SURVIVAL ANALYSIS WITH TIME-RELATED COVARIATES

by

K.L.Q. Read, R.R. Harris, A.A. Noura and Dennis Gillings

Department of Biostatistics
University of North Carolina at Chapel Hill

Institute of Statistics Mimeo Series No. 1375

January 1982

DEPARTMENT OF BIOSTATISTICS
Chapel Hill, North Carolina
SURVIVAL ANALYSIS WITH TIME-RELATED COVARIATES

by

K.L.Q. READ†, R.R. HARRIS†, A.A. NOURA† and DENNIS GILLINGS††

SUMMARY

A general approach to the analysis of survival data is developed, based on the fitting of log-linear and similar models to multidimensional contingency tables formed from such data by suitable categorisation of the time domain and the covariates (including treatment). Consideration is given to non-proportional hazards, time-dependent or sequentially realised covariates, alternative transformations of the survivor function and subsets of the parameters which may be involved non-linearly. The work is presented in terms of weighted least squares estimation, though this is not necessary to the underlying formulation and other methods of fitting may be adopted.

Some keywords: Survival analysis, categorical data, log-linear models, complementary log-log transformation, time-related covariates, weighted least squares.

† Department of Mathematical Statistics and Operational Research, University of Exeter.

†† Department of Biostatistics, School of Public Health, University of North Carolina at Chapel Hill.
1. **Introduction**

The purpose of this paper is to describe a general method for modelling and analysis of survival data with censoring when a variety of risk factors may be expected to condition the experiences of the individuals under study. This situation is in marked contrast to many industrial applications (reliability and life-testing) in which components or materials of supposedly uniform quality are tested under controlled conditions. Survival processes in animal and human populations involve heterogeneous individuals living under at least partly controlled condition. The analyst must first identify factors which, singly or in combination, may affect mortality. Secondly, the relevant factors serve to specify risk groups of the subjects or individuals under study; and thirdly it is required to model the survivor functions appropriate to these groups. In common with much published work on survival analysis, this paper was motivated by consideration of the results of a clinical trial concerned with the treatment of a chronic disease. Although this has conditioned terminology, the basic approach is applicable more generally, for example to surveys of diseased patients, or to workers exposed to long term occupational and other environmental hazards. Also 'death' may be interpreted broadly as the occurrence of some critical event in the problem under study, such as 'death or a repeat myocardial infarction'.

In a typical study, relevant factors such as age, sex, and previous medical history may be known before onset of the disease, initial symptoms may be recorded at or soon after onset, and subsequently symptoms that identify critical developments in the patient's condition may be observed. In the case of a clinical trial, treatment is itself a factor which will be realised at or after the time of onset. It is fair to assume that the processes ways in which these
factors condition survival through time will not be known a priori:
rather, it is in the hope of gaining such knowledge that survival studies
are undertaken. A robust analysis based on a minimum of structural
assumptions, expressed or implied, about the data is therefore called
for. Otherwise, strong restrictions on the form of model which are
unsupported by prior knowledge may act like a straitjacket on the
data, so that erroneous relationships may be implied and true
patterns go undetected. However, a review of the literature suggests
that the possibility of varying forms of time-dependence across risk
groups has not been fully explored. Thus a good deal of effort has been
invested in the proportional hazards (PH) model, in which the hazards
for all subjects are multiples of an underlying function of time. In
Cox's (1972) analysis, for example, no explicit form for this time
dependence need be assumed; however, although the possibility is acknow-
ledged, there appears to be relatively little work on general
alternatives to the PH model. On the other hand, the fitting of linear
models to multidimensional contingency tables of categorised data has
proved a fruitful means of catering for time-independent covariates
without building strong presuppositions of structure into the analysis.

In this paper, a synthesis of these two approaches is attempted. In
the simplest (non-sequential) situation when the definition of \( z \) does not
change with time a two-way table is set up in which the columns
correspond to successive time-intervals and the rows to different sub-
populations or risk groups. Each individual is then located in the cell
determined by his risk vector \( z \) and survival time. A full set of
parameters (one less than the number of intervals) is available to
describe the experience of each risk group through time, and there need
be no prior or built-in relationships between the parameters of different
groups. When concomitant information arises sequentially at specific points
in time after $t = 0$, it is natural to divide the time axis at these points. Since the definition of $x$ and the numbers of risk groups then vary between intervals, it is necessary to modify the above construction of the two-way table, but this entails no new problems of principle. In both cases strong prior assumptions about the effect of $x$ are avoided by the use of indicator variables, and a corresponding representation is adopted for (the cumulative integral of) $\lambda_0(t)$.

A review of relevant literature is presented in Section 2. In Sections 3 and 4 respectively the analysis framework is described for the cases on non-sequential and sequential covariates as distinguished above. A brief discussion is given in Section 5, and this is followed by an Appendix in which specific systems of suitable parametric models are considered.
2. **A Review of Some Earlier Work:**

**Comparison-invariance and Proportional Hazards**

If \( t \) denotes time measured from an arbitrary origin such as the initial exposure to an occupational or environmental risk or from the onset of a disease or experimental test, then the survivor function of a typical individual or subject may be written as

\[
p(T > t | z) = \mathbb{Q}(t | z) = \exp \left( - \int_{0}^{t} \lambda(u | z) \, du \right)
\]  

(2.1)

for a process in continuous time, where \( \lambda(t | z) \) is the hazard function at time \( t \) for a subject with concomitant vector \( z \) which for the present is assumed not to depend on time and which can therefore be specified at time \( t = 0 \).

Analysis of the survival data is aimed at modelling \( \mathbb{Q} \)
or equivalently \( \lambda \). Two aspects of this are (i) the dependence of \( \mathbb{Q} \) on \( z \) for fixed \( t \), i.e. the variation of the chances of survival to time \( t \) with the vector of risk factors \( z \); and (ii) the mortality experience through time for subjects with a given concomitant (or risk) vector \( z \), i.e. the dependence of \( \mathbb{Q}(t | z) \) on \( t \). In regard to (i) a special case arises when, if the hazards of all subjects are ordered at time \( t \), say as

\[
\lambda(t | z_1) < \lambda(t | z_2) < \ldots < \lambda(t | z_n),
\]

their functional forms entail the same ordering at all other times.

In this event the survival probabilities are in the reverse order

\[
\mathbb{Q}(t | z_1) > \mathbb{Q}(t | z_2) > \ldots > \mathbb{Q}(t | z_n),
\]  

(2.2)
determined for all time by the time-independent risks \( z \). In so far as comparisons (i.e. orderings) of the chances of survival of subjects with different risk vectors are invariant through time, (2.2) represents a property of comparison-invariance, which we shall refer to as CI.
A sufficient but not necessary condition\(^1\) for the CI of a set of survivor functions is that the hazard \(\lambda(t|z)\) be separable, i.e. factorise as the product of a function of time only by a function of \(z\) only:

\[
\lambda(t|z) = \lambda_0(t)h(z) 
\]

(2.3)

(2.3) is often referred to as the proportional hazards (PH) model; clearly the functions \(\lambda_0\) and \(h\) are arbitrary to within a constant multiple. \(\lambda_0(t)\) may conveniently be regarded as a reference or 'minimum risk' hazard function appropriate to standard or normative conditions coded as \(z_0\) (say) and for which \(h(z_0) = 1\). Furthermore, the representation (2.3) will later be used generally on the basis that \(z\) may depend on time.

**Cox's Approach and Some Developments**

In many cases, such as finding the best forms of treatment for classes of patients suffering from a degenerative or other severe disease, interest mainly resides in identifying, and then comparing outcomes for, different risk groups (= risk-factor-by-treatment combinations, altogether indexed by \(z\)). By contrast, there may be little knowledge of (or concern with) the time-dependence of \(\lambda(t|z)\), that is, given proportional hazards, of the form of \(\lambda_0(t)\). Analytical methods which do not need strong prior (and possibly wrong) structural assumptions about \(\lambda_0(t)\) (for example, that all survival time distributions belong to the exponential family) may be expected to have a major advantage of robustness. Thus the method of Cox (1972) consists in maximising a marginal likelihood which does not involve \(\lambda_0(t)\) (which could be regarded as a nuisance function). Kalbfleisch and Prentice (1973) have

---

\(^1\) Suppose that we have two groups of subjects, indexed by \(z=0\) and \(z=1\), and that \(\lambda(t|z) = \lambda(1+t)^z\) for \(t>0\), \(\lambda\) on the RHS being a positive constant. Clearly the hazards of the two groups are not proportional. However, the survivor functions \(\Psi(t|0)\) and \(\Psi(t|1)\) are respectively given by \(\exp(-\lambda t)\) and \(\exp(-\frac{\lambda t^2}{2} - \lambda t)\), and \(\Psi(t|1) < \Psi(t|0)\) for all \(t > 0\).
shown that the ordering of individuals who die (fail) by their failure times is a marginally sufficient statistic for the parameters of \( h(x) \), i.e. sufficient for these parameters in the absence of knowledge of \( \lambda_0(t) \). This leads to the formulation of the marginal likelihood \( L \) as

\[
L = \prod_{t_i \in \mathcal{R}_i} \frac{\lambda(t_i | z_i)}{\sum_{t \in \mathcal{R}_i} \lambda(t_i | z_i)}
\]

in which a total of \( n \) subjects die (and known numbers of subjects may be censored or lost to follow-up at various known times), and \( \sum_{t \in \mathcal{R}_i} \) runs over all those who survive until time \( t_i \) (or \( t_i = 0 \)), that is, over the risk set \( \mathcal{R}_i \) at time \( t_i \). With proportional hazards the null hazard \( \lambda_0(\cdot) \) cancels from (2.4) and Cox's choice \( h(x) = \exp(\beta'x) \) then gives

\[
L = \prod_{t \in \mathcal{R}_i} \frac{h(z_i)}{\sum_{t \in \mathcal{R}_i} h(z_i)} = \prod_{t \in \mathcal{R}_i} \frac{\exp(\beta'z_i)}{\sum_{t \in \mathcal{R}_i} \exp(\beta'z_i)}
\]

