NOT FOR QUOTATION
WITHOUT PERMISSION
OF THE AUTHOR

HETEROGENEITY IN COMPOSITE LINK MODELS

Camille Vanderhoeft

October 1986
WP-86-60

Working Papers are interim reports on work of the International Institute for Applied Systems Analysis and have received only limited review. Views or opinions expressed herein do not necessarily represent those of the Institute or of its National Member Organizations.

INTERNATIONAL INSTITUTE FOR APPLIED SYSTEMS ANALYSIS
A-2361 Laxenburg, Austria

FOREWORD

One of the activities of the Population Program at IIASA focuses on the analysis of social-economic and medical-demographic consequences of aging. The methodology of this research is based on the concept of heterogeneity that emphasizes the importance of differences among the individuals' susceptibilities to various forms of demographic transitions.

This paper discusses an approach to the data analysis in the presence of hidden heterogeneity. The author suggests the use of a modification of a quite general statistical model in order to take hidden heterogeneity into account. The results of the paper can be useful in many fields related to data analysis in the presence of latent variables.

The paper was produced during the Young Scientists Summer Program at IIASA in 1986.

Anatoli Yashin
Deputy Leader
Population Program

Acknowledgments

The author is grateful to the National Bank of Belgium for financial support and to IIASA and its staff of the Population Program for the opportunities provided during the summer program. Special thanks are due to Drs. Anatoli Yashin and Anatoli Mikhalski for helpful discussions on the model.

HETEROGENEITY IN COMPOSITE LINK MODELS

Camille Vanderhoeft *

1. INTRODUCTION

The purpose of this paper is to show how a quite general statistical model can be modified in order to take hidden heterogeneity into account. The author's knowledge about hidden heterogeneity finds its origins in the study of quasi-likelihood estimation techniques (Wedderburn,1974; McCullagh,1983). The methods discussed in this paper are extensions of models proposed by Williams (1982) and Breslow (1984), who introduce respectively extra-binomial variation in the binomial-logistic-linear and extra-poisson variation in the Poisson-log-linear model.

The methods of this paper reflect the author's experience with the statistical package GLIM (Baker and Nelder,1978): see the formulation of the basic model as a *composite link model* in Section 3 and the numerical example in Section 9. It is hoped that the models discussed can be used in the analysis of demographic processes, though we don't have numerical results so far. This is our main topic for further research.

This paper is rather an introduction to new estimation techniques in demographic analysis. The mathematical details are therefore not given; they may be published elsewhere.

* Camille Vanderhoeft, Interuniversity Programme in Demography, Vrije Universiteit, Pleinlaan 2, 1050 Brussel, Belgium.

2. THE DATA

Consider a sample of individuals, drawn from a population in which we want to study the occurence of some event(s), and assume that we know for each individual its age x, measured in completed units (e.g. months, years,...). The subsample of individuals aged x will be called *cohort* x . Suppose that the minimum age of individuals in the sample is x_0 and that the maximum age is x_1. We assume further that the data on the occurence of the event(s) of interest are in one of the following three forms.

a. Binomial data :

$$y_x, n_x, x \quad (x=x_0,...x_1)$$

where n_x is the total number of individuals in cohort x and y_x is the number of individuals in cohort x who have ever (i.e. by age x) experienced the event of interest. Binomial data are data on a non-renewable event.

b. Poisson data :

$$y_x, n_x, x \quad (x=x_0,...x_1)$$

where n_x is as before and y_x is the number of events of interest ever experienced by the individuals in cohort x. Poisson data are data on a renewable event.

c. Multinomial data :

$$y_{tx}, n_x, t, x \quad (x=x_0,...x_1;\ t=t_0,...x)$$

where t_0 is the minimum age (in completed units) at which an individual might experience the event of interest, n_x is as before and y_{tx} is the number of individuals in cohort x who have experienced the event of interest at age t. Multinomial data are data on a non-renewable event.

Clearly, if we also know for each individual a set of characteristics $\mathbf{X}=(X_1,...X_p)$, then we may consider *cohort-subgroups* $(x,\mathbf{X})$ and corresponding counts $n_{x\mathbf{X}}$, $y_{x\mathbf{X}}$, and $y_{tx\mathbf{X}}$. Whenever possible, covariates $\mathbf{X}$ are dropped from the notations for the seek of simplicity.

There may be several reasons for collection of the data in one of the above described ways. First, exact ages of individuals and exact ages at which events are experienced are rarely known. Ages are often given merely in completed units (e.g. months, years,...). It is therefore not unrealistic to consider grouped (aggregated) data. In other words, multinomial data often reflect all the information available. (Note that multinomial data are here only defined for non-renewable events, but we could consider some similar form for renewable events.) Second, the lack of accuracy in the data may be such that multinomial data are unreliable, while the corresponding binomial data - note that $y_x = \Sigma_{t \leq x} y_{tx}$ - are much better, although they are less detailed. For instance, the retrospective WFS data on breastfeeding duration are often inaccurate since women cannot remember the exact age at which they weaned their children. Breastfeeeding durations are therefore often given as multiples of 6 months. This is the well-known phenomenon of *heaping*. Binomial data on breastfeeding are thus more reliable than multinomial data if there is a strong effect of heaping.

3. MAXIMUM LIKELIHOOD ESTIMATION AND MODELLING

The usual procedure for analysing data as described in Section 2 is to construct a (log-) likelihood function and to maximize this with respect to the parameters involved. Thus, for the binomial data, for instance, one would assume that the counts y_x are, given n_x and π_x, independent binomially distributed, and the likelihood would be (proportional to)

$$L = \Pi_x \pi_x^{y_x}.(1-\pi_x)^{n_x-y_x}. \qquad (3.1)$$

Note that π_x is the probability that an individual in cohort x has experienced the event of interest. Unrestricted estimates of the probabilities π_x are the observed proportions y_x/n_x. But only rarely is the sequence $(y_x/n_x)_x$ increasing as it should be since $(\pi_x)_x$ is the discrete counterpart of the cumulative distribution function, F(.) say, which describes the occurence of the event by age. Moreover, unrestricted estimation does not provide estimates below the minimum age x_0 and the maximum age x_1. For those reasons, one may prefer to model the basic parameters π_x through some analytic formula.

