September, 1988

STOCKHOLM RESEARCH REPORTS IN DEMOGRAPHY
NO.49

A discrete time method for the analysis of event histories.

Elja Arjas and Pekka Kangas
University of Oulu

Abstract

The paper discusses the application of discrete time regression models to demographic life history data. General conditions are stated under which the likelihood expression obtains a simple Bernoulli product form. As an illustration, such a model with a logistic link function is fitted to Swedish third births data, earlier studied by B. and J. Hoem. We also discuss briefly aspects relating to computation and the necessary software.

Mailing address : Department of Applied Mathematics and Statistics,
University of Oulu, 90570 Oulu, Finland

ISBN 91-7820-036-9
ISSN 0281-8728

1. Outline of the method

In demographic longitudinal studies, time measurements in the data are mostly reported in a rounded form : A time unit is fixed, typically a month or a year, and then it is reported during which interval (of unit length) the event in question occurred.

Such rounding does cause some problems, however. For example it may result in a considerable number of ties, and the requirement of computer time can be excessive for methods such as Cox's proportional hazards regression. A slightly different problem arises in using the Poisson regression technique with piecewise constant baseline hazards, which requires the estimation of the times of exposure in the various groups. To do this, the rounded time measurements are conventionally modified by applying "acturial methods", thus introducing an element of spurious accuracy into the data. For example, if the length of the time interval is a month, a single event is dated to the fifteenth day, and if there are two events, one is dated to the tenth and the other to the twentieth day of the month. If the time intervals are short these somewhat ad hoc modifications of the essentially discrete data, in order to apply a continuous time method, appear rather harmless since the differences between the true, recorded, and modified time measurements are small. However, "spurious accuracy" is somewhat awkward, and it is natural to look for alternatives.

The obvious alternative is to fit discrete time models. The foundations of this approach have been discussed by Allison (1982), and recent applicatiaons include Rindfuss et al. (1984) and Morgan et al. (1988). One will then consider probabilities that the demographic event in question occurs to an individual during a considered time interval. To fix ideas, suppose that the event in question is death and that the time unit is one month. Then the life of an individual can be thought of as a sequence of "monthly Bernoulli trials", resulting in a sequence of zero's as long as the individual stays alive, and ending with a one for the month during which the individual dies. For the data as a whole, the number of trials equals the number of person months of exposure. Moreover, the estimated probabilities from such a model are easily converted into continous time intensities, relative risks, etc., if it is felt that these form a preferred way of reporting the results.

It is not difficult to accommodate modern regression techniques involving, say, generalized linear models and time dependent covariates into such a discrete time framework. The discrete time also applies in situations where the response splits into different "types" corresponding to, say, different causes of death, or where the event can repeat itself several times, as in employment studies. However, the following two conventions seem more or less unavoidable if the data are in a discrete time form: First, censoring, and possible other forms of controlling the risk set, are always thought to happen at times which are integer multiples of the chosen time unit. Second, if the model involves time dependent covariates, their measured values need to be interpreted as "prevailing conditions of exposure during the time interval (of unit length) to which the corresponding response is related". For an interesting discussion of this second aspect see Sandefur and Thuma (1987). In particular, if the occurence of the response can change the conditions of exposure these must be determined at the beginning of the time interval, before the response occurs. In a sense, the discrete time model consists of a set of short term predictions, each made at the beginning of a unit time interval.

Below, we use the notion of a history (mathematically, a σ-field) as a description of such conditions. Briefly, the history $\mathcal{H}_{t-1}$ is taken to be the set of conditions prevailing during the time interval $(t-1, t]$, and then leading to the response indexed in our discrete time model by t. The exact mathematical definition of the history $\mathcal{H}_{t-1}$ is given in the Appendix. Apart from the explicit conditioning on histories, our models below are identical to the discrete time models discussed by Allison (1982). In fact, this paper, and particularly the Appendix, can be seen as a contribution to the discussion concerning the legitimacy of the discrete time model (section "Problems with the discrete time approach" in Allison's paper). In particular we hope that the Appendix could straighten the apparent confusion in Allison's paper which regards the product form of the likelihood expression and independence. The product form is really a consequence of the chain multiplication rule of conditional probabilities.The computer program discussed in section 3, on the other hand, can be viewed as a pragmatic way to answer the practicality question raised by Allison.

We now explain the structure of the considered discrete time regression model. For a more careful discussion of the model and its properties the reader is referred to Arjas (1986), and Arjas and Haara (1987).

Let $j=1,2,\ldots$ index an individual, let $t=1,2,\ldots$ be the discrete time variable indicating the t^{th} "month" of the follow-up, so that t corresponds to the time interval $(t-1,t]$ in continuous time.

Denote

(1)
$$\Delta N_j(t) = \begin{cases} 1, & \text{if the demographic event occurs to } j \text{ during the } t^{th} \text{ month} \\ 0, & \text{otherwise} \end{cases}$$

Let $Z_j(t) = (Z_{j,1}(t), \ldots, Z_{j,p}(t))$ be a known (possibly time dependent) covariate vector for j. We now postulate that, given the observed "history" H_{t-1} corresponding to times $0,1,2,\ldots,t-1$ in the follow-up, the conditional probability that the considered demographic event occurs to j at t (i.e., during the t^{th} month) can be expressed as

(2)
$$\Pr(\Delta N_j(t) = 1 \mid H_{t-1}) = Y_j(t)\cdot g^{-1}(\beta' Z_j(t)),$$

where

(3)
$$Y_j(t) = \begin{cases} 1, & \text{if } j \text{ is in the risk set during the } t^{th} \text{ month} \\ 0, & \text{otherwise,} \end{cases}$$

$\beta ==(\beta_1, \ldots, \beta_p)$ is a vector of model parameters, and g^{-1} is the inverse of a link function in the sense of generalized linear models (McGullagh and Nelder (1983)). Specifically, we shall consider the <u>logit link</u>

(4)
$$g(x) = \log(x/(1-x)) \qquad (0 < x < 1),$$

$$g^{-1}(x) = \exp(x)/(1+\exp(x)) \qquad (-\infty < x < \infty)$$

and the cloglog-link

$$(5) \qquad g(x) \equiv \log(-\log(1-x)) \qquad (0 < x < 1),$$

$$g^{-1}(x) \equiv 1 - \exp(-\exp(x)) \qquad (x > 0).$$

Under general conditions (presented in detail in the Appendix) the likelihood corresponding to an entire follow-up data set becomes a product over j and t of the terms

$$(6) \qquad \Pr(\Delta N_j(t)=1 \mid H_{t-1})^{\Delta N_j(t)} \, [1-\Pr(\Delta N_j(t)=1 \mid H_{t-1})]^{1-\Delta N_j(t)}$$

This likelihood function is log-concave in β, which leads to an unproblematic numerical ML-estimation of the parameter vector. It is shown in Arjas and Haara (1987) that the "standard asymptotic normality results" hold for the estimates when the logit-link (4) is used.