The linear form \( \beta'x \) permits a systematic exploration of the data using likelihood ratio tests corresponding to the nested hypothesis strategy for linear models whilst the resulting ML estimates of parameters are marginally sufficient and have the usual BAN properties. Given \( \hat{\beta} \), Cox subsequently estimates \( \lambda_0(t) \) as a sum of separately estimated contributions arising from the failure points \( t_i \) not later than time \( t \), taking \( \lambda_0(t) = 0 \) except at these points. Breslow (1974) represents \( \lambda_0(t) \) as taking a series of constant values (to be estimated) between successive failure points and estimates all parameters simultaneously. Cox’s treatment of the (slightly-grouped) discrete-time case,

1 Despite the likely large number of parameters to be estimated, Breslow’s final equations for \( \hat{\beta} \) are identical with those of Cox if, in the case of simultaneous failures following slight grouping on a continuous time scale, Peto’s approximation to \( L \) replaces that of Cox (Cox (1972), discussion). It may be conjectured on the basis of this equivalence that the estimation of as many parameters as failure-points simultaneously with \( \hat{\beta} \) may not necessarily upset the inferences on \( \hat{\beta} \).
in which tied failure times may occur, is refined by Kalbfleisch and Prentice (1973), who give the marginal likelihood for $\theta$ to which Cox's likelihood is a fair approximation when the grouping intervals are narrow. Following Cox, these authors first estimate $\hat{\beta}$ and then in terms of their $\hat{\beta}$ estimate contributions to $\lambda_0(t)$ arising from each failure time. The resulting estimate of $\hat{Y}(t|z)$ is a step function related to the Kaplan-Meier (1958) product limit estimate. Kalbfleisch and Prentice also analyse with chosen (smaller) number of time (not failure) points, $\lambda_0(t)$ being assumed constant between successive time points. The ML estimates of these constants are easily obtained in terms of $\hat{\beta}$ and $\lambda_0(t)$ may thus be modelled in terms of as many or as few parameters as desired. This development yields continuous estimates of survivor functions and in the use of prespecified time intervals parallels our own approach.

The use of time-dependent covariates $z$ in the PH model was mentioned by Cox (1972). If the standard conditions or risks $z_0$ do not vary with time (e.g. if $z_0 = 0$) then $\lambda_0(t)$ still represents the corresponding reference hazard function, but comparison-invariance will not hold in general when the effect (or definition) of $z$ depends on time. Several authors have noted this possibility (for example Holt et al. (1974), Thompson (1977), Prentice et al (1978), Breslow (1978)), but we are aware of relatively few who have used it in applications. Those who have are: Cox (1972), who used a simple parametric form to illustrate a possible model of departure from the PH model; Kay (1977), (implicitly, by allowing different functions $\lambda_0(t)$ in different 'strata' of subjects); and Taulbee (1979), who characterised the hazard function by a polynomial in $t$ with coefficients depending on $z$. 
Analysis of Survival Data in Contingency Tables

Grizzle, Starmer and Koch (1969) - abbreviated as GSK - have described how weighted least squares analysis of linear regression models may be used to test hypotheses and fit simplified models to a multi-way table (such as we envisage above) of cell frequencies arising from the cross-classification of subjects by categorical variables. As stated by Koch et al (1972), linear and nonlinear functions of the cell proportions and a corresponding dispersion matrix may be estimated, and the resulting test statistics, which belong to the Minimum Modified Chi-Square (MMCS) class (Neyman, 1949) or equivalently to the quadratic form criterion of Wold (1943), have central $\chi^2$ distributions when the null hypotheses at which they are directed are true. Maximum Likelihood (ML) (Bishop (1969), Goodman (1970, 1971), Nelder (1977)) and Minimum Discrimination (MD) (Ku et al (1971)) approaches can also be used in this situation and together with the GSK method are based on BAN estimates (Koch et al (1972)) and hence asymptotically equivalent. In principle, therefore, the structural models for contingency tables of survival data put forward in this paper could be analysed by any of the GSK, ML or MD methods. Our preference for GSK over ML and MD is secondary (though not always unimportant), arising from the computational convenience of having explicit analytic formulae for the estimates. All data are taken to be (or are rendered) categorical in form, and the justification of the Normal theory $\chi^2$ statistics essentially arises from the asymptotic Normality of multinomially distributed cell frequencies when (roughly speaking) the average of these frequencies over the whole table is sufficiently large (Johnson et al (1971) suggest at least 25 observations per cell).

Illustrative applications of the GSK method to survival analysis with covariates which do not depend on time are described in Koch et al (1972) and Freeman et al (1974). Thus, in an analysis of 5-year survival rates
from breast cancer Koch et al., following Cutler and Myers (1967), considered
18 risk groups arising from 3 underlying factors (degree of skin fixation,
node status and tumour size) classified into 3, 2 and 3 levels respectively.
Two alternative saturated models involving 18 parameters were fitted to the
5-year survival rates from the 18 risk groups. Progressive simplification
by a process of exploratory hypothesis testing led to a 5-parameter model
(with a residual $\chi^2$ value of 2.76 on 13 degrees of freedom) which was used
as the basis of an empirical statistical classification of the extent of
disease in cancer of the breast. Adaptations of the GSK approach to deal
with withdrawals, losses to follow-up and period (as opposed to cohort)
studies are also discussed in the above two references and will not be
further considered here.
3. Categorical Data Analysis : Non-sequential Case

Basic Theory of the GSK method

The purpose of this section is to discuss the modelling of hazards \( \lambda(t|z) \) in which the covariates \( z \) are defined at time \( t = 0 \). For completeness of exposition we begin by repeating (with minor adaptations) elements of the basic GSK formulation as given in GSK (1969) and Koch et al (1972). For simplicity we consider the case of a follow-up study of a disease in which no losses to observation occur until the planned final time of follow-up (or censoring), \( t_k \), measured from each individual's onset of disease. We divide the time axis into \( k + 1 \) suitably defined intervals

\[
0 = t_0 < t < t_1, t_1 < t < t_2, \ldots, t_{k-1} < t < t_k, t > t_k.
\]

For subjects surviving a time \( t_k \) we note merely that \( t > t_k \), but otherwise the interval \( t_{j-1} < t \leq t_j \), or ideally the exact time\(^1\) of death is recorded. Since the concomitant data are categorical, we have a strictly limited number of possible different vectors \( z \) which can be indexed a priori as \( z_i \), \( i = 1, \ldots, N \). Subjects with the same categorical \( z \) value are treated as a homogeneous risk group, and it is assumed that deaths due to causes unrelated to the disease of interest are excluded. Then subjects may be classified in an \( N \times (k + 1) \) contingency table as envisaged in Section 2. For \( i = 1, 2, \ldots, N \), and \( j = 1, 2, \ldots, k + 1 \) the cell frequency \( n_{ij} \) denotes the number of deaths in the \( j \)th interval \( t_{j-1} < t < t_j \) (or \( t > t_k \), if \( j = k + 1 \)) occurring in the \( i \)th risk group (which is of total size \( n_i \), equal to the \( i \)th row total), \( n_{ij} \) denotes the probability that a member of this group dies in the \( j \)th time interval, and

\(^1\) For a given choice of \( k, t_1, \ldots, t_{k-1} \), the precise times of death are irrelevant for the GSK approach because we are concerned only with the chance of death in each interval, which is estimated by multinomial maximum likelihood in terms of the interval (or cell) frequencies only. However, knowledge of the exact times permits one to reanalyse with different numbers and choices of intervals and compare their effects on the results.
$P_{ij}$ denotes the corresponding multinomial estimate given by $\frac{n_{ij}}{n_i}$. Writing

$$F(t|\xi) = F_i(t)$$

for brevity, we see that, for all $i$ and $j$,

$$F_i(t_j) = 1 - \sum_{l=1}^{j} \pi_{il} = \sum_{l=j+1}^{k+1} \pi_{il},$$

with corresponding multinomial estimate

$$F_i(t_j) = 1 - \sum_{l=1}^{j} p_{il} = \sum_{l=j+1}^{k+1} p_{il}.$$ 

We now form vectors $\pi_i$ and $P_i$, where $\pi_i' = (\pi_{i1}, \ldots, \pi_{i(k+1)})$ and $P_i' = (p_{i1}, \ldots, p_i(k+1))$, and composite vectors $\pi$ and $P$ given by

$$\pi' = (\pi_1', \pi_2', \ldots, \pi_N')$$

$$P' = (P_1', P_2', \ldots, P_N').$$

Then the GSK method can in principle be applied to fit models of the form

$$E_{u1}(\pi) = X_{uv} \beta_{v1},$$

(3.1)

where $E$ is a vector of $u$ components which are twice differentiable functions of the probabilities $\pi_{ij}$ (and $u \leq N_k$, the total degrees of freedom of the data layout $\{n_{ij}\}$), $X$ is a prescribed design matrix of full rank $v$, where $v \leq u$, and $\beta$ is a vector of $v$ unknown parameters. Current versions of the GSK computer program handle functions $E$ which can be represented as a succession of linear, logarithmic or exponential operations on the vector $\pi$, and these possibilities have facilitated many practical applications.

