ProQuest LLC.
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106 – 1346

STATISTICAL MODELS FOR BIOLOGICAL GROWTH

G W Morgan

The problems of selecting and fitting models to data collected on the growth of an organism or population are considered.

The ways of identifying suitable models are reviewed and some suggestions made using smoothe d growth rates.

The problems in fitting non-linear models to growth data are considered. Use is made of measures of non-linearity and new methods developed to take into account the effects of non-linearity and also the possible presence of auto correlated errors.

Stochastic dynamic models for growth are examined and parameter estimators found. Extensive use is made of the Kalman filter.

For multivariate situations two methods by which phase plane solutions can be fitted in the bivariate case are developed. The use of the Kalman filter is also considered.

The study also considers the role of statistical modelling. A possible methodology is presented.

Examples from two areas are presented:

plant growth and child growth.

STATISTICAL MODELS FOR BIOLOGICAL GROWTH CURVES

GEOFFREY W MORGAN BSc MSc FSS

A thesis submitted to the Council for National Academic Awards
in partial fulfilment of the requirements for the degree of
Doctor of Philosophy

Sponsoring Establishment : Department of Applied Statistics
and Operational Research
Sheffield City Polytechnic

Collaborating Establishment : National Vegetable Research
Station

January 1987

ACKNOWLEDGEMENTS

I would like to thank Dr G K Kanji, my supervisor and acting Head of Department, for all his help during this study.

I would also like to thank Dr G Freeman and Prof M Preece for their helpful discussions and for providing data.

My thanks also go to Maggie Harper for typing this thesis and coping with my hand writing.

Finally I would like to express my gratitude to my wife, Margaret, for her support and also my children Rachel and Paul for their patience when 'Daddy's working'.

ABSTRACT

STATISTICAL MODELS FOR BIOLOGICAL GROWTH

G W Morgan

The problems of selecting and fitting models to data collected on the growth of an organism or population are considered.

The ways of identifying suitable models are reviewed and some suggestions made using smoothe d growth rates.

The problems in fitting non-linear models to growth data are considered. Use is made of measures of non-linearity and new methods developed to take into account the effects of non-linearity and also the possible presence of auto correlated errors.

Stochastic dynamic models for growth are examined and parameter estimators found. Extensive use is made of the Kalman filter.

For multivariate situations two methods by which phase plane solutions can be fitted in the bivariate case are developed. The use of the Kalman filter is also considered.

The study also considers the role of statistical modelling. A possible methodology is presented.

Examples from two areas are presented:

plant growth and child growth.

CONTENTS

1. Introduction

1.1 Growth Data

Data on the growth of part, whole or groups of organisms are collected in many areas of Biological Science. In agriculture the growth of crops and animals is studied, in Medicine it is the growth of individuals or the growth of tumors. In Botany, either the whole plant can be looked at or one can look at a particular part, for example a leaf or a tiller.

The data collection can be classified into two types. The first is when a single entity is studied through time and measurements are taken at various times. The second is when a sample is taken from an experimental population at various times in order to study the growth of the population as a whole. This is common in the study of crop growth when, for example, in studying grass growth over a season a sample of grass given by a randomly placed quadrate would be harvested and weighed every week.

One of the particular difficulties with this type of data, especially with the first type of collection, is that the observations will, in general, not be independent but correlated. Glasbey(1979) considers five contributions to the errors of a model of the growth of Ayrshire steer calves

(a) Variations in gut fill between weighings

(b) Seasonal variations and changes in diet

(c) Illness

(d) Errors in measuring procedure

(e) Choice of wrong parametric form of the curve.

Of these factors (b), (c) and (e) result in correlated errors.

In the case of sampling from an experimental population changes in the environment which would be common to the samples would also lead to the correlated error structure. Thus while standard regression modelling procedure involve the assumption of independent errors this will rarely be valid in the case of growth data.

Two detailed examples of the type of data that will be dealt with will be given later.

A second problem is that, over a reasonable time span, the data is not linear. The standard shape being that of a sigmoid curve, thus the most natural model form is a non-linear model.

1.2 Some Approaches to the Problem

There are three basic approaches to this problem. One is to consider functions of the observations, usually some form of difference, and assume that these are, approximately, independent. Some form of analysis is then carried out on these functions of the data (eg Radford (1967) Hunt (1978, 1982)).

Alternatively a second approach, which will be called the statistical approach, relies on approximating the growth curve by a low order polynomial. The advantage of using a polynomial is that it is statistically simpler to deal with and a procedure involving a general error structure can be arrived at which only involves linear computations. However, it is rarely possible to assign a biological meaning to the parameters of the fitted model, thus making interpretation difficult.

The third approach is to fit non-linear models whose parameters have a reasonable biological interpretation. Being non-linear these models involve greater problems in estimating the parameters than do the linear models. Numerical techniques have to be employed in order to minimise the resulting non-linear sums of squares function required by least squares procedures. Also very little work has been done on fitting non-linear models with dependent errors.

These two other approaches will now be reviewed in detail starting with the statistical approach, then looking at non-linear curves.

1.3 The Statistical Approach

One popular approach is to use a split-plot type analysis. Suppose there are r individuals on each of p treatments with observations taken at ν times, a possible model would be

$$Y_{ijk} = \mu + \alpha_i + \eta_{ij} + t_k + (\alpha t)_{ik} + e_{ijk}$$

$$i=1,\ldots,p, \quad i,j=1,\ldots,r, \quad k=1,\ldots,\nu$$

α_i - treatment effect, $\Sigma\, \alpha_i = 0$

t_k - time effect, $\Sigma\, t_k = 0$

$(\alpha t)_{ik}$ - time x treatment interaction

$$\sum_i (\alpha t)_{ik} = \sum_k (\alpha t)_{ik} = 0$$

η_{ij} - random individual effect

e_{ijk} - error term independent of η_{ij}

with

$$Var(\eta_{ij}) = \sigma_I^2 \quad \text{and} \quad Var(e_{ijk}) = \sigma_{II}^2$$

Now

$$Var(Y_{ijk}) = \sigma_I^2 + \sigma_{II}^2$$

$$Cov(y_{ijk}, y_{ijk'}) = \sigma_I^2$$

Let

$$\sigma^2 = \sigma_I^2 + \sigma_{II}^2 \text{ and } \rho = \frac{\sigma_I^2}{\sigma_I^2 + \sigma_{II}^2}$$

then

$$Var(y_{ijk}) = \sigma^2$$

$$Cov(y_{ijk}) = \rho\sigma^2$$

So the variance - covariance matrix V_i for the observations of an individual is $\sigma^2(I(1 - \rho) + J_\rho)$ ie

$$V_i = \begin{bmatrix} 1 & \rho & & \rho \\ \rho & 1 & & \\ & & \rho & \\ & & 1 & \\ \rho & & \rho & \end{bmatrix} \sigma^2$$

Rowell and Walters (1976) are critical of this approach. One of the main criticisms is that the covariance matrix is of a definite and, possibly, unlikely form since it implies that the correlation between two observations S time units apart is the same as that just one apart. In order to test this assumption, Wilks test is proposed (Wilks(1946)), see also Winer (1970) who uses a method based on Box (1950). Huynh and Feldt (1970) show that the F test is valid if V_i is of the form

$$[I + a J + \alpha 1' + 1 \alpha']$$

where a is a scalar, 1 is a column of 1's and α is such that $\alpha' 1 = 0$.

The analysis of variance is

Source	df	EMS
Treatment	p-1	$\sigma_{II}^2 + t\sigma_I^2 + rt(\Sigma \alpha^2/p-1)$
Error I	p(r-1)	$\sigma_{II}^2 + t\sigma_I^2$
Time	t-1	$\sigma_{II}^2 + pr(\Sigma t^2/t-1)$
Time x Treatment	(t-1)(p-1)	$\sigma_{II}^2 + r(\Sigma\Sigma(\alpha t)^2/(t-1)(p-1))$
Error II	p(t-1)(r-1)	σ_{II}^2
Total	tpr - 1	

(See for example Cochran and Cox or Winer).

If a block structure is introduced then the following partition is used

Source	df
Block	r-1
Treatment	p-1
Error I	(p-1)(r-1)
Time	t-1
Time x Treatment	(t-1)(p-1)
Error II	p(t-1)(r-1)
Total	tpr-1

It has also been suggested that the Error II be partitioned into

Time x block - (t-1)(r-1) df

and

Time x Treatment x Block (t-1)(p-1)(r-1) df

and use the 3-factor component to test the Time x Treatment interaction. This procedure is sometimes referred to as a split-block analysis. It gives the same test as would result from a blocks x treatments x times factorial experiment using significance tests appropriate to a mixed model in which blocks were

regarded as a random effect and treatment and times as fixed effects.

The (t-1) df for time can be further partitioned into single degree components corresponding to orthogonal polynomials, ie linear, quadratic terms etc. Also the Time x Treatment interaction can be partitioned into linear x treatment, quadratic x treatment etc components.

Another approach to the analysis of this model is that of Fuller and Battese (1973) which is discussed by Grassia and DeBoer (1980). In this approach a transformation I is found such that $\mathbf{e}^* = \mathbf{Te}$ are independent errors and the regression of TY on TX is then performed. For this problem T is of the form

$$\begin{bmatrix} T_1 & & 0 \\ & T_2 & \\ 0 & & T_{FP} \end{bmatrix}$$

with

$$T_i \underset{(txt)}{=} \begin{bmatrix} 1 & & -\alpha/t \\ & 1 \quad 0 & -\alpha/t \\ 0 & 1 & -\alpha/t \end{bmatrix} \text{ where } \alpha = 1 - [\sigma_{II}^2/\sigma_{II}^2 + t\sigma_I^2]^{\frac{1}{2}}$$

Estimating σ_{II}^2 and σ_I^2 by

$$\hat{\sigma}_{II}^2 = \text{Error II Mean square} = S_{II}^2$$

$$\hat{\sigma}_I^2 = \frac{S_I^2 - S_{II}^2}{t} \text{ where } S_I^2 = \text{error I mean square}$$

or more appropriately for this method using the method of 'fitting of constant' (see for example Searle).

To estimate σ_{II}^2 fit the full model treating η_{ij} as a fixed effect to yield a residual sum of square SS_{fr}

Then $\hat{\sigma}_{II}^2 = SS_{fr}/(p(t-1)(r-1))$

To estimate σ_I^2 regress y on all the fixed effects and obtain the residual sum of squares SS_f then

$$\hat{\sigma}_I^2 = \frac{SS_f - tp(F-1)\hat{\sigma}_{II}^2}{n^*}$$

where

$$n^* = tpr - tr\left[(X'X^{-1} \sum^{rp} (\bar{x}_i' \bar{x}_i)(rp)^2\right]$$

where $\hat{x}_i$ is the (lxp) vector having kth element

the mean of x_k for the ith individuals

= t in this case

The advantage of the second method of estimating σ_{II}^2 and σ_I^2 is that it can easily be used in an unbalanced situation.

A second approach was suggested by Wishart (1938). He fitted linear and quadratic orthogonal polynomials to the response over time of each individual. The estimated coefficients were then analysed using the analysis of variance technique. This gave an analysis of linear growth and an analysis of rate of change of growth.

Hills (1968) used first and second order finite differences rather than orthogonal polynomials. Rao (1958) considered the first finite difference but used a transformed time scale to improve the assumption of linear growth.

The two above mentioned methods involve a univariate approach, the use of multivariate techniques has also been suggested (Box (1950), Cole and Grizzle (1966)). Instead of treating time as a factor treat the observations on an individual over time as a single multivariate observation with an arbitrary variance - covariance matrix. The model is

$$\underset{nxt}{E(Y)} = \underset{nxm}{A}\ \underset{mxt}{\xi}$$

where

$$\frac{y_i}{(1xt)}$$

has a multivariate normal distribution with variance - covariance matrix Σ and rank (A) = m where m is the number of independent terms in the model.

To test the hypothesis H_0 $C\xi M = 0$

C is sxm ($s < m$) and of rank s

M is txu and rank $u < t$

Test statistics are based on the matrix ratio $S_H S_E^{-1}$ where

$S_H = M'YA(A'A)^{-1}C'(C(A'A)^{-1}C')^{-1}C(A'A)^{-1}A'YM$

$S_E = M'Y'[I - A(A'A)^{-1}A')YM$

Three statistics have been proposed

Wilks likelihood ratio $\Lambda = 1/|S_H S_E^{-1} + I|$

Product of non-zero characteristic roots $tr(S_H S_E^{-1})$

and

$$\theta_S = \frac{\lambda_m}{1 + \lambda_m} \text{ where } \lambda_m = \text{largest characteristic root of } S_H S_E^{-1}$$

For Λ tables are available for some cases. Also $-(n-\frac{1}{2}(t+S+1)\log\Lambda \sim \chi^2_{tS}$ asymptotically.

For suitable choice of **C** and **M** the required test of hypotheses can be carried out. Test for the form of the response over time can be constructed using orthogonal polynomials (see Cole and Grizzle (1966)).

In a recent simulation investigation Schwertman et al (1985) compare split plot methods with the difference methods of Hills and Rao and a multivariate test

using Hotelling's T square. They concluded that the split plot approach was best for testing parallelism of growth when Geisser and Greenhouse's (1958) method of adjusting the degrees of freedom is used.

A major advance in the analysis of growth curves come with the Potthoff and Roy (1964) paper. They introduced the idea of an extended MANOVA model

$$E(Y) = A\xi P$$

where **P** is a (qxt) matrix of known coefficients, and ξ is now (mxq). The **P** matrix allows a linear growth curve to be fitted, eg the ith row may be

$$1 \quad t_i \quad t_i^2 \quad t_i^3 \quad \ldots$$

The analysis of the model is carried out by the use of a transformation of the form

$$Z = Y\, G^{-1}P'(P\, G^{-1}P')^{-1}$$

where **G** is any symmetric positive definite matrix. This gives

$$E(Z) = A\xi$$

$$Cov(Z) = [PG^{-1}P']^{-1}PG^{-1}\Sigma G^{-1}P'(PG^{-1}P')^{-1}$$

(G may also be any non-singular matrix such that $PG^{-1}P$ is of full rank, this will slightly alter the var-cov matrix above). The transformed model can now be analysed as an ordinary MANOVA.

There are several possible choices for G

(i) G = I. The main advantage of this is the computational simplicity especially if orthogonal polynomials are used to construct **P**.

(ii) Among all estimators of the form

$$\underset{(1xn)}{d'} \quad \underset{}{Y} \quad \underset{(qx1)}{W}$$

$$b'C(A'A)^{-1}A'Y\Sigma^{-1}P'(P\Sigma^{-1}P')^{-1}Mf$$

is the minimum unbiased estimator of $b'C\xi Mf$. Comparing this with the LS estimator from the transformed model

$$(A'A)^{-1}A'Z = (A'A)^{-1}A'YG^{-1}P'(PG^{-1}P')^{-1}$$

suggests taking $G = \Sigma$. However, in general the value of Σ will not be known.

(iii) If a stochastically independent estimate of Σ is available, S_I say, then take $G = S_I$ using the argument of (ii).

Rao (1965) discusses another extension of the MANOVA model ξ is considered to be random with

$E(\xi) = X$

$Cov(\xi) = \Lambda$

Thus

$Cov(Y) = (A\Lambda A' + \Sigma)$

He suggests the following steps in the analysis of growth data

1. Replace observed values by orthogonal polynomial regression coefficients (b_j, j=0,...,t-1) for each individual. (cf Wisharts approach). These estimate the true coefficients β_O β_R.

2. Calculate the mean and corrected sums of squares and cross-products for the above coefficients, $\bar{b}$ and S say.

3. Examine to see if a subset of the coefficients is inference sufficient. This will depend on the degree of the polynomial of the growth curve and on the structure of the true dispersion matrix of $b_0,...,b_{t-1}$.

Suppose the polynomial with coefficients $b_0,...,b_k$ adequately represents the curve, then a subset of the remaining coefficient $b_{k+1},...,b_{t-1}$ whose expectation is zero may yield information of $\beta_0,...,\beta_k$ through their correlations with $b_0,...,b_k$. Improved estimates of $\beta_0,...,\beta_k$ can be obtained by using $b_{k+1},...,b_k$ as concomitant variables to the main coefficients $b_0,...,b_k$ in a covariance analysis.

In order to see if any of the $b_{k+1},...,b_{t-1}$ are required the form of the dispersion matrix Λ can be tested

$$\Lambda = \begin{bmatrix} \Lambda_k & 0 \\ 0 & I\alpha^2 \end{bmatrix}$$

where Λ_k is $k+1$ x $k+1$ matrix. This may be carried out using

$$\Lambda = \frac{|S|}{|S_{k+1}||S_2|}$$

where

$$S = \begin{bmatrix} S_1 & & S_{12} \\ & & \\ S_{12} & & S_2 \end{bmatrix} \begin{matrix} 0 \\ k \\ k+1 \\ t-1 \end{matrix}$$

$$0 \ldots k \; k+1 \ldots t-1$$

and

$$\Lambda_2 = \frac{|S_2|}{a_{k+1}a_{k+2},\ldots,a_{t-1}S^{p-k-1}}$$

where a_i = sum of squares of values of orthogonal polynomials

$$S = \left[\sum_{i=k+1}^{t-1} S_{ii}/d_i \right] / t-k-1 \text{ with } S = [S_{ii}]$$

Then $-(n-1)\log\Lambda$ is approximately χ^2 on $(t-k-1)(k+1)$ df and $-(n-1)\log\Lambda_2$ is approximately χ^2 with $\frac{1}{2}(t-k-2)(t-k+1)$ df. The covariance approach is as follows. Let U be a nxq matrix of concomitant variables

$$E(u) = 0$$

$$E(Y|u) = A\xi + u\rho$$

Then

$$\begin{bmatrix}\hat{\xi}\\ \hat{\rho}\end{bmatrix} = \begin{bmatrix}A'A & A'u\\ u'A & u'u\end{bmatrix}^{-1}\begin{bmatrix}A'\\ u'\end{bmatrix} Y = C.g\ Y$$

and

$$\hat{\Lambda} = (n-r-q)^{-1}(Y'Y - YgCg'Y)$$

the

$\text{cov}(P'\xi)$ for some P is given by

$$\text{Cov}(P'\xi) = (P'C_{mm}P)\hat{\Lambda}$$

C_{mm} being the first m rows and columns of C.

In this paper Rao is also critical of Potthoff and Roy's method. He criticizes it for ignoring information provided by concomitants and for the arbitrary choice of G if G does not equal Σ.

Katri (1966) showed that the maximum likelihood estimate of ξ for Potthoff and Roy's Model was obtained when G was replaced by S calculated from the same data. He also shows that the three test criteria given above can be used. Grizzle and Allen (1969) developed Rao's technique giving examples. They also point out that if G = S then the Potthoff and Roy method is equivalent to using all the possible concomitant variables in Rao's method.

The relationship between the two methods was further examined by Lee (1974) and Bakeslary, Corsten and Kala (1978). The latter shows that for any choice of a subset of concomitant variables there exists a value of G such that the two methods are equivalent. However, they recommend the use of Rao's method when no information about the structure of Σ is known and Potthoff and Roy's when there is some information. The structure can be examined

using the techniques given by Khatri (1973).

This model can be extended to cope with missing data. Kleinbaum (1973) introduced the extended model

$$E(Y_j) = A_j \xi P B_j \quad j=1,\ldots,u$$

where A_j has dimensions $(n_j x m)$ and B_j is an incidence matrix of 0's and 1's with dimensions $(t x t_j)$. The data matrices Y_j have dimensions $(n_j x t_j)$ that is there are n_j individuals which have observations at the same t_j times. Kleinbaum shows that the best asymptotic normal estimate for ξ is given by

$$\hat{\xi} = \left[\sum_{j=1}^{u} P B_j (B_j' \hat{\Sigma} B_j)^{-1} B_j' P' \otimes A_j' A_j \right]^{-}$$

$$\sum_{j=1}^{u} \left[P B_j (B_j' \hat{\Sigma} B_j)^{-1} \otimes A_j' \right] Y_j$$

$\hat{\Sigma}$ is any constant positive definite estimator of Σ.

Schwertman and Allen (1979) give a smoothed estimate for $\hat{\Sigma}$. Leeper and Woolson (1982) compare the method of Schwertman and Allen with two estimators suggested by Kleinbaum, one using just complete data, the other pooling across all available data for each cell of Σ separately. They conclude that if a subset of complete data of moderate size is available then $\hat{\Sigma}$ from this data should be used otherwise the smoothed estimate should be used.

An alternative approach to missing values is proposed by Liski (1984). He uses Bayesian methods to estimate the missing values, either via the linear model of Lindley and Smith (1972) or using Fearn's model as described below.

A Bayesian approach to the analysis of these models has been given by Geisser (1970) and Fearn (1975). The model is based on a two-stage approach in

which the regression coefficient for individuals come from a multivariate Normal distribution.

The individual model $Y_i \simeq N(X_i\beta_i, \sigma^2 I) \quad i=1,\ldots,n$

with second stage $\beta_i \simeq N(\mu, C)$

This gives $Y_i \simeq N(X_i\mu, X_i C X'_i + \sigma^2 I)$

Using vague prior knowledge for μ the posterior distribution for β_i is normal with

$$\text{mean} \qquad \beta_i^* = w_i\beta_i + (I-W_i)\left[\sum_{j=1}^{n} w_j\right]^{-1}\left[\sum_{j=1}^{n} w_j\beta_j\right]$$

$$\text{variance} = \left\{w_i + (I-W_i)\left[\sum_{j=1}^{n} w_j\right]^{-1} w_i\right\} \sigma^2_i\ (X_iX_i)^{-1}$$

$$w_i = (\sigma^2_i X'_i X_i + C^{-1})^{-1}\ \sigma^2_i (X'X_i)$$

$$\hat{\beta}_i = (X_iX_i)X_i'\ y_i$$

For the second stage

$$\mu^* = \frac{1}{n}\sum_{i=1}^{n} \beta_i$$

$$\text{Var} = \left[\sum_{i=1}^{n} [\sigma^2_i (X_i'X_i)^{-1} + C]^{-1}\right]^{-1}$$

If however, **C** is unknown, the problem is more complex. There is no solution to the required posterior integrals. Fearn suggests substituting estimates in the above expressions.

In a later paper Fearn (1977) shows how a two stage model will lead to Rao's model. This is obtained by giving μ the structure

$$\mu = \begin{bmatrix} \mu_0 \\ 0 \end{bmatrix}$$

A similar approach is used by Hui (1984). He assumes that

$$V_i = X_i \; C \; X'_i + \sigma^2 \; I$$

is of the form

$$V_i = \sigma^2_i \; J + \sigma^2 \; I$$

ie the form from the split-plot approach. The weighted least squares estimate of μ is then

$$\hat{\mu} = \left[\sum_{i=1}^{n} X_i' \; V_i^{-1} \; X_i \right] \quad \left[\sum_{i=1}^{n} X_i' \; V_i^{-1} \; y_i \right]$$

He shows how an iterative weighted procedure can be set up estimating σ_i^2 and σ^2 from the residuals.

The statistical aspects of the approaches described in the section have been subject to a considerable amount of investigation. In chapters 7 and 10 a more conceptual criticism of these models will be developed.

1.4 Non-linear Models

If we wish to consider more realistic models of growth, rather than the simple polynomial models, then these tend to be non-linear. Some of the most common are given below. They are derived from suitable differential equation models.

(i) Exponential

If W is the measurement under study then if the rate of growth of W is proportion to w, ie

$$\frac{dw}{dt} \alpha W$$

then $W = ae^{kt}$

This model assumes that there is no restriction on growth. This would be appropriate to certain stages in the growth of an organism before restrictions start to play a major role.

(ii) Monomolecular

Here there is a term indicating that there is an underlying steady growth rate and a term indicating a resistance to growth at a rate proportion to W

$$\frac{dw}{dt} = R(a - W) \quad a, k > 0$$

giving

$$W = a - be^{-kt}$$

As $t \to \infty$, $w \to a$, $\frac{dw}{dt} \to 0$, thus giving the curve an asymptote.

(iii) Logistic

This model gives an exponential growth with a resistance proportional to w^2

$$\frac{dw}{dt} = k(W - (1/a)W^2)$$

This could be considered as a second order Taylor's series approximation to the model

$$\frac{dw}{dt} = f(w)$$

The logistic model is

$$W = \frac{a}{1 + be^{-kt}}$$

As $t \to \infty$, $W \to a$, $\frac{dW}{dt} \to 0$, giving an asymptote, now

$$\frac{d^2W}{dt^2} = k(1 - 2/a\ W)\frac{dW}{dt}$$

so there is a point of inflection at $w = a/2$, $t = \frac{\log b}{k}$

(iv) Gompertz

Starting from the exponential growth law

$$\frac{dW}{dt} = r(t)W$$

If r decreases with time, exponentially, so that

$$\frac{dr}{dt} = - kr$$

then

$$f(t) = b' e^{-kt}$$

and

$$w = \exp(a - be^{-kt})$$

This is an asymetric curve with asymptote at w = exp a and point of inflection at t = (log b)/k.

(v) Richards or Generalised Logistic

A generalisation of the logistic function was proposed by Richards (1959, 1969). The differential equation is

$$\frac{dW}{dt} = kW\left[1 - \left[\frac{W}{A}\right]^{1/\theta}\right]$$

which gives

$$W = \frac{A}{[1 + e^{-(b+kt)/\theta}]^{\theta}}$$

This curve contains many of the above functions

$\theta = 1$ - logistic

$\theta = -1$ - monomolecular

$\theta \to 0_+$ - exponential

$\theta \to \infty$ and A a linear function of θ - Gompertz

(vi) Cui Qiwu and Lawson (1982) consider a different generalisation of the logistic. Using adsorption theory of chemical kinetics they derive the equation

$$\frac{dW}{dt} = \frac{\mu W(1 - W/W_m)}{1 - W/W_m'}$$

with solution

$$\log(W/W_0) - \log((W_m - W)/(W_m - W_0)$$

$$+ (W_m/W_m') \log((W_m - W)/(W_m - W_0)) = \mu(t - t_0)$$

using boundary conditions $w = w_0$ when $t = t_0$.

(vii) Turner et al (1976) and Pruitt et al (1979) consider the very general equation

$$\frac{dW}{dt} = \frac{\beta}{K^n} W^{1-np}(K^n - W^n)^{1+p}$$

with solution

$$W = \frac{K}{\{1 + (1 + \beta_{np}(t - t_0))^{-1/p}\}^{1/n}}$$

This function includes the Richards function.

Comparison of Curves

Of the four basic curves, (i) to (iv), only the exponential does not possess an asymptote, hence its use will be different from the other three, being used for organisms which do not reach senescence. The monomolecular does not have a

point of inflection where as both the logistic (at t = a/2) and the Gompertz (at t = $\frac{\log b}{k}$) do possess one.

The main difference between the logistic and the Gompertz is that the logistic is symmetrical whereas the Gompertz is not. Stone (1980) gives a general review of sigmoid curves.

Fitting of Curves

An important decision in fitting these non-linear curves is that between an additive error structure and a multiplicative error structure, the latter not being uncommon for this type of data.

This section will consider some of the particular aspects related to the above curves. General points on the minimisation of non-linear sums of squares will be given in Appendix A.

(i) Exponential

Under a multiplicative error, taking a log transformation a linear form is obtained

$$\log y = \log a + kt + \log e$$

For the additive error the model can easily be fitted by say the Gauss-Newton method. Initial estimates may be obtained from the linear form.

(ii) Monomolecular

With an additive error this model can be fitted by standard methods but an additional approach is as follows

If k were known to give a linear model

$$w = a + bx \quad \text{where} \quad x = e^{-kt}$$

so for given k, $\hat{a}$ and $\hat{b}$ can be found by linear least squares. A linear search (see Appendix A) on k can thus be carried out which will at the same time yield the required estimates of a and b.

For monomoecular model the log transformation only increases the non-linearity. It would therefore be better to use an iterative weighted procedure on the untransformed model using weight

$$v_i = (1/y_i)^2$$

This will lead to the same result provided the model gives a good fit to the data (ie $|\text{residual}| < |y_i|$).

(iii) Logistic

The fitting of the logistic curve has been studied extensively (eg Oliver (1964, 1966, 1982), Nelder (1961), Patton and Krause (1972), Vierra and Hoffmann (1977)). The correct parameterisation is important as it will improve the speed of convergence and lead to estimates which are not highly correlated.

Patton and Krause (1972) suggest a parameterisation which gives a diagonal asymptotic variance - covariance matrix for a true parameter model (a-known). He uses the two parameters

$$R = a/4.k$$

$$\tau = \log (b)/k$$

This is for an additive error structure.

In discussing the logarithmic version of the Richards function, Davies and Ku (1977) suggest replacing b by e^{-c} which is less non-linear than b.

Using this parameterisation of the log transformed model is

$$\log w = \log a - \log(1 + e^{-(c+kt)})$$

For both error structures the model can be fitted by standard techniques. For initial estimates a has to be estimated from a plot of the data and then if

$$\log\left[\frac{W}{\hat{a} - w}\right] \text{ is plotted against } t$$

the initial values of C and k can be obtained from the intercept and slope respectively.

(iv) Gompertz

Much less attention has been paid to the Gompertz curve than to the logistic. Vierra and Hoffmann (1977) discuss the estimation under both multiplicative and additive error terms.

Under a multiplicative error structure the log transformation yields the following form

$$\log W = a - be^{-kt} + \log c$$

this is the monomolecular function applied to log W, hence the discussion above is applicable.

Vierra and Hoffmann fit the model under an additive error by the Gauss-Newton method.

(v) Richards Function

The fitting of the Richards Function has been considered by Nelder (1961, 1962), Causton (1969) and Davies and Ku (1977). Nelder fits the curve by the method of scoring while Causton uses Newton-Raphson. In both cases a multiplicative error is assumed.

In his second paper Nelder reparameterises to give

$$W = \frac{a}{(1 + \varphi e^{-(b+kt)})^{1/\varphi}}$$

In this form the Gompertz is found as $\varphi \to 0$. Causton (1969) uses the form

$$\log w = \log a + m \log(1 + Be^{kt})$$

this is criticised by Davies and Ku (1977) who suggest the form

$$\log w = \log a + m \log(1 - e^{-(c+kt)})$$

this being less non-linear.

No results have been published for fitting the curve using an additive error.

Criticism of the above approach

While this approach provides more realistic models than does the statistical approach it has two major problems. First the problem of correlated errors is generally ignored. There have, however, been two investigations.

Finney (1958) proposed a monomolecular model with a particular form of correlated error structure. He provided estimates of the non-linear parameter based on ratios of linear and quadratic functions of the observations. The covariance structure of the model can be written as

$$\mathrm{Cov}[w_{t_1} w_{t_2}] = \left[-\tfrac{1}{2}\sigma^2/k\right]\left[e^{-k(t_2-t_1)}\right]\left[1 - e^{-2kt_2}\right]$$

These approaches are only concerned with models in which the covariance structure is a function of the parameters of the deterministic part of the model and hence have limited flexibility.

Glasbey (1979) considers the generalised logistic under an error structure with correlation

$$\text{Corr}\left[w_{t_1}\ w_{t_2}\right] = \sigma^2\rho^{|t-t'|}$$

In particular he looks at the situation of equally geared observations, then if $e_t = w_t - f_t$, where f_t is the functional value,

$$\text{log likelihood} = (n/2)\log\sigma^2 + \{(n-1)/2\}\log(1-\rho^2) + \{A + (1+\rho^2)B - 2\rho c\}/\{2\sigma^2(1-\rho^2)\}$$

where

$$A = e_1^2 + e_n^2,\quad B = \sum_{i=2}^{n-1} e_t^2,\quad C = \sum_{i=1}^{n-1} e_t e_{t+1}$$

This function is maximised by a composite process of Newton-Raphson, steepest descent and the method of scoring.

When the time between observations is not a constant,

$$\text{log likelihood} = \frac{1}{2}\sum_{i=1}^{n-1}\log\left[1 - \rho^{2x_1}\right] + (n/2)\log\sigma^2 + \left[e_1^2/\left[1 - \rho^{2x_1}\right] + e_n^2\left[1 - \rho^{2x_{n-1}}\right]\ t\ \sum_{i=2}^{n-1} e_i^2\left[1 - \rho^{2x_{i-1}}\rho^{2x_i}\right]/\left\{\left[1 - \rho^{2x_{i-1}}\right]\left[1 - \rho^{2x_i}\right]\right\} - \sum_{i=2}^{n} 2e_{i-1}\left[\rho^{x_{i-1}}\right]\right]/2\sigma^2$$

where

$$x_i = t_{i+1} - t_i.$$

Other approaches to this problem and an extension of Glasbey's results will be developed in chapter 4.

The recent work of Khatri and Shah (1982) helps to unite the approaches of sections 1.3 and this section. They consider a model of the form

$$Y = A\xi P(\lambda) + e$$

similar to Potthoff and Roy's model but P is a matrix of possibly non-linear known functions of the unknown parameters λ. In particular they look at the form

$$P = \begin{bmatrix} 1 & 1 & 1 & \dots & 1 \\ e^{-\lambda_1 t_1} & e^{-\lambda_1 t_2} & & & e^{-\lambda_1 t_t} \\ & & & & \\ e^{-\lambda_k t_1} & e^{-\lambda_k t_2} & & & e^{-\lambda_k t_t} \end{bmatrix}$$

This model would be different from Potthoff and Roy's model even if the λ_i's were known for in this case ξ will have structural zeros.

They estimate ξ and λ by Generalised Least Squares (GLS) and Maximum Likelihood (ML).

Let $$\xi \frac{\lambda \rho_i}{\partial \lambda_k} = \sum_j c_{kj}\, \xi_{kj} = C_{k'}\, \xi_{k'}$$

and $$Z = YS^{-1}P'(PS^{-1}P')^{-1} = YQ$$

where $$S = Y'[I - A(A'A)^{-1}A']^{-1}$$

For GLS estimators

$$\hat{\xi} = (P'S^{-1}P)^{-1} P'S^{-1}Y$$

and

$$\hat{\xi}_{k'} A'Y[I - QP]S^{-1}C_k = 0$$

For ML estimators

$$\hat{\xi} = (P'S^{-1}P)^{-1} P'S^{-1}Y$$

and

$$\hat{\xi}_k [I - (A'A)^{-1}A'Y(I - QP)^{-1}SY'A]^{-1}A'Y[I - QP]S^{-1}C_k = 0$$

This model is generalised to cover a number of distinct groups with entirely different parameters, in this case S becomes a pooled estimate. Confidence intervals were also found for the non-linear parameters.

The second problem is to analyse sets of curves fitted to individuals in an experiment. Several possibilities are now considered.

(i) Analyse the fitted parameters in a series of one dimensional ANOVA's. This is the extension of Wishart's approach. The values could be weighted by their variances.

(ii) Analyse all the coefficients using a MANOVA. Again the variances of the estimates could be incorporated in the analysis.

(iii) Covariates could be incorporated into the parameters ie let

$$a_i = \exp(\Sigma\, x_{ij}\, \beta_j) \text{ for the jth individual}$$

The analysis would then be based on a series of Log likelihood ratio statistics.

(iv) Krause, Seigel and Hurst (1962) examine a logistic curve in which the parameters a, b and log k have different independent log normal distributions. Structure representing the experimental design could be attached to the means of these distributions.

Further consideration of these problems will be given later.

1.5 Deterministic Approaches

A number of deterministic models for growth, particularly for populations, will be reviewed in this section.

Discrete Time Models

A number of difference equation models for growth have been developed. These are considered particularly relevant for populations with non-overlapping generations eg for anthropods (Hassell (1981)). These models may also be of use in the study of general growth particularly when the observations are taken at equally spaced time intervals.

The basic model is

$$N_t = \lambda N_{t-1}$$

with solution

$$N_t = \lambda^t N_0$$

Let $r = \log \lambda$

then $N_t = N_0 e^{rt}$

Hence this is the discrete time equivalent to the exponential growth model.

A more general, density dependent form is given by

$$N_t = N_{t-1} f(N_{t-1})$$

Various forms for $f(N_{t-1})$ are given by Pielou (1977), May et al (1974), May (1981)

(1) λN^{-b}

(2) $\lambda \exp(-\alpha N)$

(3) $(1 + r)/(1 + r(N/k))$

(4) $1 + r(1 - N/k)$

(6) $\lambda_0 + \dfrac{(\lambda_1 - \lambda_0)}{1 + \exp(S(N - J))}$

(3) is the discrete time equivalent of the logistic function.

Most interest in such models has been concerned with their behaviour. May (1976) gives a good account of the complex behaviour of models (4), (3) and (1). In this thesis we will be concerned with stochastic analogies of these models and in particular the problems of fitting and selecting such models.

These models can be extended to either multi-species populations or age structured populations by using the Leslie Matrix model (eg Usher (1972)). The basic model is

$$a_t = A\, a_{t-1}$$

where A is pxp matrix and a_t, a_{t-1} are p dimensional vectors, there being p species/age classes.

Density dependence can be introduced to the elements of A (Leslie (1948), Pennywick, Compton and Beckingham (1969)). Williamson (1974) uses the idea of jumping between two matrices A and A* depending on the value of a_{t-1}.

The stability of 2x2 density dependent models is considered by Cooke and Leon (1976).

Non-linear difference equation models for two species have also been proposed, they include

(i) $$N_{it} = \frac{\lambda_i N_{it-1}}{1 + \alpha_i N_{1t-1} + \gamma_i N_{2t-1}} \qquad i=1,2$$

(Pielou (1977))

(ii) Leslie and Gower's quoted Pielou (1977) Model

$$N_{1t} = \frac{\lambda_1 N_{1t-1}}{1 + \gamma_1 N_{2t-1}} \qquad N_1 \text{ is host}$$

$$N_{2t} = \frac{\lambda_2 N_{2t-1}}{1 + \gamma_2 N_{2t-1}/N_{1t-1}} \qquad N_2 \text{ is parasite}$$

(iii) The Nicholson-Bailey Model

$$N_{1t} = \lambda N_{1t-1} \exp(- a N_{2t-1})$$

$$N_{2t} = C N_{1t-1} (1 - \exp(- a N_{2t}))$$

or its generalisation

(iv) $$N_{1t} = N_{1t-1} \exp(r(1 - N_{1t-1}/k) - a N_{2t-1})$$

$$N_{2t} = \alpha N_{1t-1} (1 - \exp(- a N_{2t}))$$

(Both quoted May (1981)).

These models are combinations of models (2) and (3) for single species.

Most work on these models has been deterministic. In chapter 5 statistical aspects of difference equation models will be examined.

Continuous Time Population Models

The univariate models of section 1.4 can be extended to the multi-species situation.

The experimental growth model extended to p species gives

$$\dot{x} = A\, x(t)$$

where x(t) is the vector with elements x_i i=1,...,p

$\dot{x}$ is the vector with elements $\frac{dx_i}{dt}$ i=1,...,p

and A is a pxp matrix of coefficients.

The solution can be formally written as

$$x(t) = x_0 \exp\{A(t - t_0)\}$$

Numerical evaluation can be evaluated in a number of ways (eg Patten (1971)).

The interaction between species is described by the coefficients a_{ij}, the following are often used

Interaction i on j	Sign a_{ji}	Sign a_{ij}
Commensalism	+	0
Amensalism	-	0
Mutualism/symbiosis	+	+
Competition	-	-
Predator-Prey	+	-

(eg May (1973 b))

May (1973 a) and Jeffries (1974) have examined conditions for qualitative stability of such a system.

May produced the conditions

(i) $a_{ii} < 0$ $\forall i$

(ii) $a_{ii} < 0$ for at least one i

(iii) $a_{ij}a_{ji} <$ for all $i \neq j$

(iv) $a_{ij}\ a_{jk}, \ldots, a_{gr}\ a_{ri} = 0$ for any sequence of 3 or more distinct indices i, j, k, ..., g, r

(v) $\text{Det}(A) \neq 0$

The implications of these are

(i) No positive feedback

(ii) At least one species has a self regulating effect.

(iii) Commensalism, amensalism and predictor-prey are qualitative stable.

(iv) No closed loops exist.

Jeffries produced sufficient conditions for stability in terms of digraphs. The system can also possess neighbourhood stability if

(i) all eigenvalues of A have negative real parts

and

(ii) some eigenvalues have zero real parts and one distinct and the remainder have negative real parts.

(May (1973 b), Kowal (1971)).

As with the single species case a more realistic description is provided by non-linear models. Most attention has been given to the two species case.

The most famous model is the Lotka-Volterra model

$$\frac{dx_1}{dt} = a\, x_1 - \alpha\, x_1\, x_2$$

$$\frac{dx_2}{dt} = -\, b\, x_2 + \beta\, x_1\, x_2$$

where x_2 is the predator. There are many criticisms of this model. It assumes unlimited population size for prey in the absence of the predator and the solution of the equations leads to an unrealistic neutrally stable cycle.

A more realistic model is given by

$$\frac{dx_1}{dt} = r\, x_1(1 - x_1/k) - x_2\, F(x_1, x_2)$$

$$\frac{dx_2}{dt} = x_2\, G(x_1\, x_2)$$

May (1981) lists a number of different forms for F and G. They include

F	G
$\alpha\, x_1$	
$k\, x_1/(x_1 + D)$	$- b + \beta\, F(x_1, x_2)$
$k(1 - \exp(- C\, x_1))$	$S(1 - \gamma\, x_1/x_2)$

By taking various combinations of F and G or using one of the original Lotka-Volterra system a wide range of models can be produced including the Leslie and Gower Models and Holling-Tanner Model (Pielou (1977)).

Stability of such systems can be shown by two theorems. Consider the system

$$\frac{dx_1}{dt} = x_1\, F_1(x_1, x_2)$$

$$\frac{dx_2}{dt} = x_2\, F_2(x_1, x_2)$$

If $F_{ij} = \dfrac{\partial\, F_i(x_1\, x_2)}{\partial\, x_j}$

then Kolmogorov's theorem (May (1973)) states that if

(i) $F_{12} < 0$ (ii) $x_1\, F_{11} + x_2\, F_{22} < 0$

(iii) $F_{22} < 0$ (iv) $x_1\, F_{21} + x_2\, F_{22} > 0$

(v) $F_1(0, 0) > 0$

and there exists A, B and $C > 0$ with $B > C$ such that

(vi) $F_1(0, A) = 0$ (vii) $F_1(B, 0) = 0$ (viii) $F_2(C, 0) = 0$

Then the system has either a stable limit cycle or a stable equilibrium point.

Bulmer (1978) produced a set of more realistic conditions. These consist of

(a) $F_{21} > 0$ (b) $F_{22} < 0$ (c) $F_{11} < 0$ when $x_2 = 0$

(d) the conditions (i), (vi), (vii) and (viii) above and (e)

for constants α and μ

$$x_2 F_2 < \alpha[x_1 F_1(x_1, 0) - x_1 F_1(x_1, x_2)] - \mu x_2$$

The term in square brackets is the number of prey consumed in unit time by the predators, α is the maximum assimulation efficiency for conversion of prey into predator, and μ is the maximum value of a factor that includes the mortality of the predator and its energy requirements for maintenance activities.

These conditions lead to the same result.

Other models for two species interaction include competition models. May (1981) presents the simple model

$$\frac{dx_1}{dt} = r_1 x_1 (1 - (x_1 + \alpha_{12} x_2)/k_1)$$

$$\frac{dx_2}{dt} = r_2 x_2 (1 - (x_2 + \alpha_{21} x_1)/k_2)$$

where k_1 and k_2 represent carrying acpacities of the environment as seen by species 1 and 2 respectively, α_{12} and α_{21} are the competition coefficients.

As this brief review will have shown the main mathematical interest in these models has been the behaviour in relationship to stability. In chapter 6 stochastic versions of the models will be considered and the problem of parameter estimation discussed.

Systems Dynamics

The system dynamics approach to modelling was popularised by Forrester's 1961 book Industrial Dynamics. One of the most famous (infamous?)applicationswas

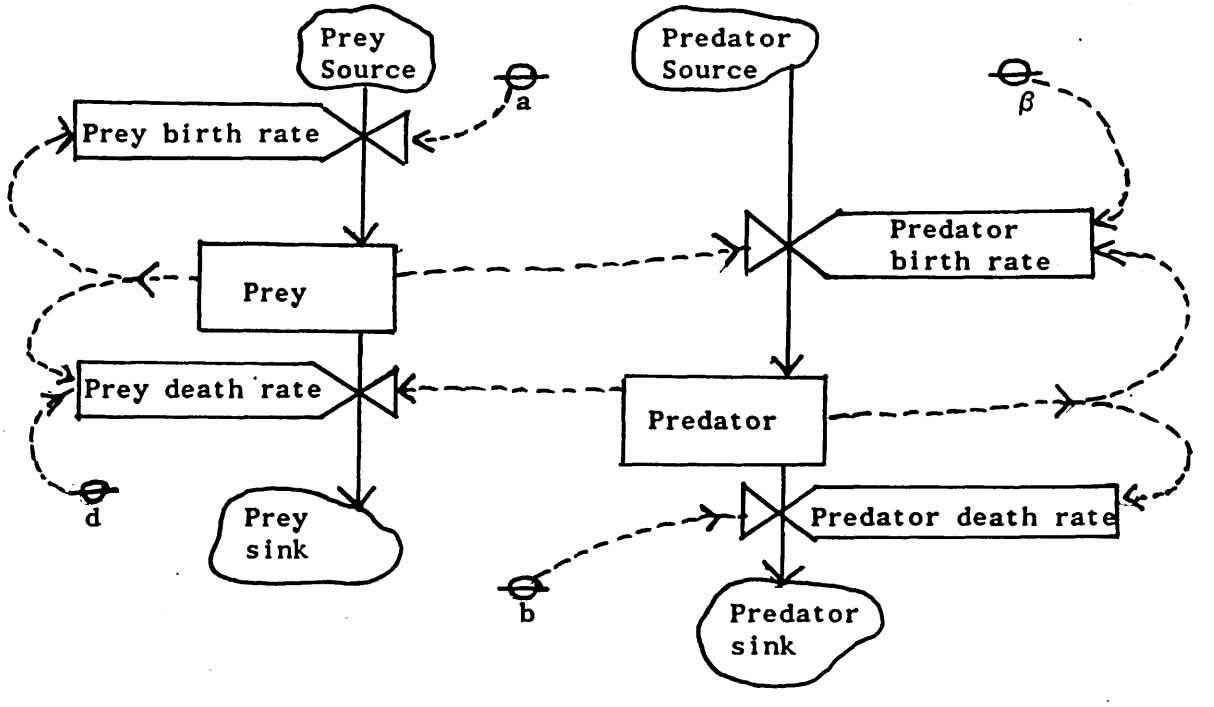
in The Limits to Growth (Meadows et al (1972)). The approach is based on linking various differential equation models to produce a model of the system. Such models can be simulated using numeric techniques. The popularity of such methods is in part due to the availability of computer packages eg DYNNAMO and DYSMAP.

A key idea in such modelling is feedback. In particular negative feedback. It is this that gives the models stability and, according to some schools of thought, makes them independent of parameter estimation problems.

An example of a system dynamic style model is given in fig 1.5(1). A discussion of the system dynamics approach will be given in chapter 7.

Fig 1.5(1)

A Forrester type diagram for the Lotka-Volterra Model



1.6 Statistical Modelling

The idea of statistical modelling as a subject has grown considerably in the last few years. For instance the books of Shapiro and Gross (1981) and Gilchrist (1984) and the Institute of Statisticians Conference on Statistical Modelling in 1985. What is meant by statistical modelling will be discussed in more detail in chapter 7, but to lay the foundation for chapters 2 to 6 the steps in statistical modelling as given by Gilchrist will be presented here

1. Identification

   This is selecting the most suitable model. The identification may be based on ideas about the situation (conceptual), the data (empirical) or a combination of both (eclectic).

2. Estimation and Fitting

   The parameters of the model are estimated using suitable criteria and the model fitted to the available data.

3. Validation

   The validity of the model is considered, this can take place at various stages in model development.

4. Application

   The use of the model, this will effect all stages of the modelling process.

5. Iteration

   The above stages are not linear but the modeller will pass back and forth between them.

1.7 Aims of Study

The general aim is to produce a wide range of statistical models for growth suitable for application in many areas. In particular the problems of estimating and fitting the models will be considered along with identification, validation and sensitivity of the models.

First, several methods of model identification will be examined in chapter 2. The problems in fitting non-linear growth models will be considered in chapter 3 and in particular the problem of auto correlated errors will be considered in chapter 4. Chapter 5 will look at dynamic models for growth rather than the static models of chapter 3. These will include both difference and differential equation models. Chapter 6 extends the work to the multivariate situation. Chapter 7 considers what is meant by statistical modelling. In particular approaches to modelling will be compared.

Two application areas will then be considered. Plant growth in chapter 8 and human growth in chapter 9.

2. Empirical Identification

2.1 Introduction

Before considering the problems of model fitting the problem of deciding upon a suitable model for a given data set has to be considered. The identification in this chapter is essentially of the empirical type and consideration of the mechanisms that could lead to the models will not be considered.

The work is primarily concerned with the static models as considered in chapter 3. The consideration of the identifiction of suitable dynamic models can be tackled in two ways

(i) Identifying the deterministic solution to the problem specially using the estimates of relative growth rates considered in section 2.3.

(ii) Identifying the density dependence in the growth model.

Both of these broad approaches are considered in this chapter.

Two basic approaches will be considered. The first, model dependent procedures are those in which the data is examined to see if one of a set of models is suitable. The second approach, model free procedures, attempts to present the data in a way that suitable models will be suggested.

Two other important aspects of model identification will also be considered. The testing for density independent growth as mentioned above, and the estimation of the inherent variability of the data.

2.2 Model Dependent Identification Procedures

One approach to identifying a model is to find a linearization and plot the transformed size against time. If ω represents size then for the principal models considered in the introduction

	Model	Transformation	
(a)	Exponential	$\log \omega$	
(b)	Monomolecular	$\log(\omega-\hat{a})$	$\hat{a}$ estimated from plot of ω vs time
(c)	Logistic	$\log\left[\frac{\omega}{\hat{a}-\omega}\right]$	$\hat{a}$ estimated from plot of w vs time
(d)	Gompertz	$\log(\log(\omega)-\hat{a})$	$\hat{a}$ estimated from a plot of $\log \omega$ vs time
(e)	Richards	$\log\left[\left[\frac{\omega}{\hat{a}}\right]^{\hat{\varphi}}-1\right]$	$\hat{a}$ estimated from a plot of ω vs time

In (e) an estimate of φ is required. One approach would be to try several values of φ to see which, if any, give a linear plot. Alternatively φ could be estimated using standard estimation and the plot used as a post estimation validation procedure.

A second approach is to look at the relative growth rates (Levenback and Reuter (1976)). If a growth rate estimator derived from a particular model is plotted against suitable functions it should give a linear plot. The following are suitable estimators for the above models.

(a) Exponential $\qquad \dfrac{\log \omega_{t_{i+1}} - \log w_{t_i}}{t_{i+1} - t_i}$

This is plotted against time and should give a horizontal line.

(b) Monomolecular $\dfrac{xy(\log(x) - \log(y))}{\tau\ \omega_{t_i}(x-y)}$

where $x = \omega_{t_{i+1}} - x_{t_i}$

$y = \omega_{t_i} - \omega_{t_{i+1}}$

$\tau = t_{i+1} - t_i$

This is plotted against $1/\omega_t$.

(c) Logistic $\dfrac{qs(\log(q) - \log(s))}{q - s}$

where $q = \left[\omega_{t_{i+1}} - \omega_{t_i}\right] / \tau\omega_{t_{i+1}}$

$s = \left[\omega_{t_i} - \omega_{t_{i-1}}\right] / \tau\omega_{t_{i-1}}$

and τ is as above.

This is plotted against ω_t

(d) Gompertz $\log\left\{\dfrac{\log\left[w_{t_{i+1}}\right] - \log\left[\omega_{t_i}\right]}{t_{i+1} - t_i}\right\}$

This is plotted against time.

(e) Richards

No corresponding plot is available since when fitting the curve to four points the resulting equations do not have a closed form solution.

Gregg et al (1964) devised a procedure based on quantities called slope characteristics. These are functions of slopes and moving averages.

The slope is estimated by linear orthogonal poloynomial and a moving average of similar extent is used to estimate the mean value. By suitable plots the different curves can be distinguished

Curve	Plot vs time	Shape
Exponential	slope/moving average	horizontal
Monomolecular	log(slope)	sloping down to right
Logistic	log(slope/(moving average)^2)	sloping down to right
Gompertz	log(slope/moving average)	sloping down to right

Holmes (1983) is critical of this method showing that even for error-free data the ability to distinguish between say the logistic and the Gompertz is slight.

Finally Holmes (1983) suggested the following procedure for data measured at equal intervals.

