambda) = \begin{cases} (y^\lambda - 1)/(\lambda \dot{y}^{\lambda-1}) & = & (\eta^\lambda - 1)/(\lambda \dot{y}^{\lambda-1}) & \lambda \neq 0 \\ \dot{y} \log y & = & \dot{y} \log \eta & \lambda = 0, \end{cases} \tag{13}$$

where, as before, the geometric mean of the observations is written as $\dot{y}$.

For fixed $\lambda \neq 0$, estimation of the parameters ψ in (13) does not depend on whether the response is $z(\lambda)$ or the nonnormalized y^λ. Multiplication of both sides of (13) by $\lambda \dot{y}^{\lambda-1}$ and simplification, leads to the model

$$y^\lambda = \{\eta(t, \psi)\}^\lambda + \epsilon. \tag{14}$$

The parameter sensitivities in this transformed model are

$$f_j^\lambda(t, \psi) = \frac{\partial \{\eta(t, \psi)\}^\lambda}{\partial \psi_j} = \lambda \{\eta(t, \psi)\}^{\lambda-1} \frac{\partial \eta(t, \psi)}{\partial \psi_j} = \lambda \{\eta(t, \psi)\}^{\lambda-1} f_j^1(t, \psi). \tag{15}$$

For fixed λ, multiplication by λ in (15) does not change the optimum design. So, for the transformation suggested by Horwitz' Rule, the sensitivities have the easily calculated form

$$f_j^H(t, \psi) = \{\eta(t, \psi)\}^{-0.86} f_j^1(t, \psi) = f_j^1(t, \psi)/\{\eta(t, \psi)\}^{0.86}. \tag{16}$$

Thus transformation of both sides for values of $\lambda < 1$ will increase the relative value of the sensitivities for times where the response is small. We can

expect that designs for Horwitz' rule will this include observations at lower concentrations than those when no transformation is needed. The results in the next sections show this to be the case.

6 Exponential Decay

The nonlinear response model resulting from first-order decay introduced in §2 comes from solution of the differential equation

$$\frac{d[A]}{dt} = -\theta[A].$$

If it is assumed that the initial concentration of A is one, the concentration of A is given by the exponential model (1). 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

$$f_A^1(t, \theta) = -t \exp(-\theta t), \tag{17}$$

whereas if $[B]$ is measured

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

both of which have their extreme value at the time $t^* = 1/\theta$. Therefore all readings should be taken at this one value of time.

Now suppose that the model needs to be transformed to give constant variance. It has already been shown that if the log transformation is appropriate and $[A]$ is measured (4) the optimum design puts all observations at the maximum possible time. Less severe transformations give less extreme designs.

From (15) the parameter sensitivity for the power transformation λ when $[A]$ is measured is

$$f_A^\lambda(t, \theta) = \{\eta_A(t, \theta)\}^{\lambda - 1} f_A^1(t, \theta) = -t \exp(-\lambda\theta t). \tag{18}$$

The optimum design is therefore at a time of $1/(\lambda\theta)$. As λ decreases, the time for the optimum design increases reaching, as we have seen, infinity when $\lambda = 0$, the log transformation.

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

$$f_B^\lambda(t, \theta) = \{\eta_B(t, \theta)\}^{\lambda - 1} f_B^1(t, \theta) = t \exp(-\theta t)\{1 - \exp(-\theta t)\}^{\lambda - 1}, \tag{19}$$

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

Figure 3 is a plot of the optimum time at which the reading of the concentration of A or B should be taken as a function of λ, when $\theta = 0.2$, so that the optimum time, in the absence of a transformation, is 5. The figure shows the strong dependence of the design on the need for a transformation. For the Horwitz value of 0.14 the optimum times are 35.7 and 1.19. Although the time when A is measured is much greater than the time of 5 when there is no need for a transformation, we have already seen in §2 that the log transformation leads to a measurement at the maximum possible time.

