Department of Statistics
University of Wisconsin
Madison, Wisconsin

Technical Report No. 137
January 1968
Estimation of Derivatives
from Discrete Data
by
William G. Hunter and Ralph D. Pollard

This research was supported in part by the National Science Foundation under grant GK-1055.
0. ABSTRACT

In engineering and scientific experimental work it is
sometimes desired to estimate certain derivatives of a re-
sponse. These derivatives may be at one or more levels of
the independent variables. It is the purpose of this paper
to investigate the derivative estimation problem and to de-
velop appropriate statistical methods of experimental design
and analysis for the case where an adequate model of the re-
sponse is known. For this problem, the number of distinct
derivatives is a more important consideration than the order
of the derivatives. Each model has a characteristic design
which is appropriate when the number of distinct derivatives
to be estimated is equal to the number of parameters in the
model. A characteristic design is appropriate for any set
containing this number of distinct derivatives. When the de-
sired set contains a smaller number of distinct derivatives,
however, a different design will be required.

1. INTRODUCTION

In engineering and scientific experimental work it is
sometimes desired to estimate derivatives of a response with
respect to time or some other independent variable. First,
second, or higher order derivatives may be desired at one
or more levels of the independent variable. Although they
may in principle be of any order, usually in practice there
is more interest in derivatives of low order. It is the
purpose of this study to investigate the derivative estimation problem and to develop appropriate statistical methods of experimental design and analysis.

It will be assumed that an adequate model for the response is known and has the form

$$ y = f(\theta, \xi, t) + \varepsilon $$

(1.1)

where $y$ is an observation of the response, $\theta$ is a $p \times 1$ vector of unknown parameters, $\xi$ is an $h \times 1$ vector of controllable variables, $t$ is an independent variable, $\varepsilon$ is a random error, and $f(\theta, \xi, t)$ is a known function. The derivatives which are considered here are derivatives of the true response (that is, the response without error) with respect to the independent variable $t$. From (1.1) it may be seen that a derivative of order $k$ of this type has the form

$$ d_k(\theta, \xi, t) = \frac{\partial^k f(\theta, \xi, t)}{\partial t^k}. $$

(1.2)

The controllable variables $\xi$ are, like $t$, independent variables. They are given a separate notation, however, to distinguish them from the variable of differentiation $t$. Although attention here is focused on derivatives with respect to a single
independent variable \( t \), the methods are also appropriate in situations where there is a vector \( \mathbf{t} \) and mixed derivatives such as \( \frac{\partial^2 f(\theta, \xi_0, t_0)}{\partial t_1 \partial t_2} \) are to be estimated.

The data consist of a set of \( n \) observations \( \{y_u\} \) at the corresponding controllable variable settings \( \{x_u\} \) and the independent variable settings \( \{t_u\} \). It will be assumed that the errors \( \{\epsilon_u\} \) are from independent, normal \((0, \sigma^2)\) distributions.

**Examples**

The type of problem considered here is illustrated by the model

\[
y = \frac{2}{\sqrt{\pi}} \int_0^{\xi/2\sqrt{\theta t}} e^{-v^2} \, dv + \epsilon
\]

(1.3)

with first derivative

\[
d_1(\theta, \xi, t) = -\frac{1}{2} \xi (\pi \theta t^3)^{-\frac{1}{2}} \exp\{-\xi^2/4\theta t\}. \quad (1.4)
\]

This model with \( p=1 \) parameter \( \theta \) and \( h=1 \) controllable variable \( \xi \) arises in heat conduction work. (See, for example, Carslaw and Jaeger (1959).) The response \( y \) is a temperature, \( \xi \) is the thickness of a piece of material being investigated, \( \theta \) is the heat conduction constant for the material, and \( t \)
is time. The derivative $d_1(\theta, \xi, t)$, then, represents the rate of change of temperature with respect to time at a particular time $t$ and for a particular thickness $\xi$.

As another example, often in chemical engineering, although it is only possible to collect discrete data on chemical reactions in the form of space velocity ($v$) versus conversion ($y$), the primary interest is in the relationship between space velocity and rate ($dy/dt$). (See, for example, Levenspiel (1962).) Koenig (1966) and Hershey, Zakin, and Simha (1967) have considered some aspects of the analysis of data in situations of this kind, but not from a statistical point of view.

Organization of Paper

In the following section, Section 2, methods of experimental design and analysis will be developed for the case where the variance $\sigma^2$ of the observations is known and a single derivative of any order (including zeroth order, the response itself) is to be estimated. A Bayesian approach will be adopted. The problem of analysis consists, essentially, of constructing the appropriate posterior density function, which contains all the relevant information about the derivative of interest. The design problem is to select settings of the controllable and independent variables so
estimates of individual quantities which may be of special interest, for example, a point estimate of the derivative and an estimate of its precision. The posterior density function for the derivative can conveniently be developed from the posterior density function for the model parameters, \( p(\theta | y) \).

**The Posterior Density Function \( p(\theta | y) \)**

Given a set of data adequately represented by a model (1.1) containing unknown parameters \( \theta = (\theta_1, \theta_2, \ldots, \theta_p)' \), a posterior density function \( p(\theta | y) \) may be constructed. If the assumptions from Section 1 are valid for the errors \( \{ u \} \) and if a locally uniform prior distribution

\[
p(\theta) \propto \text{constant} \quad (2.1)
\]

is appropriate for \( \theta \), then \( p(\theta | y) \) has the multivariate normal form

\[
p(\theta | y) = \left| X'X \right|^{-\frac{1}{2}} (2\pi)^{-\frac{p}{2}} 2^p \exp \left\{ -\frac{1}{2\sigma^2} (\theta - \hat{\theta})' X'X(\theta - \hat{\theta}) \right\} \quad (2.2)
\]
be obtained by transforming the posterior density function \( p(\theta | y) \). Appropriate new parameters \( \phi = (\phi_1, \phi_2, \ldots, \phi_p) \) will be defined in terms of \( \theta \), \( p(\phi | y) \) will be determined from \( p(\theta | y) \), and the desired density function will be obtained simply as a marginal density function associated with \( p(\phi | y) \).

If \( d_k(\theta, \xi_0, t_0) \) can adequately be represented for \( \theta \) near \( \hat{\theta} \) by

\[
d_k(\theta, \xi_0, t_0) = d_k(\hat{\theta}, \xi_0, t_0) + \sum_{i=1}^{p} \frac{\partial d_k(\theta, \xi_0, t_0)}{\partial \theta_i} \bigg|_{\theta = \hat{\theta}} (\theta_i - \hat{\theta}_i), \tag{2.5}
\]

then the technique is particularly simple since a linear transformation from \( \theta \) to \( \phi \) may be used. For linear models (2.5) is, of course, always adequate, and for many nonlinear models it is adequate over ranges of \( \hat{\theta}, \xi_0 \), and \( t_0 \) which are of interest. In what follows, it will be assumed that the situation is such that expression (2.5) is adequate.