We further define

$$V(\pi_i)_{(k+1) \times (k+1)} = \frac{1}{n_i}$$

$$= \begin{bmatrix}
\pi_{i1}(1-\pi_{i1}) & -\pi_{i1}\pi_{i2} & \cdots & -\pi_{i1}\pi_{i(k+1)} \\
-\pi_{i2}\pi_{i1} & \pi_{i2}(1-\pi_{i2}) & \cdots & -\pi_{i2}\pi_{i(k+1)} \\
\vdots & \vdots & \ddots & \vdots \\
-\pi_{i(k+1)}\pi_{i1} & -\pi_{i(k+1)}\pi_{i2} & \cdots & \pi_{i(k+1)}(1-\pi_{i(k+1)})
\end{bmatrix}.$$
V(\hat{v}_i) = \text{sample estimate of } V(\pi_i); 

V(p) = \text{block diagonal matrix having } V(\hat{v}_i) \text{ on the main diagonal;}

f_m(\pi) = \text{any twice-differentiable scalar function of the elements } \pi_{ij} \text{ of } \pi, \ m = 1, 2, \ldots, u < Nk;

f_m(p) = f_m(\bar{\pi}) \text{ evaluated at } p;

F' = (F(p))' = (f_1(p), f_2(p), \ldots, f_u(p));

\begin{bmatrix}
H \cdot \left[(k+1)N\right] = \\
\begin{bmatrix}
\frac{\partial f_m(\pi)}{\partial \pi_{ij}} & \pi_{ij} = p_{ij}
\end{bmatrix}
\end{bmatrix};

S_{uxu} = HV(p)H'.

The matrix S is the \textit{sample estimate} of the covariance matrix of F, exactly when the components f_m are linear in the p_{ij}'s and asymptotically as given by the 'delta' method otherwise. We assume that the components f_m are functionally independent and independent of the constraints

\[\sum_{j=1}^{k+1} \pi_{ij} = 1, \ i=1, \ldots, N; \quad (3.2)\]

then H and S are of full rank u. If difficulties arise from occasional zero cells, such zeros are replaced by 1/(k+1).

The \textit{weighted least squares estimate} of \hat{\beta} in the model (3.1) is \hat{b} determined to minimise the quadratic form

\[(F - X\hat{b})' S^{-1} (F - X\hat{b}),\]

namely

\[\hat{b} = (X'S^{-1}X)^{-1}X'S^{-1}F.\]

The test statistic for the fit of the model (3.1) is

\[SS(F(\pi) = X\hat{\beta}) = F'S^{-1}F - \hat{b}'(X'S^{-1}X)\hat{b},\]

which is asymptotically \(\chi^2\)-distributed on u-v DF if (3.1) holds. Assuming (3.1),
the test of $H_0 : \mathbf{C} \mathbf{b} = 0$, where $\mathbf{C}$ is an arbitrary $d \times v$ matrix of full rank $d \leq v$, is based on

$$SS(\mathbf{C} \mathbf{b} = 0) = \mathbf{b}' \mathbf{C}' \left[ \mathbf{C}(\mathbf{X}' \mathbf{S}^{-1} \mathbf{X})^{-1} \mathbf{C}' \right]^{-1} \mathbf{C} \mathbf{b}$$

which is asymptotically $\chi^2$-distributed on $d$ DF if $H_0$ is true.

Application to the Proportional Hazards (PH) Model

As the genuine PH model implies, we begin by supposing that the parameter vector $\mathbf{B}$ representing the effects of the covariates is the same in all $k+1$ time intervals. Putting $h(z) = \mathbf{B}' \mathbf{z}$ in equation (2.3) and then substituting in (2.1) gives

$$\Phi(t | \mathbf{z}) = \exp \left\{ - \int_0^t \lambda_0(u) du \exp(\mathbf{B}' \mathbf{z}) \right\}.$$

Applying the double-log transformation, we obtain

$$\ln \left[ - \ln \Phi(t | \mathbf{z}) \right] = \ln \left[ \int_0^t \lambda_0(u) du \right] + \mathbf{B}' \mathbf{z}.$$

To model the function $\lambda_0(t)$ in the intervals $t_{j-1} < t \leq t_j$, $j=1, \ldots, k$, we now introduce a subvector of additional parameters $\mathbf{\phi}_j$, where $\mathbf{\phi}_j = (\phi_1, \ldots, \phi_k)$ and

$$\exp(\phi_j) = \int_0^{t_j} \lambda_0(u) du,$$

so that

$$\ln \left[ - \ln \Phi_i(t_j) \right] = \phi_j + \mathbf{B}' \mathbf{z}_i, \quad i=1, \ldots, N; j=1, \ldots, k. \tag{3.3}$$

The model (3.4) is of the form $\Phi(y) = \mathbf{X} \mathbf{p}$, where $\mathbf{p}' = (\phi', \mathbf{B}')$ is the transposed full vector of unknown parameters, and so can be analysed using the GSK method. Explicitly,

$$\Phi(y) = \ln \left[ - \ln (\mathbf{A} \mathbf{n}) \right] \tag{3.5}$$

where $\mathbf{A}$ is a $(Nk) \times [N(k+1)]$ block-diagonal matrix

$$\mathbf{A} = \begin{bmatrix} \mathbf{B} & 0 & \cdots & 0 \\ 0 & \mathbf{B} & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & \cdots & \mathbf{B} \end{bmatrix}$$

† A similar formulation of this equation has been presented independently by Dr. R. McCullagh in a recent paper to the Royal Statistical Society.
with blocks \( B \) given by

\[
B_{kx(k+1)} = \begin{bmatrix}
0 & 1 & 1 & \cdots & 1 \\
0 & 0 & 1 & \cdots & 1 \\
& & \ddots & \ddots & \ddots \\
0 & 0 & 0 & \cdots & 1
\end{bmatrix}
\]

and \( \ln \) applied to a vector argument means the logarithm of each component.

Furthermore, from writing

\[
\varphi_i(t_j) = \sum_{l=j+1}^{k+1} \pi_{il}
\]

we find that the matrix \( H \) of first partial derivatives of the components of (3.5) with respect to the \( \pi_{ij} \)'s is a block-diagonal matrix

\[
H_{(Nk) \times (N(k+1))} = \begin{bmatrix}
H_1 & 0 & \cdots & 0 \\
0 & H_2 & \cdots & \cdots \\
& & \ddots & \ddots & \cdots \\
0 & 0 & \cdots & H_N
\end{bmatrix}
\]

in which for \( i=1, \ldots, N \) the \( k \times (k+1) \) matrix \( H_i \) is given by

\[
H_i = G_i B
\]

where the \( k \times k \) matrix \( G_i \) is given by

\[
G_i = \text{diag} \left( \gamma_{i1}, \ldots, \gamma_{ik} \right)
\]

and

\[
\gamma_{ij} = \left[ \varphi_i(t_j) \ln \varphi_i(t_j) \right]^{-1}, \quad j=1, \ldots, k.
\]

Had we written \( \varphi_i(t_j) \) as \( 1 - \sum_{l=1}^{j} \pi_{il} \) the matrix \( H_i \) would have come out slightly differently. However, the resulting asymptotic covariance matrix \( S = HVH' \) is invariant under uses of the constraints (3.2). The general form of the design matrix \( X \) is

\[
X_{(Nk) \times (k+N-1)} = \begin{bmatrix}
I_k & l_k & z_1' \\
I_k & l_k & z_2' \\
& & \ddots & \ddots & \ddots \\
I_k & l_k & z_{N-1}' \\
I_k & l_k & z_N'
\end{bmatrix}
\]
where $I_k$ denotes the $k \times k$ identity matrix, $1_k$ denotes a vector of units and the $(N-1)$-dimensional vectors $\mathbf{z}_1, \ldots, \mathbf{z}_N$ indicate the covariate levels (or values) pertaining to the $N$ risk groups. We note that, given a parametrisation $\phi$ of the underlying hazard function $\lambda_o(t)$, only $N-1$ parameters (indicator variables, components of $\mathbf{z}$) are necessary to discriminate fully between the mortality experiences of the $N$ groups.

