LIFE TABLES AND RELATIONAL MODELS : A UNIFIED APPROACH, AND APPLICATIONS IN DEMOGRAPHY

Camille Vanderhoeft

INTERUNIVERSITY PROGRAMME IN DEMOGRAPHY

I.P.D.-WORKING PAPER 1984-3

Acknowledgements :

This paper has become possible thanks to research grant nr. CP82.39A from the Population Council (International Research Awards Programme), the support of the Belgian Administration for Development Cooperation (ABOS-AGCD) and the collaboration of the Kenyan Central Bureau of Statistics. None of these institutions is responsible for the content.

CONTENTS

INTRODUCTION

This paper proposes a unification of models which have been used in various demographic disciplines, such as the analysis of mortality, intermediate fertility variables and nuptiality.

The presentation of models as relational ones is already well-known to demographers, but was used first in reliability studies, bio-assay and medical sciences. The full parametrization of the model stems from mathematical properties of probability distributions as the parametrizable relation between different distributions, and the transformation of one distribution into another through simple one-to-one functions (linear, log-, and power-transformations).

Since the data to be analysed often need to be presented in life table format, a large part of this paper is reserved for the discussion of life table techniques. In fact, the models discussed in this paper are intended for the formal (i.e. statistical) comparison of different life tables.

Finally, estimation of the model parameters is, for quite different studies, carried out with the same techniques. In the past, techniques were developped in function of the analysis to be done and of the problems related to it.

As a consequence, the methods may be attractive for the theoretician, being interested in the statistical background of the model, as well as for the researcher who wants to use the models in order to understand his data.

1. THE RELATIONAL MODEL

1.1. The functional representation

Assume that a sequence of proportions $\pi^*(t)$, for some values of t, represents the cumulative distribution of some phenomenon, and that $\pi_s(t)$ is a standard cumulative distribution for that phenomenon. Then, the functional form of the relational model is :

$$\Phi(\pi^*(t)) = \theta.\Phi(\pi_s(t)) + \beta \qquad [1]$$

That is, there exists a linear relation between the transformations, by $\Phi(.)$, of the cumulative distributions $\pi^*(t)$ and $\pi_s(t)$.

The transformation $\Phi(.)$ is assumed to be strictly increasing from $[0,1]$ into $[-\infty,+\infty]$ with $\Phi(o) = \lim_{\pi \to o} \Phi(\pi) = -\infty$ and $\Phi(1) = \lim_{\pi \to 1} \Phi(\pi) = +\infty$. The parameter θ should be positive.

Typical examples of phenomhena which can be studied through relational models, as will be shown in later sections, are roughly characterized as follows :

- $\pi^*(t)$ is the cumulated proportion of married women, married at exact age t;
- $\pi^*(t)$ is the cumulated proportion of children weaned at exact age t;
- $\pi^*(t)$ is the cumulated proportion of the total fertility, realized at exact age t.

1.2. The transformation

The following transformations frequently occur in demographic analyses(1) :

$\Phi(\pi) = \text{logit}\ (\pi) = \log\ (\frac{\pi}{1-\pi})$, used in the Brass model life tables (Brass et al, 1968);

$\Phi(\pi) = -\log\ (-\log\pi)$, used in the relational Gompertz model of fertility by age of woman (Brass, 1974);

$\Phi(\pi) = \log\ (-\log(1-\pi))$, used in analysis of breast-feeding (Vanderhoeft, 1982, 1983);

---

(1) The natural logarithms are used throughout this text, i.e. $x = \log e^x$.

$$\Phi(\pi) = F_W^{-1}(\pi), \text{ with } F_W(w) = \int_{-\infty}^{w} \frac{(.604)^{.604}}{\Gamma(.604)} e^{-.604(u+e^{-u})} du,$$

used in the Coale-McNeil model of age at first marriage, as shown in Vanderhoeft (1983)(2).

Those special transformations all belong to a very broad class of transformations, defined as follows :

$$\Phi(\pi) = F_W^{-1}(\pi) \qquad [2a]$$

where $F_W(.)$ is a cumulative exponential generalized F distribution function (with $2m_1$ and $2m_2$ degrees of freedom) :

$$F_W(w) = \int_{-\infty}^{w} \frac{\Gamma(m_1+m_2)}{\Gamma(m_1).\Gamma(m_2)} \left(\frac{m_1}{m_2}\right)^{m_1} e^{m_1 u} \left(1 + \frac{m_1}{m_2} e^{u}\right)^{-(m_1+m_2)} du \qquad [2b]$$

The special transformations listed above are obtained for special values of the parameters m_1 and m_2, which should be positive (possibly infinite). Details about this can be found in Vanderhoeft (1983) and are summarized in Appendix Table A.1.

The parametrization of the transformation Φ by [2] links several models of different kinds. It also allows for formal tests on the transformation to be used, and at the same time for the search of other (and better) ones.

From the 2-parameter class of transformations defined by [2], one can derive 1-parameter subclasses. For instance, by putting $m_1 = +\infty$, or $m_2 = +\infty$, or $m_1 = m_2$. It is interesting to note that the above listed special transformations belong to 1-parameter subclasses (see Appendix A).

1.3. The (transformed) standard

Empirical investigation, i.e. comparison of patterns describing the same process (e.g. mortality, nuptiality, fertility,....) in different populations, has shown regularities across populations. The differences can be summarized with 3 parameters :

---

(2) $\Gamma(.)$ denotes the gamma function :

$$\Gamma(a) = \int_{0}^{\infty} e^{-x} x^{a-1} dx$$

- a shift of some time-variable g(t) (parameter β);
- a scale transformation of that time-variable (parameter θ);
- a scale transformation of the (cumulated) pattern (parameter C).

Thus, for two populations 1 and 2, with cumulated patterns $\pi_1(t)$ and $\pi_2(t)$, one might have :

$$\Phi(C_1^{-1}.\pi_1(t)) = \theta_1.g(t) + \beta_1$$

$$\Phi(C_2^{-1}.\pi_2(t)) = \theta_2.g(t) + \beta_2$$

For the rescaled patterns $\pi_i^*(t) = C_i^{-1}\pi_i(t)$ we can write then :

$$\Phi(\pi_2^*(t)) = \theta.\Phi(\pi_1^*(t)) + \beta$$

Hence, π_1^* (t) can be used as a standard.

Here, one of the populations being studied becomes a standard (or reference) population. For instance, the group of illiterate women, living in rural areas may be the reference group in a study of the effects of education and residence on some phenomenon in a particular country. Such standard (sub) populations may be useful for local comparisons (e.g. within the country). But since they depend on the analysis to be carried out, they will not be widely applicable. Therefore, investigators were looking for universal standards $\pi_s(t)$.

Heather Booth (1979) derived a universal standard schedule for fertility. Lesthaeghe and Page (1980) obtained smooth standard schedules for breastfeeding and post-partum amenorrhoea. Roughly speaking, they considered a particular observed schedule from a population with data of fairly good quality, smoothed this schedule to remove irregularities due to random errors in the observations and eventually corrected it to give a good fit if used in the relational model.

Other investigators were looking for mathematical formulae describing the (standard) pattern of some process. Simple methods proceed through various plots of the transformed schedules $\Phi(\pi^*(t))$ against time t or a transformation of t (e.g. log t). For instance, Martin (1967) found that a plot of log-log transforms (i.e. $\Phi(\pi)=-\log(-\log\pi)$) of the rescaled cumulated fertility against age of women shows a linear relation. Coale and McNeil (1972) proceeded through more complex calculations to find a parametric schedule of first marriage in a female cohort from Swedish 1865-69 data.

A broad class of time-transformations, defining at the same time a standard schedule (if Φ is given) can be represented by the formula :

$$\Phi(\pi_s(t)) = \alpha_1 + \alpha_2 \frac{(t + \alpha_4)^{\alpha_3} - 1}{\alpha_3} \qquad [3]$$

One may fix the parameters α_1, α_2, α_3 and/or α_4, or estimate them from a real data set. Note that a parametric standard schedule is defined by giving some values to the parameters α_1, α_2, α_3, α_4, m_1 and m_2(3).

Finally, we shall assume throughout the text that the standard schedules are normalized, i.e. $\pi_s(t) \longrightarrow 1$ if $t \longrightarrow +\infty$, from which it follows that $C_s = 1$ or $\pi^*_s(t) = \pi_s(t)$.

---

(3) Constraints are : $\alpha_2 > 0$, $\alpha_3 \geqslant 0$, $t+\alpha_4 > 0$ if α_3 is non-integer or zero. The choice of α_3, m_1 and m_2 is crucial, since they define the shape of the distribution. Note that $\Phi(\pi_s(t)) = \alpha_1 + \alpha_2 \log (t+\alpha_4)$ if $\alpha_3=0$, and $\Phi(\pi_s(t)) = \alpha_1 + \alpha_2(t+\alpha_4-1)$ if $\alpha_3 = 1$.

2. GENERALIZED SEMI-LINEAR RELATIONAL MODELS (GESLIRM)

2.1. Multiple regression analysis

Demographic surveys, such as the World Fertility Surveys, often contain multiple measurements on each individual subject (or sample of individual subjects). For instance, apart from a woman's fertility (e.g. children ever born), information is collected on other variables which affect fertility, such as biological (e.g. sex, lactation period, contraceptive use,...) and socio-economic characteristics (e.g. education, religion..).

Simple techniques consisting of calculating measures for all groups of individuals separately may provide useful preliminary insight in the data. However, they usually suffer from basic problems as sample fragmentation, since demographic surveys are not experimentally designed.

This can be a reason for the application of a relational multiple regression model : a relational model induces a fixed underlying law (which is believed to hold in general) in all subsamples, removing irregularities in the observed underlying pattern, and a multiple regression model allows for joint effects of the predictors on the dependent variable. Thus, we shall extend [1] to a multiple regression model.

If $\pi(t;z)$ represents the process to be studied for a population with characteristics $z = (z_1,\ldots z_p)$, and if $\pi_s(t)$ is a standard schedule, then the relational multiple regression model is :

$$\Phi(C(z)^{-1}.\ \pi(t;z)) = \theta(z).\Phi(\pi_s(t)) + \beta(z) \qquad [4]$$

Thus, all populations z are assumed to follow the same general law $\pi_s(t)$, but they differ from this law in scale and location which are summarized in the parameters $C(z)$, $\theta(z)$ and $\beta(z)$ - given Φ.

Although other assumptions may be interesting too, we shall further assume that the parameters $\theta(z)$ and $\beta(z)$ are linear in the characteristics z, i.e. $\theta(z) = \theta.z'$ and $\beta(z) = \beta.z'$ where θ and β are parameter vectors $(\theta_1,\ldots\theta_p)$ and $(\beta_1,\ldots\beta_p)$, where z' is the transposed of z, and

$\theta.z' = \Sigma\theta_j z_j$ and $\beta.z' = \Sigma\beta_j z_j$. Then, [4] can be written as :

$$\Phi(C(z)^{-1}.\pi(t;z)) = \sum_{j=1}^{p} \theta_j.\ z_j\Phi(\pi_s(t)) + \sum_{j=1}^{p} \beta_j z_j \qquad [5a]$$

Thus, the regression model is linear in the parameters θ_j and β_j ($1 \leqslant j \leqslant p$) and, if $C(z)$ is known for all z, linear in the regressors z_j and $z_j \Phi(\pi_s(t))$ ($1 \leqslant j \leqslant p$). Writing $\pi^*(t;z) = C(z)^{-1}.\pi(t;z)$, it follows that :

$$\Phi(\pi^*(t;z)) = \sum_{j=1}^{p} \theta_j . \; z_j \Phi(\pi_s(t)) \; + \sum_{j=1}^{p} \beta_j z_j \qquad [5b]$$

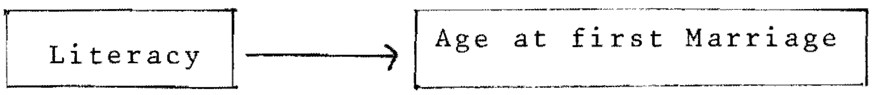
Hence, the multiple regression model for the normalized schedules $\pi^*(t;z)$ is generalized linear, where "generalized" refers to the transformation $\Phi(.)$ and "linear" to both the parameters and the predictors. In case of [5b], we shall speak of Generalized Linear Relational (multiple regression) Models (GLIRM). In the general case [5a], we speak of Generalized Semi-Linear Relational (multiple regression) Models (GESLIRM).

A last remark in this section concerns the predictors of the models [5]. Of course, the z_j could be either interval-scaled (continuous) or categorical (nominal and/or ordinal); mixtures might be considered as well. However, since interval-scaled variables (e.g. age) can be recoded into - naturally ordinal - categorical variables (e.g. age-cohort), we will deal in this text with categorical variables only. In practice, the predictors z_j in [5] are then always dummy variables. For the applications these models turn out to be flexible enough. For instance, interactions between variables can be modelled easily, and curvilinear relations can be taken into account.

Note, however, that $\Phi(\pi_s(t))$ itself is an interval-scaled variable (if fixed).

2.2. Two simple examples

Let's consider, for example, the following simple causal models. One states an effect of literacy on age at first marriage (in a female cohort, say) :



The other one states an effect of literacy on fertility (number of children ever born, say) :

From WFS data sets on all women (i.e. ever and never married women), we can subtract the following variates :

z_i = 1 if woman i is literate;
 = 0 if women i is illiterate;

t_i = age of woman i in completed years at the time of the survey;

δ_i = 1 if woman i was ever married;
 = 0 if woman i was never married;

m_i = number of children born to woman i.

By grouping of the individual data, we get for each recorded age t :

n_{1t} = number of literate (z=1) women, aged t completed years at the time of the survey;

n_{ot} = number of illiterate (z = 0) women, aged t completed years at the time of the survey;

d_{1t} = number of literate (z=1) women, aged t completed years and ever married (δ = 1);

d_{ot} = number of illiterate (z = 0) women, aged t completed years and ever married (δ = 1);

m_{1t} = number of children ever born to literate women (z = 1), aged t completed years;

m_{0t} = number of children ever born to illiterate women (z=0), aged to completed years.

Formally $d_{zt} = \sum_{\substack{i|z_i=z \\ t_i=t}} \delta_i$, and $m_{zt} = \sum_{\substack{i|z_i=z \\ t_i=t}} m_i$.

Then, the ratios d_{zt}/n_{zt} and m_{zt}/n_{zt} are observed estimates of $\pi(t+\frac{1}{2};z)$, where $\pi(t;z)$ represents the age at first marriage pattern respectively the fertility pattern in subsample z.

The GESLIRM is perfectly designed to fit a model to this kind of data : i.e. to data where information on the variable of interest is recorded by means of sequences of cumulative counts or proportions (e.g. data on age at first marriages are given as a sequence n_{zt},d_{zt} $(t=t_o,...t_1)$ of counts, two at each age t between some t_o and t_1).

In these examples we used current status data. It will be shown later how retrospective data can be used equally well. Note also that the fertility data and nuptiality data have been treated in the same way, although they collect information on very different kinds of events. More specifically, the fertility data are on a renewable event (i.e. birth of a child), while the nuptiality data are on a non-renewable event (i.e. first marriage). More details about these types of data - their differences and similarities - are discussed later on.

2.3. Interpretation of parameters and models

It is not always obvious how to give a demographic interpretation to the parameters of the models. Often, we need to transform them into a set of parameters which are more suitable for interpretations.

Roughly speaking, we can consider a group of "shape" parameters, a group of "scale" parameters and a group of "location" parameters. Since it is hard - or even impossible - to give a general outline for interpretations, we shall discuss only a few simple examples (other examples may be found in the applications - sections later on).

Let's consider the following situation :

$$\Phi(\pi_s(t)) = \alpha_1 + \alpha_2 t,$$

from which it follows that :

$$\Phi(\pi^*(t)) = \theta.(\alpha_1 + \alpha_2 t) + \beta$$
$$= \theta^+ t + \beta^+$$

where $\theta^+ = \theta.\alpha_2$ and $\beta^+ = \beta + \theta.\alpha_1$. It is then easy to see that θ^+ and β^+ are respectively a scale and a location parameter of time t. This follows from the following equations :

$$\bar{\theta}^+ t + \bar{\beta}^+ = \theta^+ \frac{t - a}{k} + \beta^+$$

where $k = \theta^+/\bar{\theta}^+$ and $a = (\beta^+ - \bar{\beta}^+)/\bar{\theta}^+$. Thus, a change in β^+ implies a shift of time t, and a change in θ^+ implies a scale transformation of time t.

Similarly, if we define "standard-time" t_s by $t_s = \alpha_1 + \alpha_2 t$, then a change in θ implies a scale transformation of standard-time t_s and a change in β implies a shift of standard-time t_s. This can also be extended to more general cases (i.e. non-parametric standard $\pi_s(t)$, or parametric standard $\pi_s(t)$ but $\alpha_3 \neq 1$ and/or $\alpha_4 \neq 1$) : define standard-time $t_s = \Phi(\pi_s(t))$.

However, since such interpretations always use a concept of "time", we use in this text a more objective terminology. We shall simply speak of "slopes" θ (or θ^+) and "intercepts" β (or β^+), having in mind the regression of $\Phi(\pi^*(t))$ on $\Phi(\pi_s(t))$.

Now, in general, $\pi^*(t)=C^{-1}\pi(t)$. Under the assumption that the standard $\pi_s(t)$ is normalized (i.e. $C_s=1$ or $\pi_s^*(t)=\pi_s(t)$), C is simply a scale parameter for the distribution itself. This parameter, however, needs more attention, as it follows from the following discussion.

First, note that the normalizing assumption (section 1.3) implies that $\pi(t;z) \rightarrow C(z)$ if $t \rightarrow +\infty$ for all populations z; also $\pi^*(t;z) \rightarrow 1$ if $t \rightarrow +\infty$. In analyses of nuptiality in a female cohort, where $\pi(t;z)$ is the proportion of women ever married by age t, C(z) is called the "ultimate proportion marrying". In breastfeeding analyses, C(z) is the "ultimate proportion of weanlings". And in fertility analyses, C(z) is usually called the "total fertility".

Now, in analyses of fertility, one often considers an age T (usually 50 years) after which definitely no more births will be recorded. The existence of such a cut off age T is of course justified by the fact that the menopause falls between ages 40 and 50 years. Hence, the reproductive period is definitely ended by age 50 (=T). Similarly, one may find that a woman cannot give her child the breast after T months (e.g. 48 months), simply because she can no longer produce milk. In practice, investigators use these facts by considering only women aged under 50 years (in fertility analyses) or children born less than 48 months before the survey (in breastfeeding analyses). Moreover, in fertility analyses one has almost always worked explicitly with $\pi_s(T) = 1$ (e.g. Booth, 1979; Zaba, 1981), and hence $\pi^*(T;z) \underset{s}{=} 1$ or $\pi(T;z) = C(z)$ for all populations z which are related to the same standard $\pi_s(t)$. In breastfeeding analyses, however, one does not so, and in nuptiality analyses the existence of a cut off age T may even be doubtful.

Thus, in certain instances, the existence of a cut off time T can be very well justified on biological grounds. We shall now discuss, however, situations wherein it might be less realistic to consider a cut off time T(4). For instance,

---

(4) It is assumed implicitly that the cut off time T is such that $\pi_s(t)<1$ for all $t<T$ and $\pi_s(t) = 1$ for all $t \geq T$. Hence, T is the smallest cut off time, and it is so in all populations. In other words $\pi^*(t;z)<1$ for all $t<T$ and $\pi^*(t;z)=1$ for all $t \geq T$, and for all z.

if breastfeeding is prolonged, one may record weanings still at ages 45, 46, 47 (if T = 48 months). If breastfeeding is not prolonged, on the other hand, then it is quite possible that all children are weaned by a much smaller age (e.g. 24 months). Hence, for two different populations the cut off point may be different. Similar situations can be met in fertility analyses : in one population, births may be recorded at ages 35 and above, while in another population all births may be recorded before age 30.

