Population Review

Volume 44, Number 1, 2005

Type: Article pp. 11-35

The Use of Flexible Parametric Duration Functions in Modelling the Tempo of Fertility: Applications to the Analysis of Birth Intervals in Rural China

Authors: Gebrenegus Ghilagaber, Ph.D., Stephen Obeng Gyimah, Ph.D., Petra Löfstedt, and Annika Johansson, Ph.D.
Affiliations: Dr. Ghilagaber, Department of Statistics, Stockholm University, Stockholm, Sweden (Dr. Ghilagaber); Department of Sociology, Queen's University, Kingston, ON, Canada (Dr. Gyimah); Department of Public Health Sciences, Division of International Health, Karolinska Institutet, IHCAR, SE-171 76 Stockhom, Sweden (Petra Lofstedt and Dr. Annika Johansson)
Corresponding author/address: Dr. Ghilagaber, Department of Statistics, Stockholm University (Dr. Ghilagaber); SE-106 91 Stockholm, Sweden (Gebre@stat.su.se)

Abstract

The birth-interval approach to the study of fertility reflects two aspects of the process of reproduction: (1) the quantum of fertility as indicated by the proportion of women who move to the next higher parity; and (2) the tempo of fertility, as measured by the time it takes to make the transition for those women who continue reproduction. In most previous empirical analyses, the focus has been on the quantum of fertility using proportional hazard models for the intensity of birth. That is, the rates at which children are born to a defined set of women within a specified unit of time is modelled as a function of covariates such that the effect of covariates is to increase or decrease such intensity relative to that of a reference category. This paper focuses on the tempo of fertility where covariates act multiplicatively on the duration itself so that their effect is to accelerate or decelerate the transition time between successive births relative to that of a reference category. We utilize the flexibility of a family of parametric duration models to select a statistically appropriate model for a given birth interval using data from rural China. The results show that the distributional shape of birth intervals depends on the birth order of the index child and that inferences concerning covariate effects on birth intervals are sensitive to model choice. The flexible parametric approach suggested in this paper provides a statistically well grounded, theoretically appropriate, and empirically evident alternative to the usually untenable quantum-based proportional hazards modelling of birth interval data.

Keywords

Tempo of fertility, quantum of fertility, child-spacing, public policy, rural China, parametric duration models, accelerated failure-time models, proportional hazards models, nested models, likelihood ratio test

1. Introduction

Studies on age at marriage and length of successive birth intervals have acquired added importance in the development of fertility theories and understanding of fertility transitions, because of their supposed relationship to fertility decline; the later the first birth (the longer the interval), the lower the total fertility (Hirshman and Rindfuss 1982; Rindfuss, Palmore, and Bumpass, 1987; Hirshman 1994; Bongaarts 1999). There is considerable evidence that both the postponement of first marriage and the lengthening of birth intervals are important components of historical fertility transitions as well as contemporary fertility declines (Hirshman 1994). Several inferences are consistent with the view that in much of the developing world, women with large families tend to have shorter birth intervals than those with smaller families.

The spacing of births also has a significant bearing on maternal and child health through the dynamics of sibling competition and maternal depletion hypotheses (Gribble, 1993; Hobcraft, McDonald & Rustein, 1985; Majumder, May & Pant, 1997; Palloni & Millman, 1986; Pedersen, 2000; Rafalimanana and Westoff, 2000; Rodriguez, Hobcraft, Menken & Trussell, 1984). According to the competition hypothesis, the birth of each successive child generates competition for scarce resources among siblings in the household which subsequently leads to a lower quality of care and attention to each child. The family resources may also be stretched to the limit, increasing the probability of children in such households becoming malnourished (Gribble, 1993). The maternal depletion syndrome contends that births in rapid succession physiologically deplete the mother of energy and nutrition which may lead to premature births or pregnancy complications thus increasing the risk of infant or maternal death, or impairing the mother’s ability to nurture her children. Additionally, women with closely spaced births may still have very young children and, as such, are less likely to attend prenatal care services which, in turn, may increase maternal and child mortality risks. Further, the early arrival of a new child often necessitates the premature weaning of the previous child, exposing the weaned one to malnutrition and increasing the child’s vulnerability to infectious and parasitic diseases. Invariably, longer birth spacing has been found to profoundly increase the probability of infant survival (Bicego & Ahmad, 1996; Defo, 1997; Pedersen, 2000). Understanding the timing and spacing of births thus provides a thorough view of attitudes toward family size, policy impacts, as well as differentials in fertility and childhood mortality levels. A deeper analysis of the dynamics of child spacing is, therefore, a topic of interest that worth investigating.

The birth interval approach to studying fertility views the family building process as consisting of a series of transitions, where women move successively from single-life to marriage; from marriage to first birth; from first to second birth; and so on, until they reach their completed family size (Rodriguez et al., 1984). While the point of entry into the process may be defined as the legitimate age to marriage or as entry into motherhood, the main focus of this analysis is on the process of transition from one stage to the next, or the intervals between successive births. The transition process is studied in terms of the birth function, defined as the cumulative proportion of women having a birth by successive duration since the previous birth (or since marriage in the case of first birth). This function reflects two aspects of the process of reproduction. The first is the quantum of fertility indicated by the proportion of women progressing to the next higher parity (parity progression ratio), and the second is the tempo of fertility measured by the time it takes to make the transition for those women who continue reproduction.

In most empirical analyses of birth interval data, the focus has been on the quantum of fertility (that is, the rates at which children are born to a defined set of women within a specified unit of time) using proportional hazard models of the type discussed in Cox (1972). The proportional hazards model (Cox, 1972) specifies the intensity of birth as a function of an unspecified time dependent baseline hazard, $\lambda_0(t)$, and the covariates,

$$\lambda(t, z) = \lambda_0(t)\exp(z\beta) \qquad (1)$$

where, z is a vector of covariates and β is a vector of unknown regression parameters. While the nuisance baseline specification makes the Cox model attractive particularly in contexts where the focus is not on the time function, the model may be too restrictive because the assumption of proportional hazards is often unrealistic in many real life situations. Also, there are instances where one's research interest centers on the distributional shape of the time function and thus calling for alternative models.