The modelling of $\lambda_o(t)$ expressed by (3.2) appears similar to, but is much less restrictive than, that of Kalbfleisch and Prentice (1973). Thus, if for $j = 1, \ldots, k$ we represent $\lambda_o(t)$ as

$$\lambda_o(t) = \lambda_j, \quad t_{j-1} < t \leq t_j,$$

then (3.2) implies that

$$\lambda_j = \frac{\phi_j - \phi_{j-1}}{t_j - t_{j-1}},$$

or

$$\phi_j = \ln \left[ \frac{j}{\sum_{l=1}^{j} \lambda_l (t_l - t_{l-1})} \right].$$

However, we do not require $\lambda_o(t)$ to be constant throughout each of a defined set of time intervals. Furthermore, specific distributional forms for $\lambda_o(t)$, which we may not have wished to assume a priori, translate into linear or nonlinear hypotheses about the subvector $\phi$. These hypotheses can be examined by graphical or other methods, and, if linear, can be tested within the GSK framework and built into the model if satisfactory. Some examples with details are listed in the Appendix.

**Extension to the Case of Time-dependent Covariate Effects**

It will be seen from the dimensions of $X$ as defined at (3.5) that (save when $k=1$) the number $(k+N-1)$ of parameters fitted in the above analysis of the PH model is less than the total number $Nk$ required to specify a saturated
model for the table of cell frequencies \( n_{ij} \). The difference represents the number of parameters which in the context of the data layout \( n_{ij} \) are available for modelling violations of the PH assumption. However, the matrix (3.5) need only be enlarged to the square matrix

\[
X^{(Nk) \times (Nk)} = \begin{bmatrix}
I_k & Z_1 \\
I_k & Z_2 \\
\vdots & \vdots \\
I_k & Z_N \\
\end{bmatrix}
\]  

(3.8)

where, for \( i = 1, 2, \ldots, N \), \( Z_i \) is a \( k \times (N-1)k \) block diagonal matrix with blocks \( Z_i' \), and the components of the parameter vector \( \psi \) are ordered as in

\[
\psi' = (\phi', \beta_1', \beta_2', \ldots, \beta_k')
\]  

(3.9)

in which for \( j = 1, \ldots, k \beta_j \) is a subvector of \( N-1 \) parameters which characterise the relative mortality experiences of the \( N \) risk groups in the \( j \)th time interval.

A model based on (3.8) necessarily fits the data layout perfectly and any reduced (unsaturated) linear model involving fewer than \( Nk \) parameters can then be tested.

**An Example**

In a given application, the specific reduced models of interest are likely to depend on the basic factors underlying the definition of the \( N \) risk groups. A natural situation, at least in the early stages of analysis, is when the risk groups correspond to cells in a cross-classification of the basic factors suitably categorised into small numbers of levels. Taking the simple case of two basic factors, suppose that age \((A)\) and relevant disease history \((H)\) are of interest, these being dichotomised as age less than 60 years \((A = 0)\) versus age 60 or over \((A = 1)\) and as presence \((H = 1)\) versus absence \((H = 0)\) of disease history. We further choose to work with four time intervals altogether: that is, \( k = 3 \) intervals up to the time of final follow-up. To avoid complications we ignore the possibility of unknown or
missing data. The N=\(2^2=4\) risk groups may be indexed by a vector \(z\), where
\[
z' = (z_1, z_2, z_3) \quad \text{and} \quad 
\]
\[
\begin{align*}
z_1 &= 1 \text{ when there is a history of the disease,} \\
&= 0 \text{ otherwise;} \\
z_2 &= 1 \text{ when the age is 60 or over} \\
&= 0 \text{ otherwise;} \\
z_3 &= z_1 z_2
\end{align*}
\]
\begin{equation}
\tag{3.10}
\end{equation}

We now display the vectors \(z_i\) corresponding to the different risk groups (which are arbitrarily numbered 1, 2, 3, 4 as shown):

Group 1 : \(H = A = 0\) : \(z_1' = (0, 0, 0)\)
Group 2 : \(H = 1, A=0\) : \(z_2' = (1, 0, 0)\)
Group 3 : \(H = 0, A=1\) : \(z_3' = (0, 1, 0)\)
Group 4 : \(H = A = 1\) : \(z_4' = (1, 1, 1)\)
\begin{equation}
\tag{3.11}
\end{equation}

Underlying this coding is a saturated (interval-specific) ANOVA-like parametrisation

\[
(\phi_j, \beta_{Hj}, \beta_{Aj}, \beta_{HAj}, j=1, 2, 3),
\]

in terms of which we have, for all (four) possible (values of) \(z\),

\[
\ln \left[ -\ln \mathcal{Q}(t_j | z) \right] = \phi_j + \beta_{Hj} z_1 + \beta_{Aj} z_2 + \beta_{HAj} z_3,
\]
or in matrix terms

\[
\begin{bmatrix}
\ln \left[ -\ln \mathcal{Q}(t_1 | z_1) \right] \\
\ln \left[ -\ln \mathcal{Q}(t_2 | z_1) \right] \\
\ln \left[ -\ln \mathcal{Q}(t_3 | z_1) \right] \\
\ln \left[ -\ln \mathcal{Q}(t_4 | z_1) \right] \\
\ln \left[ -\ln \mathcal{Q}(t_1 | z_2) \right] \\
\ln \left[ -\ln \mathcal{Q}(t_2 | z_2) \right] \\
\ln \left[ -\ln \mathcal{Q}(t_3 | z_2) \right] \\
\ln \left[ -\ln \mathcal{Q}(t_4 | z_2) \right] \\
\end{bmatrix}
= 
\begin{bmatrix}
1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\
1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\
0 & 0 & 1 & 0 & 0 & 0 & 0 & 1 & 0 \\
0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\
1 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 \\
0 & 0 & 1 & 0 & 0 & 0 & 1 & 1 & 1 \\
0 & 0 & 1 & 0 & 0 & 0 & 0 & 1 & 1 \\
\end{bmatrix}
\begin{bmatrix}
\phi_1 \\
\phi_2 \\
\phi_3 \\
\beta_{H1} \\
\beta_{A1} \\
\beta_{HA1} \\
\beta_{H2} \\
\beta_{A2} \\
\beta_{HA2} \\
\beta_{H3} \\
\beta_{A3} \\
\beta_{HA3} \\
\end{bmatrix},
\]
\begin{equation}
\tag{3.12}
\end{equation}
in which the parameter vector conforms to the ordering specified for \( \Phi \) in (3.9).

In this illustration, Group 1 (\( H=A=0 \)), for which \( z_1 = 0 \), is the standard or reference group which has hazard function \( \lambda_0(t) \) parameterised by \( \phi_1, \phi_2 \) and \( \phi_3 \). The group having the least unfavourable concomitant vector \( z \) is a natural candidate for this role. Some possible hypotheses of interest in this example are the following:

\[
H_{oa} : \beta_{H1} = \beta_{H2} = \beta_{H3}, \quad \beta_{A1} = \beta_{A2} = \beta_{A3}, \quad \beta_{HA1} = \beta_{HA2} = \beta_{HA3},
\]

corresponding to hazards proportional across all risk groups;

\[
H_{ob} : \beta_{A1} = \beta_{A2} = \beta_{A3}, \quad \beta_{HA1} = \beta_{HA2} = \beta_{HA3},
\]

corresponding to proportional hazards across levels of \( A \) for fixed \( H \);

\[
H_{oc} : \beta_{A1} = \beta_{HA1} = 0,
\]

corresponding to no effect of \( A \) in the first interval; and

\[
H_{od} : \beta_{HA1} = \beta_{HA2} = \beta_{HA3} = 0,
\]

corresponding to additivity of \( H \) and \( A \) effects on the double-log scale.

Alternatively, when patterns of significant interactions occur in the ANOVA version, a clearer picture may be obtained from a hierarchical representation based on

\[
\begin{align*}
z_1 & \text{ as defined in (3.10); } \\
z_2 & = 1 \text{ when age is 60 or over and } z_1 = 0 \text{ (and } z_2 = 0 \text{ otherwise); } \\
z_3 & = 1 \text{ when age is 60 or over and } z_1 = 1 \text{ (and } z_3 = 0 \text{ otherwise).}
\end{align*}
\]

Underlying this coding is a hierarchical parametrisation

\[
(\phi_j, \beta_{Hj}, \beta(A \text{ in } H_0)^j, \beta(A \text{ in } H_1)^j, \ j = 1, 2, 3),
\]

where 'A in H' refers to the risk group of subjects with \( A = 1 \) within the group with \( H = 0 \), and so on. On this basis the risk vector for Group 4 (\( H=A=1 \)) is now given by

\[
z_4' = (1, 0, 1), \quad (3.14)
\]
but $x_1$, $x_2$ and $x_3$ remain as defined at (3.11). Correspondingly,

$$\ln\left[-\ln \Phi(t_j | x)\right] = \phi_j + \beta_{Hj} x_1 + \beta(A \text{ in } H_0) x_2 + \beta(A \text{ in } H_1) x_3,$$

$$j = 1, 2, 3,$$

or in matrix terms (writing $\Phi_1(t_j)$ for $\Phi(t_j | x_i)$)

$$\begin{bmatrix}
\ln[-\ln \Phi_1(t_1)] \\
\ln[-\ln \Phi_1(t_2)] \\
\ln[-\ln \Phi_1(t_3)] \\
\ln[-\ln \Phi_2(t_1)] \\
\ln[-\ln \Phi_2(t_2)] \\
\ln[-\ln \Phi_2(t_3)] \\
\ln[-\ln \Phi_3(t_1)] \\
\ln[-\ln \Phi_3(t_2)] \\
\ln[-\ln \Phi_3(t_3)] \\
\ln[-\ln \Phi_4(t_1)] \\
\ln[-\ln \Phi_4(t_2)] \\
\ln[-\ln \Phi_4(t_3)]
\end{bmatrix}
= \begin{bmatrix}
1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\
1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\
1 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 \\
\end{bmatrix}
\begin{bmatrix}
\phi_1 \\
\phi_2 \\
\phi_3 \\
\phi_4 \\
\phi_5 \\
\phi_6 \\
\phi_7 \\
\phi_8 \\
\phi_9 \\
\phi_{10} \\
\phi_{11} \\
\phi_{12} \\
\phi_{13} \\
\phi_{14}
\end{bmatrix}$$

(3.15)

Again the parameter vector is ordered as in (3.9).