The conclusion is that a universal cut off time T may be justified on biological grounds, but not on non-biological (e.g. socio-economic) grounds. The solution to the problem is fairly simple. Indeed, we shall not work explicitly with a time T for which $\pi_s(T) = 1$ (and hence $\pi^*(T;z) = 1$ for all z; see also footnote (4)). On the contrary, we assume that $\pi_s(t) < 1$ for all finite t, and hence $\pi^*(t;z)<1$ or $\pi(t;z)<C(z)$ for all finite t and for all z. Cut off times can then still be defined. For instance, one considers a time T_s so that $\pi_s(T_s)=.99$, and times T_z so that $\pi^*(T_z;z)=.99$ or $\pi(T_z;z)=.99C(z)$. Such a new cut off point varies across populations due to non-biological differences. The existence of a universal biological cut off time T is still interesting. For instance, one knows that woman's reproductive history is complete if she is followed up to age 50. Thus, older women do not offer more information on fertility than women under age 50 do, and so they need not to be included in the survey.

Interpretation of other parameters (m_1, m_2, α_3 and α_4) is even less obvious. Briefly, we shall refer to these as "shape" parameters, only to distinguish them from scale and location parameters. They are, however, interesting for general interpretations of the model. I.e. special values of these parameters imply special nice properties of the model. Examples are found in Appendix Table A.2.

3. THE DATA

3.1. General format

The data for which a relational model is designed should be of the form :

$$[d_{zt},\ n_{zt},\ t,\ z\ ] \qquad [6a]$$

or of the form :

$$[p_{zt},\ n_{zt},\ t,\ z\ ] \qquad [6b]$$

where

$z = (z_1, \ldots z_p)$ is a vector of dummies (section 2.1), representing some characteristics of a sample of individual subjects;

t = duration;

n_{zt} = number of individual subjects in sample z, observed during time t;

d_{zt} = number of events experienced during time t by the n_{zt} subjects in sample z;

$p_{zt} = d_{zt}/n_{zt}$.

The duration variable t measures the time between two events, e.g. the event of becoming at risk and recording the data. Of course, t could be an exact variate, but as we are concerned here with analyses of WFS data, we shall only consider duration variables t measured in completed units (e.g. years, months). Then, an event recorded as taking place at time t, in fact takes place somewhere in the interval of unit length $[t, t+1)$.

The following sections will deal with the derivation of data in format [6] from WFS surveys.

3.2. Renewable and non-renewable events

This text is intended partially to show similarities between renewable and non-renewable events. More specifically, estimation of the parameters of the relational model for data on renewable and non-renewable events shows no basic differences. However, there are important differences between them in obtaining the appropriate data (in format [6]). This will be made clear in the next sections. Here, we will only mention some differences in the terminology.

When dealing with non-renewable events (e.g. first marriage, weaning of a child,...) the ratio p_{zt} is an observed estimate of the probability that an individual with covariates z has experienced the event by time t. For instance, in a nuptiality analysis p_{zt} is the probability that a woman with characteristics z maries before age t. Note that $0 \leq p_{zt} \leq 1$ or, equivalently, $0 \leq d_{zt} \leq n_{zt}$.

When dealing with renewable events (e.g. births, marriages,...), an individual may experience the event more than once during time t. Hence, p_{zt} is the mean number of events experienced by the n_{zt} individuals. For instance, in a fertility analysis p_{zt} is the mean number of children born to the n_{zt} women with characteristics z in their first t years of life. Note that d_{zt} can be larger than n_{zt}, or that p_{zt} can be larger than 1.

Thus, we speak of "probabilities" if the data are on non-renewable events, and we speak of "mean numbers" if the data are on renewable events.

3.3. Data on non-renewable events

3.3.1. Introduction

We shall proceed as follows in the derivation of data in format [6] from WFS surveys. First, a cross-tabulation is constructed. From this, two types of data (both in format [6]) are easy to subtract. As they result, in fact, from two life table techniques, they are compared with the more common actuarial life table technique.

Since one of the techniques, as well as the actuarial life table technique, uses retrospective data (i.e. retrospectively reported durations or ages), we assume throughout this discussion that retrospective information is recorded. If such data are not available, a cross-tabulation cannot fully be constructed, and one is then obliged to use the current status technique (which is the other technique to be discussed in what follows).

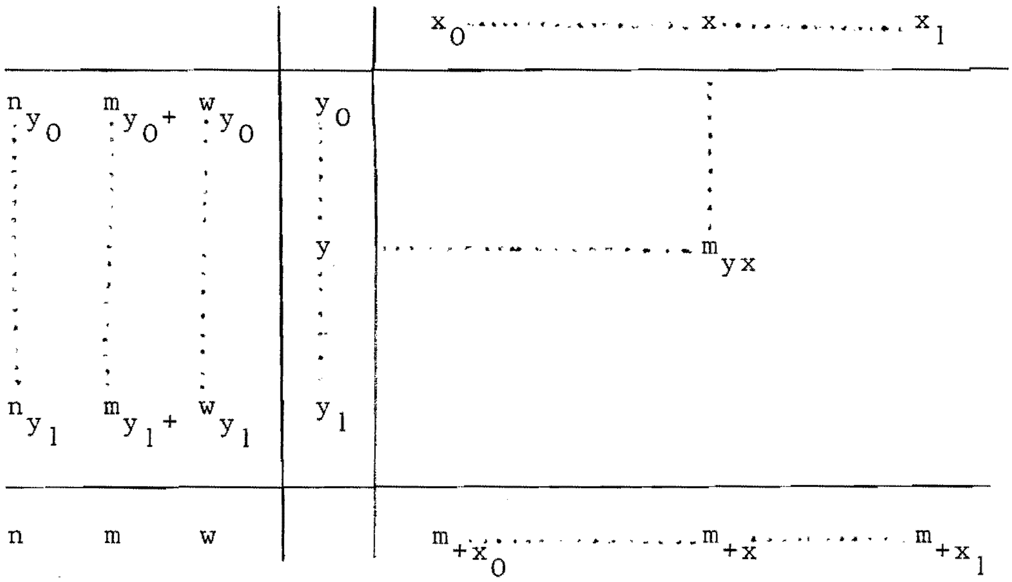
In the following sections, we use the terminology of nuptiality analyses. Of course, the discussion applies also to breast-feeding, post-partum abstinence and any other analysis of non-renewable events.

Finally, we shall omit covariates z from the notations, since the techniques discussed here are applicable to separate samples only.

3.3.2. Cross-tabulations and life tables

We consider a cohort of women whose age at the date of interview (i.e. current age, or CA) lies between y_0 and y_1, and whose age at first marriage (AM) lies between x_0 and x_1 (if they are ever married)(5). Then, we can construct a cross-tabulation as shown in Fig. 1 and the definitions below.

FIG. 1 : CROSS-TABULATION FOR NON-RENEWABLE EVENTS



We define :

m_{yx} = number of women with CA y and AM x;

w_y = number of never married women with CA y.

The following marginals can be computed(6) :

---

(5) As mentionned in section 3.1, all ages (CA and AM) are measured in completed years. Thus, "with AM x" means : married in the unit interval $[x,x+1)$. Note also that $x_0 \leqslant y_0 \leqslant y_1 = x_1$

(6) In fact, for the marginals m_{y+} the summation should be from x_0 to y. However, by a simple extension of the cross-tabulations with cells with zero entries, the summations are correct.

$$m_{y+} = \sum_{x=x_0}^{x_1} m_{yx}$$ = number of ever married women with CA y;

$$m_{+x} = \sum_{y=y_0}^{y_1} m_{yx}$$ = number of women with AM x;

$$m = \sum_{y=y_0}^{y_1} m_{y+} = \sum_{x=x_0}^{x_1} m_{+x} = \sum_{y=y_o}^{y_1} \sum_{x=x_0}^{x_1} m_{yx}$$ = number of married women;

$n_y = m_y + w_y$ = number of women with CA y;

$$w = \sum_{y=y_0}^{y_1} w_y$$ = number of never married women;

$$n = \sum_{y=y_0}^{y_1} n_y$$ = number of women in the sample.

$= m + w$

Further, we define some cumulative counts :

$$d_{yx} = \sum_{s=x_0}^{x} m_{ys}$$ = number of women with CA y and AM less than or equal to x (completed years);

$$d'_{yx} = d_{yx} - m_{yx} = \sum_{s=x_0}^{x-1} m_{ys}$$ = number of women with CA y and AM less than or equal to x-1 (completed years).

Now, if $\pi(t;y)$ is the cumulated proportion of women married by exact age t in a cohort aged y, we have :

$$\frac{d_{yx}}{n_y} \text{ estimates } \pi(x+1;y) \text{ if } x<y, \qquad [7a]$$

$$\frac{d_{yy}}{n_y} \text{ estimates } \pi(y+\tfrac{1}{2};y), \qquad [7b]$$

$$\frac{d'_{yx}}{n_y} \text{ estimates } \pi(x;y) \text{ for all } x \leqslant y \qquad [7c]$$

Note that [7a] and [7c] are the same, but we distinguished them for later purposes (i.e. when several sigle year cohorts are grouped). In fact, it is easier to work with counts d'_{yx} than counts d_{yx}.

Next sections will deal with the problem of grouping of single year cohorts and with the computation of life table estimates for the proportion $\pi(t)$ of women married by exact age t in the cohort of women aged y_0 to y_1.

Complexity in grouping single year cohorts is due to "progressive time censoring"(7). In the above problem of estimating patterns $\pi(t;y)$, only simple "time censoring" was met. We shall discuss three life table methods for (progressively) time censored data : the current status (CS) method, the actuarial (ACT) method and the retrospective zero-one (R01) method. The first two are well-known to demographers. The third method proposed in this text is intended for use with relational models(8). Fortunately, it shows clear similarities with the ACT method.

---

(7) "Time censoring" is often called "Type I censoring" (e.g. Kalbfleisch and Prentice, 1980). Elandt-Johnson and Johnson (1980) use the term "truncation" instead of "time censoring". In this text, however, truncation refers to other kinds of data. For instance, nuptiality data on ever married women are truncated.

(8) The R01 method (for non-renewable events) has been used in Vanderhoeft (1983). Unfortunately, the definitions were not exact, and caused biases.

3.3.3. Current status life tables

The CS life table estimates for the nuptiality pattern $\pi(t)$ of a cohort of women aged y_0 to y_1 are easily obtained from the estimates [7]. We have the following possibilities :

$$\text{CS-1 :} \quad \frac{d'_{yy}}{n_y} \quad \text{estimates } \pi(y),\ y_0 \leqslant y \leqslant y_1 \qquad [8]$$

$$\text{CS-2 :} \quad \frac{d_{yy}}{n_y} \quad \text{estimates } \pi(y+\tfrac{1}{2}),\ y_0 \leqslant y \leqslant y_1 \qquad [9]$$

The CS-2 method assumes that both exact CA's and exact AM's are distributed uniformly over one-year intervals. This assumption is not needed in the CS-1 method, since the n_y women are all followed up to exact age y at least and since d'_{yy} is the number of women married before that exact age y.

However, the CS-1 method is not applicable if, from the survey, only current status data (i.e. current age of women and marital status at date of interview) are available. Then, we are forced to use the CS-2 method. Unfortunately, the above assumption used in this method may have serious consequences for the accuracy of the estimates if units are larger (e.g. 5 years). We are thinking here about possible seasonality, perhaps not in nuptiality experiences, but possibly in other ones.

Note that the CS-2 method is the method used with household data (e.g. Rodriguez and Trussell, 1980).

3.3.4. Actuarial life tables

The ACT life table method uses the numbers(9)

$$N_x = \sum_{i \geqslant x} m_{+i} + \sum_{y \geqslant x} w_y \qquad [10]$$

i.e. the number of women "at risk" at exact age x. Assuming that both marriages and exact ages are uniformly distributed over each year, $N_x - \frac{1}{2} w_x$ is the mean number of women exposed to risk during interval [x,x+1), and $m_{+x}/(N_x - \frac{1}{2} w_x)$ is the conditional probability to marry in the interval [x,x+1), given not married before exact age x. Hence, we get :

(9) For remarks on the summation, see footnote (6).

$$1 - \prod_{i<x} \left(1 - \frac{m_{+i}}{N_i - \frac{1}{2}w_i}\right) \text{ estimates } \pi(x) \qquad [11]$$

Note that ACT and CS-2 methods use the same assumption, so that the same remarks on the estimates can be made.

The ACT method is mentionned here only for a comparison with CS and R01 methods. However, no data in format [6] is provided by the ACT method.

3.3.5. Retrospective zero-one life tables

R01 life table methods of estimation are using the following counts, which can, of course, all be found from the cross-tabulations as in section 3.3.2 :

$$d'_x = \sum_{y=x}^{y_1} d'_{yx} = \sum_{y=x}^{y_1} \sum_{s=x_0}^{x-1} m_{ys}$$

$$d_x = \sum_{y=x}^{y_1} d_{yx} = \sum_{y=x}^{y_1} \sum_{s=x_0}^{x} m_{ys}$$

$$n'_x = \sum_{y=x}^{y_1} n_y$$

Hence, n'_x is the number of women with CA larger than or equal to x, d'_x is the number of women among those n'_x women who are married before exact age x, and d_x is the number of women among those n'_x women who are married before exact age x+1.

Then, we get the following estimates :

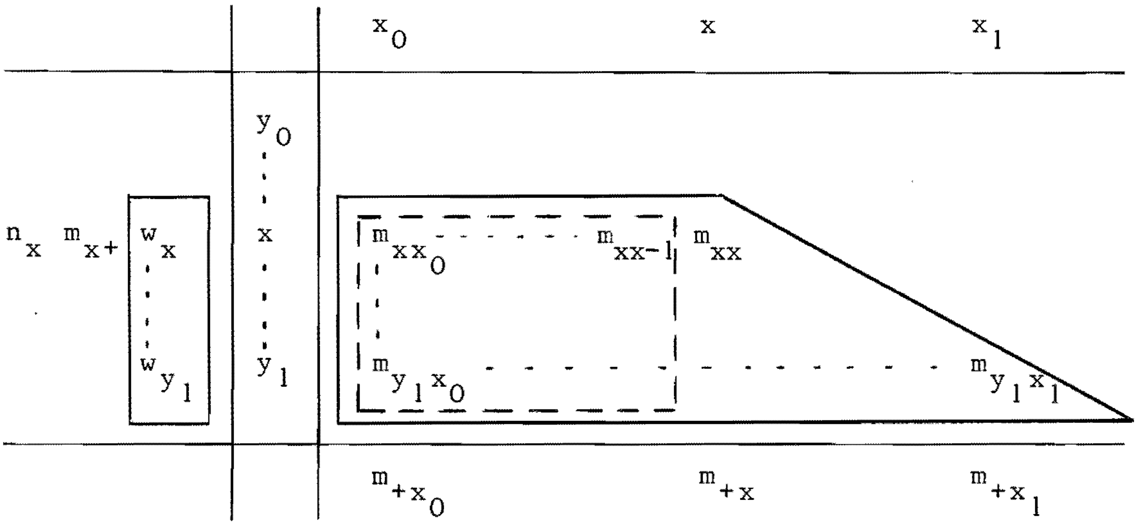
$$\text{R01-1 :} \quad \frac{d'_x}{n'_x} \text{ estimates } \pi(x) \qquad [12]$$

$$\text{R01-2 :} \quad \frac{d_x}{n'_x} \text{ estimates } \pi(x+\tfrac{1}{2}) \qquad [13]$$

Again, the assumption of uniformly distributed ages and marriages over each year is needed for the RO1-2 method but not for the RO1-1 method. Hence, remarks similar to those in section 3.3.3 can be made. Note also that [12] reduces to [7 c] for single year cohorts.

The RO1-1 approach is visualized in Fig. 2 .

FIG. 2 : RO1-1 LIFE TABLES FROM CROSS-TABULATIONS OF NON-RENEWABLE EVENTS(*)



(*) n'_x is the sum of counts in the boxes with full lines
d'_x is the sum of counts in the box with broken lines

Note that the RO1-2 method can be visualized similarly : d'_x is replaced by d_x, and only the bar with broken lines in Fig. 2 should be extended one column to the right hand side.

RO1 life tables are in fact the result of the following procedure. In order to get RO1-1 data, having CA y and AM x (in completed years) we construct for each woman a sequence of zeros and ones (called "individual RO1-1 data") as follows(10):

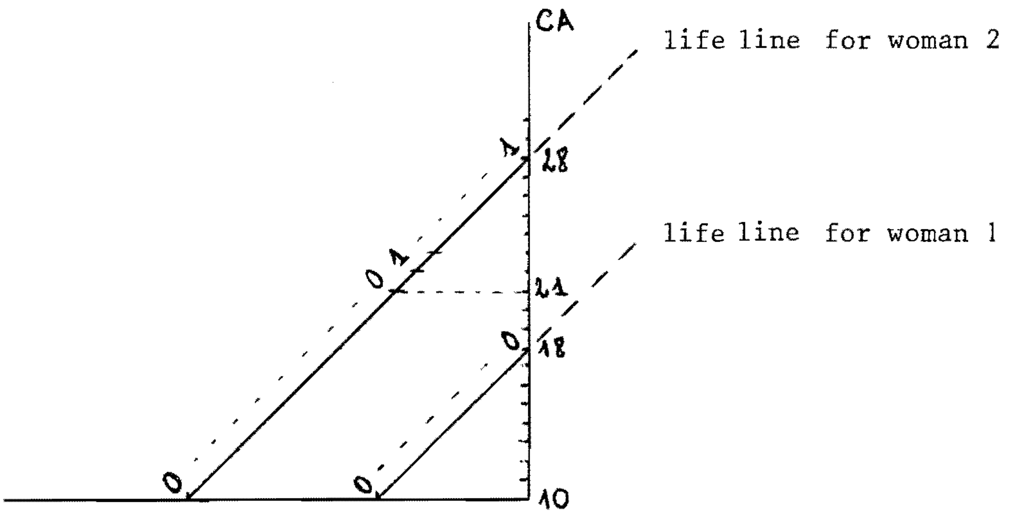
$$\delta_t = 0 \text{ if } t \leqslant x \qquad [14]$$
$$= 1 \text{ if } x < t \leqslant y$$

This is shown in Fig. 3. Woman 1 has current age 18 completed years and is not yet married. Hence, her marital status at

(10) The lower bound for t is e.g. the minimum age at first marriage in the sample; it may be lower, but it has to be equal for all women in that sample.

exact ages 10, 11,... is coded zero. Note that she may still marry before she attains exact age 19. Woman 2 has current age 28 completed years and married at age 21 completed years. Hence, her status at exact ages 10, 11,...21 is coded zero, and her status at exact ages 22, 23,...28 is coded one. Note that women whose CA and AM (in completed years) are equal, are not considered as ever being married, by this approach.

FIG. 3 : INDIVIDUAL RO1-1 DATA ON NUPTIALITY



The RO1-2 approach can be explained similarly. Now, for each woman a sequence of δ_t's is constructed as follows :

$$\delta_t = 0 \text{ if } t<x \qquad [15]$$
$$= 1 \text{ if } x\leqslant t\leqslant y$$

After constructing such a sequence for each woman in the sample, we sum up the δ_t's at any t over all women with the same current age, and so we get the counts d'_{yt} (or d_{yt}). The counts n'_y are simply obtained by counting the number of women with current age y and above.

3.3.6. Comments

First, we consider the estimation from truncated data. WFS surveys pertain for some countries to ever-married women only. Hence, the data on nuptiality is truncated. The reasoning followed in sections 3.3.2 - 3.3.5 is still appropriate, provided the following modifications :

1. Cross-tabulations of retrospective data are still possible, but now $w_y = 0$ for all y. Hence : $w = 0$, $n = m$ and $n_y = m_{y+}$ for all y.

2. For single year age cohorts (say women with CA y), we get estimates of the truncated distributions. For instance, the R01-1 method applied on the cohort aged y for which only truncated data is available, becomes :

$$\frac{d'_{yx}}{n_y} \text{ estimates } \frac{\Pi(x;y)}{\Pi(y;y)} \qquad [16]$$

3. One cannot combine single year age cohorts as we did in previous sections, since the truncation time is different across cohorts. However, truncation of the data at the same time point in all cohorts would make grouping possible. For instance, if marriages before exact age 25 are used only, then we could combine cohorts aged 25 and above, and we could find estimates for the nuptiality pattern truncated at age 25.