It is important to note that, should one prefer to present the results from a statistical analysis in the form of continous time rates (intensities) instead of probabilities, the discrete time model will provide such estimates as well. In the case of the logit link (4) we have

$$(7) \qquad \Pr(\Delta N_j(t)=1 \mid H_{t-1}) = \frac{Y_j(t)\cdot\exp(\beta' Z_j(t))}{1 + \exp(\beta' Z_j(t))} .$$

For most demographic events such probabilities are small, and therefore $\Pr(\Delta N_j(t)=1 \mid H_{t-1}) \approx Y_j(t)\cdot\exp(\beta' Z_j(t))$ holds as an approximation. It is natural to use a constant first covariate $Z_{j1}(t)\equiv 1$ so that β_1 can be viewed as an intercept. The value of β_1 will of course depend on the length of the "month", the chosen time unit. The other β-coefficients, which have the role of relative risks modulating the baseline, tend to be quite stable towards such changes. Therefore it is quite convenient to use longer intervals for provisional model fitting if the computing time otherwise becomes a problem.

On the other hand, postulating that $\Delta N_j(t)$ has a constant H_{t-1}-conditional intensity $\lambda_j(t)$, say, over the unit interval corresponding to t, we also have that for short intervals approximately $\Pr(\Delta N_j(t)=1 \mid H_{t-1}) \approx \lambda_j(t)$. Combining these two, we find that $\lambda_j(t) \approx Y_j(t)\cdot\exp(\beta' Z_j(t))$, i.e., the familiar multiplicative intensity form holds as an approximation. When the cloglog-link is used, this relationship actually becomes exact : From (5) we obtain

$$\text{(8)} \qquad \Pr(\Delta N_j(t)=1 \mid H_{t-1}) \equiv Y_j(t)\cdot(1-\exp(-\exp(\beta' Z_j(t))))$$

$$\equiv 1-\exp(-Y_j(t)\cdot\exp(\beta' Z_j(t))) \; .$$

But this probability equals $1-\exp(-\lambda_j(t))$, and so $\lambda_j(t) \equiv Y_j(t)\cdot\exp(\beta' Z_j(t))$ holds exactly.

2. A case study : B. and J. Hoem's data on third births in modern Sweden

In order to get a concrete idea about how the method explained above compares with a more conventional one in demography, the Poisson regression based on piecewise constant proportional occurrence rates and on actuarial methods for adjusting the dates, we reanalyzed the data on third births in modern Sweden (B. and J. Hoem (1987), henceforth abbreviated as H&H). Since the emphasis is on the comparison of the methods we followed the steps taken in H&H very closely, also initially employing the same covariates. As it turns out, our numerical results are in very good agreement with those obtained in H&H. We do not describe the data here nor make an attempt to assess the significance of the demographic findings but refer to H&H for those.

We started by drawing the Nelson-Aalen plots for the third births. There is one plot for each age cohort. Time t is measured in months from ten months after the second birth (Figure 1). As in H&H, there was complete censoring of follow-up times exceeding 98 months. The birth rates seem approximately proportional in the cohorts, with younger cohorts having slightly smaller rates. Corresponding to these "baseline rates", we decided to split the time axis into three intervals where the levels were approximately constant : 0-24 months, 25-72 months and 73-98 months. In the model we used the following three "baseline covariates" which are common to all individuals (we denote the indicator function by $1(\cdot)$) :

$$z_{j,1}(t) = 1$$
$$z_{j,2}(t) = 1(\ t \leq 24\ )$$
$$z_{j,3}(t) = 1(\ 73 \leq t \leq 98\ ) \ .$$

The middle interval from 25 to 72 months is therefore directly represented by the intercept β_1 (= coefficient of $z_{j,1}(t)$).

The remaining covariates depend on the individual. Following H&H, we used initially information of the interval between first two births (fixed covariate, 3 levels), the educational level of the mother at second birth (fixed covariate, 2 levels), a cohort*(age at first birth) interaction term (fixed covariate, 9 categories), and a combined employment status covariate (time dependent, 12 categories). Always including one of the levels/covariates in the baseline, we were thus led to the following first menu of covariates (cf. Table 3 in H&H) :

interval	$z_{j,4}(t) = 1$(birth interval $\leq$ 29 months)
between first	(base level : 30-53 months)
two births	$z_{j,5}(t) = 1$(birth interval $\geq$ 54 months)
educational	(base level : education "low" or "middle")
level at	$z_{j,6}(t) = 1$(education "high")
second birth	
cohort*(age	$z_{j,7}(t) = 1$(cohort 1936-40, age 16-19)
at first birth)-	$z_{j,8}(t) = 1$(cohort 1941-45, age 16-19)
interaction	$z_{j,9}(t) = 1$(cohort 1946-50, age 16-19)
	$z_{j,10}(t) = 1$(cohort 1936-40, age 20-25)
	(base level : cohort 1941-46, age 20-25)
	$z_{j,11}(t) = 1$(cohort 1946-50, age 20-25)
	$z_{j,12}(t) = 1$(cohort 1936-40, age 26-34)
	$z_{j,13}(t) = 1$(cohort 1941-45, age 26-34)
	$z_{j,14}(t) = 1$(cohort 1946-50, age 26-34)
combined	$z_{j,15}(t) = 1$(less than 25 %, full time)
employment	$z_{j,16}(t) = 1$(25-75 % , full time)
status	$z_{j,17}(t) = 1$(more than 75 %, full time)
(percent of time	(base level : less than 25 %, part time)
spent as house-	$z_{j,18}(t) = 1$(25-75 % , part time)
wife, current	$z_{j,19}(t) = 1$(more than 75 %, part time)
employment	$z_{j,20}(t) = 1$(less than 25 %, child minder)
status)	$z_{j,21}(t) = 1$(25-75 % , child minder)

$Z_{j,22}(t) = 1$(more than 75 %, child minder)
(base level : less than 25 %, housewife (*))
$Z_{j,23}(t) = 1$(25-75 % , housewife)
$Z_{j,24}(t) = 1$(more than 75 %, housewife)

(*) In the present study, unlike in H&H, the category "less than 25 %, housewife" was combined with "less than 25 %, part time" to form a baseline.