In this paper, we present a second class of models, more akin to ordinary linear regression, that specifies the covariates to act multiplicatively on tempo of fertility (or linearly on log-tempo) rather than on the quantum (rate of birth). We demonstrate how a number of common parametric duration models like the Weibull and log-normal may be embedded in a single parametric framework, and how each special-case model may be assessed relative to a more comprehensive one. This class of models is then applied on birth interval data from Yunnan province in rural China with a view to examining the distributional shape of birth intervals and the sensitivity of inferences to the choice of a model. In the development of fertility theories and understanding the timing of fertility transitions, China is of particular interest since its major family planning programs 'the later-longer-fewer (wan-xi-shao)' campaign in the 1970s, and the one-child policy, introduced in 1979, have emphasized both delayed marriage and childbearing and longer spacing between the first and second child (Feeney and Wang 1993).

The paper thus has both methodological and substantive objectives. The methodological objective centres on the application of a flexible family of parametric survival models to the analysis of birth interval data. The second set of objectives, of a substantive nature, relate to examining correlates of birth spacing in rural China. In the next section, we introduce the class of flexible parametric duration models and describe how covariates effects are estimated in such models. In Section 3, we fit this family of models to birth interval data from Yunnan province in rural China and discuss the results, while Section 4 summarises the contents of the paper.

2. Family of Flexible Parametric Duration Functions

2.1 Accelerated failure-time models for the tempo of fertility

Suppose we denote by T_0 the time (say, birth interval in months) associated with the baseline category corresponding to zero-values for the covariate ($z = 0$). Such baseline levels may, for example, be women with no education (to be later compared with those having some primary- or secondary-level education) or urban residents (to be compared with rural residents). Then, the accelerated tempo model specifies that if the vector of covariates had been z ($z \neq 0$), the corresponding time (time to next birth) is

$$T = T_0\exp(z\beta), \qquad (2)$$

or, equivalently,

$$\ln T = \ln T_0 + z\beta \qquad (3)$$

where, T is the vector of birth intervals (durations), z is a vector of covariates, and β is a vector of unknown regression parameters. Since covariates alter, by a scale factor, the rate at which an average woman traverses the time axis, (2) may be referred to as the accelerated failure time model (accelerated tempo of fertility in the context of this paper). Thus, for proportional hazards model (1), the explanatory variables act multiplicatively on the baseline rate so that their effect is to increase or decrease the rate of birth relative to the baseline rate $\lambda_0(t)$. For accelerated tempo models, on the other hand, the explanatory variables act multiplicatively on time to the event (birth in our case) so that their effect is to accelerate or decelerate transition time to birth relative to that of the baseline category (T_0).

The model in (3) is a linear model with $\ln T_0$ playing the role of an error term with an underlying baseline distribution. Usually, an intercept term α and a scale parameter δ are allowed in the model to give

$$\ln T = \alpha + z\beta + \delta \ln T_0. \tag{4}$$

In terms of the original (untransformed) times to birth, the effect of the intercept term and the scale factor are to scale and power the time to birth, respectively:

$$T = \exp(\alpha + z\beta + \delta \ln T_0) = T_0^{\delta}\exp(\alpha)\exp(z\beta). \tag{5}$$

In other words, the effect of covariates in an accelerated tempo model is to change the scale, but not the location, of a baseline distribution of birth times.[1]

2.2 The choice between alternative baseline distributions

As we saw above, the model for the response variable (4) consists of a linear effect composed of the covariates together with a random disturbance term. Such models may be rewritten more explicitly as

$$\ln T = z\beta + \delta\varepsilon \tag{6}$$

in which the intercept is incorporated in the coefficient vector β and a more conventional notation is used for the random error term. The distribution of the random error term can be taken from a class of distributions that includes the extreme-value, normal, and logistic distributions, and by using a log-transformation, exponential, Weibull, log-normal, log-logistic and gamma distributions. In general, the distribution may depend on additional shape parameter k.

Embedding competing models in a single parametric framework allows the methods of ordinary parametric inference to be used for discrimination and leads to an assessment of each competing model relative to a more comprehensive one. Stacy (1962) showed that the generalized gamma model could be useful in this regard. The generalized-gamma model is the distribution of T such that $\ln T = z\beta + \delta\varepsilon$, where the random error term ε has the density;

[1] A point worth noting at this stage is that the parameterizations in (1) and (2) are different. A positive coefficient in (1) implies an increased birth intensity (shorter interval) while in (2) it implies longer interval (decreased intensity) relative to that of the baseline level.

$$f(k,\varepsilon) = \frac{1}{\Gamma(k)} \exp\{k\varepsilon - \exp(\varepsilon)\}, \quad -\infty < z\beta < \infty,\ -\infty < \varepsilon < \infty, \text{ and } \delta, k > 0. \qquad (7)$$

Prentice (1974) showed that a transformation of the form $w = k^{½}(\varepsilon - \ln(k))$ leads to a standard normal distribution for w as $k \to \infty$. Further, he extended the generalized gamma distribution by setting $q = k^{-½}$ and by allowing the error density at -q to be a reflection, about the origin, of that of q. The parameter q = k-½ was chosen as the unique power of k that leads to finite, nonzero likelihood derivatives at the log-normal model for T.
The final model with parameters $-\infty < z\beta < \infty$, $-\infty < q < \infty$, and $\delta > 0$, can be written as

$$\ln T = z\beta + \delta\varepsilon,$$

where the error density function $f(q, \varepsilon)$ is

$$f(q,\varepsilon) = \begin{cases} \dfrac{|q|}{\Gamma(q^{-2})} (q^{-2})^{q^{-2}} \exp\left[q^{-2}\{q\varepsilon - \exp(q\varepsilon)\}\right], q \neq 0 \\ (2\pi)^{-½} \exp(-\dfrac{\varepsilon^2}{2}), q = 0 \end{cases} \qquad (8)$$