Let

\[
\phi_1 = d_k(\theta, \xi_0, t_0), \tag{2.6}
\]

\[
\hat{\phi}_1 = d_k(\hat{\theta}, \xi_0, t_0), \tag{2.7}
\]
and let \( b_1 \) be a \( p \times 1 \) vector with elements

\[
[b_1]_i = \left. \frac{\partial d_k(\theta, x_0, t_0)}{\partial \theta_i} \right|_{\theta = \hat{\theta}}.
\] (2.8)

(For linear models the elements of \( b_1 \), like the elements of \( X \), do not depend on \( \theta \).) Expression (2.5) may then be written as

\[
\phi_1 = \hat{\phi}_1 + b_1 (\theta - \hat{\theta}),
\] (2.9)

and additional parameters

\[
\phi_i = \hat{\phi}_i + b'_i (\theta - \hat{\theta}) \quad i=2,3,...,p
\] (2.10)

may be defined in terms of \( \hat{\theta} \) in a manner which is arbitrary except that the transformation between \( \theta \) and \( \phi \) must be nonsingular. In matrix form, the transformation may be written more compactly as

\[
\phi = \hat{\phi} + B (\theta - \hat{\theta}),
\] (2.11)

and the requirement that the transformation be nonsingular may be stated as the requirement that the \( p \times p \) matrix \( B \) be nonsingular. Let the \( n \times p \) matrix \( Z \) be defined by the re-
\( X = ZB. \) \quad (2.12)

Then by using (2.11) and (2.2) one can show that the posterior density function for \( \phi \) has the multivariate normal form

\[
P(\phi | y) = \frac{1}{|Z Z'|^\frac{1}{2} (2\pi)^\frac{p}{2} \sigma^{-p}} \exp\left\{ -\frac{1}{2\sigma^2}(\phi - \hat{\phi})' \frac{Z Z'}{2} (\phi - \hat{\phi}) \right\} \quad (2.13)
\]

The posterior density function for the derivative is the marginal density function for \( \phi_1 \); that is, it has the normal form

\[
P(d_1(\theta, \xi_0, t_0) | y) = p(\phi_1 | y) = \frac{1}{2} (2\pi)^{-\frac{1}{2}} \sigma^{-1} \exp\left\{ -\frac{c(\phi_1 - \phi_1)^2}{2\sigma^2} \right\}
\]

(2.14)

where \( c \) is defined by the expression

\[
1/c = \delta_1' (Z Z')^{-1} \delta_1 \quad (2.15)
\]

and where \( \delta_1 \) is a \( p \times 1 \) vector with first element equal to one and remaining elements equal to zero. Thus, \( 1/c \) is simply the upper left hand element of the matrix \((Z Z')^{-1}\) and can
be expressed as

\[ \frac{1}{c} = b_1' (X' X)^{-1} b_1, \]  

(2.16)

a more convenient form than (2.15) for some purposes.

From (2.14) it can be seen that \( \hat{\phi}_1 \) and \( c \) completely specify \( p(d_k(\theta, \xi_0, t_o) | y) \). Since this is so, the problem of determining \( p(d_k(\theta, \xi_0, t_o) | y) \) from a set of data may be viewed as the problem of determining \( \hat{\phi}_1 \) and \( c \). Furthermore, \( \hat{\phi}_1 \) (the value of \( d_k(\theta, \xi_0, t_o) \) at which the posterior density function attains its maximum value) may be used as a point estimate for the derivative, and \( \sigma^2/c \) (the posterior variance of \( d_k(\theta, \xi_0, t_o) \)) may be used as a measure of its precision.

**Examples**

The method by which the posterior density function \( p(d_k(\theta, \xi_0, t_o) | y) \) may be constructed for a set of \( n \) data points will now be illustrated with examples; in particular, it will be shown how \( \hat{\phi}_1 \) and \( c \), which completely determine \( p(d_k(\theta, \xi_0, t_o) | y) \), may be calculated. With actual data in problems of this kind, we would usually recommend that a plot of \( p(d_k(\theta, \xi_0, t_o) | y) \) be made because it would provide a simple, readily understandable picture representing the available information about the derivative of interest.
Example 2.1

Consider the three-parameter, linear model

\[ y = \theta_1 + \theta_2 t + \theta_3 t^2 + \epsilon \]  \hspace{1cm} (2.17)

where it is desired to estimate the first derivative at \( t = t_0 \)

\[ d_1(\theta, t_0) = \theta_2 + 2\theta_3 t_0 \]  \hspace{1cm} (2.18)

from a set of \( n \) data points. Then

\[
X = \begin{bmatrix}
1 & t_1 & t_1^2 \\
1 & t_2 & t_2^2 \\
\vdots & \vdots & \vdots \\
1 & t_n & t_n^2
\end{bmatrix}
\]  \hspace{1cm} (2.19)

and

\[ b_1' = [0 \ 1 \ 2t_0]. \]  \hspace{1cm} (2.20)

The maximum likelihood estimate \( \hat{\theta} \) can be determined from the expression

\[ \hat{\theta} = (X'X)^{-1} X'y, \]  \hspace{1cm} (2.21)
and using (2.7) one obtains

$$\hat{\phi}_1 = \hat{\theta}_2 + 2\hat{\theta}_3 t_o \quad (2.22)$$

The quantity $c$ can be calculated by using (2.16). Note that since the model is linear, neither $\hat{X}$ nor $b_1$ depends on $\hat{\theta}$. Consequently, the quantity $c$ does not depend on $\hat{\theta}$ either.

**Example 2.2**

Consider the single-parameter, nonlinear, exponential model

$$y = \exp \{-\theta t\} + \epsilon \quad (2.23)$$

where it is desired to estimate the first derivative at $t = t_o$

$$d_1(\theta, t_o) = -\theta \exp\{-\theta t_o\} \quad (2.24)$$

from a set of $n$ data points. Then
\[ X = \begin{bmatrix} -t_1 \exp(-\hat{\theta} t_1) \\ -t_2 \exp(-\hat{\theta} t_2) \\ \vdots \\ \vdots \\ -t_n \exp(-\hat{\theta} t_n) \end{bmatrix} \]  

(2.25)

and

\[ b'_1 = (-1 + \hat{\theta} t_0) \exp(-\hat{\theta} t_0). \]  

(2.26)

Since this model is nonlinear, the maximum likelihood estimate \( \hat{\theta} \) cannot be expressed in closed form but may be defined as that value of \( \theta \) which minimizes

\[ S(\theta) = \frac{1}{n} \sum_{u=1}^{n} [y_u - \exp(-\theta t_u)]^2. \]  

(2.27)

The quantity \( \hat{\phi}_1 \) may then be calculated by using (2.7),

\[ \hat{\phi}_1 = -\hat{\theta} \exp(-\hat{\theta} t_0), \]  

(2.28)

and the quantity \( c \) may be calculated using (2.16). Note that since this model is nonlinear both \( X \) and \( b_1 \) depend on \( \hat{\theta} \), and thus so does the quantity \( c \).
2.2 Design

The Design Criterion \( c \)

For a single derivative, the shape of the posterior density function depends only on the quantity \( c \), which is a function of the settings \( \{x_u\} \) and \( \{t_u\} \) at which the observations \( \{y_u\} \) are recorded.