The results of fitting this model ("Model 1"), by using the logit link, are displayed in Table 1 and Table 2. We also tried the cloglog-link but do not report the results here. The reason is that the results are so similar; typically differences in the coeffient estimates appeared in the third significant digit. We make two rather obvious comments :

1. The column "exp(parameter)" in Table 1 contains the numbers $\exp(\hat{\beta}_i)$, where $\hat{\beta}_i$ is the i^{th} coordinate of the ML-estimate $\hat{\beta}$. As explained in Section 1 above, these numbers are approximately the same as the relative intensities in a multiplicative intensity model. They can therefore be compared directly to the numerical results in Table 3 of H&H (reproduced here in the rightmost column of Table 1). The agreement is very good; the differences are only a few percent.

2. Since we have also produced (estimates of) the standard deviations and the correlations of the coefficient estimators, we can, using the asymptotic normality of the estimators, consider their statistical significance. Our first impression on this is that the division of the "combined employment status" covariate into 12 response categories (of which we consider 11) is too fine to let us infer reliably about the values of the individual coeffients. Only two of the ten "test statistics" have an absolute value exceeding 2. This brings up the question of finding another model with a more parsimonious parametrization. A second set of covariates where parameters could perhaps be saved is $Z_{j,7}(t)$,..., $Z_{j,14}(t)$, where we could consider only the cohort- and (age at first birth)- "main effects".

From the results and discussion in H&H we concluded that those categories in the combined employment status which could be associated to changes in employment correspond to either low or high third birth fertility, whereas those categories which correspond to a stable employment behavior seem to have an average third birth intensity. Thus we lumped the three categories corresponding to covariates $Z_{j,16}(t)$, $Z_{j,17}(t)$ and $Z_{j,19}(t)$ into a single "> 25 %, full time & > 75 %, part time"-category, expecting low third birth fertility, and the two categories corresponding to $Z_{j,20}(t)$ and $Z_{j,23}(t)$ into a "< 25 %, childminder & 25-75 %, housewife"-category; expecting high third birth fertility. The remaining five original response categories were combined with the previous baseline "< 25 %, part time". This lumping reduced the number of parameters by eight when compared to Model 1. Considering cohort and age at first birth only as "main effects" reduced the number of parameters further by four.

The results of fitting this model, called "Model 2", are shown in Table 3. The increase of the deviance is 12.73, corresponding to the 12 degrees of freedom saved. We do not make an attempt to comment on how natural this reduced model is from the point of view of demography but note that the six covariates which are common to Model 1 and Model 2 had very similar coeffient estimates.

It is obvious that from the estimated model we can also calculate, for any given covariate profile $\{ Z_j(t), t=1,2,\ldots \}$, the estimated third birth probabilities. This is done most conveniently in the form of an estimated cumulative distribution $\hat{Pr}$(time to third birth $\leq t$), $t=1,2,\ldots$. We demonstrate this possibility in Figures 2 and 3, which illustrate the estimated effects (Model 2) of cohort and employment profile (as a fixed covariate) when other covariates are at base level.

3. Computational aspects

The results presented here were produced by using a computer program which is designed to handle time dependent covariates in an efficient manner. The idea is that the covariate vector $Z_j(t)=(Z_{j,1}(t), \ldots ,Z_{j,p}(t))$ is continously updated for each subject j, instead of arranging all "monthly Bernoulli experiments" into a conventional data matrix. Time dependent covariates are updated only when their values change.

In practice this is done by dividing the covariates into three classes : fixed, preset time dependent and computed time dependent. Fixed covariates are specified in the usual way as a rectangular data matrix. Preset time dependent covariates are defined by giving their initial values and a file of update records, describing the subject, the time of change, the changing covariate and a new covariate value. Computed time dependent covariates are used to specify functions of time and may reference fixed covariates. A typical example would be the logarithm of the time elapsed after an operation on a patient.

Output consists of the model description, time grid interval used, estimated parameter values, their asymptotic standard deviations, corresponding approximate relative risks, estimated asymptotic correlation matrix of parameters and the deviance of the model. Predictions for given covariate combinations can be obtained as a separate step. It is also possible to output "the Bernoulli trial"-type data into a file for further processing by SAS, BMDP, GLIM, or other statistical software to provide influence statistics etc. Since the number of these trials can be very large (in this data there are about 86 000 months of exposure), BMDPLR or SAS PROC CATMOD can be used to combine trials with the same covariate combination. This works well with dichotomous design-type covariates but is less useful with continous covariates.

Model 1 (Table 1) with 23 covariates and intercept requires about 6 minutes of CPU-time on an IBM 3083 EX2 under VM/SP CMS-operating system, starting from zero initial values and with a time grid of 1 month. Models with less covariates (Table 3) use about 1 to 3 minutes of CPU. To save computer resources a wider time grid value of e.g. 5 months can be used to obtain good initial values of parameter estimates. In our experience a wider time grid usually leads to nearly the same estimates (excepting, of course, the intercept $\hat{\beta}_1$).

The program is written entirely in standard FORTRAN 77 without using machine-dependent features (except for time and date). The current installation works under IBM VM/SP CMS-operating system, but should be easily adapted to other environments. The program consists of 23 modules with a total of about 3300 lines of code. It is extensively commented; in fact half of the lines are comments. A short manual is provided with sample data and results. It should be noted that even though the program is self-contained, it is designed to be used with accompanying software : for example it has no facilities of subsetting the observations or transforming the covariates, nor does it have a command language. The specification of time dependent covariates usually requires the help of other software. We have utilized the versatility of SAS (Statistical Analysis System) to convert this data into a suitable form.

Readers interested in obtaining the program may request it by writing to the authors.

Appendix: Details of the statistical model

Let $(\Omega,\mathcal{F})$ be a measurable space in which the variables $Y_j(t-1)$, $\underset{\sim}{Z}_j(t-1)$ and $\Delta N_j(t)$ are defined, and let $R(t-1) = \{j:Y_j(t-1)=1\}$ be the risk set at t (i.e., during interval $(t-1,t]$). Let $\mathcal{F}_0$ be the σ-field representing "initial information"; usually $\mathcal{F}_0$ is the trivial field. Then the σ-fields $\mathcal{F}_t$ and $\mathcal{H}_{t-1}$, $t\geq 1$, defined inductively by

$$\mathcal{H}_{t-1} = \mathcal{F}_{t-1}\ \sigma\{\ R(t-1),\ \{\underset{\sim}{Z}_j(t-1);\ j \varepsilon R(t-1)\}\}\ ,$$

$$\mathcal{F}_t = \mathcal{H}_{t-1}\ \sigma\{\ \Delta N_j(t);\ j\ \varepsilon\ R(t-1)\}\ ,$$

represent the experimental history registered up to time t, $\mathcal{F}_t$ including and $\mathcal{H}_{t-1}$ excluding the responses $\Delta N_j(t)$ during $(t-1,t]$.

Consider a statistical model $\{P^\theta;\ \theta\ \varepsilon\ \Theta\}$ for the observation process $(R(s-1),\ \{\Delta N_j(s), \underset{\sim}{Z}_j(s-1); j\varepsilon R(s-1)\})_{s\geq 1}$ and a P^θ-likelihood which corresponds to data collected up to time t, $t\geq 1$.