The data is divided into overlapping blocks each block consisting of a fixed number of consecutive data points (between 10 and 25). If there are 2m+1 points the first being at t = k then the transformation

$$T_k = \{(t-k)/m\} - 1$$

is used. Then within each block the model

$$f(T_k) = \sum_{r=0}^{4} \alpha_r(k)\, g_r\,(T_k)$$

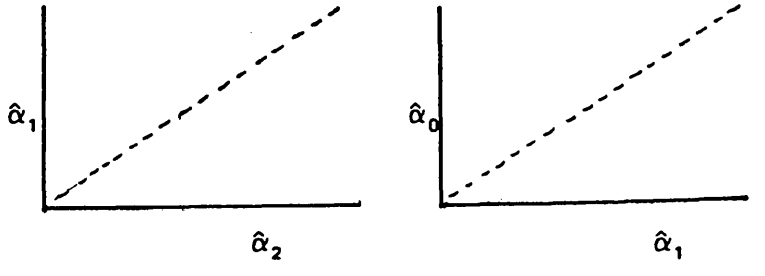
is fitted where

$\alpha_r(k)$ are unknown paramters

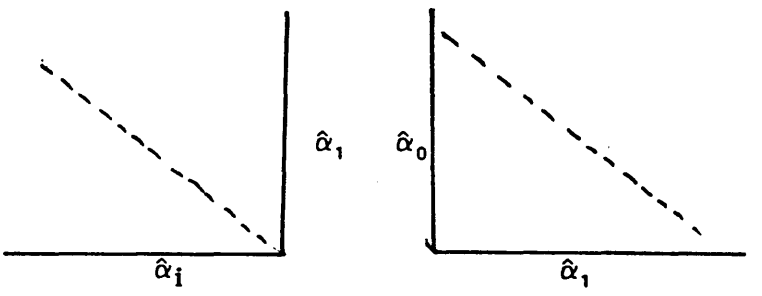
$g_r(\tau)$ are known polynomials (see Holmes)

Plots of $\hat{\alpha}_0$ vs $\hat{\alpha}_1$ and $\hat{\alpha}_1$ vs $\hat{\alpha}_2$ are made, these should give the following configurations

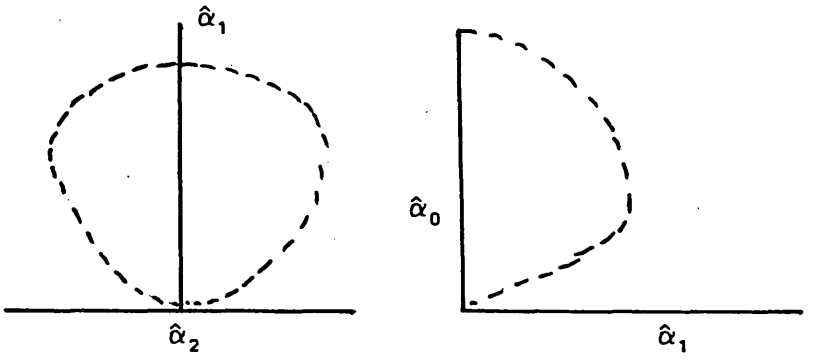
(a) Exponential



(b) Monomolecular



(c) Gompertz or Logistic



This method would seem to be of limited use in the problems considered in this work as they require a long series of observations which are generally not available for growth situations.

2.3 Model-Free Identification

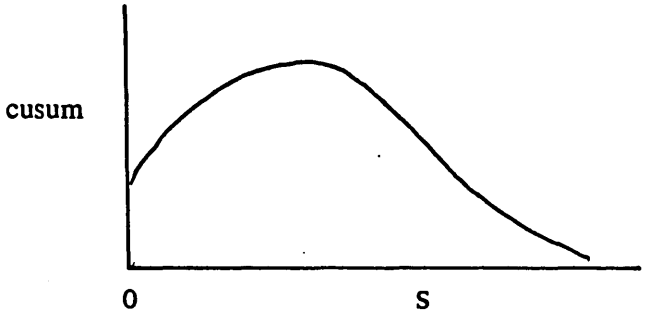
Sandland and McGilchrist suggest a method based on a cumulative sum (cusum) plot. They calculate and estimate the relative growth rate

$$R_t = \frac{\log[\omega_{t_{i+1}}] - \log[\omega_{t_i}]}{(t_{i+1} - t_i)}$$

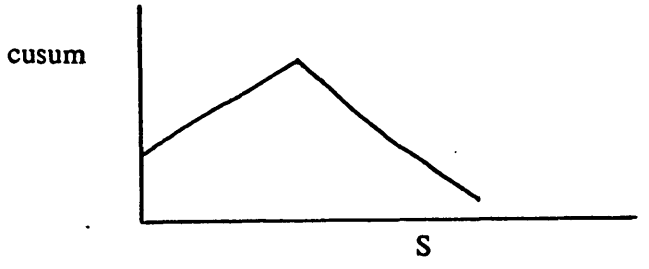
A cusum is then formed by calculating

$$C = \sum_{j=0}^{S} (\hat{R}_j - \bar{R}) \quad ; \quad \bar{R} = \sum_{j=0}^{T} \hat{R}_j/T$$

against S. These plots can indicate whether there is a smooth change in the relative growth rate, giving plots like



or a phase change indicated by plots with a sharp peak



These plots would have to be used in conjunction with plots of $\hat{R}_j$ against time.

The cusum plots are useful for indicating if a multiphase model is required, but do not distinguish clearly between the standard growth models described earlier.

Another approach to the identification problem is to find a suitable smooth estimate of the relative growth rate and examine plots for that. A plot of the actual data can give some information eg whether there is a point of inflection or if the curve is symmetrical, but the information is limited as all curves have the same basic shape, whereas the underlying equations for the relative growth rate may be quite different.

Hershey, Zakin and Sinha (1967) investigate differentiating experimental data using orthogonal polynomials. An orthogonal polynomial is fitted locally to a series of points and an estimate of the derivative at the central point is given by the derivative of the fitted polynomial. This leads to a series of

coefficients for the derivative at time t. These can be calculated from tables of orthogonal polynomials (eg Fisher and Yates)

Number of points	Degree of polynomial	Coefficients XD							D
		t-3	t-2	t-1	t	t+1	t+2	t+3	
5	2		-2	-1	0	1	2		10
5	3		1	-8	0	8	-1		12
7	2	-3	-2	-1	0	1	2	3	28
7	3	22	-67	-58	0	58	67	-22	252

Other sets of coefficients are available to provide estimates for end points. This method will only work for equally spaced observations. For data that is not equally spaced local polynomials could still be fitted and their derivatives calculated, but the calculations would be more laborious than with orthogonal polynomials. The computation would involve updating the regression at each point. (See Chambers (1971)).

The smoothed estimate of relative growth rate can then be plotted against either time or a smooth estimate of ω. The latter would be obtained by use of a local polynomial in the same way as the slope is estimated. (See table 2.3(1) for the forms of the relative growth rates).

A more sophisticated approach to the problem would be the use of cubic splines (eg Reinsch (1967), Wold (1974)). Cubic splines are local cubic polynomials that are continuous for the first two derivatives. In order to fit such curves by least squares the B-spline representation is used.

$$S(x) = \sum_{t=1}^{m+2} \lambda_i \beta_i(x)$$

$i+2$

where $$\beta_i(x) = \sum_{k=i-2} (x-\xi_k)^3 \Big/ \prod_{\substack{s=i-2 \\ s\neq k}} (\xi_k-\xi_s)$$

where

m = number of knots (or connections between polynomials)

ξ_i i=1,...,m position of the knots

and

$$\xi_k = \xi_1 - (1-k)(\xi_1-x_{min}) \qquad k < 0$$

$$\xi_k = \xi_m + (k-m)(x_{max}-\xi_m) \qquad k > m+1$$

Computational considerations are considered by M G Cox (1972, 1975, 1978).

When fitting splines a balance between smoothness and accuracy of fit needs to be obtained for the greater the number of knots the greater the accuracy but the rougher the fitted curve. This problem is discussed by Wold (1974), Powell (1967) and Woodford (1970). Wold gives the following rules of thumb

(1) Have 4 or 5 points per interval.

(2) Have not more than one extremum point and one inflection point per interval.

(3) Have extremum points centred in the intervals.

(4) Have inflection points close to knots.

Cubic splines may be fitted by algorithms in the Nag library, these are based on the work of Cox.

A comparison of results based on both the local orthogonal polynomials and the use of cubic splines indicated that the extra computation required in the use of splines was not justified.

Although the use of splines and local polynomials has been suggested for the estimation of various growth rates (eg Hunt (1982)) their use in the identification of suitable growth functions has not been fully appreciated.

Table 2.3 (i)

Relative Growth Rates

Model			
Exponential	$W = be^{kt}$	k	
Monomolecular	$W = A(1-be^{-kt})$	$\frac{bke^{-kt}}{1 - be^{-kt}}$	$k\left[\frac{A-W}{W}\right]$
Compertz	$W = Aexp(-be^{-kt})$	bke^{-kt}	$k \log\left[\frac{A}{W}\right]$
Logistic	$W = \frac{A}{1 + be^{-kt}}$	$\frac{bke^{-kt}}{1 + be^{-kt}}$	$k\left[1-\frac{W}{A}\right]$
Richards	$W = \frac{A}{\{1 + e^{-(b+kt)/\theta}\}^i}$		$k\left[1-\left[\frac{W}{A}\right]^{1/\theta}\right]$

As a result of experience with several data sets the following conclusions could be reached.

A simple method of model-free identification is to differentiate the logarithm of the observations using local polynomials. By applying different polynomials various compromises between smoothness and closeness of fit can be obtained. The order for decreasing smoothness is

OP(7, 2)

OP(3, 2)

OP(7, 3)

OP(3, 3)

where OP(n, p) is a p'th degree polynomial fitted to n points.

Plots of these against time and a smoothed value of the observations give a good indication of the type of model to be investigated. Further confirmation can be found by either the methods of section 2.2 or from a past model fitting examination.

These methods differ from those of Gregg et al in that a variety of different slope indicators are available and the aim is to give an insight into the situation rather than pick out a particular model from a particular plot.

2.4 Testing For Density Dependence

As was stated in the introduction the null model for growth is often the density independent exponential growth model

$$y(t) = y(0)e^{rt}$$

Thus if

$$x(t) = \log y(t)$$

$$x(t) = x(t') + r(t-t') \quad t>t'$$

So for equally spaced observation a plot of x(t) against x(t-1) should give a straight line with slope 1. This is known as a Morris plot. If density dependent growth is present then the slope of the line will be less than one. This suggests that a test of the regression coefficient of x(t) on x(t-1) to see if it equals 1 should test for density dependence. However, a number of authors have pointed out that such a test is biased (Eberhardt (1970), Maelzer (1970), St Amant (1970)). The problem being that for density independent growth the respected value of the slope is less than one as it is approximately the first order correlation coefficient.

Several suggestions have been made to overcome this problem. Varley and Gradwell (1968) had suggested examining both the regression of x(t) and x(t-1) and x(t-1) on x(t) and only rejecting density independence if both were significantly different from 1. Other suggestions include using functional relationship methods (see chapter 6). These are reviewed by Slade (1977).

Bulmer (1975) suggested a test for density dependency based on the von Neumann ratio

$$\frac{\sum_{i=1}^{N-1} (x_{(t+1)} - x_{(t)})^2}{\sum_{i=1}^{W} (x_{(t)} - \bar{x})^2}$$

This test is only suitable in the absence of any trend and so is not of use for the situations considered in this project.

Two recent comparisons of the tests are those of Slade (1977) and Vickery and Nudds (1984).

Slade compared the methods using three models

(i) $x(t) = r + \beta x(t-1) + e(t)$

(ii) $x(t) = r + \beta[x(t-1) - \omega(t-1)] + \omega(t)$

(iii) $x(t) = r + \beta[x(t-1) - \omega(t-1)] + \omega(t) + e(t)$

where $\omega(t)$ and $e(t)$ are random errors. The latter two can best be understood in terms of state space models considered in chapters 5 and 6. He concluded that Varley and Gradwell's method was too conservative, Bulmer's test did not work when there was a population trend and in general the major axis methods

were better than standard regression when there is a trend.

Vickery and Nudds (1984) used the simulation model

$$x(t) = x(t-1) + e(t)$$

where x(0) and e(t) were chosen to give the $y = e^x$ the same mean and variance as certain duck populations. Their conclusions were similar to Slade's, except that they come out more strongly in favour of a major axis test rather than a standard major axis.

Vickery and Nudds also suggested the use of simulation tests based on the test statistics. In a similar approach Pollard and Lakhani (1986) have considered a randomization test using the regression slope as a test statistic.

In examining tests for independence the key aspects seem to be the choice of the null and alternate models. In the study of growing individuals or populations the basic null model will be

$$x(t) = r + x(t-1) + u(t)$$

where u(t) is the random error term which may be

(i) auto correlated

(ii) density dependent (ie var (u_t) is a function of x(t)).

The alternative model is usually of the form

$$x(t) = r + \beta x(t-1) + u(t).$$

This is only one of several density dependent relationships and the form of u(t) has not been fully considered. There are two approaches which could be used

(i) A simple test based on the basic null model

$$x(t) = r + x(t-1) + e(t)$$

using either a simulation test (Vickery and Nudds (1984)) or a randomization test (Pollard and Lakhani (1986)) based on either the simple regression coefficient or the major axis estimate. The results for these tests is promising even if they are time consuming to perform.

(ii) A test based on a complete specification of the stochastic model for both the null and alternate situations. Such models are considered in chapter 5.

The first approach along with plots would be most useful in the initial identification phase and would point the modeller towards density dependent models. The second approach could be used later when more specific ideas of the most suitable possible models have been obtained.

2.5 Estimating Inherent Variability

An important aspect of model selection in multiple regression is knowledge of the underlying variability as measured by the variance of the random component, σ^2. Methods such as R^2 and c_p plotting aim to find the model that has the residual sum of squares at the level of this inherent variability with the smallest number of variables. In cases of repeated observation for certain values of the independent variables the inherent variability is estimated by the pure error estimate (Draper and Smith (1981)). In non-linear regression the adequacy of a particular model would be indicated from the comparison of its residual sum of squares and the inherent variability if that were known. Methods of estimating inherent variability in situations without replication will be considered.

In order to estimate the intrinsic variability a very flexible model has to be fitted to the data and then the residual sum of squares used as the basis of the estimate. Four possible types of flexible model could be used

(i) Moving averages

(ii) Cubic splines

(iii) Local independent polynomials

(iv) Higher order global polynomials.

Green (1971) for linear models and Breiman and Meisel (1976) for non-linear models have considered (iii). In order to use such an approach the following decisions have to be made

(a) How to split the data into segments.

(b) Selection of suitable model for each segment.

Breiman and Meisel use a first order polynomial and select the number of segments by means of an F-test. Green also recommends the use of a first order polynomial. From a simulation study Breiman and Meisel an indication of the number of segments needed can be obtained

Sample size	Number of segments
100	4 - 8
500	10 - 40
2000	25 - 100 (max)

that is about 25 points per segment. However in most growth situations a maximum of about 50 observations are available, often considerably less. In these situations a natural division of the growth period into three segments would be the most appealing approach. Breiman and Meisel estimate σ^2 by

the total residual sum of squares divided by the number of observations. This gives a better estimate than when the divisor is adjusted by the number of parameters estimated.

Method (ii) would probably be inferior to method (iii) as the additional restriction of smoothness on the model would make the estimate of variability larger than that for (iii) and the extra complication is not required.

A simple moving average may be reasonable for the growth curve situation when observations are often equally spaced. Using the various order moving averages as used in section 2.3 an indication of the variability could be obtained from the residuals.

For the relatively short series in which there are few observations in a plateau region (ie at maximum growth) then the higher order polynomial should give reasonable results. The order of the polynomial and the estimate of intrinsic variability can be decided from a plot of the value of s^2 from the regression against the order, this should level out at the intrinsic variability as the order becomes large. In many growth situations it would be advisable to logarithmically transform the data first.

Breiman and Meisel compare their method with methods (ii) and (iv) for two simulated models. Their method was superior but in consideration of their results it has to be noted that they used very large sample sizes (100-2000) and the polynomial had a highest order of only 3. It is unlikely that the method would be as superior with a smaller sample size. Further, only two models were considered. The only true way of evaluating these methods is how useful they are in practice.

2.6 Discussion

A number of technics to help the modeller to identify a suitable model have been considered in this chapter. They should be viewed not as rule producing formulae but as a guide for the modeller. They should help highlight aspects of the data which he can take into account alongside his conceptual knowledge of the situation.

The application of these methods will be illustrated in chapters 8 and 9.

3. Static Models For Growth

3.1 Introduction

First what is meant by a static model needs to be discussed. A static model does not explicitly describe the change of behaviour from one time point to the next. When fitting such a model the order in which the observations (eg y, x) occur is unimportant.

In growth studies the observations will be of the form (y, t, x), where y is the measured growth, t the time of observation and x any explanatory variables. Using the usual regression terminology y is the dependent variable and x and t are the independent variables. The models considered will be based on dynamic deterministic models of the form

$$\frac{d\eta}{dt} = g(\eta, t, x) \qquad (1)$$

However, the solution to the above differential equation will be assumed known ie

$$\eta = f(t, x) \qquad (2)$$

and observations are then

$$y = \eta + e \quad \text{or} \quad \log y = \log \eta + e \qquad (3)$$

where e are independent errors.

As was discussed in chapter 1, one of the most common form of growth function, f(t, x) is the logistic. In this chapter a generalisation, the Richards function, will be examined in detail along with other generalisations of the logistic.

3.2 Causton and Venus' Work on Richards' Function

A large proportion of Causton and Venus' book, The Biometry of Plant Growth (1981), is devoted to the examination of the Richards' function.

In biological growth it is usual for the variance to increase with the size of observation, therefore the logarithmic transformed data is considered eg

$$\log y = \log \eta + e$$

Causton and Venus consider the log transformed Richards

$$\ell = \log \eta = a - \frac{1}{n} \log(1 \pm e^{(b+kt)}) \qquad (1)$$

The negative alternative being used when n is negative.

They fit the non-linear regression model

$$\log y = \ell + e \qquad (2)$$

using the Newton-Raphson method applied to the sum of squares

$$s = \Sigma (\log y - \ell)^2$$

This method requires both the first and second derivatives of s with respect to the parameters. The weakness of the Newton-Raphson method is concerned with the matrix G such that

$$G_{ij} = \frac{\partial^2 s}{\partial \beta_i \partial \beta_j} \qquad (3)$$

where β_i i=1,...,4 represent the parameters. G is known as the Hessian matrix. The iterations are given by

$$\beta^{k+1} = \beta^k - G_k^{-1} g_k \qquad (4)$$

where

g is the gradient matrix $\left\{\frac{\partial s}{\partial \beta_j}\right\}$

and k indicates the kth iteration.

If a poor initial estimate of β is used then G will be numerically unstable. For sum of squares functions the Gauss-Newton method or a derivative such as Marquardts method are usually recommended (see Appendix A). One way of looking at the Gauss-Newton method is to see it as replacing G by 2V'V where

$V_{ij} = \frac{\partial \ell_i}{\partial \beta_j}$ in this case.

Now

$$G_{ij} = \frac{\partial}{\partial \beta_j}\left[\frac{\partial s}{\partial \beta_i}\right] = \frac{\partial}{\partial \beta_j}\left[-2 \sum_m (\log y_m - \ell_m) \frac{\partial \ell m}{\partial \beta_i}\right]$$

$$= -2 \sum_m \left\{-\frac{\partial \ell_m}{\partial \beta_j}.\frac{\partial \ell_m}{\partial \beta_i} + (\log y_m - \ell) \frac{\partial^2 \ell_m}{\partial \beta_i \partial B_j}\right\}$$

$$= 2 \sum_m \left[\frac{\partial \ell_m}{\partial \beta_j}\right]\left[\frac{\partial \ell_m}{\partial \beta_j}\right] - 2 \Sigma e_m \frac{\partial^2 \ell_m}{\partial \beta_i \partial \beta_j}$$

So

$2V'V = E(G)$.

Hence it can be considered as an approximation if the residuals are small or as using the method of scoring.

The Gauss-Newton methods require less computation than the Newton-Raphson and Marquardts method, in particular, is numerically more stable.

Causton and Venus gave the variance of the estimates as

$\text{Var}\ (\bar{\beta}) = -I^{-1}$

where

$$I = -\frac{1}{2\sigma^2} G \qquad ,\sigma^2 = \text{var}(e)$$

This is the asymptotic variance of maximum likelihood estimation which is equivalent to least squares under the normal model. The usual estimates of variance for non-linear regression models are given by

$$\text{Var}(\hat{\beta}) = (V'V)^{-1}\sigma^2$$

In the light of the previous discussion this can be considered as the expected Fisher information as compared with the observed as given above (Efron and Hinkley (1978)).

They investigate the properties of the estimates by means of a simulation experiment. Using models suitable for (a) sunflower (leaf weight) growth in harvests between days 8 and 34 inclusive, and (b) wheat (leaf area) growth with harvests between days 6 and 24 inclusive. The results of 200 simulations are given below

	Sunflower		Wheat		
Parameter	Actual	Alternate day Harvests	Actual	Daily Harvests	Alternate day Harvests
a	-1.114	-1.114	2.767	2.769	2.775
b	6.430	6.407	45.49	48.11	52.93
k	-0.3898	-0.3890	-2.675	-2.828	-3.112
n	0.4395	0.4377	3.206	3.392	3.725

The simulations were based on 10 replicates per harvest and the standard deviation of the normal error was 0.2.

The results show that for the sunflower the estimation is fine but the estimates are poor for the wheat data. They also investigated the distribution of the estimates, for the sunflower they were approximately normal but showed positive skew for the wheat.

They conclude that based on these results and previous experience that if n is in the range

$$0.05 < n < 0.6$$

then there are rarely problems of estimation. This claim will be investigated later using a different approach.

Causton and Venus also consider the problem of heteroscedastic data. They carry out a simulation for the model with increasing error variance on the log scale and a = -1.116, b = 6.442, k = -.3916 and n = 0.4417. The models were fitted using both ordinary and weighted least squares.

The results show little difference in the average value of the estimates but the variances for the weighted least squares were lower. They conclude that there is little to choose between the two methods providing n is small.

Having examined the theoretical properties of the model and its fitting the authors proceed to show its usefulness in a variety of situations in the study of plant growth.

3.3 Sensitivity of Richards' Function

3.3.1 Sensitivity

The sensitivity of a model is the effect changes in variables, parameters and model structure have upon the behaviour of the model (Gilchrist (1984), Jeffers (1982)).

For a model $y = f(\beta) + e$

where β are parameters. The sensitivity coefficient is defined as

$$^{c}\beta_i = \frac{\partial f}{\partial \beta_i} \qquad (1)$$

The Elasticity of the Sensitivity is defined as

$$^{\ell}\beta_i = {}^{c}\beta_i \frac{\beta_i}{f(\beta)} \times 100 \qquad (2)$$

(Gilchrist (1984)).

The object of these indices is to gauge the effect of a small change in the parameter on the fitted value ($^{c}\beta_i$) or the percentage change in fitted value for the percentage change in the parameter ($^{\ell}\beta_i$). The sensitivity, as defined here, has to be considered alongside the confidence interval for the estimated parameter. That is the possible variation in the parameter due to estimation uncertainty indicates how critical sensitivity can be.

The effect of sensitivity on model fitting can be viewed in two ways

(i) High sensitivity will mean that the fitted model will have to be treated cautiously and its application under any slight change of circumstances would be a dubious process.

(ii) If the model is sensitive to a parameter it would seem likely that the parameter will be accurately estimated by least squares since a slight change in the parameter will greatly effect the residual sum of squares.

In some ways (i) and (ii) counteract each other, the key to the effect will often be the data. If the data is of good quality and fully represents the area of application then the effects of sensitivity will be reduced. However if poor data is used, eg possible outliers, poor range of values, then the model will be worthless.

In the case of the Richards function the effect of model form is not critical due to the flexible nature of the model. This flexibility is most closely connected with the parameter n, so a study of n will be important in the analysis of the sensitivity of the Richards function.

3.3.2 Evaluation of Sensitivity

Using the form 3.2(1) eg

$$\ell = a - \frac{1}{n} \log(1 + e^{(b+kt)})$$

the following are obtained

$$\text{let } A = e^{(b+kt)}$$

$$c_a = 1$$

$$c_b = -\frac{1}{n} \{A/(1 + A)\}$$

$$c_k = -\frac{1}{n} \{At/(1 + A)\}$$

$$c_n = \frac{1}{n^2} \log (1 + A)$$

If the alternate form

$$\ell = a - m \log (1 + e^{(b+kt)})$$

is used then

$$c_m = \log (1 + e^{(b+kt)}) \qquad (4)$$

The most interesting elasticities are

$$\ell_n = 100\left\{\frac{an}{\log(1 + A)} - 1\right\}^{-1} \qquad (5)$$

and

$$\ell_m = 100\left\{\frac{a/m}{\log(1 + A)} - 1\right\}^{-1}$$

Three ranges of values of t will now be considered, high, low and intermediate. It should be noted that these results need to be treated with some caution due to the effect of the log scale.

For high values of t the curve will tend to its asymptote so with negative K

$A \to 0$.

This means only

$$c_a = 1$$

and

$$\ell_a = 1$$

are non zero. This is as would be expected.

For low values of t $A \to e^b$. In both the wheat and sunflower models discussed above $e^b \gg 1$ so $\log(1 + e^b) \simeq b$ and $A/(1 + A) \simeq 1$.

Hence

$$c_b \simeq -\frac{1}{n} \qquad c_k \simeq -\frac{t}{n}$$

$$c_n \simeq \frac{b}{n^2} \quad \text{and} \quad \ell_n \simeq \frac{100}{(a/b)n - 1)}$$

Evaluating c_n and ℓ_n for the sunflower data

$$c_n \simeq 33.3 \qquad \ell_n \simeq -92.9\%$$

For the wheat data

$$c_n \simeq 4.43 \qquad \ell_n \simeq -124\%$$

For intermediate values very approximately we can take $A \simeq 1$ so

$$A/(1 + A) \simeq \tfrac{1}{2} \quad \text{and} \quad \log(1 + A) \simeq \log 2$$

so

$$c_b \simeq -\frac{1}{2n} \qquad c_k \simeq -\frac{t}{2n}$$

$$c_n \simeq \frac{\log 2}{n^2} \qquad \ell_n \simeq \frac{100}{\frac{a_n}{\log 2} - 1}$$

Again evaluating for sunflower data

$$c_n \simeq 3.6 \qquad \ell_n \simeq 67\%$$

for wheat data

$$c_n \simeq 0.07 \qquad \ell_n = 8\%$$

This shows that the wheat data appears less sensitive than the sunflower, the mid time points being about day 16.

Overall the sensitivity of the sunflower model seems marginally greater than the wheat model. When the models were fitted to different data the results presented by Causton and Venus show the following coefficients of variation for the parameters

	Sunflower	Wheat
a	.47	.22
b	.74	.83
k	.49	.57
n	.67	.96

Showing that the inter-data variability of the two models is similar, the sunflower being slightly lower.

Overall we can see in this case the situation which is more sensitive is also better estimated.

3.4 Measures of Non-Linearity

3.4.1 Background

For a linear model

$$y = \sum_j \beta_j x_j + e$$

the properties of the estimators of the β_j's are exact under the normal model. Further

$$^c\beta_j = x_j$$

so the sensitivity is only scale dependent.

For non-linear models the properties of estimates are only approximate. The approximation depends on how non-linear is the model. Various measures of non-linearity will be examined.

3.4.2 Beale's Measures of Non-Linearity

Beale(1960) proposed two measures of non-linearity. Consider the model

$$y_i = f_i(\theta) + e$$

where θ is parameter vector of length p. Using a Taylor's series approximation

$$f_i(\theta) \simeq f_i(\hat{\theta}) + \sum_{j=1}^{p} v_{ij} \alpha_j + \sum_{j=1}^{p} \sum_{k=1}^{p} \omega_{ijk} \alpha_j \alpha_k$$

where

$$v_{ij} = \frac{\partial f_i}{\partial \theta_j} \quad \text{and} \quad \alpha_j = \theta_j - \hat{\theta}_j \tag{1}$$

and

$$w_{ijk} = \frac{\partial^2 f_i}{\partial \theta_j \, \partial \theta_k}$$

The usual properties of least squares estimates are based on the adequacy of a local linear approximation. That is the third term above is ignored. Intuitively the size of this term is a measure of the non-linearity of the model. Beale's first measure is

$$N_\theta = \frac{\sigma^2}{p+2} \sum_{i=1}^{n} \sum_{j=1}^{p} \sum_{k=1}^{p} \sum_{\ell=1}^{p} \sum_{m=1}^{p} \omega_{ijk} \, \omega_{i\ell m} \, \delta_{jk\ell m} \tag{2}$$

where

$$\delta_{jk\ell m} = \gamma_{jk} \, \gamma_{\ell m} + \gamma_{j\ell} \, \gamma_{km} + \gamma_{jm} \, \gamma_{k\ell} \text{ and } \gamma_{jk}$$

is the (j,k)th element of the matrix $(V'V)^{-1}$, the element of V being v_{ij}.

N_θ can be considered as the theoretical version of the measure

$$ps^2 \frac{\sum_{\ell=1}^{m} \sum_{i=1}^{n} \left[f_i(\theta_\ell) - f_i(\hat{\theta}) - \sum_{j=1}^{p} v_i \, \alpha_j \right]^2}{\sum_{\ell=1}^{m} \left[\sum_{i=1}^{n} (f_i(\theta_\ell) - f_i(\hat{\theta}))^2 \right]^2}$$

where θ_ℓ $\ell=1,\ldots,m$ are points in the neighbourhood of $\hat{\theta}$.

This measure is dependent of the parameterisation used. So a measure of intrinsic non linearity, N_φ, was also suggested, this is parameter independent.

$$N_\varphi = \frac{\sigma^2}{p+2} \sum_{i=1}^{n} \sum_{j=1}^{p} \sum_{k=1}^{p} \sum_{\ell=1}^{p} \sum_{m=1}^{p} \omega'^{*}_{ijk} \, \omega^{*}_{i\ell m} \, \delta_{jk\ell m} \tag{3}$$

where

$$\omega^{*}_{ijk} = \omega_{ijk} - \sum_{\ell=1}^{p}\sum_{m=1}^{p}\sum_{o=1}^{n} \omega_{im}\,\gamma_{\ell m}\,\omega_{o\ell}\,\omega_{ojk}$$

The difference $N_\theta - N_\varphi$ gives a measure of the effect of parameterisation on non-linearity.

The non-linearity is disastrous if

$$\hat{N}_\theta > 1/F_a(p,n) \qquad (4)$$

and the usual approximations are satisfactory if

$$\hat{N}_\theta < .01/F_\alpha(p,n) \qquad (5)$$

where $\hat{N}_\theta$ is the empirical version of N_θ.

Guttman and Meeter (1965) examined Beale's measures and concluded that for high non-linearity the empirical measures underestimated the amount of non-linearity and for both theoretical and empirical the bound given above was conservative.

Gillis and Ratkowsky (1978) tried using these measures but reported them to be unsatisfactory in that they did not detect problems found through simulation.

3.4.3 Box's Measure of Bias

Related to the problem of the measure of non-linearity is the bias in estimates. Box (1971) gave a measure of this bias.

Box shows that the expected bias is

$$b = -\frac{\sigma^2}{2}\left[V'V\right]^{-1}\sum_{i=1}^{n} V_i\,\mathrm{tr}\left[\left[V'V\right]^{-1}W_i\right] \qquad (6)$$

where V_i is the ith row of V and W_i is the pxp matrix with elements ω_{ijk} .

His measure of bias is given by

$$M = b' \frac{V'V}{p\sigma^2} b$$

and he shows that $N_\theta - N_\varphi \geq M$. .

Gillis and Ratkowsky (1978) reported favourable results for these measures in their study.

3.4.4 Bates and Watts Measures

Bates and Watts (1980) produce two measures of curvatures; the parameter effects curvature and the intrinsic curvature. For convenience they standardise with respect to a standard radius $s\sqrt{p}$ by dividing V and W by $s\sqrt{p}$.

They use the QR decomposition of V

$$V = Q \begin{bmatrix} R \\ 0 \end{bmatrix} \begin{matrix} pxp \\ n-pxp \end{matrix} \qquad (7)$$

setting $L = R^{-1}$ they compute

$$U = L' \, W \, L \qquad (nxpxp) \qquad (8)$$

and hence

$$A = Q' \, U \qquad (nxpxp) \qquad (9)$$

this is partitioned into

$$\underset{(nxpxp)}{A} = \underset{(pxpxp)}{A^T} : \underset{(n-pxpxp)}{A^N} \qquad (10)$$

Their measures are then given by

$$\text{Maximum Intrinsic curvature} = \Gamma^N = \max_d \Sigma \, d' \, A_i^N \, d \qquad (11)$$

and

$$\text{Maximum Parameter effects curvature} = \Gamma^T = \max_{d} \Sigma\, d'\, A^T\, d \qquad (12)$$

They relate their measures to Beale's measures. First they show that Beale's measures can be written as

$$N_\theta = \frac{1}{p(p+2)} \left\{ \frac{1}{4} \sum_{i=1}^{n} \left[\sum_{j=1}^{p} U_{jji} \right]^2 + \frac{1}{2} \sum_{i=1}^{p} \sum_{j=1}^{p} \left[\sum_{k=1}^{n} U_{ijk} \right]^2 \right\}$$

$N_\theta - N_\varphi$ is as for N_θ but with $\mathbf{U}$ replaced by A^T (the n summation replaced by p)

and

N_φ is as for N_θ but with $\mathbf{U}$ replaced by A^N (the n summation replaced by n-1)

They then show that N_φ is a quarter of the mean square intrinsic curvature and $N_\theta - N_\varphi$ is a quarter of the mean square parameter effects curvature.

Bates and Watts suggest comparing their measures with

$$1/\sqrt{p\, F(p,\ n-p\ .05)} \qquad (13)$$

which can be regarded as the radius of curvature of the 95% confidence region.

3.4.5 Comparison with Sensitive Measure

Both the measures of sensitivity given in 3.3.1 and the properties of the least squares estimates for a non-linear model are based on the adequacy of the approximation

$$f(\beta + \delta\beta) = f(\beta) + \sum_{j=1}^{p} \frac{\partial f}{\partial \beta_j}\, \delta\beta_j$$

Therefore both are only valid if the amount of non-linearity, as measured by the above measures, is low.

3.4.6 Measures of Bias and Non-linearity for the Richards Function

A FORTRAN program was written to compute

Beale's measures of non-linearity

Bates and Watts measures of non-linearity

and

Box estimates of bias.

The program was validated by using results given by Bates and Watts (1980) and Gillis and Ratkowsky (1978).

The results for Causton and Venus' sunflower and wheat models are given in table 3.4(1).

The results can now be compared with the simulation results of Causton and Venus given on page 55.

Bates and Watts comparison statistic $1/\sqrt{F}$ is 0.42. Beale's limits are (at 5%) .0017 and .17. The intrinsic measures (Γ^N and N_φ) give a slight cause for concern in the daily wheat model only. Γ^T is significant in all cases while N_θ is only critical (> .17) for alternate day wheat but indicates inadequacy for the other two.

The bias estimates are accurate for the sunflower model, give a reasonable indication for the daily wheat model and are inaccurate for the alternate day wheat.

Both Γ^T and N_θ pick out the poor performance of alternative day what estimation while only Γ^T picks out the daily wheat.

The following tentative conclusions could be drawn. The Bates and Watts measures are better at detecting medium non-linearity. Box's measures of bias are fine for low and medium non-linearity but poor for high non-linearity (high value of Γ^T, critical value of N_θ).

Table 3.4(1)

	Sunflower		Wheat
		Daily	Alternate Days
Γ^N	0.0185	0.4797	0.4060
Γ^T	1.3488	1.8123	7.3659
N_φ	0.0003	0.0075	0.0049
N_θ	0.0588	0.1487	1.9483
a bias	0.0006	0.0008	-0.0000
b bias	0.0289	4.9006	51.1667
k bias	-0.0017	-0.2877	-3.0063
n bias	0.0025	0.3469	3.6147

The method was also applied to a series of models fitted by Causton and Venus to leaf areas of sunflowers (table A5) and leaf weights and areas of wheat (table A6). The alternate day harvest strategies as described before were used. The results for sunflowers are given in table 3.4(2) and for wheat in table 3.4.(3). As can be seen some values are extremely large. All values of Γ^T were significant and 9 values of N_θ were critical (marked with a c). There is no apparent relationship between the value of n and the values of Γ^T and N_θ. This must call into question Venus and Caustons statement on the validity of estimation if $0.05 < n < 0.6$. What can be seen from table 3.4(3) is that the strength of non-linearity is a property of the individual leaf ie the figures for both weight and area are of the same order.

Having applied the measure of non-linearity to a range of fitted Richards functions it can be seen that in nearly all cases there is significant non-linearity and the estimates should therefore be treated with great caution.

Table 3.4(2)

Leaf Number	n	Γ^T	N_θ	
1	.4026	1382.8625	59769.9727	c
2	1.3720	1.3082	0.0638	
3	.6633	0.8932	0.0271	
4	.4261	1.4584	0.0677	
5	1.6672	7.8215	1.8827	c

Table 3.4(3)

	Leaf Number	n	Γ^T	N_θ	
	1	.1740	$1x10^7$	$7x10^{12}$	c
	2	.2963	2489	193703	c
Weight	3	2.2249	1.950	.1666	
	4	.5952	.8217	.0236	
	5	1.6993	1.635	.1011	
	6	.8367	2.688	.2254	c
	1	-.2955	$2x10^8$	$1x10^{15}$	c
	2	.5801	21318	$1x10^7$	c
Area	3	3.4076	1.8133	.1846	c
	4	.8265	.8644	.2773	c
	5	1.9607	2.0238	.1599	
	6	1.5667	1.146	.0408	

3.4.7 Further Results for Non-linear Models

Clarke (1980) examined the moments of the regression estimates for a non-linear model. As well as obtaining Box's estimates of bias he also obtained an improved estimate of variance.

Adopting Clarke's results to the notation used here, we need to define the pxpxpxp matrix C as the first p faces of

$$Q' Z$$

where Z is obtained by multiplying each of the n(pxpxp) faces of the matrix of third derivatives of f by L.

The variance is then given by

$$\mathrm{Var}(\hat{\beta}) = L(\sigma^2 I + \sigma^4(VA + VB + VC))L'$$

where

$$V A = \sum_{i=1}^{n} (u_u{}^2 - u_i(\mathrm{trace}\ u_i))$$

$$- \sum_{i=1}^{p} (A^T{}_i - A^T{}_i(\mathrm{trace} A^T{}_i))$$

$$VB_{ij} = \sum_{\ell,m}^{p} \left\{ \frac{1}{2} A^T_{i\ell m} A^T_{jim} + \frac{1}{2} A^T_{ij\ell} A^T_{\ell mm} \right.$$

$$\left. + \frac{1}{2} A^T_{ji\ell} A^T_{\ell mm} + A^T_{i\ell m} A^T_{mj\ell} + A^T_{j\ell m} A^T_{mi\ell} \right\}$$

and

$$VC_{ij} = -\frac{1}{2} \sum_{\ell}^{p} (C_{ij\ell\ell} + C_{ji\ell\ell})$$

where all matrices are in the unstandardised form.

It can be seen that if VA, VB and VC are ignored

$$\mathrm{Var}(\hat{\beta}) = L L' \sigma^2 = (V'V)^{-1} \sigma^2$$

However, as was seen in the case of the Richards function the bias estimate starts to break down at medium non-linearity. This indicates that the variance estimates will only be of use for low, but not negligible, non-linearity. Thus these results are of limited use.

Cook and Tsai (1985) examine the residuals from a non-linear model. They show that the expected values of the residuals are given by

$$E(r) = \frac{1}{2} \sigma^2 \, N \sum_{i=1}^{p} A^N_{.ii}$$

where the (nxn-p) matrix N forms an orthonormal basis for the orthogonal complement of the column space of V.

Their results show that the residuals will be biased for models with a significant intrinsic non-linearity.

They also produce some projected residuals. However, as with linear regression, the improved properties of the projected residuals are outweighed by their lack of a 1-1 correspondence with the original observations (Masters (1975)).

3.5 Design of Experiments

3.5.1 Previous Work

The work of Box and Lucas (1959) and Box (1968) laid the foundation for the extension of the ideas of optimal design from the linear case (eg Fedorov (1972)) to the non-linear situation.

The optimal design for a model will be those x values that

maximise det (V' V).

Box (1968) showed that for a number of cases this is given by the design with $^{n}/_{p}$ observations and p points.

White (1973) produced a non-linear equivalent to Keifer and Wolfowitz's General Equivalence Theorem.

As the designs are based on the local linear approximation they will only be valid in those situations without severe non-linearity. The problem of severe non-linearity will be compounded because **V** contains values of the unknown parameters. These values need to be estimated from previous experiments and will be biased, particularly in the situation of severe non-linearity.

3.5.2 Application to Richards Function

The situation envisaged is one in which harvests can be taken at any selected time and the number of harvests and number of individuals per harvest can be chosen subject to a constraint on the total number of observations. Let harvests take place between t_0 and t_3. Then if we assume we take $^{n}/_{4}$ observations at 4 points for the 4 parameters it is reasonable to take two of the points as t_0 and t_3. Thus we need to find two points $t_0 < t_1 < t_2 < t_3$ such that

$$t_1,\ t_2 \text{ minimise } \det(V'\ V)^{-1}$$

where V is the 4x4 matrix with elements

$$V_{ij} = \frac{\partial f(t_i,\ \beta_j)}{\partial \beta_j}$$

where f is the Richards function with parameters β_1, β_2, β_3 and β_4.

This was achieved by using the Nelder-Mead algorithm (Appendix A).

3.5.3 An Example

Using the parameter values given for the sunflower in section 3.2 the optimal design for a range of dates 8 days to 34 days was

$t_1 = 12.6$

$t_2 = 18.4$

The ratio of the determinants of this design to the alternate day strategy is 3.11.

The values of $(V'V)^{-1}$ for both designs are Alternate Day (14 time points)

a	.0221	-.0185	-.2288	-.0119
n		.0575	.6892	.0324
b			8.3531	.3964
k				.0190

Optimal Design

a	.0294	-.0117	-.1464	-.0084
n		.0324	.4009	.0193
b			5.0082	.2432
k				.00120

(using 140 observations in each case).

These show a considerable improvement in using an optimal design.

The effect on non-linearity was also considered. For the optimal design with 140 observations and standard deviation .2 as before, the following values were

obtained

Γ^N	$.12 \times 10^{-13}$	a bias	.0014
Γ^T	.9344	b bias	.0156
N_φ	$.63 \times 10^{-29}$	k bias	-.0009
N_θ	.0278	n-bias	.0013

Again, these values show a considerable improvement. This indicates that the optimal design has improved bias properties as well as smaller variances.

3.6 Alternative Approaches to Estimation

3.6.1 Two Approaches

In this section two non-standard approaches to estimation will be considered. They involve balancing the usual criterion of minimising the sum of squared residuals with criterion based on bias or a criterion based on the distance of the estimates from 'reasonable' values of the parameters.

3.6.2 Bayesian and Bayesian-like Methods

The standard Bayesian approach is to find the posterior distribution of the estimates as

posterior $\propto$ likelihood x prior

The posterior can then be summarised in terms of either its mode (MPD - maximum of posterior distribution) or in terms of regions of highest density (HPD regions) (Bard (1974), Box and Tiao (1973)).

Fitting non-linear models has been considered by a number of authors eg Bard (1974), Box and Tiao (1973), Katz, Azen and Schumitsky (1981), Beal (1982) and Berkey (1982).

Consider the general non-linear model -

$$y_i = f_i(\beta) + e_i \qquad (1)$$

where $\mathrm{Var}(e_i) = \sigma^2$ and β is the p dimensional vector of unknown parameters. Assuming errors to be independent and normal the logarithm of the likelihood will be

$$\text{const} - \frac{n}{2} \log \sigma^2 - \frac{1}{2\sigma^2} \Sigma (y_i - f_i(\beta))^2 \qquad (2)$$

If the prior distribution of β is also normal with mean β_0 and variance-covariance matrix Σ_0 and with a non informative prior density for σ^2 $p(\sigma^2)$ proportion to $1/\sigma^2$ (Box and Tiao (1973)) then the logarithm of the posterior density is

$$\text{const} - \log \sigma^2 - \frac{1}{n} (\beta - \beta_0)' \Sigma_0^{-1} (\beta - \beta_0) - \frac{n}{2} \log \sigma^2 - \frac{1}{2\sigma^2} \Sigma (y_i - f_i(\beta))^2 \qquad (3)$$

Maximising with respect to σ^2 we obtain a maximised log posterior for β as

$$\text{const} - \left[\frac{n}{2} + 1\right] \log \left\{\Sigma (y_i - f_i(\beta))^2\right\} - \frac{1}{2} (\beta - \beta_0)' \Sigma_0^{-1} (\beta - \beta_0) \qquad (4)$$

If σ^2 is integrated out of (3) then the log posterior is

$$-\frac{n}{2} \log \{\Sigma (y_i - f_i(\beta))^2\} - \frac{1}{2} (\beta - \beta_0)' \Sigma_0^{-1} (\beta - \beta_0) \qquad (5)$$

(O'Hagan (1976))

So to find the MPD estimates the minimum of

$$S^* = \exp \left\{\frac{-1}{n+S} (\beta - \beta_0)' \Sigma_0^{-1} (\beta - \beta_0)\right\} \Sigma (y_i - f_i(\beta))^2 \qquad (5)$$

$\delta = 0$ or 1 as appropriate

is found.

This could be considered as weighted least squares with a weight of

$$\exp\left\{\frac{-\frac{1}{2}}{n+\delta}(\beta-\beta_0)'\,\Sigma_0^{-1}\,(\beta-\beta_0)\right\} \qquad (7)$$

The minimisation can be achieved through iterative weighted least squares. One purpose of the prior in non-linear estimation is to draw the estimates towards a more 'reasonable' value. As can be seen with Causton's data in section 8.2, the almost flat ridge in the sum of squares surface draws the estimates to extremely large values with poor statistical properties. Information from a range of fitted models could be used to construct a prior that would keep the estimates in sensible range even at the loss of some fit in terms of the residual sum of squares.

As the prior can be seen as a penalty function acting on the sum of squares minimisation other penalty functions could be considered. One such function is

$$S^{\lambda} = \left[\frac{\lambda n}{p}\right](\beta-\beta_0)'\,\Sigma_0^{-1}\,(\beta-\beta_0) + \Sigma\,(y_i - f_i(\beta))^2 \qquad (8)$$

This function can be viewed in three ways

(i) As being similar to the use of a penalty function to convert a constrained optimisation problem to an unconstrained problem (Walsh (1975)). In these cases the penalty function takes a low value inside the feasible region and an increasing value outside. In the case the flexible region is 'fuzzy' but the basic concept of using a penalty function to hold the minimum to a particular region is still valid.

(ii) As a form of non-linear ridge regression. Ridge regression (eg Hoerl and Kennard (1970)) aims to overcome the problems of ill conditioned linear least squares by solving the adjusted normal equations

$$X' X + kI) \beta = X' Y$$

These estimates can have improved mean squared error properties.

Marquardt (1970) showed that ridge regression is equivalent to augmenting the X and Y matrices with p extra rows ie

$$X^* = \begin{bmatrix} X \\ H \end{bmatrix} \qquad Y^* = \begin{bmatrix} Y \\ 0 \end{bmatrix}$$

such that $H' H = kI$

Taking a simplified version of (8) with independent parameter penalties

$$S^{\lambda} = \Sigma \omega^2_j(\beta_j - \beta_{0j})^2 + \Sigma (y_i - f_i(\beta))^2 \qquad (9)$$

which is equivalent to augmenting the data by

$$y_i = 0 \quad i=n+1,\ldots,n+p$$

$$f_{i+j}(\beta) = \omega_j(\beta_j - \beta_0) \quad j=1,\ldots,p$$

The $V' V$ matrix will become

$$(V' V + \text{diag } \omega^2_j))$$

(iii) As a non-linear version of the Goldberger-Theil estimators (Toutenburg (1982)). For a linear regression model

$$Y = X \beta + e \quad , \quad \text{var}(e) = V$$

Let other information concerning β be expressed as

$$\beta_0 = \beta + \varepsilon$$

where

$$\text{var}(\varepsilon) = P$$

Combining

$$\begin{bmatrix} Y \\ \beta_0 \end{bmatrix} = \begin{bmatrix} X \\ I \end{bmatrix} \beta + \begin{bmatrix} e \\ \varepsilon \end{bmatrix} = X_1 \beta + e_1$$

$$\text{var}\begin{bmatrix} e \\ \varepsilon \end{bmatrix} = \begin{bmatrix} V & 0 \\ 0 & P \end{bmatrix} = V_1 \text{ assuming independence.}$$

Then the BLUE estimate of β is

$$\hat{\beta} = (X_1' \, V_1^{-1} \, X)^{-1} \, (X' \, W^{-1} \, y + P^{-1} \, \beta_0)$$

3.6.3 Using Bias in Estimation

As will be seen in the examination of Causton's data in section 8.2 it was possible to find a value of n such that Γ^N (and N_θ) was minimised. It would seem reasonable to use an estimation process that takes into account both fit and the bias of estimates. Using the residual sum of squares as a measure of fit and criteria of the form

$$F(\lambda) = \lambda x(\text{bias measure}) + \Sigma(y_i - f_i(\beta))^2 \qquad (10)$$

could be used.

Possible Bias Measures are

(i) Box's measure $M = (b' \, v' \, v \, b)/ \, p\sigma^2$

(ii) As Box showed that M is bounded by $N_\theta - N_\varphi$ and in most cases considered N_φ has been relatively small then N_θ could be used.

(iii) As Bates and Watts show that the bias can be calculated from the matrix A^T which is the basis of Γ^T then Γ^T could also be used as a measure.

The value of λ has to be chosen to balance the two parts of the criteria. In selecting λ the following need to be considered

(i) The size of σ^2 and hence the residual sum of squares.

(ii) The critical value for the measure of bias. As it is the parameter effects bias which is of interest, the intrinsic non-linearity provides a lower bound for the measure of bias in cases (ii) and (iii). It should be possible to find a range of values for the bias measure that would indicate by how much we would like to decrease it.

3.6.4 Application of Methods

In order to illustrate the use of the methods considered above two sets of data were simulated. They both consisted of 140 observations using the sunflower alternate day harvesting design, with a standard deviation of 0.2 as described in section 3.2.

Set A used the sunflower parameter values considered before

a = -1.114

b = 6.430

k = -0.3898

n = 0.4395

Set B used a different n value

a = -1.114

b = 6.430

k = -0.3898

n = 2.2753

To obtain suitable prior values for these parameters and suitable relative weights the means and variances from 5 leaves reported by Causton and Venus (1982) (Table A.5) were used. These gave

Parameter	a	b	k	n
Mean value (β_0)	-1.45	3.75	-.308	.327
Standard Deviation	.69	2.79	.151	.219
Weight (ω_j)	.17	.14	.81	.56

Where the weights are inversely proportional to the variances, adjusted so that

$\Sigma\omega^2_j = 1$

The adjusted sum of squares was then minimised

$$S^{\lambda} = \lambda\, n \sum_{j=1}^{4} w^2_j \left[\beta_j - \beta_{j0}\right]^2 + \sum_{i=1}^{n} \left[y_i - f_i(\beta)\right]^2$$

for values of λ = 0 .1, .2, 1, 2 using a SAS program.

The results were

Data Set A

λ	.0	.1	.2	1	2
a	-1.1606	-1.1604	-1.1606	-1.1609	-1.1621
n	0.4494	0.4512	0.4517	0.4526	0.4555
β	6.5902	6.5846	6.5783	6.5676	6.5309
k	-0.4014	-0.3992	-0.3989	-0.3984	-0.3966

Data set B

λ	0	.1	.2	1	2
a	-1.0799	-1.4762	-1.2482	-1.1095	-1.1533
n	1.9851	1.4762	1.2482	1.0281	0.8274
β	4.6366	6.2531	5.8556	5.7967	4.9690
k	-0.3229	-0.3716	-0.3713	-0.3427	-0.3276

In order to understand these results it is useful to consider

$(n/p)\omega^2_j\ (\beta_j - \beta_{j0})^2$

evaluated at the value of $\hat{\beta}_j$ and also the residual sum of squares.

Set A

	$(n/p)\omega^2_j\ (\hat{\beta}_j - \beta_{j0})^2$	
a	0.0847	
n	0.1644	Resid ss = 5.0639
β	5.3790	
k	0.2003	
	5.8284	

Set B

	n	
a	0.1385	
n	30.1763	Resid ss = 10.1203
β	0.4916	
k	0.0005	
	30.8069	

For Set A starting from the least squares estimates, the largest value of the penalty function of interest is obtained since any increase in the penalty function must also increase the residual sum of squares and hence the value of s^{λ}. This value is approximately the same as the residual sum of squares. Further the sum of squares surface is sufficiently steep for any decrease in the value of the penalty function to be outweighted by the increase in residual sum of squares for any moderate value of λ. Thus it can be seen that there is little change in the value of the parameter estimates over the range of λ considered.

For Set B the value of the penalty function is three times the residual sum of squares at the least squares estimate. Most of which occurs due to the difference between $\hat{n}$ and its prior value. This leads to $\hat{n}$ being pulled towards the prior value for increasing values of λ.

To examine the approach of using the bias in estimation data Set A was again used. Rather than minimise a composite value of bias measure and residual sum of squares (10) over all the parameters the bias measure was related to the parameter n alone using the following procedure

(i) For fixed values of n over a suitable range the least squares estimates of the remaining parameters, a β and k were obtained

and the residual sum of squares computed.

(ii) For each of the above fitted models the value of Γ^T was obtained.

(iii) Plots of residual sums of squares and Γ^T against n were drawn and the minimum of $F(\lambda)$ = Residual sum of squares + $\lambda\ \Gamma^T$ found for a suitable value of λ.

The results are shown in figures 3.6(1), (2) and (3) Figure 3.6(3) shows a plot of F against n taking $\lambda = 0.1$. The value $\lambda = 0.1$ was selected since the range of values of the residual sum of squares over $.39 < n < .31$ was 0.063 and the range for Γ^T was 0.43 thus $\lambda = 0.1$ gives a reasonable balance. From figure 3.6(3) it can be seen that the minimum value is when $n = 0.46$. Using this value the fitted model has parameter values

$a = -1.1638$

$n = 0.46$

$\beta = 6.7375$

$k = -0.4087$

A further illustration of the ideas presented here will be given in chapter 8.

3.7 Other Logistic Generalisations

3.7.1

In addition to the Richards function a number of generalisations of the basic logistic function have been considered in growth studies. This section will examine some that have been used in human growth.