The parametric model considered in this paper is as follows. Consider a non-repeatable event, let T be the random variable representing the age at which the event is experienced by an individual, and let F(.) be its cumulative distribution function. Then, we assume :

$$\Phi(F(t)/C)=\theta.[((t+b)/k)^{\delta}-1]/\delta+\beta \qquad (3.2)$$

where $0<C\leq1$, $\theta>0$, $k>0$, and where the *transformation* $\Phi(.)=\Phi(.;m_1,m_2)$ is defined by its inverse :

$$\Phi^{-1}(w)=\int_{-\infty}^{w}\frac{\Gamma(m_1+m_2)}{\Gamma(m_1)\Gamma(m_2)}\left(\frac{m_1}{m_2}\right)^{m_1}e^{m_1u}\left(1+\frac{m_1}{m_2}e^{u}\right)^{-(m_1+m_2)}du, \qquad (3.3)$$

where $\Gamma(.)$ is the gamma-function, $-\infty<w<\infty$, $m_1>0$ and $m_2>0$. Note that $C=\lim_{t\to\infty} F(t)$, so that we may call C the *ultimate proportion* (i.e. the proportion of individuals who have ever experienced or who will ever experience the event). Of course, if $C \neq 1$, then F(.) is improper, meaning that there are individuals who will never experience the event of interest (e.g. some women never marry, or do not resume menstruation,...). Further, we will call b the *shift*, k the *scale*, θ the *slope* and β the *intercept*. If the observed individuals are classified according to covariates **X**, then we will assume $C=C(\mathbf{X})$, $\theta = \theta(\mathbf{X})$, $\beta=\beta(\mathbf{X})$, $b=b(\mathbf{X})$ and $k=k(\mathbf{X})$, but the parameters m_1, m_2 and δ are assumed to be independent of covariates **X**. Obviously, if the model is used for a comparative study (i.e. if we wish to compare the schedules $F(.;\mathbf{X})$ across subgroups **X**), then the *parameters of interest* are C, θ, β, b, and k. The parameters m_1, m_2, δ may then be called *nuisance parameters*.

The model (3.2-3) can also be used to analyse the occurence of repeatable events. For instance, the event may be the birth of a child. Then C is the *total fertility rate*, and F(t) is the cumulative fertility rate up to age t. Thus, if repeatable events are to be analysed, then C is called the *total rate* and F(.) the *cumulative rate function*. Note that $C>0$ is now the only restriction on C.

If the model (3.2-3) is used in the analysis of binomial data, then we assume $\pi_x=F(x+1/2)$, substitute this in (3.1), and the problem is then to maximize $\mathbf{L}$ or $\log \mathbf{L}$ with respect to the parameters C, θ, β,... (some of which may be fixed a priori). Unfortunately, we are dealing here with a constrained maximization problem, since some of the parameters are constrained. In order to avoid this problem (partially), we use the reparametrization :

$$
\begin{aligned}
C(\mathbf{X}) &= 1/(1+\exp(-\eta_C(\mathbf{X})) && \text{for non-renewable events} \\
&= \exp(\eta_C(\mathbf{X})) && \text{for renewable events} \\
\theta(\mathbf{X}) &= \exp(\eta_\theta(\mathbf{X})) \\
b(\mathbf{X}) &= \eta_b(\mathbf{X}) \\
k(\mathbf{X}) &= \exp(-\eta_k(\mathbf{X})) \\
\beta(\mathbf{X}) &= \eta_\beta(\mathbf{X})
\end{aligned}
\qquad (3.4)
$$

where the $\eta(\mathbf{X})$ are functions which are linear in both the parameters and the covariates, that is $\eta(\mathbf{X})=\Sigma_j X_j . \eta_j$. The parameters η_j, which may be called the *linear parameters*, are essentially unconstrained.

The model (3.2-4) is a very flexible one. Indeed, a lot of special models used in various areas of research (e.g. demography, bio-assay, epidemiology, reliability,...) are covered by this general model. Such special models are obtained for special values of the parameters; for instance by considering the limit cases $m_1 \to 0$ or ∞, $m_2 \to 0$ or ∞ and/or $\delta \to 0$, and/or fixing some parameters of interest (e.g. $\theta(\mathbf{X})=1$ and $\beta(\mathbf{X})=0$ to obtain a class of *shifted-accelerated failure time models*). The more interested reader is refered to Prentice (1976), Kalbfleisch and Prentice (1980) and Vanderhoeft (1983). Two examples will be discussed in more detail at the end of this section.

Reparametrization to *linear predictors* $\eta(\mathbf{X})$ is not only useful to avoid a constrained maximization problem, but also for the following reasons. First, maximum likelihood theory states that the asymptotic distribution of the maximum likelihood estimators is normal. But since the normal distribution is the distribution of a random variable which takes negative as well as positive values, it may be expected that the asymptotic theory applies better to unconstrained parameters, providing, for instance, better confidence intervals, from which the confidence intervals for the constrained parameters are easily constructed. Second, the likelihood function may behave better as a function of the unconstrained parameters, improving convergence properties of the maximization procedure. Third, the model is an extension of the *generalized linear models* described by Nelder and Wedderburn (1972). In such models, the mean μ of an observation is related to a single linear predictor η through a *link function* g(.), where $\eta = g(\mu)$. The model (3.2-4) relates the mean μ of an observation to several linear predictors η_C, η_θ ,... . Following Thompson and Baker (1981) we may speak of *composite link functions*, and the model (3.2-4) is therefore refered to as a *composite link model* (CLM) in the rest of this paper. (Cox (1984) notes that the composite link function is in fact an *unlinking function*, since there is no longer a link or one-to-one relation between μ and the linear predictors. This has an important consequence if one looks for initial values of the parameters in the estimation procedure.) Fourth, linear predictors and parameters are also useful since they often simplify interpretation and discussion of the results, particularly in a comparative study.

We shall now close this section by considering briefly two special models, which will be refered to repeatedly in the next sections. First, we have the *Coale-Mcneil model*. This model is obtained from the general model (3.2-4) by assuming $m_1 \to \infty$, m_2=.604, δ=1, $\theta(\mathbf{X})$=1 and $\beta(\mathbf{X})$=0. Details about this model can be found in Coale and McNeil (1972), Rodriguez and Trussell (1980) and Vanderhoeft (1983). The Coale-McNeil model has successfully been used in the analysis of nuptiality, first birth and first union. This does not mean that the Coale-McNeil model is the only possible model for such analyses. Indeed, Trussell and Bloom (1981) have shown that a proportional hazards model can be a good alternative. Similarly, Vanderhoeft (1985) used a *shifted-proportional hazards model* for the analysis of first union. Those findings support the idea that other special models, obtained from (3.2-4), are useful alternatives to the Coale-McNeil model. Therefore, we will consider a class of shifted-proportional hazards models obtained from (3.2-4) by assuming that m_1=1, $m_2 \to \infty$, $\theta(\mathbf{X})$=1 and $k(\mathbf{X})$=1. Note that $\Phi(.)$ is then the complementary log-log (cloglog) transformation $\Phi(\pi)=\log(-\log(1-\pi))$. Note also that we obtain a parametric model, while Trussell and Bloom (1981) and Vanderhoeft (1985) considered semi-parametric proportional hazards models.