4. The CS method can not be applied to truncated data.

We can show, using $w_i = 0$ for all $i<y_0$, that ACT and R01-1 estimates are equal at exact ages x, where $x \leqslant y_0$. Indeed, we have :

$$N_x = \sum_{j \geqslant x} m_{+j} + \sum_{y \geqslant x} w_y = \sum_{j \geqslant x} m_{+j} + \sum_{y \geqslant y_0} w_y$$

for all $x \leqslant y_0$. It follows then that, for all $x \leqslant y_0$:

$$N_{x_0} - N_x = \sum_{j=x_0}^{x-1} m_{+j} = d'_x, \text{ and } N_{x-1} - N_x = m_{+x-1}.$$

We have also that $n'_x = N_{x_0}$ for all $x \leqslant y_0$. Hence, we get indeed, for all $x \leqslant y_0$:

$$1 - \prod_{i<x} \left(1 - \frac{m_{+i}}{N_i - \frac{1}{2} w_i}\right) = 1 - \prod_{i<x} \left(1 - \frac{m_{+i}}{N_i}\right)$$

$$= 1 - \prod_{i<x} \frac{N_i - m_{+i}}{N_i}$$

$$= 1 - \prod_{i<x} \frac{N_{i+1}}{N_i}$$

$$= 1 - \frac{N_x}{N_{x_0}} = \frac{N_{x_0} - N_x}{N_{x_0}}$$

$$= \frac{d'_x}{n'_x}$$

It can be seen also that ACT and RO1-1 estimates are close together if the numbers w_y are small. It follows thus that for both a single year cohort and for truncated data the ACT and RO1-1 method give the same estimates.

These facts may justify the RO1 method. Moreover, we note that RO1 and CS estimates are exactly the same at the oldest age y_0 in the cohort, and are close together at ages near y_0. Hence, at young ages, the RO1 estimates are close to the ACT estimates, while at older ages, the RO1 estimates are close to the CS estimates. Numerical examples will be given later on.

To obtain data in format [6] , one can thus follow two different procedures : the current status (life table) technique or the retrospective zero-one (life table) technique.
Of course, which one is actually used depends on the problem to be investigated and on the data being available. For instance, if only household data on nuptiality is recorded, then only the CS-2 method is applicable. If retrospective data is available, then both the CS and RO1 method are technically applicable. However, it will depend then on the accuracy of the data (e.g. retrospectively reported AM) and perhaps on statistical features which technique is best to be used.

3.4. Data on renewable events

3.4.1. Introduction

The steps followed in the next sections are mainly as those for non-renewable events. Therefore, the discussion is somewhat shortened, taking over equivalent results for non-renewable events. So we could concentrate more on special problems with data on renewable events.

The terminology of fertility (e.g. children ever born) analyses is used here. Another example would be an analysis of marriage (e.g. number of times ever married) : the discussion would be completely equivalent.

Again, covariates z are omitted from the notations for the same reason as it was for non-renewable events.

3.4.2. Cross-tabulations and life tables

Consider a cohort of women whose age at date of interview (i.e. CA) lies between y_0 and y_1 and suppose that their age at the birth of any of their children (i.e. AB) lies between x_0 and x_1. Again, CA and AB are recorded in completed years, and $x_0 \leqslant y_0 \leqslant y_1 = x_1$ (cfr. footnote 5). Then, the cross-tabulation of Fig. 4, according to the definitions below can be constructed.

FIG. 4 : CROSS-TABULATION FOR RENEWABLE EVENTS

			x_0	...	x	...	x_1
n_{y_0}	m_{y_0+}	y_0			⋮		
⋮	⋮	⋮			⋮		
		y	...	...	m_{yx}		
⋮	⋮	⋮					
n_{y_1}	m_{y_1+}	y_1					
n	m		m_{+x_0}	...	m_{+x}	...	m_{+x_1}

We define :

m_{yx} = number of children born to women with CA y when they had age x;

n_y = number of women with CA y.

Then, we can compute marginals :

$$m_{y+} = \sum_{x=x_0}^{x_1} m_{yx}$$ = number of children born to women with CA y;

$$m_{+x} = \sum_{y=y_0}^{y_1} m_{yx} = \text{number of children born when the mother had age } x;$$

$$m = \sum_{x=x_0}^{x_1} m_{+x} = \sum_{y=y_0}^{y_1} m_{y+} = \sum_{x=x_0}^{x_1} \sum_{y=y_0}^{y_1} m_{yx} = \text{number of children ever born;}$$

$$n = \sum_{y=y_0}^{y_1} n_y = \text{number of women.}$$

Further, we define :

$$d_{yx} = \sum_{s=x_0}^{x} m_{ys} = \text{number of children born to women wity CA } y, \text{ before they were } x+1 \text{ exact years old;}$$

$$d'_{yx} = \sum_{s=x_0}^{x-1} m_{ys} = \text{number of children born to women with CA } y, \text{ before they were } x \text{ exact years old.}$$

It is seen that the formal representation of the data on fertility (in general : renewable events) through a cross-tabulation is similar to that for nuptiality (in general : non-renewable events), except for two points :

1. The w-column in Fig. 3 has disappeared in Fig. 4.

2. The n-column is no longer a column of marginals, as it was in Fig. 3.

Nevertheless, if $\pi(t;y)$ is the cumulated fertility (i.e. mean number of children born) at exact age t for women with CA y, then we may consider estimations as [7a-c].

The problem of next sections is to find estimates for the cumulated fertility distribution $\pi(t)$ corresponding to the pooled single year age cohorts y ($y_0 \leq y \leq y_1$). As for non-renewable events, CS, ACT and R01 techniques are discussed.

3.4.3. Current status life tables

The CS life table technique for non-renewable events can simply be taken on here. Thus, [8] and [9] are still valid. Problems similar to those for non-renewable events arise, and therefore, the reader is referred to section 3.3.3.

3.4.4. Actuarial life tables

The ACT life table technique for non-renewable events seems to have an equivalent for renewable events in single year age cohorts only. If we define :

$$N'_{yx} = \sum_{s \geq x} m_{ys}, \text{ then } 1 - \prod_{i<x} \left(1 - \frac{m_{yi}}{N'_{yi}}\right)$$

is the probability of having a birth before exact age x, given there will be at least one birth before exact age y (where y is the age in completed years of the cohort considered). This probability is an estimate at t=x of the truncated fertility distribution $\pi(t;y)/\pi(y;y)$. Since, according to [8] (i.e. the CS-1 method) d'_{yy}/n_y estimates $\pi(y;y)$, we get finally :

$$\frac{d'_{yy}}{n_y} \cdot \left(1 - \prod_{i<x} \left(1 - \frac{m_{yi}}{N'_{yi}}\right)\right) \text{ estimates } \pi(x;y) \qquad [17]$$

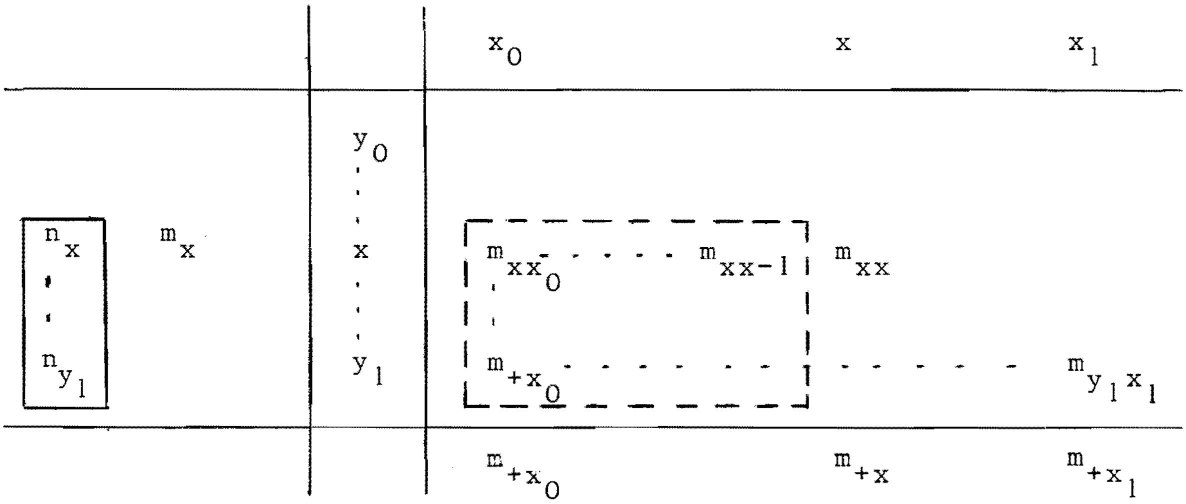
An ACT life table technique for pooled single year age cohorts could be outlined. For instance, having 5 successive single year cohorts, one should have to measure both current age and ages at birth in units of 5 years. Then, one could obtain estimates as [11], but since x is measured in units of 5 years (e.g. x=4 means that AB is between 20 years and 25 years), one would have estimates after each 5-year period only. Thus, the estimated fertility pattern would be less detailed (although it may be smoother). Such techniques would side-track us very much, as all other life table techniques discussed in this paper provide estimates after each 1-year period (if the year is the original unit of measurement).

The method of the next section, still providing estimates at each year even when single year cohorts are pooled, may thus be very useful as an alternative method for the construction of life tables where the actuarial method fails.

3.4.5. Retrospective zero-one life tables(11)

Counts d'_x, d_x and n'_x are still defined as in section 3.3.5. The interpretations, however, are now : d'_x is the number of births before exact age x for women currently aged x and above; d_x is the number of births before exact age x+1 for women currently aged x and above; n'_x is the number of women currently aged x and above. These definitions are visualized in Fig. 5.

FIG. 5 : R01-1 LIFE TABLES FROM CROSS-TABULATIONS OF RENEWABLE EVENTS(*)



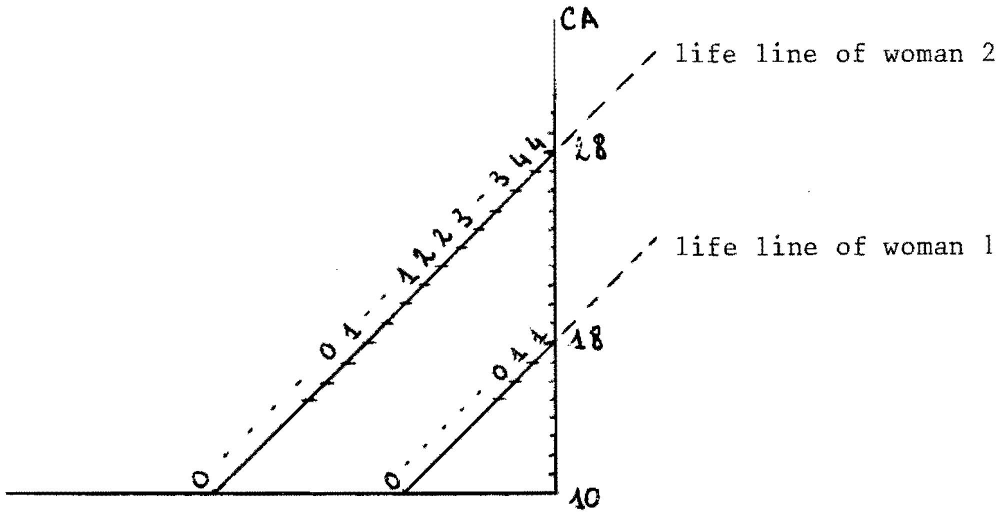
(*) n'_x is the sum of counts in the boxes with full lines
d'_x is the sum of counts in the box with broken lines

The R01 life table estimates are then exactly as in formulae [12] and [13], and again the same remarks as in section 3.3.5 can be made.

(11) From the discussion it follows that the term "zero-one" is not quite appropriate here. But since there are clear similarities between the R01 method for non-renewable events and the method of this section, we shall use the same terminology.

The construction of individual R01-1 data is shown in Fig. 6.

FIG. 6 : INDIVIDUAL R01-1 DATA ON FERTILITY



In Fig. 6, woman 1 has CA 19 and did have only 1 birth at age 16 (completed years). Woman 2 has CA 28 and did have births at ages 17, 21, 23 and 26 (completed years). A sequence of δ_t's can thus be constructed for each woman. Summation of δ_t's at any t over all women with the same current age y would then give immediately the counts d'_{yt} (or d_{yt}).

3.4.6. Comments

Truncated data on renewable events can also be considered. If the truncation time is y, then the data would only include women who did have at least one birth before age y (for computation of the n_j's $j \leqslant y$) and all births of those women before age y (for computation of the m_{ij}'s, $j \leqslant y$, $i \leqslant j$). This is completely equivalent with truncated data on non-renewable events. Hence, the remarks on truncated data made in section 3.3.6 are still valid here.

The problem of section 3.4.4., i.e. computation of actuarial estimates for pooled single year cohorts, seems to be inherent in data on renewable events. Fortunately, it are the current status or the retrospective zero-one techniques which provide the data for our relational models; actuarial life table techniques are not useful to this purpose.

4. ESTIMATION OF THE PARAMETERS

The parameters of the relational models can be estimated through several techniques. Most commonly used are graphical and (weighted) least squares methods. Maximum likelihood methods have become more popular in demography only recently.

The general model involves parameters of different kinds : intercepts β (and α_1), slopes θ (and α_2), scale parameters C, shape parameters m_1 and m_2 (and α_3, α_4). This makes it very hard to develop a general estimation procedure, although particular techniques may be easily applied under special circumstances. For instance, (weighted) least squares and even a graphical technique is easily applied if only θ and β have to be estimated.

In the past, several estimation methods have been developped, but they are only applicable in special situations. For instance, if only θ and β are to be estimated in a logit model (i.e. $m_1=m_2=1$, $\pi_s(t)$ fixed, C=1), then the minimum logit χ^2 criterion may be used. Another method, applied in logit regression to estimate θ and β, is based on maximum likelihood methods and the use of "working logits". Equivalent techniques were developped for probit regression. The reader is referred to Finney (1971) for methods in probit analysis, and to Ashton (1972) for methods in logit analysis. Zaba (1981) has developped a "ratio" method for analysis of fertility data through the relational Gompertz model (i.e. $m_1=+\infty$, $m_2=1$, $\pi_s(t)$ fixed). Rodriguez and Trussell (1980) discussed several maximum likelihood procedures for estimation of the parameters in the Coale-McNeil nuptiality model. Both Zaba (1981) and Rodriguez and Trussell (1980) considered the estimation of C, although in quite different ways.

The method being used in this paper is based on maximum likelihood and derived estimation techniques, as will be outlined in the following paragraphs. In order to simplify the notations, only one (sub)sample is considered, so that covariates z could be dropped. Attention is focussed mainly on the estimation of the parameters θ and β, all other parameters being fixed.

Consider either CS-1 or RO1-1 data $[d_t, n_t, t]$ or $[p_t, n_t, t]$ (cfr. section 3.1), where p_t is an estimate of $\pi(t)$. We may then consider the function

$$L = \prod_t \pi(t)^{d_t} (1-\pi(t))^{n_t-d_t}$$

which is briefly called the likelihood function. For both CS-1 and RO1-1 data, p_t is the unconstrained estimate of $\pi(t)$ obtained by maximization of L or logL. Note that in case of

CS-2 or RO1-2 data, $\pi(t)$ should be replaced by $\pi(t+\frac{1}{2})$; then p_t is the unconstrained estimate of $\pi(t+\frac{1}{2})$. If estimation of $\pi(t)$ is constrained through a relational model, then logL is a function of the parameters θ, β, m_1, m_2 (and possibly α_1, α_2, α_3, α_4) and is maximized with respect to the unknown parameters. As usual, this is done through derivation of logL with respect to the unknown parameters, equating this derivatives to zero, and solving the so obtained system in the unknown parameters.

The function L needs special attention. For CS data, it is exactly - or at least proportional to - the conditional likelihood under the assumption that each d_t is, conditionally on n_t, binomially distributed with mean $n_t.\pi(t)$. In fact, this follows from the independency of the sub-samples n_t at different times t. For RO1 data the subsamples n_t are clearly not independent, since n_t entirely covers n_{t+1}. Therefore, L is not a conditional likelihood. Of course, $\pi(t)^{d_t}(1-\pi(t))^{n_t-d_t}$ is still the binomial conditional probability of d_t responses out of n_t, but the product over t, - which is L - is not the joint conditional probability of the random vector d_t given the random vector n_t. The method of partial likelihood (Cox, 1975) - of which the conditional likelihood method is only a special case - seems to provide a justification for L if RO1 data are available. It is interesting to note here that Cox (1975) has outlined the construction of the partial likelihood function for "grouped life table data" which is similar to the ACT data (section 3). Having in mind the equivalence of ACT and RO1 data in case of simple time censoring (section 3.3.6), Cox's approach might then be used.

Of course, the parameters of the relational models can always be found as those who maximize the above function L, whatever the mathematical justification may be. However, problems arise when asymptotic (or large sample) properties of estimates and of likelihood (-ratio) functions are used. Some of these problems are for instance discussed in Andersen (1970) and Cox (1975).

Maximization of L turns out to be rather easy if only θ and β have to be estimated and if the GLIM (release 3) package is available on a computer (Baker and Nelder, 1978). The GLIM-program has been used intensively to obtain the estimates in subsequent sections. We also adopted the notations used in the GLIM-manual to describe the "linear predictor" of the model. For instance, if A and B are categorical and SST a continuous variate, then (A*B).SST + B, or explicitly A.SST + B.SST + A.B.SST + B, means that A and B have main effects and an interaction effect on the slope and that only B has a main effect on the intercept. In the following sections, SST stands for the transformed standard schedule $\Phi(\pi_s(t))$. If SST does not appear in the expression for the linear predictor, then it is assumed that the slope has a fixed value (usually 1). Note that the transformation Φ is called the "link function" in the GLIM-manual.

5. APPLICATION TO BREASTFEEDING DATA

5.1. The model

A model for analysis of breastfeeding data is specified as follows :

1. The transformation of the model is the complementary loglog

$$\Phi(\pi) = \log(-\log(1-\pi)) \qquad [18a]$$

2. The transformed standard schedule is a linear function of log t :

$$\Phi(\pi_s(t)) = \alpha_1 + \alpha_2 \log t \qquad [18b]$$

3. The ultimate proportion of children weaned is 1, and thus, for all covariates z, one has :

$$C(z) = 1 \qquad [18c]$$

Formula [18a] can be found from [2a-b] by taking $m_1=1$ and the limit case $m_2 = +\infty$. Formula [18b] can be obtained from [3] by taking $\alpha_4 = 0$ and the limit case $\alpha_3 = 0$. Formula [18c] expresses the fact that any child will ever be weaned if its age becomes large enough.

From [18a-b] it follows that the duration of breastfeeding t has a 2-parameter Weibull distribution. Log t has an extreme value (minimum) distribution.

Since $\pi(t;z)=\pi^*(t;z)$ by formula [18c], we get the following multiple regression model for breastfeeding :

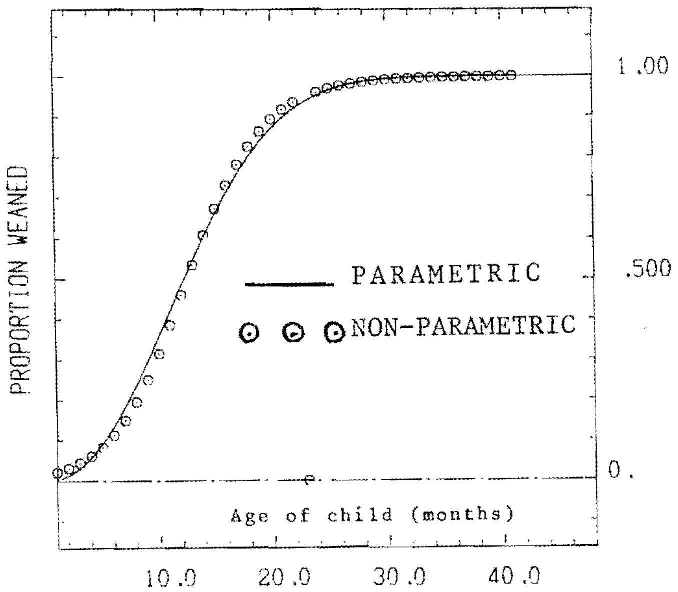
$$\log(-\log(1-\pi(t;z)))=(\theta.z')(\alpha_1+\alpha_2 t)+\beta.z' \qquad [18d]$$

The remaining unspecified parameters α_1, α_2, θ_j and β_j are not independent, and we have to fix one of the parameters α_2, θ_j and one of the parameters α_1, β_j. We have fixed α_1 and α_2. Their values are theoretically unimportant, but for practical purposes a particular set of values may be preferred. We have taken estimates $\hat{\alpha}_1=-5.908$ and $\hat{\alpha}_2=2.210$. With these values [18a-b] defines a standard schedule for breastfeeding which is close to the non-parametric standard obtained by Lesthaeghe and Page (1980) (12). The non-parametric standard $\pi_s(t)$ and

(12) In fact $\hat{\alpha}_1=-5.908$ and $\hat{\alpha}_2=2.210$ are maximum likelihood estimates, obtained by fitting [18a-b] to the standard in Lesthaeghe and Page (1980). (Note that t is measured in months)

the estimated Weibull standard $\hat{\pi}_s(t)$ are contrasted in Fig. 7 and Table 1. The mean and variance of the Weibull standard $\hat{\pi}_s(t)$ are 12.83 and 37.60 respectively.