7 Multivariate Response: Two Consecutive First-Order Reactions

The examples in this section and in §8.2 are based on the model for two consecutive reactions

$$A \stackrel{\theta_1}{\to} B \stackrel{\theta_2}{\to} C. \tag{20}$$

The kinetic differential equations for $[A], [B]$ and $[C]$, the concentrations of the chemical compounds A, B and C as functions of time t are

$$\begin{aligned} \frac{d[A]}{dt} &= -\theta_1[A]^{\nu_1} \\ \frac{d[B]}{dt} &= \theta_1[A]^{\nu_1} - \theta_2[B]^{\nu_2} \\ \frac{d[C]}{dt} &= \theta_2[B]^{\nu_2}, \end{aligned} \tag{21}$$

where θ_1 and θ_2 are the rates of reaction and ν_1 and ν_2 are the orders. Atkinson and Bogacka (1997) discuss the consequences for experimental design of this distinction between the two parts of ψ.

In this section we take both reactions as first order, that is $\nu_1 = \nu_2 = 1$. Given the initial concentrations of A, B and C, an explicit algebraic solution can be found for the concentrations as a function of time. If the initial concentration of A is one and that of B and C are zero, $\eta_A(t, \theta)$ follows the exponential decay (1) with $\theta = \theta_1$. The other concentrations are given by

$$\begin{aligned} \eta_B(t, \theta) &= \frac{\theta_1}{\theta_1 - \theta_2}(e^{-\theta_2 t} - e^{\theta_1 t}) \\ \eta_C(t, \theta) &= 1 - \eta_A(t, \theta) - \eta_B(t, \theta). \end{aligned} \tag{22}$$

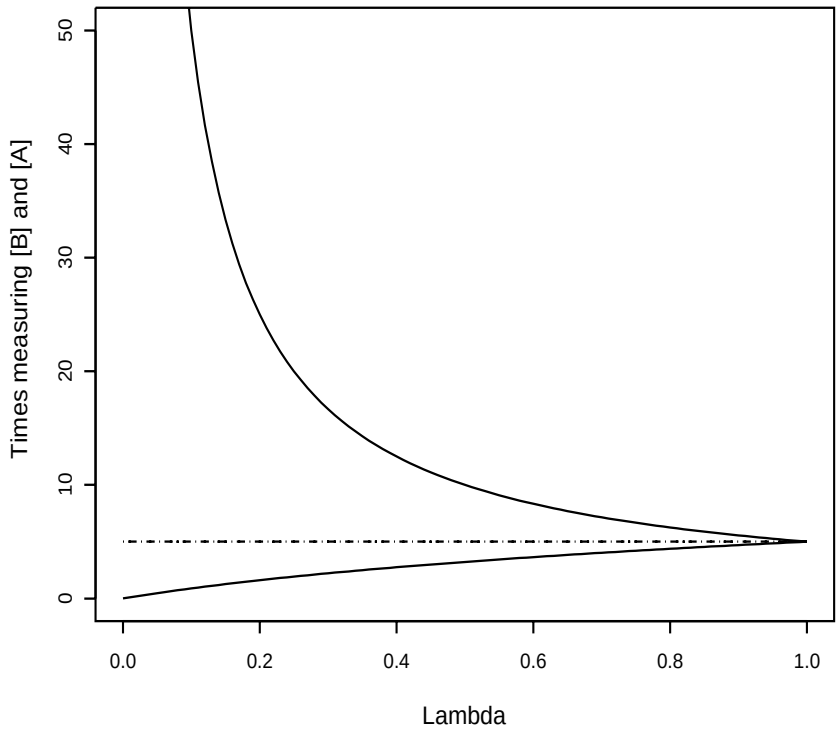

Figure 3: Exponential decay: time of the optimum measurement of concentration as a function of the transformation parameter λ when $[A]$ is measured (upper curve) and when $[B]$ is measured (lower curve)