If \( c \) is large, the density function is relatively peaked indicating that the derivative is well determined while if \( c \) is small, the density function is relatively flat indicating that the derivative is not well-determined. It therefore follows that \( c \) may be used as a design criterion, with the \( \{x_u\} \) and the \( \{t_u\} \) being chosen so as to maximize \( c \).

For nonlinear models there is the problem (which also occurs in parameter estimation work) that \( c \) also depends on the maximum likelihood estimate \( \hat{\theta} \) which is not known before the experimentation is performed. There is usually, however, some information available about \( \theta \), for example, from previous experimentation or from theoretical knowledge of the process under investigation. This information may be used to form a preliminary estimate to replace \( \hat{\theta} \) in (2.16).
Characteristic Design When p = 1

For a model with one parameter \( \theta \) there is a characteristic design which is appropriate for the estimation of any single derivative. As stated in the following result, this design is also appropriate for the estimation of the parameter \( \theta \). (Note that there may be more than one characteristic design.)

Result 1

For a model with one parameter \((p = 1)\), a design which maximizes \( c \) corresponding to one particular choice of a single derivative also maximizes \( c \) corresponding to any choice of a single derivative. This design may be obtained by maximizing \( |X'X| \) and is therefore also appropriate for the estimation of the parameter \( \theta \).

Since the proof of the more general Result 2 is also valid here, it is not necessary to give a separate proof for Result 1.

3. DISTINCT DERIVATIVES WHEN THE VARIANCE IS KNOWN

In this section the problems of experimental design and analysis will be considered for the case where the variance \( \sigma^2 \) of the observations is known and \( q \) distinct derivatives, \( d_{k_1} (\theta, \xi_{10}, t_{10}), d_{k_2} (\theta, \xi_{20}, t_{20}), \ldots, d_{k_q} (\theta, \xi_{q0}, t_{q0}) \),
where \(1 \leq q \leq p\), are to be estimated. These derivatives may all be of the same order \((k_1 = k_2 = \ldots = k_q)\) but evaluated at different values of \(\xi\) and \(t\), or they may all be of different orders but evaluated at the same values of \(\xi\) and \(t\).

\[(\xi_{10} = \xi_{20} = \ldots = \xi_{q0}, t_{10} = t_{20} = \ldots = t_{q0})\],

or they may be some other combination of orders and values of \(\xi\) and \(t\).

Whatever the combination of \(\{k_i\}, \{\xi_{10}\}, \text{and} \{t_{10}\}\), it is required that the derivatives be distinct in a sense which will be defined.

3.1 Analysis

As in the single derivative \((q = 1)\) case, it is also appropriate in the multiple distinct derivative \((1 \leq q \leq p)\) case to represent the derivative information in a set of data by a posterior density function. This density function may again be obtained from the density function for the model parameters.

The Posterior Density Function

\[
p(d_{k_1}, d_{k_2}, \ldots, d_{k_q} | \mathbf{y})
\]

The posterior density function \(p(d_{k_1}, d_{k_2}, \ldots, d_{k_q} | \mathbf{y})\) may be obtained by transforming the posterior density function \(p(\theta | \mathbf{y})\) and constructing the appropriate marginal posterior.
density function in a manner similar to that by which the single-derivative case was handled. Assuming that the \( q \) derivatives can each be adequately represented by an approximation of the type (2.5), a linear transformation may be used. Let

\[
\phi_i = d_{k_i} (\theta, \xi_i \theta, t_{i0}) \quad i = 1, 2, \ldots, q, \tag{3.1}
\]

\[
\hat{\phi}_i = d_{k_i} (\hat{\theta}, \xi_i \hat{\theta}, t_{i0}) \quad i = 1, 2, \ldots, q, \tag{3.2}
\]

and let \( b_1, b_2, \ldots, b_q \) be \( p \times 1 \) vectors with elements

\[
[b_j]_i = \left. \frac{\partial d_{k_j} (\theta, \xi_j \theta, t_{j0})}{\partial \theta_i} \right|_{\theta = \hat{\theta}}. \tag{3.3}
\]

The approximation (2.5) written in vector form for the \( q \) derivatives is then

\[
\phi_i = \hat{\phi}_i + b_i (\theta - \hat{\theta}) \quad i = 1, 2, \ldots, q \tag{3.4}
\]

and additional parameters

\[
\phi_i = \hat{\phi}_i + b_i' (\theta - \hat{\theta}) \quad i = q + 1, q + 2, \ldots, p \tag{3.5}
\]
may be defined to complete the transformation. Using an obvious notation, (3.4) and (3.5) may be written in matrix form, separately, as

\[ \phi_1 = \hat{\phi}_1 + B_1 (\hat{\theta} - \hat{\theta}) \]  

(3.6)

and

\[ \phi_2 = \hat{\phi}_2 + B_2 (\hat{\theta} - \hat{\theta}) \]  

(3.7)

and in a combined matrix form as

\[ \phi = \hat{\phi} + B (\hat{\theta} - \hat{\theta}) . \]  

(3.8)

A set of \( q \) distinct derivatives may now be defined as a set such that a nonsingular transformation of type (3.8) can be defined. Equivalently, a set of \( q \) distinct derivatives is a set such that the \( q \times p \) matrix \( B_1 \) with elements

\[ [B_1]_{ij} = \frac{\partial d_{ki}(\hat{\theta}, \hat{\xi}_i, \hat{\tau}_i)}{\partial \theta_j} \bigg|_{\theta = \hat{\theta}} \]  

(3.9)

has rank \( q \).
Using (2.2), (2.12), and (3.8) the posterior density function for $\phi$ can be shown to have the multivariate normal form

$$ p(\phi | y) = \left| \begin{array}{c} \sum_i^p \frac{1}{2} \mathbf{Z} Z^T (2\pi)^{-p/2} \sigma^{-p} \exp \left( -\frac{1}{2\sigma^2} (\phi - \hat{\phi})^T Z Z (\phi - \hat{\phi}) \right) \end{array} \right| $$

(3.10)

The posterior density function for the derivatives is then the marginal density function for $\phi_1$; that is, it has the multivariate normal form

$$ p(d_{k1}, d_{k2}, \ldots, d_{kq} | y) = p(\phi_1 | y) = \left| \begin{array}{c} \sum_i^q \frac{1}{2} \mathbf{D} (2\pi)^{-q/2} \sigma^{-q} \end{array} \right| \exp \left( -\frac{1}{2\sigma^2} (\phi_1 - \hat{\phi}_1)^T \mathbf{D} (\phi_1 - \hat{\phi}_1) \right) $$

(3.11)

where $\mathbf{D}$ is a $q \times q$ matrix defined by the relationship

$$ \mathbf{D}^{-1} = \Delta_q (\mathbf{Z} \mathbf{Z})^{-1} \Delta_q $$

(3.12)

and where $\Delta_q$ is a $q \times p$ matrix with the first $q$ elements in its main diagonal equal to 1 and all other elements equal to 0.