The model (3.15), which like (3.12) is saturated and hence fully general, may be deemed more convenient for testing hypotheses such as

$$H_{oe} : \beta(A \text{ in } H_0)1 = \beta(A \text{ in } H_1)1 = \beta(A \text{ in } H_2)1 = 0$$

which implies no effect of $A$ in the first interval at level 0 of $H$ and no effect of $A$ in the first two intervals at level 1 of $H$. In the ANOVA-like formulation $H_{oe}$ becomes

$$H_{oe} : \beta_{A1} = \beta_{HA1} = 0, \beta_{A2} + \beta_{HA2} = 0.$$  

If we do not wish to work with an ANOVA-type or hierarchical design matrix, a simple unstructured alternative is to use the $(Nk) \times (Nk)$ identity which associates one parameter with each risk-group by interval (1,...,k) combination.

For example, systematic equalities of mortality rates can easily be explored with this representation.
4. **Categorical Data Analysis: The Sequential Case**

**Generalities**

In this section we extend our approach to deal with survival data when additional concomitant information accrues at times after $t = 0$. To an initial set of factors such as age and relevant medical history which are defined for all the subjects under study at time $t = 0$, new factors may be obtained from clinical assessments or follow-up reports. When the new data arise at prescribed times in the disease episode it would seem natural to divide the time axis at these points. Then at each time $t_i$ more risk groups can be identified by cross-classifying by the newly observed factors those already defined. On this basis the number of risk groups is multiplied by the number of levels of each new factor introduced. The problems arising when several factors are used to classify a relatively small set of data would call for further consideration in any given application: whilst their obvious practical importance is not denied, they are not, however a major concern of this paper. In general, similar difficulties pertain to the discussion of survey data involving a large number of potential explanatory factors, and require the exercise of statistical substantive judgement in the context of whatever methods of analysis are adopted. Here we consider first a reformulation of the non-sequential case which allows for the subdivision of risk groups at different points in time. The general representation of the sequential case is then obtained and illustrated by an example.

**Conditional Binomial Formulation of the Non-sequential Case**

The work of this paragraph is based on the fact that the multinomial probability of observing death frequencies $n_{i1}, n_{i2}, \ldots, n_{ik}$ in the $i^{th}$ risk group can also be viewed as a product of binomial probabilities in a
conditional sequence. Thus, in the first interval $n_{i1}$ is a realisation of a binomial distribution with parameters $n_{i1}^* = n_i$ and $\pi_{i1}^* = \pi_{i1}$ (we use the notation $n_{i1} \sim B(n_i, \pi_{i1}^*)$). Subsequently, that is for $j = 2, 3, \ldots, k$, given that

$$n_{ij}^* = n_i \times \frac{j-1}{\sum_{l=1}^{j-1} n_{il} - n_i \times \left(1 - \sum_{l=1}^{j-1} \pi_{il}\right)} \tag{4.1}$$

subjects survive to enter the $j^{th}$ interval, $n_{ij}^* \sim B(n_{ij}^*, \pi_{ij}^*)$, where

$$\pi_{ij}^* = \frac{n_{ij}^*}{1 - \sum_{l=1}^{j-1} \pi_{il}} \text{ (and } 1 - \pi_{ij}^* = \frac{1 - \sum_{l=1}^{j-1} \pi_{il}}{1 - \sum_{l=1}^{j-1} \pi_{il}}) \tag{4.2}$$

Corresponding sample estimates $\hat{p}_{ij}$ are given by writing $\hat{p}_{ij} = \frac{n_{ij}}{n_i}$ for $\pi_{il}$, $l = 1, \ldots, k$, in (4.2). (In fact $p_{ij}^* = p_{il}$ and $p_{ij}^* = \frac{n_{ij}}{n_i}$, $j = 2, 3, \ldots, k$).

On this basis the data for the $i^{th}$ risk group are now laid out in a $k \times 2$ contingency table, the $j^{th}$ row of which refers to the interval $t_{j-1} < t \leq t_j$ in which $n_{ij}^*$ subjects are represented as dying or surviving with respective probabilities $\pi_{ij}^*$ and $1 - \pi_{ij}^*$. Accordingly, the model (3.5) will now be written in terms of vectors $\pi_i^*$ and $\hat{p}_i$, $\pi^*$ and $\hat{p}$, where

$$(\pi_i^*)' = (\pi_{i1}^*, \pi_{i2}^*, \ldots, \pi_{ik}^*), i = 1, 2, \ldots, N,$$

$$(\pi^*)' = [ (\pi_1^*)', (\pi_2^*)', \ldots, (\pi_N^*)' ],$$

and $\hat{p}_i$ and $\hat{p}$ are analogously defined. We obtain the representation

$$\hat{F}^* (\pi^*) = \text{kn} [- C^* \text{kn} (1 - \pi^*) ] \tag{4.3}$$

where $C^*$ is a $(Nk) \times (Nk)$ block-diagonal matrix.
\[
C^* = \begin{pmatrix}
B^* & 0 & \ldots & 0 \\
0 & B^* & \ldots & 0 \\
\vdots & \vdots & \ddots & \vdots \\
0 & 0 & \ldots & B^*
\end{pmatrix}
\]

with blocks \(B^*\) given by

\[
B^*_{k\times k} = \begin{pmatrix}
1 & 0 & 0 & \ldots & 0 \\
1 & 1 & 0 & \ldots & 0 \\
\vdots & \vdots & \iddots & \vdots & \vdots \\
1 & 1 & 1 & 1 & 1
\end{pmatrix}
\]

The appropriate 'covariance matrix', \(V^*\) say, is an \((Nk) \times (Nk)\) block diagonal matrix with blocks \(V^*_i, i=1,\ldots,N\) where \(V^*_i\) is a \(k \times k\) diagonal matrix with

with \((j,j)\) element \(\frac{\pi^*_{ij}(1-\pi^*_i)}{n^*_ij}\), in which, for \(j \geq 2, n^*_ij\) expresses the conditioning on the outcome of the preceding interval. The matrix \(H^*\) of first partial derivatives of the components of \((4.3)\) with respect to the \(\pi^*_ij\)'s is a block-diagonal matrix

\[
H^*_0(Nk) \times (Nk) = \begin{pmatrix}
H^*_1 & 0 & \ldots & 0 \\
0 & H^*_2 & \ldots & 0 \\
\vdots & \vdots & \ddots & \vdots \\
0 & 0 & \ldots & H^*_N
\end{pmatrix}
\]

in which for \(i = 1, 2, \ldots, N\) the \(k \times k\) matrix \(H^*_i\) is given by

\[
H^*_i = - G^*_i B^* P^*_i
\]

where the \(k \times k\) matrix \(G^*_i\) is given by

\[
G^*_i = \text{diag} (\gamma^*_{i1}, \ldots, \gamma^*_{ik})
\]

and

\[
(y^*_{ij})^{-1} = \frac{j}{\sum_{l=1}^j \ln (1 - \pi^*_il)} = \ln \left[ 1 - \frac{j}{\sum_{l=1}^j \pi^*_il} \right] \quad (4.4)
\]

and the \(k \times k\) matrix \(P^*_i\) is given by
\( (P^{*}_i)^{-1} = \text{diag} \left( 1 - \pi^*_{i1}, \ldots, 1 - \pi^*_{ik} \right) \).

\[
= \text{diag} \left( 1 - \pi^*_{i1}, \ldots, \frac{1 - \sum_{i=1}^{k-1} \pi^*_{ij}}{k-1} \right)
\]

It is not difficult to verify from these definitions that the corresponding sample covariance matrix of the \(\text{N} \times k\) random quantities \(\ln \left[ -\ln \varphi_i(t_j) \right]\), namely

\[ S^* = H^* V(p^*) (H^*)', \]

agrees exactly with the matrix \(S = H V(p) H'\) defined in Section 3.