The distribution of T, when the error term has the density (8) will henceforth be called the Extended Generalized Gamma (EGG) distribution. As can be seen from the lower part of (8), the EGG model reduces to the standard normal distribution for ε when the shape parameter q is equal to zero. Accordingly, T will have a log-normal distribution. When the shape parameter q equals 1, (8) reduces to $f(1,\varepsilon) = f(\varepsilon) = \exp\{\varepsilon - \exp(\varepsilon)\}$, $-\infty < \varepsilon < \infty$, which is the standard (type 1) extreme-value distribution. As lnT is a linear function of ε, it has the same (extreme-value) distribution as ε. Hence $T = \exp(z\beta + \delta\varepsilon)$ will have a Weibull distribution. If q = 1 and δ = 1, then T has the exponential distribution as a special case of the Weibull distribution. The case of q = -1 corresponds to extreme maximum-value distribution for lnT. This, in turn, corresponds to reciprocal Weibull distribution for T. The case of δ = 1 and q > 0 is also of interest. Farewell and Prentice (1977) argue that this gives the ordinary gamma distribution for T, though, in accordance with Bergström and Edin (1992) and Ghilagaber (2005), this does not hold in our case illustration. Consequently, we shall label this special case (δ = 1, q > 0) the 'gamma' distribution in our illustrative example.

Thus, five models for T are included as special cases of the EGG model. Since each of these five models is nested within the EGG model, its goodness of fit to the data, in relation to the more comprehensive EGG model, may be assessed through standard likelihood ratio tests. Another model of interest, though not a special case of the EGG model, is the log-logistic model. A log-logistic distribution is the distribution of T such that logT follows a logistic distribution. Description and applications of the log-logistic model may be found in Shoukri, Mian, and Tracy (1988), Singh, Lee, and George (1988), Diekmann (1992) or Blossfeld and Rohwer (2002).

2.2 Estimation

The practical estimation of (6) proceeds as follows. Consider survival times of n individuals $t_1, t_2, ..., t_n$ and p covariates $z_1, z_2, ..., z_p$. Let d_i take value 0 if t_i is a censoring time and value 1 if t_i represents an event time. The log-likelihood function $\ln(\ln t; z\beta, \delta, q)$, assuming a noninformative censoring mechanism, will then be proportional to

$$\sum_{i=1}^{n} d_i \{\ln f(\varepsilon; q) - \ln \delta\} + \sum_{i=1}^{n} (1-d_i) \ln S(\varepsilon_i; q), \qquad (9)$$

where f(ε;q) is given by the EGG model (8), S(ε;q) is the corresponding survivor function, and $\varepsilon_i = [\ln(\ln t_i) - z_i\beta]/\delta$. At each of several q-values the maximum likelihood estimates $\hat{\beta}(q)$, and $\hat{\delta}(q)$ are obtained by using the Newton-Raphson method to solve the normal equations arising from (9). Standard errors of coefficients may be obtained from the information matrix as usual. Below, we demonstrate the procedures described in this section through analysis of data from rural China.

3. Analysis of birth-spacing in Yunnan province, rural China.

3.1 Data Set

The data on which our analyses is based come from interviews conducted in Huaning County, Yunnan province in the southwest of China, in May 2000. The total population of Huaning in 2000 was 198,000, with over 90 percent engaged in agriculture. The main ethnic group is Han and about one third belong to minority groups, mainly Yi, Hui, and Miao. Huaning county was purposively selected, based on the local authorities willingness to cooperate.

In Rural China, each county is divided into townships and each township into administrative villages. These consist of a number of 'natural villages', which are composed of groups or 'clusters' of houses lying close to each other. Based on lists provided by local authorities, multistage cluster sampling techniques were used to randomly select three out of the five townships, ten out of 48 administrative villages and in these, half of all 'natural villages' were selected. In each 'natural village' all households were visited, which altogether included around 2000 eligible women. Of these, 1503 were at home at the time of our visit.

All women who were at home at the first visit agreed to participate in the study, after having been explained the purpose and that they were free to decline. The interviews were conducted in the women's home by female health workers from the area, with extensive experience from previous surveys. We used pre-coded questionnaires including details on marriage, births, abortions, contraceptive use, the woman's education and occupation and husband's age. In the present study, the events of interest are first marriage, and first-, second, and third births, along with their timing and sociodemographic correlates. The interviews were reviewed each day-by the research supervisors and checked for internal consistency. After cleaning the data set, checking for missing values and internal inconsistencies, 1326 cases remained out of the 1503.

3.2 Background of the Study Area

Huaning is a mountainous area located in the southeast part of Yunnan. When the one-child policy was introduced in 1979, Yunnan was one of the poorest provinces in China. It had a lower life expectancy at birth, a higher total fertility rate, higher percent of third or higher parity births and higher crude birth rate compared to China as a whole (Banister 1987; Li 1990; Bignami-Van Assche 2003; UNESCAP 2004). The family planning program in Yunnan during the 1970s was basically on a pilot work level, but the crude birth rate dropped from 38 per 1000 in 1971 to 32 in 1976. When the one-child policy was launched, the provincial natural rate of increase of 19 per 1000 was considered unacceptably high and family planning organizations were expanded. In 1984, the program was gradually applied to the whole province. In 1990, the crude birth rate had dropped to 23.60, and in 2000 it had dropped further to 20.00, but still higher than the national averages (17.50 in 1990 and 16.00 in 2000) (UNESCAP 2004).

3.3 Correlates of Birth-Spacing

We now fit the models discussed in section 2 above to our data set on interval lengths to study the distributional shape of these intervals, discriminate among special-case models. , and identify potential correlates of birth-spacing. Related questions concern the dependence of inference about these correlates on the distributional shape of the duration variable (birth intervals), and the dependence of model choice on the order of birth interval.

The main dependent variables for this analyses are transition times to marriage, to motherhood, as well as to 2nd and 3rd births, all measured in months. For comparison purposes, we shall also fit the Cox model (1) in which the dependent variable is the birth rate (or marriage rate) at a given time. Generally, differences in birth intervals can be explained through demographic, socio-economic and socio-cultural factors. On the basis of previous work (Löfstedt, Ghilagaber, Shusheng, and Johansson, 2005; Löfstedt, Ghilagaber, and Johansson, 2005), we have selected an array of theoretically relevant variables as likely covariates of birth intervals. These include ethnicity, religion, maternal education and occupation, mother's birth cohort and marriage cohort, age at first marriage, survival status of the index child, and sex (or sex composition) of previous child(ren).