Suppose that the parameter θ can be represented in the form $\theta = (\theta_1,\theta_2)$, where θ_1 is the parameter of interest and θ_2 is a nuisance parameter. Typically, we think of θ_1 as parametrizing the conditional distribution of the variables $\Delta N_j(s)$, conditioned on $\mathcal{H}_{s-1}$, and of θ_2 as the parameter associated with the conditional law of the variables $R(s)$ and $\underset{\sim}{Z}_j(s)$ $(j\ \varepsilon\ R(s))$, given $\mathcal{F}_{s-1}$. It is then a simple cosequence of the chain multiplication rule of conditional probabilities that the full likelihood corresponding to the observed values $\{r(s-1),\{\Delta n_j(s),\ \underset{\sim}{z}_j(s-1);\ j\ \varepsilon\ r(s-1)\};\ s \leq t\}$ can be expressed as the product of two terms, viz., as

$$\prod_{s\le t} P^{\theta}(R(s-1) = r(s-1), \underset{\sim}{Z}_j(s-1) = \underset{\sim}{z}_j(s-1);\ j \in r(s-1) \mid \mathcal{F}_{s-1})$$

$$\cdot \prod_{s\le t} P^{\theta}(N_j(s) = \Delta n_j(s);\ j \in r(s-1) \mid \mathcal{H}_{s-1}). \tag{A.1}$$

Following Cox (1975), the second factor can be called a partial likelihood. Ordinary ML-estimation of θ_1, the parameter of interest, can be done by considering that factor alone provided that the following condition holds:

<u>Assumption 1.</u> (i) For each $s \ge 1$, the conditional P^{θ}-distribution of $(R(s-1), \{\underset{\sim}{Z}_j(s-1);\ j \in R(s-1)\})$, given $_{s-1}$, does not depend on θ_1;
(ii) For each $s \ge 1$, the conditional P^{θ}-distribution of $(\Delta N_j(s);\ j \in R(s-1))$, given $\mathcal{H}_{s-1}$, does not depend on θ_2.

Of course, the validity of Assumption 1 depends on the model $\{P^{\theta};\ \theta \in \Theta\}$. Actual verification of this assumption would require that the model were fully specified, including the probability law of the censoring mechanism and possible random covariates. This is usually not done explicitly. However, part (ii) of Assumption 1 becomes obvious if the censoring times and the covariates are fixed, or random but $\mathcal{F}_0$-measurable. More generally, we can consider (ii) to be valid if the censoring is non-informative about θ_1 and the covariates are external (cf. Kalbfleisch and Prentice (1980)). For internal covariates more caution is needed: If (i) is not met, also the first factor in (A.1) can depend on θ_1, and then using only the second factor in the maximization is a potential source of bias. Finally, it seems that part (ii) in Assumption 1 can always be met in practice by making a convenient choice of θ_1, the parameter of interest.

For a continuous time version of Assumption 1, see Arjas and Haara (1984).

Our next assumption imposes an independence condition between the individuals and simplifies, in particular, the handling of ties.

<u>Assumption 2.</u> For each $s \ge 1$, and $\theta \in \Theta$, the random variables $\{\Delta N_j(s);\ j \ge 1\}$ are conditionally P^{θ}-independent given $\mathcal{H}_{s-1}$.

This assumption is likely to hold in practice if there are no multiple responses of common cause, or if such responses can occur but the background variable causing the failure can be included as a covariate.

Under Assumptions 1 and 2 the likelihood function (A.1) depends on θ_1 only through the factor

$$\prod_{s \le t} \prod_{j \varepsilon r(s-1)} P^{\theta}(\Delta N_j(s) = \Delta n_j(s) | \mathcal{H}_{s-1}). \qquad \text{(A.2)}$$

On the other hand, because of Assumption 1 (ii), this expression does not depend on θ_2.

It remains to specify the conditional probabilities in (A.2). Our next assumption guarantees that all relevant information in $\mathcal{H}_{s-1}$, when used as a condition for the probability of $\{\Delta N_j(s) = \Delta n_j(s)\}$, is actually contained in the p-vector $\underset{\sim}{Z}_j(s-1)$ and the indicator $Y_j(s-1)$.

<u>Assumption 3</u>. For all $s \ge 1$, $j \ge 1$ and $\theta \,\varepsilon\, \Theta$, $\Delta N_j(s)$ and $\mathcal{H}_{s-1}$ are conditionally P^{θ}-independent given $Y_j(s-1)$ and $\underset{\sim}{Z}_j(s-1)$.

As a last step, the conditional probabilities in (A.2) are specified in formula (2). There we also change the notation slightly, writing $\underset{\sim}{\beta} = (\beta_1, \ldots, \beta_p)'$ instead of θ_1 and Pr instead of P^{θ}.

<u>Acknowledgement</u>

It was Jan Hoem's initiative that we should undertake this small study. We are grateful to him, and to Britta Hoem, for letting us use their data, and for their many helpful comments during the preparation of the paper.

<u>References :</u>

(1) Allison, P.D. (1982). Discrete time-methods for the analysis of event histories. In: Social Methodology 1982 (S. Leinhardt, Ed.), Jossey-Bass, San Francisco.

(2) Arjas, E. (1986). Stanford heart transplantation data revisited : A real time approach. In : "Modern Statistical Methods in Chronic Disease Epidemiology" (S.H. Moolgavkar and R.L. Prentice, Eds.), Wiley.

(3) Arjas, E. and Haara, P. (1987). A logistic regression model for hazard. Scandinavian J. of Statistics 14,1-18 .

(4) Hoem, B. and Hoem J. (1987). The impact of female employment on second and third births in modern Sweden. Stockholm Research Reports in Demography No. 36, University of Stockholm.

(5) McGullagh, P. and Nelder, J.A. (1983). Generalized Linear Models. London: Chapman and Hall.

(6) Morgan, S.P., Lye, D.N. and Condran G.A. (1988). Sons, daughters and divorce: does the sex of children affect the marital disruption? American J. of Sociology (to appear).

(7) Rindfuss, R.R., Morgan, S.P. and Swicegood C.G. (1984). The transition to motherhood: the intersection of structural and temporal dimensions. American Sociological Review 49, 359-372.

(8) Sandefur, G.D. and Thuma, N.B. (1987). How data type affects conclusions about individual mobility. Social Science Research 16, 301-328.