4. SOURCES OF HETEROGENEITY

Consider the binomial data of Section 2. If we want to fit the CLM (3.2-4) to such data, then, if the model involves s unknown parameters, one should have at least s+1 (consecutive) cohorts $x_0,\dots x_1$ say. (We ignore covariates for a while.) For instance, in order to estimate the parameters C, b and k in the Coale-McNeil model, one should have at least 4 cohorts. This implies that trends across the cohorts $x_0,\dots x_1$ cannot be estimated. But cohort trends may be an important source of heterogeneity in the data on the cohorts $x_0,\dots x_1$, which may yield bad or misleading results. Similar remarks can be made on Poisson and multinomial data. Thus, each kind of data has some limitations, which may lead to some kind of heterogeneity which is not incorporated in the basic model (3.2-4).

Often, covariates, if included in the types of data discussed in Section 2, are measured at the cohort age x, while it might be more realistic to measure them at the time the event is experienced; or even better: to construct covariates from measures at both age x and the time the event is experienced. If, however, not enough details are available on the covariates, then one cannot avoid this problem of *misclassification*. This is another source of heterogeneity which can lead to bad fits of the model.

Another problem might be considered in connection with categorical covariates. Indeed, it can for instance be difficult to classify individuals in one or another category of a particular covariate. One usually groups such categories which are hard to distinguish in order to avoid misclassification, but this is surely another source of hidden heterogeneity in the data.

Consider data on first union. There can be really an important difference in attitudes between individuals who enter first union through cohabitation and individuals who enter first union through marriage. Thus, in general, unrecorded *causes of failure* can be another source of heterogeneity.

It is thus likely that application of the CLM (3.2-4) for analysis of data as described in Section 2 yields bad fits - and misleading results - if there is an important degree of hidden heterogeneity present. Therefore, we will now develop a model which takes this heterogeneity into account.

5. QUASI-LIKELIHOOD ESTIMATION

This section will briefly deal with the concept of quasi-likelihood (QL). We shall only mention the main ideas of this very new statistical estimation method. The reader is refered to Wedderburn (1974) and McCullagh (1983) for more details.

In traditional maximum likelihood (ML) estimation one starts with the construction of the (log-)likelihood function. Therefore, it is necessary to know the complete distribution of the observations y. Often, one may have good reasons to assume one or another particular distribution. For the binomial data, for instance, it is often reasonable to assume that the y_x are (conditionally on the n_x) independent binomially distributed. But *often* does not mean *always* and so investigators have found examples of it. In most - if not all - of these cases it was found that the variance of y_x is not the variance which is implied by the binomial distribution. More specifically, investigation of the observed mean-variance relationship may indicate that the observations y_x are not binomially distributed. Examples are found in the papers mentionned in the beginning of this section. One may then try to find a distribution which has the observed mean-variance relationship, but this would be a very tedious problem. It is at this point that the QL method becomes a very powerful tool, since one does not need the complete distribution of y_x, but merely the mean-variance relationship. Indeed, suppose that μ_x is the mean of y_x and that $V(\mu_x)$ is the variance of y_x, where V(.) is a known function. Then, the QL function $l(y_x,\mu_x)$ is simply defined by the differential equation

$$\frac{\partial\, l(y_x,\mu_x)}{\partial \mu_x} = \frac{y_x-\mu_x}{V(\mu_x)} \qquad (5.1)$$

In order to obtain the QL estimates of the parameters $\gamma_1,\ldots\gamma_m$, where $\mu_x=\mu_x(\gamma_1,\ldots\gamma_m)$, it is not needed to know the likelihood $l(y_x,\mu_x)$ or to solve (5.1). The estimates are simply found by solving the equations

$$\frac{\partial}{\partial \gamma_j}\left(\Sigma_x\, l(y_x,\mu_x)\right) = 0 \qquad (j=1,\ldots m) \qquad (5.2)$$

or, applying the chaine rule and substitution of (5.1), by solving the equivalent equations

$$\Sigma_x \frac{y_x - \mu_x}{V(\mu_x)} \frac{\partial \mu_x}{\partial \gamma_j} = 0 \qquad (j=1,...m), \qquad (5.3)$$

where it is assumed that the observations y_x are independent.

In the next sections we shall introduce heterogeneity in our CLM (3.2-4) and we shall show that, for instance for the binomial data, the usual binomial mean-variance relationship does not hold. In order to get the appropriate mean-variance relationships it will be unnecessary to derive the full distribution of the observations y_x. Estimation of the parameters of the extended CLM will then use ideas of QL estimation , as we shall briefly discuss in Sections 6 and 7.

6. POISSON-NORMAL COMPOUND MODELS

In this section we will show that the Poisson-normal compound (PNC) model discussed by Hinde (1982) can be used to introduce heterogeneity in our CLM. It will also be shown how QL methods can be used to obtain estimates of the parameters of the model. In this section we also introduce the notations used in Section 7.

Hinde (1982) considered the following model for Poisson data:

$$Y|\mathbf{X},Z \sim \mathcal{P}(\mu) \qquad (6.1a)$$
$$\log \mu = \mathbf{X}.\mathbf{B}+Z \qquad (6.1b)$$
$$Z \sim \mathcal{N}(0, \sigma^2) \qquad (6.1c)$$

Thus, Y is conditionally on **X** and Z Poisson distributed with conditional mean μ. The conditional mean μ is related to the covariates **X** and the random variable Z through (6.1b). The random variable Z is normally distributed with mean zero and variance σ^2. **X** is a row vector of covariates $(X_1,...X_p)$ and $\mathbf{B}$ is a column vector of parameters: $\mathbf{B}^T=(B_1,...B_p)$. In order to relate this model to the Poisson data of Section 2, it should be noted that **X** includes the age variable x, and that Y has realizations $y_\mathbf{x}$. Note also that the mean and variance of the random variable Z do not depend on **X**. The conditional expectation μ can be written $E(Y|\mathbf{X},Z)$ - or $E(Y|Z)$ for brievity, since the **X** are fixed - and is itself thus a random variable which takes values $E(Y|Z=z)$. This value $E(Y|Z=z)$ can be interpreted as the mean of Y in a homogeneous population with *frailty* Z=z (and with characteristics **X**, which are dropped from the notations). The random variable Z represents hidden heterogeneity, since the realizations z are not observed.