A cohort is indicative of structural factors that have shaped the life of individuals. At the macro level, similar life experiences can be detected among women belonging to the same cohort despite subtle micro level differences. Given the changing contextual factors affecting reproduction in rural China, we expect the younger cohorts, who became adolescents in a period of a more egalitarian gender role, efficient contraceptives, and higher female enrolments in formal education, to have longer intervals than older cohorts.

Age at first marriage is also of tremendous importance in fertility studies because of its inverse relation to the exposure to the risk of conception (see, e.g., Gyimah, 2003; Westoff, 1992). It also represents a number of unmeasured factors that predispose women to differential timing of births and, thus, overall fertility. Women who marry at younger ages are likely to come from disadvantaged socio-economic backgrounds and are thus more likely to be associated with higher risks of births than their counterparts whose first marriages occurs late (Gyimah, 2001). Consequently, we expect women who marry early to be associated with shorter intervals.

Also significant in determining the length of the inter-birth interval is the survival status of the index child (Montgomery and Cohen, 1998; Preston, 1978). It has been demonstrated that intervals following the death of the index child tend to be significantly shorter than intervals where the child survived, a result of biological and behavioural processes (Gyimah and Fernando, 2002). We thus expect the death of the index child to be associated with shorter intervals. Sex of previous child or sex composition of first two children is also important particularly in societies like rural China where son preference is strong. Thus, we would expect woman whose first (or first two) births resulted in girl(s) to move faster to the next parity than their counterparts whose first birth(s) results in at least one boy.

There is also considerable empirical evidence that associates high levels of maternal education with low fertility. The pathways through which these happen have been explained through an array of mechanisms including late age at marriage, greater knowledge and access to contraception, high labour force participation and alternative values regarding family size (Cochran, 1979; 1983; Martin, 1995; Ware 1984).

3.4 Descriptive results

There were 1326 usable records of women of which 1101 (83%) were married by interview time while the rest 225 (17%) were still single and treated as censored. Of the married women, 1052 (97%) have borne their first child while the rest were with zero-parity at interview time. Further, of the 1052 women with at least one child, 792 (75%) have proceeded to 2nd child, while the rest 262 (25%) were still single-parity women by the survey time. Finally, 534 (68%) of the parity-two women have had a 3rd child while the rest 254 (32%) were still at parity two. Table 1 presents a cross tabulation of frequencies of the various events of interest across the relevant covariates.[2]

3.5 Covariate effects

In Tables 2 - 5, we report results of fitting models (1) and (8) to the data on transitions to marriage, first birth, second birth, and third birth, respectively. In each table, the estimated coefficients in the first seven columns under 'parametric models' come from fitting the model in (8), as well as the log-logistic model and, hence, represent effects of the respective levels of a factor on log-interval (that of baseline level is set to 0.000 by design). Estimates given in the column 8 (labelled Cox) are related to model (1) and, as such, measure the effect of the covariates on the log-intensity of transition to marriage or to the next higher parity (birth intensity). The last column is just relative intensities (intensity ratios, or hazard ratios) obtained by exponentiating the estimates in the Cox model.

It may be worth noting that the shape and scale parameters are free (estimated from the data) in the more comprehensive EGG model, while in the five special case-models, one or both of these parameters are set to some fixed value(s) as discussed in Section 2. According to Table 2, for example, the factors that significantly extend the time to marriage (decrease the intensity of marriage) are having some education, and belonging to the younger birth cohort. Such results are reported by most models though the reciprocal Weibull model also show some ethnic and religious differentials while the exponential model is more conservative with respect to the cohort-effects. We shall examine, later, if these models have adequate goodness of fit.

The effects of the various covariates on transition to 1st, 2nd, and 3rd births are shown in Tables 3, 4, and 5, respectively. According to Table 3, those who marry older (at or after 20 years of age) have shorter durations to first birth (after marriage) as compared to those who marry as teenagers. Further, the younger the marriage cohort the shorter the interval between marriage and first birth. Lastly, women who are younger than their husbands by at least two years get their first child faster (after marriage) as compared to those women who are at least as old as their husbands. These findings are consistently reported by almost all models though they differ slightly in the strength of the effects and that the reciprocal Weibull model indicates some ethnic differentials.

In Table 4, the only important factors are age at marriage and date of 1st birth. Women who married older and those who got their first child after 1985 seem to postpone their 2nd births, while those who got their first child in the first half of the 1980s have shorter intervals to 2nd birth (after first birth). The results for those who got their first child after 1985 may, however,

[2] In Table 1, 79 of the 1101 women who made transition to marriage were excluded because they either had unknown values on some variables related to first birth (49 cases) or their first child was conceived or born before they married (30 cases). As a result, the usable records of women for the analyses of first birth was only 1022. However, the 30 women who conceived or borne their first child before marriage were included in the analysis of 2nd births, giving rise to 1052 usable records.

be partly explained by the fact that a number of these woman who married around the survey time may not have yet reached to the stage of giving rise to a 2nd child.

Table 5 shows that women with some education postpone 3rd births relative to those with no education. Those who married older, got their 2nd child after 1980, or already have at least one son among their first two children also tend to postpone 3rd births relative to their counterparts. On the other hand, women who lost their 2nd baby tend to replace it immediately. Again, these results are reported consistently by three models (EGG, reciprocal Weibull and “Gamma” models) while the other models deviate somehow. Below, we outline procedures to demonstrate, among other things, that the first set of models are equally adequate in explaining variations across durations to 3rd birth.

3.6 Discrimination among parametric models

When parametric models are nested, likelihood ratio tests can be used to assess the best fit model (Heckman and Walker, 1991; Allison, 1995). The likelihood-ratio statistics (and associated p-values) corresponding to various tests for special cases of the EGG model (8) are presented in the last 2 rows of Tables 2 – 5. These are used to test whether the corresponding special-case model is adequate enough relative to the more comprehensive EGG model.