Fig. 3.6(1)

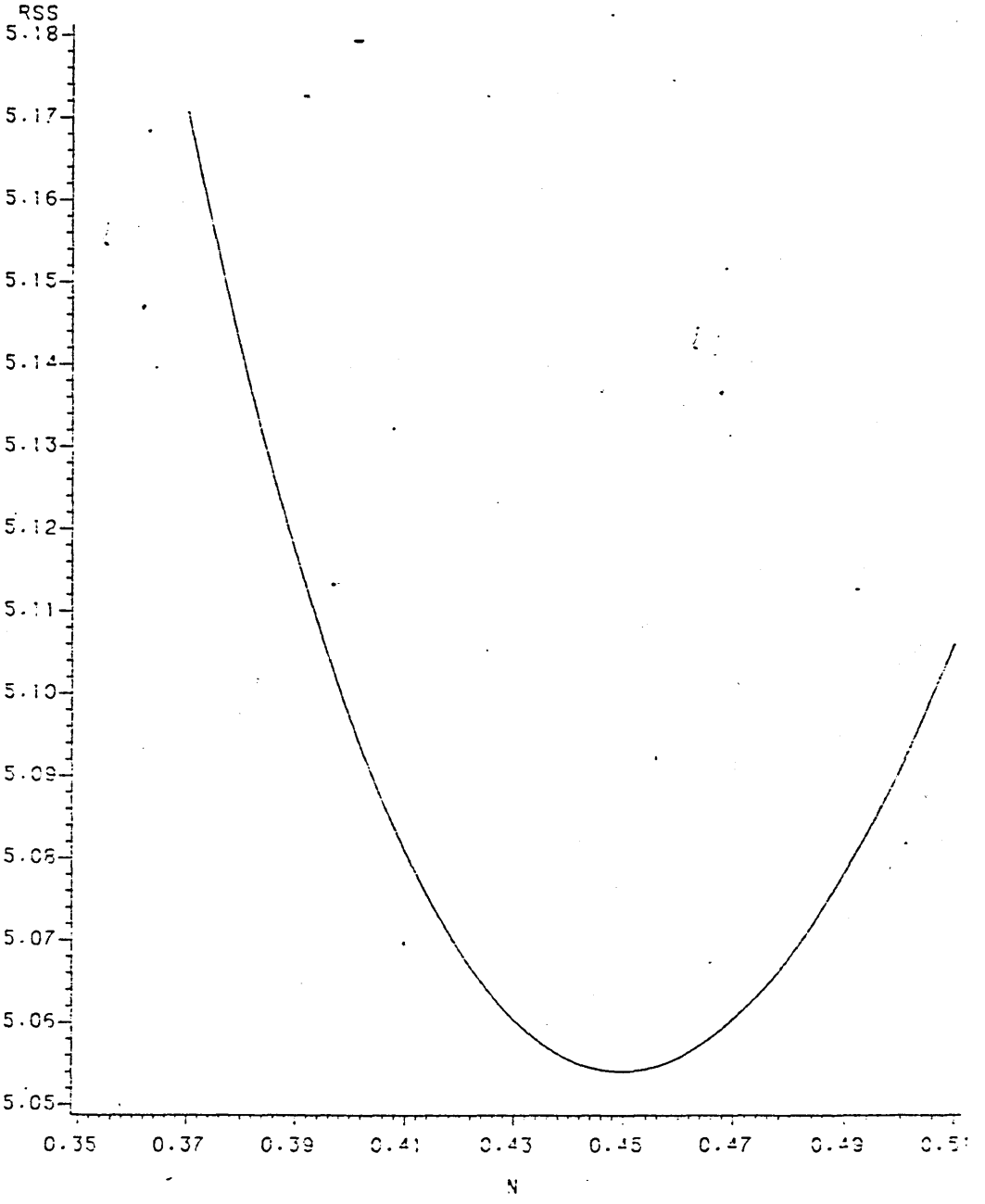

RSS
5.18
5.17
5.16
5.15
5.14
5.13
5.12
5.11
5.10
5.09
5.08
5.07
5.06
5.05
0.35
0.37
0.39
0.41
0.43
0.45
0.47
0.49
N

Fig. 3.6(2)

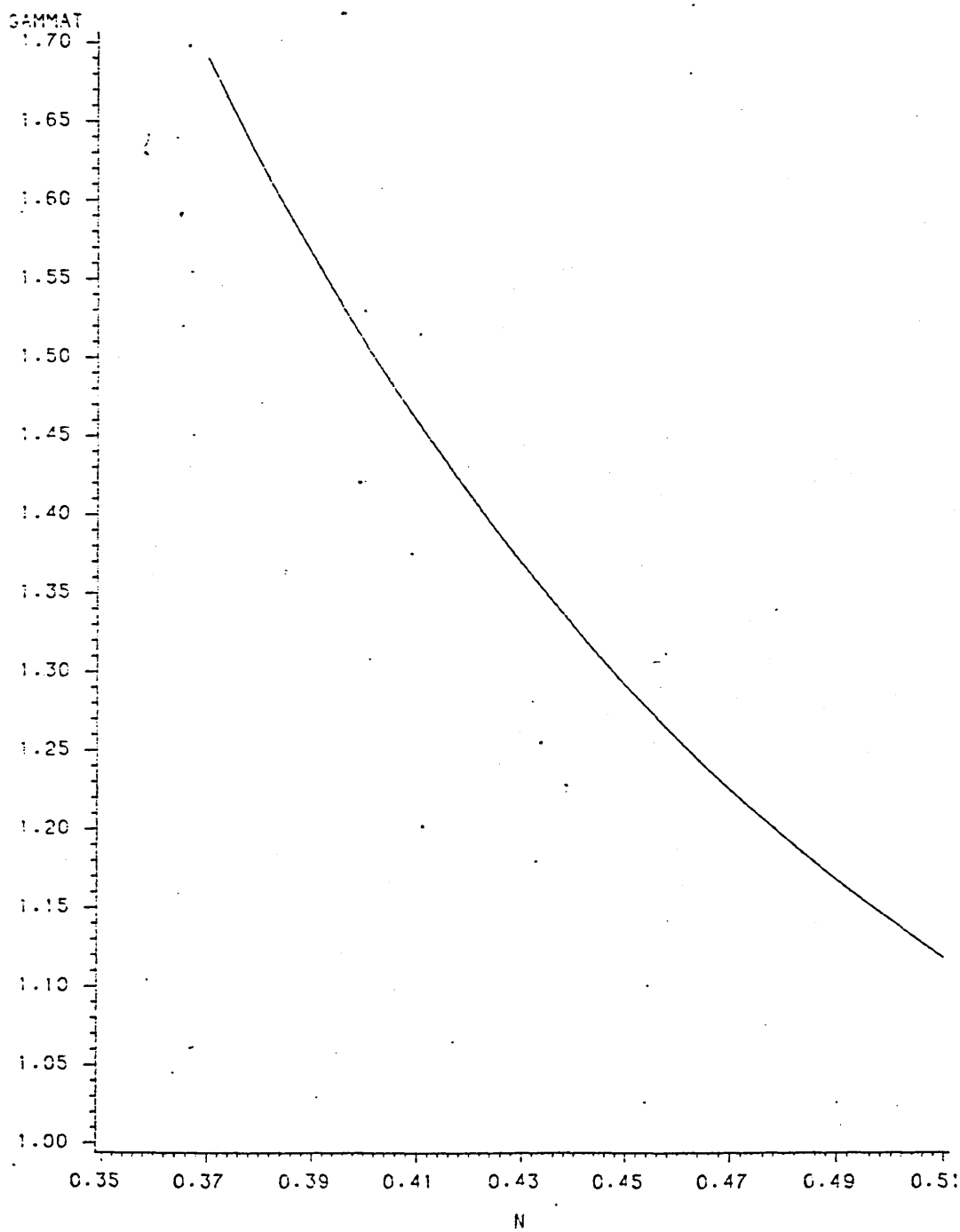
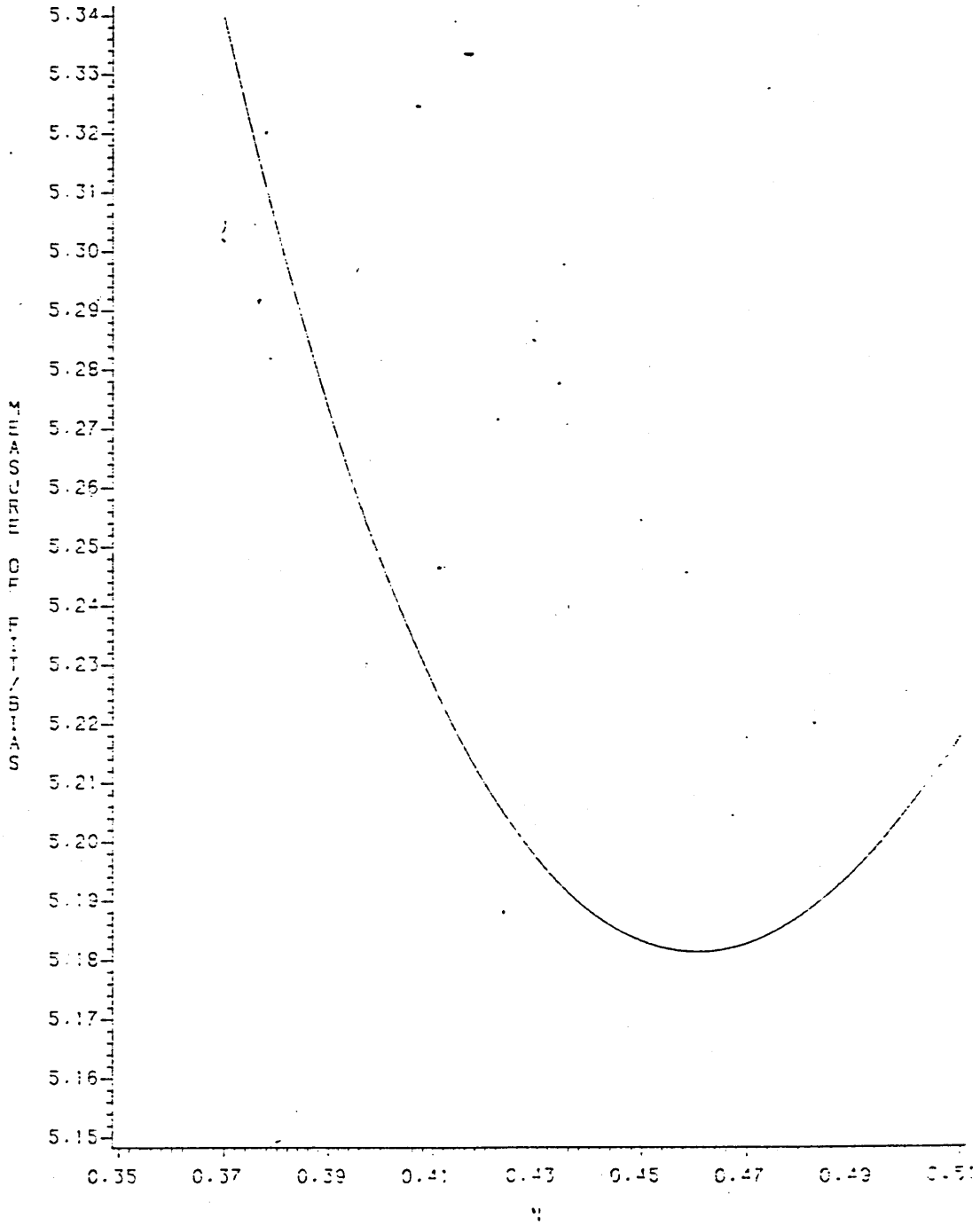

GAMMAT
1.70
1.65
1.60
1.55
1.50
1.45
1.40
1.35
1.30
1.25
1.20
1.15
1.10
1.05
1.00
0.35
0.37
0.39
0.41
0.43
0.45
0.47
0.49
0.5
N

Fig. 3.6(3)


5.34
5.33
5.32
5.31
5.30
5.29
5.28
5.27
5.26
5.25
5.24
5.23
5.22
5.21
5.20
5.19
5.18
5.17
5.16
5.15
0.35
0.37
0.39
0.41
0.43
0.45
0.47
0.49

3.7.2 Logistic Mixtures

These models are of the form

$$y = \Sigma\, \alpha_i\ f_i(t,\ \theta_i)$$

where $f_i(t,\ \theta_i)$ is the unit logistic function ie

$$\frac{1}{1 + \exp(-\beta_1(t - \beta_2))}$$

and $\Sigma\alpha_i = A$ the asymptotic value of y. They have been used in modelling human growth as will be examined in chapter 9. Examples of this type of model are the double logistic (Bock et al (1973), Preece and Baines (1978)) and the triple logistic (Bock and Thissen (1976) EL Lozy (1978), Bock and Thissen (1979)).

Preece and Baines (1978) fitted the double logistic

$$y = \frac{a}{1 + \exp(-b_1(t - c_1))} + \frac{f - a}{1 + \exp(-b_2(t - c_2))} \qquad (2)$$

to data on 35 boys and obtained the following estimates

Parameter	Mean of Estimates	Standard Deviation
t	174.1	6.0
a	148.1	5.9
b_1	.3089	.0619
c_1	2.1435	1.1070
b_2	1.0712	.1871
c_2	13.7324	.8373

The non-linearity measures were calculated based on 47 observations at times (in years)

3 3.5 4 4.5 5 5.5 6 6.5 7 7.5 8 8.5 9 9.5 10 10.5

10.75 11 11.25 11.5 11.75 12 12.25 12.5 12.75 13 13.25

13.5 13.75 14 14.25 14.5 14.75 15 15.25 15.5 15.75 16

16.25 16.5 16.75 17 17.5 18 18.5 19 20

Such data set would represent a very good range of values. Taking an estimate of σ^2 as 0.20, which is again quite good, the following figures were obtained

$\Gamma^N = .1876$ $\Gamma^T = 2.6990$ $N_\varphi = .0006$ $N_\theta = .1209$

Critical values for Γ's is 0.52 and for N_θ the values are .27 and .0027. These indicate that there is substantial non-linearity due to the parameter effects. The estimated bias in the parameters are given below

Parameter	Bias
f	.0676
a	-.0285
b_1	.0007
c_1	.0020
b_2	-.0020
c_2	.0065

Indicating that the bias is still reasonable compared with the variation between individuals.

Bock and Thissen (1979) fitted the triple logistic model

$$y = a_1 \left\{ \frac{1 - p}{1 + \exp(- b_1(x - c_1))} + \frac{p}{1 + \exp(- b_2(x - c_2))} \right\} + \frac{a_2}{1 + \exp(- b_3(x - c_3))} \qquad (3)$$

to 66 boys giving

Parameter	Mean Estimate	Standard Deviation
a_1	155.32	5.45
b_1	.82	.52
c_1	-.47	.46
b_2	.41	.06
c_2	7.12	1.41
a_2	25.28	4.31
b_3	1.14	.22
c_3	13.75	1.07

A value of p was not given, but based on the diagram of average components a reasonable value would be 0.39, this value should not be critical as far as bias calculations are concerned. This value corresponds to values given by EL Lozy (1978).

Using the design given above the following values were obtained

$\Gamma^N = .3482$ $\Gamma^T = 1.9869$ $N_\varphi = .0012$ $N^\theta = .0507$

These show substantial non-linearity. The estimated bias were

Parameter	Bias
a_1	.0350
b_1	.0006
c_1	.0014
b_2	-.0002
c_2	.0034
a_2	-.1524
b_3	-.0127
c_3	.0016
p	.0012

Again these are small relative to the size of the parameters.

3.7.3 Compound Logistics

Preece and Baines (1978) consider the logistic differential equation

$$\frac{dy}{dt} = r\, y(1 - y/k)$$

They observed that for data on child growth

$$\frac{dy}{dt} \cdot \frac{1}{k - y}$$

had a logistic shape, hence they formed the pair of equations

$$\frac{dy}{dt} = s(k - y)$$

$$\frac{ds}{dt} = \gamma(s_1 - s)(s - s_0)$$

Solving these they obtained

$$y = k - \frac{(k - y_\theta)}{\{\frac{1}{2}\exp(\gamma s_0(t - \theta) + \frac{1}{2}\exp(\gamma s_1(t - \theta))\}^{1/\gamma}} \qquad (5)$$

for $y = y_\theta$ when $t = \theta$.

They also considered two simplifications

(I) Take $\gamma = 1$

(II) Take $s = p + q$ with

$$\frac{dp}{dt} = (p_1 - p)(p - p_0) \text{ and } \frac{dq}{dt} = q(q_1 - q)$$

giving (6)

$$y = k - \frac{4(k - y_\theta)}{\{\exp(p_0(t - \theta)) + \exp(p_1(t - \theta))\}\{1 + \exp(q_1(t - \theta))\}}$$

note if $p_0 = p_1$ (I) is obtained.

Both models I and II were fitted to data on 35 boys by Preece and Baines and gave estimates

Model I			Model II		
Parameter	Mean Estimate	Std Dev	Parameter	Mean Estimate	Std Dev
k	176.6	6.0	k	174.0	5.8
y_0	162.9	5.6	y_0	164.0	5.7
s_0	.1124	.0126	p_0	.0886	.0257
s_1	1.2397	.1683	p_1	.2245	.0795
θ	14.60	.93	q_1	1.3676	.1743
			θ	14.75	.98

Using the same experimental design as in the previous section the following estimates of non-linearity were obtained

	Model I	Model II
Γ^N	.0378	.0459
Γ^T	.6049	.1766
N_φ	.00003	.00003
N_θ	.0010	.0009

These indicate that there is no significant non-linearity. Correspondingly the bias in the parameters was exceedingly small.

These models would appear to be useful, flexible generalisations of the logistic, superior to the logistic mixtures in terms of non-linearity and associated problems.

The general form of the equation was not considered as the presence of the γ

parameter would almost certainly lead to problems with non-linearity. If $s_0 = 0$ then the function reduces to a Richards function like formulation so it can be seen as a more complex form of the Richards function which has already been examined in detail and shown to suffer from problems with non-linearity.

3.8.4 Discussion of Logistic Generalisations

There are many generalisations of the logistic model. Others not discussed in this section include Cui Qiwu and Lawson's model considered in the introduction and Turner et al's generic growth model. This latter model which includes the Richards function would be too complex to be fitted meaningfully in the light of the problem encountered with the Richards function in this chapter.

The Preece and Baines models look promising and would be worth exploring in a wider context. They do have one limitation in that there is no simple logarithmic form which is often of most use in the study of growth.

3.8 Conclusions

Several points need to be made in the light of the study reported in this chapter.

1. Richards type functions, while they are flexible, are unreliable as far as the properties of the estimates are concerned.

2. The intrinsic non-linearity of the models was always considerably less than the parameter effects non-linearity. This suggests that a different

parameterisation should be used. Such a parameterisation has been found for the simple logistic (Patton and Krause (1972)) but one important aspect of these non-linear models is the interpretation of the parameters. Improved estimation of parameters of no interest to the biologist would be of little use.

3. A balance has to be achieved between the absolute fit of a model as measured by the residual sum of squares and other considerations such as bias of the estimates. In non-linear regression a small decrease in residual sum of squares may be achieved only by a large step along the solution locus, thus creating problems, particularly in interpretation, but also in other aspects such as bias, variance etc. Is the decrease worth it?

4. Other generalised logistic models need to be considered instead of the Richards.

4. Non-Linear Models with Auto Correlated Errors

4.1 Introduction

As was mentioned in 1.4 Glasbey (1979) has considered fitting logistic functions with auto correlated errors. He considers the two cases

(i) The error is an auto regressive process of order 1 ie

$$e_i = \varphi e_{i-1} + a_i$$

where a_i is independent normal zero mean and variance σ^2 .

(ii) The error has the correlation structure given by

$$\text{corr}(e_t \quad e_{t'}) = \rho^{|t-t'|}$$

By constructing the likelihood he is able to find maximum likelihood estimates in both cases. In this chapter the two error models will be extended and alternative methods of estimation will be considered.

4.2 Auto Regressive Error Models

The exact likelihood for an auto regressive process is

$$\text{Like}(\varphi, \sigma_a) = (2\pi\sigma_a^2)\left|\Sigma_n\right|^{-\frac{1}{2}} \exp - \left\{\frac{e'\Sigma_n^{-1}e}{2\sigma_a^2}\right\} \qquad \sigma_a^2\Sigma_n^{-1} = \text{var}(e), \quad \sigma_a^2 = \text{var}(a_i)$$

(Box & Jenkins (1976) p 274) which can be written as

$$\text{Like}(\varphi, \sigma_a) = \left[2\pi\sigma_a^2\right]^{-n/2}\left|\Sigma p\right|^{-\frac{1}{2}} \exp\left\{\frac{-\,s(\varphi)}{2\sigma_a^2}\right\} \qquad 4.2\ (1)$$

where

$$s(\varphi) = e_1' \Sigma_p^{-1} e_p + \sum_{i=p+1}^{n} (e_i - \varphi_1 e_{i-1} \ldots - \varphi_p e_{i-p})^2$$

and

$$\sigma_a{}^2\Sigma_p = \begin{bmatrix} \gamma_0 & \gamma_1 & \cdots & \gamma_{p-1} \\ \gamma_1 & \gamma_0 & \cdots & \gamma_{p-2} \\ \vdots & \vdots & & \vdots \\ \gamma_{p-1} & \gamma_{p-2} & \cdots & \gamma_0 \end{bmatrix}$$

the γ_i's being the theoretical auto covariances.

If

$e_i = y_i - f_i(B)$

ie the error for the model the log likelihood is given by

$$L(B, \varphi\ \sigma_a) = -\frac{n}{2} \log \sigma_a{}^2 - \frac{1}{2} \log \left|\Sigma_p\right| - \frac{s(\varphi)}{2\sigma_a{}^2} + \text{const}$$

Differentiating with respect to $\sigma_a{}^2$

$$\frac{\partial L}{\partial \sigma_a} = -\frac{n}{2\sigma_a{}^2} + \frac{s(\varphi)}{2\sigma_n{}^4}$$

giving

$$\hat{\sigma}^2 = s(\varphi)/n$$

maximising likelihood with respect to $\sigma_a{}^2$ gives

$$L(\beta, \varphi/\hat{\sigma}_a{}^2) = \text{const} - n \log s(\varphi) + \frac{1}{2} \log\left|\Sigma_p\right| \qquad 4.2\ (2)$$

In order to maximize 4.2(2) three approximations have been suggested.

4.2.1

Ignore the term

$$\frac{1}{2} \log \left|\Sigma_p\right|$$

this leads to least squares estimates.

Let $D_{ij} = e_i e_j + e_{i+1} e_{j+1} \cdots e_{n+1-j} e_{n+1-i}$

ie the sum D_{ij} contains (n+1-i-j) terms

and

$$D = \begin{bmatrix} D_{11} & -D_{12} & -D_{13} & \cdots & -D_{1,p+1} \\ -D_{12} & -D_{22} & -D_{23} & \cdots & D_{2,p+1} \\ \vdots & \vdots & \vdots & & \vdots \\ -D_{1,p+1} & -D_{2,p+1} & -D_{3,p+1} & \cdots & D_{p+1,p+1} \end{bmatrix}$$

then as

$$s(\varphi) = \sum_{i=1}^{n} \sum_{j=1}^{p} e_i e_j \gamma_{|i-j|} + \sum_{i=p+1}^{n} (e_i - \varphi_1 e_{i-1} \cdots \varphi_p e_{i-p})^2$$

It can be shown that

$$s(\varphi) = \varphi^{*\prime} D \varphi^* \quad \text{where } \varphi^* = \begin{bmatrix} 1 \\ \varphi \end{bmatrix}$$

(Box & Jenkins (1976) 275-6)

$$= \varphi' D \varphi - 2\varphi' d + D_{11}$$

where

$$d_i = D_{1,i+1}$$

Thus the minimum of $s(\varphi)$ and hence maximum of $-n \log(s(\varphi))$ is given by

$$\hat{\varphi} = D^{-1} d$$

The approximation of ignoring $\frac{1}{2} \log|\Sigma_p|$ may be reasonable if the series is long but is unlikely to be so in the growth curves considered which will be relatively short.

4.2.2

A second suggestion by Box and Jenkins is to replace D_{ij} in equation by

$$D^*_{ij} = nD_{ij}/(n+1-i-j)$$

this gives approximate maximum likelihood estimates. This method is considered by Glasbey (1980).

4.2.3

The third approach is using the Yule-Walker equations

$$\hat{\varphi} = R^{-1}r$$

R is pxp matrix $R_{ij} = r_{|i-j|}$

$$r_0 = 1, \; r_i = c_k/c_0 \quad i = 1,\ldots,p$$

$$nc_k = \sum_{n=1}^{n-k} e_i e_{i+k}$$

This is considered by Gallant and Goebel (1976).

4.2.4

All three methods can be incorporated in an iterative procedure to maximise the likelihood. Before considering that the exact likelihood will be developed.

Let

$$z_i = \begin{cases} e_i & i=1,\ldots,p \\ e_i - \hat{\varphi}_1 e_{i-1} \cdots \hat{\varphi}_p e_{i-p} & i=p+1,\ldots,n \end{cases}$$

then

$$s(\varphi) = z'\Sigma_n^{p-1}z \quad \text{where} \quad \underset{(n\times n)}{\Sigma_n^p} = \begin{bmatrix} \Sigma_p & 0 \\ 0 & I \end{bmatrix}$$

Let L be the Choleski decomposition of Σ^p_n

then if

$$L\omega = z$$

$$s(\varphi) = \Sigma\omega^2_i$$

so

$$L(\beta, \varphi|\sigma_a^2) = \text{const} - n \log \Sigma\omega_i^2 - \log|L|$$

so maximum likelihood estimates are given by minimising

$$s = \Sigma e_i^{*2} \quad \text{where} \quad e_i^* = |L|^{1/n} \omega_i$$

This is the method given by Ansley (1979).

4.2.5

The approximate procedures would involve the following steps

(1) Find initial estimates for β, $\hat{\beta}$ using OLS

(2) Calculate $\hat{e}_i = y_i - f_i(\hat{\beta})$

(3) Estimate φ by approximate method using $\hat{e}_i$'s

(4) Use $\hat{\varphi}$ to calculate z_i's and **L** and hence e_i^*'s

(5) $\min s = \Sigma e_i^{*2}$
by non-linear least-squares method with respect to β to give $\hat{\beta}$

(6) If $\hat{\beta}$ has not converged: goto 2

If $\hat{\beta}$ has converged: stop

As step 5 will involve an iterative procedure during which s will be evaluated it would be possible to incorporate steps 2, 3 and 4 into each of these substeps.

The exact likelihood method would involve the following steps

1. Find initial values for $\hat{\beta}$ and $\hat{\varphi}$
2. Minimise s using iterative procedure involving
   (i) calculate e_i's
   (ii) calculate z_i's **L** and hence e_i's
   (iii) calculate $s = \Sigma e_i^*$

   at each step.

Suitable methods are Powell's method and DuD (see Appendix A). Alternatively the minimisation could be split into two levels. This approach has been suggested by Richards (1961) and Ross (1970). The steps are

1. Find initial values for $\hat{\beta}$ and $\hat{\varphi}$ -
2. Minimise s with respect to β and φ by searching for min with respect to φ and at each step
   (i) calculate z_i's and L from current value of φ and hence e_i^*'s
   (ii) minimise $s = \Sigma e_i^{*2}$ with respect to β

This procedure is similar to that for the approximate methods. The advantage of this approach is that in minimising with respect to β it is straightforward to use a method involving gradients eg Gauss-Newton or Morquandt's method (see Appendix A) as

$$\frac{\partial s}{\partial \beta_j} = 2\Sigma e_i^* \frac{\partial e_i^*}{\partial \beta_j} = 2\left|L\right|^{1/n} \Sigma\, \omega_i \frac{\partial \omega_i}{\partial \beta_j}$$

$$\frac{\partial z_i}{\partial \beta_j} = \begin{cases} \dfrac{\partial e_i}{\partial \beta_j} = -\dfrac{\partial f_i}{\partial \beta_j} & i=1,\ldots,p \\ \dfrac{\partial e_i}{\partial \beta_j} - \varphi_1 \dfrac{\partial e_{i-1}}{\partial \beta_j} \cdots - \varphi_p \dfrac{\partial e_{i-p}}{\partial \beta_j} & \\ \quad = -\dfrac{\partial f_i}{\partial \beta_j} + \varphi \dfrac{\partial f_{i-1}}{\partial \beta_j} \cdots + \varphi_p \dfrac{\partial f_{i-p}}{\partial \beta_j} & i=p+\ldots n \end{cases}$$

and

$$\left\{\frac{\partial \omega_i}{\partial \beta_j}\right\} = L\left\{\frac{\partial z_i}{\partial \beta_j}\right\} .$$

In these calculations as the calculation of ω_i's the L matrix does not have to be formed and the calculations may be carried out iteratively. See appendix 4A for details of calculation of the Choleski decomposition of a band matrix.

Comparing the methods of 4.2.1, 4.2.2 and 4.2.3 with the exact method given in 4.2.4, there are two main differences

(i) The form of the L matrix this can be illustrated by looking at an AR(1) model.

From appendix 4A for maximum likelihood

$$L = \begin{bmatrix} \frac{1}{\sqrt{1-\varphi^2}} & 0 \\ 0 & I \end{bmatrix}$$

For method of section 4.2.1

$$L = \begin{bmatrix} \frac{\sum_{i=1}^{n-1} e_i^2}{\sum_{i=2}^{n-1} e_i^2 - \sum_{i=2}^{n} e_i e_{i-1}} & 0 \\ 0 & I \end{bmatrix}$$

for method of section 4.2.2

$$L = \begin{bmatrix} \frac{\sum_{i=2}^{n-1} e_i^2}{\sum_{i=2}^{n-1} e_i^2 - \left[\frac{n-2}{n-1}\right] \sum_{i=2}^{n} e_i e_{i-1}} & 0 \\ 0 & I \end{bmatrix}$$

and finally for method of section 4.2.3

$$L = \begin{bmatrix} \frac{\sum_{i=1}^{n} e_i^2}{\sum_{i=1}^{n} e_i^2 - \sum_{i=2}^{n} e_i e_{i-1}} & 0 \\ 0 & I \end{bmatrix}$$

(ii) The variables over which a non-linear optimisation takes place. For the three approximate methods this is the number of parameters in the functional part of the model $f_i(\beta)$ say m. These optimisation are repeated for updated values of $\hat{\varphi}$. Gallant and Goebel (1976) in fact suggest only two iterations. The maximum likelihood method requires optimising over all parameters ie m+p variables, but as has been suggested, this can be decomposed into two levels of m and p variables.

With current computing power there is no great advantage in using an approximate method to reduce the number of variables in the non-linear optimisation. Most methods are efficient for about 7 dimensions which would allow the fitting of an AP(3) and AP(4) which is adequate for the type of data examined in this project. Neither are the calculations at each iteration significantly simpler. In view of these considerations only the maximum likelihood method will be developed further.

4.2.7

An examination of the properties of fitting models with auto regressive error structures was performed using simulated data sets.

(i) For 50 observations data was generated for the model

$$y_i = 10 - \log(1 + \exp(.1 + .1 \times t)) + e_i$$

$$t=1,2,\ldots,50$$

with

I $e_i = 0.5\ e_{i-1} + a_i$

II $e_i = 0.5\ e_{i-1} - 0.5\ e_{i-2} + a_i$

III $e_i = 0.5\ e_{i-1} + 0.5\ e_{i-2} + a_i$

where a_i is a random observation from Normal distribution

Table 4.2(1)

Model $\omega_i = .5\omega_{i-1} + e$

	β_0	β_1	β_2	φ_1	φ_2	Dev	Sig
1	10.003	.0988	.0991			-318.92	
2	9.996	.0858	.1039			-343.85	
3	10.001	.1168	.0982			-314.06	
4	9.997	.1107	.1014			-350.72	
5	9.998	.0923	.1029			-345.56	
6	9.998	.1003	.0998			-342.63	
7	10.003	.0943	.0983			-339.49	
8	10.006	.0929	.0980			-337.35	
9	10.001	.0976	.1004			-332.24	
10	10.006	.0816	.0984			-301.46	
1	10.003	.0990	.0991	.0989		-319.42	NS
2	9.996	.0854	.1038	.2700		-347.80	Sig
3	10.001	.1159	.0986	.4414		-325.02	Sig
4	9.997	.1109	.1013	.1685		-352.23	NS
5	9.998	.0935	.1028	.4337		-356.44	Sig
6	9.998	.0986	.1000	.3404		-348.58	Sig
7	10.003	.0946	.0982	.3480		-346.02	Sig
8	10.006	.0936	.0978	.3721		-345.35	Sig
9	10.001	.0977	.1004	.2362		-335.24	Sig
10	10.007	.8924	.0964	.6814		-333.22	Sig
1	10.003	.0985	.0991	.0684	-.1839	-320.56	NS
2	9.996	.0867	.1038	.4054	-.4803	-359.43	Sig
3	10.001	.1163	.9085	.3659	.1234	-324.52	NS *
4	9.996	.1115	.1016	.3267	-.9943	-336.19	NS *
5	-	-	-	-	-	-	-
6	9.998	.0996	.0998	.3550	-.0725	-348.85	NS
7	9.983	.0989	.8977	.4657	-.2497	-351.80	Sig
8	10.006	.0913	.0981	.4469	-.2299	-347.02	NS
9	10.001	.0961	.1006	.1993	.0807	-335.17	NS
10	10.003	.0511	.1026	1.0000*	-.9704*	-267.74*	NS *

Table 4.2(2)

Model $\omega_t = .5\omega_{t-1} - .5\omega_{t-2} + e_t$

	β_0	B_1	β_2	φ_1	φ_2	Dev	Sig
1	10.002	.0869	.1012			-304.88	
2	10.000	.1044	.0991			-340.32	
3	9.998	.0980	.1006			-298.86	
4	10.001	.1024	.0998			-311.82	
5	10.002	.1040	.0988			-291.08	
6	9.999	.1014	.1006			-316.73	
7	10.000	.0989	.1010			-330.48	
8	10.000	.1149	.0987			-325.95	
9	10.000	.1123	.0996			-325.17	
10	10.002	.0999	.0993			-295.92	
1	10.001	.0891	.1009	.4022		-315.67	Sig
2	10.000	.1017	.0996	.2427		-343.33	NS
3	9.998	.0975	.1007	.3109		-304.00	Sig
4	10.001	.1007	.0999	.7280		-312.23	NS
5	10.002	.1054	.0984	.2455		-294.06	NS
6	9.999	.0993	.1007	.2507		-319.55	NS
7	10.000	.1011	.1008	.2821		-334.71	Sig
8	10.000	.1150	.0988	.4281		-336.56	Sig
9	10.001	.1116	.0995	.1899		-327.17	NS
10	10.001	.0985	.0998	.2669		-299.54	NS
1	10.001	.0836	.1015	-.7031	-.5729	-336.91	Sig
2	10.000	.1051	.0989	.3263	-.4074	-352.42	Sig
3	10.000	.0994	.1002	.5215	-.6596	-331.39	Sig
4	10.000	.1034	.0999	.1880	-.5633	-333.60	Sig
5	10.001	.1055	.0990	.3954	-.6998	-327.49	Sig
6	9.999	.1019	.1008	.3716	-.5467	-336.89	Sig
7	10.000	.0980	.1011	.4811	-.6060	-360.91	Sig
8	10.000	.1152	.0988	.6437	-.5031	-348.49	Sig
9	10.000	.1121	.0996	.3243	-.6075	-349.42	Sig
10	10.002	.1001	.0991	.4285	-.5970	-319.53	Sig

Table 4.2(3)

Model $\omega_t = .5\omega_{t-1} + .5\omega_{t-2} + e_t$

	β_0	β_1	β_2	φ_1	φ_2	Dev	Sig
1	9.997	.1336	.0945			-323.11	
2	9.970	.1531	.1154			-328.23	
3	10.007	.1036	.0960			-334.13	
4	9.989	.1070	.1041			-334.90	
5	9.964	.1602	.1000			-342.97	
6	10.012	.0925	.0978			-268.18	
7	9.976	.1543	.1025			-347.69	
8	10.002	.0478	.1045			-309.87	
9	9.990	.1100	.0970			-327.82	
10	9.992	.1379	.1005			-339.20	
1	9.980	.1237	.0956	.6207		-343.78	Sig
2	9.970	.1598	.1144	.5678		-347.19	Sig
3	10.007	.1034	.0946	.3512		-340.88	Sig
4	9.989	.1051	.1043	.4159		-343.46	Sig
5	9.965	.1608	.0998	.1437		-343.96	NS
6	10.009	.1034	.0982	.8524		-330.33	Sig
7	9.976	.1542	.1025	-.0604		-347.89	NS
8	10.019	.0500	.1047	.6249		-333.35	Sig
9	9.989	.1127	.0970	.3921		-335.41	Sig
10	9.992	.1376	.1005	.0717		-339.46	NS
1	9.980	.1265	.0957	.4745	.2462	-345.58	NS
2	9.970	.1607	.1138	.3503	.3558	-351.80	Sig
3	10.007	.1029	.0946	.2687	.1619	-341.26	NS
4	9.989	.1047	.1041	.3058	.2880	-347.53	Sig
5	9.967	.1682	.0978	.6767	.5930	-359.17	Sig
6	10.009	.1016	.0970	.5133	.3938	-336.39	Sig
7	9.975	.1555	.1029	-.1021	.5212	-357.54	Sig
8	10.019	.0520	.1047	.4199	.3037	-335.97	NS
9	9.989	.1149	.0968	.2755	.3040	-338.68	NS
10	9.912	.1381	.1003	-.0129	.4319	-349.15	Sig

Each set of conditions was replicated 10 times and the models fitted with

A independent error

B AR(1) error

C AR(2) error

The results are shown in table 4.2.1 (1), (2) and (3).

The deviance is the maximised value of

$-2n \log s(\beta, \varphi)$

and the difference between the deviance in nested models is the $-2 \log \ell$ where ℓ is the likelihood ratio statistic and hence will approximately have a χ^2 distribution with df given by the difference in the number of fitted parameters. The sig column indicated whether the test with the simpler model is significant at the 5% level.

Table 4.2(1)

(i) The test for an AR(1) model vs random error model was significant 8 out of 10 times.

(ii) When trying to fit an AR(2) model problems in convergence were encountered, all values marked * are suspect.

(iii) There are only small differences in the estimates of the β parameters between the three methods, the same pattern exists in each method.

Table 4.2(2)

(i) The AR(1) was significantly better than the random error model only 4 out of 10 times but the AR(2) was better than the AR(1) 9 out of 10 times.

(ii) Again the β values are only slightly affected by the chosen error structure.

Table 4.2(3)

(i) The AR(1) is significantly better than the random error model in 7 out of 10 cases and the AR(1) is significantly better than the AR(1) in 6 out of 10 cases.

(ii) The β estimates are again only slightly affected by the fitting method but there is a much higher variation in the estimates with this structure than with the other two.

Two simulations were carried out to test the effect of fitting the wrong model. Here the data was generated for the model

$$y_i = 10 + 0.5 \log(1 + \exp(0.1 + 0.1t)) + e_t \qquad t=1,2,\ldots,50$$

with

$$\text{I} \quad e_t = a_t$$

$$\text{II} \quad e_t = 0.5e_{t-1} + a_t$$

a_t as before.

The results for 10 runs for each model of fitting the same models as before are given in tables 4.2(4) and 4.2(5).

Table 4.2(4)

(i) In no cases is the AR(1) significantly better than the random model, nor the AR(2).

(ii) The β estimates give consistant results.

Table 4.2(5)

(i) 7 out of 10 times the AR(1) gives a better fit than the random model, the AR(2) is never significantly better than the AR(1).

(ii) Again estimates are consistent for the β parameters.

Table 4.2(4)

Random Error Incorrect Model

	β_0	β_1	β_2	φ_1	φ_2	Sig	Dev
1	10.004	0.9352	.0890				-339.98
2	10.003	.9619	.0890				-335.49
3	10.002	.9277	.0951				-344.25
4	10.002	.9457	.0917				-331.13
5	10.002	.9466	.0915				-338.49
6	10.004	.9461	.0892				-341.69
7	10.000	.9477	.0930				-342.20
8	10.000	.9652	.0897				-341.62
9	9.999	.9459	.0960				-338.73
10	10.003	.9533	.0890				-329.27
1	10.004	.9354	.0888	-.1575		NS	-341.10
2	10.003	.9620	.0890	.0810		NS	-335.81
3	10.002	.9274	.0950	-.0454		NS	-344.30
4	10.001	.9449	.0918	-.0083		NS	-331.11
5	10.001	.9447	.0919	-.2412		NS	-341.15
6	10.004	.9463	.0892	-.1418		NS	-342.65
7	10.000	.9483	.0929	.1272		NS	-343.03
8	10.002	.9648	.0900	-.0996		NS	-347.02
9	9.999	.9449	.0962	-.1421		NS	-339.71
10	10.003	.9535	.0892	.0857		NS	-329.63
1	10.004	.9357	.0888	-.1987	-.0665	NS	-341.98
2	10.003	.9623	.0891	.0857	-.1325	NS	-336.37
3	10.002	.9272	.0950	-.0707	-.2156	NS	-346.63
4	10.002	.9439	.0919	.0015	-.0741	NS	-331.32
5	10.002	.9480	.0915	-.1873	-.04773	NS	-340.38
6	10.004	.9457	.0892	-.0865	-.1083	NS	-342.55
7	10.000	.9461	.0933	.1277	-.1443	NS	-343.82
8	10.003	.9645	.0897	-.0959	-.0240	NS	-341.90
9	9.999	.9458	.9603	-.1991	-.1541	NS	-341.62
10	10.003	.9524	.0891	.0687	-.1277	NS	-330.17

Table 4.2(5)

Model $w_t = .5\omega_t + e_t$ Incorrect Model

	β_0	β_1	β_2	φ_1	φ_2	Dev	Sig
1	10.002	.9861	.0885			-358.20	
2	9.996	.9331	.0985			-342.41	
3	10.003	.9138	.0958			-321.65	
4	10.002	.9715	.0871			-311.72	
5	10.007	.9360	.0903			-333.09	
6	10.002	.9782	.0900			-326.29	
7	10.000	.9105	.0967			-316.82	
8	10.002	.9425	.0889			-314.05	
9	10.000	.9174	.0941			-337.86	
10	9.998	.9180	.0965			-321.47	
1	10.002	.9871	.0885	.2137		-360.64	NS
2	9.996	.9344	.0982	.2938		-346.84	NS
3	10.005	.9171	.0940	.5190		-336.41	Sig
4	10.002	.9760	.0865	.2495		-314.76	NS
5	10.007	.938	.0900	.2406		-336.69	NS
6	10.002	.942	.0895	.3125		-332.42	Sig
7	10.000	.920	.0956	.5724		-335.80	Sig
8	10.002	.9433	.0896	.4459		-324.41	Sig
9	10.000	.9118	.0942	.3380		-343.70	Sig
10	9.998	.919	.0973	.4256		-331.12	Sig
1	10.002	.9858	.0885	.1813	.0922	-360.73	NS
2	9.996	.9302	.0990	.3341	-.2149	-348.67	NS
3	10.003	.9161	.0958	.5622	-.1614	-336.94	NS
4	10.002	.9686	.0873	.2279	.0408	-315.28	NS
5	10.006	.9324	.0907	.2401	.0360	-336.70	NS
6	10.003	.9507	.0873	.7173	-.9884	-303.70	NS
7	9.999	.9040	.0971	.4927	.0799	-336.03	NS
8	10.003	.9413	.0905	.96675	-1.000*	-328.43*	NS
9	10.000	.9152	.0943	.4104	-.2660	-345.84	NS
10	9.998	.9235	.0968	.4371	-.0495	-330.85	NS

Table 4.2(5)

Further simulations were carried out for the true model and the estimates of φ are displayed below for the model $e_i = 0.5e_{i-1} + a_i$

ie $\varphi = 0.5$

$\hat{\varphi}$

.1 0 7

.2 1 3 7 9

.3 4 4 5 7

.4 2 3 3 3 4 4 5

.5 9

.6 9 8

Mean = .3847 SD = .1580.

95% confidence interval for the mean is .31 to .46.

This shows that the estimate is biased, there is also a suggestion of skewness in the distribution.

To investigate this further a sample of size 100 was used again with an AR(1) model with $\varphi = .5$. The results were

$\hat{\varphi}$

.2 4 7

.3 1 6

.4 0 1 2 3 4 4 5 6 7 8 9

.5 0 1 1 6

The results are better but there is still a small bias and evidence of a slight skew.

For a smaller sample size of twenty five the following estimates of φ were obtained

$\hat{\varphi}$

-.1 2 6

-.0 0 2 3 3 5 5 9

+.0 2 3 3 4 5 6

.1 0 1 1 3 4 5 7 7 8 9

.2 0 0 1 1 1 3 3 4 4 4 5 5 6 7 9

.3 0 0 2 2 2 2 3 4 4 4 4 5 6 6 8 8 8 8 9 9 9 9 9 9

.4 0 0 1 1 1 1 1 1 1 1 1 2 2 3 3 3 4 5 6 6 6 8 9

.5 0 0 3 4 5 5 5 5

.6 1 3 3 3 5

Mean .301 SD .186

95% confidence interval for mean is .265 to .337.

This shows an increasing bias with small sample size.

Finally an investigation of the situation when the sample size was only 10 was carried out. In the situation the error model was again

$$e_t = 0.5e_{t-1} + a_t$$

(i) The AR(1) was only significantly better than the random error model in 4 out of 5 cases.

(ii) The estimate of φ was very poor as illustrated below.

$\hat{\varphi}$

-.8 4

-.7 7 3

-.6 5 8

-.5 0 3 3

-.4 0 5 7 8

-.3 0 1 2 3 5 9 9

-.2 5 6 9

-.1 0 1 3 6 6 8 9

-.0 4 5 5 6 8

+.0 3 6 8

.1 1 5 6 8

.2 6 8

.3 4

.4 1 4

.5 0

The estimates of β_2 and β_3 were also examined when fitted in model assuming random error structure when it was an AR(1) with $\varphi = 0.5$. These results were

True value $\beta_2 = 0.1$

$\beta_3 = 0.5$

$\hat{\beta}_2$

0.04 4

0.05 1 1 8

0.06 3 3 3 5 6

0.07 0 2 3 4 6 6 7 7 7 8

0.08 0 0 2 3 3 3 3 4 6 6 6 8 9

0.09 3 3 4 4 4 4 5 6 7 8 9 9 9 9

0.10 0 0 2 2 2 2 3 6 6 6 8 9 9 9

0.11 0 2 3 4 4 5 5 6 7 8 9

0.12 0 1 2 2 2 4 6 7

0.13 0 2 3 3 4 5 6 8

0.14 0 1 1 1 2 3 7 9

0.15 2 5 5 8

0.16 0 4 4

$\hat{\beta}_3$

.45 0 8

.46 1 2 3 7 9

.47 0 1 1 2 3 3 3 4 6 8 8 9

.48 1 1 4 5 5 5 7 7 7 7 8 8 9 9 9

.49 1 1 3 3 3 3 3 4 4 4 5 5 5 5 6 6 6 6 7 7 9 9 9 9

.50 0 0 0 0 2 3 3 3 4 4 5 5 5 5 6 7 8 8 9 9

.51 0 0 1 1 2 4 4 4 7 9

.52 0 0 1 3 7

.53 3 3 5 5 7 7 8

The asymptotic variances with random error would be 0.020 for β_2 and 0.014 for β_3 using the formula

$$\text{Var} = (v' v)^{-1}\sigma^2 \quad \text{where} \quad v_{ij} = \frac{\partial f_i}{\partial \beta_j}$$

As can be seen there is a reasonable correspondance between the observed distribution of $\hat{\beta}_1$ and $\hat{\beta}_2$ and the expected using the above formulae.

The following conclusions can be drawn from the above simulations, allowing for an amount of uncertainty due to the limited number of simulations carried out

(i) The inclusion of an auto regressive term in the model does not greatly affect the estimates of the parameters of the functional part of the model ie the β's. There will be a reduction in the estimated variance of the parameters (see Gallant and Goebel 1976) if the AR model is fitted since the estimated variance will be smaller (reduced residual sums of squares), but for small samples this will be unreliable (see (iii)). The observed variances of β correspond to what would be expected if the AR terms were ignored and so the fitting seems reasonably robust to auto correlated errors.

(ii) For medium size samples (about 50) the likelihood ratio test provides a reasonable test for the inclusion of AR terms. For small samples (say 10) this test does not perform at all well.

(iii) The estimates of the AR parameters show small sample bias. It is only with large samples (> 100) that the bias is sufficiently small, for small samples it is very bad. Indeed the estimates also have a very large variance.

The simulations and hence the conclusions have concentrated on small samples as the samples used in biological growth tend to be at most 50. The poor performance of the method for these sample sizes however should not discourage its use for larger samples where it looks more promising, but further investigation would be needed.

4.3 Other Error Structures

In addition to AR models full ARMA models could be considered. There are several algorithms for calculating the likelihood for an ARMA model Newbold (1974), Dent (1977) and Pearlman (1980) have produced algorithms but Ansley's (1979) method appears to be the most efficient (Pearlman (1980)) for smaller models. Anleys method was used in the previous chapter to fit non-linear models with auto regressive errors it would also be used to fit MA and ARMA models. The difference would be in the structure of the L matrix, the Σ^{n}_{pq} matrix for the covariance of the modified errors eg

$$z_i = \begin{cases} e_i & i=1,\ldots,p \\ e_i - \varphi_1 e_{i-1} - \varphi_p e_{i-p} & i=p+1,\ldots,n \end{cases}$$

would be of the form

$$\begin{bmatrix} \Omega_p & 0 \\ 0 & \beta \end{bmatrix}$$ where β is a banded matrix of width q.

ie a pxp matrix of the covariances of z_i, i=1,...,p and a band matrix of width q where q is the number of MA parameters.

Details for individual model are given in appendix 4A.

This approach was not developed for the following reasons

(i) When attempting to fit such models problems with convergence were encountered. It was important to obtain good initial values before starting the iterative phase. Pure auto regressive models seem more robust in their convergence to the initial values.

(ii) Considering the short series that will be modelled complex ARMA models could not be considered therefore there would be less to gain by expanding from the simpler AR models.

(iii) The interpretaiton of an MA model would be less obvious than an AR model, see later, although a case could be made for an ARMA(1,1) model treated as the sum of an AR(1) + random error.

4.4 Standard Errors of Parameter Estimates

The estimates of standard errors can be obtained by using

$$\text{var}(\hat{\beta},\hat{\sigma}^2,\hat{\varphi},\hat{\theta}) = -I^{-1}$$

where I is the matrix of the second derivatives of the log likelihood given by 4.2(1).

The main interest will be in the β parameters and so the variances could be approximated by conditioning on the estimated values of the nuisance parameters φ and θ. Given φ and θ the problem of estimating β is a weighted least squares problem. The sum of squares

$$s_c = \Sigma\omega_i^2$$

is to be minimised with respect to β. Let T be the nxn matrix such that

$$z = Te$$

then

$$\omega = L^{-1}Te$$

Thus using the result used in section 4.2.4, with $\mathbf{v}$ as the nxp matrix of derivatives of f with respect to β then

$$\text{var}(\hat{\beta}) \simeq [\mathbf{v}'(\mathbf{T}'\mathbf{L}'^{-1})(\mathbf{L}^{-1}\mathbf{T})\mathbf{v}]^{-1}\sigma^2$$

Investigation of the properties of these estimated variances is not within the scope of this project.

4.4 Models for Continuous Time

In this section the model

$$y_t = f_t(\beta) + e_t$$

where

$$E(e_t e_{t+\tau}) = \rho(\tau\,\theta), \qquad \theta \text{ unknown parameters}$$

is considered.