The parameter sensitivities for $\eta_B(t,\theta)$ are

$$\begin{aligned} f^1_{B1}(t,\theta) &= \{-(\theta_2/\theta_1)\eta_B(t,\theta) + \theta_1 t e^{-\theta_2 t}\}/(\theta_1-\theta_2) \\ f^1_{B2}(t,\theta) &= \{\eta_B(t,\theta) - \theta_1 t e^{-\theta_2 t}\}/(\theta_1-\theta_2) \end{aligned} \tag{23}$$

and, since $[A]+[B]+[C]=1$,

$$f^1_{Cj}(t,\theta) = -f^1_{Aj}(t,\theta) - f^1_{Bj}(t,\theta). \tag{24}$$

In (24) $f^1_{A1}(t,\theta)$ is given by (17) with $\theta=\theta_1$ and $f^1_{A2}(t,\theta)=0$. The parameter sensitivities for the transformed model are then found from (15).

Information on θ_2 cannot be obtained from measurements only on $[A]$. The value of $[B]$ rises from zero and falls again to zero, providing information on both parameters. By comparison, $[C]$ which rises from zero to one provides rather imprecise information about the value of θ_1.

To illustrate the properties of multiresponse designs we take the two responses as having the same variance and zero covariance, with $\theta_1=0.7$ and $\theta_2=0.2$. The resulting optimum designs are given in Table 1. In all designs there are only two design points. When a single response is measured the weights are 0.5 at each point. When both responses are measured the weights are no longer quite equal.

The first part of the table gives designs when no transformation is necessary. The two design points when $[B]$ and $[C]$ are both measured are either side of the time for maximum concentration which is at $t=2.51$. This design is similar to that when only $[B]$ is measured, which has a relative efficiency of 68% as measured by the square root of the ratio of determinants of the information matrices for the two designs. Both times for the design for measurement of $[C]$ only are above the time of maximum concentration of $[B]$.

The effect of the transformation in the second part of the table is to move all design points to lower concentrations of the response being measured. The upper time for measurement of only $[B]$ therefore increases, while the lower one decreases. The same is true for the design for both $[B]$ and $[C]$. Measurement of $[C]$ does not add so much information in this case and the design for measurement of only $[C]$, in which both times are now smaller, has an efficiency of only 7.2% relative to that when both responses are measured.

8 Designs Requiring Numerical Derivatives

8.1 The Direct Method

If the orders ν_i in (21) are not both one, the differential equations for the responses have to be solved numerically. Solution of the equations does not

Untransformed

Measured Response	t_1 (w_1	t_2 w_2)	Determinant	Efficiency Percent.
$[B]$	1.23	6.86	0.164E00	68.1
	(0.5	0.5)		
$[C]$	3.37	9.98	0.826E-2	15.3
	(0.5	0.5)		
$[B]$ and $[C]$	1.34	6.55	0.354E00	100.
	(0.511	0.489)		

Horwitz

Measured Response	t_1 (w_1	t_2 w_2)	Determinant	Efficiency Percent.
$[B]$	0.33	37.62	0.436E02	83.2
	(0.5	0.5)		
$[C]$	0.71	6.57	0.323E00	7.2
	(0.5	0.5)		
$[B]$ and $[C]$	0.48	38.32	0.630E02	100.
	(0.517	0.483)		

Table 1: D-optimum designs for both parameters in the consecutive first-order model (22), both without and with transformation.

provide information on the sensitivities. These could in principle be found by the crude method of solving the differential equations for two adjacent values of the parameters and then using differencing to approximate the derivatives. A better procedure is to augment the kinetic equations with a second set of differential equations for the sensitivities. This numerical procedure for finding the values of the parameter sensitivities is called the "direct method" (Valko and Vajda 1984). Examples in experimental design in chemical engineering are given by Bauer et al. (2000) and in chemical kinetics by Atkinson and Bogacka (2001). A comparison of numerical methods for finding sensitivities is given by Uciński (1999).