The results for marriage (Table 2), for instance, show that all special cases are rejected in favour of the more general EGG model. This is in accordance with the estimated value of the shape parameter (0.397) under the EGG model. This estimate is closer to the assertion of the log-normal where the shape parameter is fixed to 0 than to any value set by the other alternative distributions. The question is whether or not this value is statistically different from 0. A simple guide is to standardise it through dividing it by its standard error (not shown in Table 2): 0.397/0.056 = 7.12, which is by far larger than any table value in the standard normal distribution. Thus, the estimated shape parameter (0.397) is significantly different from 0. In fact the value of the Chi-square statistic reported at the bottom of Table 2 (corresponding to the log-normal column) is the square of this standardised value (7.12**2 = 51). The same is true in Table 3 - all special case distributions are rejected in favour of the EGG model. Once again, the log-normal model is the closest to the EGG model, and this is consistent with the estimate of the shape parameter under the EGG model (0.136) which is closest to 0. But how close should it be to zero? Again, a standardisation (0.136/0.055 = 2.47) shows that it is marginally different from 0 at 5% significance level, but not at 1% level of significance. Note again that the square of this value (2.47**2 = 6.10) is equal to the corresponding value of the Chi-square as reported in Table 3 where the marginality of the significance is shown by the small p-value (0.013).

The results for 2nd birth (Table 4) also show that the EGG model is the most parsimonious model. The estimate of the shape parameter (-0.317) is close to 0 (suggesting a log-normal model), but again it is significantly different from zero (-0.317/0.070 = -4.51) and that (–4.51)**2 = 20.30 is equal to the corresponding value of the Chi-square shown in Table 4 (save some round-off errors).

A different picture is shown in Table 5 which reports results related 3rd births. Here, the reciprocal-Weibull and the “Gamma” models are as adequate as the parent EGG model in explaining variations in 3rd birth intervals. As a result, one can save one degree of freedom by fixing the shape parameter to –1 or the scale parameter to 1. This is shown by the insignificant Chi-square statistics corresponding to these models as shown at the bottom of Table 5 (0.218 with a p-value of 0.641, and 0.162 with a p-value of 0.687, respectively).

The equivalence of these three models is also clearly indicated by the estimated shape and scale parameters in the EGG model in Table 5. The estimated shape parameter, -1.10, is very close to –1 (corresponding to the assertion of the reciprocal-Weibull model) while the estimated scale parameter, 1.02 is close to 1 (corresponding to the assertion of the “Gamma”). The tests outlined above also indicate that these values are close enough to their asserted values.

Table 6 contains a summary of the different models under the null and alternative hypotheses and the corresponding Chi-square statistics (Likelihood Ratio Tests) together with their associated p-values.

3.7 Determinants of time-to-marriage and child-spacing

As we noted in the above section(s) the most adequate model for a given set of duration data depends on the specific data set at hand. Thus, for our data set, the EGG model is the most parsimonious model for marriage duration and durations between 1st and 2nd births. On the other hand, the reciprocal Weibull and “Gamma” models are as adequate as the EGG model in explaining variation across durations between 2nd and 3rd births. Lastly, the log-normal model is marginally close to the EGG model in explaining variations across duration between marriage and 1st birth.

More importantly, we have noted that inferences concerning covariate effects on a given time variable depend on model choice. We have, for instance, found, in Table 4, that the number of variables with significant effect on the 2nd birth interval and their strength differs between the columns of the EGG and the Weibull models.

In Table 7, therefore, we have provided a summary of the results for the four intervals, all obtained from the EGG model in order to facilitate comparisons. Although the number and type of covariates included in the models are not the same across the four intervals, some general comparisons can be attempted.

We note, from Table 7, that women from younger birth cohorts and those with some education tend to postpone marriage. Women with some education do not, however, seem to compensate for this late marriage by shortening the interval to first birth after marriage. The category of women who tend to shorten the time to first birth are those who married older, belong to the younger marriage cohorts (specially after 1980) and those who are “too younger” than their husbands (younger by at least two years). Those who married at age 24 or older, and those who got their first child after 1985 also postpone 2nd births. Postponement of 3rd births is also common among those who married old and those who got their 2nd child after 1980. Education does not seem to be important for time to the first two births, while it shows a significant association with time to 3rd birth. Women with some education tend to postpone 3rd birth. Women who married older, those who got their 2nd child after 1980, and those with at least one son among their first two children also have longer durations to 3rd birth, while those who lost their 2nd child seem to replace it shortly.

3.8 Hazard model or duration model?

The main goals in the analyses of survival (duration) data are: to describe the distribution of the time (duration) variable, to compare the survival experiences (distributions) of different groups of a population, and to investigate explanatory variables that could affect the survival (duration). The survival experience of a given group of individuals is often described by three different but equivalent functions: the density function, denoted by f(t), from which the probability of experiencing the event of interest within a given interval of time is obtained; the hazard function, denoted by $\lambda(t)$, which is the instantaneous rate at which the event of interest occurs; and the survivor function, denoted by S(t), which is the probability of surviving beyond time t (not experiencing the event of interest until time t). These functions are computed for our specific data set and shown in Figures 1, 2, and 3, respectively. The figures clearly indicate that the four time intervals under study do not have the same form of distribution – something worth considering while fitting some model to the data sets.

Further, as described earlier hazard-rate models measure the quantum of the event under investigation (marriage or child-birth in our present case) and express the rate of occurrence of such event as a function of some covariates. In duration models (accelerated failure time models), on the other hand, the interest is the tempo of the event and, thus, the dependent variable is the duration of the event.

The most popular hazard-rate model is the Cox proportional hazards model (Cox, 1972; 1975). This model leaves the baseline hazard in Equation (1) unspecified. This is a very attractive property and it is mainly for this reason that almost all investigators in the applied sciences feel comfortable in using this model for their data analytic purposes.

However, not many investigators care of what is behind the scene – namely the assumption of proportional hazards across the various levels of a covariate. For instance, a proportional hazards assumption across the three levels of Education (None, Compulsory or less, Above Compulsory, in our data set) would mean that the hazard functions of these levels would look like what is shown in Figure 4. A plot of the actual hazard functions for our data set, however, is what is shown in Figure 5. A comparison of Figures 4 and 5 shows that the assumption of proportional hazards is violated for the covariate Education. Is it then advisable to use the Cox proportional hazards model in such circumstances? While there are some studies that indicate that the model is robust to modest violations of the assumption, the current authors are of the opinion that one should resort to the more appealing alternative models described in this paper.