FIG. 7 : PARAMETRIC AND NON-PARAMETRIC BREASTFEEDING STANDARDS



If the model is fitted under the constraints $\theta.z'=1$ for all covariates z, it can be seen from Appendix Table A.2 that the hazard functions $\lambda(t;z)$ for different z-values are proportional and the process of weaning is accelerated or decelerated. Formally, [18d] becomes

$$\log(-\log(1-\pi(t;z)))=\log(-\log(1-\hat{\pi}_s(t)))+\beta.z' \quad [19a]$$

which is equivalent to one of the following formulae(13) :

$$S(t;z)=\hat{S}_s(t)^{e^{\beta.z'}} \quad [19b]$$

$$\lambda(t;z)=e^{\beta.z'}\hat{\lambda}_s(t) \quad [19c]$$

---

(13) S denotes the survival function : $S(t)=1-\pi(t)$, and λ denotes the hazard function :

$$\lambda(t) = \frac{d}{dt}\log S(t).$$

TABLE 1 : NON-PARAMETRIC AND PARAMETRIC BREASTFEEDING STANDARDS

Duration t in months	$\pi_s(t)$ (*)	$\hat{\pi}_s(t)$ (**)	Duration t in months	$\pi_s(t)$	$\hat{\pi}_s(t)$
1	.021	.003	25	.969	.965
2	.031	.012	26	.976	.974
3	.044	.030	27	.981	.981
4	.061	.057	28	.985	.986
5	.083	.091	29	.988	.990
6	.113	.133	30	.991	.993
7	.149	.182	31	.993	.995
8	.195	.236	32	.994	.997
9	.250	.295	33	.995	.998
10	.314	.356	34	.996	.999
11	.386	.420	35	.997	.999
12	.460	.483	36	.997	.999
13	.535	.545	37	.998	>.999
14	.608	.604	38	.998	-
15	.673	.660	39	.998	-
16	.731	.712	40	.999	-
17	.782	.759	41	.999	-
18	.826	.801	42	>.999	-
19	.863	.838	43	-	-
20	.893	.870	44	-	-
21	.917	.897	45	-	-
22	.935	.919	46	-	-
23	-	.938	47	-	-
24	.960	.953	48	-	-

(*) $\pi_s(t)$ is the cumulated proportion of children weaned by age t as given by Lesthaeghe and Page (1980, Table 1)

(**) $\hat{\pi}_s(t)=1-\exp(-\exp(\hat{\alpha}_1+\hat{\alpha}_2 \log t))$, where $\hat{\alpha}_1=-5.908$ and $\hat{\alpha}_2=2.210$.

$$\pi(t;z)=\hat{\Pi}_s(e^{\beta.z'/\hat{\alpha}_2}.t) \qquad [19d]$$

$$\lambda(t;z)=e^{\beta.z'/\hat{\alpha}_2}.\hat{\lambda}_s(e^{\beta.z'/\hat{\alpha}_2}.t) \qquad [19e]$$

Formulae [19b-c] show that the model is now a proportional hazards (PH) model, and subsample z has - relative to the standard $\hat{\pi}_s(t)$ - a relative risk $e^{\beta.z'}$ at any time t. Formulae [19d-e] show that the model is now an accelerated failure time (AFT) model, and subsample z has - relative to the standard $\hat{\pi}_s(t)$ - an acceleration factor $e^{\beta.z'/\hat{\alpha}_2}$.

5.2. Data and life tables

The data on breastfeeding durations are extracted from the Kenya Fertility Survey 1977-78 (KFS). The computations presented in this paper are not on the entire set of 8100 women interviewed in the KFS, but on women residing in Central and Eastern provinces only (N=2585).

In this paper, we are interested in the effect of female education on fertility. Two covariates have been constructed as follows. The individual education variable (IED) defines 4 groups of women, resp. with 0 years (IED=1), 1 to 4 years (IED=2), 5-8 years (IED=3) and more than 9 years (IED=4) of education (variable V704 in the KFS; women for which years of education is not stated, i.e. V704=99, are excluded).
The contextual education variable (CED) groups sampling strata into 3 blocks : the mean number of years of education of women in a sampling stratum can be less than 3 (CED=1), from 3 to 4.9 (CED=2), or more than 5 (CED=3). Central and Eastern provinces contain 25 sampling strata; according to the CED-levels as defined above, there are resp. 7, 9 and 9 sampling strata (with resp. 794, 1019 and 772 women). Combination of 4 IED-groups with 3 CED-groups yields 12 subgroups.

More details and an argumentation in support of these covariates are found in Lesthaeghe et al (1983).

Further, the data on breastfeeding durations are on ever married women only. Moreover, since breastfeeding durations were asked for the open and the last closed pregnancy interval only, we only considered births corresponding to those intervals. Non-live births were excluded, as well as children whose duration of breastfeeding was not stated or who were breastfed until they died. Finally, the analyses were restricted to

TABLE 3A : LIFE TABLE ESTIMATES OF THE CUMULATIVE DISTRIBUTION $\pi(t)$ OF BREASTFEEDING DURATIONS FROM DATA ON THE TWO MOST RECENT PREGNANCIES OF MARRIED WOMEN IN CENTRAL AND EASTERN PROVINCES(*)

	CS-1 METHOD			ACT METHOD					RO1-1 METHOD		
t	n_t	d'_{tt}	$p_t^{(1)}$	N_{t-1}	m_{+t-1}	w_{t-1}	${}_1p_{t-1}$	$p_t^{(2)}$	n'_t	d'_t	$p_t^{(3)}$
1	36.008	0.000	0.000	1891.335	39.661	–	.979	.021	1891.335	39.661	.021
2	41.910	1.053	.025	1851.674	2.650	36.008	.999	.022	1855.327	42.311	.023
3	39.733	2.541	.064	1813.016	9.100	40.857	.995	.027	1813.417	50.358	.028
4	41.106	1.155	.028	1763.059	18.743	36.812	.989	.038	1773.684	66.180	.037
5	48.970	1.463	.030	1707.504	17.307	39.951	.990	.048	1732.578	82.332	.048
6	48.196	1.420	.029	1650.246	30.192	47.507	.981	.065	1683.608	111.061	.066
7	51.056	1.843	.036	1572.547	50.622	46.776	.967	.096	1635.412	160.263	.098
8	47.574	1.625	.034	1475.149	28.629	46.956	.980	.114	1584.356	184.792	.117
9	48.462	3.578	.074	1399.564	63.869	43.724	.954	.155	1536.782	244.811	.159
10	44.554	8.865	.199	1291.971	64.421	44.884	.949	.198	1488.320	305.654	.205
11	33.776	2.332	.069	1182.666	38.085	33.446	.967	.224	1443.766	332.631	.230
12	56.375	8.474	.150	1111.135	25.372	31.165	.977	.242	1409.990	355.392	.252
13	48.968	15.325	.313	1054.598	310.579	45.216	.699	.470	1353.615	654.812	.484
14	44.769	22.291	.498	698.803	32.130	31.025	.953	.495	1304.647	668.999	.513
15	47.292	20.369	.431	635.648	62.194	22.478	.900	.545	1259.878	708.902	.563
16	41.464	17.772	.429	550.976	66.091	24.074	.877	.601	1212.586	751.775	.620
17	28.263	15.563	.551	460.811	43.102	23.692	.904	.639	1171.122	777.105	.664
18	56.340	35.474	.630	394.017	28.102	12.700	.928	.665	1142.859	789.644	.691
19	39.498	29.094	.737	353.215	80.481	20.866	.765	.744	1086.519	834.651	.768
20	37.875	25.515	.674	251.868	19.898	10.404	.919	.765	1047.021	825.455	.788
21	38.735	27.545	.711	221.566	31.199	12.360	.855	.799	1009.146	831.139	.824
22	31.256	23.853	.763	178.007	11.219	11.190	.935	.812	970.411	814.813	.840
23	37.194	32.167	.865	155.598	6.849	5.784	.955	.820	939.155	796.190	.848
24	33.361	28.013	.840	142.965	3.525	5.027	.975	.825	901.961	767.548	.851
25	34.663	31.736	.916	134.413	64.521	4.158	.512	.910	868.600	802.866	.924
26	39.460	35.260	.894	65.734	5.989	1.721	.908	.919	833.937	775.913	.930
27	46.033	41.303	.897	58.024	4.338	4.200	.922	.925	794.477	744.991	.938
28	38.974	34.310	.880	49.486	.582	4.730	.988	.926	748.444	704.270	.941
29	41.555	38.171	.919	44.174	2.013	4.471	.952	.929	709.470	671.780	.947
30	71.876	63.008	.877	37.690	2.066	1.318	.944	.933	667.915	633.609	.949
31	44.620	44.620	1.000	34.306	2.706	8.868	.909	.939	596.039	573.307	.962
32	38.610	36.218	.938	22.732	.582	0.000	.974	.941	551.419	529.269	.960
33	34.743	34.743	1.000	22.150	0.000	2.392	1.000	.941	512.809	493.051	.961
34	27.184	26.049	.958	19.758	1.745	0.000	.912	.946	478.066	460.053	.962
35	33.131	31.411	.948	18.013	0.000	1.135	1.000	.946	450.882	434.004	.963
36	32.140	32.140	1.000	16.878	2.190	.581	.868	.953	417.751	403.644	.966
37	31.777	31.777	1.000	14.107	3.833	0.000	.728	.966	385.611	375.337	.973
38	30.297	30.297	1.000	10.274	0.000	0.000	1.000	.966	353.834	343.560	.971
39	36.633	35.613	.972	10.274	0.000	0.000	1.000	.966	323.537	313.263	.968
40	35.937	32.929	.916	10.274	0.000	1.020	1.000	.966	286.904	277.650	.968
41	28.667	26.077	.910	9.254	.595	2.413	.926	.968	250.967	244.721	.975
42	77.056	75.019	.974	6.246	0.000	2.590	1.000	.968	222.300	218.644	.984
43	26.804	26.804	1.000	3.656	0.000	2.037	1.000	.968	145.244	143.625	.989
44	39.257	39.257	1.000	1.619	0.000	0.000	1.000	.968	118.440	116.821	.986
45	14.509	14.509	1.000	1.619	0.000	0.000	1.000	.968	79.183	77.564	.980
46	18.298	18.298	1.000	1.619	0.000	0.000	1.000	.968	64.674	63.055	.975
47	19.233	17.614	.916	1.619	0.000	0.000	1.000	.968	46.376	44.757	.965
48	27.143	27.143	1.000	1.619	0.000	1.619	1.000	.968	27.143	27.143	1.000

(*) t=age of child in months; n_t=number of children with current age t completed months; d'_{tt}=number of children among those n_t who are weaned before exact age t; N_{t-1}=number of children at risk at exact age t-1; m_{+t-1}=number of children weaned if age is in [t-1,t); w_{t-1}=number of children censored if age is in [t-1,t); ${}_1p_{t-1}=m_{+t-1}/(N_{t-1}-\frac{1}{2}w_{t-1})$=probability to be weaned after exact age t if not be weaned before exact age t-1; n'_t=number of children with current age larger than or equal to t; d'_t=number of children among those n'_t who are weaned before exact age t; $p_t^{(1)}=d'_{tt}/n_t$=CS-1 estimate of $\pi(t)$, according to [8]; $p_t^{(2)}=1-\prod_{i<t} {}_1p_t$=ACT estimate of $\pi(t)$, according to [11] ; $p_t^{(3)}=d'_t/n'_t$=RO1-1 estimate of $\pi(t)$, according to [12]

TABLE 3B : LIFE TABLE ESTIMATES OF THE CUMULATIVE DISTRIBUTION $\pi(t)$ OF BREASTFEEDING DURATIONS FROM DATA ON THE TWO MOST RECENT PREGNANCIES OF MARRIED WOMEN IN CENTRAL AND EASTERN PROVINCES, AND WITH CED=2 AND IED=1 (*)

	CS-1 METHOD			ACT METHOD					R01-1 METHOD		
t	n_t	d'_{tt}	$p_t^{(1)}$	N_{t-1}	m_{+t-1}	w_{t-1}	$_1P_{t-1}$	$p_t^{(2)}$	n'_t	d'_t	$p_t^{(3)}$
1	4.682	0.000	0.000	337.581	12.814	–	.962	.038	337.581	12.814	.038
2	8.036	0.000	0.000	324.767	0.000	4.682	1.000	.038	332.899	12.814	.038
3	4.524	1.054	.233	320.085	1.239	8.036	.996	.042	324.863	14.053	.043
4	13.983	0.000	0.000	310.810	4.755	3.470	.985	.056	320.339	17.754	.055
5	11.840	1.270	.107	302.585	0.000	13.983	1.000	.056	306.356	17.754	.058
6	16.643	1.141	.069	288.602	3.263	10.570	.988	.067	294.516	19.747	.067
7	10.209	0.000	0.000	274.769	6.952	15.502	.974	.092	277.873	25.558	.092
8	3.393	0.000	0.000	252.315	3.520	9.054	.986	.105	267.664	27.923	.104
9	5.387	0.000	0.000	239.741	6.638	2.252	.972	.129	264.271	33.420	.126
10	6.922	2.067	.299	230.851	11.344	5.387	.950	.173	258.884	44.764	.173
11	5.992	0.000	0.000	214.120	8.789	4.855	.958	.207	251.962	51.486	.204
12	5.267	0.000	0.000	200.476	5.744	5.992	.971	.230	245.970	57.230	.233
13	11.134	2.082	.187	188.740	55.224	5.267	.703	.459	240.703	112.454	.467
14	9.716	4.843	.498	128.249	3.229	7.897	.974	.473	229.569	112.446	.490
15	6.649	1.155	.174	117.123	13.539	4.873	.882	.535	219.853	121.142	.551
16	4.306	2.248	.522	98.711	14.121	5.494	.853	.603	213.204	134.108	.629
17	3.349	3.349	1.000	79.096	9.372	2.058	.880	.651	208.898	141.232	.676
18	10.296	3.459	.336	67.666	2.175	0.000	.968	.662	205.549	140.058	.681
19	9.018	7.994	.886	65.491	15.742	6.837	.746	.748	195.253	152.341	.780
20	3.480	2.241	.644	42.912	1.090	1.024	.974	.754	186.235	145.437	.781
21	6.807	4.511	.663	40.798	3.247	1.239	.919	.774	182.755	146.443	.801
22	8.171	6.901	.845	36.312	3.875	2.296	.890	.799	175.948	145.807	.829
23	3.876	3.876	1.000	30.141	0.000	1.270	1.000	.799	167.777	138.906	.828
24	3.643	3.643	1.000	28.871	0.000	0.000	1.000	.799	163.901	135.030	.824
25	2.208	2.208	1.000	28.871	16.153	0.000	.441	.911	160.258	147.540	.921
26	4.401	4.401	1.000	12.718	0.000	0.000	1.000	.911	158.050	145.332	.920
27	5.958	4.677	.785	12.718	2.163	0.000	.830	.927	153.649	143.094	.931
28	3.168	3.168	1.000	10.555	0.000	1.281	1.000	.927	147.691	138.417	.937
29	6.099	5.086	.834	9.274	0.000	0.000	1.000	.927	144.523	135.249	.936
30	18.486	18.486	1.000	9.274	1.013	0.000	.891	.935	138.424	130.163	.940
31	11.593	11.593	1.000	8.261	0.000	0.000	1.000	.935	119.938	111.677	.931
32	9.070	7.842	.865	8.261	0.000	0.000	1.000	.935	108.345	100.084	.924
33	4.532	4.532	1.000	8.261	0.000	1.228	1.000	.935	99.275	92.242	.929
34	6.643	6.643	1.000	7.033	1.164	0.000	.834	.945	94.743	88.874	.938
35	3.348	3.348	1.000	5.869	0.000	0.000	1.000	.945	88.100	82.231	.933
36	3.062	3.062	1.000	5.869	0.000	0.000	1.000	.945	84.752	78.883	.931
37	4.511	4.511	1.000	5.869	3.610	0.000	.385	.979	81.690	79.431	.972
38	3.247	3.247	1.000	2.259	0.000	0.000	1.000	.979	77.179	74.920	.971
39	6.621	5.601	.846	2.259	0.000	0.000	1.000	.979	73.932	71.673	.969
40	3.231	3.231	1.000	2.259	0.000	1.020	1.000	.979	67.311	66.072	.982
41	11.595	10.356	.893	1.239	0.000	0.000	1.000	.979	64.080	62.841	.981
42	26.663	26.663	1.000	1.239	0.000	1.239	1.000	.979	52.485	52.485	1.000
43	4.856	4.856	1.000	–	–	–	–	–	25.822	25.822	1.000
44	9.014	9.014	1.000	–	–	–	–	–	20.966	20.966	1.000
45	2.044	2.044	1.000	–	–	–	–	–	11.952	11.952	1.000
46	3.396	3.396	1.000	–	–	–	–	–	9.908	9.908	1.000
47	1.029	1.029	1.000	–	–	–	–	–	6.512	6.512	1.000
48	5.483	5.483	1.000	–	–	–	–	–	5.483	5.483	1.000

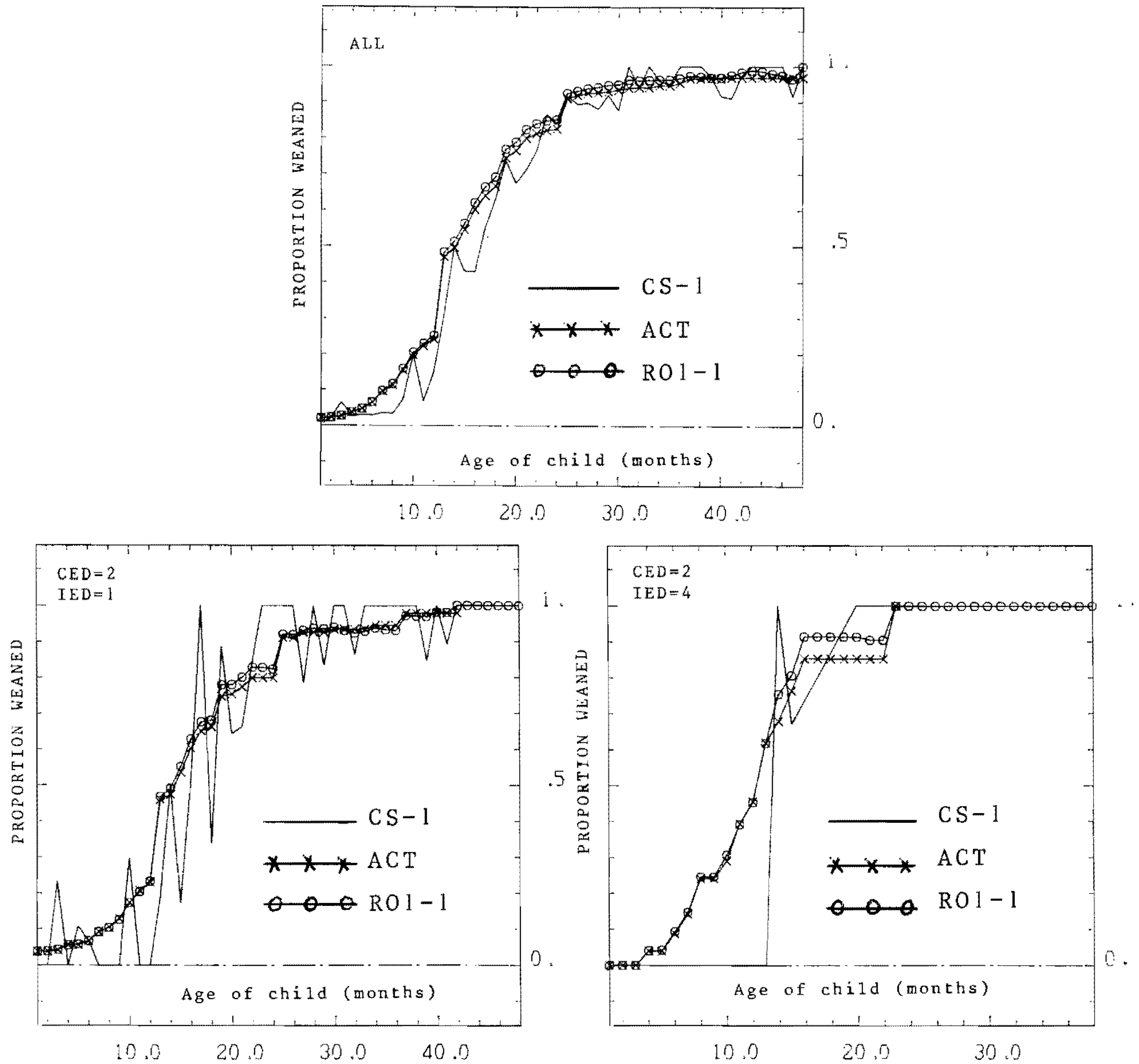
(*) See Table 3A.