Since risk groups are independent, it is enough to show agreement for the \(k \times k\) submatrices, \(S_R, S_R^*\) say, relating to a general risk group \(R\) with the identifying suffix (i hitherto) omitted. The general \((j, \ell)\) element of \(S_R\) is then obtained from (3.6) and following results in Section 3, and (4.4) as

\[
S_{j\ell} = \frac{c}{\Sigma p_m} \cdot \frac{1}{\ln \left( 1 - \Sigma p_m \right)} \cdot \frac{\gamma^*_{j\ell}}{\gamma^*_{m1}} \cdot \frac{\Sigma p_m}{\ln \left( 1 - \Sigma p_m \right)} \cdot \frac{1}{\ln \left( 1 - \Sigma p_m \right)}
\]

where \(c = \min (j, \ell)\); and that of \(S_R^*\) is given by

\[
S_{j\ell}^* = \gamma^*_{j\ell} c \cdot \frac{p_m^*}{\Sigma p_m} \cdot \frac{1}{n^* (1 - p_m^*)} \cdot \frac{\gamma^*_{m1}}{\gamma^*_{j\ell}} \cdot \frac{c}{\Sigma p_m} \cdot \frac{1}{\ln \left( 1 - \Sigma p_m \right)} \cdot \frac{1}{\ln \left( 1 - \Sigma p_m \right)}
\]

Equality is easily proved by induction on \(c\), being obvious for \(c = 1\) and resulting for general \(c\) from the fact that

\[
\frac{\Sigma p_m}{c+1} \cdot \frac{c}{1 - \Sigma p_m} = \frac{c}{c+1} \cdot \frac{1}{(1 - \Sigma p_m)(1 - \Sigma p_m)}
\]

It follows from the above results that the binomial reformulation applied to all \(N\) risk groups yields an \((Nk) \times 2\) table of data, analysis of which
terms of the 'starred' model (4.3) yields estimates and inferences precisely equivalent to those obtained for the multinomial model treated in Section 3.

The Model for the Sequential Case: Outline of General Construction

Our starting point for modelling data involving sequentially defined covariates is the (Nk) x 2 layout of binomial data categorised by the N risk groups distinguished at time t=0. For convenience, we begin by reordering the rows of this table so that they are grouped by time interval instead of by risk group. For consistency, the same permutation is then applied to the rows of \( F^* \) as given at (4.3) and to the rows of the design matrix X of Section 3. We now consider a general risk group defined in the \((j-1)^{th}\) interval, \(j = 2, 3, \ldots, k\), which is to be split on the basis of new data into two or more groups in the \(j^{th}\) interval. The numbers surviving from the original (or parent) group at \(t_{j-1}\) are divided into their new groups. In each new group the numbers dying in and surviving the \(j^{th}\) interval are represented as binomially distributed and the rows pertaining to these groups then replace the row corresponding to the original (undivided) group in the \(j^{th}\) interval. Similarly, the survival probabilities of these groups replace the original survival probability for the original group in the \(j^{th}\) interval, and corresponding rows identifying the mortality experiences of these groups up to and within the \(j^{th}\) interval are inserted into the matrix C in place of the row corresponding to the original group. The inserted rows will of course agree with the row being replaced in regard to all elements related to the common history of survival of the new groups through the \(1^{st}, 2^{nd}, \ldots, (j-1)^{th}\) intervals. By working successively through the \(2^{nd}, 3^{rd}, \ldots, k^{th}\) intervals, each time using the vector of survival probabilities (\(1 - \psi\)) and the matrix \(C^*\) just obtained, the initial \((Nk) \times 2\) table and the associated representation (4.3) may be expanded vertically as often as necessary to include risk groups discriminated sequentially at the times \(t_1, t_2, \ldots, t_{k-1}\). There is of course no point in
introducing new groups at the final follow-up time $t_k$ because, by definition, after this time no further mortality data arise by which such groups may be distinguished in our analysis.

Reordering and division of the original risk groups as described leads to an $(N_0 = N_1 + N_2 + \ldots + N_k) \times 2$ table, where, for $j = 1, 2, \ldots, k$, $N_j$ is the number of risk groups finally identified in the $j^{th}$ interval. The resulting table is 'stratified' by interval - that is, by time - in that the first $N_1$ rows of data pertain to the first interval, the next $N_2$ rows to the second, and so on. The accompanying representation in terms of the (vertically expanded) equation (4.3) may then be parametrised as $X\psi^*$, where in the usual initial case of a saturated model $X$ is a $N_0 \times N_0$ design matrix and $\psi^*$ is a corresponding vector of $N_0$ parameters, and analysed by means of the general GSK theory quoted in Section 3.

An Illustration

A full statement of the general layout of the sequential case, with arbitrary numbers and levels of factors each arbitrarily subdivided at each time point, is not given here on account of the excessively complicated notation required. Instead we give a simple illustration involving one initial covariate ($F^0$ say) defined at three levels (coded 0, 1 and 2) and two binary (0, 1) covariates $F^i$ defined respectively at times $t_i$, $i = 1, 2$. Incorporation of $F^1$ and $F^2$ in the model demands at least four time-intervals ($k \geq 3$): for simplicity we take $k=3$ and the time of final follow-up is then denoted by $t_3$. In the simple case when the risk groups in the 2nd and 3rd intervals are defined by exhaustive cross-classification, we then have $N_1=3$, $N_2=6$, $N_3=12$ and $N_0=3+6+12=21$. Accordingly, the data are laid out in a $21 \times 2$ table. To clarify matters, this table, together with the corresponding probability vector $\pi$ and its sample estimate $\hat{p}$, is displayed below. To be consistent with the earlier work of this section the n's, n's and p's should carry *'s but, again for the sake of clarity, the *'s have been omitted.
<table>
<thead>
<tr>
<th>Interval</th>
<th>Dying</th>
<th>Surviving</th>
</tr>
</thead>
<tbody>
<tr>
<td>1st</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>2nd</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>10</td>
<td>10</td>
</tr>
<tr>
<td></td>
<td>11</td>
<td>11</td>
</tr>
<tr>
<td></td>
<td>20</td>
<td>20</td>
</tr>
<tr>
<td></td>
<td>21</td>
<td>21</td>
</tr>
<tr>
<td>3rd</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>10</td>
<td>10</td>
</tr>
<tr>
<td></td>
<td>11</td>
<td>11</td>
</tr>
<tr>
<td></td>
<td>20</td>
<td>20</td>
</tr>
<tr>
<td></td>
<td>21</td>
<td>21</td>
</tr>
</tbody>
</table>

In the table, the left column (d) contains the numbers dying, and the right column (s) the numbers surviving, within the interval considered. Each
row of the table corresponds to a unique risk group which is identified by the superscripted levels of the hitherto defined covariates \((F^0\) in the 1st interval, \(F^0\) and \(F^1\) in the 2\(^{nd}\) and \(F^0\), \(F^1\) and \(F^2\) in the 3\(^{rd}\)). In the 1\(^{st}\) interval the numbers \(n_1^i, n_1^i \sim n_1^i\) surviving in each of the previously defined (original) risk groups are subdivided as \(n_1^{0}, n_1^{1}\) corresponding to the values 0 and 1 of the factor \(F^1\) which is defined on all surviving subjects at time \(t_1\). The numbers \(n_1^{0}, n_1^{1}, i = 0, 1, 2\) then comprise the 6 row totals pertaining to the 2\(^{nd}\) interval. The numbers surviving the 2\(^{nd}\) interval are similarly subdivided according to the value of \(F^2\) realised at time \(t_2\), giving rise to the 12 row totals for the 3\(^{rd}\) interval as shown. The double-log transformation of the survivor function is given by

\[
F(\pi) = \ln[-C \ln(1 - \pi)]
\]

(compare (4.3)) where \(C\) is a 21 x 21 'block triangular' matrix in which all the
omitted elements are zero. A saturated hierarchical model (for example), in which parameters refer specifically to the effects of newly defined covariates within each of the risk groups previously defined, is given by \( X\beta \) as shown below, zero elements of \( X \) being omitted.

\[
X\beta = \begin{bmatrix}
1 \\
1 & 1 \\
1 & 1 \\
1 & & & 1 \\
1 & & & & & & 1 \\
1 & & & & & & & & & & & 1 \\
\end{bmatrix}
\]

\[
\begin{bmatrix}
\phi_1 \\
\beta_1 \\
\phi_2 \\
\beta_2 \\
\beta_{01} \\
\phi_3 \\
\beta_3 \\
\beta_{011} \\
\beta_{101} \\
\beta_{0111} \\
\beta_{1111} \\
\end{bmatrix}
\]

In this representation \( \beta_j^i \) denotes the effect (on the double-log scale) in the \( j^{th} \) interval of \( F^0 \) at level \( i, j = 1, 2, 3; i = 1, 2 \). Similarly, \( \beta_{j}^{i1} \) denotes the effect in the \( j^{th} \) interval of \( F^1 \) at level \( 1 \) when \( F^0 \) is at level \( i, j = 2, 3; i = 0, 1, 2 \); and in the \( 3^{rd} \) interval \( \beta_{2}^{i1} \) denotes the effect of \( F^2 \) at level \( 1 \).
when \( F^0 \) and \( F^1 \) are at respective levels \( i \) and \( m, m = 0, 1; i = 0, 1, 2 \). The underlying structure may also be displayed in a tree diagram as shown on page 30. When estimates of parameters and of \( \ln[-\ln \Psi] \) (and of their standard errors) are inserted at appropriate points in the tree, the resulting diagram is often a valuable aid in the process of model reduction. The technique is of course also useful in the non-sequential case.