In order to estimate the parameters $\mathbf{B}$ and σ^2 of the PNC model (6.1), Hinde used the EM algorithm. He showed that the iterative procedure is rather simple, but, nevertheless the method heavily relies on the assumed normal distribution of Z. In some circumstances we can avoid this disadvantage as follows. It is easy to show that the model (6.1) implies the unconditional mean and variance of Y :

$$E(Y) = \exp(\mathbf{X}.\mathbf{B}+\sigma^2/2) \qquad (6.2a)$$
$$\text{var}(Y) = E(Y)+[\exp(\sigma^2)-1].E(Y)^2 . \qquad (6.2b)$$

Now, if σ^2 is small, we have approximately :

$$E(Y) \approx \exp(\mathbf{X}.\mathbf{B}) = \mu_0 \quad (6.3a)$$
$$\text{var}(Y) \approx \mu_0 + \sigma^2.\mu_0^2. \quad (6.3b)$$

Note that μ_0 may be interpreted as the mean of Y in the homogeneous population with frailty Z=0. We see thus that the variance of Y is a function of the mean of Y (at least approximately), and that the mean only involves the parameters **B**. Hence, if σ^2 would be known, (approximate) QL methods could be used for estimation of **B**. If σ^2 is not known, one can construct an iterative procedure as follows :

(1) Choose an initial estimate σ_0^2 (e.g. $\sigma_0^2=0$).
(2) (Re)estimate the parameters **B**, using QL methods (which reduces to ML estimation if $\sigma_0^2=0$).
(3) Reestimate σ^2, replace σ_0^2 by this new estimate and return to step (2) (untill some convergence criterion is satisfied).

Thus, if σ^2 is small, we have the above alternative to the EM algorithm.

Extensions of the PNC model (6.1) are easily constructed. Indeed, (6.1b) can be rewritten as $\mu=\exp(\mathbf{X}.\mathbf{B}).\exp(Z)$. If we now replace the factor $\exp(\mathbf{X}.\mathbf{B})$ by $g^{-1}(\mathbf{X}.\mathbf{B})$, where g(.) corresponds to the link function in generalized linear models, then we can still show that

$$E(Y) = g^{-1}(\mathbf{X}.\mathbf{B}).\exp(\sigma^2/2) \quad (6.4a)$$
$$\text{var}(Y) = E(Y)+[\exp(\sigma^2)-1].E(Y)^2 \quad (6.4b)$$

or, if σ^2 is small :

$$E(Y) \approx g^{-1}(\mathbf{X}.\mathbf{B}) = \mu_0 \quad (6.5a)$$
$$\text{var}(Y) \approx \mu_0 + \sigma^2.\mu_0^2 \quad (6.5b)$$

The parameters **B** and σ^2 can still be estimated with the EM algorithm as discussed by Hinde (1982), or, if σ^2 is small, with the alternative method based on QL.

The PNC model can also be extended for composite link models. Let there be k linear predictors $\eta_j=\mathbf{X}_j.\mathbf{B}_j$, j=1,...k, and suppose that the mean of Y is $\mathcal{F}(\eta_1,... \eta_k)$ if there would be no hidden heterogeneity. Let $\mathcal{F}$ be the supermatrix (i.e. a block diagonal matrix)

$$\mathcal{X} = \begin{bmatrix} X_1 & 0 & & \\ 0 & X_2 & & \\ & & \cdot & \\ & & & X_k \end{bmatrix}$$

and $\mathcal{B}$ be the supervector

$$\mathcal{B} = \begin{bmatrix} B_1 \\ \cdot \\ \cdot \\ B_k \end{bmatrix},$$

so that

$$\mathcal{X}.\mathcal{B} = \begin{bmatrix} X_1.B_1 \\ \cdot \\ \cdot \\ X_k.B_k \end{bmatrix} = \begin{bmatrix} \eta_1 \\ \cdot \\ \cdot \\ \eta_k \end{bmatrix} = \boldsymbol{\eta}$$

and $\mathcal{F}(\eta_1,... \eta_k)=\mathcal{F}(\boldsymbol{\eta})$. The PNC model for composite link models (and for Poisson data) is thus:

$$Y|\mathcal{X}, Z \sim \mathcal{P}(\mu) \qquad (6.6a)$$
$$\mu = \mathcal{F}(\boldsymbol{\eta}).\exp(Z) \qquad (6.6b)$$
$$Z \sim \mathcal{N}(0, \sigma^2) \qquad (6.6c)$$

The parameters $\mathcal{B}$ and σ^2 can still be estimated with the EM algorithm as discussed by Hinde (1982). If σ^2 is small, we can also apply the alternative method based on QL and on the approximate relations

$$E(Y) \approx \mathcal{F}(\boldsymbol{\eta})=\mu_0 \qquad (6.7a)$$
$$var(Y) \approx \mu_0+\sigma^2.\mu_0^2. \qquad (6.7b)$$

As mentionned already, a disadvantage of Hinde's estimation procedure is that it is based on the assumption that Z is normally distributed. The alternative method (for small σ^2), however, does not use the distribution of Z, since relations like (6.7) can be shown to hold for any distribution of Z (with zero mean and (small) variance σ^2). Thus, it is sufficient to assume that

$$Y|\mathcal{X}, Z \sim \mathcal{P}(\mu) \qquad (6.8a)$$

$$\mu = \mathcal{F}(\eta).\exp(Z) \qquad (6.8b)$$

$$Z \sim (0, \sigma^2) \qquad (6.8c)$$

where (6.8c) means that Z has some distribution with mean zero and variance σ^2. The model (6.8) leads to the approximate relations (6.7), for small σ^2, and the method based on QL can be applied in order to estimate the parameters $\boldsymbol{\beta}$ and σ^2. The advantage of a model like (6.8) is that no assumption is made about the shape of the distribution of Z. (Note that we can, therefore, not derive the unconditional distribution of Y.) In the next section, we will consider other extensions which use the same idea.

7. CLM'S WITH MULTIDIMENSIONAL HETEROGENEITY

In the PNC model (6.1) of Hinde (1982), the heterogeneity variable Z may be considered as *acting linearly on the linear predictor* $X.\beta$. Alternatively, we may say that Z *acts multiplicatively on the mean of* Y. In the more general model (6.8) we can still say that Z acts multiplicatively on the mean of Y, but we cannot say that Z acts linearly on the linear predictor(s). The latter is only true if there is only one linear predictor (k=1) and if $\mathcal{F}(\eta)=\exp(\eta)$, i.e. under the condition (6.1b). In this section we present some models wherein heterogeneity acts linearly on each of the linear predictors. We therefore consider a vector-valued random variable $Z=(Z_1,\ldots Z_k)^T$, so that each Z_j will act linearly on the corresponding linear predictor $\eta_j=X_j.\beta_j$.