The response functions $\eta_i(t, \psi)$, $i = 1, \ldots, m$, are solutions of the set of m differential equations

$$\frac{d\eta_i(t,\psi)}{dt} = g_i(\eta, t, \psi) \tag{25}$$

with given initial conditions. The differential equations satisfied by the parameter sensitivities are found by differentiation of (25) to be, in matrix notation,

$$\frac{df_j(t,\psi)}{dt} = \frac{\partial g(\eta,t,\psi)}{\partial \eta} f_j(t,\psi) + \frac{\partial g(\eta,t,\psi)}{\partial \psi_j}, \qquad j = 1, \ldots, p. \tag{26}$$

These have to be solved together with equations (25) for given prior values of the parameter vector ψ.

8.2 Two Consecutive Second-Order Reactions

As an example of the use of numerical sensitivities, consider design for the consecutive reaction (21) when both reactions are second order, that is $\nu_1 = \nu_2 = 2$. Partial solutions for the concentrations and the differential equations satisfied by the concentration of B and by the sensitivities are given in the Appendix. They were solved using the Maple procedure "`dsolve`", which produces answers at a grid of time points. Interpolation used the Fortran subroutines `locate` and `polint` from Press et al. (1992, Chapter 3).

To illustrate the effect of transformation on design again let $\theta_1 = 0.7$ and $\theta_2 = 0.2$. The design region is such that the maximum time of measurement is 50. Only the concentration of B is measured. The optimum design in the absence of transformation for these four parameters, given in Table 2, puts two trials either side of the time of maximum response 3.3. One observation is at the maximum allowable time of 50. If a transformation is needed the observations are again taken at times of lower concentration, except for $t = 50$. The effect is most noticeable in the time of the first observation.

	t_1	t_2	t_3	t_4
Untransformed	0.57	2.65	9.68	50.0
Horwitz	0.14	2.13	12.61	50

Table 2: D-optimum designs for all four parameters in the second-order consecutive model (21), both without and with transformation.

9 Discussion

This paper shows that the design of experiments for the parameters in mechanistic models depends strongly on whether the response needs to be transformed to give constant variance. Of the three examples in the paper, two are non-standard involving multiple responses and the numerical calculation of sensitivities. The paper shows how to design experiments when such transformations are needed.

Transformations were introduced to provide statistical models in which the observations had constant variance. An alternative (Bogacka and Wright 2001) is to use weighted least squares with weights proportional to $E(Y)^{-(1-\lambda)}$. The resulting parameter sensitivities and so designs, are identical to those of §5.

The weights in this form of weighted least squares include the parameters of the linear model. Information on these parameters is therefore obtained from the change of the variance with experimental conditions as well as from the change of mean. Atkinson and Cook (1995) find D-optimum designs for heteroscedastic linear models, including the special case when the linear model for the structure in the variance is the same as that for the mean. The resulting information matrix is the sum of two matrices, one of which is that for weighted least squares. An example of a design for a nonlinear model when the variance is of this special form is given by Downing, Fedorov, and Leonov (2001).

An important difference between such methods and the transformation studied here is in the model implied for fitting the data. Weighted least squares is the analogue of the method of Sheiner and Beal (1980) for linearising nonlinear models for parameter estimation. The error distributions are therefore treated as being symmetrical, although with non-constant variance. But, in the transformation model, the original observations will have skewed distributions, which become symmetrical, with constant error variance, after the appropriate transformation.

Acknowledgements

I am grateful to Dr Barbara Bogacka of Queen Mary, University of London for, amongst other matters, the numerical solution of the differential equations in §5

A Appendix: Sensitivity Equations for the Reaction $A \rightarrow B \rightarrow C$

The vector of the expected multivariate response in §5.2 is given by the solution of (21). The first equation can be solved analytically and the solution substituted in the second equation giving the differential equation for $[B]$

$$\frac{d[B]}{dt} = \theta_1\{1-(1-\nu_1)\theta_1 t\}^{\frac{\nu_1}{1-\nu_1}} - \theta_2[B]^{\nu_2}. \qquad \text{(A1)}$$