All saturated models are formally equivalent (although not necessarily equally useful as starting points for interpreting a given set of data), and many alternative design matrices \( X \) (for example, of the 'ANOVA type' mentioned in Section 3) are possible. However, even saturated models are sometimes not exhaustively cross-classified, for example when such is logically precluded by the definitions of the covariates. Suppose that \( F^1 = 1 \) cannot arise when \( F^0 = 0 \) and that \( F^2 = 1 \) cannot arise when \( F^1 = 0 \). Then the corresponding rows are removed from the tableau (4.5) and the 6 categories asterisked in the tree are deleted, along with the specifically related parameters \( \beta_{01}, \beta_{011}, \beta_{011}, \beta_{101}, \beta_{101} \) and \( \beta_{301} \) and the corresponding \( (5^{th}, 11^{th}, 12^{th}, 13^{th}, 15^{th} \) and \( 19^{th} \) \) rows and columns of the matrices \( C \) and \( X \) defined at (4.6) and (4.7). The result is a saturated model appropriate to the non-exhaustive classification, and this can be investigated in the usual way.
5. **Discussion**

In this paper we have described an approach to the analysis of survival data with arbitrarily time-related covariate information. If, as seems desirable, we wish to avoid strong implicit or explicit prior assumptions of structure underlying the covariate effects of their dependence on time, then a natural step is to categorise the data in the form of multidimensional contingency tables. In Sections 3 and 4 we have developed a formulation in which arbitrarily general linear models may be fitted to such tables, with or without transformations of the cell frequencies. Proportional hazards models and various known survival time distributions are accessible as special cases of this representation (see Section 3 and the Appendix) and sequentially realised concomitant data can be handled as discussed in Section 4. Primarily for the convenience of having explicit analytic formulae for the estimates, we have worked in terms of the GSK (Grizzle, Starmer and Koch) methodology for modelling and testing, based on minimum modified $\chi^2$ (MMCS) statistics. As noted in Section 2, however, maximum likelihood (ML) and other BAN asymptotically equivalent methods can in principle be applied to fit models within our formulation, and with equal generality to that of GSK. Elsewhere (Read et al. (1980)) we describe a practical application - a clinical trial of randomised duration of bedrest in the treatment of males suffering from an acute myocardial infarction - and compare results obtained by MMCS and ML methods. In that paper we also review and discuss modelling strategy.
References


References Cont.


Wald, A. (1943). Test of statistical hypotheses concerning several parameters when the number of observations is large. *Transactions of the American Mathematical Society* 54, pp. 426-489.
Appendix: Some Parametric Models for the Reference Hazard Function $\lambda_o(t)$

A.1 Generalities

The purpose of this appendix is to present some specific models for the standard or reference hazard function $\lambda_o(t)$ (or the corresponding survivor function $\varphi_o(t)$) which may serve as useful reductions of saturated linear models fitted to the double-log transformation of the survivor function as in Section 3. We are thus considering non-sequential situations in which, although the effects of the covariates may vary arbitrarily in successive time intervals, the definition of the concomitant vector $z$ is invariant through time. Since any saturated model fits the data layout perfectly, there is in one sense no need to consider transformations other than the double-log form, or models which are not linear in all the parameters. However, in practice interest invariably lies in simplified models with substantially reduced numbers of parameters, and in any particular case there is no reason why the double log (rather than some other) transform of $\varphi$ should necessarily have the simplest or most economical representation in terms of linear (rather than non-linear) models. In the course of our account, other transforms and non-linear models will therefore arise.

As in Section 3 we work with a total parameter vector $\psi$, where $\psi^t = (\psi^t, \beta^t)$, in which the subvectors $\psi^t$ and $\beta^t$ respectively parametrise $\lambda_o(t)$ and the effects of the covariates. Here we consider replacing the $k$-vector $\psi^t$ by a smaller subvector $\phi$ of parameters identifying members of a relatively simple family of distributions. In the current (not necessarily saturated) linear model fitted, say

$$ F(p) = X\psi, $$

we may partition $X$ as $[X_\phi, X_\beta]$ where $X_\phi$ consists of the first $k$ columns of $X$ and $X_\beta$ stands for the remaining columns. Then

$$ F(p) = X_\phi \phi + X_\beta \beta $$

(A.1)
Before we deal with specific cases, two general approaches are worth noting. The first, which can be applied within the double-log formulation, consists in replacing $X_{\psi}$ in (A.1) by a vector function $g(t|\theta)$, where $t' = (t_1, \ldots, t_k)$, which normally has some explicit parametric dependence on time expressed through the reduced subvector $\theta$. Then if $\theta$ is involved nonlinearly, through some function $g(t|\theta)$, where $t' = (t_1, \ldots, t_k)$, weighted least squares analysis implies minimising with respect to the components of $\psi$ a function of the general form

$$Q = \left[ F(p) - g(t|\theta) - X_{\beta} \right]' S^{-1} \left[ F(p) - g(t|\theta) - X_{\beta} \right].$$  (A.2)

Although computationally unattractive, the nonlinear minimisation of $Q$ may be justified by the performance of the resulting model. No essential difficulties of principle are raised since, if the form of model expressed in (A.1) is correct, $Q$ is still asymptotically $\chi^2$-distributed. Furthermore, if $g(t|\theta)$ agrees closely with the linear form $X_{\phi}$ which it replaces, it is likely that $\hat{\theta}$ will have been little changed by this substitution. At least as a first step, estimating $\hat{\theta}$ conditional on the original (linear model) estimate of $\hat{\theta}$ may reduce the computational burden considerably.

Secondly, a linear models formulation which embraces a number of cases of interest is that in which $\mathcal{F}$ is a conveniently invertible function, $G$ say, of a function, $h(t, z|\psi)$ say, which is linear in the components of the vector $\psi$ of unknown parameters. We may then write

$$\mathcal{F}(t|z) = G(h),$$

where

$$G^{-1}(\mathcal{F}) = h(t, z|\psi) = \sum_{l=1}^{L} h_l(t, z) \psi_l$$  (A.3)

where $L$ is the dimensionality of $\psi$ and $h_1, \ldots, h_L$ are known, prespecified functions of $t$ and $z$ which are such that within some region of parameter space
\[ \varphi \text{ is a legitimate survivor function. The corresponding hazard is} \]

\[ \lambda(t | z) = \frac{\partial}{\partial t} \left[-\ln \varphi(t | z)\right] \]

A subsystem of (A.3) which preserves the linear combination \( \beta'z \) intact is given by

\[ h(t, z | \psi) = g(t | \theta) + \tau(t) \beta'z, \tag{A.4} \]

where \( \tau(t) \) is a known prespecified function of \( t \), \( g \) is linear in the components of the subvector \( \theta \) and \( \psi \) is given by \( \psi' = (\theta', \beta') \). (A.3) and (A.4) are not CI in general, although the separable product case,

\[ \tau(t)(\theta + \beta'z), \tag{A.5} \]

obtained by restricting \( \theta \) to be a scalar and putting \( g(t | \theta) = 0 \), \( \tau(t) \), clearly is CI (assuming that covariate effects subsumed in \( \beta \) do not vary between intervals). The original double-log system is seen to fall within the scope of (A.4) if we take \( g^{-1}(\varphi) = -\ln(-\ln \varphi) \), replace \( \theta \) by \( \hat{\theta} \) as defined at (3.3), take \( \tau(t) \equiv 1 \) and define \( g(t | \hat{\theta}) \) as \( \phi_j \) when \( t_{j-1} < t < t_j \), \( j = 1, \ldots, k \).

### A.2 Analyses for Particular Families of Distributions

Our starting point is the \( N \times (k+1) \) data layout of Section 3, the reference or standard hazard function \( \lambda_0(t) \) being parametrised by the subvector \( \hat{\theta} \) defined at (3.3). With proportional hazards, covariate effects are parametrised by the subvector \( \beta \) as at (3.4); otherwise, when hazards are not proportional, \( \beta \) is composed of further subvectors \( \beta_1, \ldots, \beta_k \) representing covariate effects in the time interval \( t_{j-1} < t < t_j \), \( j = 1, \ldots, k \), as at (3.9). In either case it may be of interest to elicit from the vector estimate \( \hat{\theta} \), and then test, a simple parametric form for \( \lambda_0(t) \) and hence for the corresponding standard survivor function \( \varphi_0(t) \). When proportional hazards do not apply, it may also be useful to look for similar simple time
dependences relating covariate effects in different time intervals

As need hardly be said, the list which follows is illustrative, not exhaustive. Not unnaturally, our choice of examples has been conditioned by mathematical convenience.

(i) The Weibull Distribution

For the two parameter Weibull distribution we have

\[ \Phi_0(t) = \exp[-(\rho t)^\alpha], \quad t > 0, \quad \alpha > 0, \quad \rho > 0, \]

so that

\[ \ln[-\ln \Phi_0(t)] = \alpha \ln \rho + \alpha \ln t, \]

which is of the form \( \theta_1 + \theta_2 \ln t \). If \( \Phi_0(t) \) is of Weibull form then a plot of the estimates \( \hat{\beta}_j \) against \( \ln t_j \) should be well fitted by a straight line. If proportional hazards apply, this could be formally tested by fitting the (still linear) reduced model expressed for \( i = 1, \ldots, N; \ j = 1, \ldots, k \) by

\[ \ln[-\ln \Phi_1(t,j)] = \theta_1 + \theta_2 \ln t_j + \beta'_i z_i; \quad (A.6) \]

otherwise \( \beta' \) in (A.6) should be replaced by \( \beta'_j \), the \( j \)th subvector of the extended \( \beta' \) defined in (3.9). The further specialisation \( \theta_2 = 1 \), corresponding to the exponential distribution, can be treated in the same way.