TABLE 3C : LIFE TABLE ESTIMATES OF THE CUMULATIVE DISTRIBUTION $\pi(t)$ OF BREASTFEEDING DURATIONS FROM DATA ON THE TWO MOST RECENT PREGNANCIES OF MARRIED WOMEN IN CENTRAL AND EASTERN PROVINCES, AND WITH CED=2 AND IED=4 (*)

	CS-1 METHOD			ACT METHOD					RO1-1 METHOD		
t	n_t	d'_{tt}	$p_t^{(1)}$	N_{t-1}	m_{+t-1}	w_{t-1}	${}_1p_{t-1}$	$p_t^{(2)}$	n'_t	d'_t	$p_t^{(3)}$
1	4.585	0.000	0.000	32.199	0.000	–	1.000	0.000	32.199	0.000	0.000
2	1.141	0.000	0.000	32.199	0.000	4.585	1.000	0.000	27.614	0.000	0.000
3	1.020	0.000	0.000	27.614	0.000	1.141	1.000	0.000	26.473	0.000	0.000
4	1.020	0.000	0.000	26.473	1.054	1.020	.959	.041	25.453	1.054	.041
5	1.164	0.000	0.000	24.399	0.000	1.020	1.000	.041	24.433	1.054	.043
6	–	–	–	23.379	1.141	1.164	.950	.089	23.269	2.195	.094
7	–	–	–	21.074	1.226	0.000	.942	.142	23.269	3.421	.147
8	–	–	–	19.848	2.285	0.000	.885	.240	23.269	5.706	.245
9	1.053	0.000	0.000	17.563	0.000	0.000	1.000	.240	23.269	5.706	.245
10	2.170	0.000	0.000	17.563	1.090	1.053	.936	.289	22.216	6.796	.306
11	–	–	–	15.420	2.170	1.029	.854	.393	20.046	7.825	.390
12	–	–	–	12.221	1.239	0.000	.899	.454	20.046	9.064	.452
13	2.252	0.000	0.000	10.982	3.308	0.000	.699	.619	20.046	12.372	.617
14	1.141	1.141	1.000	7.674	1.020	2.252	.844	.678	17.794	13.392	.753
15	3.198	2.144	.670	4.402	1.164	0.000	.736	.763	16.653	13.415	.806
16	–	–	–	3.238	1.020	1.054	.624	.852	13.455	12.291	.913
17	–	–	–	1.164	0.000	0.000	1.000	.852	13.455	12.291	.913
18	–	–	–	1.164	0.000	0.000	1.000	.852	13.455	12.291	.913
19	–	–	–	1.164	0.000	0.000	1.000	.852	13.455	12.291	.913
20	1.239	1.239	1.000	1.164	0.000	0.000	1.000	.852	13.455	12.291	.913
21	–	–	–	1.164	0.000	0.000	1.000	.852	12.216	11.052	.905
22	–	–	–	1.164	0.000	0.000	1.000	.852	12.216	11.052	.905
23	2.390	2.390	1.000	1.164	1.164	0.000	0.000	1.000	12.216	12.216	1.000
24	1.053	1.053	1.000	–	–	–	–	–	9.826	9.826	1.000
25	1.053	1.053	1.000	–	–	–	–	–	8.773	8.773	1.000
26	–	–	–	–	–	–	–	–	7.720	7.720	1.000
27	1.029	1.029	1.000	–	–	–	–	–	7.720	7.720	1.000
28	–	–	–	–	–	–	–	–	6.691	6.691	1.000
29	3.622	3.622	1.000	–	–	–	–	–	6.691	6.691	1.000
30	–	–	–	–	–	–	–	–	3.069	3.069	1.000
31	1.029	1.029	1.000	–	–	–	–	–	3.069	3.069	1.000
32	–	–	–	–	–	–	–	–	2.040	2.040	1.000
33	–	–	–	–	–	–	–	–	2.040	2.040	1.000
34	–	–	–	–	–	–	–	–	2.040	2.040	1.000
35	–	–	–	–	–	–	–	–	2.040	2.040	1.000
36	–	–	–	–	–	–	–	–	2.040	2.040	1.000
37	1.020	1.020	1.000	–	–	–	–	–	2.040	2.040	1.000
38	1.020	1.020	1.000	–	–	–	–	–	1.020	1.020	1.000
39	–	–	–	–	–	–	–	–	–	–	–
40	–	–	–	–	–	–	–	–	–	–	–
41	–	–	–	–	–	–	–	–	–	–	–
42	–	–	–	–	–	–	–	–	–	–	–
43	–	–	–	–	–	–	–	–	–	–	–
44	–	–	–	–	–	–	–	–	–	–	–
45	–	–	–	–	–	–	–	–	–	–	–
46	–	–	–	–	–	–	–	–	–	–	–
47	–	–	–	–	–	–	–	–	–	–	–
48	–	–	–	–	–	–	–	–	–	–	–

(*) See Table 3A.

FIG. 8 : LIFE TABLE ESTIMATES OF THE CUMULATIVE DISTRIBUTION $\pi(t;z)$ OF BREASTFEEDING DURATIONS FROM DATA ON THE TWO MOST RECENT PREGNANCIES OF MARRIED WOMEN IN CENTRAL AND EASTERN PROVINCES, OR IN SUBSAMPLES z SPECIFIED BY COVARIATES CED AND IED.



children born within 4 years before the survey, and not born in the month of the survey. Thus, children's age could vary from 1 to 48 months.

Implications - such as selection biases - of the use of breastfeeding data restricted to the two most recent pregnancies are for instance discussed in Page et al (1982). In this paper, however, we are forced to use data on the two most recent pregnancies only. The reason is that the KFS does not provide breastfeeding durations for other pregnancies, and a comparison is to be made here between CS and R01 life table approaches. A R01 life table approach is only possible if breastfeeding durations are available for all births included in the analysis.

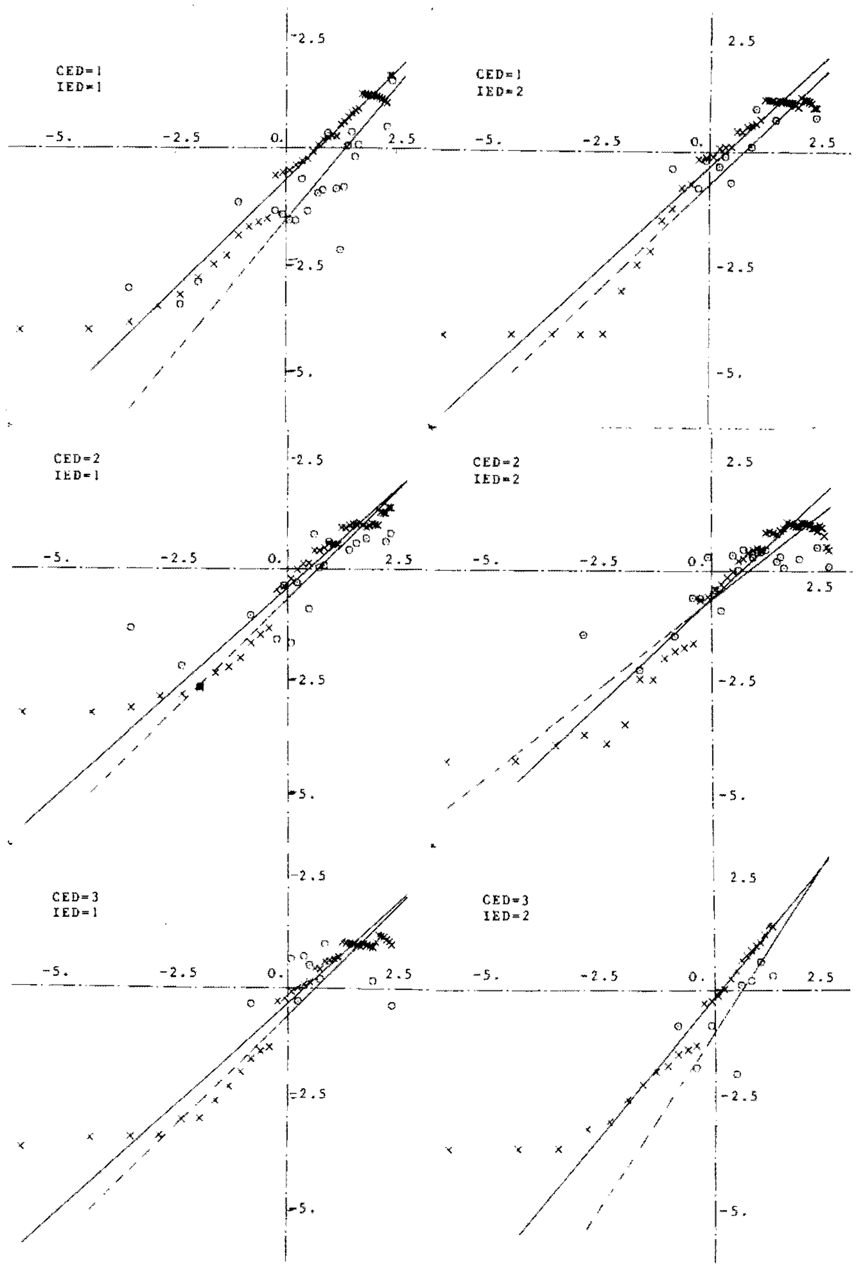
The KFS is not self-weighting and provides a sample weight variable (V006). Our analyses used weighted data : in general, the numbers n_{zt} and d_{zt} in sections 3.1-2 are weighted numbers (i.e. sums of individual weights). Similarly, all numbers n, m, w, d and N (with appropriate subscripts and/or apostrophes) are weighted numbers or sums of individual weights. The δ's used in section 3 should be multiplied with the appropriate individual weight. Weighted sample sizes are shown in Table 2.

TABLE 2 : WEIGHTED NUMBER OF BIRTHS IN SUBSAMPLES DEFINED BY COVARIATES CED AND IED

		IED				
		1	2	3	4	TOTAL
CED	1	283.012	80.652	89.783	16.333	469.780
	2	337.581	245.706	246.309	32.199	861.795
	3	161.228	163.358	193.106	42.068	559.760
TOTAL		781.821	489.716	529.198	90.600	1891.335

The weaning of a child is a non-renewable event, and hence the theory of section 3.3 is applicable. Thus, we constructed for each particular subsample of births (defined by covariates CED and/or IED) a cross-tabulation as shown in Fig.1, and then we computed CS, ACT and R01 life tables. CS-1 (formula [8]), ACT (formula [11]) and R01-1 (formula [12]) life tables are shown in Table 3A for the whole of central and Eastern provinces, in Table 3B for the subsample specified as CED=2 and IED=1 (the largest one), and in Table 3C for the subsample specified as CED=2 and IED=4 (one of the smallest). The life table estimates of $\pi(t)$ are also plotted in Fig. 8.

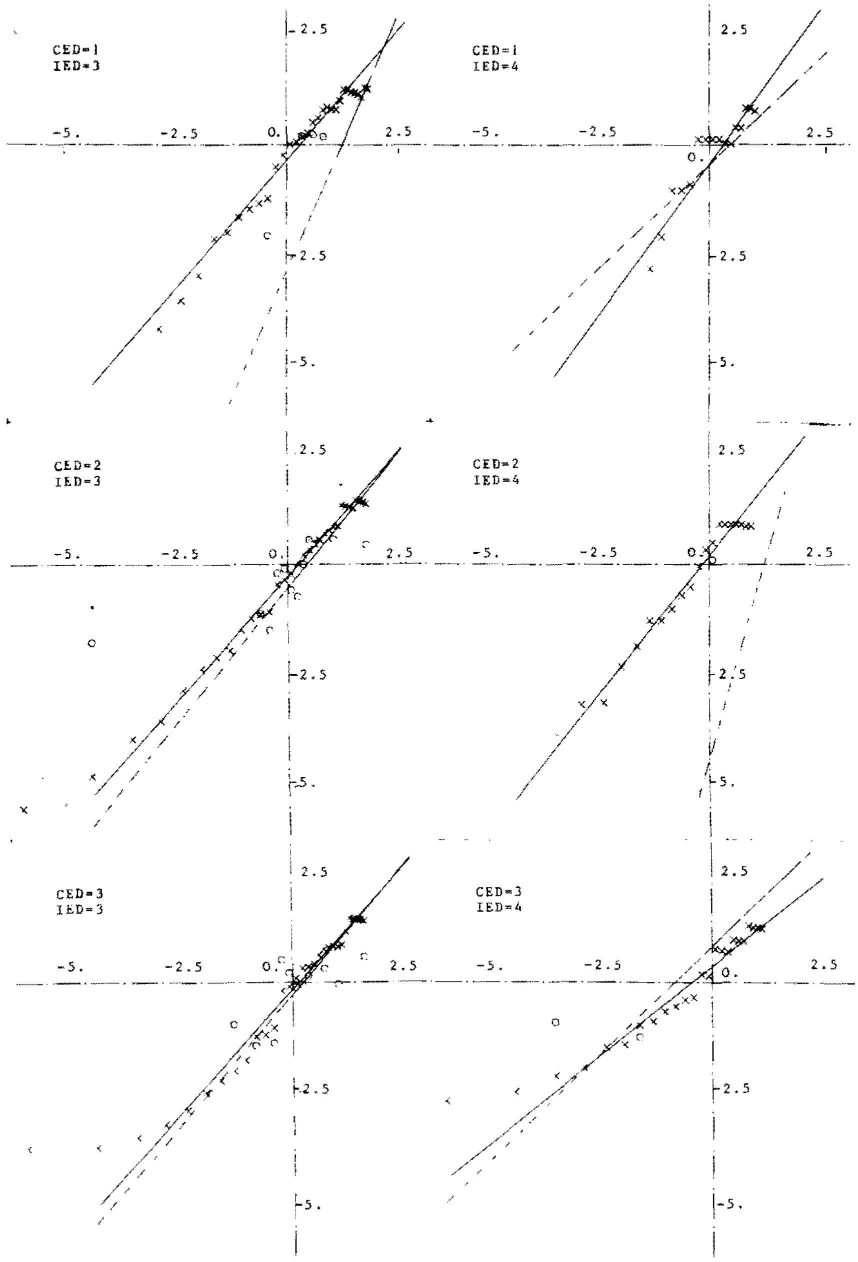
FIG. 9 : COMPLEMENTARY LOG-LOG TRANSFORMS OF BREASTFEEDING SCHEDULES $\pi(t;z)$ AGAINST THE COMPLEMENTARY LOG-LOG TRANSFORM OF THE PARAMETRIC STANDARD $\hat{\pi}_s(t)$, FOR SUB-SAMPLES SPECIFIED BY CED AND IED(x)



(x) X X X RO1-1 OBSERVED; O O O CS-1 OBSERVED;

——— RO1-1 ESTIMATED;----- CS-1 ESTIMATED.

FIG. 9 : continued



From the tables and the figures, we can see that the RO1-1 life table estimate is close to the ACT life table estimate if the sample size is large enough and/or if the number of censored cases is small relative to the total number of cases. Both RO1-1 and ACT estimated curves are steep between ages 12 and 13, 18 and 19, 24 and 25, 36 and 37. This shows the effect of heaping of retrospectively reported breastfeeding durations at multiples of 6 months. This effect is not seen in CS estimates. However, CS estimated curves are much more irregular as a consequence of sample fragmentation at different ages t.

5.3. Analysis by GLIRM's

The generalized linear relational models for breastfeeding durations as discussed in section 5.1, can now be fitted to both CS-1 and RO1-1 data.

Whether the models will fit adequately, can to some extent be examined graphically. Therefore, we have plotted in Fig. 9 the complementary log-log transformations of the observed CS-1 and RO1-1 life table estimates of $\pi(t;z)$ for each of the 12 subsamples z defined by covariates CED and IED against the complementary log-log transformation of the Weibull standard $\hat{\pi}_s(t)$. For instance, for the subsample CED=2, IED=4 we computed the transformations by Φ of the observed cumulative proportions $p_t^{(1)}$ and $p_t^{(3)}$ shown in Table 3C and plotted these against $\Phi(\hat{\pi}_s(t))$ in Fig. 9h(14). The plotted points should be approximately on a straight line, in order to have a good fit of the proposed model. Fig. 9 generally shows such linearity. Discrepancies are found mainly at the tails of the distributions (i.e. at low or at high ages) and for CS estimates, which are more irregular by their nature.

Thus, Fig. 9 indicates that the GLIRM [18d] may be fitted, but we can also expect that the fit is better for RO1-1 than for CS-1 data. Table 4 shows some results if model [18d] is applied to each subsample separately - where a subsample is defined by covariates CED and/or IED (or none of them if the whole sample is considered). The regression lines are plotted in Fig. 9 for each subsample defined by CED and IED. From the figures and the table, we see that the fit is generally better for RO1-1 than for CS-1 data : estimates from CS-1 data are, for instance, more distorted by sample fragmentation. We can

---

(14) If an observed proportion is 0. or 1., the corresponding point is not shown in the figure since $\Phi(0.)=-\infty$ and $\Phi(1.)=+\infty$.

also see that both CED and IED have a negative effect on the duration of breastfeeding, with a larger effect for IED than for CED. A combined effect of CED and IED (i.e. an interaction effect) seems to be absent.

The results shown in Table 4 could also have been found by fitting the multiple regression models CED*IED*SST, CED*SST, IED*SST and SST to both R01-1 (Panel A) and CS-1 (Panel B) data. Other models can then be fitted too, and they can be compared with the above ones. Since both CED and IED have a significant effect on the duration of breastfeeding - whether or not we have controlled for the other variable - we started with the "full" model CED*IED*SST and tried to simplify this model. Since R01-1 data are smoother than CS-1 data we shall first discuss the selection of a parsimonious model for R01-1 data.

A comparison with the model (CED+IED)*SST showed that interaction effects of CED and IED (on both slopes and intercepts) were not significant : the largest difference between corresponding estimated means was less than 2.20 months, and this occurred for the subsample with the smallest size (i.e. CED=1, IED=4). Model (CED+IED)*SST has then been compared to model SST+(CED+IED) : the estimated means all differed less than .20 months, showing that the effects of covariates CED and IED on the slopes were not significant. Then, a comparison of model SST+(CED+IED) with SST+(CED*IED) again showed that the interaction effect of covariates CED and IED (on the intercept) was not significant. Next, noting that the (overall) slope in the model SST+(CED+IED) was estimated as .995 with standard error .009, we could obviously fit proportional hazards (PH) models. A comparison of SST+(CED+IED) with the PH model CED+IED showed no significant differences. Further, we found also that the interaction effect was not significant in the PH model, but that main effects of CED and IED were both significant (whether or not controlled for one variable). Hence, a parsimonious model for R01-1 breastfeeding data is the additive PH model CED+IED. Results from the fit of this model are shown in Table 5, Panel A. This table gives also results from the fit of the PH models CED, IED and the one wherein no effects of covariates are estimated. Notice that, as demonstrated in section 5.1, the PH models are AFT models as well.

Following the same procedure, we can find a parsimonious model for CS-1 data too. We found that generally the same remarks as for R01-1 data had to be made. Thus, the PH (or AFT) model CED+IED adequately fitted to CS-1 data. Results from the fit of this model and other PH (or AFT) models to CS-1 data are shown in Table 5, Panel B.