3.9 Are Inferences Sensitive to Model-Choice?

In order to compare the results from our duration models with those obtained from the Cox PH model, we have also reported results from fitting the Cox model (1) in each of the tables in the Appendix. As indicated earlier (see footnote 1), a positive effect in the duration models should imply a negative effect (and hence a relative hazard that is less than 1) in the hazard models (and vice-versa). A closer look at Table 6 shows that this is the case in most, but not in all, the comparable pair of columns. Even when the directions of the effects are in the expected directions, there are situations where the strength of effects of some covariates differs across the two sets of models. In situations when they are not compatible, we recommend that inferences be based on the results from the appropriate duration model because there is at least a statistical evidence supporting the choice of such models.

4. Summary

A natural question arises as to which model to fit or which procedure to use when one is confronted with a specific data-analysis problem. As with most statistical or demographic methods, it is rather difficult to codify the procedures involved in choosing a model. There are many factors, such as mathematical convenience, theoretical appropriateness, and empirical evidence that should legitimately enter the decision and none can be easily quantified. Given the wide range of fertility models in the literature, it is worth asking whether conclusions are sensitive to the particular model chosen. The answer to this question is unknown until results obtained with one method have been compared to those obtained by another method. Such comparisons have been one of the objectives of the present paper. Our empirical results indicate that the distributional shape of birth intervals is different depending on whether we refer to the first or higher-order birth intervals. More importantly, our results demonstrate that inferences concerning covariate effects on birth intervals are sensitive to the choice of the distributional shape.

In sum the results indicate that the choice of the appropriate distribution of birth intervals is of crucial importance in order to make valid inference and, thereby, suggest sound and effective policy interventions. This paper has outlined a statistically well grounded, theoretically appropriate, and empirically evident procedure on how to identify the most appropriate model for a given data set on duration times. Based on such procedure, we have also identified determinants of time-to-marriage and child-spacing in a remote county of rural China.

This study is, however, not without limitations. Unlike death, the events studied in this paper (marriage and childbearing) are not certain events to all individuals in the long run – there may be long-term survivors in the sense of Maller & Zhou (1996). Accordingly, alternative models that allow for this feature, such as Li and Choe (1997), Yamaguchi (2003), or Land, Nagin, and McCall (2001), could be appropriate. It is our ambition to address this issue in the near future. At this stage, however, we believe that such long-term survivors, if any, are too few in the context of our study that failure to address them does not bring about any substantial differences in our results.

Meanwhile, it is our hope that the findings in this paper bring into the surface the importance of how to specify duration phenomena. This, in turn, is expected to motivate researchers to look for stronger links between the underlying reality and the models we present. One such issue is to investigate what behavioural or biological processes are better represented by one model than another and what sorts of bias would one expect to observe in estimated effects if those processes are not appropriately modelled.

References

Allison, Paul D. (1995), *Survival Analysis Using the SAS System*. Cary N.C.,: SAS Institute Inc.

Banister J. (1987*), China's Changing Population*. Stanford: Stanford University Press, Stanford, p. 227-269.

Bicego, G., and O. B. Ahmad (1996), Infant and Child Mortality. *DHS Comparative Studies* **No. 20**. Macro International, Demographic and Health Surveys: Calverton, Maryland.

Bignami-Van Assche S. "Estimating life expectancy at birth for China's provinces, 1975-1990: A Note. " Available: http://www.ssc.upenn.edu/~sbignami/Papers/Note-Estimates%20of%20e(0)%20in%20China.pdf (Accessed: 2004, December 01)

Bergström, R., and P. A. Edin (1992), Time aggregation and the distributional shape of unemployement duration. *Journal of Applied Econometrics* **7**, 5-30.

Blossfeld, H. –P., and Rohwer, G. (2002), *Techniques of event history modeling : new approaches to causal analysis* (2nd edition). Mahwah, NJ : Lawrence Erlbaum Associates,

Bongaarts J. (1999), The fertility Impact of Changes in the Timing of Childbearing in the Developing World. *Population Studies* **53**, 277-289.

Cochrane, S. (1979), *Fertility and Education: What Do We Really Know?* Baltimore: John Hopkins University Press.

Cochrane, S. (1983), Effects of education and urbanization on fertility. In R.A. Bulatao and R.D. Lee (eds.), *Determinants of Fertility in Developing Countries. A Summary of Knowledge*, Part B. New York: Academic Press.

Cox, D. R. (1972), Regression models and life-tables (with discussion). *Journal of the Royal Statistical Society - Series B* **34**, 187-220.

Cox, D. R. (1975), Partiall Likelihood. *Biometrika* **62**, 269-276.

Defo, K. B. (1997), Effects of infant feeding practices and birth spacing on infant and child survival: A reassessment from retrospective and prospective data. *Journal of Biosocial Science* **29**, 303-323.

Diekmann, A. (1992), The log-logistic distribution as a model for social diffusion processes. *Journal of Scientific and Industrial Research* **51**, 285.

Farewell, V. T., and R. L. Prentice (1977), A study of distributional shape in life testing. *Technometrics* **19**, 69-75.

Feeney, G., and F. Wang (1993), Parity Progression and Birth Intervals in China: The Influence of Policy in Hastening Fertility Decline. *Population and Development Review* **19**, 61-100.

Ghilagaber, G. (2005), The Extended Genralized Gamma Model and its Special Cases: Applications to Modelling Marriage Durations. *Quality & Quantity – International Journal of Methodology*, 71 – 85.

Gribble, J.N. (1993), Birth intervals, gestational age and low birth weight: Are the relationships confounded? *Population Studies* **47**,133-146.

Gyimah, S. O (2001), *Childhood mortality and reproductive behavior in Ghana and Kenya: An Examination of Frailty and Non-frailty Models*. Ph.D Dissertation, Faculty of Graduate Studies, The University of Western Ontario, London, Canada.

Gyimah, S. O. and Fernando, R. (2002), The physiological effects of infant death on fertility: A comparative analysis of DHS data from Ghana and Kenya. *Social Biology* **49**, 44-57.