In other words, $\mathbf{D}^{-1}$ is the $q \times q$ upper left submatrix of $(\mathbf{Z} \mathbf{Z})^{-1}$.

The matrix $\mathbf{D}$ may also be expressed in the following form, which is more convenient for some purposes,

$$ \mathbf{D}^{-1} = \mathbf{B}_1 (\mathbf{X} \mathbf{X})^{-1} \mathbf{B}_1^T $$

(3.13)
From (3.11) it can be seen that \( \hat{\phi}_1 \) and \( D \) completely specify the posterior density function, so the problem of constructing this function may be viewed mathematically as the problem of determining \( \hat{\phi}_1 \) and \( D \).

3.2 Design

The Design Criterion \( c \)

For multiple distinct derivatives \( (1 < q < p) \) the shape of the posterior density function, aside from constants, depends only on the matrix \( D \). Since, for a given set of data, \( D \) is a function of the settings \( \{\xi_u\} \) and \( \{t_u\} \) at which the observations \( \{y_u\} \) are made, it follows that it is reasonable to use some function of \( D \) as an experimental design criterion. By proper choice of the \( \{\xi_u\} \) and the \( \{t_u\} \) a value of \( D \) will result which corresponds to a posterior density function which is in some sense optimal.

A function of \( D \) which may be used is the determinant

\[
c = |D|.
\]

(3.14)

It can be shown that if the \( \{\xi_u\} \) and the \( \{t_u\} \) are chosen so as to maximize \( c \) then a posterior density function will result with highest posterior density regions which have
smallest possible volumes. (For a discussion of highest posterior density regions, see Box and Tiao (1965).) A similar determinant criterion has previously been used in other contexts for purposes of parameter estimation. (See, for example, Wald (1943), Box and Lucas (1959), and Box and Hunter (1963).)

It should be noted that the criterion $c$ defined in (3.14) for the multiple derivative design problem, reduces, for $q=1$, to the $c$ defined in (2.16) for the single derivative design problem. Like the single derivative criterion, the multiple derivative criterion depends, in the case of nonlinear models, on the maximum likelihood estimate $\hat{\theta}$ which is not available before the experiments are performed. It is therefore necessary for nonlinear models to use some preliminary estimate for $\theta$ in place of $\hat{\theta}$.

**Characteristic Design When $q=p$**

For a model with $p$ parameters $\theta = (\theta_1, \theta_2, \ldots, \theta_p)'$, there is a characteristic design which is appropriate for the estimation of any set of $q=p$ distinct derivatives. As stated in the following result, which is a generalization of Result 1, this design is also appropriate for the estimation of $\theta$. (Note that there may be more than one characteristic design.)
Result 2

For a model with p parameters, a design which maximizes the c corresponding to one set of p distinct derivatives maximizes the c corresponding to any set of p distinct derivatives. This design may be obtained by maximizing $|X'X|$ and is therefore also appropriate for the estimation of $\theta$.

Proof:

When p distinct derivatives are to be estimated the criterion c may be written in the form

$$c = d^{-2} |X'X|$$  \hspace{1cm} (3.15)

where $d = |E_1|$ does not depend on the $\{e_i\}$ or the $\{t_u\}$.

4. NONDISTINCT DERIVATIVES WHEN THE VARIANCE IS KNOWN

In this section the problems of experimental design and analysis will be considered for the case where the variance $\sigma^2$ of the observations is known and q nondistinct derivatives are to be estimated. As implied in the last section, a set of q derivatives $d_{k1}(\theta,e_{10},t_1)$, $d_{k2}(\theta,e_{20},t_2)$, ..., $d_{kq}(\theta,e_{q0},t_q)$ is nondistinct if the q\times p matrix $B_1$ with elements
\[ [B_1]_{ij} = \frac{\partial d_{k_i}^{\theta', \xi_i, t_{i0}}}{\partial \theta_j} \]  

(4.1)

has rank less than q. It follows that a single derivative can never be nondistinct (except for the degenerate case where the derivative is functionally independent of the parameter) and that a set of q > p derivatives is always non-distinct.

4.1 Analysis

For the nondistinct case it is not possible to employ, unaltered, the transformation technique used previously to construct a joint posterior density function for all q derivatives since the transformation would be singular. With the transformation technique it is nevertheless possible to construct separate one-dimensional marginal posterior density functions for each of the q derivatives or joint posterior density functions for as many as \( q_1 \) distinct derivatives where \( q_1 < q \) is the size of the largest distinct subset of derivatives, \( q_1 \) will be at least 1 (except in degenerate cases) but not larger than p. In practice, degenerate cases cause no difficulty because the derivatives under consideration can be calculated exactly.
4.2 Design

When designing an experiment for the estimation of q nondistinct derivatives it is only necessary to consider one distinct subset of size $q_1$ and apply the design methods of subsections 2.2 and 3.2 to this subset. As shown in the following result, an appropriate design for any distinct subset of $q_1$ derivatives will then be obtained. This design is also appropriate for obtaining estimates of $q > q_1$ nondistinct derivatives.

Result 3

Let $q_1$ be the size of the largest distinct subset which is contained in a particular set of $q$ nondistinct derivatives. Let $\{d^k\}$ be one such distinct subset of size $q_1$. Then a design which is obtained by maximizing the criterion $c = |D|$ corresponding to the subset $\{d^k\}$ is also obtained by maximizing the criterion $c = |D|$ corresponding to any other distinct subset of size $q_1$.

Proof:

Let

$$
1/c_a = |B_{1a}(X'X)^{-1} B'_{1a}|
$$

(4.2)
define the criterion corresponding to one subset of $q_1$ distinct derivatives and let

$$1/c_\beta = |B_{\beta\beta} (X'X)^{-1} B'_{1\beta}|$$ \hspace{1cm} (4.3)$$

define the criterion corresponding to a second subset of $q_1$ distinct derivatives. There exists a $q_1 \times q_1$ nonsingular matrix $H$ whose elements do not depend on $\{e_u\}$ and $\{t_u\}$ such that

$$B_{1\alpha} = H B_{1\beta}.$$ \hspace{1cm} (4.4)$$

It then follows that

$$1/c_\alpha = |H|^2 (1/c_\beta)$$

It should be noted that the $c$ corresponding to one subset of size $q_1$ may have a simpler form than the $c$'s corresponding to other subsets of the same size. Therefore, in practice, it will sometimes be worthwhile to consider several different subsets of size $q_1$ in order to find one with which it is particularly convenient to work. Similar remarks apply to situations where Result 1 and Result 2 are appropriate.
5. ESTIMATION OF DERIVATIVES WHEN THE VARIANCE IS UNKNOWN

In this section the problems of experimental design and analysis will be considered for the case where the variance \( \sigma^2 \) of the observations is unknown. As will be seen, only slight modifications to the variance known procedures are needed.