TABLE 4 : FITTING MODELS CEDxIEDxSST, CEDxSST, IEDxSST AND SST TO BREASTFEEDING DATA

PANEL A : RESULTS FROM FITS ON ROI-1 DATA

	IED=1	IED=2	IED=3	IED=4	NOT CONTROLLED FOR IED
CED=1	.992(a) -.710(b) 17.74 (c) 8.54 (d)	.946 -.340 15.10 7.59	1.165 -.359 14.79 6.16	1.388 -.430 14.90 5.31	.993 -.575 16.67 8.02
CED=2	.910 -.425 15.86 8.25	.944 -.628 17.33 8.72	1.130 -.279 14.37 6.16	1.270 .191 12.05 4.65	.948 -.438 15.82 7.94
CED=3	.925 -.340 15.61 8.00	1.220 -.170 13.72 5.49	1.108 -.166 13.75 5.99	.808 .367 10.50 6.08	1.004 -.209 14.10 6.72
NOT CONTROLLED FOR CED	.937 -.521 16.50 8.36	.971 -.437 15.73 7.72	1.126 -.250 14.21 6.11	.985 .194 11.74 5.69	.964 -.406 15.52 7.67

PANEL B : RESULTS FROM FITS ON CS-1 DATA

	IED=1	IED=2	IED=3	IED=4	NOT CONTROLLED FOR IED
CED=1	1.217 -1.611 23.45 9.40	.974 -.732 18.02 8.82	2.315 -2.884 23.41 5.25	.955 -.447 15.86 7.90	1.115 -1.195 20.88 9.05
CED=2	.994 -.673 17.43 8.37	.792 -.654 18.76 11.06	1.220 -.535 15.71 6.28	3.582 -4.389 23.74 3.56	.937 -.607 17.21 8.72
CED=3	1.003 -.610 16.89 8.05	1.543 -.947 17.18 5.57	1.164 -.283 14.36 5.99	.973 .811 8.80 4.31	1.050 -.412 15.33 7.01
NOT CONTROLLED FOR CED	1.031 -.961 19.56 9.10	.928 -.701 18.06 9.24	1.289 -.507 15.42 5.87	.985 .162 11.91 5.77	.980 -.681 17.58 8.56

NOTES : (a) estimated slope $\hat{\theta}$; (b) estimated intercept $\hat{\beta}$; (c) estimated mean duration of breastfeeding; (d) estimated standard deviation of duration of breastfeeding.

TABLE 5 : FITTING PH (AFT) MODELS CED+IED, CED, IED AND THE PH (AFT) MODEL WITHOUT COVARIATES TO BREASTFEEDING DATA

PANEL A : RESULTS FROM FITS ON ROI-1 DATA

	IED=1	IED=2	IED=3	IED=4	NOT CONTROLLED FOR IED
CED=1	-.657(a) 17.27 (b) 8.26 (c) .743(d) .518(e)	-.588 16.74 8.00 .767 .556	-.365 15.14 7.23 .848 .694	.038 12.61 6.03 1.017 1.038	-.580 16.68 7.97 .769 .560
CED=2	-.576 16.65 7.96 .771 .562	-.507 16.14 7.71 .795 .602	-.284 14.59 6.97 .879 .753	.118 12.16 5.81 1.055 1.126	-.468 15.86 7.58 .809 .626
CED=3	-.359 15.09 7.21 .850 .699	-.289 14.63 6.99 .877 .749	-.067 13.22 6.32 .970 .935	.336 11.02 5.27 1.164 1.399	-.208 14.10 6.74 .910 .812
NOT CONTROLLED FOR CED	-.562 16.55 7.91 .775 .570	-.454 15.76 7.53 .814 .635	-.216 14.15 6.76 .907 .806	.196 11.74 5.61 1.093 1.216	-.425 15.55 7.43 .825 .654

PANEL B : RESULTS FROM FITS ON CS-1 DATA

	IED=1	IED=2	IED=3	IED=4	NOT CONTROLLED FOR IED
CED=1	-1.152 21.61 10.33 .594 .316	-1.121 21.30 10.18 .602 .326	-.678 17.44 8.33 .736 .508	-.180 13.92 6.65 .922 .835	-1.048 20.61 9.85 .623 .351
CED=2	-.835 18.72 8.95 .685 .434	-.804 18.46 8.82 .695 .447	-.361 15.11 7.22 .849 .697	.136 12.06 5.77 1.063 1.146	-.674 17.41 8.32 .737 .510
CED=3	-.587 16.73 8.00 .767 .556	-.556 16.50 7.89 .778 .574	-.113 13.50 6.45 .950 .893	.385 10.78 5.15 1.190 1.469	-.373 15.19 7.26 .845 .689
NOT CONTROLLED FOR CED	-.921 19.47 9.30 .659 .398	-.785 18.31 8.75 .701 .456	-.331 14.90 7.12 .861 .718	.160 11.94 5.71 1.075 1.173	-.703 17.63 8.43 .728 .495

NOTES : (a) estimated intercept $\hat{\beta}$; (b) estimated mean duration of breastfeeding; (c) estimated standard deviation of duration of breastfeeding; (d) estimated acceleration factor $\exp(\hat{\beta}/\hat{\alpha}_2)$; (e) estimated relative risk $\exp(\hat{\beta})$

Both Table 4 and Table 5 show systematic differences between estimates from RO1-1 data and the corresponding estimates from CS-1 data. The differences between the estimated mean durations of breastfeeding are shown in Table 6.

TABLE 6 : DIFFERENCES BETWEEN ESTIMATED MEAN DURATIONS OF BREASTFEEDING FROM CS-1 AND RO1-1 DATA, AFTER FITTING PH MODELS

	IED=1	IED=2	IED=3	IED=4	NOT CONTROLLED FOR IED
CED=1	4.34	4.56	2.30	1.31	3.93
CED=2	2.07	2.32	0.52	-0.10	1.55
CED=3	1.64	1.87	0.28	-0.24	1.09
NOT CONTROLLED FOR CED	2.92	2.55	0.75	0.20	2.08

The higher estimated mean durations of breastfeeding for CS-1 data can be explained as follows. Estimates of $\pi(t)$ at low ages t are based on recent births and hence on short periods only, while estimates of $\pi(t)$ at higher ages t are based on earlier births and hence on longer periods. Therefore, errors in misreporting ages may contribute relatively more in estimating $\pi(t)$ at low ages t than estimating $\pi(t)$ at high ages t. For instance, children with low ages but already being weaned may be shifted to higher ages, and hence the proportion of weaned children is underestimated at low ages. This leads to an upwards bias of the mean duration of breastfeeding.
RO1-1 data seems to reduce this bias, which can be explained by the simple fact that the estimate of $\pi(t)$ at low ages t is based on both recent and earlier births, hence on short and longer periods (cfr. Fig. 2). Notice that the overestimation of mean durations of breastfeeding from CS-1 data is large for less-educated women, but decreases with increasing level of education - which seems to be quite reasonable.

6. APPLICATION TO NUPTIALITY DATA

6.1. The model

The model for analysis of age of entry into first marriage is specified as follows :

1. The transformation of the model is

$$\Phi(\pi) = F_w^{-1}(\pi) \qquad [20a]$$

$$\text{where } F_w(w) = \int_{-\infty}^{w} \frac{m_2^{\,m_2}}{\Gamma(m_2)} e^{-m_2(u+e^{-u})} du \qquad [20b]$$

2. The transformed standard schedule is a linear function of age t :

$$\Phi(\pi_s(t)) = \alpha_1 + \alpha_2 . t \qquad [20c]$$

Formula [20b] is the limit case $m_1 = +\infty$ of [2b]; Φ is the inverse cumulative distribution function of an exponential reciprocal gamma variate (Appendix Table A.1). Formula [20c] can be obtained from [3] by taking $\alpha_4 = 1$ and $\alpha_3 = 1$. From [20a-c] it follows that age of entry into first marriage t has an exponential reciprocal gamma distribution. The model is precisely the Coale-McNeil model. The construction of the model is presented in Coale and McNeil (1972), Rodriguez and Trussell (1980) have discussed its application to WFS-surveys and Vanderhoeft (1983) reformulated it as a generalized relational model. The reader is referred to those papers for more technical details.

The multiple regression model for nuptiality becomes :

$$\Phi(\pi^*(t;z)) = (\theta.z')(\alpha_1 + \alpha_2.t) + \beta.z' \qquad [20d]$$

or

$$\Phi(\pi(t;z)/C(z)) = (\theta.z')(\alpha_1 + \alpha_2 t) + \beta.z' \qquad [20e]$$

The original formulation of the model was done through its proper density function :

$$f^{*}(t;z) = \frac{\lambda(z)}{\Gamma(\alpha(z)/\lambda(z))} e^{-\alpha(z)(t-\xi(z))-e^{-\lambda(z)(t-\xi(z))}} \quad [21]$$

where $f^{*}(t;z) = \frac{d}{dt} \pi^{*}(t;z) = \frac{1}{C(z)} f(t;z)$

It can be shown that the original and present formulation are related as follows :

$$\begin{cases} m_2 = \alpha(z)/\lambda(z) \\ \lambda(z) = \alpha_2 \theta.z' \\ \xi(z) = (\log m_2 - \alpha_1 \theta.z' - \beta.z')/(\alpha_2 \theta.z') \end{cases} \quad [22]$$

A standard schedule of nuptiality is defined through [20a-c] and values of the parameters m_2, α_1 and α_2. From Rodriguez and Trussell (1980), we obtained the estimates m_2=.604, $\hat{\alpha}_1$=-2.2495 and $\hat{\alpha}_2$=.288 (and c_s=1). This is equivalent to λ_s=.288, α_s=.174 and ξ_s=6.06. Those parameter values specify a standard $\hat{\pi}_s(t)$ with mean 11.36 and variance 43.36, which is close to an emperically derived standard from Swedish data recorded between 1865 and 1869 (Coale and McNeil, 1972).

From Appendix Table A.2 it follows that the cumulative distribution functions satisfy the relation for "translated-accelerated failure time" (TAFT) models :

$$\pi^{*}(t;z) = \pi_s\left(\frac{t-a_0(z)}{k(z)}\right) \quad [23]$$

where new parameters $a_0(z)$ and $k(z)$ are related to the parameters of the generalized relational model through the equations :

$$\begin{cases} k(z) = 1/(\theta.z') \\ a_0(z) = \frac{\alpha_1}{\alpha_2}\left(\frac{1}{\theta.z'} - 1\right) - \frac{\beta.z'}{\alpha_2 \theta.z'} \end{cases} \quad [24]$$

The parameters a_0, k and C were originally used by Coale (1971) in his model schedules of nuptiality. They can thus computed very easily after fitting the generalized relational model [20d-e].

The model as presented above involves essentially 4 parameters m_2, C, θ and β. However, as noted by Rodriguez and Trussell (1980) the parameter m_2 (which is the ratio a/λ) should be constant. Since its value for the standard is .604, m_2 takes this value for any population. Hence, the model involves only 3 parameters C, θ and β, and those are equivalent to three parameters C, a_0 and k used by Coale (1971). The difference between the approach in Coale (1971) and the present approach (being the Coale-McNeil (1972) model) is the distribution of age at first marriage : the former does not rely on a specific distribution, while the latter relies on a special mathematical distribution function. In fact, it follows from Coale (1971) that this mathematical distribution function is common to all populations (i.e. [20a-b] with w=t and constant m_2=.604). It is shown in that paper that the same underlying law for entry into first marriage is applicable to a large variety of populations.

6.2. Data and life tables

As for the analysis of breastfeeding, the data on nuptiality are extracted from the KFS. Again, women living in central and Eastern provinces only are used. Both ever married and single women are included in the analysis in order to have untruncated data (section 3.3.6), making possible a comparison between CS and ROI approaches. Three women with recorded age at first marriage (V109) of 9 completed years were excluded. Hence, the minimum age at first marriage is 10 completed years.

Covariates CED and IED are defined as in section 5.2. The data are again weighted with the sample weight variable V006. Weighted sample sizes are shown in Table 7.

TABLE 7 : WEIGHTED NUMBER OF WOMEN IN SUBSAMPLES DEFINED BY CED AND/OR IED

		IED				
		1	2	3	4	TOTAL
	1	355.891	110.124	156.349	35.306	657.670
CED	2	392.803	278.201	406.799	94.662	1172.465
	3	185.635	178.027	290.037	126.538	780.237
TOTAL		934.329	566.352	853.185	256.506	2610.372

As in the analyses of breastfeeding (section 5), we have constructed cross-tabulations for each subsample defined by CED and/or IED. From those cross-tabulations we have computed CS-1, RO1-1 and ACT life table estimates of the cumulative distributions of nuptiality $\pi(t;z)$, where z depends on CED and/or IED. Since entry into first marriage is a non-renewable event, the theory behind the above computations can be found in section 3.3. Together with life table estimates of $\pi(t;z)$, we obtained, of course, also CS-1 and RO1-1 data, appropriate for analysis through GESLIRM's (or GLIRM's).

The life table estimates of $\pi(t)$, which is the cumulated nuptiality schedule for the whole sample of observations, are plotted in Fig. 10. It is seen that ACT and RO1-1 life table estimates are very close, the RO1-1 estimate being somewhat higher than the ACT estimate near the centre of the distribution. The CS-1 estimate is lower than both the ACT and the RO1-1 estimate. Assuming that women have reported their status of first marriage at the time of the interview without error, the low CS-1 estimates at low ages may reflect misreporting of ages itself. Particularly, ever married women may tend to overestimate their ages.

FIG. 10 : LIFE TABLE ESTIMATES OF THE CUMULATED NUPTIALITY SCHEDULE $\pi(t)$ IN CENTRAL AND EASTERN PROVINCES (KFS)

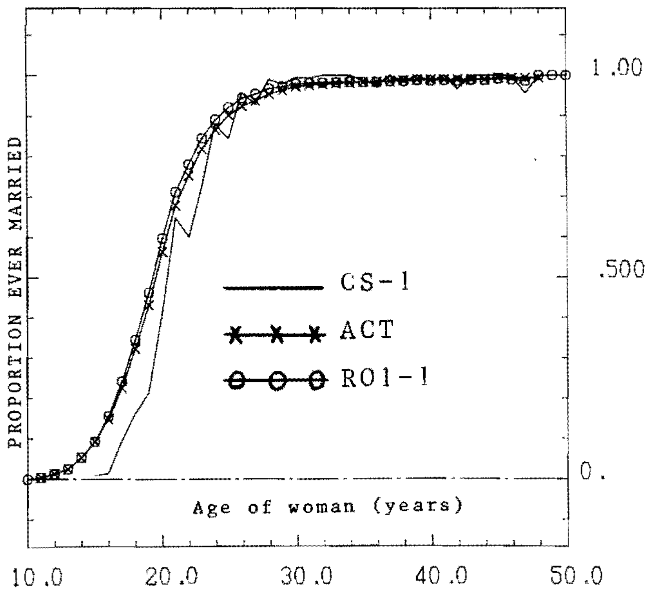


Fig. 10 also indicates that almost all women residing in Central and Eastern provinces are ever married or will ultimately marry.

We made similar plots for all subsamples defined by CED and/or IED. Those for subsamples CED=2, IED=4 and CED=3, IED=1 are shown in Fig. 11. If the sample size was large enough (see Table 7), generally the same conclusions as for the whole sample could be drawn from these plots. RO1-1 and ACT estimates are, for instance, almost the same in subsample CED=3, IED=1 (Fig. 11). This was due to the small number of censorings in that sample (cfr. section 3.3.6) : all women, except 1.208 with current age 16 and 1.812 with current age 24, had ever been married. The ultimate proportion of married women in this subsample seemed to be 1. as well.

FIG. 11 : LIFE TABLE ESTIMATES OF THE CUMULATED NUPTIALITY SCHEDULE $\pi(t;z)$ FOR SUBSAMPLES z, SPECIFIED BY CED AND IED, OF WOMEN RESIDING IN CENTRAL AND EASTERN PROVINCES (KFS)

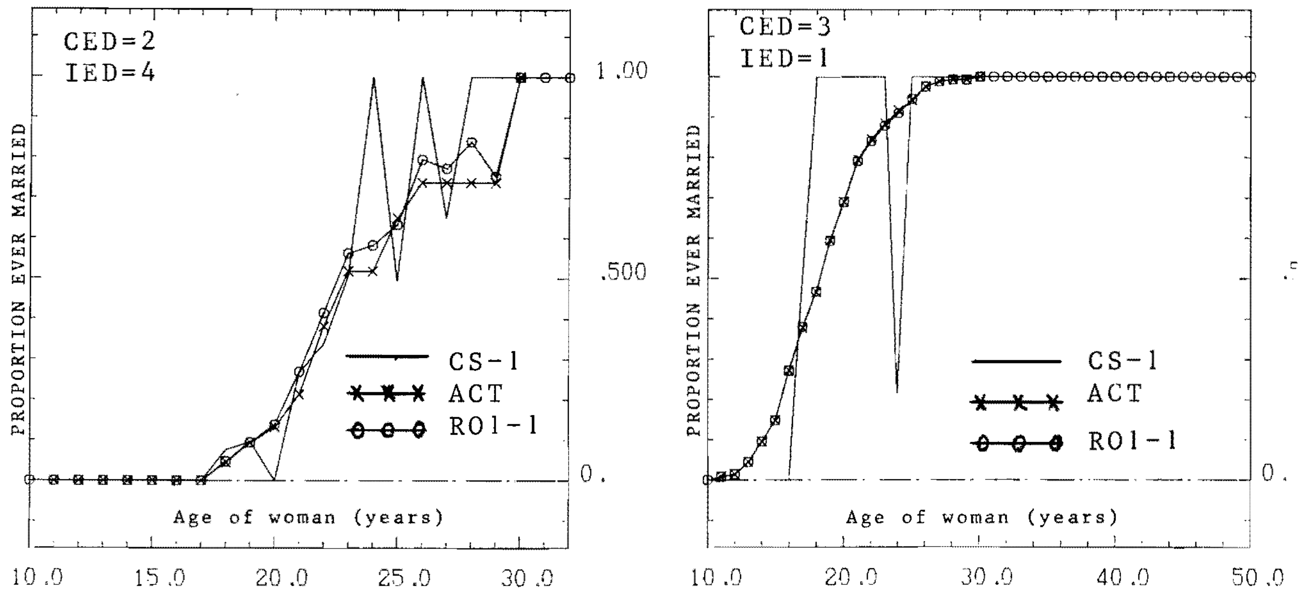


Fig. 11 also shows the life table estimates for subsample CED=2, IED=4, which has sample size 94.662. For such small subsamples, the estimated schedules were more irregular. However, the RO1-1 and ACT estimated schedules followed closer the CS-1 estimated schedule, in the sense that the CS-1 schedule is no longer systematically below the RO1-1 and ACT schedule. In subsample CED=2, IED=4 this is clearly due to the lack of women with current age higher than 32 years : the RO1-1 estimate at low ages depends on women with higher ages too (cfr. Fig. 2), but here we didn't have older women.

Estimation of life tables for subsample CED=3, IED=1 made clear that CS data can be very poor, while RO1 data can contain enough information for a reliable estimation. Hence, use of RO1 data in multiple regression may be quite advantageous.

6.3. Analysis by GESLIRM's (or GLIRM's)

The models to be fitted to nuptiality data essentially involve 3 parameters : slopes θ, intercepts β and ultimate proportions C. Hence, we are dealing with GESLIRM's. However, visual investigation of life table estimates (section 6.2) indicates that the ultimate proportions C can be given a fixed value 1. Thus, in [20e] we take $C(z)=1$, implying $\pi(t;z)=\pi^{*}(t;z)$ and GLIRM's for $\pi(t;z)$ itself. These models could easily been fitted as they involved only 2 linear parameters $\theta(z)$ and $\beta(z)$.

First we fitted the models CED*IED*SST, CED*SST, IED*SST and SST to both RO1-1 and CS-1 nuptiality data. The results are shown in Table 8.