The log likelihood for the model is

$$L(\beta\theta) = -\frac{n}{2}\log 2\pi - \frac{1}{2}\log\left|\Sigma\right| - \frac{n}{2}\log\sigma^2 - \frac{1}{2}\sigma^2\, e'\Sigma^{-1}e$$

where $\Sigma\,\sigma^2$ is the var-cov matrix of e

now

$$|\Sigma| = \begin{vmatrix} 1 & \rho(\tau_1) & \rho(\tau_1+\tau_2) & \cdots & \rho(\tau_1+\tau_2 \cdots \tau_{n-1}) \\ \rho(\tau_1) & 1 & \rho(\tau_2) & \cdots & \rho(\tau_2 \cdots \tau_{n-1}) \\ \rho(\tau_1+\tau_2) & \rho(\tau_2) & 1 & & \\ & & & & \\ & & & & \rho(\tau_{n-1}) \\ \rho(\tau_1+\tau_2\ldots\tau_{n-1}) & & & & 1 \end{vmatrix}$$

where $\tau_i = t_{i+1} - t_i$

If $\rho(\tau+\tau') = \rho(\tau).\rho(\tau')$

then $|\Sigma| = \begin{vmatrix} 1 & \rho(\tau_1) & \rho(\tau_1+\tau_2) & \cdots & \rho(t_1+\tau_2 \cdots \tau_{n-1}) \\ 0 & 1-\rho(2\tau_1) & & & \\ 0 & 0 & 1-\rho(2\tau_2) & & \\ & & & & \\ 0 & 0 & & & 1-\rho(2\tau_{n-1}) \end{vmatrix}$

$$= \prod_{i=1}^{n-1} (1-\rho(2\tau_i)) = \Delta$$

The cofactors of Σ are

$$c_{11} = \Delta$$

$$c_{ii} = \Delta/(1-\rho(2\tau_{i-1}))$$

For c_{ij} $\quad i < j-1$

$$c_{ij} = \begin{vmatrix} 1 & \rho(\tau_1) & & & & & & \\ & 1 & & & & & & \\ & & & & & & & \\ i-1 & & \rho(\tau_{i-2}) & 1 & \rho(\tau_{i-1}) & & & \\ i+1 & & & & \rho(\tau_i) & 1 & & \\ & & & & & \rho(\tau_{i+1}) & 1 & \\ & & & & & & & \\ J & & & & & & \rho(\tau_{i-1}) & 1 \\ J+1 & & & & & & 1 & \\ & & & & & & & \\ & & & & & & & 1 \end{vmatrix}$$

$$= \begin{vmatrix} 1 & \rho(\tau_i) & & & & & \\ & 1-\rho(2\tau_1) & & & & & \\ & & & & & & \\ i-1 & & 1-\rho(2\tau_{i-2}) & & & & \\ & & & - & 0 & & \\ & & & & 0 & & \\ J & & & & & & \\ J+1 & & & & & 0 & \\ & & & & & & 1-\rho(2\tau_{j-1}) \end{vmatrix}$$

$= 0$ (ie zeros on diagonal of triangular matrix)

By symmetry $c_{ji} = 0 \qquad i < j-1$

$$c_{ii} \begin{vmatrix} 1 & \rho(\tau_1 & & & & \\ & & & & & \\ i-1 & \rho(\tau_{1-2}) & 1 & \rho(\tau_{i-1}+\tau_i) & & \\ i+1 & & \rho(\tau_1+\tau_{i-1}) & 1 & \rho(\tau_{i+1}) & \\ & & & & 1 & \end{vmatrix}$$

$$= \begin{vmatrix} 1 & \rho(\tau_1) & & & \\ 0 & 1-\rho(2\tau_i) & & & \\ & & & & \\ i-1 & & 1-\rho(2\tau_{i-2}) & & \\ i+1 & & & 1-\rho(2\tau_i+2\tau_{i-1}) & \\ & & & & 1-\rho(2\tau_{i+1}) \end{vmatrix}$$

$$= \frac{\Delta\ (1-\rho(2\tau_i+2\tau_{i-1}))}{(1-\rho(2\tau_{i-1}))(1-\rho(2\tau_i)}$$

$$
c_{i\,i+1} = \begin{vmatrix}
 & 1 & 2 & & i-1 & & i & & i+2 \\
1 & 1 & \rho(\tau_1) & & & & & & \\
 & & & & & & & & \\
i-1 & & & & 1 & & \rho(\tau_{i-1}) & & \rho(\tau_{i-1}+\tau_{i+1}) \\
i+1 & & & & \rho(\tau_i+\tau_{i-1}) & & \rho(\tau_i) & & \rho(\tau_{i+1}) \\
i+2 & & & & & & \rho(\tau_{i+1}+\tau_i) & & 1
\end{vmatrix}
$$

$$
= \begin{vmatrix}
1 & \rho(\tau_1) & & & \\
 & 1-\rho(2\tau_2) & & & \\
 & & & & \\
i-1 & & 1-\rho(2\tau_{i-2}) & & \\
i+1 & & & \rho(\tau_i) - \rho(2\tau_i + \tau_{i+1}) & \\
 & & & & 1-\rho(2\tau_{i+1})
\end{vmatrix}
$$

$$
= \frac{\triangle \rho(2\tau_i)(1-\rho(2\tau_{i-1})}{(1-\rho(2\tau_{i-1})(1-\rho(2\tau_i)} = \frac{\triangle \rho(2\tau_i)}{1-\rho(2\tau_i)}
$$

Hence Σ^{-1} is a tridiagonal matrix with ith row

$$
0 \ldots 0, \frac{-\rho(2\tau_{i-1})}{1-\rho(2\tau_{i-1})}, \frac{1-\rho(2\tau_i+2\tau_{i-1})}{(1-\rho(2\tau_{i-1})(1-\rho(2\tau_i)}, \frac{-\rho(2\tau_i)}{(1-\rho(2\tau_i)}, 0 \ldots 0
$$

Thus

$$
L = \text{const} - \frac{1}{2} \sum_{i=1}^{n} \log (1-\rho(2\tau_i)) - \frac{\sigma^2}{2}\left\{\frac{e_1^2}{1-\rho(2\tau_i)} + \frac{e_n^2}{1-\rho(2\tau_{n-1})}\right.
$$

$$
+ \sum_{i=2}^{n-1} e_i^2 \,(1-\rho(2\tau_i+2\tau_{i-1})/(1-\rho(2\tau_{i-1}))(1-\rho(2\tau_i))
$$

$$
\left. - 2 \sum_{i=2}^{n} e_i\, e_{i-1}\, \rho(2\tau_{i-1})/(1-\rho(2\tau_{i-1}))\right\}
$$

If

$$
\rho(\tau_i) = \rho^{\tau_i}
$$

then this reduces to the model and likelihood discussed by Glasbey (1979).

Other functions could be used for example

$$\rho(\tau) = \exp(-\lambda_1 |\tau| - \lambda_2 |\tau|^2)$$

However in the light of the previous discussions this approach will not be developed further.

Choleski decomposition of a band matrix

1. Band width 1

$$\begin{bmatrix} a_{11} & a_{12} & & & \\ a_{21} & a_{22} & a_{23} & & \\ & a_{32} & a_{33} & & a_{n-1,n} \\ & & & a_{n,n-1} & a_{nn} \end{bmatrix} = \begin{bmatrix} \ell_{11} & & & & \\ \ell_{21} & \ell_{22} & & & \\ & \ddots & \ell_{32} & & \\ & & & \ell_{n,n-1} & \ell_{nn} \end{bmatrix} = \begin{bmatrix} \ell_{11} & \ell_{21} & & & \\ & \ell_{22} & \ell_{32} & & \\ & & & & \\ & & & & \ell_{nn} \end{bmatrix}$$

$$\ell_{ii}^2 = a_{ii} = \ell_{i,i-1}^2 \qquad i < 2 \; ; \; \ell_{11}^2 = a_{11}$$

$$\ell_{i,i-1} = \frac{1}{\ell_{1-i,i-1}} \left\{ a_{i,i-1} \right\}$$

$$L' \; x = b \qquad x_1 = b_1/\ell_{11}$$

$$x_i = \frac{1}{\ell_{ii}} (b_i - \ell_{i,i-1} x_{i-1})$$

2. Band width 2

$$\begin{bmatrix} a_{11} & a_{12} & a_{13} & & \\ a_{21} & a_{22} & a_{23} & a_{24} & \\ a_{31} & a_{32} & a_{33} & a_{34} & a_{35} \\ & a_{42} & a_{43} & a_{44} & a_{45} & a_{46} \end{bmatrix} = \begin{bmatrix} \ell_{11} & & & & \\ \ell_{21} & \ell_{22} & & & \\ \ell_{31} & \ell_{32} & \ell_{33} & & \\ & \ell_{42} & \ell_{43} & \ell_{44} & \end{bmatrix}$$

$$\begin{bmatrix} \ell_{11} & \ell_{12} & \ell_{13} & & \\ & \ell_{22} & \ell_{23} & \ell_{24} & \\ & & \ell_{33} & \ell_{34} & \ell_{35} \\ & & & \ell_{44} & \end{bmatrix}$$

$$\ell_{11}^2 = a_{11} \; , \; \ell_{21} = a_{21}/\ell_{11}$$

$$\ell_{22}^2 = a_{22} - \ell_{21}^2 \; , \; \ell_{31} = a_{31}/\ell_{11} \; , \; \ell_{32} = \frac{1}{\ell_{22}} \left\{ a_{32} - \ell_{12}\ell_{31} \right\}$$

$$\ell_{ii}^2 = a_{ii} - \left\{ \ell_{i,i-1}^2 + \ell_{i,i-2}^2 \right\}$$

$$\ell_{i,i-1} = \frac{1}{\ell_{i-1,i-1}} \left\{ a_{i,i-1} - \ell_{i,i-2}\ell_{i-1,i-2} \right\}$$

$$\ell_{i,i-2} = a_{i,i-2}/\ell_{i-2,i-2}$$

$L'x = b$. $\quad x_1 = b_1/\ell_{11}$, $x_2 = \frac{1}{\ell_{22}}(b_2 - \ell_{21}x_1)$

$$x_i = \frac{1}{\ell_{ii}}(b_i - \ell_{i,i-1}x_{i-1} - \ell_{i,i-2}x_{i-2})$$

AR(1) $\quad \omega_t = \varphi\omega_{t-1} + a_t$

$$z_1 = \omega_1$$

$$z_i = \omega_i - \varphi\omega_{i-1}$$

$$\gamma_0 = \frac{1}{1-\varphi^2} \quad \Omega = \begin{bmatrix} \gamma_0 & \\ & I \end{bmatrix}$$

$$L = \begin{bmatrix} \frac{1}{\sqrt{1-\varphi^2}} & \\ & I \end{bmatrix}$$

so $\quad e_1 = z_1/\sqrt{\varphi_0} = \omega_1 \sqrt{1-\varphi^2}$

$$e_i = z_i = \omega_i - \varphi\omega_{i-1}$$

$$|L| = \frac{1}{\sqrt{1-\varphi^2}} \quad .$$

AR(2)

$$z_1 = \omega_1$$

$$z_2 = \omega_2$$

$$z_i = \omega_i - \varphi_1\omega_{i-1} - \varphi_2\omega_{i-2}$$

$$\gamma_0 = \left[\frac{1-\varphi_2}{1+\varphi_2}\right] \frac{1}{(1-\varphi_1-\varphi_2)(1+\varphi_1-\varphi_2)}$$

$$\gamma_1 = \frac{\varphi_1}{1-\varphi_2}\gamma_0 \quad , \quad \varphi_2 = \left[\varphi_2 + \frac{\varphi_1^2}{1-\varphi_2}\right]\gamma_0$$

$$\Omega = \begin{bmatrix} \gamma_0 & \gamma_1 & & \\ \gamma_1 & \gamma_0 & & \\ & & 1 & \\ & & & 1 \end{bmatrix} \qquad \begin{aligned} \ell_{11} &= \sqrt{\gamma_0} \\ \ell_{12} &= \gamma_1/\sqrt{\gamma_0} \\ \ell_{22} &= \sqrt{\gamma_0} - \gamma_1^2/\gamma_0 \end{aligned}$$

$$e_1 = z_1/\sqrt{\gamma_0} = \omega_1/\sqrt{\gamma_0}$$

$$e_2 = (z_2 - (\gamma_1/\sqrt{\gamma_0}).(z_1/\sqrt{\gamma_0}))/\sqrt{\gamma_0} - \gamma_1^2/\gamma_0)$$

$$= (z_2 - \gamma_1/\gamma_0\, z_1)/\sqrt{\gamma_0} - \gamma_1^2/\gamma_0$$

$e_i = z_i$

MA(1)

$$z_t = \omega_t$$

$\gamma_0 = (1+\theta^2)\sigma_a^2 \qquad \gamma_1 = \theta\sigma_a^2$

$\ell_{11} = \sqrt{1+\theta^2} \qquad \ell_{ii}^2 = (1+\theta^2) - \ell_{i,i-1}^2$

$\ell_{i,i-1} = \theta/\ell_{i+i-1}$

$e_1 = \omega_t/\sqrt{1+\theta^2}$

$e_t = \frac{1}{\ell_{tt}}(\omega_t - \ell_{t,t-1}e_{t-1})$

MA(2)

$$\gamma_0 = (1 + \theta_1^2 + \theta_2^2)\sigma_a^2$$

$$\gamma_1 = (\theta_1 + \theta_2\theta_1)\sigma_a^2$$

$$\gamma_2 = \theta_2\sigma_a^2$$

$$\ell_{11}^2 = (1 + \theta_1^2 + \theta_2^2) \ , \ \ell_{21} = \frac{\theta_1 + \theta_2\theta_1}{\sqrt{1 + \theta_2^2 + \theta_2^2}}$$

$$\ell_{22}^2 = 1 + \theta_1^2 + \theta_2^2 - \frac{\theta_1 + \theta_2\theta_1}{1 + \theta_1^2 + \theta_2^2}$$

$$\ell_{ii}^2 = (1 + \theta_1^2 + \theta_2^2) - (\ell_{i,i-1}^2 + \ell_{i,i-2}^2)$$

$$\ell_{ii-1} = ((\theta_1 + \theta_2\theta_1) - \ell_{i,i-2}\ell_{i-1,i-2})/\ell_{i-1,i-1}$$

$$\ell_{i,i-2} = \theta/\ell_{i-2,i-2}$$

$e_1 = \omega_1/\ell_{11}$

$e_2 = (\omega_2 - \ell_{21}e_1)/\ell_{22}$

$e_t = (\omega_t - \ell_{t,t-1}x_{t-1} - \ell_{t,t-2}x_{t-2})/\ell_{tt}$

ARMA (1, 1)

$$z_1 = \omega_1 = (\varphi\omega_0 + \theta a_0) + a_1$$

$$z_t = \omega_t - \varphi\omega_{t-1} \qquad t > 1$$

$$= a_t - \theta a_{t-1}$$

$$\Omega = \begin{vmatrix} \sigma_\omega^2 & \theta & \\ \theta & 1+\theta^2 & \theta \\ & \theta & \end{vmatrix} \qquad E(z_1 z_2) = \theta \text{ var}(a) = \theta\sigma_a^2$$

$$\sigma_\omega^2 = \frac{1 + 2\theta\varphi + \theta^2}{1 - \varphi^2}\,\sigma_a^2$$

$$\ell_{11} = \frac{1 + 2\theta\varphi + \theta^2}{1 - \varphi^2} \qquad \ell_{21} = \theta/\ell_{11}$$

$$\ell_{ii}^2 = 1 + \theta^2 - \ell_{i,i-1}^2 \qquad \ell_{i,i-1} = \theta/\ell_{i-1,i-1}$$

$$e_i = \omega_1/\ell_{11}$$

$$e_t = (z_t - \ell_{t,t-1}e_{t-1})/\ell_{tt}$$

5. Dynamic Models of Growth

5.1 Introduction

In this chapter models which describe how the size of an organism is changing will be examined. Two classes of model will be considered

(a) stochastic difference equations,

and

(b) stochastic differential equations.

The former can be considered as approximations to the latter formed in discrete time rather than continuous time. These difference models may be more tractable than the differential models and may also be more realistic at the level of complexity available. For example the growth of a plant is continuous in continuous time but over a short period of time will vary in a complex manner due to temperature and light fluctuations. These variations may well average out over a longer period of time thus making, say, weekly growth increments, in a sense, simpler to model. Also all measurements will be in discrete time and the model will directly match the observations.

The emphasis in this chapter will be on deriving results that will be useful in fitting and validating the models rather than the more usual concern of model behaviour.

5.2 Stochastic Difference Equation Models

Both linear and non-linear equations will be considered in this section. A key tool for the estimation is the Kalman Filter, this will be described in more detail and subsequently applied to both linear and non-linear models.

5.2.1 Some Simple Linear Difference Models

The first order deterministic linear difference model is

$$x_t = \lambda x_{t-1} \tag{1}$$

with solution

$$x_t = \lambda^t x_0 = x_0 e^{rt} \text{ with } r = \log \lambda \tag{2}$$

A simple stochastic version is

$$x_t = \lambda x_{t-1} + e_t \tag{3}$$

where e_t is an independent error with zero mean. This corresponds to an auto regressive model, but as $\lambda > 1$ for growth, it is not stationary and so usual auto regressive time series results do not hold. The solution to 5.2 (3) is

$$x_t = \lambda^t x_0 + \sum_{i=0}^{t-1} \lambda^i e_{t-i} \tag{4}$$

(eg Bartlett (1978)).

Hence

$$E(x_t) = \lambda^t x_0$$

$$\text{var}(x_t) = \sum_{i=0}^{t-1} \lambda^i \text{ var}(e_{t-i}) = \sigma^2 \sum_{i=0}^{t-1} \lambda^{2i} \tag{5}$$

$$\text{if var}(e_t) = \sigma^2 \ \forall t$$

$$= \sigma^2(\lambda^{2t} - 1)/(\lambda^2 - 1) \tag{6}$$

This shows the increase in variance usually observed with biological growth.

$$\text{cov}(x_t x_{t'}) = \sigma^2 \lambda^{t-t'} \sum_{i=0}^{t'-1} \lambda^{2i} = \sigma^2 \lambda^{t-t'}(\lambda^{2t'} - 1)/(\lambda^2 - 1) \tag{2}$$

Using (6) and (7) a generalised least squares estimate of λ could be obtained. However it is simpler to consider

$$e_t = x_t - \lambda x_{t-1}$$

and minimise

$$\Sigma e^2 \qquad (8)$$

with respect to λ and, if x_0 is unknown, with respect to x_0. In this use of the model the value of x_0 is important since $E(x_0) \neq 0$ as is the case in the stationary time series model Minimising 5.2 (8) with respect to λ gives

$$\hat{\lambda} = \frac{\sum_{1=2}^{n} x_t x_{t-1} + x_1 x_0}{\sum_{i=2}^{n} x_{t-1}^2 + x_0^2}$$

and $\hat{x}_0$ is found by minimising

$$\sum_{i=1}^{n} (x_t - \hat{\lambda} x_{t-1})^2 + (x_1 - \hat{\lambda} x_0)^2 \qquad (10)$$

with respect to x_0 via a non-linear minimisation algorithm (see Appendix A).

Alternative formulations for the stochastic version of (1) are

$$x_t = \lambda x_{t-1} + \sqrt{x_t} e_t \qquad (11)$$

$$x_t = \lambda x_{t-1} + x_{t-1} e_t \qquad (12)$$

with $\text{var}(e_t) = \sigma^2$.

The two formulations allow for the random effect to have variance proportion to size or standard deviation proportional to size. A general form of this model is

$$x_t = \lambda x_{t-1} + g(x_t) e_t \qquad (13)$$

where g( ) is a known function. The solution being

$$x_t = \lambda^t x_0 + \sum_{i=0}^{t-1} g(x_{t-1-i}) \lambda^i e_{t-i}$$

Estimation of λ and x_0 can be obtained by minimising

$$\sum_{i=1}^{n} (x_t - \lambda x_{t-1})^2 / g^2(x_{t-1}) \qquad (14)$$

with respect to x_0 and λ in an analogous way to the simpler model.

A parametric form of g( ) could be considered.

If

$$g(x_{t-1}) = x_{t-1}^{\theta} \qquad (15)$$

then with $\theta = 0, \frac{1}{2}, 1$ models (3), (11) and (12) are obtained. Thus by minimising (14) with respect to x_0, λ and θ a form of automatic model selection is obtained, and by testing the value of θ using approximate F or t tests a model test can be obtained.

A non independent error could also be considered eg

$$u_t = \varphi e_{t-1} + e_t \qquad (16)$$

or a more complex auto regressive moving average model. Model (16) would give

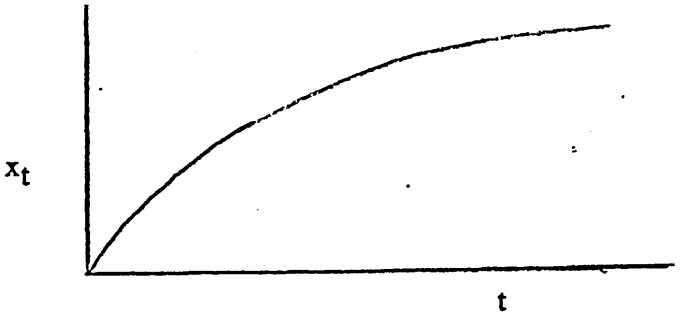
$$x_t = \lambda x_{t-1} + \varphi e_t + e_t \qquad (17)$$

This could be considered as a non stationary auto regressive moving average model. Estimation for such a model will be considered later.

[Notation

As many difference models are non stationary version of the normal stationary time series ARMA models, they will be referred to as NSARMA models.]

Second order difference equation models are also of interest. These provide a more varied behaviour pattern than is possible with the first order equations. In particular the common decrease in growth rate eg



The equation

$$x_t = \lambda_1 x_{t-1} + \lambda_2 x_{t-2} + e_t \tag{18}$$

has solution

$$x_t = A\mu_1^t + B\mu_2^t + \sum_{i=0}^{r-2} \frac{\mu_1^{i+1} - \mu_2^{i+1}}{\mu_1 - \mu_2} e_{t-i} \tag{19}$$

where μ_1 and μ_2 are solutions of

$$z^2 - \lambda_1 z - \lambda_2 = 0 \tag{20}$$

and A and B are given by initial conditions (Bartlett (1978)).

Again as this is a model for growth the general results for stationary models (ie auto regressive models of order 2) are not applicable.

Now

$$E(x_t) = A\mu_1^t + B\mu_2^t \tag{21}$$

and

$$\begin{aligned} \text{var}(x_t) &= \sigma^2 \sum_{i=0}^{r-2} \left[\frac{\mu_1^{i+1} - \mu_2^{i+1}}{\mu_1 - \mu_2}\right]^2 \\ &= \frac{\sigma^2}{(\mu_1 - \mu_2)^2} \left\{\frac{\mu_1^{2t} - 1}{\mu_1 - 1} - 2\,\frac{(\mu_1\mu_2)^t - 1}{\mu_1\mu_2 - 1} + \frac{\mu_2^{2t} - 1}{\mu_2^2 - 1}\right\} \\ &= \sigma_t^2(\mu_1, \mu_2) \end{aligned} \tag{22}$$

Estimation of λ_1, λ_2 and x_{-1}, x_0 is carried out by minimising

$$\sum_{i=1}^{n} (x_t - \lambda_1 x_{t-1} - \lambda_2 x_{t-2})^2 \qquad (23)$$

with respect to λ_1 λ_2 g x_0 and x_{-1}. From these values estimates of A, B, μ_1 and μ_2 could also be obtained. Again with this model x_0 and x_{-1} are important aspects and need to be fully estimated and not treated as nuisance parameters.

More complex error structures could be considered, but this will be left to later.

A useful way of placing the above and more complex model in a compact framework is to use the Kalman filter, this will now be discussed.

5.2.2 Kalman Filter

The Kalman Filter originates in the control engineering field eg Kalman (1960), Kalman and Bucy (1961). It is concerned with the following model

Discrete Time	Continuous Time	
$x_t = f_t(x_{t-1}) + g_t(x_{t-1})\varepsilon_t$	$\dot{x}_t = f_t(x_t) + g_t(x_t)\varepsilon_t$	(24)
$y_t = h_t(x_t) + e_t$	$y_t = h(x_t) + e_t$	

The first equation is the relationship between the unobserved state-space through time, the second is the measurement equation between the observed y and the unobserved x. ε_t and e_t are independent errors, all variables may be vectors.

In this section we will look at the discrete time situation and when the functions are linear ie

$$x_t = F_{t-1}x_{t-1} + G_{t-1}\varepsilon_t$$
$$y_t = H_{t-1}x_t + e_t \qquad (25)$$

F, G and H being matrices. Further

$$\text{var}(\varepsilon_t) = Q_t \quad \text{var}(e_t = R_t \qquad (26)$$

The problem is to estimate x_t given $y_1,...,y_{t-1}$, the solution is given by the following recursive formulae.

Now

$$\hat{x}_{t-} = F\hat{x}_{t-1} \qquad (27)$$

$$P_{t-} = F_{t-1}P_{t-1}F'_{t-1} + G_{t-1}Q_{t-1}G'_{t-1} \qquad (28)$$

and

$$\hat{x}_t = \hat{x}_{t-} + K_t(y_t - H_t\hat{x}_{t-}) \qquad (29)$$

where

$$K_t = P_{t-}H'_t(H_tP_{t-}H'_t + R_t)^{-1}, \text{ the Kalman gain matrix} \qquad (30)$$

Also

$$P_t = [I - K_tH_t]P_{t-} \text{ is the variance of } \hat{x}_t \qquad (31)$$

(eg Gelb (1974) Maybeck (1979)).

It can be seen that the estimate $\hat{x}_t$ is a weighted combination of the value predicted from time t-1 using the state space equations, $\hat{x}_{t-}$ and the residual from the observed and predicted values of y_t. There are several ways of deriving the above estimates. As the procedure involves using prior information the natural setting is Bayesian. It can be shown that $\hat{x}_t$ is the mean of the posterior distribution if one assumes both errors and initial prior information are Normal. Harrison and Stevens (1976) use it as the core of the 'Bayesian Forecasting'.

Other interpretations of the Kalman Filter are also possible. Duncan and Horn (1972) consider it in terms of regression analysis. They start with the wide-sense random - β regression model ie

$y = x\beta + e$ (measurement equations)

$\beta = \mu + \varepsilon$ (state-space equations)

where e and ε have any distribution with zero mean and variance Σ and Σ_β respectively (ie wide-sense distribution). Further μ is a known prior mean for β. The rest are as for standard regression. They show that the Kalman Filter estimate b_t ($\equiv \hat{x}_t$ in previous notation) is the minimum mean square error estimate of β.

A further interpretation can be given in terms of the Goldberger-Theil estimator for the linear model as discussed in chapter 3. Now as before

$$\begin{bmatrix} y \\ r \end{bmatrix} = \begin{bmatrix} X \\ R^* \end{bmatrix} \beta + \begin{bmatrix} e \\ \varepsilon \end{bmatrix} \qquad (32)$$

Let $r = F\beta_{t-1}$, $R^* = I$, $\beta = \beta_t$ then the model is

$\beta_t = F\beta_{t-1} - \varepsilon$

$y = X\beta + e$

so the Goldberger-Theil estimator with

$w = R_t$

$\sigma^2 v = P_{t-}$

is

$$\hat{\beta}_t = \left[X'R_t^{-1}X + P_{t-}^{-1}\right]^{-1}\left[X'R_t^{-1}y + P_{t-}^{-1}F\beta_{t-1}\right]$$

$$= F\beta_{t-} + \left[X'R_t^{-1}x + P_{t-}^{-1}\right]$$

$$\left[X'R_t^{-1}y + \left[P_{t-}^{-1} - \left[X'R_t^{-1}X + P_{t-}^{-1}\right]\right]F\beta_{t-1}\right]$$

$$= F\beta_{t-1} + \left[X'R_t^{-1}X + P^{t-1}\right]^{-1}X'R_t^{-1}\left[y - XF\beta_{t-1}\right]$$

Now pre multiplying

$$\left[X'R_t^{-1}X + P_{t-}^{-1}\right]^{-1}X'R_t^{-1}$$

and

$$K_t = P_{t-}X' \left[XP_{t-}X' + R_t\right]^{-1} \text{ by } \quad \left[X'R_t^{-1}X + P_{t-}^{-1}\right]$$

and post multiplying by $(XP_{t-1}X' + R_t)$ both give $(X'R_t^{-1}XP_{t-} + I)X'$ as

$$X'R_t^{-1}\ (XP_{t-}X' + R_t) = X'R_t^{-1}XP_{t-}X' + X'$$

and

$$(X'R_t^{-1}X + P_{t-}^{-1})P_{t-}X' = X'R_t^{-1}XP_{t-}XP_{t-}X' + X'$$

ie

$$K_t = (X'R_t^{-1}X+P_{t-}^{-1})^{-1}X'R_t^{-1}$$

hence

$$\hat{\beta}_t = \hat{\beta}_{t-} + K_t\ (y - X\hat{\beta}_{t-}) \qquad (33)$$

where

$$\hat{\beta}_{t-} = F\beta_{t-1}$$

This relationship between the two estimators has also recently been shown by Diderrich (1985).

The recursive regression estimator (Brown, Durbin and Evans (1974)) can also be put in Kalman Filter form as

$$\hat{b}_t = \hat{b}_{t-1} + (X_t'X_t)^{-1}(y_t - x_t'\hat{b}_{t-1}) \qquad (34)$$

with $H_t = x_t'$, $R = F = I$ and noting thatnat

$$(X_tX_t)^{-1} = (X_{t-1}'X_{t-1})^{-1} - \frac{(X_{t-1}'X_{t-1})^{-1}\ x_tx_t'(X_{t-1}'X_{t-1})^{-1}}{1 + x_t'(X_{t-1}'X_{t-1})^{-1}x_t}$$

since

$$P_{t-1} = (X_{t-1}'X_{t-1})^{-1}$$

and

$$P_{t-} = I\ P_{t-1}\ I + Q = P_{t-1}$$

$$(X_tX_t)^{-1}x_t = P_{t-}(I + I(H\ P_{t-}\ H_t') - H_t'H_tP_{t-})(1 + H_tP_{t-}H_t')^{-1}H'$$

$$= P_{t-} H'(I + H_t P_{t-} H_t')^{-1} = K_t$$

as HP_tH' is a scalar so

$$(I(HP_{t-}H) - H_t'H_tP_{t-})H_t' = H_t'(H_tP_{t-}H_t') - H'(H_tP_{t-}H_t')$$
$$= 0$$

So we have the Kalman-Filter result.

5.2.3 Further Properties

It can be shown that the quantities

$$z_t = y_t - H\hat{x}_{t-}$$

under the Normal model are Independent Normal with zero mean and variance

$$H_tP_{t-}H_t' + R_t \qquad (35)$$

(eg Maybeck (1979)).

This result can be used in two ways

(i) The validation of the model. The distribution of the standardised residuals

$$z^*_t = \frac{y_t - H\hat{x}_{t-}}{\sqrt{(H_tP_{t-}H_t' + R_t}}, \text{ in the scalar y case,} \qquad (36)$$

can be examined via time plots, Normal probability plots and the examination of the auto correlation of z_t.

(ii) The likelihood can be formed.

Mehra (1972) and Maybeck (1979) review methods estimating the parameters of the model. The matrices

F, G, H, Q and R

may all contain unknown parameters. Using the property given above it can be readily seen that the likelihood in the scalar y case can be written as

$$\text{Log Like} = \text{const} - \frac{1}{2}\sum_t \log\,(H_t P_{t-} H_t' + R_t)$$

$$- \frac{1}{2}\sum_t (y_t - H\hat{x}_{t-})^2/(H_t P_{t-} H_t' + R_t) \qquad (37)$$

This can be maximised using a derivative free non-linear minimisation function eg Nelder-Mead and the variance of the estimates obtained using

$$\text{var}(\text{params}, \theta) = - \left[\frac{\partial^2 \log \text{like}}{\partial\theta_i \partial\theta_j}\right]^{-1} \qquad (38)$$

where the second derivatives may be obtained using numerical methods (see appendix A).

Other methods of estimation are available (Mehra (1972), Maybeck (1979)). These are based on the correlation of the y's/z's and assume time invariant parameters.

(i) Output Correlation Methods

Now

$$E(z_t z_{t-k}) = \begin{cases} H\,\Sigma\,H' + R & k = 0 \\ H\,F\,\Sigma\,H' & k > 0 \end{cases}$$

where $\Sigma = E(x_t x'_{t-k})t$

Unless y_t can be assumed to be stationary and equally Σ is a function of k only satisfying

$$\Sigma = F\,\Sigma\,F' + Q$$

then the complexity of the calculation of the $E(z_t z_{t-k})$ would remove any advantage this method would have over maximum likelihood.

In the simpler stationary case the theoretical and sample auto covariances/correlations can be equated and estimates of the parameters thus obtained.

However, in the growth situation the y's are not stationary and hence this method would be of little use.

(ii) Innovation Correlation Method

As was stated above the sequence z_t should be independent. If an incorrect filter is used z_t will be auto correlated hence values of parameters may be found that give zero auto correlation. Again this is only efficient if the x_t's are stationary.

(iii) Covariance Matching

In the case of vector y_t the observed and expected values of the variance-covariance of y_t can be compared. This is of limited use as only a small number of parameters could be estimated.

A further approach to estimating is by including the unknown parameters as states in an augmented state vector and include them in the filtering (Friedland (1969)). This method is suitable for on-line applications with a small number of unknown parameters. However in the uses presented here it is of little use, but the idea of the concept of the state space being both unknown values of physical states (ie as in engineering applications) and unknown parameters (ie as in the regression approach given above) may well be of use in certain

applications.

Missing Values

The discrete Kalman Filter is applied to a sequence of equally spaced observations. However, it is easily adapted to the situation in which an observation in the sequence is missing.

Let y_{t+1} be missing then

$$x_{t+1} = F\ x_t + G\ e_{t+1}$$

$$x_{t+2} = F\ x_{t_1} + G\ e_{t+2}$$

$$= F.F\ x_t + F\ G\ e_{t+1} + G\ e_{t+2}$$

Hence

$$\hat{x}_{t+2} = F.F\ \hat{x}_t$$

$$\text{var}(x_{t+2}) = F\ F\ \text{var}(x_t)\ F'\ F' + F\ G\ Q\ G'\ F' + G\ Q\ G'$$

$$= F(F\ \text{var}\ (x_t)F' + G\ Q\ G)F' + G\ Q\ G'$$

So

$$P_{t+1-} = F\ P_t\ F' + G\ Q\ G'$$

$$P_{t+2-} = F\ P_{t+1-}\ F' + G\ Q\ G'$$

Thus the Kalman-Filter can 'jump' a missing observation.

5.2.4 Applications of Kalman Filter to Growth Models

Many of the models discussed in section 5.2 can be considered as non stationary auto regressive moving average msodels. It is well known that an auto regressive moving average model can be expressed in the form (24)

For an ARMA(p,q) model eg Harvey and Phillips (1979)

Let r = max(p,q+1) then

$$F = \begin{bmatrix} \varphi_1 & \\ \varphi_2 & I_{r-1} \\ \vdots & \\ \varphi_r & 0 \end{bmatrix} \quad G = \begin{bmatrix} 1 \\ \theta_1 \\ \vdots \\ \theta_{r-1} \end{bmatrix} \qquad (39)$$

and

$x_t' = (x_t, x_{1t}, \ldots, x_{r-1,t})$ -

$x_{1,t}, \ldots, x_{r-1,t}$ being dummy state variables with

$H = (1, 0'_{r-1})$, $R = 0$.

This formulation does not require a stationary model and so can be used for growth models. The models can also be generalised by allowing non-zero R.

Models (12) etc cannot be included in this form as the terms '$x_{t-1}\ e_t$' etc make the model non-linear, these models are covered by later results but will not be examined in detail in relation to the Kalman Filter.

In using the Kalman Filter we need values of x_0 and P_0 to start the recursions as stated before x_0 has to be estimated as it represents parameters of interest. There are several approaches for obtaining values of P_0

(i) Use of information from the population from which the individual came, eg if x_0 represents the height of a plant one week after germination the variation in height of 1 week old plants from the same experiment could be used to give P_0.

(ii) Set $P_0 = 0$. As x_0 is to be estimated on the basis of all the data, it could be considered as 'known' hence not contributing to the uncertainty of the true size at any time. If one is studying an organism that has considerable growth during the study period the size of x_0 compared with later values of x_t will be small and consequently show small variation, this

suggests that this would be a reasonable approximation.

(iii) Where x_0 is estimated the variance of the estimate can also be obtained, this could be used as P_0 for a subsequent iteration of the estimation procedure. As this estimate of variation depends on all the data it will make the properties of the Kalman Filter only approximate.

Kalman Filter and Second Order Models

Model 1 NSAR(2)

$$x_t = \lambda_1 x_{t-1} + \lambda_2 x_{t-2} + \varepsilon_t$$

$$y_t = x_t$$

Hence

$$\mathbf{H} = [1 \quad 0], \quad \mathbf{G} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \quad \mathbf{R} = 0 \text{ and}$$

$$\mathbf{x}_t = \begin{bmatrix} x_t \\ x_{1t} \end{bmatrix} \qquad \mathbf{F} = \begin{bmatrix} \lambda_1 & 1 \\ \lambda_2 & 0 \end{bmatrix}$$

Let

$$\mathbf{P}_{t-} = \begin{bmatrix} \pi_1 t & \pi_2 t \\ \pi_2 t & \pi_3 t \end{bmatrix} \quad \text{hence} \quad \mathbf{K}_t = \begin{bmatrix} 1 \\ \pi_2 t / \pi_1 t \end{bmatrix}$$

so

$$\hat{\mathbf{x}}_t = \begin{bmatrix} \lambda_1 x_{t-1} + x_{1t} \\ \lambda_2 x_{t-1} \end{bmatrix} + \begin{bmatrix} 1 \\ \pi_2 t / \pi_1 t \end{bmatrix} (y - \lambda_{t-1} x_{t-1} - x_{1t})$$

and

$$\mathbf{P}_t = \begin{bmatrix} 0 & 0 \\ -\pi_2 t / \pi_1 t & 1 \end{bmatrix} \quad \mathbf{P}_{t-} = \begin{bmatrix} 0 & 0 \\ 0 & (\pi_3 t - \pi_2 t^2 / \pi_1 t) \end{bmatrix}$$

Finally

$$\mathbf{P}_{t+1-} = \begin{bmatrix} P_3 t + Q & 0 \\ 0 & 0 \end{bmatrix} \text{ where } P_3 t = \pi_3 t - \pi_2 t^2 / \pi_1 t$$

Thus $\pi_2{}^{t+1} = \pi_3{}^{t+1} = 0$ and $\pi_1{}^{t+1} = P_3{}^t + Q$

so

$$K_{t+1} = \begin{bmatrix}1\\0\end{bmatrix} \quad P_{t+1} = \begin{bmatrix}0 & 0\\0 & 0\end{bmatrix} \quad \text{ie } P_3{}^{t+1} = 0$$

and $\hat{x}_{t+1} = y_{t+1}$ so $z_{t+2} = y_{t+2} - \lambda_2 y_{t+1} - x_{1t+1}$

If we let $P_0 = 0$ then

$$z_t = y_t - \lambda_1 y_{t-1} - \lambda_2 y_{t-2} \text{ for } t=1,\ldots,n$$

and

$$\log \text{Like} = -\frac{n}{2} \log Q - \frac{1}{2Q} \Sigma\, 2_t{}^2$$

ie the usual form for maximum likelihood/least squares estimation.

If we let

$$P_0 = \begin{bmatrix}P_1 & P_2\\P_2 & P_3\end{bmatrix}$$

$$P_{1-} = \begin{bmatrix}Q + \lambda_1{}^2 P_1 + 2\lambda_1 P_2 + P_3 & \lambda_2(\lambda_1 P_1 + P_2)\\ \lambda_2(\lambda_1 P_1 + P_2) & \lambda_2{}^2 P_1\end{bmatrix}$$

giving

$$K_1 = \begin{bmatrix}1\\ \lambda_2(\lambda_1 P_1 + P_2)/(Q + \lambda_1{}^2 P_1 + 2\lambda_1 P_2 + P_3)\end{bmatrix}$$

$$P_1 = \begin{bmatrix}0 & 0\\0 & \lambda_2{}^2\left[P_1 - \dfrac{(\lambda_1 P_1 + P_2)^2}{Q + \lambda_1{}^2 P_1 + 2\lambda_1 P_2 + P_3}\right]\end{bmatrix} = \begin{bmatrix}0 & 0\\0 & P_3'\end{bmatrix}$$

but $K_2 = \begin{bmatrix}1\\0\end{bmatrix}$ as before.

So $z_1 = y_1 - (\lambda_1 \hat{x}_0 + \hat{x}_{10})$

$$\hat{x}_1 = y_1$$

$$\hat{x}_{11} = \lambda_2 \hat{x}_{10} + \frac{\lambda_2(\lambda_1 P_1 + P_2)}{Q + \lambda_1{}^2 P_1 + 2\lambda_1 P_2 + P_3} \quad (y_1 - (\lambda_1 \hat{x}_0 + \hat{x}_{10}))$$

$$z_2 = y_2 - (\lambda_1 y_1 + \hat{x}_{11}) \qquad P_{2-} = \begin{bmatrix} P_3' + Q & 0 \\ 0 & 0 \end{bmatrix}$$

$$\hat{x}_2 = y_2$$

$$\hat{x}_{12} = \lambda_2 y_1$$

$$z_3 = y_3 = \lambda y_2 - \lambda_2 y_1$$

As $HP_{1-}H' = Q + \lambda_1{}^2 P_1 + 2\lambda_1 P_2 + P_3 = QA$

and $HP_2H' = P_3' + Q = Q(\lambda_2{}^2(P_1 + (P_1P_3 - P_2{}^2)/Q^2)/A + 1)$

$$\text{Log Like} - \frac{n}{2} \log Q - \frac{1}{2Q} \sum_{t=2}^{n} z_t{}^2$$

$$- \log A - \log(\lambda_2{}^2(P_1 + (P_1P_3 - P_2{}^2)/Q^2)/A + 1)$$

$$- z_1{}^2/2AQ - z_2{}^2/(\lambda^2(P_1 + (P_1 + (P_1P_3 - P_2{}^2)/Q')/A + 1)Q.$$

Model 2 NSARMA (2, 1)

$$x_t = \lambda_1 x_{t-1} + \lambda_2 x_{t-2} + \varepsilon_t + \theta \varepsilon_{t-1}$$

$$y_t = x_t$$

H_1 R_1 x_t and F are as before but $G = \begin{bmatrix} 1 \\ \theta \end{bmatrix}$

so $$P_{t+1-} = \begin{bmatrix} P_3{}^t + Q & \theta Q \\ \theta Q & \theta^2 Q \end{bmatrix}$$

$$= Q \begin{bmatrix} P_3{}^*t + 1 & \theta \\ \theta & \theta^2 \end{bmatrix} \quad \text{where } Q\, P_3{}^{*t} = P_3 t$$

and $$k_t = \begin{bmatrix} 1 \\ 1/(1 + P_3{}^{*t}) \end{bmatrix}$$

If P_0 is assumed to be zero $P_{1-} = Q \begin{bmatrix} 1 & \theta \\ \theta & \theta^2 \end{bmatrix}$

$$k_1 = \begin{bmatrix} 1 \\ Q \end{bmatrix} \qquad P_1 = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}$$

ie $k_2 = \begin{bmatrix} 1 \\ \theta \end{bmatrix}$ $P_{t-} = Q\begin{bmatrix} 1 & \theta \\ \theta & \theta^2 \end{bmatrix}$

So log likelihood is

$$-\frac{n}{2}\log Q - \frac{1}{2Q}\Sigma z_t^2 \text{ where}$$

$$z_1 = y_1 - \lambda_1\hat{x}_0 - \hat{x}_{10}$$

$$\hat{x}_1 = y_1$$

$$\hat{x}_{11} = \lambda_2\hat{x}_0 + \theta z_1$$

$$z_2 = y_2 - \lambda_1 y_1 - \hat{x}_{11}$$

$$\hat{x}_2 = y_2$$

$$\hat{x}_{12} = \lambda_2 y_1 + \theta z_2$$

in general

$$z_t = y_t - \lambda y_{t-1} - \hat{x}_{1t-1}$$

where

$$\hat{x}_{1t} = \lambda_2 y_{t-1} + \theta z_t$$

Alternatively if $P_0 = \begin{bmatrix} P_1 & P_2 \\ P_2 & P_3 \end{bmatrix}$

$$P_{1-} = \begin{bmatrix} Q + \lambda_1^2 P_1 + 2\lambda_1 P_2 + P_3 & Q(\theta + \lambda_2(\lambda_1 P_1 + P_2)) \\ Q(\theta + \lambda_2(\lambda_1 P_1 + P_2)) & Q(\theta^2 + \lambda_2^2 P_1) \end{bmatrix}$$

so $K_1 = \begin{bmatrix} 1 \\ Q(\theta + \lambda_2(\lambda_1 P_1 + P_2))/(Q + \lambda_1^2 P_1 + 2\lambda_1 P_2 + P_3) \end{bmatrix}$

$$P_1 = \begin{bmatrix} 0 & 0 \\ 0 & P_3' \end{bmatrix} \text{ where } P_3' = Q(\theta + \lambda_2^2 P_1) - \frac{(\theta + \lambda_2(\lambda_1 P_1 + P_2))^2 Q^2}{Q + \lambda_1^2 P_1 + 2\lambda_1 P_2 + P_3}$$

$$P_{2-} = \begin{bmatrix} P_3' + Q & Q\theta \\ Q\theta & Q\theta^2 \end{bmatrix} \quad \text{and} \quad K_2 = \begin{bmatrix} 1 \\ \frac{\theta}{Q + P_3'} \end{bmatrix}$$

In general $K_t = \begin{bmatrix} 1 \\ Q\theta/(Q + P_3^{t-1}) \end{bmatrix}$

$$HP_{t-1}H' = P_3^{t-1} + Q$$

and

$$P_3^t = \theta^2(1 - 1/(P_3^{t-1} + Q))$$

Letting $QP_3^{*t} = P_3^t$ then

$$P_3^{*t} = \theta^2 \left[\frac{P_3^{*t-1}}{1 + P_3^{*t-1}}\right]$$

$$K_t = \begin{bmatrix} 1 \\ \theta/(1 + P_3^{*t-1}) \end{bmatrix}$$

and $HP_{t-1}H' = Q(1 + P_3^{*t-1})$

Thus log likelihood is

$$\text{const} - \frac{n}{2} \log Q - \frac{1}{2} \Sigma \log (1 + P_3^{*t-1}) - \frac{1}{2Q} \Sigma z_t^2/(1 + P_3^{*t-1})$$

Model 3

$$x_t = \lambda_1 x_{t-1} + \lambda_2 x_{t-2} + \varepsilon_t$$

$$y_t = x_t + e_t$$

Here F, x_t, H are as before $G = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$ and $R \neq 0$

So $K_t = \begin{bmatrix} \pi_1^t/(\pi_1^t + R) \\ \pi_2^t/(\pi_1^t + R) \end{bmatrix}$

$$P_t = \begin{bmatrix} \pi_1^t R/(\pi_1^t + R) & \pi_2^t R/(\pi_1^t + R) \\ \pi_2^t R/(\pi_1^t + R) & \pi_3^t - \pi_2^2/(\pi_1^t + R) \end{bmatrix} = \begin{bmatrix} P_1^t & P_2^t \\ P_2^t & P_3^t \end{bmatrix}$$

and $P_{t+1-} = \begin{bmatrix} Q + \lambda_1^2 P_1^t + 2\lambda_1 P_2^t + P_3^t & Q\lambda_2(\lambda_1 P_1^t + P_2^t) \\ Q\lambda_2(\lambda_1 P_1^t + P_2^t) & Q\lambda_2^2 P_1 \end{bmatrix}$

If $P_0 = 0$ $\quad P_{1-} = \begin{bmatrix} Q & 0 \\ 0 & 0 \end{bmatrix}$

so $\quad K_1 = \begin{bmatrix} QR/(Q + R) & 0 \\ 0 & 0 \end{bmatrix}$

and $\quad P_{2-} = \begin{bmatrix} Q + \lambda_1{}^2QR/(Q + R) & \lambda_2 Q\lambda_1 QR/(Q + R) \\ \lambda_2 Q\lambda_1 QR/(Q + R) & Q\lambda_2{}^2 QR/(Q + R) \end{bmatrix}$

So apart from the first iteration the full Kalman filter matrix operations are required. One simplification can be incorporated

Let

$$Q = qR \quad \text{and} \quad P^*_{t-} = P_{t-}/R$$

Thus

$$\mathbf{H\,P\,H'} + R = R(\pi_1{}^{*t} + 1)$$

$$= R(q + 1 + \lambda_1{}^2 P_1{}^{*t} + 2\lambda_1 P_2{}^{*t} + P_3{}^{*t})$$

The log likelihood becomes

$$-\frac{n}{2}\log R - \frac{1}{2}\Sigma \log(\pi_1{}^{*t} + 1) - \frac{1}{2R}\Sigma z^2{}_t/(\pi_1{}^{*t} + 1)$$

and R can be estimated by

$$\frac{1}{n}\Sigma z_t{}^2/(\pi_1{}^{*t} + 1)$$

and the likelihood maximised over R calculated as

$$\text{const} - \frac{n}{2}\log(\Sigma z_t{}^2/(\pi_1{}^{*t} + 1)) + \frac{1}{2}\Sigma \log(\pi_1{}^{*t} + 1)$$

5.2.5 A Simulation Study

In order to gain insight into the behaviour of the above mentioned models a simulation study was carried out. The model was chosen so as to give data which approximated the observations given by Day (1966) on the weight gain of pregnant women. The basic parameters used were

$\lambda_1 = 0.8$

$\lambda_2 = .933$

$x_0 = 110$

$x_{-1} = 110$

These were obtained by fitting the basic solution

$$y = Ak^t + Bc^t + e$$

to Day's data using the non-linear regression procedure in SAS. Minor adjustments were made to give the desired behaviour for data simulated from the model.

For several formulations a 100 samples of 42 weeks data was simulated using Nag routines to generate the random normal deviates. The sample size is the maximum length of pregnancy usually recorded. The relatively small number of samples used is due to the complex calculations involved and the iterative nature of each fit making a large simulation (say 1000 samples) not practical on the available computer system.

Computer programmes were written to fit three variations on the chosen model

(a) Basic model (NSAR(2))

(b) Model with auto correlated errors, NSAEMA(2, 1)

(c) The full state-space model with measurement errors (NSAR(2)

with ME)

Data was also simulated for the above three models.

Simulation Result

Basic Model (NSAR(2))

Using a standard deviation of 0.5 the simulations gave results that are summarised in table 5.2(1). The increase of variance with size can clearly be seen.

Fitting the correct model (a) the following results were obtained

Number of samples for which the converged solution to maximisation was obtained = 92

Parameter	Mean	Standard Deviation
λ_1	.078	.011
λ_2	.935	.001
x_0	110.056	.579
x_{-1}	109.921	.514

The standard deviations for λ_1 and λ_2 were as high as they were because of 3 values in each case which were deviant from the rest.

Table 5.2(1)

Week	Mean	Std Dev	Minimum	Maximum
1	111.388	.463	110.30	112.20
2	111.587	.546	110.30	112.60
3	112.855	.675	110.90	114.50
4	113.120	.711	111.30	115.00
5	114.288	.786	112.30	116.10
6	114.682	.734	112.70	116.20
7	115.758	.794	113.30	117.60
8	116.210	.874	113.60	118.10
9	117.287	.896	114.40	119.00
10	117.797	.960	115.50	120.20
11	118.845	.949	115.70	120.80
12	119.334	1.065	116.70	122.30
13	120.487	1.055	117.70	122.90
14	121.055	1.086	118.10	123.90
15	122.069	1.052	119.10	124.70
16	122.661	1.203	119.20	125.80
17	123.749	1.053	121.40	126.10
18	124.364	1.185	120.90	127.70
19	125.432	1.032	123.10	127.70
20	125.778	1.210	122.80	129.50
21	127.051	1.223	124.90	129.30
22	127.756	1.273	123.90	131.20
23	128.821	1.234	126.10	131.60
24	127.502	1.272	126.20	132.70
25	130.555	1.316	127.50	133.40
26	131.311	1.351	128.60	134.70
27	132.265	1.409	129.10	135.10
28	133.055	1.462	130.30	136.60
29	134.153	1.412	130.50	137.30
30	134.853	1.520	131.80	138.50
31	135.968	1.450	132.60	139.30
32	136.661	1.557	133.10	140.40
33	137.785	1.529	134.10	141.30
34	138.470	1.611	135.00	141.70
35	139.574	1.550	135.40	144.30
36	140.459	1.610	136.90	143.80
37	141.424	1.624	136.90	146.30
38	142.374	1.730	138.30	145.90
39	143.379	1.695	140.10	148.70
40	144.339	1.827	139.50	148.00
41	145.288	1.704	141.50	150.80
42	146.323	1.910	141.30	150.60

Fitting model (b) gave the following results

Number of solutions = 94

Parameter	Mean	Standard Error
λ_1	.080	.010
λ_2	.933	.010
θ	.324	.177
x_0	111.382	.470
x_{-1}	102.650	1.223

Again results for λ_1, λ_2 and θ were affected by deviant values. For θ, 35 samples gave values greater than two standard deviations, indicating the possibility of fitting an over complex model. There seems to have been some compensation for θ by lowering x_{-1}, but why this should happen is unclear.

Model with Auto correlated errors (NSARMA(2, 1))

Table 5.2(3) gives a summary of the simulated data for the case where θ = 0.4, standard deviation of error = 0.5.

Fitting the correct NSARMA(2, 1) model the results were

Number of estimates = 93

Table 5.2(3)

Week	Mean	Std Dev	Minimum	Maximum
1	111.314	1.368	107.70	114.90
2	111.348	1.339	108.10	115.40
3	112.818	1.852	109.00	116.90
4	112.941	1.686	108.50	117.40
5	114.313	2.183	108.20	118.80
6	114.570	1.740	110.70	119.50
7	115.901	2.171	109.60	121.50
8	116.110	1.993	110.20	121.30
9	117.167	2.507	111.90	125.70
10	117.930	2.471	112.30	124.20
11	118.774	2.407	113.30	127.20
12	119.422	2.412	113.50	126.20
13	120.472	2.590	114.50	127.60
14	121.070	2.493	114.60	129.30
15	122.185	2.888	114.30	130.10
16	122.538	2.838	114.50	129.80
17	123.777	3.014	116.40	132.50
18	124.369	2.621	118.40	132.00
19	125.549	2.934	118.70	134.40
20	126.109	2.823	120.40	133.10
21	127.176	3.325	120.30	136.40
22	127.783	3.066	122.10	135.10
23	128.884	3.104	122.10	138.10
24	129.360	3.192	122.30	138.40
25	130.520	3.359	123.00	140.30
26	131.222	3.251	125.80	140.70
27	132.358	3.145	124.30	142.00
28	132.764	3.500	125.80	143.80
29	134.064	3.249	126.20	145.00
30	134.462	3.281	127.90	143.20
31	135.831	3.258	128.40	146.10
32	136.632	3.742	130.10	146.20
33	137.530	3.831	127.80	151.10
34	138.407	3.542	131.90	147.20
35	139.494	3.620	131.10	152.40
36	140.260	3.680	132.60	149.80
37	141.398	3.501	134.10	156.70
38	142.141	3.935	133.60	151.20
39	143.247	3.719	134.20	154.80
40	144.198	4.121	136.20	157.30
41	145.180	3.762	138.20	158.10
42	146.249	3.926	138.70	156.70

Parameter	Mean	Standard Deviation
λ_1	.082	.001
λ_2	.932	.001
θ	.346	.168
x_0	111.402	.587
x_{-1}	102.45	.542

This time in 36 cases θ would not have been significant.

Fitting the simpler NSAR(2) model the results were

Number of Estimates = 94

Parameter	Mean	Standard Deviation
λ_1	.112	.088
λ_2	.901	.089
x_0	109.934	.643
x_{-1}	110.051	.702

Note the relatively poor estimation of λ_1 and λ_2.