The models for binomial, Poisson and multinomial data discussed in Section 2 are as follows.

a. Binomial data

$$Y|\mathcal{X}, Z \sim \mathcal{B}(n, \mu) \qquad (7.1a)$$
$$\mu = n.\mathcal{F}(\eta+Z) \qquad (7.1b)$$
$$Z \sim (0, \Sigma) \qquad (7.1c)$$

where (7.1c) means that Z has some (multidimensional) distribution with mean vector 0 and dispersion matrix Σ. The following approximations can then be found:

$$E(Y) \approx n.\mathcal{F}(\eta) \qquad (7.2a)$$
$$\mathrm{var}(Y) \approx n.\mathcal{F}(\eta).(1-\mathcal{F}(\eta))+n(n-1)J^T.\Sigma.J \qquad (7.2b)$$

where J is the vector of partial derivatives $\mathcal{F}_j=\partial\mathcal{F}/\partial\eta_j$ evaluated in η :

$$J = \begin{bmatrix} \mathcal{F}_1(\eta) \\ . \\ . \\ \mathcal{F}_k(\eta) \end{bmatrix} .$$

Note that $\mathcal{F}(\eta+Z)=\mathcal{F}(X_1.\beta_1+Z_1,\ldots X_k.\beta_k+Z_k)$.

b. Poisson data

$$Y|\mathcal{X}, Z \sim \mathcal{P}(\mu) \qquad (7.3a)$$
$$\mu = \mathcal{F}(\boldsymbol{\eta}+Z) \qquad (7.3b)$$
$$Z \sim (0, \Sigma) \qquad (7.3c)$$

Then we can find the approximations

$$E(Y) \approx \mathcal{F}(\boldsymbol{\eta}) \qquad (7.4a)$$
$$var(Y) \approx \mathcal{F}(\boldsymbol{\eta})+J^T.\Sigma.J \qquad (7.4b)$$

where J is as for the binomial data.

c. Multinomial data

First, we introduce some notations. We consider a vector $\mathbf{Y}=(Y_1,...Y_t,...Y_l)^T$, where Y_t has realization y_t (covariates, including cohort x, are again dropped from the notations). Further, let $\boldsymbol{\mu}=(\mu_1,...\mu_t,...\mu_l)^T$ and $\mathbf{D(Y)}$ be the dispersion matrix of Y. The model is:

$$Y|\mathcal{X}, Z \sim \mathcal{M}(\mu,n) \qquad (7.5a)$$
$$\boldsymbol{\mu} = n.\mathcal{F}(\boldsymbol{\eta}+Z) \qquad (7.5b)$$
$$Z \sim (0, \Sigma) \qquad (7.5c)$$

where $\mathcal{F}$(.) is now the vector-valued function

$$\mathcal{F}(\boldsymbol{\eta}) = \begin{bmatrix} {}_1\mathcal{F}(\boldsymbol{\eta}) \\ . \\ . \\ {}_l\mathcal{F}(\boldsymbol{\eta}) \end{bmatrix}$$

and where (7.5a) means that Y is, conditionally on $\mathcal{X}$ and Z (and on $n=Y_1+...+Y_l$), multinomially distributed with mean-vector $\boldsymbol{\mu}$. We can then find the approximations

$$E(Y) \approx n.\mathcal{F}(\boldsymbol{\eta}) \qquad (7.6a)$$
$$D(Y) \approx n.V(\boldsymbol{\eta})+n(n-1)J^T.\Sigma.J \qquad (7.6b)$$

where J is the jacobian matrix evaluated in $\boldsymbol{\eta}$:

$$J = \begin{bmatrix} {}_1\mathcal{F}_1(\eta) & \cdots & {}_l\mathcal{F}_1(\eta) \\ \cdot & & \cdot \\ \cdot & & \cdot \\ {}_1\mathcal{F}_k(\eta) & \cdots & {}_l\mathcal{F}_k(\eta) \end{bmatrix}$$

where ${}_t\mathcal{F}_j$ is the j-th partial derivative of ${}_t\mathcal{F}$ (t=1,...l; j=1,...k), and where $V(\eta)$ is the matrix :

$$V(\eta) = \begin{bmatrix} {}_1\mathcal{F}(\eta).(1-{}_1\mathcal{F}(\eta)) & -{}_2\mathcal{F}(\eta).{}_1\mathcal{F}(\eta) & \cdots & -{}_l\mathcal{F}(\eta).{}_1\mathcal{F}(\eta) \\ -{}_1\mathcal{F}(\eta).{}_2\mathcal{F}(\eta) & {}_2\mathcal{F}(\eta).(1-{}_2\mathcal{F}(\eta)) & \cdots & -{}_l\mathcal{F}(\eta).{}_2\mathcal{F}(\eta) \\ \cdot & \cdot & & \cdot \\ -{}_1\mathcal{F}(\eta).{}_l\mathcal{F}(\eta) & -{}_2\mathcal{F}(\eta).{}_l\mathcal{F}(\eta) & \cdots & {}_l\mathcal{F}(\eta).(1-{}_l\mathcal{F}(\eta)) \end{bmatrix}.$$

From the general models (7.1), (7.3) and (7.5) one can obtain special models by assuming special forms of the dispersion matrix Σ. For instance, one might assume that Σ is a diagonal matrix, which means that the heterogeneity variables are not correlated. The most simple case is obtained when $\Sigma = \sigma^2.I$, where I is the identity matrix of order k. It is worth noting that the approximate expressions for the variance of Y (or the dispersion of Y) are then much simpler. For instance, in the binomial case we get :

$$\begin{aligned} var(Y) &\approx n.\mathcal{F}(\eta).(1-\mathcal{F}(\eta))+n(n-1)\sigma^2 J^T.J \\ &= n.\mathcal{F}(\eta).(1-\mathcal{F}(\eta))+n(n-1)\,\sigma^2(\Sigma_j\, \mathcal{F}_j(\eta)^2). \end{aligned}$$

Note also that under the condition $\Sigma = \sigma^2.I$ the models are equivalent to the model in which there is only one heterogeneity variable Z which acts, however, linearly on each linear predictor η_j.

The models which we have discussed here are generalizations of very special models discussed by Williams (1982) and Breslow (1984). In the same spirit as those authors constructed algorithms for estimation of the parameters B and Σ, we can consider the following procedures. The first algorithm assumes that the approximate variance of Y (or dispersion of Y) can be written as a function of the approximate mean. Then, if Σ would be known, we could estimate B by QL methods. If Σ is unknown, we propose the following iterative procedure :

(1) Choose an initial estimate $\mathcal{I}_0$ (e.g. $\mathcal{I}_0 = \mathcal{O}$).

(2) (Re)estimate the parameters $\mathcal{B}$, using QL methods (which reduces to ML estimation if $\mathcal{I}_0 = \mathcal{O}$).

(3) Reestimate $\mathcal{I}$, replace $\mathcal{I}_0$ by this new estimate and return to step (2) (untill some convergence criterion is satisfied).