TABLE 1 Third births model 1 : Deviance 5004.26 (logit link)

Parameter index	Covariate name	Parameter estimate	Estim. std. dev.	Param. /std.dev.	Exp(parameter)	Relative intensity in H&H
1	Constant	-5.44	0.17	-31.98	0.0043	
2	Indicator(time <= 24)	-0.32	0.11	-2.98	0.73	
3	Indicator(time >= 73)	-0.98	0.21	-4.70	0.37	
4	Birth-interval <= 29	0.72	0.10	6.97	2.06	2.06
5	Birth-interval >= 54	-0.98	0.22	-4.37	0.37	0.38
6	Education "high"	0.55	0.15	3.64	1.73	1.73
7	Cohort 1936-40, Age 16-19	0.49	0.25	1.95	1.62	1.63
8	Cohort 1941-45, Age 16-19	0.67	0.19	3.61	1.95	1.94
9	Cohort 1946-50, Age 16-19	0.52	0.18	2.82	1.68	1.68
10	Cohort 1936-40, Age 20-25	0.27	0.15	1.84	1.32	1.32
11	Cohort 1946-50, Age 20-25	-0.49	0.16	-3.02	0.61	0.61
12	Cohort 1936-40, Age 26-34	-0.12	0.23	-0.55	0.88	0.89
13	Cohort 1941-45, Age 26-34	-0.52	0.22	-2.35	0.59	0.60
14	Cohort 1946-50, Age 26-34	-0.39	0.36	-1.11	0.67	0.70
15	<25 %, Full time work	-0.11	0.21	-0.51	0.90	0.92
16	25-75 %, Full time work	-0.41	0.40	-1.03	0.66	0.69
17	>75 %, Full time work	-0.82	0.72	-1.13	0.44	0.45
18	25-75 %, Part time work	-0.22	0.24	-0.92	0.80	0.83
19	>75 %, Part time work	-1.21	0.52	-2.33	0.30	0.31
20	<25 %, Childminder	0.48	0.52	0.92	1.62	1.67
21	25-75 %, Childminder	-0.01	0.41	-0.02	0.99	1.02
22	>75 %, Childminder	-0.15	0.60	-0.25	0.86	0.87
23	25-75 %, Housewife	0.57	0.19	2.94	1.77	1.79
24	>75 %, Housewife	0.13	0.15	0.84	1.14	1.15

TABLE 2 Third births Model 1 : Estimated asymptotic correlation matrix

	1	2	3	4	5	6	7	8	9	10	11	12
1	1.000											
2	-0.187	1.000										
3	-0.136	0.170	1.000									
4	-0.195	-0.043	-0.009	1.000								
5	-0.207	-0.005	0.005	0.226	1.000							
6	-0.341	-0.031	0.026	-0.056	0.063	1.000						
7	-0.163	-0.010	0.001	-0.083	-0.035	0.093	1.000					
8	-0.253	-0.035	0.007	-0.031	-0.015	0.147	0.219	1.000				
9	-0.258	-0.028	0.011	-0.048	-0.039	0.145	0.217	0.288	1.000			
10	-0.369	-0.003	-0.008	-0.026	0.006	0.100	0.253	0.331	0.331	1.000		
11	-0.353	-0.010	0.042	-0.089	0.046	0.038	0.215	0.289	0.293	0.362	1.000	
12	-0.219	0.007	-0.015	0.013	0.033	-0.098	0.139	0.185	0.183	0.250	0.233	1.000
13	-0.272	-0.016	0.028	-0.051	0.038	-0.113	0.135	0.182	0.184	0.252	0.254	0.193
14	-0.142	-0.116	0.019	-0.036	0.031	-0.110	0.079	0.108	0.110	0.153	0.162	0.126
15	-0.392	0.007	-0.015	-0.047	-0.019	-0.096	-0.002	-0.038	-0.011	-0.005	0.037	-0.029
16	-0.274	0.025	-0.029	-0.014	0.017	0.068	0.005	-0.037	-0.013	0.018	0.048	0.014
17	-0.149	0.008	-0.046	0.000	0.014	0.044	-0.029	-0.017	-0.009	0.002	0.015	0.001
18	-0.433	0.056	-0.056	-0.049	0.013	0.108	0.008	-0.030	-0.036	0.033	0.013	0.032
19	-0.216	0.029	-0.060	-0.015	0.014	0.074	-0.013	-0.009	-0.000	0.016	0.030	0.011
20	-0.219	-0.005	-0.019	-0.015	0.016	0.070	0.022	0.031	-0.024	0.057	0.026	0.036
21	-0.273	0.052	0.001	-0.061	0.010	0.112	-0.008	-0.088	-0.029	0.043	0.022	0.021
22	-0.191	0.017	0.004	-0.021	-0.005	0.081	0.012	-0.030	-0.005	0.021	0.011	0.010
23	-0.519	-0.093	0.023	-0.046	0.021	0.180	0.004	-0.007	-0.023	0.012	0.021	0.005
24	-0.696	-0.032	0.038	-0.119	0.096	0.267	-0.077	-0.049	-0.035	0.010	0.084	0.008

	13	14	15	16	17	18	19	20	21	22	23	24
13	1.000											
14	0.147	1.000										
15	0.052	0.022	1.000									
16	0.055	0.037	0.181	1.000								
17	0.028	0.017	0.100	0.064	1.000							
18	0.069	0.040	0.299	0.184	0.104	1.000						
19	0.035	0.020	0.138	0.090	0.052	0.146	1.000					
20	0.042	0.013	0.134	0.084	0.047	0.141	0.067	1.000				
21	0.029	0.015	0.175	0.118	0.065	0.191	0.092	0.087	1.000			
22	0.029	0.019	0.118	0.078	0.043	0.127	0.061	0.059	0.084	1.000		
23	0.055	0.040	0.362	0.217	0.121	0.357	0.168	0.171	0.223	0.153	1.000	
24	0.108	0.075	0.460	0.290	0.164	0.471	0.227	0.221	0.306	0.203	0.594	1.000

TABLE 3 Third births Model 2 : Deviance 5017.58 (logit link)

Parameter index	Covariate name	Parameter estimate	Estim. std. dev.	Param. /std.dev.	Exp(parameter)
1	Constant	-5.43	0.11	-48.77	0.0044
2	Indicator(time <= 24)	-0.29	0.10	-2.70	0.75
3	Indicator(time >= 73)	-1.02	0.21	-4.91	0.36
4	Birth-interval <= 29	0.71	0.10	6.91	2.03
5	Birth-interval >= 54	-1.01	0.22	-4.54	0.36
6	Education "high"	0.48	0.14	3.46	1.62
7	Cohort 1936-40	0.22	0.12	1.85	1.24
8	Cohort 1946-50	-0.32	0.12	-2.63	0.72
9	Age 16-19	0.68	0.12	5.63	1.98
10	Age 26-34	-0.37	0.15	-2.43	0.69
11	>75%,Full t. & > 75% Part t.	-0.91	0.32	-2.81	0.40
12	<25%,Childm. & 25-75% Housew	0.53	0.15	3.54	1.69