Measurement Error Model (NSAR(2) + me)

Two situations were simulated

(i) $Q = R = 1.0$, see table 5.2(3), $(q = 1.0)$

(ii) $Q = 3^2$, $R = .5^2$, see table 5.2(4), $(q = .36)$

Table 5.2(4)

Week	Mean	Std Dev	Minimum	Maximum
1	111.475	.638	109.90	112.90
2	111.571	.593	110.40	113.10
3	112.887	.660	110.80	114.20
4	113.144	.571	111.40	114.20
5	114.475	.800	112.20	116.20
6	114.766	.642	113.20	116.70
7	115.935	.776	113.60	117.70
8	116.381	.725	114.60	118.10
9	117.487	.866	115.50	119.60
10	117.877	.788	115.90	119.70
11	118.951	.804	117.50	121.00
12	119.537	.782	118.00	122.10
13	120.605	.897	118.40	122.60
14	121.163	.918	119.20	123.50
15	122.194	1.008	119.40	124.50
16	122.830	.911	120.70	125.10
17	123.844	.969	121.30	126.10
18	124.451	.912	122.60	126.70
19	125.525	.945	123.40	128.10
20	126.070	1.097	123.10	128.70
21	127.163	.885	125.00	129.30
22	127.773	1.178	124.40	130.40
23	128.965	1.023	126.10	131.70
24	129.551	1.147	125.80	131.90
25	130.561	.937	127.90	132.90
26	131.354	1.125	127.90	134.90
27	132.404	1.070	130.00	135.40
28	132.962	1.170	130.20	136.10
29	134.137	1.127	131.20	136.80
30	134.854	1.208	132.00	137.80
31	135.921	1.139	133.30	138.50
32	136.775	1.216	133.50	139.90
33	137.812	1.107	135.10	141.00
34	138.740	1.305	135.50	142.00
35	139.598	1.237	136.90	143.20
36	140.593	1.125	137.80	143.00
37	141.548	1.227	138.60	144.10
38	142.416	1.210	139.50	145.30
39	143.425	1.231	140.80	146.60
40	144.435	1.321	141.10	148.00
41	145.347	1.283	142.40	148.70
42	146.309	1.244	143.80	149.20

In both cases there is a non-monotonic increase in variance with time due to the measurement error.

The results of fitting the correct model were

	Simulation (i)		Simulation (ii)	
Number of Estimates	94		93	
Parameter	Mean	Std Dev	Mean	Std Dev
λ_1	.082	.001	.082	.001
λ_2	.931	.002	.931	.001
q	1.363	.170	1.356	.090
x_0	111.394	1.345	111.458	.520
x_{-1}	102.264	1.300	102.487	.487

The estimation of λ_1, λ_2 was satisfactory but the estimation of q was poor in both cases.

Finally the NSAR(2) and NSARMA(2, 1) models were fitted to simulations (ii) this gave

	NSAR(2)		NSARMA(2, 1)	
No. of Estimates	94		94	
Parameters Mean		Std Dev	Mean	Std Dev
λ_1	.107	.089	.086	.038
λ_2	.905	.089	.927	.038
θ	-	-	.022	.132
x_0	109.785	.732	111.501	.629
x_{-1}	110.166	.844	102.045	4.278

The NSAR(2) model gives poor estimates of λ_1 , λ_2 the NSARMA(2, 1) is better although the θ parameter is usually non-significant.

Conclusions

The selection of the correct (or a more complex) model leads to good estimation of the growth parameters λ_1 and λ_2 . However, the estimation of the additional parameters themselves (θ and c_1) is poor. There is also a problem in the estimation of x_{-1}, this has a mean of about 102 in the more complex models. The reason for this needs further investigation, but the parameter is not of general interest. There was no evidence of non-normality of parameter estimates.

Table 5.2(2)

Week	Mean	Std Dev	Minimum	Maximum
1	111.425	.589	109.80	113.00
2	111.542	.547	110.40	112.90
3	112.924	.788	111.10	114.60
4	113.095	.822	110.70	115.30
5	114.484	.864	112.00	116.70
6	114.662	.962	112.10	117.10
7	116.069	1.003	113.40	118.60
8	116.392	1.062	113.90	118.80
9	117.677	1.233	114.50	120.20
10	118.086	1.256	114.80	120.60
11	119.327	1.358	115.60	122.00
12	119.669	1.381	115.90	122.30
13	120.950	1.401	117.50	124.10
14	121.325	1.537	117.00	124.50
15	122.465	1.582	118.20	126.20
16	123.048	1.617	118.20	126.70
17	124.153	1.591	120.00	127.80
18	124.836	1.730	119.30	128.90
19	125.927	1.672	121.40	130.40
20	126.558	1.841	121.30	130.90
21	127.620	1.795	122.90	132.10
22	128.311	1.909	123.40	132.30
23	129.351	1.945	123.80	134.30
24	130.011	1.999	124.20	134.10
25	131.033	2.141	124.50	136.10
26	131.819	2.113	126.10	136.70
27	132.865	2.306	127.00	137.50
28	133.593	2.198	128.20	138.10
29	134.664	2.368	129.10	139.80
30	135.387	2.291	129.40	140.60
31	136.449	2.359	130.80	141.70
32	137.258	2.359	131.10	142.80
33	138.329	2.410	131.80	144.00
34	139.174	2.427	132.40	144.80
35	140.197	2.529	133.00	145.80
36	141.003	2.613	134.20	147.30
37	142.093	2.634	134.70	148.00
38	142.867	2.686	135.90	148.40
39	144.017	2.692	136.30	149.80
40	144.767	2.727	137.40	150.20
41	145.972	2.785	138.80	152.00
42	146.749	2.924	138.50	152.50

5.2.6 Non-Linear Models and the Kalman Filter

Little attention has been paid to the non-linear discrete time Kalman Filter. The model is (in scalar case)

$$\begin{aligned} x_t &= f_t(x_{t-1}) + g_t(x_{t-1})\varepsilon_t \\ y_t &= h_t(x_t) + e_t \end{aligned} \qquad (40)$$

where f_t, g_t and h_t may be non-linear functions.

The problem with non-linear filtering is that

$$E(x_t) \neq f_t(E(x_{t-1}))$$

and there is no simple relationship between the var (x_t) and var (x_{t-1}).

The problem can be overcome using Taylor's Series approximations (Kendall and Stuart (1969))

$$\begin{aligned} E(x_t) &\simeq f_t\,(E(x_{t-1})) + \frac{1}{2}\frac{\partial^2 f_t(x)}{\partial x^2}\bigg|_{x = E\,(x_{t-1})} \text{x var } (x_{t-1}) \\ \text{var } (x_t) &\simeq \left[\frac{df_t(x)}{dx}\right]^2 \bigg|_{x = E(x_t)} \text{var } (x_t) \end{aligned} \qquad (41)$$

If the measurement equation is linear, say

$$y_t = x_t + e_t$$

then the following procedure can be used

$$\begin{aligned} \hat{x}_{t-} &= f_t(\hat{x}_{t-1}) + \frac{1}{2}\frac{\partial^2 f_t(x)}{\partial x^2}\bigg|_{x = \hat{x}_{t-1}} P_{t-1} \\ P_{t-} &= \left[\frac{df_t}{dx}\right]^2 \bigg|_{x = \hat{x}_{t-1}} P_{t-1} + Q_t \\ K_t &= P_{t-}/(R_t + P_{t-}) \\ z_t &= y_t - \hat{x}_{t-} \\ \hat{x}_t &= \hat{x}_{t-} + K_t\, z_t \\ P_t &= (1 - K_t)P_{t-} \end{aligned} \qquad (42)$$

We will now look at three non-linear difference models

(i) $$x_t = \frac{a\, x_{t-1}}{1 + b\, x_{t-1}}$$

This model is considered by Nash (1977) and also May (1981), May et al (1974), Pielou (1977) and by letting $a = e^r$ and $b = (e^r - 1)/K$ the logistic model is obtained.

Using the stochastic model

$$x_t = \frac{a\, x_{t-1}}{1 + b\, x_{t-1}} + \varepsilon_t$$

The following Kalman filter equations are obtained

$$\hat{x}_t = \frac{a\, \hat{x}_{t-1}}{1 + b\, \hat{x}_{t-1}} + \frac{1}{2}\left[\frac{-2\, a\, b}{(1 + b\, \hat{x}_t)_1^3}\right] P_{t-1}$$

$$P_{t-} = \frac{a^2}{(1 + b\, \hat{x}_{t-1})^4} P_{t-1} + Q$$

the remaining equations are the same.

Estimation is simplified if

$$R\, P_t^* = P_t,\ R\, P_{t-}^* = P_t \quad \text{and} \quad R\, q = Q$$

then $$P_{t-}^* = \frac{a^2}{(1 + b\, \hat{x}_{t-1})^4} P_{t-1}^* + q$$

$$K_t = P_{t-}^*/(P_{t-}^* + 1)$$

and $$P_t^* = (1 - K_t)P_{t-}^* = P_{t-}^*/(P_{t-}^* + 1)$$

The Log likelihood becomes

$$\text{Log Like} = \text{const} - \frac{1}{2}\sum_t \log(R_t) - \frac{1}{2}\sum_t \log(P^*_{t-} + 1)$$

$$- \frac{1}{2}\sum_t (y_t - \hat{x}_{t-})^2/R_t(P^*_{t-} + 1)$$

$R_t = R$ independent of time.

$$\text{Log Like} = \text{const} - \frac{n}{2}\log R - \frac{1}{2}\sum_t \log(P^*_{t-} + 1)$$

$$- \frac{1}{2R}\sum_t (y_t - \hat{y}_{t-})^2/(P^*_{t-} + 1)$$

giving

$$\hat{R} = \frac{1}{n}\sum_t (y_t - \hat{x}_{t-})^2/(P^*_{t-} + 1)$$

and maximised log likelihood with respect to R as

$$\text{log Like max} = \text{cost} - \frac{n}{2}\log\left(\Sigma\,(y_t - \hat{x}_{t-})/(P_t^* + 1)\right)$$

$$- \frac{1}{2}\log\,(P^*_{t-} + 1)$$

This can then be maximised with respect to a, b and λ_1.

The assumption behind the likelihood is that $z_t = y_t - \hat{x}_{t-}$ has a normal distribution. This can be proved in the linear case but may only be approximate in the non-linear case. An alternative criterion would be to use weighted least-square, ie

$$\text{minimise} \quad \Sigma(y_t - \hat{x}_{t-})^2/(P^*_{t-} + 1)$$

with respect to q, a and b.

A second formulation is to allow the variance of the random component to be related to the size, this gives

$$x_t = \frac{\bar{a}\, x_{t-1}}{1 + b\, x_{t-1}} + \frac{x_{t-1}}{1 + b\, x_{t-1}}\, \epsilon_t$$

This could also be interpreted as considering the 'a' parameter to be random with mean $\bar{a}$ and variance Q.

In this case the equation for $\hat{x}_{t-}$ is as before but

$$P_{t-} = \frac{(1 + \bar{a^2})}{(1 + b\, \hat{x}_{t-1})^4} P_{t-} + \frac{\hat{x}^2_{t-1}}{(1 + b\, \hat{x}\quad)} Q_{t-1}^{\ 2}$$

(ii) $$x_t = (1 + r)x_{t-1} - {}^{r}/_{K}\, x^2_{t-1}$$

This is considered by May (1981), May et al (1974), Nash (1977) and Pielou (1977).

This equation can be seen as the discrete time version of the logistic equation. As in (i) a constant or density dependent error could be included. However, as r represents a growth rate which could be random due to environmental changes we will consider the random r model.

Let $r_t = \rho + \varepsilon_t$, ρ is mean rate

then

$$x_t = x_{t-1} + \rho x_{t-1}(1 - x_{t-1}/K) + x_{t-1}(1 - x_{t-1}/K)\varepsilon_t$$

This gives rise to the approximate propagation equations of

$$x_t = x_{t-1} + \rho x_{t-1}(1 - x_{t-1}/K) - \rho\frac{P_{t-1}}{K}$$

$$P_{t-} = \left\{1 + \rho\left[1 - {}^{2x_{t-1}}/_{K}\right]\right\}^2 P_{t-1} + \left\{1 - {}^{2x_{t-1}}/_{K}\right\}^2 P_{t-1}$$

$$+ x^2_{t-1}\,(1 - x_{t-1}/K)^2\, Q$$

The update equations are as before.

(iii) $$x_t = x_{t-1} \exp(r(1 - x_{t-1}/K))$$

This model is considered by May (1981), May et al (1974) and Pielou (1972). Again r is a measure of growth rate.

The multiplicative nature of the model suggests considering a transformed version with

$$\omega_t = \log x_t$$

and

$$y_t = \omega_t + et$$

This gives the equation

$$\omega_t = \omega_{t-1} + r\left[1 - e^{\omega_{t-1}/K}\right]$$

Again using the random r model with $r_t = \rho + \varepsilon_t$ we obtain

$$\omega_t = \omega_{t-1} + \rho(1 - e^{\omega t}/K) + (1 - e^{\omega t}/K)\varepsilon_t$$

To clarify the model let c = log K hence

$$\omega_t = \omega_{t-1} + \rho\left[1 - e^{\omega_{t-1} - c}\right] + \left[1 - e^{\omega_{t-1} - c}\right]\varepsilon_t$$

The propagation equations for this model are

$$\omega_t = \omega_{t-1} + \rho\left[1 - e^{\omega_{t-1} - c}\right] - \frac{h}{2} e^{\omega_{t-1} - c} P_{t-1}$$

$$P_{t-} = e\left\{1 - \rho\, e^{\omega_{t-1}-c}\right\}^2 P_{t-1} + e^{2(\omega_{t-1}-c)} P_{t-1} + \left\{1 - e^{\omega_{t-1}-c}\right\}^2 Q$$

These three models illustrate the way stochastic non-linear difference equation models could be developed. The selection of a suitable model would be based on the techniques shown in chapter 2 and the validation of the model by considering the prediction errors

$$z_t = y_t - \hat{x}_{t-}$$

5.2.7 A Simulation Study

The model $$x_{t+1} = \frac{a\ x_t}{1 + b\ x_t}$$

as studied with a = 1.100 and b = 0.001. Two forms of error structure were considered

(i) $$x_{t+1} = \frac{a\ x_t}{1 + b\ x_t} + \varepsilon_t$$

(ii) $$x_{t+1} = \frac{a\ x_t}{1 + b\ x_t} + \frac{a\ x_t}{1 + b\ x_t}\ \varepsilon_t$$

In both cases the measurement equation

$$y_t = x_t + e_t$$

was used.

With x_0 = 1, 100 samples of 25 observations were generated with Q = R = .04. Summaries of the runs are given in tables 5.2(5) and 5.2(6) respectively.

For situation (i) the following results were obtained

Number of converged estimates = 87 (81*)

Parameter	Mean	Std Dev	Mean*	Std Dev*
a	1.098	.036	1.099	.036
b	.002	.010	.002	.010
q	7.649	43.683	.835	.604
x_0	1.044	.236	1.046	.235

The '*' results were obtained when some of the more extreme values were omitted. This illustrates a weakness with a simulation approach with a non-linear minimisation process. In practice one would try several different starting values to check for global minimisation, this is not practical when carrying out 100 or more simulation. It may well be that more careful analysis of the extreme cases would lead to better estimates.

The residuals from the models were also examined, the results given in table 5.2(7) showed no sign of departure from normality.

For situation (ii) the results were

Number of converged estimates = 71 (59*)

Parameter	Mean	Std Dev	Mean*	Std Dev*
a	1.150	.145	1.125	.117
b	.012	.025	.006	.012
q	703.556	3466.706	1.589	1.159
x_0	.952	.263	.977	.273

Table 5.2(5)

Week	Mean	Std Dev
1	1.124	.266
2	1.209	.396
3	1.363	.434
4	1.526	.502
5	1.621	.573
6	1.802	.618
7	1.911	.735
8	2.130	.801
9	2.275	.860
10	2.535	.975
11	2.756	1.096
12	3.071	1.240
13	3.322	1.360
14	3.634	1.560
15	3.952	1.749
16	4.368	1.928
17	4.813	2.143
18	5.225	2.410
19	5.692	2.602
20	6.267	2.850
21	6.794	3.110
22	7.401	3.385
23	8.111	3.712
24	8.788	4.038
25	9.555	4.402

Table 5.2(6)

Week	Mean	Std Dev
1	1.059	.288
2	1.260	.399
3	1.333	.480
4	1.501	.656
5	1.733	.843
6	1.903	1.127
7	2.123	1.167
8	2.249	1.501
9	2.575	1.839
10	2.868	2.214
11	3.067	2.487
12	3.291	2.592
13	3.574	2.768
14	3.929	3.311
15	4.439	3.781
16	4.655	3.781
17	4.800	3.644
18	5.219	4.152
19	5.919	5.276
20	6.635	6.121
21	7.170	7.033
22	7.915	7.670
23	8.171	8.373
24	8.689	8.718
25	9.340	9.679

Again the problem of deviant results was encountered. The results for the 'better' values are still poor compared with the model (i). This is partly to be expected due to the increased variance present in this model.

Analysis of the residuals showed that even on eleminating the more extreme residuals (outside ± 4) there was still evidence of non-normality, see table 5.2(8).

In both cases there was some evidence of non-normality in the parameter estimates. The sample size of 25 is, however, at the lower end of range at which asymptotic results become valid.

Table 5.2(7)

Count	Midpoint	One symbol equals approximately 8.00 occurrences
1	-1.1	
1	-1.0	
7	-.9	*
10	-.8	*
22	-.7	***
48	-.6	******
80	-.5	**********
151	-.4	********************
187	-.3	************************
254	-.2	*********************************
306	.1	**
333	.0	**
326	.1	***
260	.2	**********************************
207	.3	***************************
145	.4	*******************
85	.5	***********
37	.6	*****
26	.7	***
10	.8	*
4	.9	*

I....+....I....+....I....+....I....+....I....+....I....+....I

0 80 160 240 320 400

Histogram Frequency

Mean	.001	Std Err	.006	Median	.007
Mode	.033	Std Dev	.301	Variance	.091
Kurtosis	-.043	SE Kurt	.098	Skewness	-.058
SE Skew	.049	Range	2.006	Minimum	-.1083
Maximum	.923				

Table 5.2(8)

Count	Midpoint	One symbol equals approximately 16.00 occurrences
2	-4.1	
13	-3.7	*
9	-3.3	*
14	-2.9	*
8	-2.5	*
26	-2.1	**
54	-1.7	***
77	-1.3	*****
144	-.9	*********
359	-.5	**********************
735	-.1	**
570	.3	************************************
229	.7	**************
85	1.1	*****
52	1.5	***
23	1.9	*
22	2.3	*
13	2.7	*
6	3.1	
11	3.5	*
4	3.9	

I....+....I....+....I....+....I....+....I....+....I....+....I

0 160 320 480 640 800

Histogram Frequency

Mean	-.026	Std Err	.018	Median	-.002
Mode	.293	Std Dev	.898	Variance	.806
Kurtosis	4.062	SE Kurt	.099	Skewness	-.166
SE Skew	.049	Range	7.855	Minimum	-3.997
Maximum	3.857				
Valid Cases	2456				

Conclusions

The results are encouraging. For the model with a simple additive error the estimates are reasonable. For the model with error variance related to size the results showed a certain amount of bias and quite seven problems with convergence to a reasonable value. But the variance involved with the simulation was high and the sample size low, relatively, so under more suitable conditions the results should be reasonable.

5.3 Stochastic Differential Equations

5.3.1 Introduction

Before examining stochastic differential growth models we need to consider the basics of stochastic calculus.

The random variable Y(t) is said to be mean square differentiable with differential coefficient $\dot{Y}(t)$ if

$$\underset{h \to 0}{\text{Lt}}\ E\left\{\left|\frac{Y(t+h)-Y(t)}{h} - \dot{Y}(t)\right|^2\right\} = 0$$

a condition for this is that $\mu(t,s) = E(Y(t)Y(s))$ has partial derivatives

$$\frac{\partial \mu}{\partial t},\ \frac{\partial \mu}{\partial s},\ \frac{\partial^2 \mu}{\partial t\ \partial s} \text{ at } t = s$$

(Bartlett 1978).

Further

$$E(\dot{Y}(t)) = \frac{\partial}{\partial t} E(Y(t)),\ E(\dot{Y}(t)Y(s)) = \frac{\partial \mu(t,\ s)}{\partial t}$$

and (1)

$$E(\dot{Y}(t)\dot{Y}(s)) = \frac{\partial^2 \mu(t,\ s)}{\partial t\ \partial s}$$

The mean square integral is

$$\int_a^b f(t)\, Y(t)\, dt \qquad (2)$$

where f(t) is a real bounded function, piecewise continuous. A necessary and sufficient condition for this integral is that

$$\int_a^b \int_a^b f(t)\, f(s)\, \mu(s, t)\, ds\, dt \qquad (3)$$

exists in the ordinary Riemann sense.

A Reimann-Stieltjes integral

$$\int_a^b f(t)\, dY(t) \qquad (4)$$

may also be defined if and only if

$$\int_a^b \int_a^b f(t)\, f(s)\, d\mu(s, t) \qquad (5)$$

exists as an ordinary Riemann-Stieltjes integral.

Many applications of stochastic calculus especially in the engineering applications involve the use of a white noise process, ω_t.

For ω_t the covariance is given by the dirac delta function.

So for such a process (5) and hence (4) does not exist.

Following Jaswinski (1970) white noise can be considered as the formal derivative of Brownian motion, then formally

$\omega_t \sim d\beta_t/dt$

For a random function $g_t(\omega)$ we need

$$\int_a^b g_t(\omega)\ d\beta_t$$

Two definitions of the integral are given in the literature.

(i) Itô Stochastic Integral

Partition the interval [a, b] into

$$a = t_0 < t_1 < \ldots < t_n = b$$

and consider the step functions

$$g_t(\omega) = \begin{cases} 0 & t < t_0 \\ g_i(\omega) & t_1 \leqslant t < t_{i+1} \\ 0 & t \geqslant t_n \end{cases} \qquad (6)$$

where $g_i(\omega)$ is independent of $\beta_{t_k} - \beta_{t_\ell}$ $t_i \leqslant t_\ell \leqslant t_k \leqslant b$

and $E(|g_i(\omega)|^2) < \infty$

then the Itô integral is defined by

$$\int_a^b g_t(\omega)\ d\beta_t = \underset{n\to\infty}{\text{Lt}} \sum_i g_i(\omega)(\beta_{t_{i+1}} - \beta_{t_i}) \qquad (7)$$

This definition of the integral has the following properties

$$E\left[\int_a^b g_t(\omega)\ d\beta_t\right] = 0 \qquad (8)$$

and for a second random function $g_t^*(\omega)$

$$E\left[\int_a^b g_t(\omega)\,d\beta_t \int_a^b g_t^*(\omega)\,d\beta_t\right] = \sigma^2 \int_a^b E(g_t\ g_t^*)\,dt \qquad (9)$$

Further let $\varphi(x)$ be a twice continuously differentiable real scalar function of the real variable x with

$$\varphi x = \frac{d\varphi}{dt} \quad \varphi xx = \frac{d^2\varphi}{dx^2}$$

then

$$\int_a^b \frac{d\varphi}{dt}\, d\beta_t = \varphi(\beta_b) - \varphi(\beta_a) - \frac{\sigma^2}{2} \int_a^b \varphi xx(\beta_t)\, dt \qquad (10)$$

(ii) Stratonovich Stochastic Integral

With the interval [a, b] partitioned as above consider $g_t(\omega)$ as an explicit function of β_t denoted by $g(\beta_t, t)$ then the Stratonovich integral is defined by

$$\$\int_a^b g(\beta_t, t)\, d\beta_t = Lt\ \Sigma g\left[\frac{\beta t_j + \beta t_{j+1}}{2}, t_i\right]\left[\beta t_{i+1} - \beta t_i\right] \qquad (11)$$

The Stratonovich calculus satisfies all the formal rules of calculus and is related to the Itô integral by

$$\$\int_a^b g(\beta_t, t)\, d\beta_t = \int_a^b g(\beta_t, t)\, d\beta_t + \frac{\sigma^2}{2} \int_a^b g_\beta(\beta_t, t)\, dt \qquad (12)$$

where

$$g_\beta(\beta, t) = \frac{\partial g}{\partial \beta} \qquad \text{(Jaswinski (1970))}$$

Jaswinski prefers the Itô integral to the Stratonovich because

(a) It is easier to compute expectations of the Itô integral as (8) and (9) do not hold for the Stratonovich.

(b) The Itô integral has nicer mathematical properties (Doob (1959) ch 1X).

(c) The Itô integral is defined for a much broader class of functions.

(d) The Stratonovich is not applicable to non-linear filtering theory.

Smith and Tuckwell (1974) state that it is better to adopt the Stratonovich calculus if the stochastic equation represents a 'physical' random process.

Turelli (1977) also discusses the selection of a stochastic calculus. The ambiguity presented by the two approaches can be overcome if

(a) The underlying differential equation is replaced by a difference equation

or

(b) The white noise is replaced by auto correlated noise.

They consider the solutions that are given by the limits of the unambiguous situations.

When considering difference equations the limits tend to the Îto integral for a limited number of situations in which results are available. The auto correlated noise situation tends to either case depending on how the progression from auto correlated to white noise is made.

One method of finding the pdf of the solution to a stochastic differential equation is by using diffusion methods.

Jaswinski (1970) considers the equation

$$dx_t = f(x_t, t)dt + g(x_t, t)d\beta_t \qquad (13)$$

where $d\beta_t$ is the formal derivative of a Weiner process. He shows that the Fokker-Plank equation for the model is

$$\frac{\partial p}{\partial t} = - \frac{\partial (Mp)}{\partial x} + \frac{v}{2} \frac{\partial^2 (pg^2)}{\partial x} \qquad (14)$$

where (i) p is the transition probability density function

(ii) $E(\beta_t^2) = vdt$

(iii) $M = f$ for the Itô calculus

$M = f + \frac{v}{2} g \frac{\partial g}{\partial x}$ for the Stratonovich calculus

if g = constant then both calculi have the same solution.

If the model is time homogeneous ie

$f(x_t, t) = f(x_t)$ and $g(x_t, t) = g(x_t)$

So reduces to

$$\frac{\partial p}{\partial t} = (v(g'^2 + g'') - M')p + (2vgg' - M')p' + \frac{v}{2} g^2 f'' \qquad (15)$$

where prime indicates differentiation with respect to x.

Bailey (1964) shows that this equation can be solved either by taking Laplace transforms or by forming a solution

$$p = \sum_r A_r x_r(x) e^{-\lambda r t}$$

where A_r and λ_r are determined by boundary conditions.

The following common processes are described by their Fokker-Plank equations

(i) Weiner process

$$\frac{\partial p}{\partial t} = - \mu \frac{\partial P}{\partial x} + \frac{\sigma^2}{2} \frac{\partial^2 P}{\partial x^2} \qquad (16)$$

with solution

$$P = \frac{1}{\sqrt{2\pi\sigma^2 t}} \exp\left\{ - \frac{(x - \mu t)^2}{2\sigma^2 t} \right\} \qquad (17)$$

when starting at the origin with no boundaries.

(ii) Limit of a branching process

$$\frac{\partial p}{\partial t} = - \alpha \frac{\partial (xP)}{\partial x} + \frac{1}{2} P \frac{\partial^2 (xP)}{\partial x^2} \qquad (18)$$

with initial condition $x = x_0 \quad t = t_0$

$$P = \frac{2\alpha}{\beta(e_{\alpha t} - 1)} \left[\frac{x_0 \, e^{\alpha t}}{x}\right]^{\frac{1}{2}} - \exp\left\{\frac{- 2\alpha(x_0 \, e^{\alpha t} + x)}{\beta(e^{\alpha t} - 1)}\right\}$$

$$x \; I_1 \left[\frac{4\alpha(x_0 \; x \; e^{\alpha t})^{\frac{1}{2}}}{\beta(e^{\alpha t} - 1)}\right] \qquad (19)$$

where I_1 is a Bessel function of first order.

(iii) A Weiner process starting at x_0 with absorbing barrier at $x = 0$

$$P = \frac{1}{\sqrt{2\pi\sigma^2 t}} \left\{\exp\left[- \frac{(x - x_0^{-\mu t})^2}{2\sigma^2 t}\right] - \exp - \frac{2\rho x_0}{\sigma^2} \left[- \frac{(x - x_0^{-\mu t})^2}{2\sigma^2 t}\right]\right\} \qquad (20)$$

(iv) An Ornstein-Uhlenbeck process (OU)

$$\frac{\partial p}{\partial t} = \beta \frac{\partial}{\partial x}(xp) + \frac{1}{2} \sigma^2 \frac{\partial^2 p}{\partial x^2} \qquad (21)$$

without barriers, starting from x_0 gives solution

$$P = \frac{1}{\sqrt{2\pi\sigma^2(1 - e^{-2\beta t})/2\beta}} \exp\left\{\frac{- \beta}{\sigma^2(1 - e^{-2\beta t})} \left[x - x_0 \, e^{-\beta t}\right]^2\right\} \qquad (22)$$

(v) An OU process with absorbing barrier at $x = 0$

$$P = \frac{1}{\sqrt{2\pi\sigma^2(1 - e^{-2\beta t})/2\beta}} \left\{\exp\left[\frac{- \beta}{\sigma^2(1 - e^{-2\beta t})} \left[x - x_0 \, e^{-\beta t}\right]^2\right.\right.$$

$$\left.\left. - \exp\left[- \frac{\beta}{\sigma^2(1 - e^{-2\beta t})} \left[x + x_0 \, e^{-\beta t}\right]^2\right]\right\} \qquad (23)$$

Associated with these processes is their differential equation models.

Using the Itô calculus

(i) Weiner process $dx_t = \mu dt + d\beta t$ (24)

(ii) Branching process $dx_t = \alpha x dt + \beta \sqrt{x}\, d\beta_t$ (25)

(iii) OU process $dx_t = -\beta x dt + d\beta_t$ (26)

Using the Stratonovich calculus the Weiner and OU processes are the same, the branching process becomes

$$dx_t = (\alpha x - {}^{\beta}/_{u})dt + \sqrt{\beta/x}\, d\beta_t \quad (27)$$

A wider class of models can be transformed to the above simple forms. Tuckwell (1974) suggests the transformation

$$y_t = \int^t f(x)^{-1}\, dx \quad (28)$$

for the equation

$$dx_t = f(x_t)dt + f(x_t)d\beta_t \quad (29)$$

Using the Stratonovich calculus

$$dy_t = dx_t/f(x_t)$$

hence (29) reduces to

$$dy_t = dt + d\beta_t$$

ie a Wiener process (24). However if the Îto calculus is used then following Jaswinski (1970)

$$dy_t = dx_t/f(x_t) + \frac{\sigma^2}{2}\left\{f(x_t)\right\}^2 \left\{\frac{\partial f(xt)^{-1}}{\partial x_t}\right\}dt$$

$$= dx_t/f(x_t) + \frac{\sigma^2}{2}\,\frac{\partial f(xt)}{\partial x_t}\,dt$$

which gives

$$dy_t = \left\{1 - \frac{\sigma^2}{2}\,\frac{\partial f(x_t)}{\partial x_t}\right\}dt + d\beta_t \quad (30)$$

which only leads to a Weiner process if $f(x_t) = cx_t$. If

$$f(x_t) = cx_t^2 + \frac{2}{\sigma^2}\,x_t$$

then it would lead to an OU process.

If the transformation is applied to the model

$$dx_t = (f(x_t) - \beta \int^{x_t} f(x)dx)dt + f(x_t)d\beta_t, \quad B > 0 \tag{31}$$

then it gives an OU process (Tuckwell (1974)).

A further set of transformations could be used. If x_t has been transformed to give y_t as above and

$$dy_t = f(t)dt + g(t)d\beta_t \tag{32}$$

where f(t) and g(t) are functions of t only, then using the transformation

$$z_t = y_t - \int^t f(\tau)d\tau$$

z_t has the diffusion equation

$$\frac{\partial p}{\partial t} = \frac{\sigma 2}{2} \frac{\partial^2 pg^2}{\partial t} \tag{34}$$

with solution

$$p = \frac{1}{\sqrt{2\pi\sigma^2 G}} \exp\left[-\frac{1}{2\sigma^2 G}(z_t{}^2) - z_0\right]^2 \tag{35}$$

where

$$G = \int_0^t g^2(t)dt$$

hence the equation

$$x_t = f(t)f_t(x_t)dt + g(t)f(x_t)d\beta_t$$

has solution, using the Stratonovich calculus

$$p = \frac{|f(x_t)|^{-1}}{\sqrt{2\pi\sigma^2 G}} \exp\left[-\left[\int_{x_6}^{x_t} f_t(x)^{-1}\,dx - \int^t f(\tau)d\tau\right]^2 / 2\sigma^2 G\right] \tag{36}$$

$$\int^{x_t} f(x)^{-1}\,dx > 0$$

This is a generalisation of the result given by S anland and McGilchrist (1979).

This result can be used to form a likelihood, given a sample of n observation x_i at times t_i

$$\log \text{like} = \text{const} - \frac{n}{2} \log \sigma^2 - \sum_{i=1}^{n} \log |f(x_i)| - \frac{1}{2} \sum_{i=1}^{n} \log G(t_i) \qquad (32)$$

$$- \frac{1}{2\sigma^2} \sum_{i=1}^{n} \left\{ \left[\int_{x_{i-1}}^{x_i} f(x)dx - \int^{t} f_t(\tau)d\tau \right]^2 / G(t_i) \right\}$$

This can be used to find estimates of the parameters of the functions $f(x_t)$, $f_t(t)$ and $g(t)$ and their variances.

If $f(x_t)$ is a known function of x_t then a weighted non-linear sum of squares approach could be used ie

$$\min s = \sum_{i=1}^{n} \left[\int_{x_{i-1}}^{x_i} f(x)dx - \int^{t} f_t(\tau)d\tau \right]^2 / G(t_i) \qquad (38)$$

Further if $g(t) = 1$ then $G = t$ hence

$$s = \sum_{i=1}^{n} \left[\int_{x_{i-1}}^{x_i} f(x)dx - \int^{t} f(\tau)d\tau \right]^2 / t \qquad (39)$$

ie weighted least squares estimation for

$$\int_{x_{i-1}}^{x_i} f(x)dx \quad \text{with model} \quad \int^{t} f(\tau)d\tau \quad \text{and weight } t.$$

The above results will now be applied to growth models.

5.3.2 Linear Growth Model

Starting with the basic deterministic model

$$\frac{dx}{dt} = rx$$

a stochastic component can be added in several ways

(i) Additive random component

$$\frac{dx}{dt} = rx + w_t$$

where w_t is normal with variance σ^2. This equation can be written in the form

$$dx_t = rx\ dt + d\beta_t$$

and is the same for both the Stratonovich and Îto calculi. However using (2) the solution is

$$x_t = x_0\ e^{rt} + e_t$$

where (A)

$$e_t = \int_0^t e^{r(t-s)}\ w_s\ ds$$

As w_t is normal e_t will be normal and the integral can be interpreted in the mean square sense. So expectation and integration can be interchanged. Hence

$$E(e_t) = \int_0^t e^{r(t-s)}\ E(w_s)ds = 0$$

d for $t > \tau$

$$E(e_t\ e_\tau) = \int_0^t \int_0^\tau e^{r(t-s)}\ e^{r(\tau-s')}\ E(w_s\ w_{s'})ds'ds$$

$$= \sigma^2 \int_0^\tau e^{r(t-s')}\ e^{r(\tau-s')}\ ds'$$

$$= \frac{\sigma^2}{2r}\left[e^{r((t+\tau)} - e^{r(t-\tau)}\right]$$

$$\text{var}(e_t) = \frac{\sigma^2}{2r}\left[e^{2rt} - 1\right]$$

From (A) it can be seen that

$$y_{t_i} = x_{t_i} - x_{t_{i-1}} e^{r\delta_i}$$

form an independent sequence of normal variables

$$E\left[y_{t_i}\right] = x_0\left[e^{rt_i} - e^{rt_{i-1}} e^{r\delta_i}\right.$$

$$= 0$$

where $\delta_i = t_i - t_{i-1}$

$$\text{var}\left[y_{t_i}\right] = \sigma^2 \int_{t_{i-1}}^{t_i} e^{2r(t_i - s)} ds$$

$$= \frac{\sigma^2}{2r}\left[e^{2r\delta_i} - 1\right]$$

Hence

$$\log \text{ like} = \text{const} - \frac{n}{2}\log\sigma^2 + \frac{n}{2}\log r - \frac{n}{2}\log\left[e^{r\delta_i} - 1\right]$$

$$- \frac{\sigma^2}{2r} \sum \left[y_i\right]^2 \Big/ \left[e^{r\delta_i} - 1\right]$$

Mitchell (1968) considers estimation in the related model with $\rho^* = e^{r}$. She gives the maximum likelihood equation for ρ^* and also considers estimates based on ratios of linear and quadratic polynomials of the observations first suggested by Finney (1958).

(ii) Random r

Let r be a process with mean ρ and variance σ^2 then

$$\frac{dx}{dt} = r(t)x = \rho x + x w_t$$

where w_t is a process with zero mean.

Lewontin and Cohen (1969) gave the solution

$$x_t = x_0 \exp \int_0^t r(t)dt$$

Levins (1969) applied the central limit theorem to

$$\int_0^t r(t)dt$$

and hence deduced that x_t/x_0 has a lognormal distribution

$$f\left[\frac{x}{x_0}\right] = \frac{x_0}{\sigma x\sqrt{2\pi}} \exp\left\{-\frac{1}{2t\sigma^2}\left[\log\left[\frac{x}{x_0}\right] - \rho t\right]^2\right\}$$

The equation can be written as

$$dx_t = \rho\, x\, dt + x\, d\beta_t$$

using the transformation $y = \log x$, $-\infty < y < \infty$

$$dy_t = \rho dt + d\beta_t$$

Hence y_t is Weiner process and the solution is

$$f(y,\ t\,|y_0) = \frac{1}{\sqrt{2\pi\sigma^2 t}} \exp\left\{-\frac{(y - y_0 - \rho t)^2}{2\sigma^2 t}\right\}$$

transforming this gives

$$f(x,t \mid x_0) = \frac{1}{x\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{(\log(x/x_0) - \rho t)^2}{2\sigma^2 t}\right\}$$

If instead of using the Stratonovich calculus the Itô calculus is used then the solution becomes

$$f(x,t \mid x_0) = \frac{1}{x\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{(\log(x/x_0) - (\rho - \frac{\sigma^2}{2})t)^2}{2\sigma^2 t}\right\}$$

As the observed x follow a log normal distribution this suggests using the log transformed observations

$$y_{t_i} = \log(x_{t_i})$$

Further $w_{t_i} = y_{t_i} - y_{t_{i-1}}$ will be independent with

$$E(wt_i) = \rho(t_i - t_{i-1}) = \rho\sigma_i$$

$$var(w_{t_i}) = \sigma^2(t_i - t_{i-1}) = \sigma^2\ \delta_i$$

Hence an estimate of ρ is

$$\frac{1}{n} \Sigma \, {}^{w_{t_i}}/_{\delta_i}$$

with variance

$$\frac{\sigma^2}{n} \Sigma \, {}^{1}/_{\delta_i}$$

and will be normally distributed.

(iii) Levins (1969) proposed the model

$$\frac{dx}{dt} = \rho x + \sqrt{v(1 - v)}\, e(t) \sqrt{x}$$

where e(t) is a random variable with unit variance. Using the transformation

$$y = x^{\frac{1}{2}}$$

he shows that

$$E(y) = y_0 \, e^{\frac{1}{2}\rho t}$$
$$\text{var}(y) = \frac{1}{4\rho} v(1 - v)(e^{\rho t} - 1)$$

and

$$E(x) = x_0 e^{\rho t} + \frac{1}{4r} v(1 - v)(e^{\rho t} - 1)$$

Further as y will be approximately normal x will be approximately distributed as a non-central χ^2. The term $\sqrt{v(1 - v)}$ is used to reflect sampling variation when v is mean viability of an individual. The term $\sqrt{v(1 - v)}\, e(t)$ could be replaced by $\varepsilon(t)$ where $\varepsilon(t)$ has 'variance' σ^2.

Writing the model in the form

$$dx = \rho x \, dt + \sqrt{x} \, d\beta_t$$

and using the above transformation an OU process is obtained for y (Tuckwell (1974)). Now the OU has to lie in the range $[0, \infty]$ with 0 being an absorbing state hence the pdf is

$$p(x_t,\ t|x_0) = \frac{x_t^{-\frac{1}{2}}\ e^{-\frac{1}{2}\rho t}}{\left[2\pi\sigma^2\left[\frac{1\ -\ e^{-\rho t}}{\rho}\right]\right]^{\frac{1}{2}}}\left[\exp\left[-\frac{2(\sqrt{x_t}\ e^{-\frac{1}{2}\rho t}\ -\ \sqrt{x_0})^2}{\sigma^2\left[\frac{1\ -\ e^{-\rho t}}{\rho}\right]}\right]\right.$$

$$\left. -\ \exp\left[-\frac{2(x_t^{\frac{1}{2}}e^{-\frac{1}{2}\rho t}+\sqrt{x_0})^2}{\sigma^2\left[\frac{1\ -\ e^{-\rho t}}{\rho}\right]}\right]\right] \quad \text{(Tuckwell (1974))}$$

giving

$$E(x_t) = \sigma\ \exp\left\{\frac{-2x_0\ \rho}{\sigma^2(1\ -\ e^{-\rho t})}\right\}\left[\frac{x_0\ e^{\rho t}(e^{\rho t}\ -\ 1)}{2\pi r}\right]^{\frac{1}{2}}$$

$$+\ \left\{x_0\ e^{\rho t}\ +\ \frac{\sigma^2}{4}\left[\frac{e^{\rho t}\ -\ 1}{\rho}\right]\right\}\ \Phi\ \left[\left[\frac{2x_0\rho}{\sigma^2(1\ -\ e^{-\rho t})}\right]^{\frac{1}{2}}\right]$$

Tuckwell points out that the difference between this result and Levins is due to Levins ignoring the problem of negative population. However if σ^2 is small compared to ρ the above gives

$$E(x_t) \simeq x_0\ e^{\rho t}\ +\ \frac{\sigma^2}{4\rho}(e^{\rho t}\ -\ 1)$$

- Levins result. This is the situation when the variation does not lead to a negative growth rate.

Kiester and Barakat (1974) use the mean square calculus, assuming the $\varepsilon(t)$'s are correlated. They show that $y = \sqrt{x}$ has a normal distribution with

$$\text{mean} = y_0\ e^{\frac{1}{2}\rho t}$$

$$\text{variance} = (\sigma^2/_4)e^{\rho t}\ G(t,t)$$

where

$$G(t_1,\ t_2) = \int_0^{t_2}\int_0^{t_1} e-\ E(e(s_1)e(s_2))ds_1ds_2$$

which reduces to

$$\frac{1}{\rho}\ (1\ -\ e^{-\rho t})$$

when the errors tend to independence, ie Levins result.

In order to estimate ρ if we assume that σ^2 is small compared to ρ and

that the probability of y being negative when considered as an unrestricted OU process is negligible then y has a Normal distribution with

$$E(y) = y_0 \, e^{\rho t}$$

$$\text{variance}(y) = \frac{\sigma^2}{4\rho} \, (e^{\rho t} - 1)$$

from above.

Thus the log likelihood is given by $\Sigma \log f(y_{t_i} | y_{t_{i-1}})$

$$= \text{const} - \frac{n}{2} \log \sigma^2 + \frac{n}{2} \log \rho - \Sigma \log(e^{\rho \delta i} - 1)$$

$$- \frac{2\rho}{\sigma^2} \Sigma \frac{(y_{t_i} - y_{t_{i-1}} \, e^{\rho \delta i})^2}{e^{\rho \delta i} - 1}$$

where $\delta_i = t_i - t_{i-1}$

This can be maximised using a non-linear optimisation technique.

5.3.3 Density Dependent Models

These models are of the form

$$\frac{dx}{dt} = rx\ F(x)$$

(i) $F(x) = x^{b-1} \qquad 0 < b < 1$

Using the transformation

$$y = \int^{x_t} x^{-b}\ dx = \frac{x^{1-b}}{1-b}, \quad y > 0$$

the equation

$$dx = \rho\ x^b\ dt + x^b\ d\beta_t$$

reduces to the Weiner model with absorbing state at x = 0.

$$f(x_t,\ t\,|\,x_0) = \frac{x^{-b}}{\sqrt{2\pi\sigma^2 t}}\left[\exp\left[-\frac{1}{2\sigma^2 t}\,\frac{(x_t^{1-b} - x_0^{1-b} - \rho t)^2}{(1-b)^2}\right]\right.$$

$$\left. - \exp\left[-\frac{2\rho x_0^{1-b}}{\sigma^2(1-b)}\right]\exp\left[-\frac{1}{2\sigma^2 t}\left[x_t^{1-b} + x_0^{1-b} - \rho t\right]^2\right]\right]$$

(This result was given by Smith & Tuckwell (1974)).

As the transformation depends on the unknown parameter b the analysis of y is not helpful. The likelihood relating to x can be used for estimation but this is extremely complex. However its close relationship to the likelihood for a mixutre of two normal distribution.

This model is of interest in that it is the alternate hypothesis for the density dependent tests discussed in section 2.4. The above shows the difficulty in setting up a test of b = 1.

(ii) Gompertz $F(x) = \log(k/x)$

This gives the equation

$$dx_t = \rho\, x \log(k/x)dt + x \log(k/x)d\beta_t$$

Smith and Tuckwell (1974) examine this model. The transformation to a Wiener process is

$$y = -\log(\log(k/x))$$

giving

$$f(x_t, t|x_0) = \frac{[x\log(k/x_t)]^{-1}}{\sqrt{2\pi\sigma^2 t}} \exp\left[-\frac{1}{2\sigma^2 t}\left[x\log\left[{}^{k}/_{x_t}\right] - x_0\log\left[{}^{k}/_{x_0}\right] - \rho t\right]^2\right]$$

The log likelihood can be formed via

$$\Sigma \log f(x_t, t|x_{t-1})$$

$$= \text{const} - \frac{n}{2}\log\sigma^2 - \Sigma \log x_t - \Sigma \log\left[\log\left[{}^{k}/_{x_{t-1}}\right]\right]$$

$$- \frac{1}{2\sigma^2}\Sigma\left[x_t \log\left[{}^{k}/_{x_t}\right] - x_{t-1}\log\left[{}^{k}/_{x_{t-1}}\right] - \rho\delta t\right]^2/t_i$$

where $\delta_t = t_i - t_{i-1}$.

If k and x_0 are known and $z_t = y_t - y_{t-1}$ then the maximum likelihood/weighted least squares estimate of ρ is given by

$$\min \Sigma (z_t - \rho\, \delta t)^2/t_i$$

$$\hat{\rho} = \frac{\Sigma\, z_t\, \delta_t/t}{\Sigma\, \delta_t{}^2/t}$$

(iii) Logistic $F(x) = 1 - {}^{x}/_{k}$

This gives

$$dx_t = \rho x_t(1 - x_t/k)dt + x_t(1 - x_t/k)d\beta_t$$

Using the transformation

$$y_t = \log(x_t/(1 - x_t/k))$$

It reduces to a Weiner process on $(-\infty, \infty)$ giving

$$f(x_t, t|x_0) = \frac{1}{(\sqrt{2\pi\sigma^2 t})x_t(1 - x_t/k)}\, x$$

$$\exp\left\{-\frac{1}{2\sigma^2 t}\left[\log\left[\frac{x_t(k - x_0)}{x_0(k - x_t)}\right] - \rho t\right]^2\right\}$$

From this result the likelihood can be formed

$$\Sigma \log(f(x_t, t|x_{t-1})$$

The above pdf was given by Tuckwell (1974), Kiester and Barakat also consider the model but with an auto correlated noise. In this case y_t has a normal distribution with

$$\text{mean} = \rho t + y_0$$

$$\text{variance} = \sigma^2 \int_0^t \int_0^t \text{corr}(t_1, t_2) dt_1\, dt_2$$

where $\text{corr}(t_1, t_2)$ is the correlation function of r, applying the transformation the above is obtained except for the $\sigma^2 t$ term being replaced by the variance expression given.

(iv) Richards $F = \frac{1}{n}\left\{1 - (x/A)^n\right\}$

$$\frac{dx}{dt} = r\, x\{1 - (x/A)^n\}/n$$

Let $\omega^n = x^n \Rightarrow d\omega = n\, x^{n-1}\, dx$

Hence

$$\frac{d\omega}{dt} = r\, \omega(1 - \omega/A^n)$$

ie a logistic with $K = A^n$

Hence solution is given by

$$f(x_t|x_0) = \frac{n\, x^{n-1}}{(\partial\, 2\pi\sigma^2 t)x_t(1 - (x_t/A)^n)^x}$$

$$\exp\left\{-\frac{1}{2\sigma^2 t}\left[\log\left[\frac{x^n(A^n - x_0^n)}{x_0^n(A^n - x^n)}\right] - \rho t\right]^2\right\}$$

The likelihood being found in a similar way to that for the logistic.

With the last three models instead of assuming a random rate of growth a random carrying capacity K could be examined. Tunelli (1977) considers some of the problems with random K models. The main object of this project is to consider the analysis of growth and hence only random r will be examined as we will assume that during the period of interest the carrying capacity is constant or can be modelled using environmental variables.

The distributions given above in general do not give a closed form for the moments, however, these may be obtained by numerical integration. the complication is due to the non-linearity of the equations. Under the mean square calculus

$$E\left[\frac{dx}{dt}\right] = \frac{d\ E(x)}{dt} = E(f(x,\ t)$$

However if f(x, t) is not linear then

$$E(f(x,\ t)) \neq f(E(x),\ t)$$

A Taylors series expansion may be used to obtain

$$E(f(x,\ t)) \triangleq f(E(x),\ t) + \frac{1}{2}\frac{\partial^2 f}{\partial x^2}\ var(x)$$

Thus the approximate value for the mean is given by the solution to

$$\frac{d\mu}{dt} = f(\mu,\ t) + \frac{1}{2}\frac{\partial^2 f}{\partial x^2}\ var(x) \Big|_{\ x\ =\ \mu}$$

This differential equation will, in general, not have an explicit solution.

5.3.4 Further Analysis

In most of the models the transformation y has lead to $y_t - y_{t-1}$ having a normal distribution independent of $z_t = y_t' - y_{t-1}'$ $t \neq t$. So

$$\xi_t = \frac{z_t - \hat{\rho}t}{\hat{\sigma} \sqrt{t}} \text{ is approximately } N(0,1)$$

Hence the suitability of the model can be examined by means of a Normal probability plot. In the general case

$$\xi_t = \frac{\int_{x_{j-1}}^{x_j} \hat{f}_x(x)dx - \int^{t} \hat{f}_t(\tau)d\tau}{\sigma^2 \hat{G}(t_i)} \quad \text{is approximately } N(0,1)$$

Standard errors of estimates may be obtained from the second derivative log likelihood matrix using either analytic or numerical derivatives.

5.3.5 Continuous Time Kalman Filter

We shall first consider the linear case. The situation of interest in when the underlying dynamics are in continuous time, ie differential equations, but we sample at discrete times. Maybeck (1979) points out that there are two approaches

(a) Use a continuous time Kalman-Filter and discretize the result

(b) Change the continuous time model to a discrete time model via

Continuous time model $\quad \dot{x}_t = f(t)x_t + g(t)\omega(t)$

Discrete time model $\quad x_{t+1} = \varphi(t+1,t)x_t + \int_t^{t+1} \varphi(t+1,\tau)g(\tau)\omega(\tau)d\tau$

where $\quad \frac{d}{dt}\varphi(t,\ t^*) = f(t)\varphi(t,\ t^*)$

and $\varphi(0,\ 0) = 1$

Maybeck gives two reasons for preferring (b) to (a)

(i) (b) is better behaved because the solution to (a) involves solving the Riccati differential equations which are difficult (see later).

(ii) The discretization of a continuous filter is an approximation to an optimal discrete time filter whereas a discrete time filter based on an equivalent discrete time model involves no approximations.

The continuous Kalman filter is described by the following equations (eg Gelb (1974)) in the scalar case

$\dot{x}_t = f(t)x_t + g(t)\omega_t$ - system equation

where ω_t is white noise,

$$E(\omega_t\ \omega_{t'}) = \delta(t - t')Q \qquad (40)$$

$y_t = h(t)x_t + e_t$ - measurement equation

$var(e_t) = R.$

Then

$$\dot{\hat{x}}_t = f(t)\hat{x}_t + K_t[y_t - h(t)\hat{x}_t] \qquad (41)$$

$$\dot{P}_t = 2f(t)P_t + g^2(t)Q - K_t^2 R \quad \text{(The Riccati Equation)} \qquad (42)$$

where

$K_t = P_t h(t)/R$

if ω_t and e_t are independent (43)

or $K_t = (P_t\ h(t) + g(t)c(t))/R$

if $E(\omega_t\ e_t^*) = c(t)\ \delta(t - t^*)$

The discretized version is given by using (41)

$\hat{x}_{t+1-} = \varphi(t+1,\ t)\hat{x}_t$

and

$$P_{t+1-} = \varphi^2(t+1,\ t)P_t + Q \qquad (44)$$

where

$$Q_\varepsilon = \text{var}(\varepsilon) \qquad \varepsilon = \int_t^{t+1} \varphi(t+1,\ \tau)g(\tau)\omega(\tau)d\tau \qquad (45)$$

also

$$K_t = P_{t-}\ h(t)/(h^2(t)P_{t-} + R) \qquad (46)$$

and

$$P_t = [1 - K_t\ h(t)]P_{t-}$$

giving

$$\hat{x}_t = \hat{x}_t(y - h(t)\hat{x}_{t-})$$

The solution of the Riccati equation when both f and g are time invariant is given by

$$P_t = (P_0 + g^2Q/2f)e^{2ft} - g^2Q/2f \qquad (47)$$

In general time varying coefficients will require numerical solution to the equation.

Imbedding the continuous time in the discrete formulation for the constant coefficient case gives

$$\varphi(t,\ t^*) = e^{f(t - t^*)}$$

and

$$\int_t^{t+1} \varphi(t+1,\ \tau)g.\omega(\tau)d\tau = g\ e^{f(t+1)} \int_t^{t+1} e^{-f\tau}\ \omega(\tau)d\tau =$$

and

$$\text{var}(\varepsilon) = g^2 \cdot \frac{Q}{2f}\ (e^{2f} - 1) \text{ from (40)}$$

Thus from (45)

$$P_{t-} = \varphi^2 P_{t-1} + \text{var}(\varepsilon)$$

$$= \left[P_{t-1} + \frac{g^2Q}{2f}\right]e^{2f} - g^2Q/2f$$

which is identical with (47) if $t = 1$, $P_0 = P_{t-1}$.