5.1 Analysis

The Posterior Density Function \( p(\theta | x) \)

When the variance \( \sigma^2 \) of the observations is unknown it is possible to construct a posterior density function \( p(\theta | x) \) by "integrating out" \( \sigma \) in (2.2), the posterior density function for \( \theta \) when \( \sigma \) is known. Recall that in order to obtain that density function it was assumed that the errors \( \{ \varepsilon_u \} \) were from independent, normal \( (0, \sigma^2) \) distributions and that a locally uniform prior distribution for \( \theta \) was appropriate. If now it is further assumed that the prior distribution for \( \sigma \) is independent of the prior distribution for \( \theta \) and has the form

\[
p(\sigma) \propto 1/\sigma ,
\]  
(5.1)
the posterior density function for $\theta$ is

$$p(\theta | y) = \frac{\Gamma \left(\frac{p}{2}\right) \lvert x' x \rvert^{\frac{1}{2}}}{\Gamma \left(\frac{n-p}{2}\right) \pi^{\frac{p}{2}} [S(\hat{\theta})]^{\frac{p}{2}}} \left\{ 1 + \frac{(\theta - \hat{\theta}) x' x (\theta - \hat{\theta})}{S(\hat{\theta})} \right\}^{-\frac{n-p}{2}}$$

(5.2)

where $\hat{\theta}$ and $\bar{x}$ are as defined in subsection 2.1 and where

$$S(\hat{\theta}) = \sum_{u=1}^{n} [y_u - f(\hat{\theta}, \xi_u, t_u)]^2.$$ 

(5.3)

Note that the density function (5.2) is related to the multivariate $t$ distribution; specifically, $(n-p)^{1/2}[S(\hat{\theta})]^{-1/2}(\theta - \hat{\theta})$ has a multivariate $t$ distribution with $n-p$ degrees of freedom.

Note also that this density function is an approximation for nonlinear models in the same sense that (2.2) is an approximation for nonlinear models when the variance is known.

The Posterior Density Function For Derivatives

The technique for obtaining the posterior density function for derivatives from the posterior density function for the model parameters is the same when the variance is unknown as when the variance is known. In both cases appropriate new parameters $\phi = (\phi_1, \phi_2, \ldots, \phi_p)'$ are defined and $p(\phi | y)$ is obtained by a transformation from $p(\theta | y)$. 
For the single derivative situation if \( \phi \) and \( Z \) are as defined in subsection 2.1, it may be shown that

\[
p(\phi | y) = \frac{\Gamma \left( \frac{n}{2} \right) |Z|^2}{\Gamma \left( \frac{n-p}{2} \right) \pi^{n/2} [S(\hat{\phi})]^2} \left\{ 1 + \frac{(\phi - \hat{\phi})' Z Z (\phi - \hat{\phi})}{S(\hat{\phi})} \right\}^{-\frac{n}{2}}.
\]

(5.4)

The posterior density function for the derivative is then the marginal density function for \( \phi_1 \); that is,

\[
p(d_k(\hat{\theta}, \xi_0, t_0) | y) = p(\phi_1 | y)
\]

\[
= \frac{\Gamma \left( \frac{n+1-p}{2} \right) c^{\frac{1}{2}}}{\Gamma \left( \frac{n-p}{2} \right) \pi^{\frac{n}{2}} [S(\hat{\phi})]^\frac{1}{2}} \left\{ 1 + \frac{c (\phi_1 - \hat{\phi}_1)^2}{S(\hat{\phi})} \right\}^{-\frac{n+1-p}{2}}
\]

(5.5)

Note that this density function is completely specified once the three quantities, \( \hat{\phi}_1 \), \( c \), and \( S(\hat{\phi}) \) are determined.

The problem of constructing \( p(d_k(\theta, \xi_0, t_0) | y) \) from a set of data may then be viewed as the problem of determining \( \hat{\phi}_1 \), \( c \), and \( S(\hat{\phi}) \).

For the multiple distinct derivative situation if \( \phi \) is defined as in subsection 3.1, \( p(\phi | y) \) has the form (5.4) and the posterior density function for derivatives is the marginal density function for \( \phi_1 \); that is,
\[ p(d_{k_1}, d_{k_2}, \ldots, d_{k_q} | y) = p(\phi_1 | y) \]

\[ = \frac{r \left( \frac{n+q-p}{2} \right) \lvert D \rvert^{1/2}}{\frac{q}{2} \pi^{q/2} [S(\hat{\theta})]^q} \left[ 1 + \frac{(\phi_1 - \hat{\phi}_1)' D (\phi_1 - \hat{\phi}_1)}{S(\hat{\theta})} \right]^\frac{-n+q-p}{2} \]

(5.6)

This density function is completely specified by three quantities, \( \hat{\phi}_1, D, \) and \( S(\hat{\theta}) \).

Essentially, then, the only difference between the variance known and variance unknown situations is the form of the posterior density functions. In the variance known case the density function is normal while in the variance unknown case it is related to the \( t \) distribution. In both cases the same linear approximations are used in obtaining the density functions, and the concept of distinct and nondistinct derivatives does not change.

6.2 Design

For multiple distinct derivatives the shape of the posterior density function depends, aside from constants, only on the \( q \times q \) matrix

\[ G = D/S(\hat{\theta}), \]  

(5.7)
which, for a set of experimental data, is a function of the 
\{\xi_u\} and of the \{t_u\} at which the observations \{y_u\} are taken. 
It would seem appropriate, therefore, to use some function of 
\(G\) as a design criterion. (For \(q=1\) the expression (5.7) reduces to 
the quantity

\[ G = c/S(\hat{\theta}). \]  
(5.8)

By proper choice of the \{\xi_u\} and the \{t_u\} a value of \(G\) would 
result corresponding to a posterior density function which 
was in some sense optimal. An overall criterion which would, 
if maximized, result in H.P.D. regions of smallest possible 
volume is once again, as in the variance known cases, a de- 
terminant, this time the determinant \(|G|\).