Figure 1

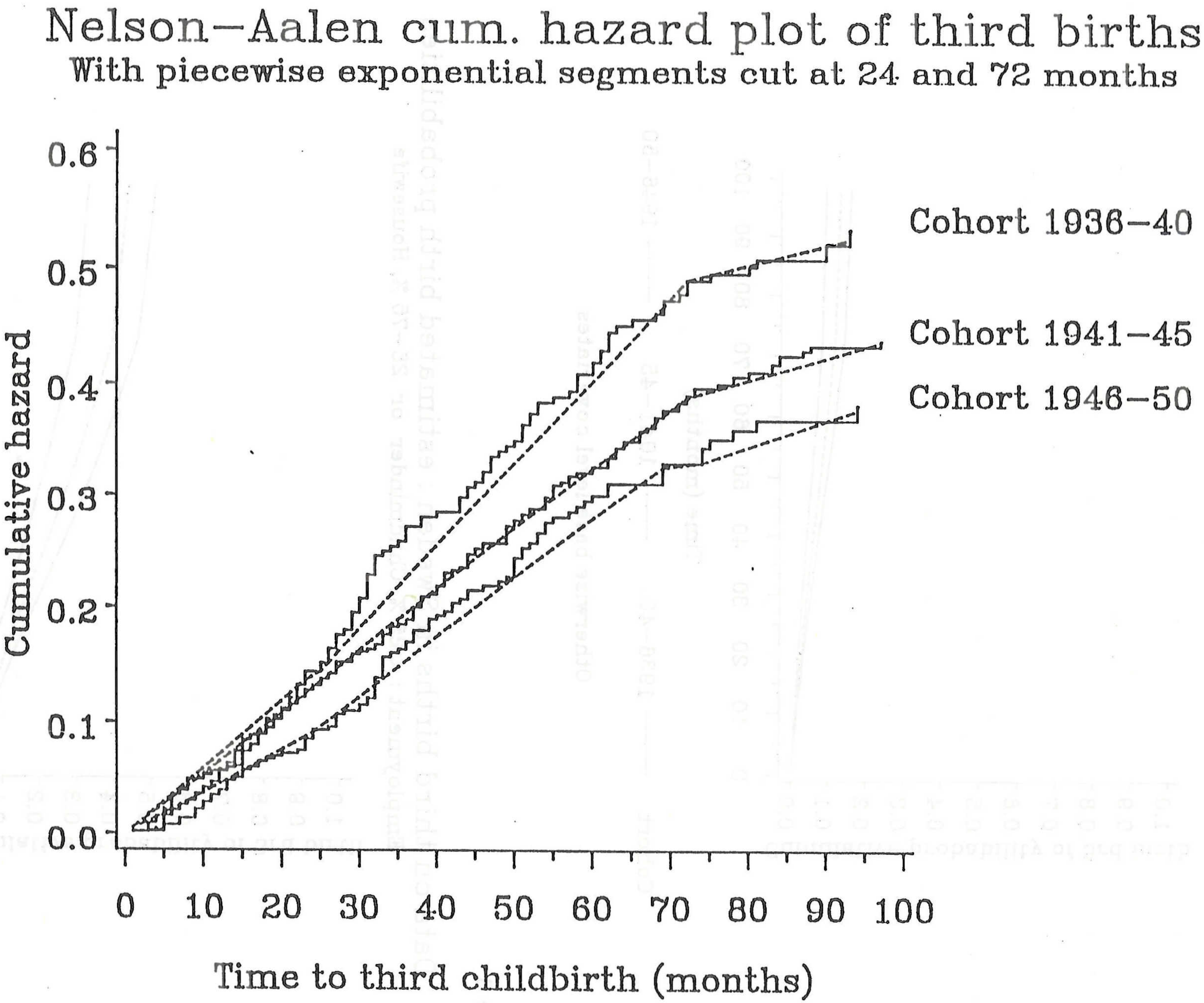
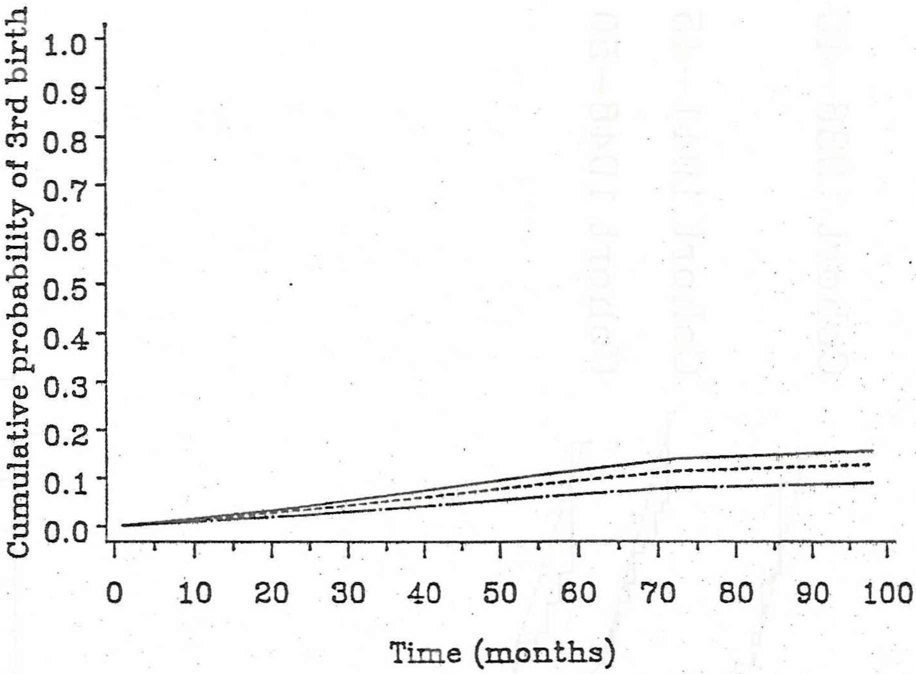

Nelson–Aalen cum. hazard plot of third births
With piecewise exponential segments cut at 24 and 72 months
Cumulative hazard
0.0
0.1
0.2
0.3
0.4
0.5
0.6
Cohort 1936–40
Cohort 1941–45
Cohort 1946–50
0
10
20
30
40
50
60
70
80
90
100
Time to third childbirth (months)

Figure 2

Data on third births in Sweden : estimated birth probabilities
Employment : > 75 %, Full time or > 75 %, Part time