Gyimah, S.O. (2003). A cohort analysis of timing of the first birth and cumulative fertility in Ghana. *Population Research and Policy Review*, **22**, 251-266.

Heckman, J. and J. Walker (1991), Economic models of fertility dynamics: A case study of Swedish fertility. *Research in Population Economics* **7**, 3-91.

Hirshman C. (1994), Why Fertility Changes. *Annual Review of Sociology* **20**, 203-233.

Hirshman, C., and R. R. Rindfuss (1982), The Sequence of Timing of Family Formation Events in Asia. *American Sociological Review* **47**, 660-680.

Hobcraft, J., McDonald, J. and S. Rustein (1985), Demographic determinants of early infant and early child mortality: A comparative analysis. *Population Studies* **39**, 363-385.

Land, K. C., Nagin, D. S., and McCall, P. L. (2001), Discrete-time hazard regression models with hidden heterogeneity - the semiparametric mixed Poisson regression approach. *Sociological Methods and Research* **29**, 342-373.

Li, Bohua (1990), Levels, Trends, and Determinants of Fertility in China: 1973-1987. *Asia-Pacific Population Journal* **5**, 3-16.

Li, L., and Choe, M. K. (1997), A mixture model for duration data: Analysis of second births in China. *Demography* **34**, 189-197.

Löfstedt, P., Ghilagaber, G., Shusheng, L., and Johansson, A. (2005), Gender Inequality and the Transition from First Marriage to First Birth in Huaning County, Yunnan Province, China. To appear in *South East Asian Journal of Medicine and Tropical Hygiene*.

Löfstedt, P., Ghilagaber, G., and Johansson, A. (2005), Birth spacing patterns in Huaning Country, Yunnan Province, PRC. Manuscript under review by *Asia-Pacific Population Journal*.

Majumder, A. H., May, M. and P. D. Pant (1997), Infant and child mortality determinants in Bangladesh: Are they changing?, *Journal of Biosocial Science* **29**, 385-399.

Maller, R. A., and Zhou S. (1996), *Survival Analysis with Long-Term Survivors*, John Wiley & Sons, New York.

Martin, T. (1995), Women's education and fertility: Results from 26 Demographic and Health Surveys. *Studies in Family Planning* **26**,187-202.

Montgomery, M. and B. Cohen (eds.)(1998), *From Death to Birth: Mortality Decline and Reproductive Change*. Washington, D.C.: National Academy Press.

Palloni, A. and S. Millman (1986), Effects of inter-births intervals and breastfeeding on infant and early age mortality. *Population Studies* **40**, 215-236.

Pederson, J. (2000), Determinants of infant and child mortality in the West Bank and Gaza strip. *Journal of Biosocial Science* **32**, 527-546.

Prentice, R. L. (1974), A log-gamma model and its maximum likelihood estimation. *Biometrika* **61**, 539-544.

Preston, S.H. (ed.) (1978), *The Effects of Infant and child Mortality on Fertility*. New York: Academic Press.

Rafalimanana, H., and C. E. Westoff (2000), Potential effects on fertility and child health and survival of birth-spacing preferences in sub-Saharan Africa. *Studies in Family Planning* **31**, 99-110.

Rindfuss RR, Palmore JA, Bumpass LL. (1987), Analyzing Birth Intervals: Implications for Demographic Theory and Data Collection. *Sociological Forum* (2) 4:811-226.

Rodriguez, G., Hobcraft, J., Menken, J., and J. Trussell (1984), Analysis of the determinants of birth intervals. *WFS Comparative Studies* **No 30**.

Shoukri, M. M., Mian, I. U. H., and D. S. Tracy (1988), Sampling properties of estimators of the log-logistic distribution with application to Canadian precipitation data. *The Canadian Journal of Statistics* **16**, 223

Singh, K. P., Lee, C. M. -S., and E. O. George (1988), On generalized log-logistic model for censored survival data. *Biometrical Journal* **7**, 843.

Stacey, E. W. (1962), A generalization of the gamma distribution. *Annals of Mathematical Statistics* **33**, 1187-1192.

Ware, H. (1984), Effects of maternal education, women's roles and child care on child mortality. in Henry Mosley and Lincoln C. Chen (eds.), *Child Survival: Strategies for Research*. The Population Council: New York.

Westoff, C. F (1992), Age at marriage, age at first birth, and fertility in Africa. *World Bank Technical Paper*, **No. 169**, World Bank: Washington, D.C.

UNESCAP, Yunnan Province. Available: http://www.unescap.org/esid/psis/population/database/chinadata/yunnan.htm (Accessed 2004, November 30)

Yamaguchi, K. (2003), Accelerated failure-time mover-stayer regression models for the analysis of last-episode data. *Sociological Methodology* **33**, 81-110.