For the three parameter Weibull distribution, (A.5) becomes

\[ \ln[-\ln \Phi_1(t,j)] = \theta_1 + \theta_2 \ln(t_j - \theta_3) + \beta'_i z_i, \]

leading to a nonlinear problem of the form (A.1). \( \theta_2 = 1 \) corresponds to a displaced exponential distribution.

(ii) The Generalised Rayleigh Distribution

This distribution has a polynomial hazard function,

\[ \lambda_0(t) = \sum_{l=0}^{p} \theta_l t^l, \]

where \( \theta_0, \theta_1, \ldots, \theta_p \) are such that \( \lambda_0(t) \geq 0 \) for all \( t > 0 \). It follows that
dependences relating covariate effects in different time intervals

As need hardly be said, the list which follows is illustrative, not exhaustive. Not unnaturally, our choice of examples has been conditioned by mathematical convenience.

(i) The Weibull Distribution

For the two parameter Weibull distribution we have
\[ F_o(t) = \exp[-(\rho t)^\alpha], \quad t > 0, \alpha > 0, \rho > 0, \]
so that
\[ \ln[-\ln F_o(t)] = \alpha \ln \rho + \alpha \ln t, \]
which is of the form \( \theta_1 + \theta_2 \ln t \). If \( F_o(t) \) is of Weibull form then a plot of the estimates \( \hat{\theta}_j \) against \( \ln(t_j) \) should be well fitted by a straight line. If proportional hazards apply, this could be formally tested by fitting the (still linear) reduced model expressed for \( i = 1, \ldots, N; j = 1, \ldots, k \) by
\[ \ln[-\ln F_i(t_j)] = \theta_1 + \theta_2 \ln t_j + \beta' z_i; \quad (A.6) \]
otherwise \( \beta' \) in (A.6) should be replaced by \( \beta'_j \), the \( j^{th} \) subvector of the extended \( \beta' \) defined in (3.9). The further specialisation \( \theta_2 = 1 \), corresponding to the exponential distribution, can be treated in the same way.

For the three parameter Weibull distribution, (A.5) becomes
\[ \ln[-\ln F_i(t_j)] = \theta_1 + \theta_2 \ln(t_j - \theta_3) + \beta' z_i, \]
leading to a nonlinear problem of the form (A.1). \( \theta_2 = 1 \) corresponds to a displaced exponential distribution.

(ii) The Generalised Rayleigh Distribution

This distribution has a polynomial hazard function,
\[ \lambda_o(t) = \sum_{l=0}^{p} \theta_l t^l, \]
where \( \theta_0, \theta_1, \ldots, \theta_p \) are such that \( \lambda_o(t) \geq 0 \) for all \( t > 0 \). It follows that
\[ \hat{\mathcal{C}}_i(t_j) = \exp\left\{ -\frac{p}{\sum_{l=0}^{\ell} \frac{\theta_l t_j^{l+1}}{l+1}} \right\} \]

and that

\[ \exp(\phi_j) = \frac{p}{\sum_{l=0}^{\ell} \frac{\theta_l t_j^{l+1}}{l+1}}, \quad j=1, \ldots, k, \]

indicating an obvious graphical way of assessing this form of distribution (preferably with small \( p \)) in the light of estimates \( \hat{\theta}_1, \ldots, \hat{\theta}_k \). Within the double-log system this distribution gives a nonlinear model of the form (A.2).

Alternatively, a general system of linear models based on (A.2) and (A.3) can be conveniently investigated by means of a single log transformation, namely,

\[ \ln[\hat{\mathcal{C}}_i(t_j)] = \frac{p}{\sum_{l=0}^{\ell} \frac{\theta_l t_j^{l+1}}{l+1}} + \tau(t_j) \beta' z_i, \]

or

\[ \ln[\hat{\mathcal{C}}_i(t_j)] = \frac{p}{\sum_{l=0}^{\ell} \frac{\theta_l t_j^{l+1}}{l+1}} - \tau(t_j) \beta' z_i, \]

where \( \tau(0) = 0 \) and \( \tau(t) > 0 \) when \( t > 0 \);

by choosing \( p = k - 1 \) this can be made a saturated model of generality equal to that of the whole double-log system. A corresponding CI subsystem of reduced models can then be derived from (A.4) if required.

(iii) Logistic Systems

The standard logistic distribution function is \( (1+e^t)^{-1}, -\infty < t < \infty \).

This may be adapted so that time runs from 0 (the median of the standard distribution), giving the survivor function

\[ \mathcal{S}_0(t) = 2(1 + e^{\theta t})^{-1}, \quad t > 0, \ \theta > 0, \]

for which the appropriate logistic transformation is

\[ \ln \left[ \frac{2 - \mathcal{S}_0(t)}{\mathcal{S}_0(t)} \right] = \theta t \]  

(A.7)
and
\[ \lambda_0(t) = \frac{\theta e^{\theta t}}{1 + e^{\theta t}} \]

It follows that, in the usual notation of the double-log formulation,
\[ \ln \left[ \ln \mathcal{A}_1(t) \right] = \ln \left[ \ln \left( \frac{1 + e^{\theta t}}{2} \right) \right] + B'z_i, \]

which is a nonlinear model of the form (A.2). Alternatively, if we apply (A.3) a generalised linear model can be set up using the adapted logit transformation (A.7), namely,
\[ \ln \left[ \frac{2 - \mathcal{A}_1(t)}{\mathcal{A}_1(t)} \right] = \sum_{k=1}^{L} h_k(t, z) \psi_k, \]

one of many possible examples being
\[ \ln \left[ \frac{2 - \mathcal{A}_1(t)}{\mathcal{A}_1(t)} \right] = t_j(\theta + B'z_i). \]

It would be interesting to compare the results of using these models with those obtained by an early use of the multiple logistic function in the Framingham study of coronary heart disease reported by Truett et al (1967).

(iv) Inverse Power Laws

For the inverse power survivor function
\[ \mathcal{A}_0(t) = \left(1 + \frac{\theta_2}{\theta_1} t\right)^{-\theta_1}, \quad t > 0, \quad \theta_1 > 0, \quad \theta_2 > 0, \]

we have
\[ \frac{1}{\theta_1} \left[ \mathcal{A}_0(t) \right]^{-1} = \theta_1 + \theta_2 t \]

and
\[ \lambda_0(t) = \frac{\theta_1 \theta_2}{\theta_1 + \theta_2 t} \]
Since $\phi_j = \int_0^{t_j} \lambda_0(u) \, du$, we have

$$\exp \left[ \frac{1}{\theta_1} e^{\theta_1 t_j} \right] - 1 = \frac{\theta_2}{\theta_1} t_j, \quad j = 1, \ldots, k \quad (A.9)$$

If we know or can guess at or are prepared to try possible values for $\theta_1$,
(A.9) indicates an obvious graphical method for assessing this form of distribution;
or, for the reference group of subjects with $z = 0$, a linear models analysis
could be based on (A.8). From the double-log transformation we obtain the
nonlinear form

$$\ln[-\ln \Psi_i(t_j)] = \ln \theta_1 + \ln \ln(1 + \frac{\theta_2}{\theta_1} t_j) + \beta' z_i.$$ 

As in (iv) reparametrisation of $\theta$ with $\theta_1' = \ln \theta_1$ may be helpful.

As in (i), a displacement parameter can be incorporated if required.

If the index of the power law is known, say $k$, models of the form

$$\left[ \Psi_i(t_j) \right]^{-1/k} = 1 + \sum_{i=1}^{L} h_i(t_j, z_i) \psi_i,$$

based on (A.3) with $G(x) = x^{-k}$, represent a wide range of possibilities which
are still linear in the parameters.

(v) The Compertz Force of Mortality

In this model, which has been widely used in actuarial studies, we have

$$\lambda_0(t) = \theta_1 e^{\theta_2 t}, \quad t > 0, \theta_1 > 0, \theta_2 > 0$$

or

$$\Psi_0(t) = \exp \left[ - \frac{\theta_1}{\theta_2} (e^{\theta_2 t} - 1) \right]$$

Corresponding nonlinear models accessible via the double-log transformation
are of the form

$$\ln[-\ln \Psi_i(t_j)] = \ln \theta_1 + \ln \left[ \frac{e^{\theta_2 t_{j-1}}}{\theta_2} \right] + \beta' z_i.$$
Evidently we may work in terms of \( \theta_1' = \ln \theta_1 \) instead of \( \theta_1 \), and then only \( \theta_2 \) is involved nonlinearly.

The approach represented by (A.3) may be less helpful here, since the invertible transformation of \( \gamma \) which is required demands knowledge of the ratio \( \frac{\theta_1}{\theta_2} \).