Comparison of SST+CED with CED*SST and SST+CED+IED with CED*SST+IED for both RO1-1 and CS-1 data has shown no significant difference between corresponding estimated means. Thus, the effect of CED on the slopes can be ignored. Comparison of SST+IED with IED*SST and SST+CED+IED with IED*SST+CED was shown no significant difference between corresponding estimated means for CS-1 data, but a significant difference for RO1-1 data. Thus, the effect of IED on slopes cannot be ignored. Since both CED and IED seem to have significant effects on the intercept, we conclude that at least the model IED*SST+CED should be fitted. For RO1-1 data, models IED*SST+CED and IED*SST+CED*IED showed only small differences between corresponding estimated means. For CS-1 data, those models showed some large differences between the estimated means. This was, for instance, the case in subsample CED=3, IED=1 for which the estimated mean in the latter model was 16.4 years, compared with 19.7 years in the former model. However, Fig. 11 has shown that the CS-1 data for that subsample was very poor, which explains the low estimate for that subsample (see also the corresponding estimate of 17.64 years in Table 8). Hence, we conclude that the interaction effect of CED and IED on the intercept can be ignored.

A parsimonious model for both RO1-1 and CS-1 data is thus the model IED*SST+CED. Results for this model are shown in Table 9.

TABLE 8 : FITTING MODELS CED*IED*SST, CEDxSST, IEDxSST AND SST TO NUPTIALITY DATA

PANEL A : RESULTS FROM FITS ON RO1-1 DATA

	IED=1	IED=2	IED=3	IED=4	NOT CONTROLLED FOR IED
CED=1	1.354(a)	1.544	2.081	1.557	1.404
	.162(b)	.128	-.978	-1.494	-.033
	19.02 (c)	18.82	20.15	22.42	19.42
	4.87 (d)	4.26	3.16	4.23	4.69
CED=2	1.295	1.548	1.904	1.452	1.422
	.125	-.031	-.942	-1.938	-.217
	19.22	19.17	20.39	23.89	19.84
	5.08	4.25	3.46	4.54	4.63
CED=3	1.626	1.555	1.284	.957	1.418
	.263	.002	-.615	-1.139	-.258
	18.43	19.09	21.24	24.65	19.95
	4.05	4.23	5.13	6.88	4.65
NOT CONTROLLED FOR CED	1.373	1.543	1.586	1.049	1.413
	.163	.009	-.761	-1.296	-.179
	18.98	19.09	20.71	24.49	19.76
	4.80	4.27	4.15	6.28	4.66

PANEL B : RESULTS FROM FITS ON CS-1 DATA

	IED=1	IED=2	IED=3	IED=4	NOT CONTROLLED FOR IED
CED=1	1.788	2.009	2.292	1.966	1.961
	-.415	-1.016	-1.627	-2.330	-1.202
	19.60	20.33	20.83	22.73	20.75
	3.68	3.28	2.87	3.35	3.36
CED=2	1.410	1.568	2.082	1.459	1.678
	-.501	-.370	-1.435	-1.979	-1.155
	20.56	19.89	20.91	23.95	21.31
	4.67	4.20	3.16	4.51	3.92
CED=3	2.085	2.266	1.565	1.681	1.707
	.521	-1.061	-1.008	-2.458	-1.303
	17.64	20.00	21.32	24.00	21.54
	3.16	2.91	4.21	3.92	3.86
NOT CONTROLLED FOR CED	1.597	1.806	1.892	1.590	1.743
	-.346	-.687	-1.295	-2.185	-1.208
	19.78	20.10	21.06	23.82	21.25
	4.12	3.65	3.48	4.14	3.78

NOTES : (a) estimated slope θ; (b) estimated intercept β;
(c) estimated mean age at first marriage;
(d) estimated standard deviation of age at first marriage.

TABLE 9 : FITTING MODELS IED*SST+CED, SST+CED AND SST+IED TO NUPTIALITY

PANEL A : RESULTS FROM FITS ON RO1-1 DATA

	IED=1	IED=2	IED=3	IED=4	SST+CED
CED=1	1.374(a)	1.547	1.593	1.057	1.416
	.183(b)	.027	-.751	-1.308	-.033
	18.93 (c)	19.04	20.67	24.47	19.40
	4.79 (d)	4.26	4.13	6.23	4.65
	10.66 (e)	11.70	13.54	13.72	11.38
CED=2	1.374	1.547	1.593	1.057	1.416
	.109	-.047	-.825	-1.382	-.216
	19.12	19.21	20.84	24.71	19.85
	4.79	4.26	4.13	6.23	4.65
	10.85	11.87	13.71	13.96	11.82
CED=3	1.374	1.547	1.593	1.057	1.416
	.239	.083	-.695	-1.252	-.259
	18.79	18.92	20.55	24.28	19.95
	4.79	4.26	4.13	6.23	4.65
	10.52	11.58	13.42	13.53	11.93
SST+IED	1.430	1.430	1.430	1.430	
	.166	.008	-.714	-1.432	
	18.89	19.27	21.02	22.77	
	4.60	4.60	4.60	4.60	
	10.95	11.33	13.08	14.83	

PANEL B : RESULTS FROM FITS ON CS-1 DATA

	IED=1	IED=2	IED=3	IED=4	SST+CED
CED=1	1.594	1.802	1.892	1.587	1.747
	-.372	-.723	-1.343	-2.234	-1.073
	19.85	20.17	21.15	23.93	20.98
	4.13	3.65	3.48	4.15	3.77
	12.72	13.87	15.15	16.78	14.47
CED=2	1.594	1.802	1.892	1.587	1.747
	-.328	-.679	-1.300	-2.191	-1.199
	19.75	20.09	21.07	23.84	21.23
	4.13	3.65	3.48	4.15	3.77
	12.63	13.78	15.07	16.68	14.72
CED=3	1.594	1.802	1.892	1.587	1.747
	-.289	-.640	-1.261	-2.152	-1.333
	19.67	20.01	21.00	23.75	21.49
	4.13	3.65	3.48	4.15	3.77
	12.54	13.71	15.00	16.60	14.99
SST+IED	1.765	1.765	1.765	1.765	
	-.525	-.669	-1.221	-2.370	
	19.85	20.14	21.22	23.48	
	3.73	3.73	3.73	3.73	
	13.42	13.70	14.79	17.05	

NOTES : (a)-(d) see Table 8;
(e) estimated origin $\hat{a}_0$

In fact, the effects of covariate IED may be due mainly to the differences between subsamples with IED=4 and IED=1,2 or 3. Since subsample IED=4 only contains less than 10% of the observed women, the effect of IED may be doubtful. If subsample IED=4 is omitted from the analysis, much simpler parsimonious models may be detected.

7. APPLICATION TO LIFE TIME FERTILITY DATA

7.1. The model

The model for analysis of life time fertility data is specified as follows :

1. The tranformation of the model is the inverse cumulative Gompertz distribution :

$$\Phi(\pi)=-\log(-\log\pi) \qquad [25a]$$

2. The (normalized) standard schedule $\pi_s(t)$ is derived by Booth (1979). It is shown in Table 10.

TABLE 10 : NON-PARAMETRIC STANDARD OF FERTILITY

AGE t	$\pi_s(t)$	AGE t	$\pi_s(t)$	AGE t	$\pi_s(t)$
10	-	24	.32829	38	.88354
11	.00000	25	.37731	39	.90816
12	.00000	26	.42597	40	.93019
13	.00002	27	.47371	41	.94925
14	.00045	28	.52013	42	.96480
15	.00313	29	.56517	43	.97698
16	.01168	30	.60861	44	.98591
17	.03043	31	.65016	45	.99188
18	.05826	32	.68968	46	.99555
19	.09428	33	.72722	47	.99782
20	.13584	34	.76275	48	.99915
21	.18187	35	.79618	49	.99982
22	.22993	36	.82751	50	-
23	.27897	37	.85663		

Formula [25a] can be found from [2a-b] by taking $m_2=1$ and the limit case $m_1=+\infty$ (cfr. Appendix Table A.1). Since a non-parametric standard schedule is used, t has no specific (mathematical) distribution. However, "standard-time" $\Phi(\pi_s(t))$ (section 2.3) has a Gompertz distribution, which follows from [a] and the relation

$$\Phi(\pi^*(t;z))=(\theta.z').\Phi(\pi_s(t))+\beta.z' \qquad [25b]$$

or $\Phi(\pi(t;z)/C(z))=(\theta.z').\Phi(\pi_s(t))+\beta.z'$ [25c]

The mean age at childbearing in the standard schedule is 28.28 years, and the variance of age at childbearing t is 52.70.

7.2. Data and life tables

The data on life time fertility are, as those on breast-feeding and on nuptiality, extracted from the KFS. The same women sample as for nuptiality analyses is used here, except that now also the three women with reported age at first marriage of 9 years are included.

The data are again weighted with the sample weight variable V006. Covariates CED and IED are defined as in section 5.2.

Since we are dealing with a renewable event, the techniques of section 3.4 were used for the construction of cross-tabulations, life tables and CS-1 and RO1-1 data. The CS-1 and RO1-1 life tables for the whole sample are shown in Table 11. Notice that ACT life table estimates could not have been computed, since several age-cohorts of women had to be pooled (section 3.4.4). The CS-1 and RO1-1 life table estimates of the cumulated fertility in the whole sample of women residing in Central and Eastern provinces are plotted in Fig. 12. The estimates of the cumulated fertility for women in the sub-samples defined as IED=1, IED=2, IED=3 and IED=4 are plotted in Fig. 13.

FIG. 12 : LIFE TABLE ESTIMATES OF THE CUMULATED FERTILITY SCHEDULE $\pi(t)$ FOR WOMEN RESIDING IN CENTRAL AND EASTERN PROVINCES (KFS)

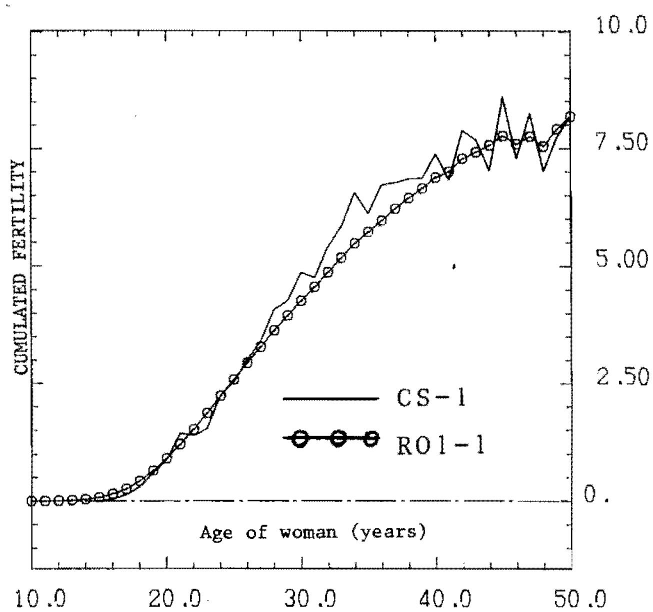
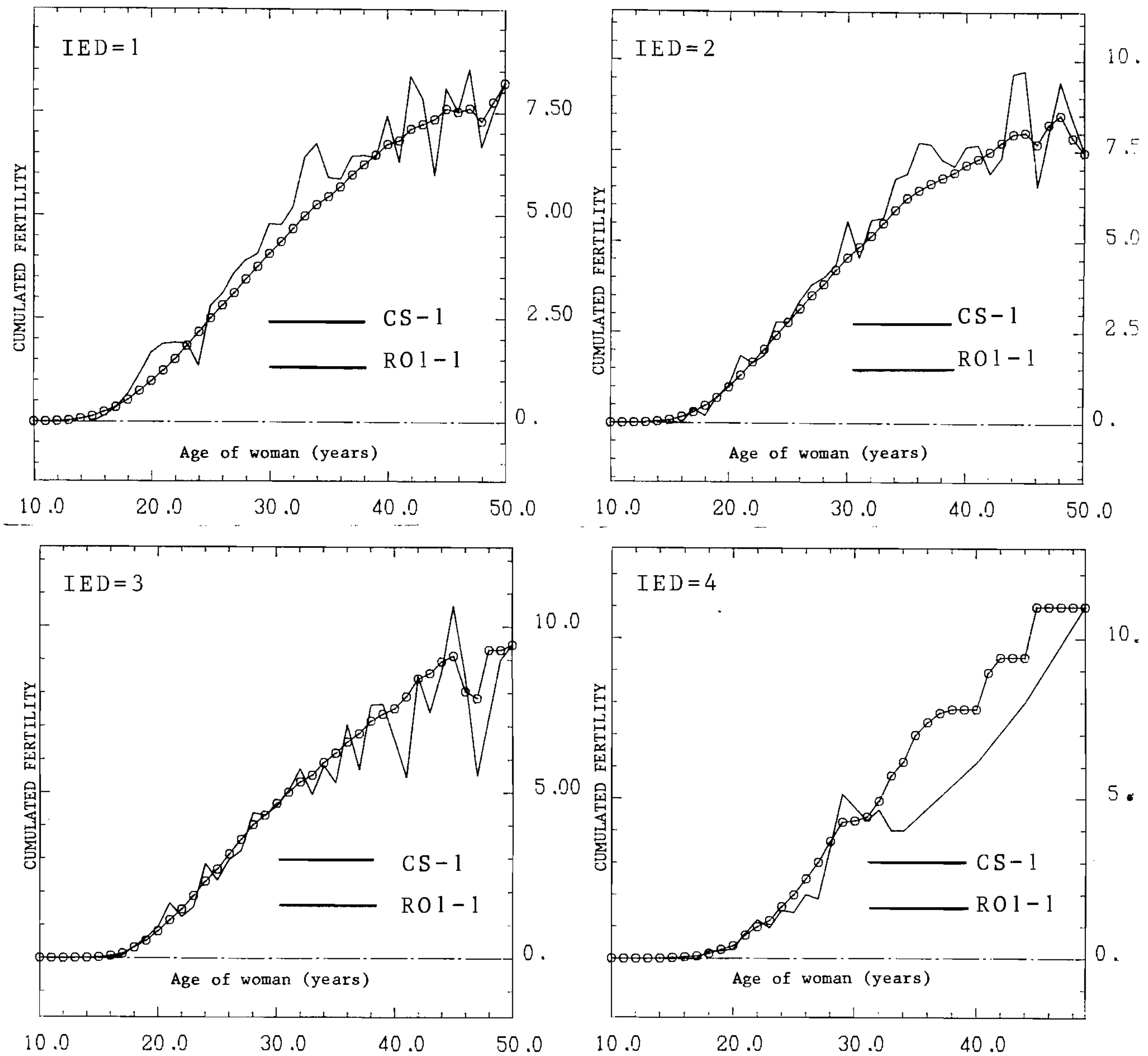


TABLE 11 : LIFE TABLE ESTIMATES OF THE CUMULATED LIFE TIME FERTILITY $\pi(t)$ FROM DATA ON WOMEN IN CENTRAL AND EASTERN PROVINCES (*)

	CS-1 METHOD			RO1-1 METHOD		
t	n_t	d'_{tt}	$p_t^{(1)}$	n'_t	d'_t	$p_t^{(3)}$
10	-	-	-	2613.342	0.000	0.000
11	-	-	-	2613.342	3.685	.001
12	-	-	-	2613.342	18.627	.007
13	-	-	-	2613.342	46.013	.018
14	-	-	-	2613.342	104.636	.040
15	157.603	.595	.004	2613.342	196.006	.075
16	143.284	5.393	.038	2455.739	373.334	.152
17	101.819	13.440	.132	2312.455	585.341	.253
18	123.828	36.757	.297	2210.636	944.451	.427
19	112.361	67.048	.597	2086.808	1344.663	.644
20	88.036	76.812	.873	1974.447	1788.464	.906
21	69.707	99.903	1.433	1886.411	2277.795	1.207
22	73.682	102.226	1.387	1816.704	2762.793	1.521
23	82.287	126.652	1.539	1743.022	3261.920	1.871
24	73.614	166.601	2.263	1660.735	3734.750	2.249
25	115.434	293.234	2.540	1587.121	4124.220	2.599
26	54.513	164.984	3.027	1471.687	4349.789	2.956
27	80.752	275.930	3.417	1417.174	4677.578	3.301
28	103.283	423.035	4.096	1336.422	4877.556	3.650
29	87.961	376.375	4.279	1233.139	4892.548	3.968
30	106.190	517.405	4.872	1145.178	4903.854	4.282
31	44.429	211.977	4.771	1038.988	4751.390	4.573
32	77.908	421.862	5.415	994.559	4853.898	4.880
33	42.228	246.670	5.841	916.651	4751.992	5.184
34	65.053	427.006	6.564	874.423	4799.552	5.489
35	100.156	613.143	6.122	809.370	4635.881	5.728
36	43.694	293.922	6.727	709.214	4234.759	5.971
37	77.008	521.659	6.774	665.520	4144.533	6.228
38	70.585	483.821	6.854	588.512	3797.903	6.453
39	43.763	300.698	6.871	517.927	3447.590	6.657
40	76.008	561.116	7.382	474.164	3265.160	6.886
41	46.090	315.226	6.839	398.156	2792.175	7.013
42	35.158	277.014	7.879	352.066	2565.500	7.287
43	22.661	174.273	7.690	316.908	2353.883	7.428
44	22.281	156.644	7.030	294.247	2228.990	7.575
45	71.087	610.936	8.594	271.966	2113.473	7.771
46	45.182	329.107	7.284	200.879	1526.402	7.599
47	47.596	392.388	8.244	155.697	1207.993	7.759
48	43.832	307.385	7.013	108.101	816.107	7.549
49	39.531	306.272	7.748	64.269	508.722	7.916
50	24.738	202.450	8.184	24.738	202.450	8.184

(*) t=age of woman; n_t=number of women with current age t completed years; d'_{tt}=number of children born to those n_t women before exact age t; n'_t=number of women with current age larger than or equal to t; d'_t=number of children born to those n'_t women before exact age t; $p_t^{(1)}=d'_{tt}/n_t$ = CS-1 estimate of $\pi(t)$, according to [8] ; $p_t^{(3)} = d'_t / n'_t$ = RO1-1 estimate of $\pi(t)$, according to [12] .

FIG. 13 : LIFE TABLE ESTIMATES OF THE CUMULATED FERTILITY SCHEDULE $\pi(t;z)$ FOR SUBSAMPLES z, SPECIFIED BY IED, OF WOMEN RESIDING IN CENTRAL AND EASTERN PROVINCES (KFS)



These plots show that the RO1-1 life table estimate is close to the CS-1 estimate. If the sample is large enough, then the RO1-1 estimate is also smooth and regular, except at the upper tail where it tends to the CS-1 estimate.

7.3. Analysis by GESLIRM's or (GLIRM's)

The model defined in section 7.1 can now be fitted to CS-1 and to R01-1 data. The model essentially involves 3 parameters θ, β and C. However, C will be fixed, and θ and β will then be estimated applying a GLIRM to the series $\pi^*(t;z)=\pi(t;z)/C$.

Fig. 12 and Fig. 13 show that the life time fertility schedules are to some extent flattened at the end of the age range. It is believed that this is mainly caused by omission of births, i.e. women tend to omit births of children who moved away or died, and this omission-error increases with age of women (Brass, 1980). Hence, older women underestimate their current fertility. Consequently, total fertility will also be underestimated. However, we can extrapolate the smooth R01-1 curves in Fig. 12 and Fig. 13 by hand - ignoring the data points at higher ages - and so we get a more reasonable idea about the total fertility C. It is seen then that C is larger than 8. Although this value of C may be somewhat low, we have used it in further analysis of the data, which is intended to demonstrate the methodology, rather than to obtain extremely accurate results.

With fixed $\hat{C}=8.00$, we normalized the data as follows : $n^*_{zt}=8.00\ n_{zt}$. Then, the GLIRM [25b] for $\pi^*(t;z)=\pi(t;z)/8.00$ is fitted to the data $[d_{zt}, n^*_{zt}, t, z]$ (cfr. section 3.1). Notice that : n_{zt} is the number of women with covariates z and current age t completed years (when dealing with CS data) or larger (when dealing with R01 data); d_{zt} is the number of births those women already had before age t; and n^*_{zt} is the number of births those women would have ultimately if they would all attain the total fertility level of 8.00.

First, we fitted the GLIRM's CED*IED*SST, CED*SST, IED*SST and SST to both CS-1 and R01-1 data. Results are shown in Table 12. Looking at the estimated mean (or median) ages at childbearing, it is difficult to see the effects of CED and IED. This may be due to the fact that the estimates for samples containing less educated women (IED=1, and IED=2) are biased because of the above mentionned ommission-error. Indeed, samples with IED=1 (and IED=2) contain most of the older women, causing an underestimation of the level of fertility at high ages or, equivalently, causing an upwards bias for the mean (or median) age at childbearing. Notice, however, that the bias is more serious for R01 than for CS data. This can be explained by the fact that the R01 estimates at low ages too depend on data for older women.