This algorithm is clearly an extension of the algorithm discussed in Section 6. If, however, the approximate variance of Y (or dispersion of **Y**) cannot be written as a function of the approximate mean, then this algorithm is not applicable. We then proceed as follows. For the binomial case, for instance, we write

$$\mathrm{var}(Y) \approx n.\mathcal{F}(\eta).(1-\mathcal{F}(\eta))/w$$

where

$$w^{-1} = 1 + [n(n-1)J^T.\mathcal{I}.J]/[n.\mathcal{F}(\eta).(1-\mathcal{F}(\eta))],$$

or, more compact, but easier to deal with in generalizations :

$$\mathrm{var}(Y) \approx v/w,$$

where $v = n.\mathcal{F}(\eta).(1-\mathcal{F}(\eta))$ and w is as before. Note that if there is no heterogeneity, then w=1 and v is the unconditional variance of Y. Thus, we write the approximate variance of Y as the ratio of the variance of Y if there would be no heterogeneity, and weights w. Similar techniques can be applied for the Poisson and the multinomial cases : for the Poisson case we have $v = \mathcal{F}(\eta)$ and $w^{-1} = 1 + [J^T.\mathcal{I}.J]/\mathcal{F}(\eta)$, and for the multinomial case we have $D(Y) \approx \boldsymbol{V}.\boldsymbol{W}^{-1}$ where $\boldsymbol{V}$ and $\boldsymbol{W}$ are matrices defined as $\boldsymbol{V} = n.V(\eta)$ and $\boldsymbol{W}^{-1} = \boldsymbol{I} + n(n-1)\boldsymbol{V}^{-1}.J^T.\mathcal{I}.J$. The estimation procedure may then be as follows:

(1) Choose initial estimates for both $\mathcal{I}$ and $\mathcal{B}$, and calculate w (or $\boldsymbol{W}$).

(2) (Re)estimate $\mathcal{B}$ by QL methods, given the fixed weights w (or $\boldsymbol{W}$).

(3) Reestimate $\mathcal{I}$, recalculate w (or $\boldsymbol{W}$) using the new estimate for $\mathcal{I}$ and $\mathcal{B}$, and return to step (2) (untill some convergence criterion is satisfied).

8. FURTHER RESEARCH

We intend to use the models discussed in the previous sections in future research. First, we will use them in simulation studies in order to find answers on some general questions. Next, we also intend to apply the techniques to real data sets. This section deals with some problems connected with the models considered and indicate how simulation studies may be useful in solving them.

Even if there is no hidden heterogeneity or if hidden heterogeneity can be ignored, some problems arise when one uses the CLM of Section 3. As mentionned in the last paragraph of Section 3, there may indeed exist a good alternative to the Coale-McNeil model in the class of shifted-proportional hazards models. This supports the following study: simulate data under the Coale-McNeil model and try to obtain good fits to this data with a shifted-proportional hazards model. The problem will be to determine a suitable value for the nuisance parameter δ . Such a simulation can be done without considering covariates. If covariates are included, some additional problems arise in view of the structure of the linear predictors. Suppose, for instance, that data are generated under the Coale-McNeil model and under the assumption of additive effects of covariates only (i.e. no interaction effects in the linear predictors). The question is then: is this additivity assumption also valid under a shifted-proportional hazards model? More generally, one may try to find nuisance parameters m_1, m_2 and δ such that all or some specified linear predictors have relatively simple structures. Or, in other words, what may be the effect of changing the values of the nuisance parameters on the structure of the linear predictors. Note that simplicity in the linear predictors will simplify the interpretation of the results, particularly in a comparative study. A problem which is related to this is that of robustness of *summary measures* such as the mean, the median, the variance and the quartiles of age at which an event is experienced. I.e. how important can be the effect of changing the values of the nuisance parameters on the summary measures?

It is worth to mention at this point some studies in related topics. For instance, quite a lot of literature exists on the problem of whether one should use the probit, the logit or the complementary log-log transformation (i.e. the choice of $\Phi(.)$ in our CLM); see Genter (1982) and the reference list in that dissertation. Another problem is that of using the power transformation (also called the Box-Cox transformation, see

Box and Cox (1964)) as a transformation to normality, symmetry and/or additivity (Hinkley, 1975; Breslow and Storer, 1985). Note that the power transformation is used on the r.h.s. of equation (3.2).

Similar problems may arise when heterogeneity is incorporated in the model. It is interesting to mention here the report by Vaupel and Yashin (1985), in which the authors show that even if hazard rates for individuals follow simple patterns, the estimated hazard rate for a heterogeneous cohort can show a much more complex pattern. This gives rise to some interesting problems. For instance, is it possible that simple (e.g. additive) structures for the linear predictors in the CLM are found if the model incorporates heterogeneity, while this may not be so if the model does not incorporate heterogeneity?

The models for heterogeneity discussed in this paper should also be compared with models with a longer tradition. Indeed, remember that the approximate models of Section 7 are likely to be more appropriate when the variances and covariances of the heterogeneity variables $Z_1, \ldots Z_k$ are *small*. The problem is to find some indication on how small these parameters should be. Moreover, as the approximate models do not use any distribution for the heterogeneity variables, one should compare those models with models where Z is assumed to have a particular distribution. This might be studied as follows: simulate heterogeneous data using a particular distribution of Z and find out what is lost if an approximate model is fitted to those data.

9. NUMERICAL EXAMPLE

In this section we illustrate the applicability of the methods by reanalyzing the seed data used first by Crowder (1978) and later by Williams (1982). The seed data can be found in the GLIM program in the Appendix. The *factors* X1 and X2 stand for type of seed (X1=1 for *O. aegyptiaca 75* and X1=2 for *O. aegyptiaca 73*) and root extract (X2=1 for *bean* and X2=2 for *cucumber*); there are five or six *replicates* for each (X1,X2) combination. There are 21 batches of size NX. The seeds are brushed onto a plate covered with a particular root extract and the number DX of germinated seeds in each batch is counted. The purpose of the analysis is to take heterogeneity between replicates into account.

Crowder (1978) and Wiliams (1982) used the logit scale in their analyses of the seed data. Crowder, however, suggested the use of other scales. But modification of Crowder's approach for different scales is likely to become computationally cumbersome, since his approach is based on the full unconditional distribution. We first show that there are no such problems with the approach of Section 7.