Hence the two methods will be identical in this case.

Applying either method to the model

$$\frac{dx}{dt} = rx + \omega_t$$

ie $f = r$ and $g = 1$

Also take $H(t) = 1$ then

$\hat{x}_{t-} = \hat{x}_{t-1}\, e^{r\,\delta t}$ where δ_t = time interval

$z_t = y_t - \hat{x}_{t-}$

$P_{t-} = (P_{t-1} + Q/_{2r})e^{2r\ \delta t} - Q/_{2r}$

$K_t = P_{t-}/(P_{t-} + R)$

and

$\hat{x}_t = \hat{x}_{t-} + K_t\, z_t$.

5.3.6 Non-Linear Kalman Filter in Continuous Time

The general form of the equations will be

$$\dot{x}_t = f(x_t, t) + g(x_t, t)\omega_t$$

or

$$dx_t = f(x_t, t)dt + g(x_t, t)d\beta_t$$

with measurement equation

$$y_t = h(x_t, t) + e_t$$

When developing the Kalman filter for such models the Itô calculus is used (eg Maybeck (1982)).

As with the non-linear difference equation, using a Taylor's series expansion a truncated second order filter can be obtained (Maybeck (1982)).

The propagation equations are

$$\dot{\hat{x}}_t = f(\hat{x}_t, t) + \frac{1}{2}\frac{\partial^2 f}{\partial x^2} P_t$$

$$\dot{P}_t = 2\frac{\partial f}{\partial x} P_t + g^2(\hat{x}_t, t)Q + \frac{\partial g(\hat{x}_t, t)^2}{\partial x} P_t Q \qquad (48)$$

$$+ g(\hat{x}_t, t)\frac{\partial^2 g(\hat{x}_t, t)}{\partial x^2} P_t Q$$

If the measurement equation is

$$y_t = x_t + e_t$$

then

$$K_t = P_{t-}/(P_{t-} + R_t) \qquad (49)$$

$$\hat{x}_t = \hat{x}_{t-} + K_t(y_t - \hat{x}_{t-})$$

$$P_t = P_{t-} - K_t P_{t-}$$

As in the non-linear discrete time case the residuals

$$z_t = y_t - \hat{x}_{t-}$$

will have variance, approximately equal to

$$P_{t-} + R_t$$

but they may not be normal, so parameters may be estimated by minimising

$$\sum_t z_t^2/(P_{t-} + R_t) \qquad (50)$$

with respect to the parameters.

5.3.7 Application to Growth Models

(i) $$\frac{dx}{dt} = rx = \rho x + x\,\omega_t$$

$$f = \rho x \qquad g = x$$

So

$$\dot{P}_t = 2\rho P_t + x_2 Q + P_t Q = (2\rho + Q)P_t + x^2 Q$$

$$\Rightarrow \quad P_{t-} = \left[P_{t-1} + \left[\frac{x^2\,Q}{2\rho + Q}\right]\right] e^{(2\rho + Q)\,\delta t} - \frac{x^2\,Q}{2\rho + Q}$$

$$\hat{x}_{t-} = x_{t-1}\, e^{\rho\,\delta t}$$

$$\left.\begin{aligned} z_t &= y_t - \hat{x}_{t-} \\ K_t &= P_{t-}/(P_{t-} + R_t) \\ \hat{x}_t &= \hat{x}_{t-} + K_t\, z_t \\ P_t &= (1 - K_t)P_{t-} \end{aligned}\right\} \qquad (51)$$

In this model the non-linearity comes from the form of the state-space error term.

(ii) Logistic $$\frac{dx_t}{dt} = rx(1 - x_t/K) = \rho x(1 - x_t/K) + x(1 - x_t/K)\omega_t$$

$$f = \rho x(1 - x_t/K)P_t \qquad g = x_t(1 - x_t/K)$$

so

$$\dot{P}_t = 2\rho(1 - {}^{2x_t}/_K)P_t + x_t^2(1 - {}^{x_t}/_K)^2Q + (1 - {}^{2x_t}/_K)^2P_tQ + x(1 - {}^{x_t}/_K)(1 - {}^{2}/_K)P_tQ$$

$$= (2\rho(1 - {}^{2x_t}/_K)/Q + (1 - {}^{2x_t}/_K)^2 - \frac{2x_t}{K}(1 - {}^{x_t}/_K))QP_t + x_t^2(1 - {}^{x_t}/_K)^2Q$$

$$= AQP_t + BQ \qquad (52)$$

giving

$$P_{t-} = (P_{t-1} + {}^{B}/_A)e^{AQ\delta t} - {}^{B}/_A \qquad (53)$$

with A and B as functions of x_t.

Now

$$\dot{\hat{x}}_t = \rho\hat{x}_t(1 - \hat{x}_t/K) - {}^{\rho}/_{K}P_{t-}$$

However, P_{t-} depends on x_t and the resulting equation could only be solved using numerical methods. If P_{t-} is assumed fixed then

$$\dot{\hat{x}}_t = - {}^{\rho}/_K(\hat{x}_t^2 - K\hat{x}_t + P_{t-})$$

$$\Rightarrow \quad \int_{\hat{x}_0}^{\hat{x}_t} \frac{dx}{x^2 - Kx + P_{t-}} = - \rho(t - t_0)/K$$

$$\Rightarrow \quad \left[\frac{1}{((K^2 - 4P_{t-})^{\frac{1}{2}}} \log\left[\frac{2x - (K + (K^2 - 4P_{t-})^{\frac{1}{2}})}{2x - (K - (K^2 - 4P_{t-})^{\frac{1}{2}})}\right]\right] = - \rho(t-t_0)/K$$

Let

$$2A = K + (K^2 - 4P_{t-})^{\frac{1}{2}}$$

$$2B = K - (K^2 - 4P_{t-})^{\frac{1}{2}}$$

and

$$C^2 = 1 - 4P_{t-}/K^2$$

then

$$\frac{\hat{x}_t - A}{\hat{x}_t - B} \quad \frac{\hat{x}_0 - A}{\hat{x}_0 - B} = e^{-PC(t-t_0)}$$

$$\Rightarrow \quad \hat{x}_t = \frac{A - B\left[\frac{\hat{x}_0 - A}{\hat{x}_0 - B}\right]e^{-PC(t-t_0)}}{1 - \left[\frac{\hat{x}_0 - A}{\hat{x}_0 - B}\right]e^{-PC(t-t_0)}} \qquad (54)$$

If the usual logistic boundary conditions of $x_0 = K/2$ is used this reduces to

$$\hat{x}_t = \frac{A + B\, e^{-PC(t-t_0)}}{1 + e^{-PC(t-t_0)}}$$

Further if $P_{t-} = 0$ then $A = K$, $B = 0$ and $c = 1$

so this reduces to the usual logistic.

The results (53) and (54) can be combined in a number of ways

(a) Use the value of $\hat{x}$ at t-1 to evaluate P_{t-} and hence $\hat{x}_{t-}$

(b) Use the value of $\hat{x}_{t-}$ from (i) to give a value for P_{t-} and hence a modified estimate for $\hat{x}_{t-}$

(c) Use a weighted estimate of $\hat{x}_{t-1}$ and $\hat{x}_{t-}$ from (i) to obtain value of P_{t-} and hence a further estimate of $\hat{x}_{t-}$.

Given estimates of P_{t-} and $\hat{x}_{t-}$ the equations can be used to perform the steps of the Kalman filter.

(iii) Logistic with density independent error ie

$$\frac{dx_t}{dt} = \rho x(1 - x/K) + \omega_t$$

$$f = \rho x(1 - x_t/K) \qquad g = 1$$

So

$$\dot{P}_t = 2\rho(1 - 2x_t/K)P_t + Q = AP_t + Q$$

giving

$$P_{t-} = (P_{t-1} + Q/A)e^{A\,\delta t} - Q/A \qquad (55)$$

This is then combined with (54) in the way described above.

(iv) Gompertz

$$\frac{dx_t}{dt} = r\,x_t\,\log(K/x_t) = \rho\,x_t\,\log(K/x_t) + x_t\,\log(K/x_t)\omega_t$$

$$f = \rho\,x_t\,\log(K/x_t) \qquad g = x_t\,\log(K/x_t)$$

So

$$P_{t-} = (P_{t-1} + B/A)eAQ\,\delta t - B/A \qquad (56)$$

where

$$A = \frac{2\rho}{Q}(\log(K/x_t) - 1) + (\log(K/x_t) - 1)^2 + x_t\,\log(K/x_t)(-1/x_t)$$

$$= \frac{2P}{Q}(\log(K/x_t) - 1) + \log^2(K/x_t) - 3\,\log(K/x_t) + 1$$

and

$$B = x_t^2\,\log^2(K/x_t)$$

Also

$$\frac{d\hat{x}_{t-}}{dt} = \rho\,\hat{x}_{t-}\,\log\left[K/\hat{x}_{t-}\right] - \frac{P_{t-}}{2\hat{x}_{t-}} \qquad (57)$$

This equation can only be solved numerically. However, if the bias correction term is ignored then the Gompertz differential equation can, of course, be solved.

(v) Richards

$$\frac{dx_t}{dt} = \frac{r}{n}\,x_t\left[1 - \left[\frac{x_t}{K}\right]^n\right]$$

$$= \frac{\rho}{n}\,x_t\left[1 - \left[\frac{x_t}{K}\right]^n\right] + \frac{x_t}{n}\left[1 - \left[\frac{x_t}{K}\right]^n\right]\omega_t$$

Let

$$a_1 = \frac{1}{n} - \frac{n+1}{n}\left[\frac{x_t}{K}\right]^n$$

and

$$a_2 = -(n+1)\, x_t^{n-1}/K^n$$

Then

$$\dot{P}_t = \left[\frac{2\rho}{Q} a_1 + a_1^2 + a_1 a_2\right] QP_t + \frac{x_t^2}{n^2}\left[1 - \left[\frac{x_t}{K}\right]^n\right]\omega_t$$

$$= AQP_t + BQ \qquad (58)$$

with solution as before.

Also

$$\frac{d\hat{x}_{t-}}{dt} = \frac{\rho}{n} x_t\left[1 - \left[\frac{x_t}{K}\right]^n\right] - \frac{(n+1)x_t^{n-1}}{2K_n} \qquad (59)$$

Again the equation can only be solved numerically, however, if the bias correction term is ignored the equation can be solved to give

$$\hat{x}_{t-} = K\left\{1 + \exp(\beta - r\delta t)\right\}^{-\frac{1}{n}}$$

where $\beta = \log\left[1 + \left[\frac{K}{x_0}\right]^n\right]$ (60)

The two solutions could be combined as considered previously.

This reduced form of the Kalman filter is known as the extended Kalman filter.

(vi) Further Models

In principle any growth model can be put in the above form. However, in general the equations for P_{t-} and $\hat{x}_{t-}$ will have to be solved by numerical methods.

A Second Approach

In the above the measurement equation of the form

$$y_t = x_t + e_t$$

was used and the non-linearity coming from the system equation.

We could consider a linear system equation for the rate of growth r ie

$$r_t = r_{t-1} + \varepsilon_t$$

or

$$\dot{r}_t = \omega_t$$

and have a non-linear measurement equation

$$y_t = f(r, t) + e_t$$

However such a model would not maintain the continuity of the growth, but if instead of modelling the growth, the growth increment is modelled such a set up is feasible.

For the logistic model

$$\left[\log\left[\frac{x - k}{x}\right]\right]_{x_1}^{x_2} = - r(t_2 - t_1) = - r \text{ if at equal time intervals}$$

writing as $f_t^{-1}(x_2 - x_1) = r$

$$(x_2 - x_1) = f(r)$$

so

$$y_t = f_t(r) + e_t \text{ as above.}$$

5.4 Discussion

This chapter has aimed to show the availability of models and techniques that allow a dynamic approach to growth. Such approaches often give greater insight than the more usual static models. They also allow changes in the environment to be incorporated into the models by simply changing parameter values at different points in time.

The use of the Kalman filter approach to modelling in biology would seem to have great potential.

6. Multivariate Growth Models

6.1 Introduction

In this chapter the situation in which more than one measurement is made on a growing organism or population is considered. Both dynamic and static models have been used to describe such situations, some of the dynamic models for multi species populations have already been considered in the introduction. Static models for the relationship between parts of an organism will now be reviewed.

6.1.1 The Allometric Relationship

The most common relationship between two parts η and ξ, is

$$\eta = a\xi^{\beta} \qquad (1)$$

or

$$\log \eta = \log a + \beta \log \xi \qquad (2)$$

hence this is known as the linear allometric relationship. An insight into the relationship can be gained by considering the relative growth rates R_{η} and R_{ξ}. From 6.1.1(2) it can be seen that

$$R_{\eta} = \beta R_{\xi} \qquad (3)$$

thus the relative growth rate of one aspect is a constant proportion of the relative growth rate of the other under this model.

This model has been found to give a reasonable fit for a wide range of different situations.

6.1.2 Fitting the Linear Allometric Relationship

Let y_i, x_i be the observed values of log η_i and log ξ_i for an independent sample with

$$y_i = \log \eta_i + e_i \qquad (4)$$

$$x_i = \log \xi_i + \epsilon_i$$

with

$$\log \eta_i = \alpha + \beta \log \xi_i$$

Then the problem of estimating α and β is that of fitting a functional relationship. This is a well known problem (eg Sprent (1969), Kendall and Stuart (1973), Causton and Venus (1981)). The solution requires some knowledge of the variances of e_i and ε_i. The usual assumption is that

$$\varphi = \frac{\text{var}(e_i)}{\text{var}(\varepsilon_i)}$$

is known in which case it can be shown that if e_i and ε_i are independent

$$\hat{\beta} = \frac{S_{yy} - \varphi S_{xx} + \sqrt{\{(S_{yy} - \varphi S_{xx})^2 + 4\varphi S^2_{xy}\}}}{2S_{xy}} \qquad (5)$$

where S.. represents the corrected sum of squares/cross products. If the errors are not independent but have variance - covariance matrix = W then Sprent (1969) shows that

$$\hat{\beta} = \frac{S_{xy} - \lambda W_{12}}{S_{xx} - \lambda W_{11}}$$

where λ is the smaller root of (6)

$$|S - \lambda W| = 0$$

where

$$S = \begin{bmatrix} S_{xx} & S_{xy} \\ S_{xy} & S_{yy} \end{bmatrix}$$

This reduces to (5) if $W_{12} = 0$ and further if $\varphi = 1$ then to the first principal component of S.

6.1.2 Multivariate Generalisations

The above approach can be generalised to more than two dimensions. Jolicoeur (1963 and 1984) considers the first principal component of the variance - covariance matrix. He showed that if all the proportions of the organism were to remain constant then the first principal component would be

$$\alpha' = \left[\frac{1}{\sqrt{p}} \cdots \frac{1}{\sqrt{p}}\right] \qquad (7)$$

This hypothesis can be tested using standard tests (Anderson (1963), Morrison (1967))

$$\chi^2_{(p-1)} = (N - 1)\left[\lambda\, \alpha'\, S^{-1}\, \alpha + \frac{\alpha\, S\, \alpha'}{\lambda} - 2\right] \qquad (8)$$

Jolicoeur (1984) proposed an altern ative test

$$F_{(p-1,\ N-p)} = (N - p)(\alpha\, S\, \alpha'\, \alpha\, S^{-1}\, \alpha' - 1)/(p - 1) \qquad (9)$$

An alternative approach suggested by Hopkins (1966) is to use a factor analysis type model

$$S = \psi + \triangle$$

where $\triangle$ is the variance - covariance matrix of the independent variations in each dimension. Various assumptions can be made about the form of $\triangle$ which are needed before inferences can take place.

Both approaches are reviewed by Sprent (1968 and 1972).

6.1.3 Time Dependent Relationships

Hole et al (1984) propose a series of models for growing crops in which the relationship between parts varies in time. The particular interest is the shoot/storage root relationship in root crops. Their principal model is

$$y_i = \alpha - \Gamma\,(\text{time}) + \beta\, x_i$$

They also consider polynomials in x. However, the problem of variation in both variables is ignored and ordinary regression analysis carried out.

6.2 Non-Linear Allometric Relationships and Phase Planes

6.2.1 Models

If the growth of two elements can be described by the model

$$\begin{aligned} \frac{dy_1}{dt} &= -\beta_1\, f(y_1, y_2, t)y_1 \\ \frac{dy_2}{dt} &= -\beta_2\, f(y_1, y_2, t)y_2 \end{aligned} \qquad (1)$$

then the phase plane portrait (Arrowsmith and Place (1982)) is the solution of

$$\frac{\beta_2\, dy_1}{y_1} = \frac{\beta_1\, dy_2}{y_2} \qquad (2)$$

ie $\beta_2 \log y_1 - \beta_1 \log y_2 = \text{const}$

the linear allometric relationship.

The relationship 6.2 can be generalised in a number of different ways, for instance the Lotka-Volterra equations

$$\begin{aligned} \frac{dy_1}{dt} &= \beta_1\, y_1(1 - \alpha_{12}\, y_2) \\ \frac{dy_2}{dt} &= \beta_2\, y_2(1 - \alpha_{21}\, y_1) \end{aligned} \qquad (3)$$

these have phase plane portraits given by

$$\beta_2 \log y_1 - \beta_2 \alpha_{21} y_1 - \beta_1 \log y_2 + \beta_2 \alpha_{12} y_2 = \text{const} \qquad (4)$$

Turner (1978) considers further generalisations. First

$$\frac{dy_1}{dt} = -\beta_1 y_1 \left[\frac{f(y_1, y_2\ t)}{1 - d_{21} y_1}\right] \qquad (5)$$

$$\frac{dy_2}{dt} = -\beta_2 y_2 \left[\frac{f(y_1, y_2\ t)}{1 - d_{12} y_2}\right]$$

which will have the same phase portrait as (4).

Secondly

$$\frac{dy_1}{dt} = \beta_1 y_1 \{k_1 - f(y_1, y_2, t)\}$$

$$\frac{dy_2}{dt} = \beta_2 y_2 \{k_2 - f(y_1, y_2, t)\} \qquad (6)$$

giving phase portraits

$$\beta_2 \log y_1 - \beta_1 \log y_2 = \beta_1 \beta_2 (k_1 - k_2)t + \text{const} \qquad (7)$$

Combining the ideas of (4) and (7) the general phase portrait of

$$\Sigma \beta_i \log y_2 + \Sigma \alpha_i y_i + \gamma t = 1 \qquad (8)$$

is derived for a multi-dimensional problem.

Using a general set of equations considered in the introduction

$$\frac{dy_1}{dt} = y_1 F_1(y_1, y_2)$$

$$\frac{dy_2}{dt} = y_2 F_2(y_1, y_2) \qquad (9)$$

Then phase portrait is obtained by solving

$$\frac{dy_1}{dy_2} = \frac{y_1 F_1(y_1, y_2)}{y_2 F_2(y_1, y_2)} \qquad (10)$$

If no analytic solution is available for (10) then a numerical method could be employed (eg Runge-Kutta method) checking for singularities given by

$F_2(y_1, y_2) = 0.$

6.2.2 Model Fitting

Turner (1978) suggested fitting (8) using least squares with the constant 1 as the dependent variable. This is criticised by Griffiths and Sandland (1982 and 1984) because the independent variables are stochastic and that the regression is not location invarient. Further they point out that the inference procedure will be distorted by the high correlation between y_i and log y_i. In the 1984 paper they considered alternate estimation procedures

(i) Using a generalised principal component approach. This requires solution of

$$|S - \lambda \Delta| = 0 \qquad (11)$$

where S is the variance-covariance matrix for both y_i's and log y_i's and Δ is a block diagonal matrix with 2x2 elements

$$\begin{bmatrix} 1 & v_i \\ v_i & v_i^2 \end{bmatrix} \qquad v_i = {}^{1}/_{E(y_i)} \qquad (12)$$

This is based on a first order Taylor's series approximation for the variance of log y_i and a local correlation of unity between y_i and log y_i.

(ii) Their main approach is to consider

$$F_i = \alpha_i\, y_i + \beta_i \log y_i \qquad (13)$$

$$\text{and to assume } F(t_j) \sim N(\theta_j, V) \qquad (14)$$

where

$$F'(t_j) = (F_1(t_j) \ \ldots \ F_k(t_j)), \ \theta'_j = (\theta_{ij} \ \ldots \ \theta_{kj})$$

and

$$\sum_{i=1}^{k} \theta_{ij} + \gamma\, t_j = 1 \quad \forall_j$$

The assumption of normality is only justified on pragmatic grounds and the fact that both y_i and $\log y_i$ cannot be normal.

Three forms of v are considered

$$\begin{aligned} &(a) \quad v_{ij} = (\alpha_i\, y_{ij} + \beta_j)^2\, \sigma^2 \\ &(b) \quad v_{ij} = (\alpha^2_i\, y^2_{ij} + \beta^2_i)\, \sigma^2 \\ &(c) \quad v_{1j} = v_{2j} = \sigma^2 \end{aligned} \tag{15}$$

Assuming that $\gamma = 0$ the log likelihood is in the case k=2

$$\begin{aligned} &- n \log 2\pi - n \log \sigma^2 - \frac{1}{2} \sum_{j=1}^{n} \log(v_{1j}\, v_{2j}) \\ &- \frac{1}{2\sigma^2} \left[\sum_{j=1}^{n} \frac{(F_{1j} - \theta_{ij})^2}{v_{1j}} + \sum_{j=1}^{n} \frac{(F_{2j} - 1 + \theta_{ij})^2}{v_{2j}} \right] \\ &+ \sum_{j=1}^{n} \sum_{i=1}^{2} \log(\alpha_i + \beta_i / y_{ij}) \end{aligned} \tag{16}$$

giving

$$\hat{\sigma}^2 = \frac{1}{2n} \sum_{j=1}^{n} \left[\frac{(F_{1j} - \hat{\theta}_{ij})^2}{v_{1j}} - \frac{(F_{2j} - 1 + \hat{\theta}_{ij})^2}{v_{2j}} \right] \tag{17}$$

and

$$\hat{\theta}_{1j} = \{v_{2j}\, F_{1j} + v_{1j}(1 - F_{2j})\}/(v_{1j} + v_{2j}) \tag{18}$$

noting that

$$\hat{\theta}_{2j} = 1 - \hat{\theta}_{1j} \tag{19}$$

These estimates can be substituted into (16) and the result maximised by a non-linear procedure.

6.2.3 Alternate Estimation Procedures

Two alternate estimation procedures will now be proposed. The bivariate case only will be considered for simplicity and also as it represents the main practical situation.

Recall

$$\beta_2(\log y_1 - \alpha_{21} y_1) - \beta_1(\log y_2 + \alpha_{12} y_2) = \text{const}$$

Then given α_{21} and α_{12}, letting

$$\begin{aligned} z_1 &= \log y_1 - \alpha_{21} y_1 \\ \text{and} \quad z_2 &= \log y_2 - \alpha_{12} y_2 \end{aligned} \qquad (20)$$

the problem reduces to the linear functional relationship for the z's as considered in section 6.1.2. In practice the transformation $z_i = y_i + \lambda_i \log y_i$ was used. This can be considered as a linearising transformation analagous to say the Box-Cox transformations. In order to select a suitable value of λ_i the following procedure was used.

The values of λ_1 and λ_2 were selected simultaneously so as to optimise a suitable linearity criterion. Such a criterion is given by Cox and Small (1978).

$$C(\lambda_1, \lambda_2) = [Q_{21} \quad Q_{12}] \begin{bmatrix} 1 & (2 - 3r^2) \\ r(2 - 3r^2) & 1 \end{bmatrix}^{-1} \begin{bmatrix} Q_{21} \\ Q_{12} \end{bmatrix} \qquad (21)$$

where Q_{ij} is the student test statistic for the regression coefficient of z_i on z^2_j and r is the correlation coefficient for z_1 and z_2.

The function $C(\lambda_1, \lambda_2)$ was minimised using the Nelder-Mead algorithm, the values of Q_{ij} and r being calculated by Nag routines G02BAF and G02CGF. In finding the global minimum a contour plot of C was found useful.

Having found the suitable values of λ_i hence z_i the values of β_1 and β_2 were

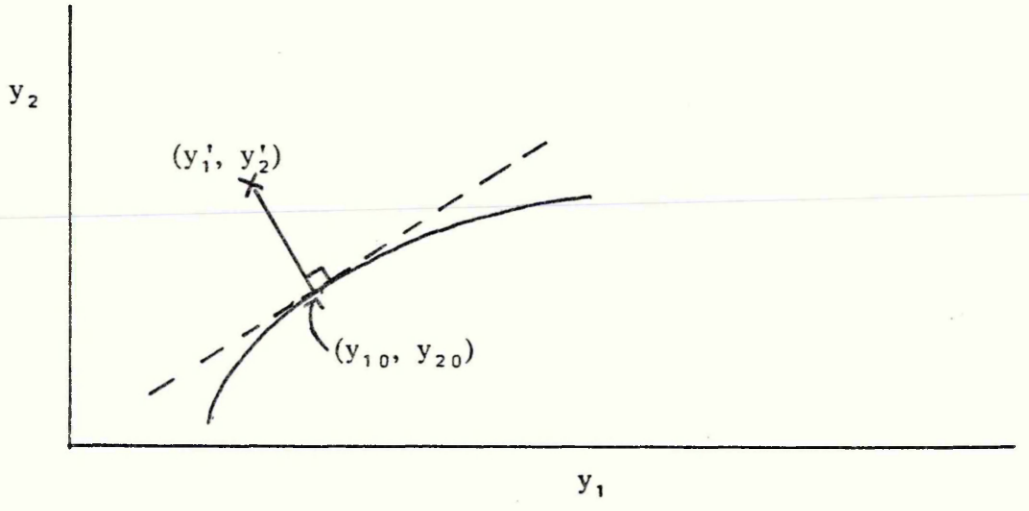
found from using (5).

The second approach was to consider the shortest distance from the observation in the (x,y) co-ordinates to the curve in the phase plane.

Considering model (9) we have

$$\frac{dy_1}{dy_2} - \frac{y_1}{y_2} \cdot \frac{F_1(y_1, y_2)}{F_2(y_1, y_2)}$$

Now let (y_{10}, y_{20}) be the point on the curve given by the solution to (10) that is closest to the observed point (y_1', y_2'). Then (y_{10}, y_{20}) must lie at the foot of the perpendicular from the curve through (y_1', y_2') as shown in the diagram



The equation of the perpendicular is

$$(y_2 - y_2') \left[\frac{-1}{\frac{dy_2}{dy_1}} \Bigg|_{y_{10}, y_{20}} \right] (y_1 - y_1')$$

ie

$$(y_2 - y_2') + \frac{y_{10}}{y_{20}} \cdot \frac{F_1(y_{10}, y_{20})}{F_{20}(y_{10}, y_{20})} (y - y_1') = 0 \qquad (22)$$

Let $P(y_1, y_2) = 0$ be the phase plane solution (ie the solution of (10) then in addition

$$P(y_{10}, y_{20}) = 0 \qquad (23)$$

Solving (22) and (23) y_{10} and y_{20} can be obtained for values of the parameters of $F_1(y_1, y_2)$, $F_2(y_1, y_2)$ and $P(y_1, y_2)$.

A generalised distance of the point (y_1', y_2') from the model may be obtained

$$d^2 = (y_1' - y_{10}, y_2' - y_{20}) \begin{bmatrix} v_1 & v_2 \\ v_2 & v_3 \end{bmatrix} \begin{bmatrix} y_1' - y_{10} \\ y_2' - y_{20} \end{bmatrix}$$

here v_1, v_2 and v_3 are chosen in relation to the variance - covariance of the stochastic errors for y_1 and y_2. Given d the parameters of the model are chosen so as to minimise

$$S^* = \sum_{i=1}^{n} d_i^2$$

The solution curve $P(y_1, y_2)$ can in some cases be obtained by analytic methods. Failing this numerical methods could be used eg fourth order Runge-Kutta method (Conte and DeBoor (1972), Patten (1971)). In addition the solution to (22) and (23) may require numerical methods although they will often reduce to a single non-linear equation.

6.2.4 Examples

The data to be used was considered by Turner (1978) and Griffiths and Sandland (1982, 1984). It is for the growth of mandibles of the stag-beetle, Cyclommatus tarandus. The data was originally collected by Dudich (1923) and was examined by Huxley (1927, 1932). The data is given in table 6.2(1). The model to be fitted is, using Griffiths and Sandland's notation,

$$\beta_1 \log y_1 + \beta_2 \log y_2 + \alpha_1 y_1 + \alpha_2 y_2 = 1$$

Using the transformation method, first the values of λ_1 and λ_2 that minimised (21) had to be found. The complexity of the $C(\lambda_1, \lambda_2)$ surface is shown by figure 6.2(1). Further the transformation

$$z = y + \lambda \log y$$

is not monotonic for all values of λ, y. Now

$$\frac{dz}{dy} = 1 + {}^{\lambda}/_{y}$$

Therefore, for a monotonic transformation

$$\lambda \neq - y^*$$

for any y^* in the range of the data considered. This means that for the beetle data

$$\lambda_1 > - 3.88 \quad \text{or} \quad \lambda_1 < - 34.5$$

and

$$\lambda_2 > - 16.5 \quad \text{or} \quad \lambda_2 < - 39.5$$

as shown in figure 6.2(2), the non-monotonic region being shaded.

Table 6.2(1)

Mandibular Length (mm) (y_1)	Body Length (mm) (y_2)
3.88	16.50
5.31	18.70
6.33	20.05
7.32	20.44
8.17	21.48
9.73	22.47
10.71	22.40
11.49	23.52
12.08	24.05
12.73	24.59
14.11	24.33
14.70	24.56
15.84	25.50
17.39	25.83
18.83	26.68
19.19	27.13
19.92	27.36
20.79	27.61
21.53	28.51
22.54	28.96
23.25	29.25
23.96	30.27
25.38	30.63
28.49	33.37
30.69	35.37
32.00	37.00
34.50	39.50

Figure 6.2(1)

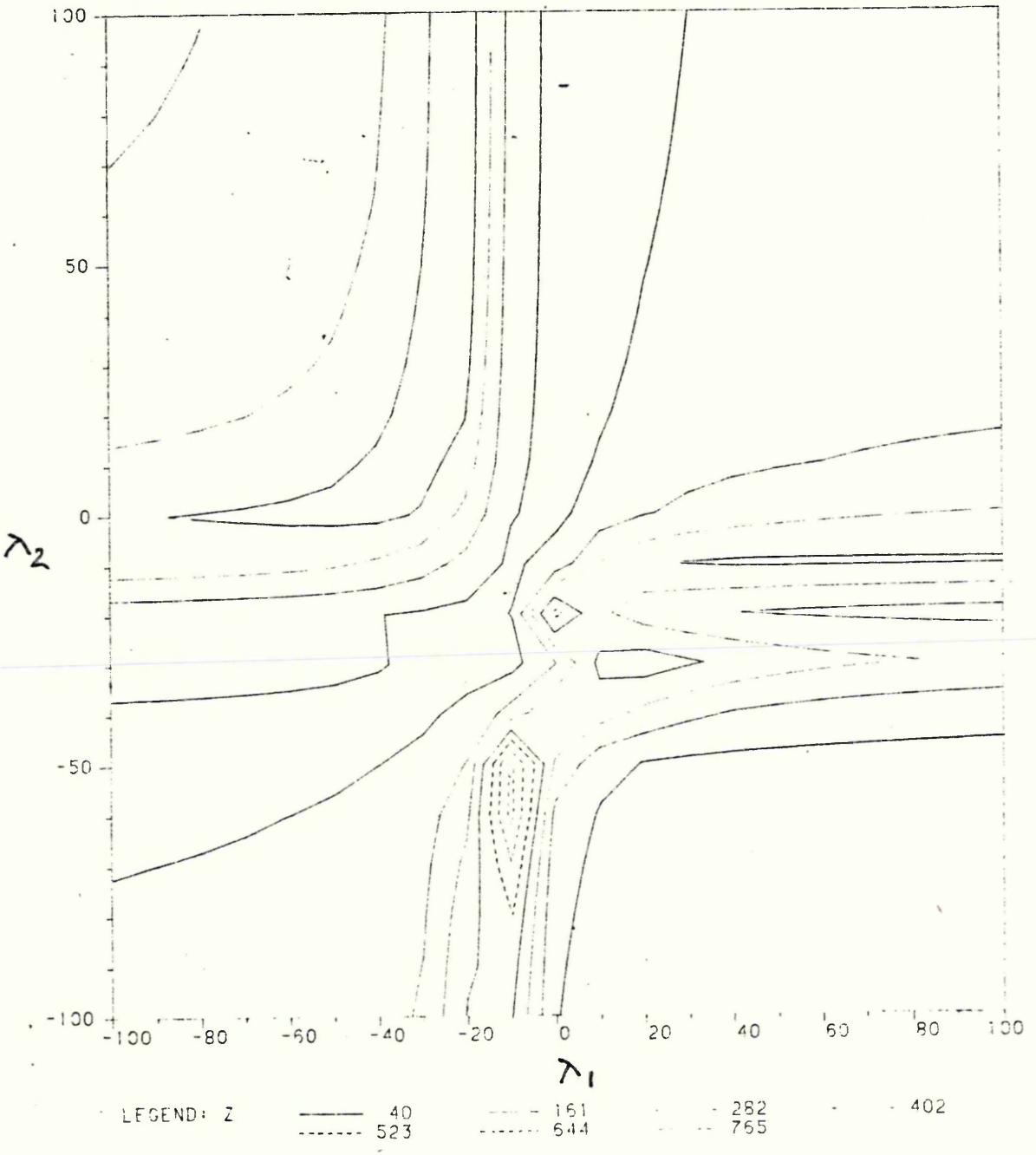
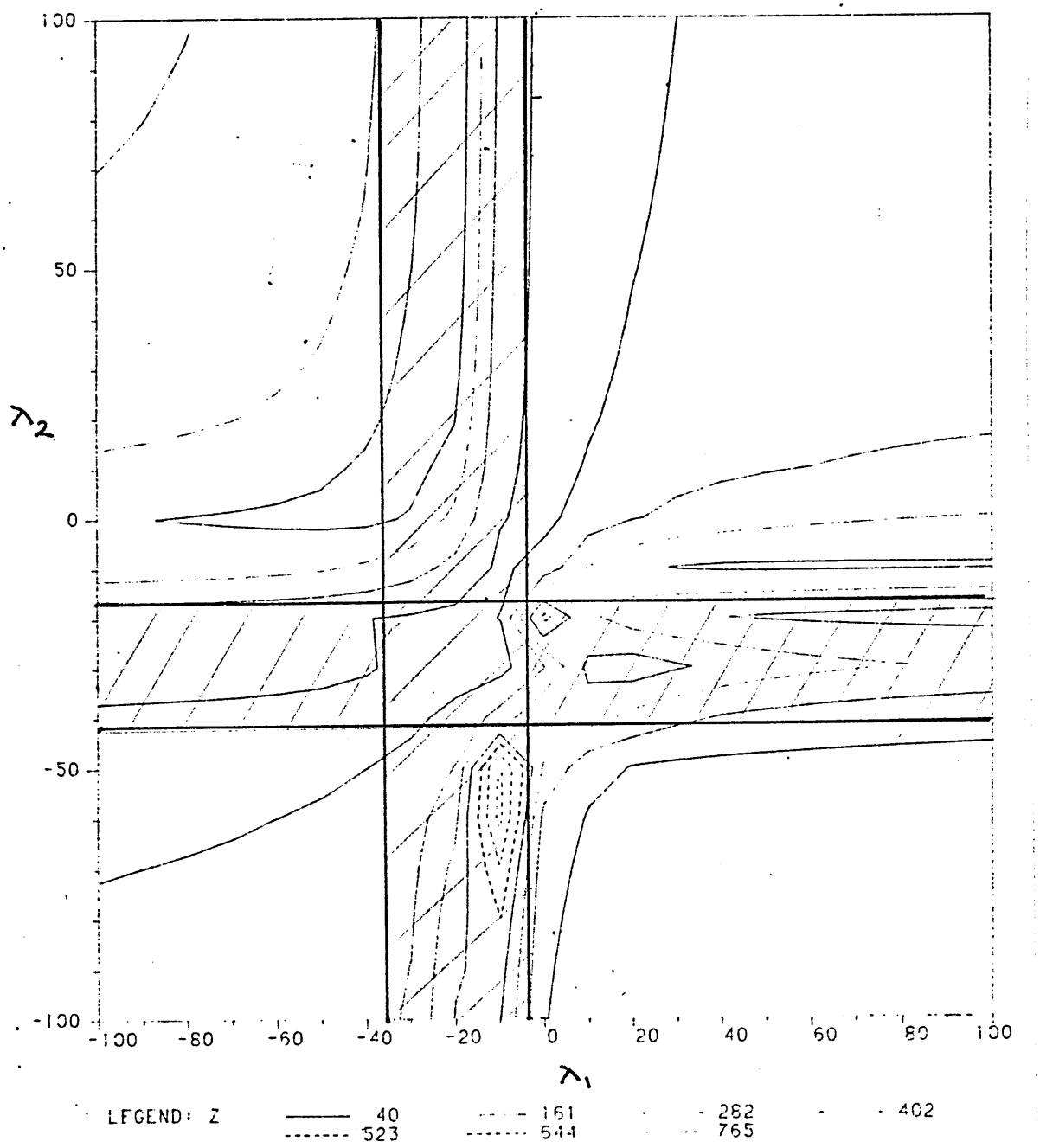

100
50
0
-50
-100
-100
-80
-60
-40
-20
0
20
40
60
80
100
λ2
λ1
LEGEND: Z
40
161
282
402
523
644
765

Figure 6.2(2)


100
50
0
-50
-100
-100
-80
-60
-40
-20
0
20
40
60
80
100
λ2
λ1
LEGEND: Z
40
161
282
402
523
644
765

Having used several starting values the minimum was found to be at

$$\lambda_1 = -161.94675 \qquad \lambda_2 = -44.80763$$

The functional relationship

$$z_1 = \beta_0 + \beta_1 z_2$$

was then fitted using (5) for values of $\varphi = 0.5, 1.0, 1.5$. The results were

φ	β_1	β_0
0.5	22.15297	2222.813
1.0	22.15265	2222.775
1.5	22.15233	2222.737

Indicating that, in this case, the result is not sensitive to the chosen value of φ. Taking $\varphi = 1.0$ the fitted model becomes

$$-.07286 \log y_1 + .44656 \log y_2 + .00045 y_1 - .00997 y_2 = 1$$

Using the second approach, the generalised least squared distances it is useful to return to the original Lotka-Volterra type equations

$$\frac{dy_1}{dt} = a_1 y_1 + a_2 y_1 y_2$$

$$\frac{dy_2}{dt} = a_3 y_2 + a_4 y_1 y_2$$

giving

$$\frac{dy_1}{dy_2} = \frac{a_1 y_1 + a_2 y_1 y_2}{a_3 y_2 + a_4 y_1 y_2}$$

with solution, taking the constant as 1,

$$a_3 \log y_1 + a_4 y_1 - a_1 \log y_2 - a_2 y_2 = 1 \qquad \text{(I)}$$

Thus comparing with Griffiths and Sandland's model

$$\beta_1 = a_3 \qquad \alpha_1 = a_4$$

$$\beta_2 = -a_1 \qquad \alpha_2 = -a_2$$

Let the observed values be ω_1 and ω_2 then the equation of the perpendicular form (ω_1, ω_2) to the curve is, from (22)

$$(y_2 - \omega_2) + \frac{y_1}{y_2}\left[\frac{a_1 + a_2 y_2}{a_3 + a_4 y_1}\right](y_1 - \omega_1) = 0$$

This gives

$$y_2^2 + y_2\left[\frac{a_2 y_1(y_1 - \omega_1)}{(a_3 + a_4 y_1)} - \omega_2\right] + \frac{a_1 y_1}{a_3 + a_4 y_1}(y_1 - \omega_1) = 0 \qquad (II)$$

Solving the quadratic (II) for y_2, the solution in terms of y_1 can then be substituted in (I) giving a single non-linear equation for y_1. This equation was solved using the secant method (Conte and de Boor (1972)). Thus obtaining values of y_1 and, substituting this value back into the solution to (II), the value of y_2.

From these values

$$r_1 = \omega_1 - y_1$$

$$r_2 = \omega_2 - y_2$$

were calculated. Assuming the form

$$d^2 = r_1^2 + r_2^2$$

Σd^2 was minimised using the Nelder-Mead algorithm.

In order to solve (II) an initial value of $y_1 = \omega_1$ was chosen, this gave $y_2 = \omega_2$. In general (II) will have 1, 2 or 0 solutions. In the most common situation of 2 solutions the solution nearest the observed value was chosen. The case of no solution arises due to the shape of the fitted function and the value of y_1 at that point in the iterative process (see later). This situation can be overcome by starting from a new initial value.

In order to find the minimum of Σd^2 for the beetle data several starting values were used, including the estimates given by Griffith and Sandland (1984), see

table 6.2(2). The minimum of Σd^2 was given by

$a_1 = -.44451$

$a_2 = .01079$

$a_3 = -.06059$

$a_4 = -.00018$

Table 6.2(2)

Parameter	Fitting Procedure							
				Griffiths and Sandland				
	Turner	Principal Component	Generalized Eigenvalue	1	2	3	Transformation Method	Generalized least squared distances
α_1	.00323	.00335	.00360	.00019	-.00047	-.00062	.00045	-.00018
B_1	-.07594	-.07582	-.08520	-.06056	-.06154	-.06433	-.07286	-.06059
α_2	-.01527	-.01542	-.01838	-.01080	-.00945	-.00891	-.00997	-.01079
B_2	.47664	.47792	.49780	.44446	.43781	.47651	.44656	.44451

Table 6.2(2) combines the results of Turner (1978), Griffiths and Sandland (1984) and those of this report. The generalised least squared distances gave a solution very similar to Griffiths and Sandland. The transformation method gives a result between the Griffiths and Sandland results and the others. There is, however, an important difference between the result that can be obtained by the transformation method and the other methods. As reported by Griffiths and Sandland (1984) the phase plane solution can have several shapes, including

(A) A monotonic single branched trajectory

(B) A two branched non-monotonic function with asymptotes parallel to either axis

(C) A closed loop.

Turner's method gives solution of type B with asymptotes parallel to the body length axis. The generalised eigenvalue method also gives a solution of type B but with the asymptotes parallel to the manibular length axis.

The transformation method can only give solution of type A and this is the type required for allometric modelling as with the beetle data.

It was the discontinuity in the solution of type B that can lead to the imaginary roots of equation (II). This will happen if

$$y_2 = \frac{1 - \alpha_1 y_1 - \beta_1 \log y_1}{\alpha_2} \text{ and } \frac{\beta_2}{\alpha_2} < 0$$

for some y_1 $(y_1, y_2 > 1)$

for example in the generalised eigenvalue solution at $y_1 = 15$.

The generalised least squares approach can give the two branched solution but in this case does not.

The advantages of the new methods are

(i) The transformation approach leads to the correct form of solution for the allometric problem and is conceptually simple to understand.

(ii) The generalised least squares approach minimises error on the observed variables y_1 and y_2 not on other, less obvious, functions.

6.3 Dynamic Models

6.3.1 Multivariate OU Processes

Cumberland and Rohde (1977) show how a Multivariate Ornstein-Uhlenbeck process can be used modelling multivariate populations. They consider two models

$$(i) \quad \frac{dy_i}{dt} = u_i(t)\, y_i \qquad (1)$$

and

$$(ii) \quad \frac{dy_i}{dt} = u_i(t)\, y_i\, (1 - y_i/k_i) \qquad (2)$$

where u_i is a k-dimensional OU process. The interaction between populations is provided the correlations in the OU process. They provide conditions for the existance of such a process but no information on the behaviour of the models other than a simple simulation.

The OU process itself satisfies the equation

$$\frac{du}{dt} = -\, r\, U + \quad (t) \quad , \quad (t) \text{ being white noise} \qquad (4)$$

where as the random growth rates are normally modelled by

$$r = \rho + \quad (t), \quad \rho \text{ being constant} \qquad (5)$$

In view of this the approach using state space models and the Kalman filter for single variable models will be developed for the multivariate situation.

6.3.2 State Space Models

For compactness only two dimensions will be considered but the results could be extended to higher dimensions.

The continuous time state space models are of the form

$$x_t = f(x_t) + G(x_t)_t$$
$$y_t = h(x_t) + e_t \qquad (6)$$

where x_t, y_t, ε_t and e_t may be vectors.

There are two ways in which these models could be used

(i) Use a single x to represent a growth factor. The functions $h(x_t)$ would relate the size of the individual components to this factor. This could be seen as a dynamic non-linear factor analysis approach.

If x is used to model the logarithm of the growth factor then $\dot{x}_t$ becomes the relative growth rate giving the equations (ignoring noise)

exponential growth $\dot{x} = r$

logistic growth $\dot{x} = r\left[1 - \frac{1}{k} e^x\right]$

The function $h(x_t)$ could be assumed linear if the usual allometric relationship is used.

If x is modelled using the non-linear equations considered in section 5.3, then if using the allometric relationship $h(x_t)$ will be a logarithmic function.

Applying the Kalman filter to these situations the non-linear relationships in f( ) and h( ) will have to be approximated using Taylor's series expansions. As modelling the logarithms of the growth factor involves only one approximation this would seem better even if the differential equations are more complex in some situations.

In this model there will be an arbitrary choice of one element of the linear function h( ) as the scale of the general growth factor could be chosen at will.

(ii) The second approach would be to use a full dynamic model for all components and h( ) reduces to an identity relationship.

As in both cases a linear h( ) will be used a second order truncated filter can be used. Representing h( ) by the matrix H then the up date equations are

$$K = P_{t-} H' (H P_{t-} H' + R)^{-1}$$

$$\hat{x}_t = \hat{x}_{t-} + K (y_t - H \hat{x}_{t-}) \qquad (7)$$

$$P_t = P_{t-} - k H p_{t-}$$

The propagation equations will be

$$\dot{\hat{x}}_t = f(\hat{x}_t) + b_t$$

where

$$\hat{b}_t = \frac{1}{2} tr \left\{ \left[\frac{\partial^2 f}{\partial x^2} \right] P_t \right\} \qquad (8)$$

and

$$P_t = F_t P_t + P_t F_t' + E(G Q G')$$

where

F is the nxn matrix with elements $\frac{\partial f}{\partial x}$

and

$$E(G\ Q\ G)_{ij} = \sum_{k=1}^{S}\sum_{\ell=1}^{S}\left[G_{ik}\ Q_{k\ell}\ G_{\ell j} + tr\left[\frac{\partial G_{ik}}{\partial x}\ Q_{k\ell}\ \frac{\partial G'_{\ell j}}{\partial x}\right] P\right.$$

$$+ \frac{1}{2}\ G_{ik}\ Q_{k\ell}\ tr\left[\frac{\partial^2 G_{\ell j}}{\partial x^2}\ P\right] + \frac{1}{2}\ tr\left[\right.$$

$$\left.\left. - P\ \frac{\partial\ G_{ik}}{\partial x^2}\right] Q_{k\ell}\ G'_{\ell j}\right]$$

(Maybeck (1982)) where S is the dimension of ε_t

6.3.3 Example

To illustrate the use of the above technique the simple Lotka-Valterra model will be used

$$\frac{dx_1}{dt} = \alpha_1\ x_1 + \alpha_2\ x_1\ x_2 + x_1\ \omega_1$$

$$\frac{dx_2}{dt} = \alpha_3\ x_3 + \alpha_4\ x_1\ x_2 + x_2\ \omega$$

where ω_1 is a single white noise term. The above formulation assumes the standard deviation is proportional to size. For this model (8) gives

$$b_t = \begin{bmatrix}\alpha_2 & p_2\\ \alpha_4 & p_2\end{bmatrix}$$

where

$$P_t = \begin{bmatrix}p_1 & p_2\\ p_2 & p_3\end{bmatrix}$$

$$F \text{ is } \begin{bmatrix}\alpha_1 + \alpha_2\ x_2 & \alpha_2\ x_1\\ \alpha_4\ x_2 & \alpha_3 + \alpha_4\ x_1\end{bmatrix}$$

Hence

$$F\ P' + P\ F' = \begin{bmatrix}2(\alpha_1 + \alpha_2\ x_2)p_1 + \alpha_2\ x_1\ p_2) & \\ (\alpha_1 + \alpha_2\ x_2)p_2 + \alpha_2\ x_1\ p_3 + \alpha_4\ x_2\ p_1 + (\alpha_3 + \alpha_4\ x_1)p_2 & \\ & 2(\alpha_4\ x_2\ p_2 + (\alpha_3 + \alpha_4\ x_1)p_3\end{bmatrix}$$

and

$$E(G\ Q\ G') = \left[\begin{bmatrix} x_1^2 & x_1\ x_2 \\ x_1\ x_2 & x_2^2 \end{bmatrix} + \begin{bmatrix} p_1 & p_2 \\ p_2 & p_3 \end{bmatrix}\right] q$$

These are combined to give the equation for $\dot{P}$.

This leads to 5 non-linear differential equations to solve

$$\dot{x}_1 = \alpha_1\ x_1 + \alpha_2\ x_1\ x_2 + \alpha_2\ p_2$$

$$\dot{x}_2 = \alpha_3\ x_2 + \alpha_4\ x_1\ x_2 + \alpha_4\ p_2$$

$$\dot{p}_1 = (x_1^2 + p_1)q + 2((\alpha_1 + \alpha_2\ x_2)p_1 + \alpha_2\ x_1\ p_2)$$

$$\dot{p}_2 = (x_1\ x_2 + p_2)q + (\alpha_1 + \alpha_2\ x_2 + \alpha_3 + \alpha_4\ x_4)p_2 + \alpha_2\ x_1\ p_3 + \alpha_4\ x_2\ p_1$$

$$\dot{p}_3 = (x_2^2 + p_3)q + 2(\alpha_4\ x_2\ p_2 + (\alpha_3 + \alpha_4\ x_1)p_3)$$

These equations can then be solved using a Runge-Kutta procedure (eg Nag routine D02BAF). A suitable measurement equation is needed, the simplest being

$$y_1 = x_1 + e_1 \qquad \text{var}(e_1) = \sigma_1^2 \qquad \text{cov}(e_1\ e_2) = 0$$

$$y_2 = x_2 + e_2 \qquad \text{var}(e_2) = \sigma_2^2$$

giving

$$H = \begin{bmatrix} 1 & 0 \\ 1 & 0 \end{bmatrix} \quad \text{and} \quad R = \begin{bmatrix} \sigma_1^2 & 0 \\ 0 & \sigma_2^2 \end{bmatrix}.$$

The parameters may then be found by minimising

$$\sum_t (y_t - H\ \hat{x}_{t-})^1\ R^{-1} (y_t - H\ \hat{x}_{t-})$$

there being 9 parameters (4 α's, q, σ_1^2, σ_2^2 and two initial values $x_{1\,0}$ and $x_{2\,0}$).

6.4 Discussion

Two approaches to the analysis of multivariate growth have been presented

(i) A static model approach based on phase plane solutions to the system of differential equations.

(ii) A dynamic model approach using the Kalman filter.

The choice between the two approaches will take into consideration the following points

(a) The availability of the time series data needed for the dynamic approach. This may not always be available or the ages may only be approximate. In the case of the beetle data considered earlier only the static approach was available.

(b) The availability of a closed form for the phase plane solution. Such a solution would simplify the static approach and if no closed form solution were available in the time domain this would justify the exclusion of the time information from the modelling process.

Either of the approaches could be applied to both the growth of parts of a single organism or the competition between species in a population dynamics situation.

7. Statistical Modelling

7.1 Introduction

The previous chapters have been concerned with the problems of identifying and fitting models to data. In this chapter the objectives of modelling and the various approaches to modelling will be considered. Questions such as

Do I need a model?

What type of model do I need?

What approach to modelling shall I take?

What are the steps in modelling?

need to be considered.

Models come in many different forms. This chapter will only deal with what can be termed Mathematical models. These can be considered as a subset of symbolic models in Ackoff and Sasceni (1968) categories of

Conceptual Models

Iconic Models

Analogue Models

Symbolic Models

The models considered in chapters 3 - 6 are only some of the many different mathematical models that could be considered, but along with those discussed in the introduction, provide a cross-section of various approaches to modelling.

7.2 Purpose and Types of Models

A list of reasons for using models could be that they

Communicate fact and ideas

Generate new ideas

Predict behaviour

Provide insight into the behaviour

Clarify thinking

(Based on OU notes on Systems Behaviour)

Jeffers (1982) gives various reasons for using ecological models including providing

An orderly and logical representation of the underlying relationships.

A means of communication between different research workers.

A synthesis of available information.

Thus there are three main themes in the purposes of models

(i) Communication

(ii) Prediction

(iii) Understanding

The emphasis placed on the three aspects will depend on the particular situation. A contrast between (i) and (ii) and (iii) is often set up. Curnow (1977) contrasted descriptive and explanatory models

Descriptive

Summarising for consideration

Decision Making

Isolation of important factors in above.

Explanatory

Increased understanding leading to further advances (extrapolations) and improved decision making.

He associated static models with descriptive purposes and dynamic models with explanatory purposes.

Gilchrist (1984) describes a contrast between conceptual and empirical aspects. The conceptual approach uses "logical reasoning, 'known theory', to obtain a model". Whereas the empirical uses only the empirical evidence, the data. In practice a mixture of the two approaches is used, what Gilchrist calls the eclectic approach.

A similar contrast is the mechanistic-empirical. The mechanistic model aims to describe how things happen.

The models considered in previous sections can be placed under the broad heading as follows

Descriptive/communicative/empirical

- Polynomial regression models
- Spline models
- Growth increment

Explanatory/conceptual/mechanistic

- Differential equation models
- Some of the difference equation models
- Systems dynamics models
- Non-linear regression models derived from differential equations.