When only $[B]$ is measured, $m = 1$ and the system (26) simplifies to one equation for each of the four parameters. Then (A1) needs to be solved together with each of the following

$$\begin{aligned}
\frac{df^1_{B1}}{dt} &= -\theta_2\nu_2[B]^{\nu_2-1}f_1 + (1-\theta_1 t)\{1-(1-\nu_1)\theta_1 t\}^{\frac{\nu_1}{1-\nu_1}-1} \\
\frac{df^1_{B2}}{dt} &= -\theta_2\nu_2[B]^{\nu_2-1}f_2 - [B]^{\nu_2} \\
\frac{df^1_{B3}}{dt} &= -\theta_2\nu_2[B]^{\nu_2-1}f_3 + \frac{\theta_1}{(1-\nu_1)^2}\{1-(1-\nu_1)\theta_1 t\}^{\frac{\nu_1}{1-\nu_1}} log\{1-(1-\nu_1)\theta_1 t\} \\
&+ \frac{\nu_1\theta_1^2 t}{1-\nu_1}\{1-(1-\nu_1)\theta_1 t\}^{\frac{\nu_1}{1-\nu_1}-1} \\
\frac{df^1_{B4}}{dt} &= -\theta_2\nu_2[B]^{\nu_2-1}f_4(t,\psi) - \theta_2[B]^{\nu_2}\log[B].
\end{aligned}$$

The sensitivities $f^H_{Bj}(t,\psi)$ are again found from (16).

References

Asprey, S. M. and S. Macchietto (2000). Statistical tools for optimal dynamic model building. *Computers and Chemical Engineering 24*, 1261–1267.

Atkinson, A. C. and B. Bogacka (1997). Compound, D- and D_s-optimum designs for determining the order of a chemical reaction. *Technometrics 39*, 347–356.

Atkinson, A. C. and B. Bogacka (2001). Compound and other optimum designs for systems of nonlinear differential equations arising in chemical kinetics. *Chemometrics and Intelligent Laboratory Systems*. (In press).

Atkinson, A. C. and R. D. Cook (1995). D-optimum designs for heteroscedastic linear models. *Journal of the American Statistical Association 90*, 204–212.

Atkinson, A. C. and R. D. Cook (1996). Designing for a response transformation parameter. *Journal of the Royal Statistical Society B 59*, 111–124.

Atkinson, A. C. and A. N. Donev (1992). *Optimum Experimental Designs*. Oxford: Oxford University Press.

Atkinson, A. C. and M. Riani (2000). *Robust Diagnostic Regression Analysis*. 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.

Bogacka, B. and F. Wright (2001). A non-linear design problem in a chemical kinetic model with non-constant error variance. (submitted).

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.

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.

Fedorov, V. V. (1972). *Theory of Optimal Experiments*. New York: AcademicPress.

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

Lindsey, J. K. (2001). *Nonlinear Models in Medical Statistics*. Oxford: Oxford University Press.

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

Mentré, F., A. Mallet, and D. Baccar (1997). Optimal design in random-effects regression models. *Biometrika 84*, 429–442.

Press, W. H., S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery (1992). *Numerical Recipes in Fortran (2nd edition)*. Cambridge, England: Cambridge University Press.

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

Sheiner, L. B. and S. L. Beal (1980). Evaluation of methods for estimatng population pharmacokinetic parameters I. Michaelis-Menten model: Routine clinical pharmacokinetic data. *Journal of Pharmacokinetics and Biopharamaceutics 8*, 553–571.

Stroud, J. R., P. Müller, and G. L. Rosner (2001). Optimal sampling times in population pharmacokinetic studies. *Applied Statistics 50*, 345–359.

Uciński, D. (1999). *Measurement Optimization for Parameter Estimation in Distributed Systems*. Zielona Góra: Technical University Press.

Valko, P. and S. Vajda (1984). An extended ODE solver for sensitivity calculations. *Computers and Chemistry 8*, 255–271.