For the variance unknown case, there is an added pro-
blem, however, in that \(G\) and therefore \(|G|\) depend not only 
on the \{\xi_u\} and on the \{t_u\} but also on \(\hat{\theta}\) and directly on the 
observations \{y_u\}. For linear models the dependence on \(\hat{\theta}\) and 
on the \{y_u\} is due only to \(S(\hat{\theta})\) while for nonlinear models the 
matrix \(D\) also depends on \(\hat{\theta}\). An estimate for \(\hat{\theta}\) could be ob-
tained as was previously done for the variance known case 
with nonlinear models but some way of handling the yet-to-be- 
oberved \{y_u\} must also be devised.
There is a simple method of handling this problem when p or more of the observations have already been made such that the resulting $\mathbf{X}$ matrix has rank p. Let $\mathbf{y}_1 = (y_1, y_2, \ldots, y_{n_1})'$ be an $n_1 \times 1$ vector of observations already taken and let $\mathbf{y}_2 = (y_{n_1+1}, y_{n_1+2}, \ldots, y_{n_1+n_2})'$ be an $n_2 \times 1$ vector of contemplated observations. As will be shown, it is possible to construct, with suitable assumptions, the posterior density function $p(\mathbf{y}_2 | \mathbf{y}_1)$ and to calculate the expected value

$$c_e = E \left| \mathbf{y}_2 \right|_{\mathbf{y}_2}$$

(5.9)

(The concept of a density function $p(\mathbf{y}_2 | \mathbf{y}_1)$ has been used previously by a number of authors. See, for example, Guttman (1967).) The expected value $c_e$ may then be used as a design criterion, and as shown in the following result this is equivalent to use of the criterion $c = |D|$. One consequence of this fact is that Results 1, 2, and 3 which are concerned with $c$ in the variance-known case also are valid in the variance-unknown case.

**Result 4**

If observations $\mathbf{y}_1$ are available and further observations $\mathbf{y}_2$ are contemplated, then use of the criterion
\[ c_e = E \left\{ |D/S(\hat{\theta})| \right\} \tag{5.10} \]

is equivalent to use of the criterion \( c = |D| \).

Proof:

Let \( X_1 \) be the \( n_1 \times p \) matrix corresponding to \( Y_1 \), let \( X_2 \) be the \( n_2 \times p \) matrix corresponding to \( Y_2 \), and let the elements of \( X_1 \) and \( X_2 \) be evaluated at the maximum likelihood estimate \( \hat{\theta}^{(1)} \) calculated using only \( Y_1 \). Further, let \( f_2 \) be the \( n_2 \times 1 \) vector with elements

\[ (f_2)_i = f(\hat{\theta}^{(1)}, \xi_{n_1 + i}, \xi_{n_1 + i}) \tag{5.11} \]

and let the \( n_2 \times n_2 \) matrix \( M \) be defined

\[ M = \left[ I_{n_2} + X_2 (X_1^{\prime} X_1)^{-1} X_1 \right] \tag{5.12} \]

If the prior distributions for \( \theta \) and \( 1/\sigma \) are locally uniform and independent, then

\[
p(Y_2|Y_1) = \frac{\Gamma \left( \frac{n-p}{2} \right) \left[ M^{-1} \right]^{\frac{1}{2}}}{n_2^{\frac{n_2}{2}}} \left\{ 1 + \frac{(Y_2 - f_2) (M^{-1} (Y_2 - f_2))}{S(\hat{\theta}^{(1)})} \right\}^{\frac{n-p}{2}} \tag{5.13}
\]
where \( n = n_1 + n_2 \). The sum of squares calculated using all \( n \) observations is

\[
S(\hat{\theta}) = S(\hat{\theta}^{(1)}) + (Y_2 - \hat{f}_2)' M^{-1} (Y_2 - \hat{f}_2).
\]  

(5.14)

Note that (5.13) and (5.14) are approximations for nonlinear models in the sense that it is necessary to assume that an adequate representation for the function \( f(\theta, \xi, t) \) is given by

\[
f(\theta, \xi, t) = f(\hat{\theta}^{(1)}, \xi, t) + \sum_{i=1}^{p} \frac{\partial^2 f(\theta, \xi, t)}{\partial \theta_i^2} \bigg|_{\theta=\hat{\theta}^{(1)}} (\theta_i - \hat{\theta}_i^{(1)}).
\]  

(5.15)

If, for nonlinear models, it is further assumed that the elements of \( D \) are evaluated at \( \theta = \hat{\theta}^{(1)} \), then it can be shown that

\[
E\{D/S(\hat{\theta})\} = \frac{\Gamma\left(\frac{n-P}{2}\right) \Gamma\left(\frac{n_1-p+2q}{2}\right)}{\Gamma\left(\frac{n_1-p}{2}\right) \Gamma\left(\frac{n-p+2q}{2}\right)} \frac{|D|}{[S(\hat{\theta}^{(1)})]^q}
\]  

(5.16)

6. DESIGN EXAMPLES

In this section the use of the design criterion \( c \) (see (2.16) and (3.14)) will be illustrated with examples.

Example 6.1

Consider the single-parameter linear model

\[
y = \theta t + \epsilon
\]  

(6.1)
where it is desired to estimate the first derivative at 
\[ t = t_0 \]

\[ d_1 (\theta , t_0) = \theta \]  \hspace{1cm} (6.2)

using a set of \( n \) data points. The \( n \times 1 \) matrix \( X \) has elements \( t_u \) and the \( 1 \times 1 \) vector \( b_1 \) has a single element equal to unity. Using (6.1), one obtains for the design criterion

\[ c = \sum_{u=1}^{n} t_u^2. \]  \hspace{1cm} (6.3)

If repeat points are allowed and if the restriction

\[ 0 \leq a \leq t_u \leq b \hspace{1cm} u = 1, 2, \ldots, n \]  \hspace{1cm} (6.4)

is imposed, the design which maximizes \( c \) is

\[ t_u = b \hspace{1cm} u = 1, 2, \ldots, n, \]  \hspace{1cm} (6.5)

that is, all observations to be taken at \( b \), the maximum allowable value of \( t \). This design illustrates an interesting point; it is not necessarily appropriate to take observations at more than one level of \( t \) when estimating a derivative with respect to \( t \). Recall, however, that we are assuming that
we know the correct form of the model. In practice the form of the model may be unknown, and in such cases this result would not hold.

Since \( p = 1 \), from Result 1 it is known that this design is also appropriate when the response itself (the zeroth order derivative) is to be estimated at any point \( t_0 \). This illustrates the fact that when estimating the response at a point \( t_0 \), it is not necessarily appropriate to take all (or, for that matter, any) observations at \( t = t_0 \).

**Example 6.2**

Consider the single-parameter, nonlinear, exponential model

\[
y = \exp(-\theta t) + \epsilon
\]  

(6.6)

where it is desired to estimate the first derivative at \( t = t_0 \)

\[
d_1(\theta, t_0) = -\theta e^{-\theta t_0}
\]  

(6.7)

using a set of \( n \) data points. This model was previously discussed in subsection 2.1, and \( X \) and \( b_1 \) are shown in (2.25) and (2.26). The design criterion is
\[ c = \frac{\sum_{u=1}^{n} t_u^2 \exp\{-2\hat{\theta}t_u\}}{(-1+\hat{\theta}t)^2 \exp\{-2\hat{\theta}t_o\}} \]  
(6.8)

and the design which maximizes \( c \) is

\[ t_u = \frac{1}{\hat{\theta}} \quad u = 1, 2, \ldots, n. \]  
(6.9)

As expected, the criterion and design for this nonlinear model depend on the unknown (before experimentation) value of \( \hat{\theta} \). Thus, some estimate of \( \theta \) must be made, based on prior knowledge or previous experimentation, before the design can be determined. (At \( t = 1/\hat{\theta} \) the criterion \( c \) is undefined.