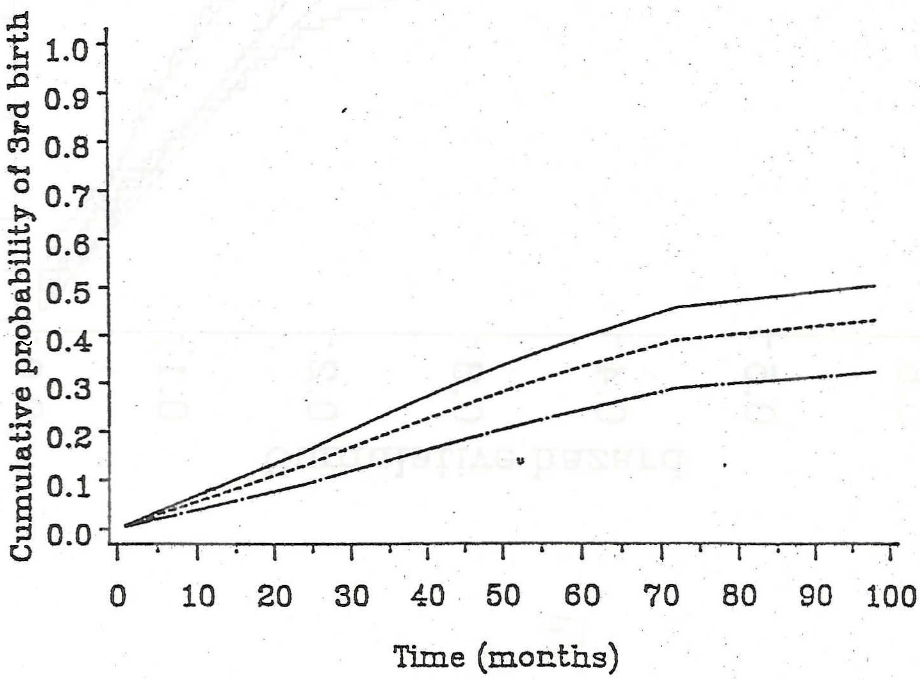


Cohort —— 1936–40 -------- 1941–45 —·— 1946–50

Otherwise base level covariates

Figure 3

Data on third births in Sweden : estimated birth probabilities
Employment : < 25 %, Childminder or 25–75 %, Housewife



Cohort —— 1936–40 -------- 1941–45 —·— 1946–50

Otherwise base level covariates

STOCKHOLM RESEARCH REPORTS IN DEMOGRAPHY

SECTION OF DEMOGRAPHY, UNIVERSITY OF STOCKHOLM

S-106 91 STOCKHOLM, SWEDEN

SRRD-1 Hoem, Jan M. and Randi Selmer: The interaction between premarital cohabitation, marriage, and the first two births in current Danish cohorts, 1975. (March, 1982.)

SRRD-2 Hoem, Jan M. and Ulla Funck Jensen: Multistate life table methodology: A probabilistic critique. (March, 1982.) Pp. 155-264 in K. C. Land and A. Rogers (eds.), Multidimensional Mathematical Demography. New York: Academic Press.

SRRD-3 Hoem, Jan M.: Balancing bias in vital rates due to an informative sampling plan. (April, 1982.) Pp 81-88 in Statistical Review 1983:5. Essays in honour of Tore E. Dalenius.

SRRD-4 Hoem, Jan M.: Distortions caused by nonobservation of periods of cohabitation before the latest. (May, 1982.) Demography 20 (4), 491-506, 1983.

SRRD-5 Hoem, Jan M.: The reticent trio: Some little-known early discoveries in insurance mathematics by Oppermann, Thiele, and Gram. (July, 1982.) International Statistical Review 51, 213-221, 1983.

SRRD-6 Hoem, Jan M. and Bo Rennermalm: Biases in cohabitational nuptiality rates caused by nonobservation of periods of cohabitation before the latest: An empirical note. (Dec., 1982.)

SRRD-7 Funck Jensen, Ulla: An elementary derivation of moment formulas for numbers of transitions in time-continous Markov chains. (Dec., 1982.)

SRRD-8 Hoem, Jan M. and Bo Rennermalm: Cohabitation, marriage, and first birth among never-married Swedish women in cohorts born 1936-1960. (Dec., 1982.) Published as "Modern family initiation in Sweden: Experience of women born between 1936 and 1960". European Journal of Population 1, 81-112, 1985.

SRRD-9 Funck Jensen, Ulla: The Feller-Kolmogorov differential equation and the state hierarchy present in models in demography and related fields. (Dec., 1982.)

SRRD-10 Hoem, Jan M.: Multistate mathematical demography should adopt the notions of event-history analysis. (Feb., 1983.)

SRRD-11 Hoem, Jan M.: Weighting, misclassification, and other issues in the analysis of survey samples of life histories. (Feb., 1983.) Pp. 249-293 in James J. Heckman and Burton Singer (1985), Longitudinal analysis of labor market data. Cambridge: Cambridge University Press.

SRRD-12 Lyberg, Ingrid: Nonresponse effects on survey estimates in the analysis of competing exponential risks. (April, 1983.)

SRRD-13 Hartmann, Michael: Past and recent experiments in modeling mortality at all ages. (Nov., 1983.) Journal of Official Statistics 3, 19-36, 1987.

SRRD-14 Lyberg, Ingrid: The effects of sampling and nonresponse on estimates of transition intensities: Some empirical results from the 1981 Swedish fertility survey. (Dec., 1983.)

SRRD-15 Hoem, Jan M.: A flaw in actuarial exposed-to-risk theory. (Jan., 1984.) Scandinavian Actuarial Journal 1984, 187-194.

SRRD-16 Hoem, Jan M.: Statistical analysis of a multiplicative model and its application to the standardization of vital rates: A review (Feb., 1984.) International Statistical Review 1987, 55, 119-152.

SRRD-17 Lindberg, Bengt: A numerical algoritm for the calculation of coefficients in moving averages (March, 1984.)

SRRD-18 Hoem, Jan M., Bo Rennermalm, and Randi Selmer (revised version) Restriction biases in the analysis of births and marriges to cohabiting women from data on the most recent conjugal union only. (April, 1986.) Pp. 555-600 in Karl Ulrich Mayer and Nancy Tuma (eds.). Applications of Event History Analysis in Life Course Research. Berlin: Max-Planck-Institut fur Bildungsforschung, 1987.

SRRD-19 Hoem, Jan M.: Marriages connected with first births among cohabiting women in the Danish fertility survey of 1975. (April, 1984.)

SRRD-20 Bernhardt, Eva: Den sociala bakgrundens betydelse för barnafödande och giftermål inom informellt samboende bland svenska kvinnor födda 1936-60. (Oct., 1984.)

SRRD-21 Etzler, Cecilia: Första steget i familjebildningen - Gifta sig, börja sambo eller få barn: Utveckling och skillnader bland svenska kvinnor med olika social bakgrund födda 1936-60. (Nov., 1984.)