The r.h.s. of equation (7.2b) can be written as v/w, as discussed in Section 7. We consider here only simple link functions (i.e. one linear predictor, or k=1) : the logit link $\mathcal{F}(\eta)=1/(1+\exp(-\eta))$, the cloglog link $\mathcal{F}(\eta)=1-\exp(-\exp(\eta))$ and the probit link $\mathcal{F}(\eta)=\int_{-\infty}^{\eta}(2\pi)^{-1/2}\exp(-x^2/2)dx$. Then the matrices **J** and Σ are scalars : $\mathbf{J}=\mathcal{F}'(\eta)=d\mathcal{F}(\eta)/d\eta$ and $\Sigma=\sigma^2$, so that

$$w^{-1}=1+\sigma^2 n(n-1)\mathcal{F}'(\eta)^2/v$$
$$=1+\sigma^2(1-1/n)(n\mathcal{F}'(\eta))^2/v.$$

It is now convenient to define $u=(n.\mathcal{F}'(\eta))^2/v$, so that

$$w^{-1}=1+\sigma^2(1-1/n)u. \qquad (8.1)$$

Indeed, the u's are the *iterative weights* in the *iteratively reweighted least squares* (IRLS) algorithm used in GLIM in order to find the estimates of the linear parameters. Computation of w is then easily done as W=1/(1+%P*(1-1/%BD)*%WT), where %P is an estimate of σ^2 (see the GLIM program in the Appendix). Reestimation of σ^2, if estimates of the linear parameters are given, is based on the (approximate) formula

$$E(X^2)=\Sigma\, w^*(1-w^*u^*q^*)\{1+\sigma^2(1-1/n)u^*\}, \qquad (8.2)$$

which is an extension of Williams' formula (3.4); the summation is over all units (i.e. the batches in the seed data), $E(X^2)$ is an estimate of the Pearson chi-squared statistic, q is the variance of the linear predictor, and * denotes evaluation at the current estimates of the linear parameters. The generality of formulas (8.1) and (8.2) makes it easy to write a flexible GLIM program : i.e. in the program shown in the Appendix one has to change merely the LINK option (G, C or P for the logit, cloglog or probit link, respectively; if one would like to use a non-standard link - through the OWN facility - then two macros should be modified).

The GLIM program shown can be used in various ways by simple modification of only a few *program-parameters* :

i. the contents of macro MOD;
ii. the LINK option;
iii. the (initial) value of σ^2 (%P)
iv. the value of %F to specify whether σ^2 is fixed (%F=0) or reestimated (%F=1); and/or
v. the number of iterations (%R).

Notice that one can also change the contents of the macro OUTP if other results from the fits are required. Table 1 shows the program-parameters and some results for 18 different runs of the program. Column 8 shows the estimated Pearson statistic for the ordinary model that ignores heterogeneity between replicates (i.e. σ^2=0 or w=1); column 7 gives the statistic if the model takes heterogeneity between replicates into account. ANOVA-like tables are shown in Table 2; panels A and B are obtained from columns 7 and 8 of Table 1, respectively. Notice that the values in panel A for the logit link are different from those given by Williams (1982; last paragraph) : Williams has set the weights w in fitting restricted models (e.g. X1+X2) equal to the estimates from the maximum model X1*X2, while we have reestimated the weights w (i.e. only σ^2 is fixed in restricted models). The last three lines of Table 1 show what happens if σ^2 is reestimated from a restricted model (e.g. X1+X2) : then the heterogeneity variable Z incorporates both the heterogeneity between replicates and the variance due to ommission of terms (e.g. the interaction).

The most important results from our analysis are : (1°) that different links give similar results ($\hat{\chi}^2$ values, significance testing), but (2°) that the estimates of σ^2 depend on the link. The first result corresponds to the well-known fact that it is generally difficult to discriminate between links if the sample sizes are not large. The second result shows that one has to be careful in interpreting σ^2 quantitatively; a similar remark holds for the linear parameters! Table 3 shows the estimated conditional proportions $\mathcal{F}(\hat{\eta}+Z)$ for different values of Z : the values of $\hat{\eta}$ +Z are quite different for different links, the values of $\mathcal{F}(\hat{\eta}+Z)$ are not!

Table 1. GLIM program options and results.

Contents of MOD (1)	LINK option (2)	Initial σ^2 (%P) (3)	Fix/est. σ^2 (%F) (4)	Nbr. of iter.(%R) (5)	$\hat{\sigma}^2$ (6)	Heter. $\hat{\chi}^2$ (7)	No heter. $\hat{\chi}^2$ (8)
X1*X2	G	.0	1	5	.1075	17.00	31.65
X1*X2	C	.0	1	5	.0563	17.00	31.65
X1*X2	P	.0	1	5	.0409	17.00	31.65
X1+X2	G	.1075	0	5	-	20.69	38.31
X1+X2	C	.0563	0	5	-	20.26	37.32
X1+X2	P	.0409	0	5	-	20.71	38.31
X1	G	.1075	0	5	-	40.70	91.55
X1	C	.0563	0	5	-	39.96	91.56
X1	P	.0409	0	5	-	41.29	91.55
X2	G	.1075	0	5	-	23.04	41.22
X2	C	.0563	0	5	-	22.66	41.23
X2	P	.0409	0	5	-	23.09	41.22
-	G	.1075	0	5	-	42.67	93.82
-	C	.0563	0	5	-	42.56	93.82
-	P	.0409	0	5	-	43.31	93.82
X1+X2	G	.0	0	5	.1494	18.00	38.31
X1+X2	C	.0	0	5	.0751	18.00	37.32
X1+X2	P	.0	0	5	.0570	18.00	38.31

Table 2. ANOVA tables.

Panel A : Heterogeneity between replicates.

Source	d.f.	Link : Logit	Cloglog	Probit
Interaction	1	3.69	3.26	3.71
X1	1	2.35	2.40	2.38
X2	1	20.01	19.70	20.58
Main effects	2	21.98	22.30	22.60

Panel B : No heterogeneity between replicates.

Source	d.f.	Link : Logit	Cloglog	Probit
Interaction	1	6.66	5.67	6.66
X1	1	2.91	3.91	2.91
X2	1	53.24	54.24	53.24
Main effects	2	55.51	56.50	55.51

Table 3. Estimates $F(\hat{\eta}+Z)$.

		Logit Z=			Cloglog Z=			Probit Z=		
X1	X2	$-\hat{\sigma}$	0	$\hat{\sigma}$	$-\hat{\sigma}$	0	$\hat{\sigma}$	$-\hat{\sigma}$	0	$\hat{\sigma}$
1	1	.2967	.3693	.4483	.3044	.3689	.4421	.2960	.3693	.4476
1	2	.3113	.3855	.4655	.3202	.3870	.4623	.3110	.3856	.4647
2	1	.6141	.6883	.7540	.6038	.6909	.7743	.6139	.6885	.7561
2	2	.4287	.5102	.5911	.4301	.5098	.5951	.4303	.5106	.5905

REFERENCES

Baker R.J. and J.A. Nelder (1978) : The GLIM-System. Release 3. Numerical Algorithms Group : Oxford.