This corresponds to the fact that the linear approximation (2.5) for \( d_1(\theta, t_o) \) is not adequate when \( t_o = 1/\hat{\theta} \). At this value of \( t_o \), (2.5) implies that \( d_1(\theta, t_o) \) is equal to a constant and does not depend on \( \theta \).

**Example 6.3**

Consider the two-parameter nonlinear model

\[ y = \theta_1 [1 - \exp\{-\theta_2 t\}] + c \]  
(6.10)
where it is desired to estimate the first derivative

\[ d_1(\theta, t_o) = \theta_1 \theta_2 \exp(-\theta_2 t_o) \]  \hspace{1cm} (6.11)

and the second derivative

\[ d_2(\theta, t_o) = -\theta_1 \theta_2^2 \exp(-\theta_2 t_o) \]  \hspace{1cm} (6.12)

at a point \( t = t_o \). Since this is a two-parameter model, it follows from Result 2 that an appropriate design for these two derivatives (or for any two distinct derivatives) may be obtained by maximizing the criterion

\[ c_{12} = |X'X|. \]  \hspace{1cm} (6.13)

For this example the \( n \times 2 \) matrix \( X \) has the form

\[
X = \begin{bmatrix}
[1 - \exp(-\theta_2 t_1)] & \hat{\theta}_1 t_1 \exp(-\theta_2 t_1) \\
[1 - \exp(-\theta_2 t_2)] & \hat{\theta}_1 t_2 \exp(-\theta_2 t_2) \\
\vdots & \vdots \\
[1 - \exp(-\theta_2 t_n)] & \hat{\theta}_1 t_n \exp(-\theta_2 t_n)
\end{bmatrix}
\]  \hspace{1cm} (6.14)
and the resulting $c_{12}$ is not a simple function of the $\{t_u\}$. Designs were therefore obtained with the aid of a digital computer.

Table 1 shows a 20-point design which was generated in a sequential fashion on a computer. The initial design points, $t_1$ and $t_2$, were obtained by making a two dimensional scan of the region ($t_2 < t_1 \leq 250$, $0 \leq t_2 < t_1$) to determine $t_1$ and $t_2$ (to the nearest integer) which maximize $c_{12}$ assuming that $\hat{\theta}_1 = 10.0$ and $\hat{\theta}_2 = 0.01$. Additional points $t_3$, $t_4$, ..., $t_{20}$ were then generated sequentially one-at-a-time by scanning the interval ($0 \leq t \leq 250$) to determine $t$ which maximizes $c_{12}$ using the same values of $\hat{\theta}_1$ and $\hat{\theta}_2$.

Note that using the same value of $\hat{\theta}_1$ and $\hat{\theta}_2$ at each stage is equivalent to assuming that the errors $\{e_u\}$ are zero. (When the errors $\{e_u\}$ are zero the model fits the data exactly, and the parameter estimates do not change at each stage.)

Additional designs were generated in the same manner using the criterion $c_1$ appropriate when only the first derivative is to be estimated and using the criterion $c_2$ appropriate when only the second derivative is to be estimated. These designs are shown in Tables 1, 2, and 3 for $t_0 = 0$, $t_0 = 60$, and $t_0 = 100$. Note that the design generated using $c_{12}$ is the same for all values of $t_0$. 
As figures of merit, $1/c_1$, $1/c_2$, and $1/c_{12}$ are shown for each of the designs. Using these figures of merit for comparison it can be seen that the design for estimating two distinct derivatives is fairly good when only a single derivative is to be estimated.

From Table 1 and Table 3 it may be seen that the sequential design for the estimation of $d_1(\theta,0)$ is identical to the design for $d_2(\theta,100)$. This may be explained by noting that $1/c_1$ may be written

$$\frac{1}{c_1} = e^{-2\hat{\theta}_2 t_o} \{ \hat{\theta}_2^2 v^{11} + 2\hat{\theta}_1 \hat{\theta}_2 (1-\hat{\theta}_2 t_o)v^{12} + \hat{\theta}_1^2 (1-\hat{\theta}_2 t_o)^2 v^{22} \}$$

(6.15)

where $v^{11}$, $v^{12}$, and $v^{22}$ are elements of the symmetric matrix $(X'X)^{-1}$ and that $1/c_2$ may be written

$$\frac{1}{c_2} = \hat{\theta}_2^2 e^{-2\hat{\theta}_2 t_o} \{ \hat{\theta}_2^2 v^{11} + 2\hat{\theta}_1 \hat{\theta}_2 (2-\hat{\theta}_2 t_o)v^{12} + \hat{\theta}_1^2 (2-\hat{\theta}_2 t_o)^2 v^{22} \}.$$ 

(6.16)

The term $(1-\hat{\theta}_2 t_o)$ plays the same role in (6.15) that $(2-\hat{\theta}_2 t_o)$ does in (6.16). It may be seen that $1/c_1$ at $t_o = s$ is
minimized by the same \( \{ t_u \} \) which minimize \( 1/c_2 \) at \( t_o = s + 1/\hat{\theta}_2 \).

In this example \( \hat{\theta}_2 = .01 \) is used and a design appropriate
for \( d_1(\theta, t_o) \) is also appropriate for \( d_2(\theta, t_o + 100) \). This
is an illustration of the point that when an adequate model
is known there is no fundamental difference between estimating
a derivative of one order and a derivative of another order.
References


Table 1

Sequential Designs At $t_0 = 0$

<table>
<thead>
<tr>
<th>Experiment Number</th>
<th>Design For Two Derivatives</th>
<th>Design For $d_1(\theta, 0)$</th>
<th>Design For $d_2(\theta, 0)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>250</td>
<td>250</td>
<td>250</td>
</tr>
<tr>
<td>2</td>
<td>78</td>
<td>72</td>
<td>69</td>
</tr>
<tr>
<td>3</td>
<td>250</td>
<td>69</td>
<td>64</td>
</tr>
<tr>
<td>4</td>
<td>78</td>
<td>66</td>
<td>60</td>
</tr>
<tr>
<td>5</td>
<td>78</td>
<td>63</td>
<td>57</td>
</tr>
<tr>
<td>6</td>
<td>250</td>
<td>61</td>
<td>250</td>
</tr>
<tr>
<td>7</td>
<td>78</td>
<td>59</td>
<td>62</td>
</tr>
<tr>
<td>8</td>
<td>250</td>
<td>58</td>
<td>60</td>
</tr>
<tr>
<td>9</td>
<td>250</td>
<td>56</td>
<td>58</td>
</tr>
<tr>
<td>10</td>
<td>78</td>
<td>250</td>
<td>250</td>
</tr>
<tr>
<td>11</td>
<td>250</td>
<td>63</td>
<td>61</td>
</tr>
<tr>
<td>12</td>
<td>78</td>
<td>62</td>
<td>60</td>
</tr>
<tr>
<td>13</td>
<td>78</td>
<td>61</td>
<td>59</td>
</tr>
<tr>
<td>14</td>
<td>250</td>
<td>60</td>
<td>250</td>
</tr>
<tr>
<td>15</td>
<td>78</td>
<td>59</td>
<td>61</td>
</tr>
<tr>
<td>16</td>
<td>250</td>
<td>58</td>
<td>60</td>
</tr>
<tr>
<td>17</td>
<td>78</td>
<td>250</td>
<td>59</td>
</tr>
<tr>
<td>18</td>
<td>250</td>
<td>62</td>
<td>250</td>
</tr>
<tr>
<td>19</td>
<td>78</td>
<td>61</td>
<td>61</td>
</tr>
<tr>
<td>20</td>
<td>250</td>
<td>61</td>
<td>60</td>
</tr>
</tbody>
</table>