In relation to the two aspects is the idea of 'biologically meaningful parameters'. That is the parameters involved in a model should have some biological interpretation. The problem with these parameters is that the statistical estimates of such parameters often do not have 'nice' statistical properties, they are often complex non-linear functions of the natural mathematical parameterisation of the model (see Causton and Venus (1981)). Hunt (1982) argues caution on rejecting a model simply because its parameters cannot be given any general biological significance. Information is often not supplied by the parameters but through their derivatives. As Hunt states 'parameters are messengers of reality, not reality itself'. Gilchrist also considers the problem of reality, 'a model is only a limited, and possibly distorted, picture of reality'. 'A model can be seen as truth insofar as it makes "unhidden" aspects of the situation being modelled that were previously hidden'. He goes on to say that the important aspect is the adequacy of the model to reveal some aspects of reality.

A further comment on the empirical-mechanistic distinction. Hunt is surely right when he says that this distinction hinges on the level of organisation. The population biologists mechanism may well be empirical to the organismal biologist (see Hunt (1982)).

When one looks at a flexible function such as the Richards function the empirical-mechanistic divide seem even fuzzier. It is a mechanistic model because it comes from a certain differential equation model with possible mechanistic interpretations? Or is it just a useful empirical model? Obviously the fit of the model alone does not necessarily imply the underlying mechanism. There has to be an interpretation of the mechanism, but how satisfactory does it need to be? There are general ideas about growth behind the models considered but are these ideas good enough for any situation?

There is a second problem with flexible models, that of falsification. The concept of a hypothesis or model being falsifiable is central to the standard scientific approach. If a model is very flexible then it becomes difficult to falsify it, so it is always right. For some model such as splines this is not a problem because no interpretation is placed on them, they are purely smoothing functions. For some of the generalised logistic models there is an implicit mechanism behind them. How useful are these 'always right' models. This brings us to the next subject to examine, criteria for models.

7.3 Criteria for Models

What makes a good model? Randers (1980) listed the following desirable characteristics of a model

(a) Insight generating capacity

(b) Descriptive realism

(c) Mode reproduction ability

(e) Transparency

(f) Relevance

(g) Ease of enrichment

(h) Fertility (of new ideas, experiments etc)

(i) Formal correspondence with data

(j) Point predictive ability.

These criteria were primarily set up from a systems dynamics viewpoint. The above can be used to set up some general criteria which may be more in sympathy with statistical modelling.

(A) Data Correspondence

Does the model make use of all the available data? Is its use of data consistent, making use of pre selected criteria to judge closeness of fit. Have regions of inadequate data fit been shown to have no substantial effect on the overall model?

(B) Justifiability

Is the complexity of the model required? (Model parsimony). Can the model be falsified given the quantity of information available? If not, is the model sensitive to these conjectures?

(C) Applicability

Can model predict behaviour if required? Is it relevant to the end user?

D) Insight

Does the model increase understanding of the modelled system? Does it indicate areas of inadequate understanding?

These are general ideas and many problems are raised by them.

In (A) traditional model fit has been taken to be least squares or maximum likelihood. As in a previous section of the report the notion of an absolute minimum (maximum) of the selected criterion has to be called into question with complex models. In some cases where near plateaux or ridges exist in the criterion surface large changes in parameter values may have a very small effect on the criteria value. In these situations other aspects need to be taken into consideration. The fit in terms of say sum of squares needs to be a constraint (eg at least 99.9% of optimal) rather than the objective. Other aspects of fit (eg χ^2 goodness of fit tests, residual plots) need to be taken into consideration.

The problem is trying to quantify the different aspects so that an overall best model can be selected. I do not think such a procedure is practical, and not even desirable. Ultimately the selection will have to be subjective and this subjectivity needs to be clearly stated. However one important aspect has to be the consistency in the use of data. In systems dynamics it is generally considered that because of the negative feedback loops incorporated in the models they are not sensitive to parameter estimation. (See Randers (1980) passim). This leads to the situation in which diverse sets of data are used in diverse ways to estimate parameters, there being little consistency between them. As has been shown in chapters 5 and 6 the Kalman filter permits the estimation of parameters in both linear and non-linear dynamics systems models. It seems reasonable that such a consistent method should be used.

The second criteria (B) is to stop parts of models being only justified in the mind of the modeller. Care should be taken over the arbitrary use of significance levels in the rejection of terms and that considered in (A) always has to be taken into account. In statistics there are two different aspects of the model that are tested. One is the need for extra terms in the model (eg quadratic term in a polynomial regression). In these situations their non requirement is assumed until the data proves otherwise. The second aspects are normality, independence etc. These are assumed as part of the model until proven otherwise (eg by a Durbin Watson test). A better attempt is needed to put both these aspects on the same footing. This can be achieved in part by a more flexible attitude to significance tests and a more acute awareness of the basic assumptions of the model. In some circumstances it is important to realise if the assumptions have been justified or not, the actual situations will be related to the sensitivity of the model to these assumptions. Thus the model sensitivity has to be taken into account at this stage.

The use of the model must provide a criterion for its assessment (C) as is stated by Gilchrist the application will influence all stages. Obviously a model, if successful, will meet the objectives of the originator of the problem for which the model was constructed. However, this has to be considered alongside (D). There is an inherent conservativeness in (C) alone, in that it often seeks to discover what it already feels to be true. The information, via a different model perhaps, may have something new to say. One illustration of this is the statistical view of outliers. They are generally to be discarded so that the data then confirms the pre selected model form. In some cases the outlier can give new information about the situation or the model being used. Insight can be sacrificed for a 'good fitting simple model'.

In research, in particular, it is the insight generated by the model that is the prime criterion for judging the model. This insight must be valid so the model has to be commensurate with the information available (primarily the data). Unfortunately there is no quantitative way of fully describing insight, therefore more qualitative comparisons are required.

In some situations the main interest is in comparing two groups, different treatments perhaps. The statistical methods using polynomial curves described in the introduction are an example of a type of model designed for this purpose. Such models are not adequate in terms of the insight they generate in regards to the response over time of the groups. However, if they fit reasonably then they do compare the components of growth, linear, quadratic etc. The question these methods raise is how important are such summaries of a complex growth form? Are aspects that have been discarded by using such a model also important? More information may be obtained from such experiments if a more detailed modelling procedure were adopted and a careful comparison of meaningful parameters made given obvious caveats as regards the meaning of

parameters.

7.4 Methodologies for Modelling

As was discussed in the introduction Gilchrist (1984) presents a methodology for statistical modelling. This is based on that used by Box and Jenkins (1976) for ARMA modelling. Recall that the steps are

1 Identification

2 Estimation

3 Validation Iteration

4 Application

This approach works well in areas such as linear regression modelling and time series modelling in which you are dealing with a well defined family of models and selecting the best from that family. If a broader view is taken and the subjective nature of the modelling process is to be fully considered then a modified methodology is required.

Several problem solving methodologies have been developed for soft systems ie systems involving qualitative variables and subjective judgement. One such methodology is due to Checkland (1972). The essential stages are

(i) Analysis

(ii) Root definition of relevant systems

(iii) Conceptualisation

(iv) Comparison of definitions of possible changes

(v) Selections

(vi) Design and implementation

(vii) Appraisal

These stages can be seen as

(i) Obtaining information about current system

(ii) Reducing the system to its basic purposes

(iii) Constructing models

(iv) Comparing results of models with the situation to suggest changes which can then be selected (v) and applied (vi) and (vii) appraised

Using the ideas of this methodology a new methodology for statistical modelling can be developed.

(1) Conceptual Analysis

The object of this phase is to produce a conceptual or 'ideas' model of the situation to be examined. It will contain all possible relationships and their forms and the data available. The use of diagrams such as systems maps and influence diagram are an important tool at this stage.

(2) Model Type Generation

Using the insight gained from (1) a number of possible types of model to be used are considered (eg stochastic differential equation models, linear regression models, non linear regression models).

(3) Model Building

At this stage the models for each type are constructed. Within this stage Gilchrist's methodology (1 - 3) can be used for each model type. A single model need not emerge from each type as there may be several competing models with little objective distinction between them. The model building will be controlled by the criteria (A) and (B) of section 7.3.

(4) Comparison of Models

The models from (3) will be compared with respect to both direct application and generated insight. This comparison will primarily be a subjective comparison.

(5) Generation of Further Model Types

As a result of (4) improved models may be suggested.

6) Application and Appraisal

The results of the above model will be used and critically evaluated.

The above methodology can be placed within the modelling-data collection circle and one of the applications may be data generation.

It is hoped that such a methodology will lead to a more flexible approach to statistical modelling.

7.5 Statistical Modelling of Biological Growth - Some Guidelines

The aim of this section is to combine the general approach and methodology developed in this Chapter with the models and techniques described and developed in Chapters 1 to 6. The steps suggested in Section 7.4 will be used as a framework for guidance as to the most appropriate model to use in different circumstances.

(1) Conceptual Analysis

(i) The first aspect to be considered is the reasons for the investigation and what goals have been set. The goal might be as vague as 'to increase understanding of the growth of x' or it might be more specific such as 'to compare the effects of treatments A and B on the growth of x'. The type of goal will effect the choice of model type as considered in (2).

(ii) Data structure and availability has also to be considered. The number of observations will effect the type of models that can be considered. If only a small number of observations are available then the fitting of complex Kalman filter type models would not be advisable. Model identification also becomes a problem with a small number of observations per growth series. While it is possible to fit, say, the Richards function to 4 data points, indeed the optimal design considered in Section 3.5 consists of repeated observations at 4 distinct points, this does not give any indication as to the fit of the model. Only if there were strong apriori reasons for using the model would it be reasonable so to do under

these circumstances. In which case these reasons form an important aspect of the modelling process.

(iii) The correlated error structures considered in Chapters 1 and 4 will need to be considered in all growth studies. While they will be of greatest importance when an individual is followed through time, even in the situation of sampling from a growing population there is still the possibility of correlated errors due to correlations in the variation of the common environment. The form of the error structure will be shown to be the key to the choice between certain model types and careful consideration is needed at an early stage.

(iv) Prior knowledge forms an important aspect of the modelling process. Previously used model may effect the choice of model and the experimental design for data collection. However it is important not to be too restricted by past models. For instance if the linear models considered in Chapter 1 had been used in the past the design of an experiment suitable for such a model would probably exclude the fitting of a Kalman filter type model. An openness to the broad range of models considered in this study is important at all stages.

In some situations if there is accurate knowledge of the value of some of the basic growth parameters this can be included in the fitting stage of the modelling process.

(v) Other external variables eg temperature, nitrogen availability, may be required to be included in the model. This will also be the case when a comparison is to be made.

The variables to be examined and how these variables enter the model needs to be considered in detail. The simplest structure is to use a linear predictor

$$\eta = \sum_i \beta_i \, x_{ij}$$

where x_i's are the external variables. The η function can then be related to a simple model parameter through a link function eg

$$\alpha = \exp(\eta) \quad \text{or} \quad \alpha = \alpha_0 + \eta.$$

The situation and the evidence given by the data will help to select the appropriate model. It may also be required to relate the x's to more than one parameter. Knowledge of the relationships from other studies is important here.

(vi) Finally the choice of growth variable to be modelled has to be decided upon, also whether the possible variables are to be considered separately or jointly.

Having assembled the above ideas and information the modeller can then proceed to the next step.

2. Model Type Generation

The model types considered in this study are:

(a) Linear models (Section 1.3)

(b) Static models, or non-linear regression models, with or without correlated errors (Chapters 3 and 4)

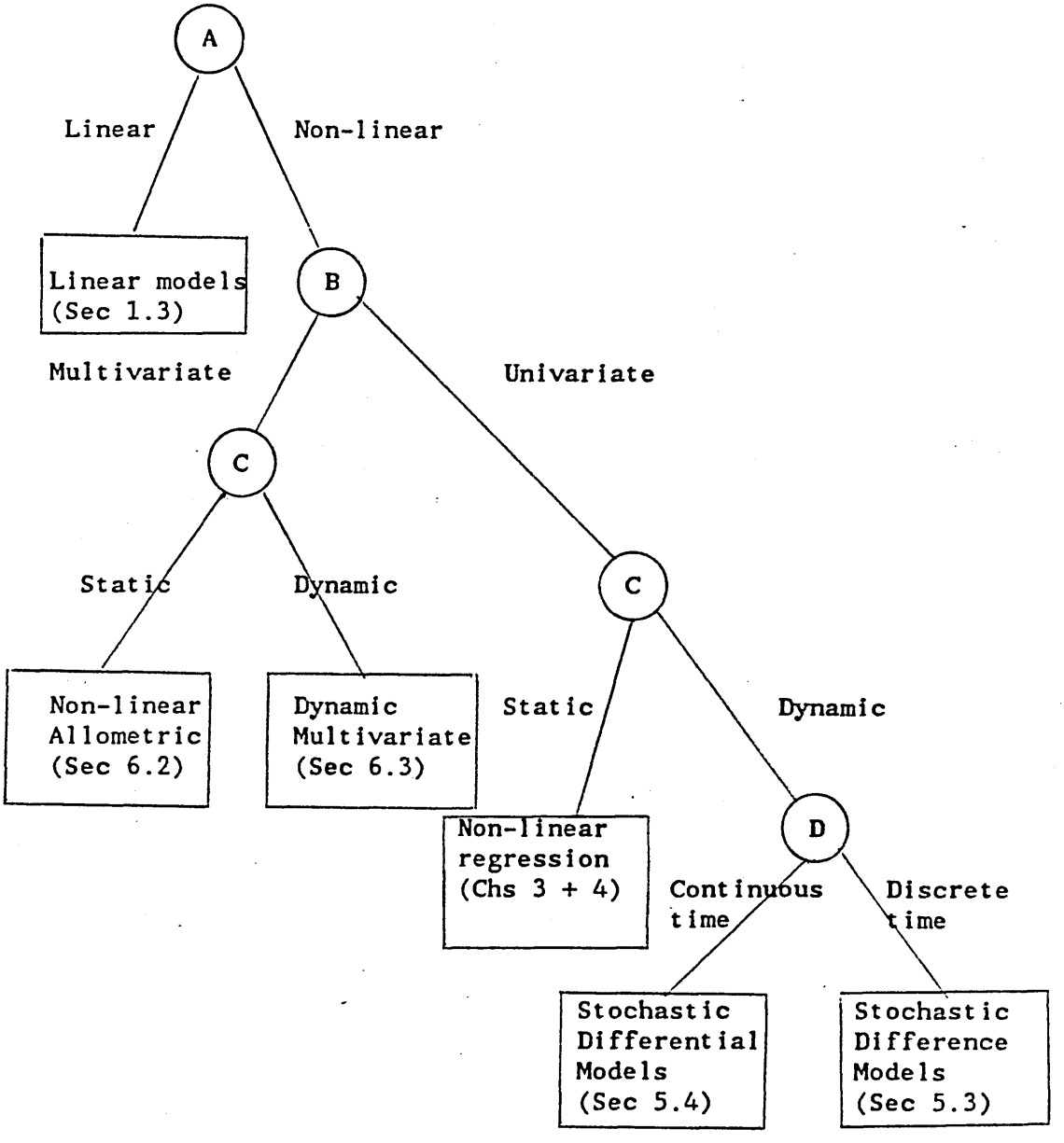
(c) Dynamic models in discrete time (Section 5.2)

(d) Dynamic models in continuous time (Section 5.3)

(e) Multivariate models (Chapter 6).

In order to select one of these model types the following four questions need to be considered:

Figure 7.5 (i)



(A) Linear or non-linear?

(B) Univariate or multivariate?

(C) Static or dynamic?

(D) Continuous time or discrete time?

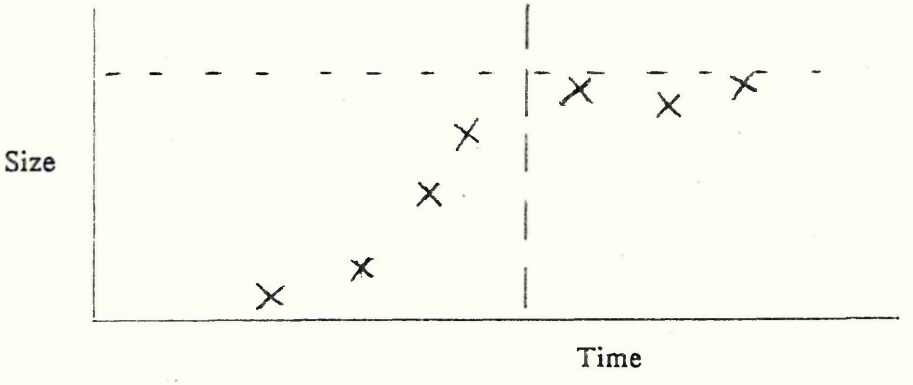
These lead to the decision tree shown in figure 7.5(i). Each of these decisions will now be considered.

A. Linear or non-linear?

The following points need to be considered.

(i) The amount of data available.

A small number of intervals in the recorded growth series, say less than 8, would exclude the use of non-linear models unless there were other reasons for selecting a particular model. The situation in which one might consider a non-linear model is if at least 3 of the points were clearly close to the maximum growth eg



Here a polynomial model would give a poor fit. However the data could be split into two as shown in the diagram. The first part could then be analysed using a polynomial model, the second part

by a constant model.

(ii) The purpose of modelling.

If the purpose is just to compare different groups then the polynomial model may be adequate. In these models linear and quadratic components of growth are compared. If such comparisons are what is required and are meaningful in the given situation then the relative simplicity of the linear models would favour their use. Higher order comparisons are also possible with the polynomial models but their interpretation is more problematic.

In the situation where a more detailed comparison involving more biologically meaningful parameters is required the linear polynomial may well be useful in an initial screening process giving a rough indication as to which curves are different. If the aim of the model is for it to be incorporated into a larger system model then there will probably be advantages in using the differential equation based non-linear models. For instance the differential equations on which the model is based could be used in a larger systems dynamic model.

(iii) Statistical properties versus reality.

The main advantage of the linear polynomial models is in the well known properties of the estimators. Also the models can incorporate a general error structure giving a flexibility in that respect that the non-linear models do not possess. However there is no straightforward understanding of these models in terms of growth dynamics. They also represent a dead end as far as model building is concerned since extensions and refinements to these models is

limited.

The above show what needs to be considered in selecting between linear and non-linear models.

B. Univariate or Multivariate

In most situations interest is focussed on a particular dimension of growth thus the univariate models would be used. These may also help to identify suitable multivariate models.

The linear models can easily be adapted to a multivariate situation and the fitting of non-linear models has been considered in Chapter 6. However in the latter case while it is possible to fit models to large dimensions this may prove difficult in practice due to the large number of possible parameters. Some of these problems would be overcome if a simplified structure was used in the model on apriori grounds. In two dimensions such modelling is reasonable as shown in Chapter 6.

C. Static or dynamic?

In order to consider this choice the state space model as used in the Kalman filter modelling needs to be considered. (Using continuous time only for convenience)

$$\frac{dx}{dt} = g(x,t) + \varepsilon_t, \text{var}(\varepsilon_t) = Q$$

$$y = h(x,t) + e_t, \text{var}(e_t) = R$$

In particular the following measurement equation can be considered

$$y = x + e_t$$

Now if Q is negligible and $x = f(t)$ is the solution to the deterministic equation

$$\frac{dx}{dt} = g(x,t)$$

then

$$y = f(t) + e_t$$

ie a non-linear regression (static) model. If R is negligible then $y = x$ so

$$\frac{dy}{dt} = g(y,t) + \varepsilon_t$$

ie the stochastic differential equation models considered in Chapter 5. So both can be seen as submodels of the Kalman filter type model. The choice between such models then reduces to the consideration of the two errors; the propogation error ε_t and the measurement error e_t.

A further relationship between the non-linear models considered in this study can be seen by considering the simple discrete time model presented in Section 5.2.1

$$x_t = \lambda x_{t-1} + \varepsilon_t \text{ , } \varepsilon_t \text{ independent}$$

This has solution

$$x_t = \lambda^t x_0 + \sum_{i=0}^{t-1} \lambda^i \varepsilon_{t-i}$$

combining with a measurement model

$$y_t = x_t + e_t$$

gives

$$y_t = \lambda^t x_0 + u_t$$

where

$$u_t = \sum_{i=0}^{t-1} \lambda^i \varepsilon_{t-i} + e_t$$

ie a non-linear regression model with an error that can be considered as an ARMA error. In a more general case the expected value solution to the stochastic state space equation will not be the same as the solution to the deterministic part of the equation. Approximate solutions can be obtained using Taylor's series approximations as considered in Chapter 5.

Thus for practical purposes many of the Kalman filter type models could be replaced by suitable static models with auto correlated errors. Given the above the choice between static and dynamic models on the following considerations.

(i) If the propogation error variance, Q, is negligible compared to the size of the variable being considered and the measurement error variance, R, then a suitable non-linear regression model may be used. This situation could partly be identified by the errors having the appearance of being independent and also by a consistency in the curves between individuals. Such observations chould be made by the initial fitting of non-linear regression models as a guide to form and to provide an estimate of error. Alternatively spline curves, high order polynomials or moving average models may fulfil this role in certain cases.

(ii) If by considering the way the data was collected it can be assumed that the measurement error variance is negligible then a stochastic differential/difference equation models can be used.

(iii) Otherwise the Kalman filter type model should be used or possibly a non-linear model with ARMA errors.

D. Continuous time or discrete time?

The choice between continuous and discrete time formulations will generally only occur for dynamic models since the static difference models can be solved, analytically or numerically, and the solutions treated as an ordinary static model. As was discussed before, the discrete time models are most natural when considering populations with non-overlapping generations, but can be used for any growth situation in which observations are taken at equally spaced time points. The stochastic difference equation model is more easily understood than the differential equation model, the stochastic differential being a rather complex mathematical concept. Hence there may be some who will prefer to use such models rather than the more common differential equation based models. Models of at least approximately the same behaviour from both classes can be found so it reduces to a matter of personal preference.

The above provides guidelines by which classes of model should be investigated. At this stage more than one class of model may be investigated and, if required, final selection being made at stage (4).

(3) Model Building

Having selected suitable classes of model for further investigation the most appropriate model from each selected class has to be chosen.

For linear models the order of the polynomial model can be selected by standard tests on additional terms in the model (eg test for cubic over quadratic).

For non-linear models the three stages of identification, fitting and validation need to be considered in more detail.

(i) Identification

The identification of a suitable model can be assisted by the graphical method given in Section 2.3. This will indicate a suitable form for the growth rate equation or relative growth rate equation.

A second important aspect will be to acquire a knowledge of what is the intrinsic variation in the data. The methods of Section 2.5 provide an aid. This will show when one has a reasonably fitting model.

(ii) Model fitting

Consideration of fitting has been covered in Chapters 3 to 6. Prior knowledge of approximate parameter values can be of considerable use and the method of Section 3.6.2 can then be used.

In Kalman filter type models it may be necessary to try to obtain a reasonable idea of the values of the variances Q and R in order to find satisfactory estimates of parameters of most interest. The fitting of the simpler non-linear regression models, perhaps with

auto-correlated errors, will often be useful here as will the methods of Section 2.5 on estimating the inherent variability.

(iii) Validation

For simple non-linear growth models the methods described in Section 2.2 provide both model identification and model validation techniques. Validation for the dynamic models can be carried out using the independent normal residuals considered in Chapter 5 and in particular for the Kalman filter on page 134.

One other important aspect of the validation of a non-linear model is to examine the effect of non-linearity. It was shown in Section 3.3 that the examination of the sensitivity coefficients in the case of the Richards function was not very informative and higher order measures were needed (Section 3.4.5). The measures of Bates and Watts discussed in Section 3.4 were seen to be reasonable when compared with the estimates of bias produced by Causton and Venus using simulation. The evaluation of the measures along with Box's estimates of bias should be routine in the fitting of non-linear regression models. If the measures indicate more than low non-linear and bias it is doubtful if the model should be used. For this reason flexible but complex models like the Richards which are vulnerable to high non-linearity should be treated with caution and may not prove as versatile as they may seem.

In theory it should be possible to evaluate such measures for the Kalman filter type models fitted by least squares. However, the required derivatives would have to be calculated numerically and the method may prove too complex in practice.

The above considerations may lead to more than one model of a particular type. In view of the problems that may be encountered with over complex models it may be better at this stage to have two different but relatively simple models of a particular type rather than one over complex model.

(4) Comparison of Models

If more than one model type has been considered or if as a result of the stages described in (3) more than one model of a particular type is under consideration then at this stage models are to be compared. The results from the different models have to be checked for consistency and any inconsistent results carefully checked and examined. The most appropriate model for the particular aim of each part of the investigation may then be selected.

(5) Generation of further models

If the models are felt to be inadequate further models may be considered and fitted. Models of a different type may need to be considered if those of the types considered either do not produce reasonable estimates or do not fit adequately. This may be due to being either over complex for the situation or of the wrong structure. If bias in the non-linear regression models has been found to be a problem then the methods of Section 3.6.3 could be used.

(6) Application and appraisal

The final critical evaluation of the modelling exercise is whether or not it has been helpful to the biologist. The biologist needs to be both open to the insights the models may generate and also critical of assumptions and intensions

of the modeller who is helping him. It is important that the modelling is an integrated part of the whole problem situation and not seen as a pleasant pastime for statisticians.

8. Plant Growth Modelling

8.1 Introduction

There are many aspects of the growth of a plant which one would be interested in modelling. These include

(i) Plant height

(ii) Plant weight (usually dry matter)

(iii) Leaf size (weight or area)

(iv) Chemical content of plant eg nitrogen, carbon, protein etc.

The plant is a complex system and there are many ways in which it can be broken down into various elements, some of the possible sub-systems are

(i) Roots/shoots

(ii) Structural dry weight/storage dry weight

(iii) Maintenance respiration/growth respiration

(iv) Carbon pool/Nitrogen pool

(v) Vegetative/reproductive

The plant can therefore be considered at various levels. One approach is to consider only one level and to model that independently of the other aspects of the plant. Another approach is to try to combine elements from different levels to produce a model for the system (or a particular sub-system). This approach will now be considered.

8.1.2 Modelling the whole plant

There are several basic components of a model for a whole plant

(i) Uptake of nutrients

(ii) Production of assimilates (photosynthesis)

(iii) Partition of assimilate

(iv) Respiration

(v) Redistribution of plant resources in relation to reproductive growth or leaf death

The uptake of nutrients is related to water availability, the soil structure and the form of the elements, and the root structure of the plant.

Models for photosynthesis vary in complexity from the simple rectangular hyperbola to complex chemical based models such as those of Charles-Edwards and Ludwig (1974 and 1975) and Hahn (1984). Total photosynthesis will be related to leaf area. One important aspect of the partition of the photosynthetic products is between shoot growth which will increase the leaf area, and hence the amount of photosynthesis, and the storage of materials in roots. This is of particular importance in crops like carrots and beet.

Respiration will contain a component dependent on size, concerned with the maintenance of metabolic activity. A second component is related to the growth of the plant.

The various parts of the model are connected to produce a model that simulates the behaviour of the whole plant though the individual parts will be based on data at the different levels indicated above.

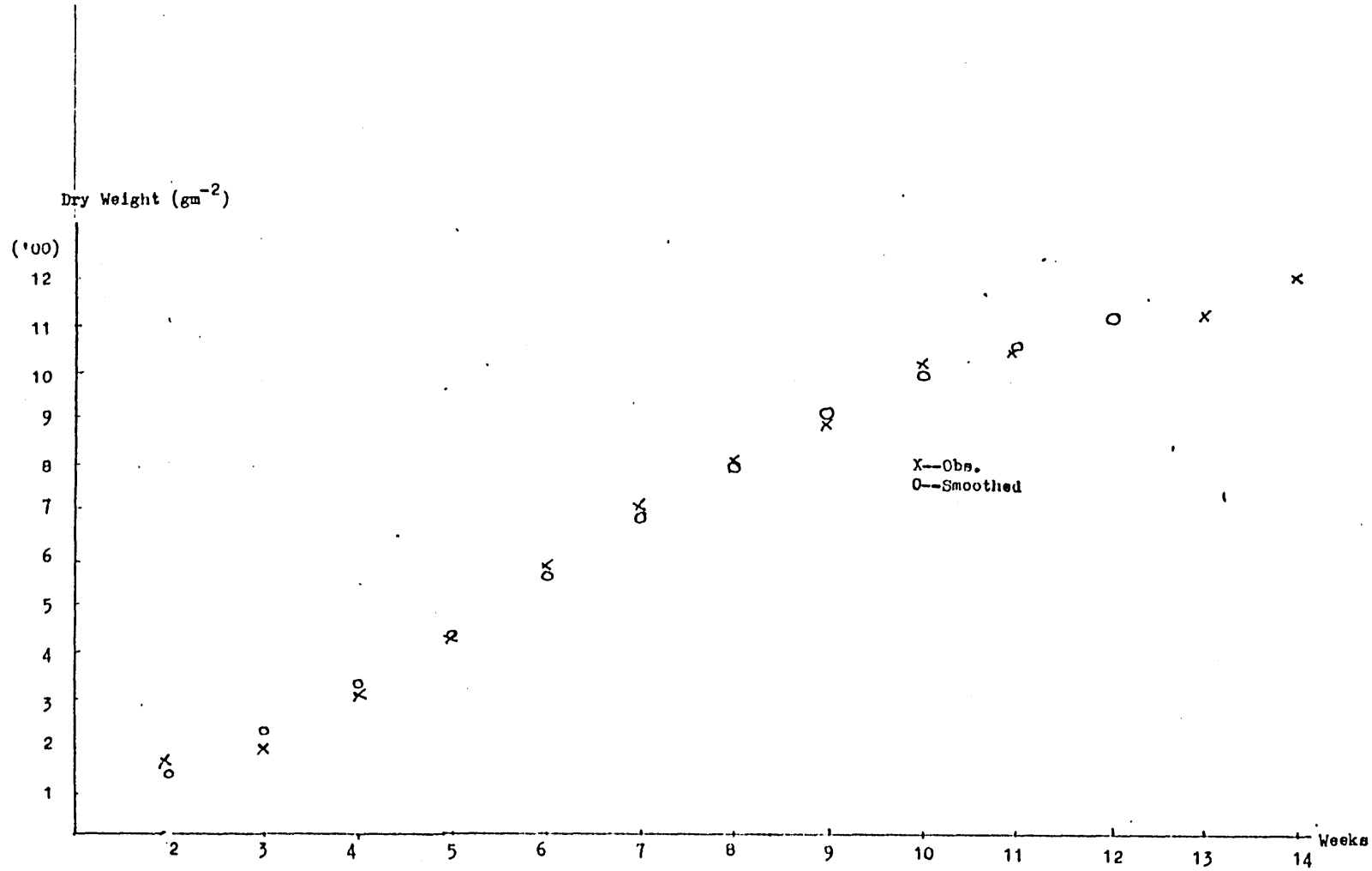

Dry Weight (gm^{-2})
('00)
12
11
10
9
8
7
6
5
4
3
2
1
X--Obs.
O--Smoothed
2
3
4
5
6
7
8
9
10
11
12
13
14
Weeks

FIGURE 8.2(1)

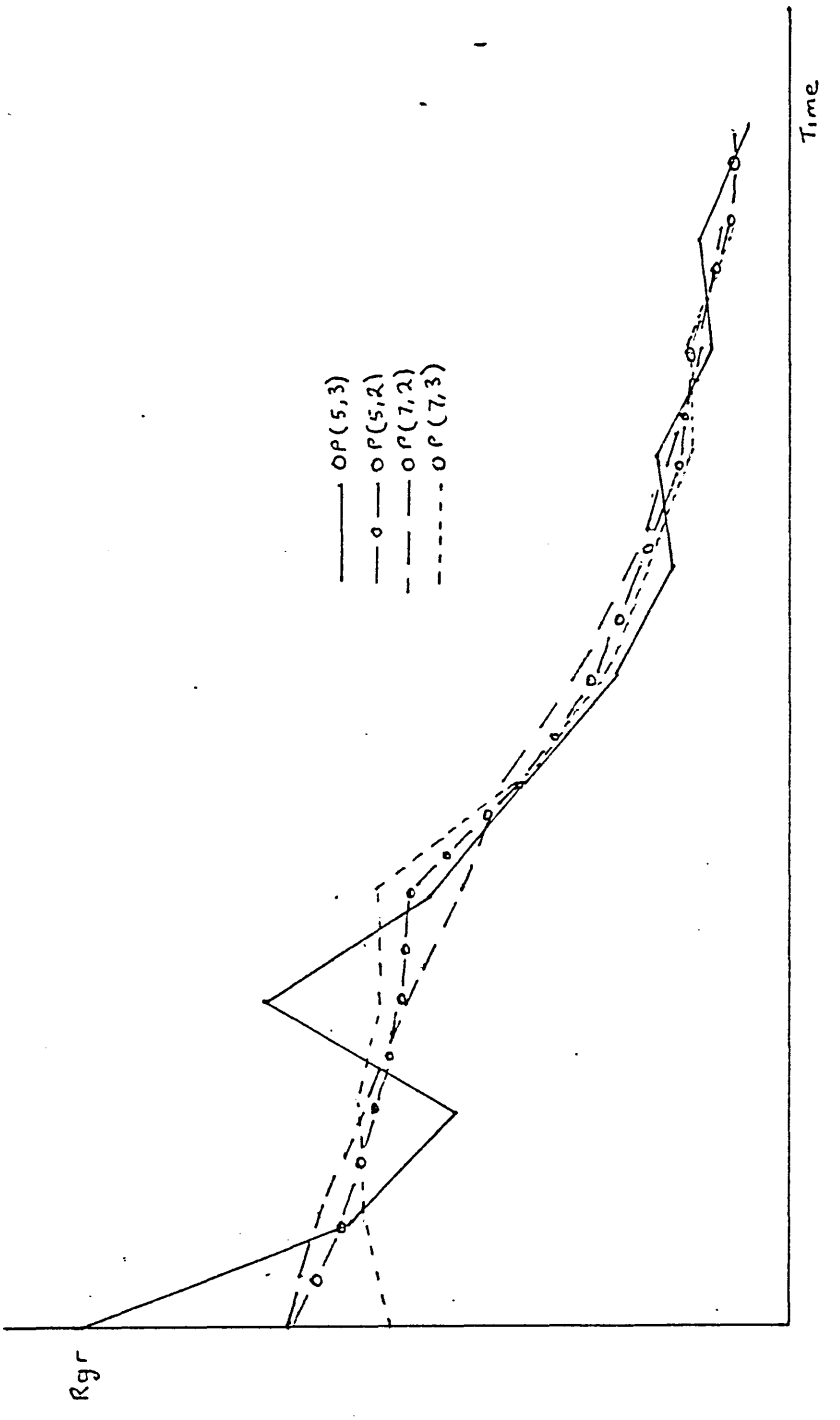


FIGURE 8.2(2)

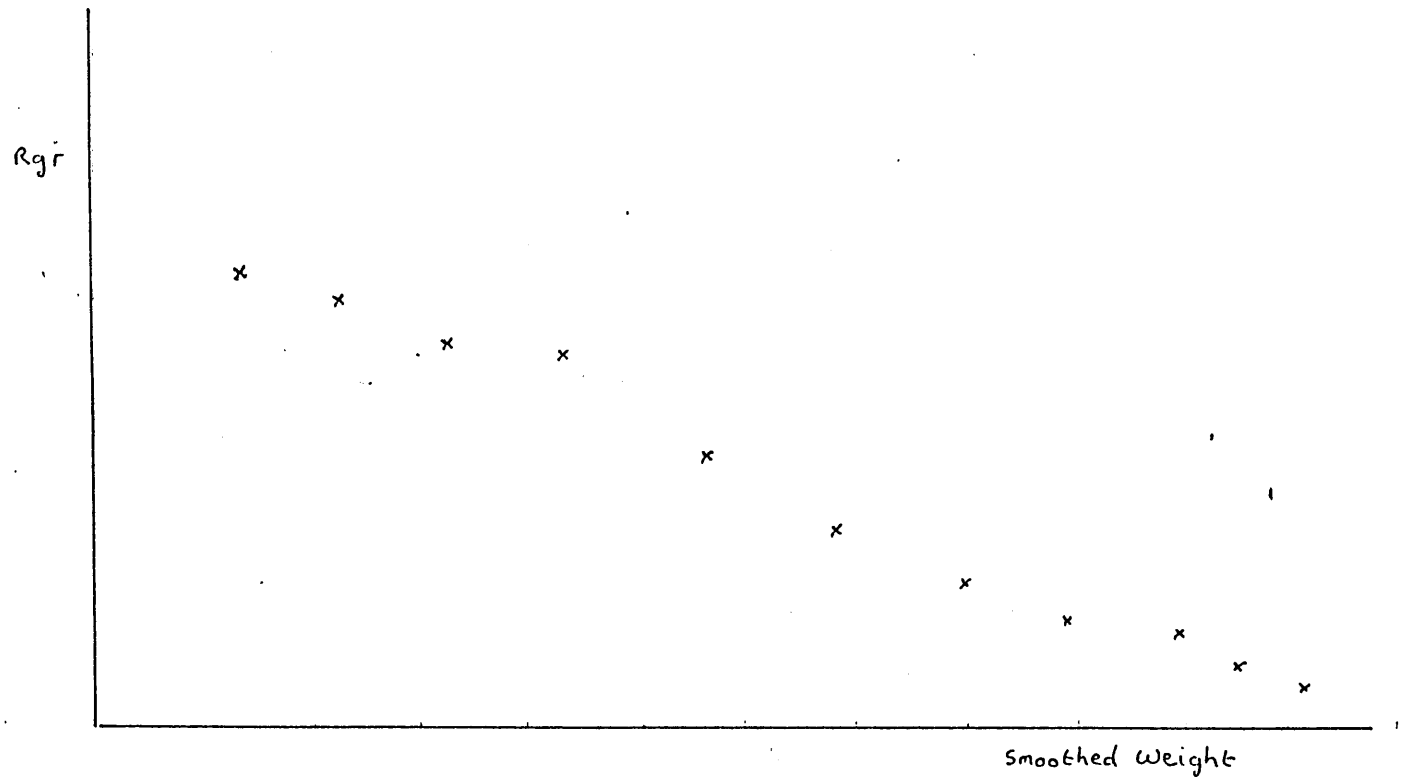


FIGURE 8.2(3)

8.1.3 Examples

From the vast range of possible examples the following have been chosen to illustrate both the techniques considered in earlier chapters and the different approaches available to the modeller. First, two examples of plant growth will be considered, then a discussion of modelling approaches will be undertaken, illustrated with reference to leaf growth.

8.2 Plant Growth - Barley

The data to be considered is given in table 8.2(1). It is the total dry weight per metre squared of ground for Proctor Barley measured at weekly intervals.

8.2.1 Model Identification

A plot of the data is given in figure 8.2(1) and smoothed estimates of the relative growth rate are shown plotted against time in figure 8.2(2). These were obtained by using the method of section 2.3. Local orthogonal polynomials of degrees 2 and 3 were fitted to either 7 or 5 running points [they will be referred to as OP(n,p) where n = number of points, p = degree of the polynomial].

Except for OP(5, 3) they give a fairly smooth downward curve. OP(5, 3) is less smooth as would be expected and shows a small local peak.

Figure 8.3(3) shows OP(5, 2) plotted against a smoothed estimate for weight, W. This was obtained using a 5 point moving average with end adjustment. The smoothed curve is also shown in figure 8.2(1). The relative growth rate in figure 8.2(3) shows an approximate linear decreasing trend. This indicates the

use of a logistic model

$$y = K/(1 + \exp(-(\beta + rt))$$

Table 8.2(1)

Date	Total Dry Weight (gm^{-2})
24 April	81
3 May	151
10 May	172
17 May	294
24 May	424
31 May	576
7 June	705
14 June	808
21 June	894
28 June	1017
5 July	1048
12 July	1189
19 July	1133
26 July	1203

The intrinsic variability was estimated using polynomial models as suggested in section 2.5. Fitting a polynomial model to all the data the following result was obtained.

Degree of Polynomial	Residual Sum of Squares
0	2169000
1	57360
2	36670
3	11080
4	8458
5	8447

suggesting $\sigma^2 \simeq 850$

If the final 3 observations are omitted

Degree of Polynomial	Residual Sum of Squares
0	1278000
1	17600
2	17460
3	3182
4	2981
5	2811

suggesting $\sigma^2 \simeq 450$

The variance of the last 3 points is 1372, thus indicating possible variance increase with size. Therefore, using a logarithmic scale for all the data

Degree of Polynomial	Residual Sum of Squares
0	10.10
1	1.322
2	.07808
3	.06163
4	.05539
5	.05414
6	.05414

suggesting $\sigma^2 \simeq 0.006$

8.2.2 Fitting Static Models

Using SAS the logistic model was fitted to the data using the natural and logarithmic forms. This gave results

Parameter	Natural	Logarithmic
k(log(k))	1223.24 (7.1093)	1199.91 (7.0900)
β	-2.9401	-3.0616
r	0.4540	0.4823
s^2	932.8936	0.0050

Plots of the residuals from both models did not give any clear preference. However, in view of the results of the previous section and comparing estimates of σ^2 the logarithmic version is to be preferred.

The logarithmic version was fitted with auto regressive errors as considered in chapter 3. With an AR(1) error the estimates are

Parameter	Estimate	
φ	-0.5038	
log k	7.0914	(k = 1201.59)
β	-3.047	
r	0.4784	

The change in deviance for the auto regressive term was 4.05 which was significant at the 5% level.

When trying to fit an AR(2) model the estimates did not converge. Given the low significance of the AR(1) this indicates that the second term is not needed.

8.2.3 Dynamic Models

The difference equation model

$$x_t = \frac{a\, x_{t-1}}{1 + b\, x_{t-1}} + \frac{x_{t-1}}{1 + b\, x_{t-1}}\, \varepsilon_t \quad , \quad \text{var}(\varepsilon_t) = q$$

was considered in section 5.2.6 and is related to the logistic with

$$a = e^r \quad \text{and} \quad b = (e^r - 1)/k$$

This was fitted using the Kalman filter as described in section 5.2.6 with measurement equation

$$y_t = x_t + e_t \quad , \quad \text{var}(e_t) = \sigma^2$$

writing $q = \lambda\, \sigma^2$ the following estimates were obtained

Parameter	Estimate	
a	1.6050	(r = 0.4731)
b	0.000498	(k = 1214.86)
λ	0.5280	
x_0	58.3396	
σ^2	560	

The stochastic logistic model

$$dx_t = \rho x_t(1 - x_t/k)dt + x_t(1 - x_t/k)d\beta_t$$

was considered in section 5.3.3. The pdf can be written as

$$f(x_t, t|x_0) = \frac{1}{(\sqrt{2\pi\sigma^2})x_t(1 - x_t/k)} \times$$

$$\exp\left\{-\frac{1}{2\sigma^2 t}\left[\log\left[\frac{x_t(k - x_0)}{x_0(k - x_t)} - r^t\right]^2\right\}\right.$$

and

$$\text{log likelihood } L = \sum_t \log\,(f(x_t, t|x_{t-1})$$

If the observations are equally spaced with time interval τ then letting

$$g_t = \log(x_t/(1 - x_t/k))$$

$$L = \text{const} - \frac{n}{2}\log\sigma^2 + n\log k + \Sigma g_t$$

$$- \frac{1}{2\sigma^2\tau}\Sigma(g_t - g_{t-1} - r\,\tau)$$

Let

$$s = \Sigma(\hat{g}_t - \hat{g}_{t-1} - \hat{r}\,\tau)^2$$

then

$$\hat{\sigma}^2 = s/n\,\tau$$

$$\hat{r} = (\hat{g}_n - \hat{g}_0)/n\,\tau$$

Substituting these results in L, $\hat{x}_0$ and $\hat{k}$ can be obtained by maximising

$$L^* = -\frac{n}{2} \log s + n \log k + \sum_t g_t$$

This was carried out using the Nelder-Mead algorithm (Appendix A).

The following estimates were obtained

Parameter	Estimate
r	.3513
k	1408.70
σ^2	.05741
x_0	57.7941

(Note estimate of σ^2 is similar to that obtained on the log scale)

As the value of $\hat{r}$ is highly dependent on the final observtion, the last two observations were deleted and the results were

Parameter	Estimate
r	.4156
k	1387.38
σ^2	.03255
x_0	54.5476

Finally the application of the Kalman filter to the logistic model was considered. This was examined in section 5.3.7. Problems were encountered in minimising 5.3(50) ie

$$\Sigma\, z_t^2/(p_{t-} + R)$$

over the 5 parameters of the model, r, k, x_0 , Q and R. In the light of this, particularly with the small size of the data set, suitable values of Q and R were selected and 5.3(50) minimised over r, k and x_0 only. Some of the results obtained using a single propogation step with $P_0 = 0$ were

Parameter	Estimates		
	R = 450 Q = 200	R = 450 Q = 300	R = 600 Q = 200
k	1282	1308	1281
r	.4767	.4670	.4781
x_0	95	88	70

Similar results were obtained from more complex propogations. The value of x_0 is greatly affected by the choice of R and Q, but this parameter is of little interest. The r parameter appears to be reasonably stable.

A larger data set is needed for such a complex model.

8.2.4 Comparison of Results

Table 8.2(2) summarises the estimates of the parameters r and k using the different methods considered in this section.

There is little empirical evidence to choose between the different models. A test showed that (3) is to be preferred to (2) in spite of independent sampling. The results from the Kalman filter depend on the values of Q and R, if Q was negligible the results should be similar to the least squares estimate (1). However, more data is required for a successful use of this method. The stochastic differential equation model fitted to all the data gave a result out of

line with the rest. When the final two observations were removed a more reasonable result was obtained. It would be reasonable to assume that the stochastic nature of the situation would change as the crop reaches senescence. The difference equation model was most compatible with the static models.

The distinction will have to be made between the models at a conceptual level rather than at an empirical level.

Table 8.2(2)

Number	Method	$\hat{r}$	$\hat{k}$
1	Least Squares	.4540	1223
2	Least Squares (log model)	.4823	1200
3	AR(1) error model	.4784	1202
4	Difference Equation	.4731	1215
5	Stochastic diff eqn (14 obs)	.3513	1409
6	Stochastic diff eqn (12 obs)	.4156	1387
7	Kalman Filter 1	.4767	1282
8	Kalman Filter 2	.4670	1308
9	Kalman Filter 3	.4781	1281

8.3 Plant Growth - Sycamore

8.3.1 Introduction

The data to be examined was presented by Causton (1969). It illustrates the problems in fitting the Richards function and motivated much of the work of chapter 3.

The data was for the growth of first year seedlings of sycamore (Acer pseudoplatancis). Harvests were taken at 2-weekly intervals. The data is given in table 8.3(1).

Table 8.3(1)

Weight of seedling of sycamore

Week	Weight (gms)
2	0.3898
4	0.8110
6	1.293
8	1.840
10	2.590
12	3.770
14	5.610
16	9.698
18	12.56
20	21.85
22	23.72
24	20.03

8.3.2 Causton's Original Work

Causton (1969) fitted the Richards function in the form

$$\log \omega = a - \frac{1}{n} \log (1 + Be^{kt})$$

which gave values

Parameter	Estimate
a	3.0772
β	7.7020×10^{45}
k	-5.2460
n	25.0180

The more usual parameterisation with

$\beta = \log (B)$

would give

$\hat{\beta} = 105.6578$

Also

$\hat{m} = \frac{1}{n} = 0.040$

8.3.3 Davies and Ku's Work

Davies and Ku (1977) examined the problem and showed that for a range of different starting values different minimums were obtained. Their results are reproduced in table 8.2(2). They concluded that

> 'The only stable parameter is a. There is an extensive, almost stationary ridge given by $^{b}/_{k} = -20.1$, $mk = 0.309$, $b > 25$ and any point on this ridge, depending on the starting point, could be reached on applying the Newton-Raphson method'.

Work carried out by Drajek (1985) partly confirms these results but showed that they were machine dependent, in that different results were obtained from different computers.

Table 8.2(2)

Davies and Ku's Results

Set	Starting Values		"Converged" Values		m	a	Res Sum of Squares
	β	k	β	k			
1	10.0	- 0.5	10.14	- 0.504	-0.4341	3.200	0.167348
2	20.0	- 1.0	20.17	- 1.003	-0.2119	3.108	0.135960
3	50.0	- 2.5	50.18	- 2.490	-0.0844	3.084	0.119630
4	100.0	- 5.0	100.2	- 4.975	-0.0422	3.078	0.116020
5	200.0	- 10.0	200.3	- 9.951	-0.0211	3.074	0.114818
6	500.0	- 25.0	500.6	- 24.89	-0.0084	3.070	0.114399
7	1000.0	- 50.0	1001	- 49.81	-0.0042	3.069	0.114356
8	1500.0	- 75.0	1502	- 74.73	-0.0028	3.069	0.114354
9	2000.0	-100.0	2003	- 99.64	-0.0021	3.069	0.114354
10	2500.0	-125.0	2504	-124.6	-0.0017	3.069	0.114354
11	3000.0	-150.0	3004	-149.5	-0.0014	3.069	0.114354
12	3500.0	-175.0	3505	-174.4	-0.00120	3.069	0.114354
13	4000.0	-200.0	4006	-199.3	-0.00105	3.069	0.114354
14	5000.0	-250.0	5008	-249.2	-0.00084	3.069	0.114354
15	6000.0	-300.0	6010	-299.0	-0.00070	3.069	0.114354
16	12000.0	-600.0	12020	-598.0	-0.00035	3.069	0.114354

8.3.3 Measures of Non-linearity

For a selected number of results from Davies and Ku the measures of non-linearity and bias estimates were calculated for the usual model with the parameter n and also the model with the m parameter. These results are given in tables 8.3(3) and 8.3(4) respectively. In all cases the measures of bias are highly significant and there is considerable bias. Beyond solution (4) the results show massive bias as one would expect. The results for both models show a similar bias.

Table 8.3(3)

Solution	Γ^N	γ^T	N_θ	N_φ	Bias			
					a	n	b	k
1	.8923	17.8779	9.9184	.0255	.0757	.8101	3.3948	- .1709
2	1.4713	13.3223	5.9112	.0751	.0439	5.237	22.1028	- 1.099
3	1.4732	17.3725	11.9079	.681	.0252	73.5383	310.4573	- 15.3859
4	0.1078	32.8041	48.6092	.0004	.0021	711.7800	3005.9041	-149.1237
6	.7522	691.7368	21366	.0177	.0000	1×10^6	6×10^6	-312629
9	7×10^5	2×10^6	2×10^{11}	2×10^{10}	.1743	2×10^{13}	9×10^{13}	-4×10^{12}
14	1×10^5	2×10^5	3×10^8	4×10^8	.0000	9×10^{11}	4×10^{12}	-2×10^{11}

Table 8.3(4)

Solution	Γ^N	Γ^T	N_θ	N_φ	a	m	b	k
1	.8921	118.9	441.6	.0254	.0757	.0752	3.393	-.1708
4	.1079	6171	1×10^6	.0004	.0021	.1771	3007.6	-.149.2
6	.3753	1×10^6	7×10^{10}	.0044	.0000	53.1847	6×10^6	-3×10^5

Can the bias be minimised? To attempt to minimise the bias fixed values of m were taken and the remaining parameters were estimated by least squares. Using all four parameters the bias and non-linearity measures were computed. The minimum non-linearity was when m = .568 this gave

$\Gamma^T = 111.20$ $N_\theta = 386.6$

both very large, however the estimated bias was

Parameter	Bias
a	.1049
m	-.1319
β	1.9453
k	-.0997

The large values of Γ^T and N_θ must be compared with the smallest values from table 8.3(4). The m parameterisation gives a much higher non-linearity measure than the n parameterisation.

The above results suggest that the best of Davies and Ku's solutions is (1) which will have minimum bias. It is worth allowing the residual sum of squares to increase from .1143, the minimum, to .1673, an increase of .053 for the superior properties of the estimates.

8.3.4 Using the Method of Section 3.6.2

Instead of minimising the residual sum of squares the function S^λ (3.6(8)) was minimised. As the key to the fitting is the m (or n) parameter only this was considered. Hence

$$S^\lambda = 12\lambda(m - m_0)^2 + \Sigma(\log \omega - f(\beta))^2$$

The value of m_0 was selected as the value for minimum bias, $m_0 = 0.568$. For a range of values of λ the following results were obtained

λ	.1	.5	1
a	3.2503	3.2729	3.2756
β	8.5904	8.0492	7.9910
k	-0.4263	-0.3991	-0.3962
m	0.5205	0.5591	0.5636

(It is interesting that for Davies and Ku's set 1 results as starting values and $\lambda = 0$ the SAS program did not converge).

The results illustrate that even for a low value of λ the results are confined to a more acceptable region.

8.4 Modelling Leaf Growth

As was mentioned in chapter 3 Causton and Venus (1981) make extensive use of the Richards function as a model for the growth of a leaf both in terms of area and dry weight using the form

$$\log y = \log \alpha - \frac{1}{n} \log \left[1 + e^{-(\beta + kt)}\right]$$

Interpretations of these parameters are

(i) α is the maximum size of the leaf

(ii) n controls the shape of the growth curve (Causton and Venus (1981) p 93)

(iii) β has no biological significance it only shifts the time axis

(iv) k is a growth rate constant but its meaning depends on the value of n

Several biologically relevant combinations of parameters have been suggested (Richards (1959), Causton and Venus (1981)). These include

(v) a weighted mean relative growth rate over the whole growth period

$$\bar{R} = k/(n + 1)$$

(vi) a weighted mean absolute growth rate

$$\bar{G} = \alpha k/2(n + 2)$$

(vii) the time required for the major portion of growth, sometimes called the duration of growth

$$d = 2(n + 2)/k$$

In addition to these parameters the relative growth rate as a function can be derived from the fitted Richards curve.

Instead of using a Richards function other statistical models have been suggested, these being fitted to the logarithm of size

Polynomials

Segmented Polynomials

Splines (cubic)

These will often provide a simpler estimation procedure for the estimation of growth rates (Nicholls and Calder (1973), Hunt and Parsons (1974), Elias and Causton (1976), Hunt and Parsons (1977), Parsons and Hunt (1981), Hunt (1982 a, b)).