Box G.E.P. and D.R. Cox (1964) : An Analysis of Transformations. J. Royal Statist. Soc. B, 26, pp. 211-252.

Breslow N.E. (1984) : Extra-Poisson Variation in Log-linear Models. Applied Statistics, 33, No. 1, pp. 38-44.

Breslow N.E. and B.E. Storer (1985) : General Relative Risk Functions for Case-Control Studies. Am. J. Epidemiol., 122, pp. 149-162.

Coale A. and D. McNeil (1972) : The Distribution by Age at First Marriage in a Female Cohort. J. American Stat. Ass., 67, No. 340, pp. 743-749.

Cox. C (1984) : Generalized Linear Models - The Missing Link. Applied Statistics, 33, No. 1, pp. 18-24.

Crowder M.J. (1978) : Beta-binomial Anova for Proportions. Applied Statistics, 27, No. 1, pp. 34-37.

Genter F.C. (1982) : A Generalized Regression Model for Ordinal Response Variables. Ph. D. Dissertation, University of Washington.

Hinde J. (1982) : Compound Poisson Regression Models. GLIM 82 : Proceedings of the International Conference on Generalized Linear Models. R. Gilchrist (ed.), Springer-Verlag : Berlin.

Hinkley D.V. (1975) : On Power Transformations to Symmetry. Biometrika, 62, No. 1, pp. 101-111.

Kalbfleisch J.D. and R.L. Prentice (1980) : The Statistical Analysis of Failure Time Data. Wiley : New York.

McCullagh P. (1983) : Quasi-likelihood Functions. The Annals of Statistics, 11, No. 1, pp. 59-67.

Nelder J.A. and R.W.M. Wedderburn (1972) : Generalized Linear Models. J. R. Statist. Soc. A, 135, No. 3, pp. 370-384.

Prentice R.L. (1976) : A Generalization of the Probit and Logit Methods for Dose Response Curves. Biometrics, 32, pp. 761-768.

Rodriguez G. and J. Trussell (1980) : Maximum Likelihood Estimation of the Parameters of Coale's Model Nuptiality Schedule from Survey Data. Technical Bulletin, No. 7, World Fertility Survey : London.

Stirling W.D. (1984) : Iteratively Reweighted Least Squares for Models with a Linear Part. Applied Statistics, 33, No. 1, pp. 7-17.

Thompson R and R.J. Baker (1981) : Composite Link Functions in Generalized Linear Models. Applied Statistics, 30, No. 2, pp.125-131.

Trussell J. and D. Bloom (1981) : Estimating the Covariates of Age at Marriage and First Birth.

Vanderhoeft C. (1983) : A Unified Approach to Models for Analysis of Zero-One Data with Applications to Intermediate Fertility Variables. IPD-Working Paper 1983-5, Interuniversity Programme in Demography : Brussels.

Vanderhoeft C. (1985) : Stratified Proportional Hazards Models. A GLIM oriented Approach, with Special Reference to the Problem of Competing Risks. IPD-Working Paper 1985-5, Interuniversity Programme in Demography : Brussels.

Vanderhoeft C. (1986) : NEGO 4 : An Analysis of Premarital Cohabitation. (Manuscript)

Vaupel J.W. and A.I. Yashin (1985) : Heterogeneity's Ruses : Some Surprising Effects of Selection on Population Dynamics. The American Statistician, vol. 39.

Wedderburn R.W.M. (1974) : Quasi-Likelihood Functions, Generalized Linear Models , and the Gauss-Newton Method. Biometrica, 61, No. 3, pp. 439-447.

Williams D.A. (1982) : Extra-binomial Variation in Logistic Linear Models. Applied Statistics, 31, No. 2, pp. 144-148.

APPENDIX : GLIM-PROGRAM

```
$C MACROS

$M MOD X1*X2 $ENDM

$M ESTP $EXT %VL $CA WVQ=%PW*(1-%PW*%WT*%VL) !
 : %P=(%X2-%CU(WVQ))/%CU(WVQ*A) !
 $PR :: " NEW SIGMA2 = " %P : $$ENDM

$M HET !
 $CA %R=%R-1 : A=(1-1/%BD)*%WT !
 $SW %F ESTP !
 $CA W=1/(1+%P*A) $W W !
 $F #MOD $PR :: " NEW CHI2    = " %X2 : $$ENDM

$MACRO OUTP $D MAR !
 $CA RES=(%YV-%FV)*%SQRT(%PW/(%SC*%FV*(1-%FV/%BD))) !
 : X=X1*X2+%GT(X1,X2) !
 $P RES X $$ENDMAC

$M PROP !
 $EXT %PE $VAR 4 L1 LO L2 !
 $CA LO(1)=%PE(1) : LO(2)=%PE(1)+%PE(2) : LO(3)=%PE(1)+%PE(3) !
 : LO(4)=%PE(1)+%PE(2)+%PE(3)+%PE(4) : L1=LO-%SQRT(%P) !
 : L2=LO+%SQRT(%P) !
 : LO=1/(1+%EXP(-LO)) : L1=1/(1+%EXP(-L1)) : L2=1/(1+%EXP(-L2)) !
 $LO L1 LO L2 $$ENDM
$C 'MAIN' PROGRAM

$UNITS 21
$DATA DX NX X1 X2 $READ
 10 39 1 1
 23 62 1 1
 23 81 1 1
 26 51 1 1
 17 39 1 1
  5  6 1 2
 53 74 1 2
 55 72 1 2
 32 51 1 2
 46 79 1 2
 10 13 1 2
  8 16 2 1
 10 30 2 1
  8 28 2 1
 23 45 2 1
  0  4 2 1
  3 12 2 2
 22 41 2 2
 15 30 2 2
 32 51 2 2
  3  7 2 2
$FACTOR X1 2 X2 2
$YVAR DX $ERR B NX $LINK G

$C CHOOSE (INITIAL) VALUE FOR SIGMA2 : %P
   CHOOSE PROGRAM CONSTANTS : %F=0 IF SIGMA2 IS FIXED
                                 1 IF SIGMA2 IS REESTIMATED
                              %R=MAX NBR. OF ITERATIONS
                                 0 IF NO HETEROGENEITY IS TAKEN INTO
                                   ACCOUNT (I.E. %P=%F=0)

$CALC %P=.0000 : %F=1 : %R=5

$PR :: " (INITIAL) SIGMA2 = " %P :
$CALC W=1 $WEIGHT W
$FIT #MOD $USE PROP
$PR :: " CHI2 (NO HETER.) = " %X2 :
$WHILE %R HET $USE PROP
$STOP
```