| $1/c_1$           | $1.15 \times 10^{-4}$       | $7.96 \times 10^{-5}$       | $8.37 \times 10^{-5}$       |
| $1/c_2$           | $6.76 \times 10^{-8}$       | $6.12 \times 10^{-8}$       | $5.54 \times 10^{-8}$       |
| $1/c_{12}$        | $2.12 \times 10^{-7}$       | $4.41 \times 10^{-7}$       | $3.02 \times 10^{-7}$       |
Table 2
Sequential Designs At $t_o = 60$

<table>
<thead>
<tr>
<th>Experiment Number</th>
<th>Design For Two Derivatives</th>
<th>Design For $d_1(\theta,60)$</th>
<th>Design For $d_2(\theta,60)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>250</td>
<td>167</td>
<td>250</td>
</tr>
<tr>
<td>2</td>
<td>78</td>
<td>157</td>
<td>70</td>
</tr>
<tr>
<td>3</td>
<td>250</td>
<td>162</td>
<td>66</td>
</tr>
<tr>
<td>4</td>
<td>78</td>
<td>162</td>
<td>62</td>
</tr>
<tr>
<td>5</td>
<td>78</td>
<td>161</td>
<td>59</td>
</tr>
<tr>
<td>6</td>
<td>250</td>
<td>162</td>
<td>57</td>
</tr>
<tr>
<td>7</td>
<td>78</td>
<td>162</td>
<td>250</td>
</tr>
<tr>
<td>8</td>
<td>250</td>
<td>162</td>
<td>62</td>
</tr>
<tr>
<td>9</td>
<td>250</td>
<td>162</td>
<td>61</td>
</tr>
<tr>
<td>10</td>
<td>78</td>
<td>162</td>
<td>59</td>
</tr>
<tr>
<td>11</td>
<td>250</td>
<td>162</td>
<td>58</td>
</tr>
<tr>
<td>12</td>
<td>78</td>
<td>162</td>
<td>250</td>
</tr>
<tr>
<td>13</td>
<td>78</td>
<td>161</td>
<td>61</td>
</tr>
<tr>
<td>14</td>
<td>250</td>
<td>162</td>
<td>60</td>
</tr>
<tr>
<td>15</td>
<td>78</td>
<td>162</td>
<td>59</td>
</tr>
<tr>
<td>16</td>
<td>250</td>
<td>162</td>
<td>250</td>
</tr>
<tr>
<td>17</td>
<td>78</td>
<td>162</td>
<td>62</td>
</tr>
<tr>
<td>18</td>
<td>250</td>
<td>162</td>
<td>61</td>
</tr>
<tr>
<td>19</td>
<td>78</td>
<td>162</td>
<td>60</td>
</tr>
<tr>
<td>20</td>
<td>250</td>
<td>162</td>
<td>59</td>
</tr>
</tbody>
</table>

| $1/c_1$           | $2.95 \times 10^{-6}$       | $2.34 \times 10^{-6}$       | $4.87 \times 10^{-6}$       |
| $1/c_2$           | $8.48 \times 10^{-9}$       | $1.43 \times 10^{-5}$       | $6.58 \times 10^{-9}$       |
| $1/c_{12}$        | $2.12 \times 10^{-7}$       | $3.69 \times 10^{-4}$       | $3.53 \times 10^{-7}$       |
Table 3

Sequential Designs At $t_0 = 100$

<table>
<thead>
<tr>
<th>Experiment Number</th>
<th>Design For Two Derivatives</th>
<th>Design For $d_1(\theta, 100)$</th>
<th>Design For $d_2(\theta, 100)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>250</td>
<td>250</td>
<td>250</td>
</tr>
<tr>
<td>2</td>
<td>78</td>
<td>56</td>
<td>72</td>
</tr>
<tr>
<td>3</td>
<td>250</td>
<td>250</td>
<td>69</td>
</tr>
<tr>
<td>4</td>
<td>78</td>
<td>56</td>
<td>66</td>
</tr>
<tr>
<td>5</td>
<td>78</td>
<td>250</td>
<td>63</td>
</tr>
<tr>
<td>6</td>
<td>250</td>
<td>250</td>
<td>61</td>
</tr>
<tr>
<td>7</td>
<td>78</td>
<td>59</td>
<td>59</td>
</tr>
<tr>
<td>8</td>
<td>250</td>
<td>58</td>
<td>58</td>
</tr>
<tr>
<td>9</td>
<td>250</td>
<td>60</td>
<td>62</td>
</tr>
<tr>
<td>10</td>
<td>78</td>
<td>250</td>
<td>63</td>
</tr>
<tr>
<td>11</td>
<td>250</td>
<td>250</td>
<td>60</td>
</tr>
<tr>
<td>12</td>
<td>78</td>
<td>60</td>
<td>59</td>
</tr>
<tr>
<td>13</td>
<td>78</td>
<td>250</td>
<td>61</td>
</tr>
<tr>
<td>14</td>
<td>250</td>
<td>250</td>
<td>60</td>
</tr>
<tr>
<td>15</td>
<td>78</td>
<td>60</td>
<td>58</td>
</tr>
<tr>
<td>16</td>
<td>250</td>
<td>60</td>
<td>250</td>
</tr>
<tr>
<td>17</td>
<td>78</td>
<td>250</td>
<td>62</td>
</tr>
<tr>
<td>18</td>
<td>250</td>
<td>60</td>
<td>61</td>
</tr>
<tr>
<td>19</td>
<td>78</td>
<td>250</td>
<td>61</td>
</tr>
<tr>
<td>20</td>
<td>250</td>
<td>60</td>
<td>61</td>
</tr>
</tbody>
</table>

$$\frac{1}{c_1} = 4.88 \times 10^{-6}$$

$$\frac{1}{c_2} = 1.56 \times 10^{-9}$$

$$\frac{1}{c_{12}} = 2.12 \times 10^{-7}$$