Venus and Causton (1979) and Causton and Venus (1981) defend the use of the Richards function on the grounds

(a) more biological meaningful parameters

(b) no problem with model choice

In agreement with (a) the Richards function (or other suitable asymptotic function) will provide an estimate of the maximum size. This is the major weakness with polynomial based models and only with careful placement of knots can a reasonable fit to the asymptote be obtained. Given a measure of maximum size, useful indicators giving the time of growth to a given proportion of that size can be obtained. As was shown in chapter 3 log α is the most stable parameter of the Richards function to estimate, so this aspect is an advantage for the Richards function approach.

The other advantage the parameters of the Richards has is in giving a simple indicator of shape, which is not possible for the polynomial models. However as shape is given by the parameter n and this parameter is the one that causes all the estimation problems this advantage is less than it may at first seem. Any comparison of the values of n needs to be treated with great caution.

Although more choice is required for polynomial models, degree and/or position of knots, they as a consequence make the user think about his data. There is a great danger in using an automatic model like the Richards specially when it has severe estimation problems.

An alternative approach to leaf growth modelling is provided by Charles-Edwards (1979). He produces a mechanistic model. He aims to use information on the chemical components of leaves based on experiments using approaches such as ^{14}C labelling. The structure of the model is given in figure 8.4(1). There are a total of 11 variables in the model

1 Mass of labile nitrogen (available ions)

2 Mass of labile carbohydrate

3 Mass of tissue water

4 Mass of degradable

5 Mass of non-degradable structure

6 Leaf volume

7 Leaf area

8 Leaf thickness

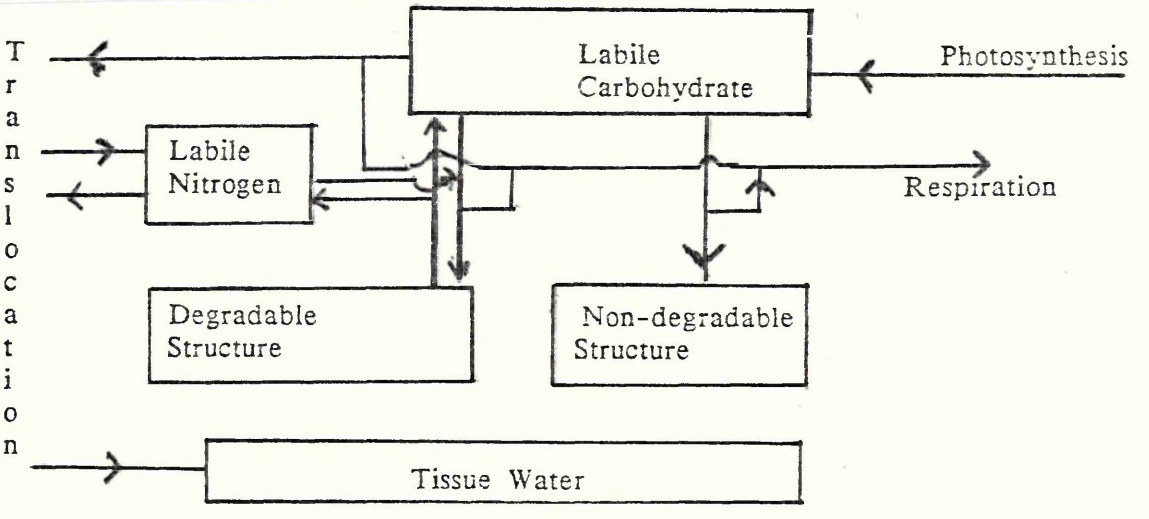
9 Ambient CO_2

10 Light flux density

11 Time

These are related by a series of differential equations.

Figure 8.4(1)



There are also a total of 16 parameters in the model. Charles-Edwards carries out simulations of the model for given values of the parameters, initial values of variables 1 to 8 and values of the exogenous variables 9 to 11.

Estimates of the 16 parameters would have to be obtained from a number of different experiments and it is open to question whether these can be combined in one model. There is no account for the inherent variation in the model but a stochastic differential equation analogue would be possible. As some of the differential equations are non-linear the expected value of the stochastic model would not be the same as the solution to the deterministic model.

Estimation of some of the parameters could be achieved by recording values of variables 6 to 11 and using the Kalman filter estimation procedures considered earlier. This would require a great amount of data. It may also be possible to record other variables such as leaf mass, photosynthetic rate etc which could be used to estimate parameters as such variables play an implicit role in the model

The mechanistic model and the empirical model both have a part to play in the understanding of the growth of a leaf. The mechanistic approach still has many statistical problems to overcome, indeed the Richards function is not free from them. The use of the simpler polynomial based models will often provide a useful check on any results or conclusions derived from a more sophisticated model.

8.5 Conclusions

This chapter illustrates the way a wide range of stochastic models can be used. The different types of models will often have different levels of interpretation and these should be viewed as complementary rather than in conflict. The importance of the range of models to be considered needs to be emphasised at the data collection stage. Adequate data for mechanistic type stochastic models can then be obtained.

It is to be hoped that complete systems models using the ideas considered in the first part of this chapter will one day be available to the statistical modeller.

9. Statistical Models for Human Growth

9.1 Introduction

There has always been a fascination with recording the growth of a child. Indeed one of the most famous sets of data was recorded by the Count de Montbeillard between 1759 and 1777.

One of the main features of the growth of an individual is the adolescent spurt. The peak growth velocities for girls is between 12 and 13 years of age and for boys between 14 and 15 years of age. In general for early maturers the growth process goes more quickly and more intensely so that a greater total growth is achieved. There is some evidence that economic privation may slow the rate of skeletal development. There is also a seasonal effect to growth. Growth in height being fastest in spring while growth in weight is fastest in autumn. The fundamental control of the rate of growth seems to be genetical but malnutrition can have an enormous effect.

Many models for the growth of an individual have been proposed to reflect the complex pattern that is observed. these will be reviewed in this chapter and two new models developed. These models will be applied to data on the growth of a child and the way in which dynamic models could be used will be considered.

9.2 Models for Growth

9.2.1 Pre-adolescent Growth Models

Two basic models have been used for pre-adolescent growth

(i) The Count model

$$y = \beta_0 + \beta_1 t + \beta_2 \log t + e$$

(Count (1943), Guire and Howalski (1979), Berkey (1982))

(ii) The Jenss and Bayley model

$$y = \beta_0 + \beta_1 t - \exp(\beta_2 + \beta_3 t) + e$$

(Jenss and Bayley (1937), Guire and Howalski (1979), Berkey (1982)).

These models may then be combined with a Gompertz curve for the adolescent period.

Berkey (1982) compared the two models using data from the Longitudinal Studies of Child Health and Development at the Harvard School of Public Health (Stuart and Reed (1959)). She concluded that the Count model does not adequately fit either weight or length for the period three months to six years, but does discriminate reliably between individuals. The Jenss and Bayley model fitted well except for a minor problem near six months. The main advantage of the Count model over the Jenss and Bayley is that it is linear in the parameters.

Models for the entire growth periods are the mixture of logistics or the compounded logistics as considered in chapter 3.

9.2.2 Bock Models

Bock et al (1973) introduced the double logistic model

$$y = \frac{a_1}{1 + \exp(-b_1(t - c_1))} + \frac{f - a_1}{1 + \exp(-b_2(t - c_1))}$$

where

a_1 - upper limit of prepubertal growth

b_1 - initial RGR of prepubertal growth

c_1 - location in time of prepubertal growth

f - mature size

$a_2 = (f - a_1)$ - contribution of adolescent growth

b_2 - initial RGR of adolescent growth

c_2 - age at maximum rate of adolescent growth

A development of this model is the triple logistic Bock and Thissen (1976)

$$y = a_1 \left\{ \frac{1 - p}{1 + \exp(- b_1(t - c_1))} + \frac{p}{1 + \exp(- b_2(t - c_2))} \right\} + \frac{a_2}{1 + \exp(- b_3(t - c_3))}$$

where

a_1 - upper limit of prepubertal growth

b_1 - initial RGR of early childhood growth

c_1 - age at maximum rate of early childhood growth

b_2 - initial RGR of middle childhood growth

c_2 - age at maximum rate of middle childhood growth

p - proportion of prepubertal growth attributable to the middle childhood component

a_2 - contribution of the adolescent component to total size

b_3 - initial RGR of adolescent growth

c_3 - age at maximum rate of adolescent growth

Both these models and the Count/Jenss and Bayley-Gompertz combination try to model different parts of the growth period with basic growth models.

9.2.3 Preece and Baines Models

Preece and Baines (1978) start by noting that plots of

$$s = \frac{dy}{dt}\frac{1}{y_{max} - y}$$

have a sigmoid shape and proposing the model

$$\frac{ds}{dt} = \gamma(s_1 - s)(s - s_0) \quad \text{ie logistic model}$$

with

$$\frac{dy}{dt} = s(y_{max} - y)$$

This leads to the model

$$y = y_{max} - \frac{y_{max} - y_\theta}{\left\{\frac{1}{2}\exp(\gamma s_0(t - \theta) + \frac{1}{2}\exp(\gamma s_1(t - \theta))\right\}^{1/\gamma}}$$

using boundary conditions $y = y_\theta$ when $t = 0$.

A simplified version being obtained when $\gamma = 1$. This simpler version can then be made more flexible by considering s to be the sum of two logistics

$$s = p + q$$

$$\frac{dp}{dt} = (p_1 - p)(p - p_0) \qquad \frac{dq}{dt} = (q_1 - q)(q - q_0)$$

setting $q_0 = 0$ did not seriously affect the generality so the model

$$y = y_1 - \frac{4(y_1 - y_0)}{\{\exp(p_0(t - \theta) + \exp(p_1(t - \theta))\}\{1 + \exp(q_1(t - \theta))\}}$$

9.2.4

As was considered for plant growth the problem of biologically meaningful parameters has to be considered. The main parameters of interest are

Adult height

Age at start of adolescent growth spurt

Size at start of adolescent growth spurt

Growth rate at start of adolescent growth spurt

Age at peak growth rate

Peak growth rate

These values can be derived from the model parameters, sometimes via a closed form.

Other approaches include the use of polynomials (Joossens and Brems-Heyns (1975), Splines (Largo et al (1978) and Rao's method of transforming the time scale (Wingard (1970)).

Discussion of these different models and approaches will be left until after the examination of the data and suggested models of the next section.

All the models assume an additive error. Examination of residuals has shown no evidence of heteroscedasticity (eg Preece and Bains (1978)). The problem of non-independence seems to be more serious (Preece and Bains (1978), Bock and Thissen (1980)).

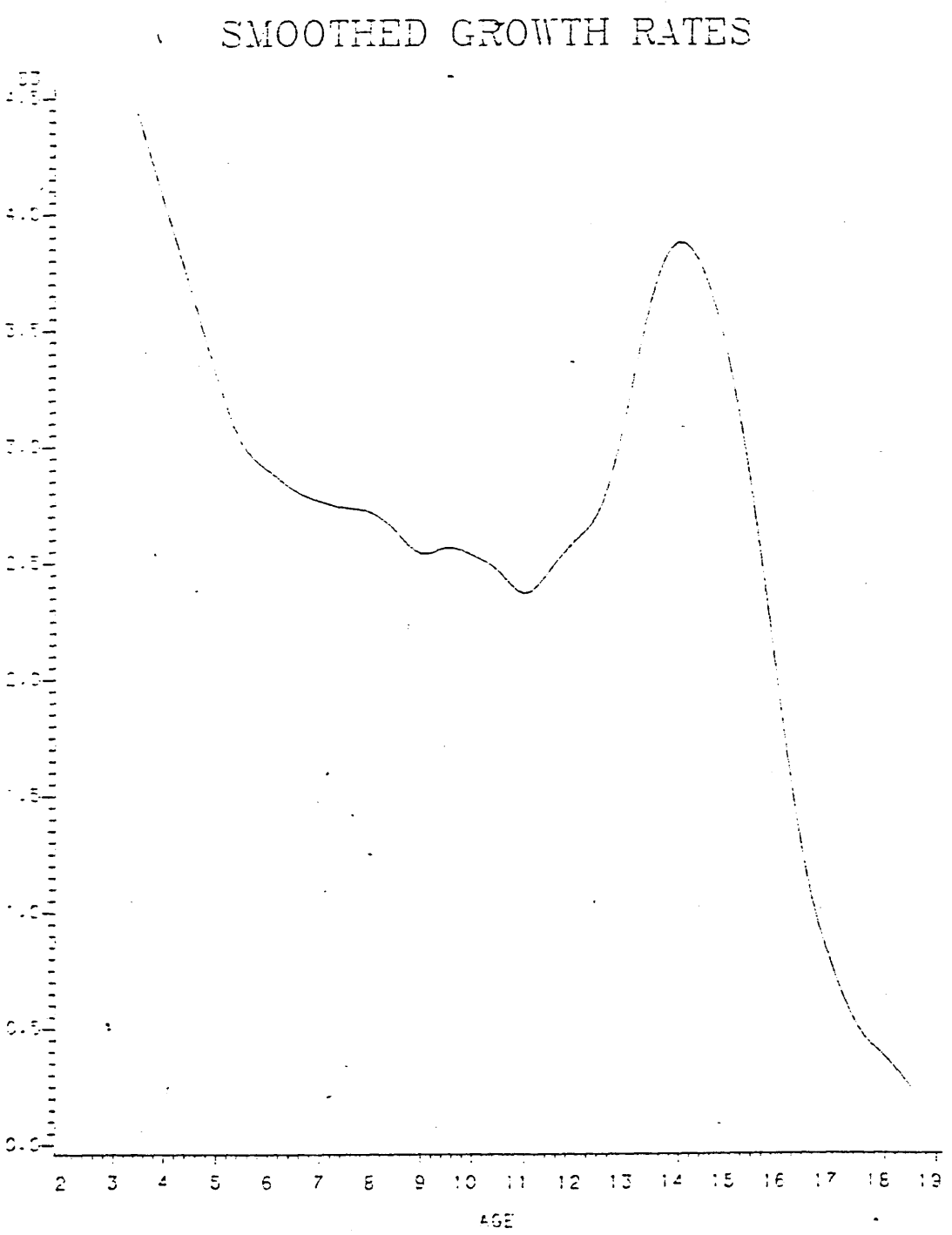
9.2 Two New Models

Plots of the growth rate of children show a characteristic shape as shown in figure 9.3(1). The shape suggests a mixture of two functions

e^{-kt} and e^{-kt^2}

at different origins. If r is the growth rate then two possible models are

Figure 9.3(1)


SMOOTHED GROWTH RATES
2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19
AGE

$$r = \alpha\, e^{-kt} + \beta\, e^{-\gamma(t-\theta)^2} \qquad (A)$$

and

$$r = \alpha\, e^{-kt} + \beta\, t\, e^{-\gamma(t-\theta)^2} \qquad (B)$$

As

$$y_t = \int_0^t r(t)dt$$

(A) gives

$$y = \alpha/k(1 - e^{-kt}) + \beta \int_0^t e^{-\gamma(t-\theta)^2}dt$$

$$\text{as } t \to \infty \quad y \to {}^{\alpha}/_{k} + \frac{1}{2}\,\beta({}^{\pi}/_{\gamma})^{\frac{1}{2}}$$

This suggests the reparameterisation

$$y = h_0\,(1 - e^{-kt}) + 2h_1 \sqrt{\frac{\gamma}{\pi}} \int_0^t e^{-\gamma(t-\theta)^2}$$

so that total growth is $h_0 + h_1$ where h_1 is the adolescent component.

Similarly for (B)

$$y = h_0(1 - e^{-kt}) + h_1(1 - e^{-\gamma(t-\theta)^2}$$

then replacing $h_0 + h_1$ by h.

The main difference between the two models is that (A) has a symetric adolescent growth spurt while (B) is skewed. For (A) the peak velocity is at $t = \theta_0$ while for (B) it is at

$$t = \frac{1}{\gamma\sqrt{2}} + \theta$$

The expression

$$\sqrt{\frac{\gamma}{\pi}} \int_0^t e^{-\gamma(t-\theta)^2}dt$$

does not have a closed form but as it is

$$\Phi(\sqrt{2}\ \gamma(t - \theta)) - \Phi(-\sqrt{2}\ \gamma\ \theta)$$

where Φ is the standard normal integral. Generally $\Phi(-\sqrt{2}\ \gamma\ \theta)$ will be negligible so the model can be written as

$$y = h_0(1 - e^{-kt}) + 2h_1\ \Phi(\sqrt{2}\ \gamma(t - \theta))$$

There are many numerical approximations for $\Phi(\ )$ (Abramowitz and Stegun (1965)).

One final variation can be considered, greater flexibility would be achieved by replacing

$$h_0(1 - e^{-kt})$$

by

$$h_0 - \exp(\beta - kt)$$

The models are now

(A) $y = h_0 - \exp(\beta - kt) + 2h_1\ \Phi(\sqrt{2}\ \gamma(t - \theta))$

(B) $y = h - \exp(\beta - kt) - h_1\ \exp(-\gamma(t - \theta)^2)$

$h, h_0, h_1, k, \gamma, \theta > 0$

These can be seen as extensions of the Jenss and Bayley model.

9.3 Example

9.3.1 The Data

The data is given in table 9.3(1). It is 34 observations on the height of a boy from the age of $3\frac{1}{2}$ to the age of 20. The observtions are not equally spaced. They are initially twice a year, then at the time of the adolescent spurt they are four times a year, finally the frequency reduces to yearly. The data was provided by Dr M Preece of the Department of Growth and Development, The

Institute of Child Health.

9.3.2 Identification

As the data is not equally spaced it had to be adjusted before the orthogonal polynomial method of section 2.3 could be used. The quarterly observations were dropped and some half year observtions 'invented' for the final part, though this was not vital. An example of the resulting smoothed growth rate is shown in figure 9.3(1). This shows the characteristic shape of a decline followed by a peak centred on age 14.

Table 9.3(1)

Age	Height (cms)
3.501	92.5
4.033	97.3
4.501	100.9
5.096	104.2
5.499	107.4
6.052	110.2
6.510	112.7
7.047	115.6
7.510	118.6
8.044	121.2
8.485	123.9
9.041	126.4
9.501	129.1
10.038	131.5
10.496	133.7
11.033	136.9
11.312	139.0
12.030	141.0
12.260	142.4
12.510	143.1
12.759	145.0
13.049	146.9
13.258	148.2
13.545	149.9
14.047	153.3
14.333	155.7
14.501	158.1
14.770	160.9
15.038	163.0
15.499	165.8
16.036	168.1
17.033	169.7
18.049	170.5
19.101	171.1
20.041	171.1

9.3.3 Static Models

A number of models were fitted using SAS. The results are given below

(i) Model A

Parameter	Estimate	Std Error
h_0	167.3	6.04
h_1	8.884	1.106
β	4.725	0.03123
k	0.1145	0.01402
γ	0.5717	0.04323
θ	13.8629	0.04503

Resid ss = 34.7281

(ii) Model B

Parameter	Estimate	Std Error
h	179.42	13.21
log h_1	3.1149	1.1034
β	5.002	.1753
k	.1702	.1537
γ	.03540	.03251
θ	9.5295	1.4556

Resid ss = 93.0888

The parameter transformation to log h_1 had to be used as the iterative procedure (Gauss-Newton or Marquardt) gave negative values of h_1 in the unrestricted minimisation.

(iii) Bock Double Logistic

Parameter	Estimate	Std Error
a_1	178.53	2.0105
f	160.28	4.5500
b_1	0.1781	0.0132
b_2	-1.3701	0.2282
c_1	-0.1304	0.04582
c_2	19.1136	3.1851

Resid ss = 20.7029

(iv) Preece and Baines Model $\gamma = 1$

Parameter	Estimate	Std Error
y_{max}	171.3038	0.3274
y_θ	158.64	0.2745
s_0	0.1029	0.001540
s_1	1.1716	.05085
θ	14.5413	.05253

Resid ss = 10.1522

(v) Preece and Baines Model 3

Parameter	Estimate	Std Error
y_1	170.78	0.2003
y_0	160.66	0.3131
p_0	0.06892	0.0009094
p_1	0.2438	0.01738
q_1	1.3957	0.07224
θ	14.7895	0.04556

Resid ss = 3.9062

(Note all Std Errors are approximate).

Looking at the residual sum of squares as well as plots of the residuals the best fitting model is clearly Preece and Baines Model 3. The non-linearity of this model was then checked giving results

$\Gamma^T = .0746$

$\Gamma^N = .0049$

$N_\theta = .0001$

$N_\varphi = .0000$

Parameter	Bias
y_1	$.27 \times 10^{-4}$
y_0	$.73 \times 10^{-4}$
p_0	$.23 \times 10^{-6}$
p_1	$-.23 \times 10^{-5}$
q_1	$.16 \times 10^{-5}$
θ	$.31 \times 10^{-4}$

As can be seen there is negligible non-linearity.

In order to fit this model with an ARMA error the data set was reduced to 24 equally spaced observations. Fitting the model, assuming independent error, to

this reduced data set gave the results

Parameter	Estimate	Std Error
y_1	169.90	0.8595
y_0	160.38	0.5142
p_0	0.07118	0.002405
p_1	0.2680	0.02779
q_1	1.5672	0.1618
θ	14.7563	0.07755

showing slight differences from the model fitted to the complete data set.

Both an AR(1) and an MA(1) were considered as possible models for the error structure. Only the MA(1) showed any improvement and gave the following results

Parameter	Estimate
MA	.1883
y_1	169.93
y_0	160.34
p_0	0.07120
p_1	0.2680
q_1	1.5587
θ	14.7513

The change in deviance was only 0.34 so was not significant.

It can be concluded that for this reduced data set there is no evidence of correlation in the errors about this model.

Due to the complexity of the growth over the period considered it would not be possible to find a simple dynamic model to provide adequate fit. However, if time dependent parameters are used a suitable model may be found. This is what Preece and Baines used. To illustrate how a stochastic dynamic model with time dependent parameters could be used the following simple model was derived.

Using the basic linear growth model

$$\frac{dx}{dt} = \lambda(t)x$$

here $\lambda(t)$ varies with time. This model can be recast in discrete form and a stochastic error included as was considered in section 5.3.5

$$x_t = \exp\left[\int_{t-\tau}^{t} \lambda(u)du\right] x_{t-\tau} + \varepsilon_t$$

Now

$$\lambda(t) = \frac{1}{x}\frac{dx}{dt} = \frac{d}{dt}\log(x)$$

is the relative growth rate and can be estimated by the usual formula

$$\hat{\lambda}t = \frac{\log y_t - \log y_{t-\tau}}{\tau}$$

A simple approximation to the behaviour of $\lambda(t)$ for a child is a cubic polynomial, p(t). Fitting such a polynomial to the estimated relative growth rates for the data being considered, with the final two observations removed gave

$$p(t) = 1.9087 - 0.5930t + .06547t^2 - .0021136t^3$$

The final two observations were removed because there was negligible growth at

this stage and the observations were a year apart, thus removing them would give a better fit by the cubic polynomial.

If

$$A(t) = \int_{t-\tau}^{t} P(u)du$$

when the model can be written in state space form as

$$x_t = \exp(\beta\ A(t)\ x_{t-\tau} + \varepsilon_t$$

$$y_t = x_t + e_t$$

with

$$\text{var}(\varepsilon_t) = \sigma_2{}^2 \quad \text{and} \quad \text{var}(e_t) = \sigma_e{}^2$$

The parameters, β, σ_ε^2 and $\sigma_e{}^2$, can be estimated by maximum likelihood using a linear discrete time Kalman filter as in sections 5.2.2 and 5.2.3. The function A(t) can be seen as transforming the time scale so as to allow the use of a linear model, a similar approach to Rao (1958).

The results were

Parameter	Estimate
β	1.001
σ_e	2.676
σ_ε	0.00046

The σ_ε estimate being unstable.

A simulation from this model is shown in table 9.3(2) showing a not unreasonable result.

Table 9.3(2)

95.1
97.8
100.4
103.1
105.8
108.5
111.2
113.9
116.6
119.3
122.1
124.8
127.6
130.3
133.1
135.9
138.6
141.3
144.0
146.7
149.4
152.2
154.9
157.6
160.3
162.9
165.6
168.3
170.8
173.1
175.1

This model is only a simple attempt at a stochastic dynamic model for child growth.

A more suitable function for A(t) could be found. The fact that the observations are not equally spaced and that the relative growth rate is not constant should be incorporated in the variance of the error term. Finally any parameters of A(t) which have to be estimated from the data could be estimated at the Kalman filter stage, not separately as was carried out in this study.

The dynamic approach has potential but needs further development.

9.4 Conclusions

This chapter has shown the wide range of models available to the modeller of child growth. New models have been introduced that illustrate how the growth rate may be modelled using both static and dynamic models.

10. Discussion

The aim of this study was to examine various models and modelling procedures for biological growth situations. One of the key reasons for modelling, as was discussed in chapter 7, is to give understanding of the situation being studied. In the introduction the statistical approach using linear models was considered in some detail. This approach has the advantage of closed form solutions and, often, known sampling distributions. However, the unrealistic form of the models do not generate much insight into the situation. In general one can only talk about linear growth and deviations from linear growth.

This study has concentrated on looking at two alternatives to the linear statistical models. First non-linear models derived from deterministic dynamic models. The problem of non-independent errors has been considered as well as the effects of non-linearity on the properties of the estimates. While doubts have been cast over the usefulness of the Richards function others, eg the Preece-Baines model, are very promising.

The second approach was stochastic dynamic models. In the use of these models the Kalman filter has proved most useful. It is to be hoped that there will be a greater use of such models in biology. This may happen as either statisticians start to use dynamic models rather than just static regression models or mathematical modellers, who often use dynamic models, start to take the stochastic element seriously.

The three types of models: linear, non-linear and stochastic-dynamic, need not be seen as separate alternatives but as complementary, each giving its own insight into the situation. This study has set out to begin to show how this is to be done.

There is a great deal more work to be done. The properties of the estimates considered will have to be studied further. Also the properties of the standard errors of the estimates. The standard errors will often have to be obtained by using numerical differentiation of the log likelihood function. Methods exist for doing this (eg Lyness (1966, 1969)) but their properties need to be examined.

Aside from the pure statistical problems, any study of modelling needs to be firmly rooted in the situation being modelled. It should involve the interaction between the specialist, eg plant physiologist, and the statistical modeller. I have been aware that much of the work in this study has been 'back room' work away from the application area. The next stage in development of any of the ideas should be in the field.

APPENDIX A

In this appendix algorithms used to minimise non-linear functions will be considered. These have been used throughout this project.

A.1 Linear Searches

It is often useful to minimise a function in a single given direction without using derivatives. These minimisations may form part of a more complex multi-dimensional minimisation procedure.

Powell's method is based on fitting a cubic polynomial to three evaluations of the function and finding its minimum (Box et al (1969), Walsh (1975). It is used in his sum of squares procedure (see A.3).

A.2 General Search Methods

One of the most efficient methods to search in more than one direction is that of Nelder and Mead (1965). This involves the use of a non-regular simplex, whose shape can be changed depending on the change encountered. Termination takes place when the standard deviation of the function values of the vertices of the simplex are sufficiently small.

A FORTRAN subroutine was given by O'Neill (1971) with corrections by Chambers and Ertel (1974), Benyon (1976) and Hill (1978).

This routine was used extensively when the derivatives of the functions would have been complex to evaluate thus making it preferable to derivative based methods such as the Newton-Raphson or Davidon-Fletcher-Powell methods (Walsh (1975)). An alternative would be to use these methods with finite differences instead of derivatives (eg Stewart (1967)). For the exploratory work of this project the Nelder-Mead algorithm was found to be simple and effective. For implementation of the methods it is likely that the most efficient may be the use of finite differences with the Davidon-Fletcher-Powell method (Stewart (1967), Himmelblau and Lindsay (1980)).

A.3 Minimising Sums of Squares

The standard routine for fitting a non-linear regression model is the Gauss-Newton method. This uses a local linear approximation to the non-linear function and the required updates are given by the least squares estimates of a linear regression model (eg Draper and Smith (1981), Bard (1974)). A related method is that of Marquardt (1963) which adjusts the Gauss-Newton directions by an amount in the direction of steepest decent of the sum of squares surface. Both these methods require the derivative of the function, but again these can be replaced by finite differences.

A derivative free method was given by Powell (1965). Based on the Gauss-Newton idea it maintains the orthogonality of the directions in which the finite difference approximations to the derivatives are evaluated. More recently Ralston and Jennrich (1978) produced their DUD algorithm which replaces the tangent approximation of the Gauss-Newton method by a secant approximation.

It is generally true that a specialist sum of squares algorithm will be more efficient than a general minimisation method at minimising a sum of squares function.

A FORTRAN program for Powell's method is given by Kuester and Mitse (1971) and was used in the project.

The SAS procedure Proc NLIN provides the Gauss-Newton, Mardquardt and DUD methods.

A Gauss-Newton method provided by Nag (E04GEF) was also used.

APPENDIX B

A number of FORTRAN programs were written in the course of this study. They were developmental programs and were not intended for general use. However, a listing of the programs is available on request.

The programs are currently run under the FORTRAN 77 compiler but many were written for the FORTRAN 4 compiler and can be easily converted. Some of the more recent programs make use of FORTRAN 77 facilities.

All programs were run on the IBM Mainframe computer at the Polytechnic. Until very recently there was no batch facility on the machine so large simulation runs were not possible.

Where suitable, the statistical package SAS was used, in particular the NLIN procedure. SAS programs for some of the methods of section 1.3 are given by Allen (1983).

REFERENCES

Abramowitz M and Stegun I A (1965); Handbook of Mathematical Functions; Dover, New York

Ackoff R L and Sasceni M W (1968); Fundamentals of Operational Research; John Wiley and Sons

Allen O B (1983); A Guide to the analysis of growth curve data with special reference to SAS; Computers and Biomedical Research, 16, 101-115

Anderson T W (1963); Asymptotic theory for principal component analysis; Ann Math Statist, 34, 122

Ansley C F (1979); An algorithm for exact likelihood of a mixed autoregressive moving average process; Biometrika, 66, 59-65

Arrowsmith D K and Place C M (1982); Ordinary Differential Equations; Chapman and Hall

Bailey N T J (1964); The Elements of Stochastic Processes with application to the natural sciences; John Wiley & Sons

Baksalary J K, Corsten L C A and Kala R (1978); Reconciliation of two different views on estimation of growth curve parameters; Biometrika, 65, 662-64

Bard Y (1974); Nonlinear Parameter Estimation; Academic Press

Bartlett M S (1978); An Introduction to Stochastic Processes; C U P, Cambridge

Bates D M and Watts D G (1980); Relative curvature measures of nonlinearity (with Discussion); J R Statist Soc, B, 42, 1-25

Beal S L (1982); Bayesian analysis of nonlinear models; Biometrics, 38, 1089-1091

Beale E M L (1960); Confidence regions in nonlinear estimation; J R Statist Soc, B, 22, 41-76

Benyon P R (1976); Remark AS R15; Applied Statist, 25, 97

Berkey C S (1982)a; Comparison of two longitudinal growth models for preschool children; Biometrics, 38, 221-234

Berkey C S (1982)b; Bayesian approach for a nonlinear growth model; Biometrics, 38, 953-962

Bock R D and Thissen D (1976); Fitting multi-component models for growth in stature; Proceedings of the 9th International Biometric Conference, 1:431-442

Bock R D and Thissen D T (1979); Statistical Problems of fitting individual growth curves; Proceedings Report, 265-290

Bock R D, Wainer H, Petersen A, Thissen D, Murray J and Roche A (1973); A parameterisation for individual human growth curves; Human Biology, 45, 63-80

Box G E P (1950); Problems in the analysis of growth and wear curves; Biometrics, 6, 362-89

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

Box G E P and Jenkins G M (1976); Time Series Analysis Forecasting and Control, 2nd Ed; Holden-Day, San Francisco

Box G E P and Tiao G L (1973); Bayesian Inference in Statistical Analysis; Addison-Wesley, Massachusetts

Box M J (1968); The occurrence of replications in optimal designs of experiments to estimate parameters in nonlinear models; J R Statist Soc, B, 290-302

Box M J, Davies D and Swann W H (1969); Nonlinear Optimisation Techniques; ICI Ltd Monograph No 5, Oliver and Boyd

Brieman L and Meisel W S (1976); General estimates of the intrinsic variability of data in nonlinear regression models; J Amer Statist Assoc, 74, 301-307

Brown R L, Durbin J and Evans J M (1974); Techniques for testing the constancy of regression relationships over time; J R Statist Soc, 149-163

Bulmer M G (1975); The statistical analysis of density dependence; Biometrics, 31, 901-911

Bulmer M G (1976); The theory of prey-predator oscillations; Theor Pop Biol, 9, 137-150

Capocelli R M and Ricciardi L M (1974); A diffusion model for population growth in random environment; Theor Pop Biol, 5, 28-41

Causton D R (1969); A computer program for fitting the Richards function; Biometrics, 25, 401-9

Chambers J M (1971); Regression updating; J Amer Statist Assoc, 66, 744-748

Chambers and Ertel (1974); A remark on algorithm AS47.AS R11; Applied Statist, 23, 250-251

Charles-Edwards D A (1979); A model for leaf growth; Ann Bot, 44, 523-535

Charles-Edwards D A and Ludwig L J (1974); A model for leaf photosynthesis by c_3 plant species; Ann Bot, 38, 921-30

Charles-Edwards D A and Ludwig L J (1975); A model of leaf carbon metabolism; Ann Bot, 39, 819-29

Checkland P B (1972); Towards a system-based methodology for real world problem solving; J Systems Eng, 3, 2 : Reprinted in Systems Behaviour (ed Open Systems Group), Harper & Row London

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

Cochran W G and Cox M C (1957); Experimental Designs; John Wiley & Sons

Cole J W L and Grizzle J E (1966); Applications of multivariate analysis of variance to repeated measurements experiments; Biometrics, 22, 11810-828

Conte S D and De Boor C (1972); Elementary Numerical Analysis; McGraw-Hill

Cook R D and Tsai C L (1985); Residuals in nonlinear regression; Biometrika, 72, 23-9

Cooke D and Leon J A (1976); Stability of population growth determined by 2x2 Leslie Matrix with density dependent elements; Biometrics 32, 435-442

Count E W (1943); Growth patterns of human physique : an approach to kinetic anthropometry; Human Biology, 15, 1-32

Cox D R and Small N J H (1978); Testing multivariate normality; Biometrika, 65, 263-72

Cox M G (1972); The numerical evaluation of B-Splines; J Inst Maths Applics, 10, 134-149

Cox M G (1975); An algorithm for spline interpolation; J Inst Maths Applics, 15, 95-108

Cox M G (1978); The numerical evaluation of a spline from its B-spline representation; J Inst Maths Applics, 21, 135-143

Cui Qiwu and Lawson G J (1982); Study on models of single populations ; an expansion of the logistic and exponential equations; J Theor Biol, 98, 645-659

Cumberland W G and Rohde C A (1977); A multivariate model for growth of populations; Theor Pop Biol, 11, 127-139

Davies O L and Ku J Y (1977); Re-examination of the fitting of the Richards growth function; Biometrics, 33, 546-547

Day N E (1966); Fitting curves to longitudinal data; Biometrics, 22, 276-291

Dent W (1977); Computation of the exact likelihood function of an ARIMA process; J Statist Comp and Simul, 5, 193-206

Diderrich G T (1985); The Kalman filter from the perspective of Goldberger-Theil estimators; American Statistician, 39, 193-198

Doob J L (1959); Stochastic Processes; John Wiley & Sons, New York

Drajek L (1985); Problems in fitting nonlinear models; unpublished project report

Draper N and Smith H (1981); Applied Regression Analysis (2nd Ed); John Wiley & Sons

Duncan D B and Horn S D (1972); Linear dynamic recursive estimation from the viewpoint of regression analysis; J Amer Statist Assoc, 67, 815-821

Eberhardt L L (1970); Correlation, regression and density dependence; Ecology, 51, 306-310

Efron B and Hinkley D V (1978); Assessing the accuracy of the maximum likelihood estimator : observed versus expected Fisher information; Biometrika, 65, 457-87

Elias C O and Causton D R (1976); Studies on data variability and the use of polynomials to describe plant growth; New Phytologist, 77, 421-30

El Lozy M (1978); A critical analysis of the double and triple logistic growth curves; Annals of Human Biology, 5, 389-394

Fearn T (1975); A Bayesian approach to growth curves; Biometrika, 62, 89-100

Fearn T (1977); A two-stage model for growth curves which leads to Rao's covariance adjusted estimators; Biometrika, 64, 141-3

Fedorov V V (1972); Theory of Optimal Experiments; Academic Press

Finney D J (1958); The efficiencies of alternative estimators for an asymptotic regression equation; Biometrika, 45, 370-88

Forrester J W (1961); Industrial Dynamics; Massachusetts Institute of Technology

Friedland B (1969); Treatment of bias in recursive filtering; IEEE, AC-14, 359-367

Fuller W A and Battese G E (1973); Transformations for estimation of linear models with nested-error structure; J Amer Statist Ass, 68, 626-632

Gallant A R and Goebel J J (1976); Nonlinear regression with autocorrelated errors; J Amer Statist Assoc, 71, 961-967

Geisser S (1970); Bayesian analysis of growth curves; Sankhya, 32, 53-64

Geisser S and Greehouse S W (1958); An extension of Box's results on the use of the F distribution in multivariate analysis; Annals of Mathematical Statistics, 29, 885-891

Gelb A (ed)(1974); Applied Optimal Estimation; The MIT Press

Gilchrist W G (1984); Sttistical Modelling; John Wiley and Sons

Gillis P R and Ratkowsky D A (1978); The behaviour of estimators of the parameters of various yield-density relationships; Biometrics, 34, 191-198

Glasbey C A (1979); Correlated residuals in nonlinear regression applied to growth data; Appl Statist, 28, 251-259

Glasbey C A (1980); Nonlinear regression with autoregressive time series errors; Biometrics, 36, 135-140

Grassia A and DeBoer E S (1980); Some methods of growth curve fitting; Math Scientist, 5, 91-103

Green J R (1971); Testing departure from a regression without using replication; Technometrics, 13, 609-615

Gregg J V, Mossell C H and Richardson J T (1964); Mathematical Trend Curves : an aid to forecasting; ICI Monograph No 1, Oliver and Boyd

Griffiths D A and Sandland R L (1982); Allometry and multivariate growth revisited; Growth, 46, 1-11

Griffiths D A and Sandland R (1984); Fitting generalized allometric models to multivariate growth data; Biometrics, 40, 139-150

Grizzle J E and Allen D M (1969); Analysis of growth and dose response curves; Biometrics, 25, 357-381

Guire K E and Howalski C J (1979); Mathematical description and representation of developmental change functions on the intra and interindividual levels; In Longitudinal Research on the Study of Behaviour and Development, 89-109, Academic Press

Guttman I and Meeter D A (1965); On Beale's measure of nonlinearity; Technometrics, 7, 623-637

Hahn B D (1984); A mathematical model of leaf carbon metabolism; Annals of Botany, 54, 325-339

Harrison P J and Stevens C F (1976); Bayesian Forecasting; J R Statist Assoc, B, 38, 205-227

Harvey A C and Phillips G D A (1979); Maximum likelihood estimation of regression models with autoregressive moving average disturbances; Biometrika, 66, 49-58

Hassell M P (1981); Anthropod predator-prey systems in May R M (ed) Theoretical Ecology; Blackwell, Oxford

Hershey, Zakin and Sinka (1967); Differentiation of experimental data; I and EC Fundamentals, 6, 414-420

Hill I D (1978); A remark on algorithm AS47 : function minimization using a simplex method - AS R28; Appl Statist, 27, 280-282

Hills M (1968); A note on the analysis of growth curves; Biometrics, 24, 192-196

Himmelblau D M and Lindsay J W (1980); A evaluation of substitute methods for derivatives in unconstrained optimization; Operations Research, 28, 668-686

Hoerl A E and Kennard R W (1970); Ridge regression : Biased estimation for non orthogonal problem; Technometrics, 12, 55-67

Hole C C, Thomas T H, Barnes A, Scott P A and Rankin W E F (1984); Dry matter distribution between shoot and storage root of carrot, parsnip, radish and red beet; Annals of Botany, 53, 625-631

Holmes D I (1983); A graphical identification procedure for growth curves; Statistician, 32, 405-416

Hopkins J W (1966); Some considerations in multivariate allometry; Biometrics, 22, 747-60

Hui Siu L (1984); Curve fitting for repeated measurements made at irregular time-points; Biometrics, 40, 691-697

Hunt R (1978); Plant Growth Analysis Studies in Biology No 96; Edward Arnold, London

Hunt R (1982); Plant Growth Curves; Edward Arnold, London

Hunt R (1982 b); Plant growth analysis : second derivatives and compounded second derivatives of splined plant growth curves; Ann Bot, 50, 317-328

Hunt R and Parsons I T (1974); A computer program for deriving growth functions in plant analysis

Hunt R and Parsons I T (1977); Plant growth analysis : further applications of a recent curve fitting program; J Appl Ecol, 14, 965-8

Huxley J S (1924); Constant differential growth ratios and their significance; Nature, 114, 895-896

Huxley J S (1932); Problems of Relative Growth; Methuen & Co Ltd, London

Huynh H and Feldt L S (1970); Conditions under which mean square ratios in repeated measurement design have exact F-distributions; J Amer Statist Soc, 65, 1582-589

Jaswinski A H (1970); Stochastic Processes and Filtering Theory; Academic Press, New York

Jeffers J N R (1982); Modelling; Chapman and Hall

Jeffries C (1974); Qualitative stability and digraphs in model ecosystems; Ecology, 55, 1415-1419

Jenss R M and Bayley N (1937); A mathematical for studying the growth of a child; Human Biology, 9, 556-563

Jolicoeur P (1963); The multivariate generalization of the allometry equation; Biometrics, 19, 497-9

Jolicoeur P (1984); Principal components, factor analysis and multivariate allometry ; A small sample detection test; Biometrics, 40, 685-690

Joossens J V and Brems-Heyns E (1975); High power polynomial regression for the study of distance, velocity and acceleration of growth; Growth, 39, 535-551

Kalman R E (1960); A new approach to linear filtering and prediction problems; J Basic Eng (ASME), Vol 82D, 35-45

Kalman R E and Bucy R S (1961); New results in linear filtering and prediction theory; Trans ASME J Basic Engng Ser D, 83, 95-108

Katz D, Azen S P and Schumitzky A (1981); Bayesian approach to the analysis of nonlinear models : implementation and evaluation; Biometrics, 37, 137, 142

Kendall M G and Stuart A (1969); The Advanced Theory of Statistics; Vol 1, Distribution Theory 3rd Ed; Charles Griffin & Co, London

Kendall M G and Stuart A (1973); The Advanced Theory of Statistics; Vol 2 Inference and Relationship; Charles Griffin & Co Ltd

Khatri C G (1966); A note on a MANOVA model applied to problems in growth curve; Ann Inst Statist Math, 18, 75-86

Khatri C G (1973); Testing some covariance structures under a growth curve model; J Mult Anal, 3, 102-16

Khatri C G and Shah K R (1982); Estimation in nonlinear growth curves; Unpublished report

Kiester A R and Barakat R (1974); Exact solutions to certain stochastic equation models of population growth; Theor Pop Biol, 6, 199-216

Kleinbaum D G (1973); A generalization of the growth curve model which allows missing data; J Mul Anal, 3, 117-124

Kowal N E (1971); A rationale for modelling dynamic ecological systems in Patten B C (ed); Systems Analysis and Simulation in Ecology Vol 1; Academic Press

Kuester J L and Mize J H (1973); Optimization techniques with Fortran; McGraw-Hill

Largo R H, Gasser T H and Prader A (1978); Analysis of the adolescent growth spurt using smoothing spline functions; Annals of Human Biology, 5, 421-434

Lee Y H K (1974); A note on Rao's reduction of Potthoff and Roy's generalized linear model; Biometrika, 61, 349-351

Leeper J D and Woolson R F (1982); Testing hypotheses for the growth curve model when the data are incomplete; J Statist Comp Simul, 15, 97-107

Leslie P H (1948); Some further notes on the use of matrices in population mathematics; Biometrika, 35, 213-45

Levenback H and Reuter B E (1976); Forecasting trending time series with relative growth rate models; Technometrics, 18, 261-272

Levins R (1969); The effect of random variations of different types on population growth; Proc N A S, 62, 1061-1065

Lewontin R C and Cohen D (1969); On population growth at randomly varying environment; Proc N S A, 62, 1056-1060

Lindley D V and Smith A F M (1972); Bayes estimates for the linear model (with discussion); J R Statist Soc B, 34, 1-41

Liski E P (1984); Missing data problems in growth curves; Compstat 1984, 479-84

Lyness J N and Moler C B (1966); Van der Monde systems and numerical differentiation; Num Math, 8, 458-464

Lyness J N and Moler C B (1969); Generalised Romberg methods for integrals of derivatives; Num Math, 14, 1-14

Maelzer D A (1970); The regression of log N_{n+1} on Log N_n as a test of density dependence : an exercise with computer constructed density independent populations; Ecology, 51, 810-822

Marquardt D W (1963); An algorithm for least squares estimation of nonlinear parameters; J Siam, 11, 431-441

Marquardt D W (1970); Generalised inverses, ridge regression, biased linear estimation and nonlinear estimation; Technometrics, 12, 591-612

May R M (1973 a); Qualitative stability in model ecosystems; Ecology, 54, 638-641

May R M (1973 b); Stability and Complexity in Model Ecosystems; Princeton University Press

May R M (1976); Simple mathematical models with very complicated dynamics; Nature, 261, 459-67

May R M (1968); Theoretical Ecology Principles and Applications; Blackwell, Oxford

May R M, Conway G R, Hassell M P and Southwood T R E (1974); Time delays, density dependence and single species oscillations; J Anim Ecol, 43, 747-70

Maybeck P S (1979); Stochastic Models, Estimation and Control, Vol 1; Academic Press

Maybeck P S (1982); Stochastic models, estimation and control; Academic Press

Meadows D H (1980); The unavoidable a piori : in Randers, J (ed) Elements of the System Dynamics Method; The MIT Press, Massachusetts

Meadows D H, Meadows D L, Randers J and Behrens W W (1972); The Limits to Growth; Pan Books, London

Mehra R K (1972); Approaches to adaptive filtering; IEEE, AC-17, 693-698

Mitchell A F S (1968); Exponential regression with correlated observations; Biometrika, 55, 149-162

Morrison D F (1967); Multivariate Statistical Methods; McGraw-Hill

Nash J C (1977); A discrete alternative to the logistic growth function; Appl Statist, 26, 9-14

Nelder J A (1961); The fitting of a generalisation of the logistic curve; Biometrics, 17, 89-110

Nelder J A (1962); An alternative form of a generalisation of the logistic equation; Biometrics, 18, 614-616

Nelder J A and Mead R (1965); A simplex method for function minimisation; Computer J, 7, 155-162

Newbold P (1974); The exact likelihood function for a mixed autoregressive moving average process; Biometrika, 61, 423-6

Nicholls A O and Calder D M (1973); Comments on the use of regression analysis for the study of plant growth; New Phytologist, 7, 571-81

O'Hagan A (1976); On posterior joint and marginal modes; Biometrika, 63, 329-33

Oliver F R (1964); Methods of estimating the logistic growth function; Appl Statist 13, 57-66

Oliver F R 91966); Aspects of maximum likelihood estimation of the logistic growth function; J Amer Statist Ass, 61, 697-705

Oliver F R (1982); Notes on the logistic curve for human populations; J R Statist Soc A, 145, 359-363

O'Neill (1971); Function minimisation using a simplex procedure, AS 47; Appl Statist, 20, 338-345

Parsons I T and Hunt R (1981); Plant growth analysis : curve fitting program using the method of B-splines; Ann Bot, 48, 341-52

Patten B C 91971); Systems Analysis and Simulation in Ecology, Vol 1; Academic Press

Patton R A and Krause G F (1972); A parameterisation of the logistic function which admits uncorrelated maximum likelihood estimators; Biometrics, 28, 469-474

Pennycuick C J, Compton R M and Beckingham L (1969); A computer model for simulating the growth of a population, or of two interacting populations; J Theoret Biol, 18, 316-329

Pearlman J G (1980); An algorithm for the exact likelihood of a high order autoregressive moving average process; Biometrika, 67, 232-237

Pielou E C 1977); Mathematical Ecology; John Wiley & Sons

Pollard E and Lakhani K H (1986); The detection of density-dependence from a series of annual censuses; Mathematical Ecology Group, November 1985

Potthoff R F and Roy S N (1964); A generalised multivariate analysis of variance model useful especially for growth curve problems; Biometrika, 51, 313-26

Powell M J D (1965); A method for minimising a sum of squares of nonlinear functions without calculating derivatives; Computer J, 7, 303-307

Powell M J D (1967); Curve fitting by cubic splines; Rep TP 307, Atomic Energy Res Est, Harwell, England

Preece M A and Baines M J (1978); A new family of mathematical models describing the human growth curve; Annals of Human Biology, 5, 1-24

Pruitt K M, De Muth R E and Turner M E (1979); Practical application of generic growth theory and the significance of the growth curve parameters; Growth, 43, 19-35

Radford P J (1967); Growth analysis formulae - their use and abuse; Crop Science, 7, 171-5

Ralson M L and Jennrich R I (1978); Dud, a derivative-free algorithm for nonlinear least squares; Technometrics, 20, 7-14

Randers J (ed)(1980); Elements of the Systms Dynamics Method; The MIT Press, Massachusetts

Rao C R (1958); Some statistical methods for the comparison of growth curves; Biometrics, 14, 1-17

Rao C R (1965); The theory of least squares when the parameters are stochastic and its application to the analysis of growth curves; Biometrika, 52, 447-58

Reinsch C H (1967); Smoothing by spline functions; Numerische Mathematik, 10, 177-183

Richards F J (1959); A flexible growth function for empirical use; Journal of Experimental Biology, 10, 290-300

Richards F J (1969); The quantitative analysis of growth. Plant Physiology, Vol VA (Ed Steward F C) 3-76

Richards F S G (1961); A method of maximum likelihood estimation; J R Statist Soc B, 23, 469-476

Ross G J S (1970); The efficient use of function minimisation in nonlinear maximum-likelihood estimation; Appl Statist, 19, 205-221

Rowell J G and Walters D E (1976); Analysing data with repeated observations on each experimental unit; J Argric Sci Camb, 87, 423-432

Sandland R L and McGilchrist C A (1979); Stochastic growth curve analysis; Biometrics, 35, 255-272

Schwertman N C and Allen D M (1979); Smoothing an indefinite variance-covariance matrix; J Statist Comp Simul, 9, 183-194

Schwertman N C, Flynn W, Stein S and Schenk K L (1985); A Monte Carlo study of alternative procedures for testing the hypothesis of parallelism for complete and incomplete growth curve data; J Statist Comp Simul, 21, 1-37

Searle S R (1971); Linear Models; John Wiley & Sons

Slade N A (1977); Statistical detection of density dependence from a series of sequential censuses; Ecology, 58, 1094-1102

Smith C E and Tuckwell H C (1974); Some stochastic growth processes; In Lecture Notes in Biomathematics, 211-225, S Levin (ed) Springer-Verlag

Sprent P (1968); Linear relations in growth and size studies; Biometrics, 24, 639-56

Sprent P (1969); Models in Regression and Related Topics; Methuen & Co Ltd

Sprent P (1972); Mathematics of size and shape; Biometrics, 28, 23-37

St Amant J L S (1970); The detection of regulation in animal populations; Ecology, 51, 823-828

Stewart G W (1967); A modification of Davidon's minimisation method to accept difference approximation of derivatives; JACM, 14, 72-83

Stone R (1980); Sigmoids; Bull App Statist, 7, 59-119

Stuart H C and Reed R B (1959); Longitudinal studies of child health and development, Harvard School of Public Health, Series II, No 1, description of project; Pediatrics, 24, 875-885

Toutenburg H (1982); Prior Information in Linear Models; John Wiley & Sons

Tuckwell H C (1974); A study of some diffusion models of population growth; Theor Pop Biol, 5, 345-357

Turelli M (1977); Random environments and stochastic calculus; Theor Pop Biol, 12, 140-178

Turner M E (1978); Allometry and multivariate growth; Growth, 42, 434-50

Turner M E, Bradley E L, Kirk K A and Pruitt K M (1976); A theory of growth; Mathematical Biosciences, 29, 367-373

Varley G C and Gradwell G R (1968); Predatory insects as density dependent mortality factors; Proceedings of the XVI International Congress of Zoology, Washington, 1, 240

Venus J C and Causton D R (1979); Plant growth analysis : a re-examination of the methods of calculation of relative growth and net assimilation rates without using fitted functions; Ann Bot, 43, 633-638

Vickery W L and Nudds T D (1984); Detection of density-dependent effects in annual duck censuses; Ecology, 65, 96-104

Vierra S and Hoffman R (1977); Comparison of the logistic and Gompertz growth functions considering additive and multiplictive error terms; Appl Statist, 26, 143-148

Walsh G R (1975); Methods of Optimisation; John Wiley & Sons

White L A (1973); An extension of the General Equivalence Theorem to nonlinear models; Biometrika, 60, 345-348

Wilks S S (1946); Sample criterion for testing equality of means, equality of variances, and equality of covariances in a normal multivariate distribution; Ann Math Statist, 17, 257-81

Williamson M (1974); The analysis of discrete time cycles : in Ecological Stability; M B Usher and M H Williamson (eds); Chapman and Hall

Winer B J (1970); Statistical Principles in Experimental Design; McGraw-Hill

Wingerd J (1970); The relation of growth from birth to two years to sex, parental size and other factors, using Rao's method of the transformed time scale; Human Biology, 42, 105-131

Wishart J (1938); Growth-rate determinations in nutrition studies with the Bacon Pig and their analysis; Biometrika, 30, 16-28

Wold S (1974); Spline functions in data analysis; Technometrics, 16, 1-11

Woodford C H (1970); An algorithm for data smoothing using spline functions; BIT, 10, 501-510