SRRD-22 Hoem, Britta: Regional utveckling i det moderna samboendet, analysmetoder och resultat. (Nov., 1984.)

SRRD-23 Bernhardt, Eva and Britta Hoem: Cohabitation and social background: Trends observed for Swedish women born 1936-60. (Feb., 1985.) <u>European Journal of Population</u> 1, 375-395, 1985.

SRRD-24 Holmlund, Bertil: Fertility and the opportunities of children: Evidence on quantity-quality interactions in Sweden. (Sept., 1984.) <u>Scandinavian Population Studies</u> 7,100-107, 1985.

SRRD-25 Hoem, Britta: Ett barn är inte nog. Vad har hänt med svenska ettbarnskvinnor födda 1936-60? (June, 1985.)

SRRD-26 Kravdal, øystein: Flytting fra foreldrehjemmet: Regionale og sosiale forskjeller blant svenske kvinner f¿dt 1936-60. (June, 1985.)

SRRD-27 Hoem, Jan M.: The impact of education on modern union initiation (Aug., 1985.) <u>European Journal of Population</u> 2, 113-133, 1986.

SRRD-28 Bernhardt, Eva M.: Women's home attachment at first birth: The case of Sweden. (Oct., 1985.) European Journal of Population 2 , 5-29, 1986.

SRRD-29 Etzler, Cecilia: Barnafödande inom äktenskapet: Det föräktenskapliga samboendets betydelse för förstfödslar inom äktenskapet bland svenska kvinnor födda 1936-60. (Jan., 1986.)

SRRD-30 Hoem, Britta: Tredjebarnsfödslar bland svenska kvinnor födda 1936-50. (Aug., 1986.)

SRRD-31 Borgan, Ørnulf and Jan M. Hoem: Demographic reproduction rates and the estimation of an expected total count per person in an open population. (Dec., 1986.) To appear in The Journal of the American Statistical Association.

SRRD-32 Hoem, Britta: Sysselsättningshistoriens betydelse för tvåbarns-mödrars fortsatta barnafödande bland svenska kvinnor födda 1936-50. (Jan, 1987.)

SRRD-33 Bernhardt, Eva M: Continued home attachment versus transition to employment: An intensity regression analysis of Swedish one-child mothers. (Jan., 1987.)

SRRD-34 Etzler, Cecilia: Education, Cohabitation and the First Child: Some Empirical Findings from Sweden. (March, 1987.) To appear in print.

SRRD-35 Pleiborn, Maria: Overgångar in i och ut ur utbildning bland unga svenska kvinnor födda 1936-1960. (March, 1987.)

SRRD-36 Hoem, Britta and Jan M. Hoem: The impact of female employment on second and third births in modern Sweden. (April, 1987.) To appear in Population Studies.

SRRD-37 Hoem, Jan M. and Per Linnemann: The tails in moving average graduation. (May, 1987.) To appear in The Scandinavian Actuarial Journal.

SRRD-38 Hoem, Britta: Graviditetens betydelse för barnlösa svenska kvinnors civilståndsändringar. (May, 1987.)

SRRD-39 Hoem, Jan M.: The issue of weights in panel surveys of individual behavior. (May, 1987.)

SRRD-40 Bernhardt, Eva M: The choice of part-time work among Swedish one-child mothers. (Sept., 1987.)

SRRD-41 Bernhardt, Eva M: Labour force participation and childbearing: The impact of the first child on the economic activity of Swedish women. (Sept., 1987.)

SRRD-42 Hoem, Britta and Jan M. Hoem: Patterns of deferment of first births in modern Sweden. (Sept, 1987.)

SRRD-43 Hoem, Britta and Jan M. Hoem: The Swedish family: Aspects of contemporary developments. (Oct., 1987.) To appear in the Journal of Family Issues.

SRRD-44 Etzler, Cecilia: Första barnet. En demografisk studie av barnafödandet bland svenska kvinnor födda 1936-60. (Dec., 1987.)

SRRD-45 Hoem, Britta and Jan M. Hoem: Dissolution in Sweden. The break-up of conjugal unions to Swedish women born in 1936-60. (March 1988)

SRRD-46 Bernhardt, Eva M.: Changing family ties, women's position and low fertility. (Sept. 1988)

SRRD-47 Hoem, Britta: Early phases of family formation in contemporary Sweden. (August 1988)

SRRD-48 Klingberg, Lars: Bartlett-type correction terms for tests in intensity regression models. (November 1988)

SRRD-49 Arjas, Elja and Kangas, Pekka: A discrete time method for the analysis of event histories. (September 1988)

SRRD-50 Hoem, Britta: En simuleringsstudie av fördelningen till likelihoodkvottestet för en intensitetsmodell med få observationer. (December 1988)

SRRD-51 Hoem, Britta: I vilken utsträckning har tvåbarnsmödrarna i den svenska fruktsamhetsundersökningen från 1981 fått barnen vid önskade tidpunkter och vad har det betytt för tredjebarnsfruktsamheten? (December 1988)

Support from the Swedish Council for Research in the Humanities and Social Sciences for the publication of the Stockholm Research Reports in Demography is gratefully acknowledged.

Also available on request:

Bernhardt, Eva (1985). Nya samlevnadsformer: Arbetardöttrar, stockholmskor och norrländskor bildade förtrupp. Välfärdsbulletinen, SCB, 1985:1, 18-19.

Hoem, Jan M.: Sveriges hundraåriga fruktsamhetsfall. Pp. 185-212 in Barn? Författare, forskare och skolungdom diskuterar varför det föds så få barn. Stockholm: Liber Förlag, 1983.

Hoem, Jan M.: Forsikringsmatematik. Pp. 229-235 in Københavns Universitet 1479-1979, Bind XII. København: G.E.C. Gads Forlag, 1983.

Hoem, Jan M.: A contribution to the statistical theory of linear graduation. Insurance: Mathematics and Economics 3, 1-17, 1984.

Hoem, Jan M. and Randi Selmer: The negligible influence of premarital cohabitation on marital fertility in current Danish cohorts, 1975. Demography 21 (2), 193-206, 1984.

Gustafsson, Siv: Equal opportunity policies in Sweden, in Schmid, G. and Renate Weitzel (eds.). Sex Discrimination and Equal Opportunity. The Labour Market and Employment Policy. Gower, Hampshire England, 1984.

Kravdal, Øystein. Flickorna flyttar hemifrån allt tidigare. Välfärdsbulletinen, SCB, 1985:3, 14-15.

Leighton, Linda and Siv Gustafsson: Differential patterns of unemployment in Sweden. Research in Labor Economics 6, 251-285, 1984.

Klingberg, Lars: Some large sample properties of the Cox regression model: A two-state example. Report no 146. Nov. 1987.