TABLE 12 : FITTING MODELS CED*IED*SST, CED*SST, IED*SST AND SST TO FERTILITY DATA

PANEL A : RESULTS FROM FITS ON RO1-1 DATA

	IED=1	IED=2	IED=3	IED=4	NOT CONTROL-LED FOR IED
CED=1	.640 (a)	.869	1.113	.753	.687
	-.286 (b)	-.048	-.049	-.261	-.262
	31.72 (c)	28.95	28.31	30.91	31.26
	9.03 (d)	7.90	6.72	8.31	8.74
	32.20 (e)	28.38	27.62	30.94	31.94
CED=2	.762	.841	.988	.811	.835
	-.094	-.093	-.110	-.538	-.116
	29.64	29.34	28.99	32.53	29.52
	8.50	8.01	7.22	7.50	8.02
	29.32	28.89	28.43	32.82	29.12
CED=3	.835	.963	.970	1.134	.928
	-.141	-.059	-.124	-.238	-.127
	29.69	28.74	29.13	29.31	29.28
	7.99	7.39	7.30	6.49	7.50
	29.34	28.12	28.60	28.80	28.79
NOT CON-TROLLED FOR CED	.727	.886	1.001	.970	.817
	-.177	-.073	-.107	-.314	-.160
	30.41	29.06	28.93	30.31	28.89
	8.61	7.78	7.17	7.09	8.07
	30.34	28.52	28.35	30.02	29.60

PANEL B : RESULTS FROM FITS ON CS-1 DATA

	IED=1	IED=2	IED=3	IED=4	NOT CONTROL-LED FOR IED
CED=1	.481	.894	1.084	.891	.687
	.019	-.107	-.068	-.309	-.168
	29.98	29.26	28.49	30.60	30.53
	11.04	7.70	6.83	7.47	8.90
	30.16	28.77	27.82	30.41	30.55
CED=2	.788	.859	.977	.906	.980
	.044	.023	-.141	-.442	-.106
	28.55	28.49	29.21	31.38	28.99
	8.48	8.03	7.25	7.21	7.27
	27.88	27.81	28.70	31.34	28.43
CED=3	.487	1.202	.909	.965	.980
	.317	.011	-.136	-.448	-.122
	27.19	27.80	29.40	31.15	29.08
	11.27	6.41	7.59	6.93	7.25
	25.58	27.03	28.94	31.00	28.54
NOT CON-TROLLED FOR CED	.638	1.009	.984	.953	.913
	.068	.012	-.124	-.410	-.126
	28.85	28.19	29.08	30.97	29.32
	9.63	7.23	7.23	7.04	7.58
	28.32	27.45	28.54	30.80	28.85

(a) estimated slope $\hat{\beta}$; (b) estimated intercept $\hat{\beta}$; (c) estimated mean age at childbearing; (d) estimated standard deviation of age at childbearing; (e) estimated median age at childbearing.

TABLE 13 : FITTING THE MODEL (CED*SST).SST+IED TO FERTILITY DATA

PANEL A : RESULTS FROM FITS ON RO1-1 DATA

	IED=1	IED=2	IED=3	IED=4	
CED=1	.641 (a)	.856	1.070	.718	
	-.178 (b)	-.076	-.108	-.319	
	30.84 (c)	29.18	28.75	31.52	
	9.23 (d)	7.95	6.86	8.43	
	31.04 (e)	28.68	28.14	31.77	
CED=2	.752	.845	.989	1.008	
	-.178	-.076	-.108	-.319	
	30.30	29.21	28.97	30.19	
	8.45	8.01	7.22	6.92	
	30.17	28.73	28.41	29.86	
CED=3	.834	.961	.979	1.073	
	-.178	-.076	-.108	-.319	
	29.95	28.86	29.00	29.96	
	7.95	7.39	7.27	6.65	
	29.66	28.27	28.44	29.57	

PANEL B : RESULTS FROM FITS ON CS-1 DATA

	IED=1	IED=2	IED=3	IED=4	
CED=1	.463	.914	1.049	.806	
	.076	-.012	-.128	-.423	
	29.53	28.58	28.92	31.78	
	11.32	7.69	6.93	7.73	
	29.48	27.92	28.35	31.94	
CED=2	.775	.867	.982	.923	
	.076	-.012	-.128	-.423	
	28.36	28.71	29.12	21.19	
	8.60	7.95	7.24	7.16	
	27.63	28.09	28.58	31.09	
CED=3	.574	1.209	.912	.985	
	.076	-.012	-.128	-.423	
	29.02	27.91	29.34	30.91	
	10.19	6.37	7.58	6.89	
	28.60	27.17	28.87	30.72	

NOTES : (a)-(e) : see Table 12.

On the whole, IED seems to have a (slight) increasing effect, while CED has a (slight) decreasing effect on the mean (or median) age at childbearing, and an interaction effect may be present as well. More obvious, however, are the negative effects of both CED and IED on the standard deviations (or variances) of age at childbearing. This may be explained by the fact that more educated women start having children at later ages than less educated women do.

Other models have been fitted and compared with the above ones. The model (CED*IED).SST+IED may provide an adequate fit. Thus, most significant are : a main effect of IED on the intercept, a main effect of both CED and IED on the slope, and an interaction effect of CED and IED on the slope. (Results are shown in Table 13.) If the interaction effect of CED and IED is dropped, the estimated mean ages (or median) at childbearing are only slightly changed; the effect of IED on the standard deviations becomes negative for both RO1-1 and CS-1 data; the effect of CED on the standard deviations becomes negative for RO1-1 data, but slightly U-shaped for CS-1 data.

Of course, the results are not quite reliable, as they ignore ommission-errors and as they rely on a very rough overall estimate of the total fertility.

REFERENCES

Andersen E.B.(1970) : Asymptotic properties of conditional maximum likelihood estimators. Journal of the Royal Statistical Society, Series B, 32, pp. 283-301.

Ashton W.D.(1972) : The logit transformation with special reference to its uses in bioassay. Griffin's Statistical Monographs and Courses no. 32, London.

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

Booth H.(1979) : The estimation of fertility from incomplete cohort data by means of the transformed Gompertz model. PhD.Thesis, University of London.

Brass W. et.al (1968) : The demography of Tropical Africa. Princeton University Press.

Brass W. (1974) : Perspectives in population prediction : illustrated by the statistics of England and Wales. J. Roy. Statist. Soc., Series A, 137, nr. 4, pp. 532-583.

Brass W. (1980) : Sreening procedures for detecting errors in maternity history data. WFS-Occasional Papers, No. 22, pp. 30-49.

Coale A.J. (1971) : Age patterns of marriage. Population Studies, XXV, no. 2, pp. 193-214.

Coale A.J. and McNeil D.R. (1972) : The distribution by age of the frequency of first marriage in a female cohort. Journal of the American Statistical Association, vol. 67, nr. 340, pp. 743-49.

Cox D.R. (1975) : Partial likelihood. Biometrika, 62, 2, pp. 269-76.

Elandt-Johnson R.C. and Johnson N.L. (1980) : Survival models and data analysis. Wiley, New York.

Finney D.J. (1971) : Probit analysis (3rd edn.), Cambridge, University Press.

Kalbfleisch J.D. and Prentice R.L. (1980) : The Statistical analysis of failure time data. Wiley, New York.

Lesthaeghe R. and Page H.J. (1980) : The post-partum non-susceptible period : development and application of model schedules. Population Studies, vol. 34, nr. 1, pp. 143-69.

Lesthaeghe R., Vanderhoeft C., Becker S. and Kibet M. (1983) : Individual and contextual effects of female education on the Kenya marital fertility transition. IPD-Working Paper 9, Brussels.

Martin M.P.D. (1967) : Une application des fonctions de Gompertz à l'étude de la fécondité d'une cohorte. Population 22, no. 6, pp. 1085-96.

Page H., Lesthaeghe R. and Shah I. (1982) : Illustrative analysis : breastfeeding in Pakistan, WFS-Scientific Report, no. 37, London.

Rodriguez G. and Trussell J. (1980) : Maximum likelihood estimation of the parameters of Coale's model nuptiality schedule from survey data. WFS-Technical Bulletin no. 7, London.

Smith D. and Keyfitz N. (1977) (ed.) : Mathematical Demography; selected papers. Springer, New York.

Vanderhoeft C. (1982) : Accelerated failure time models : an application to current status breast-feeding data from Pakistan. Genus, vol. 38, nr. 1-2, pp. 135-157.

Vanderhoeft C. (1983) : A unified approach to models for analysis of zero-one data with applications to intermediate fertility variables. IPD-Working Paper 5, Brussels.

Zaba B. (1981) : Use of the relational Gompertz model in analysing fertility data collected in retrospective surveys. CPS Working Paper 81-2, London School of Hygiene and Tropical Medicine.

TABLE A.1 : ONE-PARAMETER SUBCLASSES OF DISTRIBUTIONS $\Phi^{-1}(.)$, WITH RELATED DISTRIBUTIONS AFTER EXPONENTIATION, AND WITH THE INVERSE FOR SOME SPECIAL VALUES OF THE PARAMETER.

Parameters m_1, m_2	c.d.f. of W $\Phi^{-1}(w)=F_W(w)=\ldots$	c.d.f. of $V=e^W$ $F_V(v)=\ldots(v>0)$	special values of parameter(s)	c.d.f. of W $\Phi^{-1}(w)=F_W(w)=\ldots$	inverse c.d.f. of W $\Phi(\pi)=\ldots$	c.d.f. of $V=e^W$ $F_V(v)=\ldots(v>0)$
$m_1=m_2=m$ (a)			$m=+\infty$	normal $\int_{-\infty}^{w}\frac{1}{\sqrt{2\pi}}e^{-\frac{u^2}{2}}du$	probit (π)	log-normal $\int_0^v \frac{1}{\sqrt{2\pi}}x^{-1}e^{-\frac{(\log x)^2}{2}}dx$
$m_1=m_2=m$ (a), $0<m<+\infty$	exponential beta-prime (b) $\int_{-\infty}^{w}\frac{\Gamma(2m)}{(\Gamma(m))^2}e^{mu}(1+e^u)^{-2m}du$	beta-prime $\int_0^v \frac{\Gamma(2m)}{(\Gamma(m))^2}x^{m-1}(1+x)^{-2m}dx$	$m=1$	logistic $\frac{e^w}{1+e^w}$	logit (π) $=\log(\frac{\pi}{1-\pi})$	log-logistic $\frac{v}{1+v}$
$0<m_1<+\infty$, $m_2=+\infty$	exponential gamma $\int_{-\infty}^{w}\frac{m_1^{m_1}}{\Gamma(m_1)}e^{m_1(u-e^u)}du$	gamma $\int_0^v \frac{m_1^{m_1}}{\Gamma(m_1)}x^{m_1-1}e^{-m_1x}dx$	$m_1=1$	extreme value (minimum) (c) $1-e^{-e^w}$	$\log(-\log(1-\pi))$	exponential (d) $1-e^{-v}$
			$m_1=\frac{1}{2}$	exponential chi-square with 1 degree of freedom $\int_{-\infty}^{w}(\sqrt{2}\Gamma(\frac{1}{2}))^{-1}e^{\frac{1}{2}(u-e^u)}du$	(e)	chi-square with 1 degree of freedom $\int_0^v(\sqrt{2}.\Gamma(\frac{1}{2}))^{-1}x^{-\frac{1}{2}}e^{-\frac{x}{2}}dx$
$m_1=+\infty$, $0<m_2<+\infty$	exponential reciprocal gamma $\int_{-\infty}^{w}\frac{m_2^{m_2}}{\Gamma(m_2)}e^{-m_2(u+e^{-u})}du$	reciprocal gamma $\int_0^v \frac{m_2^{m_2}}{\Gamma(m_2)}x^{-m_2-1}e^{-m_2v^{-1}}dv$	$m_2=1$	Gompertz (f) $e^{-e^{-w}}$	$-\log(-\log\pi)$	log - Gompertz $e^{-v^{-1}}$
			$m_2=.604$	$\int_{-\infty}^{w}(.4983)e^{-.604(u+e^{-u})}du$	(g)	$\int_0^v(.4983)x^{-1.604}e^{-.604\,x^{-1}}dx$

(a) Those are symmetric around w=0

(b) Or generalized logistic distribution

(c) Or log-Weibull distribution

(d) A special Weibull distribution

(e) No explicit formula

(f) Or extreme value (maximum) distribution

(g) No explicit formula, but useful for the nuptiality model

TABLE A.2 : SOME SPECIAL MODELS : PROPORTIONAL HAZARDS MODELS AND (TRANSLATED-) ACCELERATED FAILURE TIME MODELS

Parameters m_1, m_2	$\Phi(\pi)=\ldots$	$\Phi(\pi_s(t))=\ldots.$	Parameters $\theta(z),\beta(z)$	Parameters (a) $\theta^+(z),\beta^+(z)$	Name of model; description (d)	Special functions of parameters
$m_1=1$, $m_2=+\infty$	$\log(-\log(1-\pi))$	n.s.[b]	$\theta(z)=1$	$\theta^+(z)=\alpha_2$	Proportional hazards model (PHM) $\lambda^*(t;z)=e^{\beta(z)}.\lambda_s(t)$	Relative hazard : $e^{\beta(z)}$
n.s.	n.s.	$\alpha_1+\alpha_2\frac{t^{\alpha_3}-1}{\alpha_3}$ $(\alpha_3\neq 0,\neq 1)$	$(\theta(z)-1).(\alpha_2-\alpha_1\alpha_3)$ $-\beta(z).\alpha_3=0$	$\theta^+(z)-\beta^+(z).\alpha_3=\alpha_2-\alpha_1.\alpha_3$	Accelerated failure time model (AFTM) $\pi^*(t;z)=\pi_s(\frac{t}{k(z)})$ $\lambda^*(t;z)=\frac{1}{k(z)}\lambda_s(\frac{t}{k(z)})$	$k(z)=\theta(z)^{-1/\alpha_3}$
		$\alpha_1+\alpha_2.t$	$(\theta(z)-1)\alpha_1+\beta(z)=0$	$\beta^+(z)=\alpha_1$		$k(z)=\theta(z)^{-1}$
		$\alpha_1+\alpha_2.\log t$	$\theta(z)=1$	$\theta^+(z)=\alpha_2$		$k(z)=e^{-\beta(z)/\alpha_2}$
n.s.	n.s.	$\alpha_1+\alpha_2\frac{(t+\alpha_4)^{\alpha_3}-1}{\alpha_3}$ $(\alpha_3\neq 0,\neq 1)$	$(\theta(z)-1).(\alpha_2-\alpha_1\alpha_3)$ $-\beta(z).\alpha_3=0$	$\theta^+(z)-\beta^+(z).\alpha_3=\alpha_2-\alpha_1\alpha_3$	Translated - Accelerated failure time model (TAFTM): $\pi^*(t;z)=\pi_s(\frac{t-a_0(z)}{k(z)})$ $\lambda^*(t;z)=\frac{1}{k(z)}\lambda_s(\frac{t-a_0(z)}{k(z)})$	$k(z)=\theta(z)^{-1/\alpha_3}$ $a_0(z)=\alpha_4.(\theta(z)^{-1/\alpha_3}-1)$
		$\alpha_1+\alpha_2.t$	n.c. (c)	n.c.		$k(z)=\theta(z)^{-1}$ $a_0(z)=\frac{\alpha_1-\beta^+(z)}{\theta^+(z)}$
		$\alpha_1+\alpha_2.\log(t+\alpha_4)$	$\theta(z)=1$	$\theta^+(z)=\alpha_2$		$k(z)=e^{-\beta(z)/\alpha_2}$ $a_0(z)=\alpha_4.(e^{-\beta(z)/\alpha_2}-1)$

(a) If $\Phi(\pi_s(t))=\alpha_1+\alpha_2\frac{(t+\alpha_4)^{\alpha_3}-1}{\alpha_3}$, then we define $\theta^+(z)=\alpha_2.\theta(z)$ and $\beta^+(z)=\beta(z)+\alpha_1.\theta(z)$

(b) n.s. = no specification required

(c) n.c. = no condition

(d) $\lambda^*(t;z)$ is the hazard function; $\lambda_s(t)$ is the hazard function of the standard population

IPD-WORKING PAPERS 1983

1983-1. R. Lesthaeghe : "A Century of Demographic and Cultural Change in Western Europe : An Exploration of Underlying Dimensions".

X 1983-2. R. Lesthaeghe, C. Vanderhoeft, S. Becker, M. Kibet : "Individual and Contextual Effects of Education on Proximate Determinants and on Life Time Fertility in Kenya".

1983-3. F. Eelens : "Impact of Breast-feeding on Infant and Child Mortality with varying Incidence of Malaria - Evidence from the Kenya Fertility Survey 1977-78".

1983-4. S. Becker, A. Chowdhury, S. Huffman : "Determinants of Natural Fertility in Matlab, Bangladesh".

1983-5. C. Vanderhoeft : "A Unified Approach to Models for Analysis of Zero-One Data with Applications to Intermediate Fertility Variables".

1983-6. S. Wijewickrema, R. Bulté : "Migration Impact on Population Growth in Belgium : A Multiregional Investigation with Detailed Projection Results for the Period 1976-2001".

1983-7. S. Wijewickrema et Alii : "Marital Status Trends in Belgium, 1961-77 : Application of Multi-State Analysis".

1983-8. H.J. Page : "Fertility and Family : New Currents and emerging Emphases in Research and Policy" (Overview of the UN Expert Group Meeting on Fertility and Family (New Delhi, 1983) preparatory to the International Conference on Population).

1983-9. R. Lesthaeghe, C. Vanderhoeft, S. Becker, M. Kibet : "Individual and Contextual Effects of Female Education on the Kenyan Marital Fertility Transition" (abbreviated non-technical version of Working Paper 1983-2).

X = OUT OF PRINT

IPD-WORKING PAPERS 1984

1984-1. F. Rajulton : "An Age Dependent Semi-Markov Model of Marital Status in Belgium : An Application of Littman's Algorithm to Period Data, 1970".

1984-2. R. Lesthaeghe : "Fertility and its Proximate Determinants in sub-Saharan Africa : The Record of the 1960's & 70's".

1984-3. C. Vanderhoeft : "Life Tables and Relational Models : A Unified Approach, and Applications in Demography".

1984-4. R. Lesthaeghe : "Demographic Renewal and Problems of Social Organization".

1984-5. S. Becker : "A Response Bias in the reporting of Month of Birth in Pregnancy History Surveys".

1984-6. R. Lesthaeghe : "On the Adaptation of sub-Saharan Systems of Reproduction".

IPD-WORKING PAPERS 1985

1985-1. R. Lesthaeghe and F. Eelens : "Social Organization and Reproductive Regimes : Lessons from sub-Saharan Africa and historical Western Europe".

1985-2. P. Willems and S. Wijewickrema : "The Evolution of Nuptiality in Belgium from 1954 to 1981".

1985-3. F. Eelens and L. Donné : "The Proximate Determinants of Fertility in Sub-Saharan Africa: A Factbook based on the Results of the World Fertility Survey".

1985-4. Ronald C. Schoenmaeckers : "Current Fertility Behaviour in Africa. Results from a Birth Interval Analysis of WFS Data".

1985-5. Camille Vanderhoeft : Stratified Proportional Hazards Models. A GLIM oriented approach, with special reference to the problem of competing risks.

1985-6. Stanley Wijewickrema : "Childlessness in Belgium and Flanders".

1985-7. R. Lesthaeghe : "Value Orientations, Economic Growth and Demographic Trends - Toward a Confrontation?

1985-8. Fernando Rajulton and Stan Wijewickrema : "An Enquiry into the Two Basic Assumptions of Multistate Increment-Decrement Life Tables".

IPD-WORKING PAPERS 1986

1986-1. F. Rajulton : "Marriageability and divorceability : a simulation of the unobservables through the conditional gaussian diffusion process.

1986-2. H.J. Page : "Child-bearing versus Child-rearing : Co-residence of Mothers and Children in Sub-Saharan Africa".

1986-3. R. Lesthaeghe, G. Kaufmann, D. Meekers : "The Nuptiality Regimes in Sub-Saharan Africa".

1986-4. R. Lesthaeghe, G. Kaufmann, D. Meekers : "The Nuptiality Regimes in Sub-Saharan Africa : Data Files NUPFILE 1 and NUPFILE 2".