Table 1: Summary Statistics of Variables for the Different Transitions

	Marriage		First Birth		Second Birth		Third Birth	
	N	% event	N	% event	N	% event	N	% event
Ethncity								
Han	1200	83	926	98	952	76	718	32
Yi	87	90	69	94	73	73	52	29
Other	39	74	27	93	27	67	18	33
Religion								
No Religion	1208	82	919	97	937	73	680	28
Budha	74	97	65	100	74	95	70	67
Don't know	44	95	38	95	41	93	38	37
Education								
None	157	97	137	100	152	94	143	71
Compulsory	1103	82	851	97	865	72	623	24
Above complusory	66	62	34	94	35	66	22	9
Occupation								
Farm	1279	83	984	97	1014	76	762	32
Non-farm	47	74	37	97	38	68	26	42
Birth cohort								
1936-40	31	100	-	-	-	-	-	-
1941-45	35	100	-	-	-	-	-	-
1946-50	72	100	-	-	-	-	-	-
1951-55	124	100	-	-	-	-	-	-
1956-60	81	99	-	-	-	-	-	-
1961-65	173	100	-	-	-	-	-	-
1966-70	234	100	-	-	-	-	-	-
1971-75	238	93	-	-	-	-	-	-
1976-80	226	57	-	-	-	-	-	-
1981-85	112	2	-	-	-	-	-	-
Age at Marriage			153	99	286	90	257	59
Before 20 years	-	-	307	96	349	74	258	24
20-21	-	-	407	98	262	69	179	16
22-23	-	-	155	94	155	61	94	11
24 years and above	-	-						
Marriage cohort								
Before 1970	-	-	124	100	109	99	134	93
1970-79	-	-	150	100	176	98	161	69
1980-84	-	-	106	100	107	98	110	10
1985-00	-	-	642	95	660	62	383	2
Age difference with spouse								
Wife older	-	-	104	95	-	-	-	-
Same age	-	-	239	97	-	-	-	-
Husband 2-4 yrs older	-	-	423	97	-	-	-	-
Husband 5+years older	-	-	256	99	-	-	-	-
Introduction								
Self	-	-	448	96	-	-	-	-
Parents/ friends	-	-	180	98	-	-	-	-
Match maker	-	-	394	98	-	-	-	-
Sex of First child								
Boy	-	-	-	-	550	76		
Girl	-	-	-	-	502	75		
Sex of 1st & 2nd child								
Two girls	-	-	-	-			190	45
One each	-	-	-	-			394	28
Two boys	-	-	-	-			204	29
Survival Status of first child								
Alive	-	-	-	-	1030	75	780	32
Dead	-	-	-	-	22	100	8	88
TOTAL	1326	83	1022	97	1052	75	788	68

See also footnote 2 in Section 3.4 of the text for further details on Table 1

Table 2: Estimated Effects of Covariates on Log Duration to Marriage (months after age 15) and on the Log Hazard of Marriage

COVARIATES	PARAMETRIC MODELS (Effects on Log Duration)							Cox PH Model (Effects on Log hazard)	
	Extended Generalized Gamma	Reciprocal Weibull	Log Normal	Weibull	Gamma	Exponential	Log logistic	Coefficient	Hazard Ratio
Constant	4.031***	3.450***	3.851***	4.299***	3.838**	4.030***	3.867***	.	.
Scale parameter	0.378***	0.516	0.404	0.363	1.000	1.000	0.214	.	.
Shape parameter	0.397***	-1.000	0.000	1.000	-0.073	1.000	-	.	.
Ethncity									
Han (reference)	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	1.000
Yi	0.042	0.034	0.031	0.471	0.004	0.019	0.265	-0.159	0.853
Other	-0.047	-0.290***	-0.017	-0.057	-0.095	-0.061	-0.042	0.043	1.044
Religion									
No Religion (reference)	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	1.000
Budha	-0.079*	-0.095	-0.074	-0.083*	-0.067	-0.074	-0.089*	0.194	1.214
Don't know	-0.083	0.018	-0.052	-0.134**	-0.055	-0.093	-0.038	0.309	1.362*
Education									
None (reference)	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	1.000
Compulsory	0.117***	0.262***	0.162***	0.029	0.162*	0.111	0.165***	-0.272	0.761***
Above complusory	0.362***	0.483***	0.417***	0.247***	0.511**	0.493***	0.412***	-0.930	0.394***
Occupation									
Farm (reference)	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	1.000
Non-farm	0.018	0.104	0.019	0.005	0.036	0.028	0.037	-0.048	0.953
Birth cohort									
1936-40 (reference)	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	1.000
1941-45	-0.197**	-0.430***	-0.161***	-0.319***	-163.000	-0.194	-0.088	0.970	2.637***
1946-50	-0.070	0.175	0.011	-0.181**	0.019	-0.063	0.019	0.558	1.747***
1951-55	0.134*	0.291***	0.174***	0.081	0.176	0.136	0.166**	-0.228	0.796
1956-60	0.308***	0.494***	0.349***	0.330***	0.355*	0.317	0.339***	-0.545	0.579**
1961-65	0.283***	0.518***	0.355***	0.212***	0.355*	0.284	0.350***	-0.481	0.618**
1966-70	0.376***	0.575***	0.442***	0.274***	0.450**	0.377*	0.437***	-0.719	0.487***
1971-75	0.344***	0.457***	0.413***	0.250***	0.447**	0.398*	0.409***	-0.674	0.509***
1976-80	0.332***	0.602***	0.427***	0.228***	0.660***	0.724***	0.384***	-0.607	0.545***
1981-85	0.526***	0.716***	0.543****	0.716***	1.648***	3.307***	0.507***	-1.596	0.203**
Log Likelihood	-623	-898	-649	-675	-1181	-1291	-602		
Chi-square (df)	-	549.226	51.768	102.784	1114.382	1335.322	-		
p-value	-	0.000	0.000	0.000	0.000	0.000	-		

Notes: Number of women=1326; Number of Events (Marriages)=1101 (83%); Censored cased=225 (17%)
*indicates the corresponding effect is statistically significant at 10%, ** indicates significance at 5%; *** indicates significance at 1%.

Table 3: Estimated Effects of Covariates on Log Duration Between Marriage and First Birth and on the Log Hazard of First Birth

COVARIATES	PARAMETRIC MODELS (Effects on Log Duration)							Cox PH Model (Effects on Log hazard)	
	Extended Generalized Gamma	Reciprocal Weibull	Log Normal	Weibull	Gamma	Exponential	Log logistic	Coefficient	Hazard Ratio
Constant	3.335***	2.942***	3.261***	3.899***	3.279***	3.684***	3.205***	.	.
Scale parameter	0.667***	0.785	0.672	0.699	1.000	1.000	0.349	.	.
Shape parameter	0.136*	-1.000	0.000	1.000	0.04 (0.07)	1.000	-	.	.
Ethncity									
Han (reference)	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	1.000
Yi	-0.030	0.106	-0.015	-0.137	-0.014	-0.089	-0.014	0.100	1.105
Other	0.125	0.310**	0.144	-0.013	0.154	0.066	0.162	-0.118	0.889
Religion									
No Religion (reference)	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	1.000
Budha	0.045	-0.113	0.041	0.024	0.042	0.036	0.097	-0.032	0.969
Don't know	-0.079	-0.241*	-0.093	-0.009	-0.081	-0.014	-0.073	0.065	1.067
Education									
None (reference)	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	1.000
Compulsory	-0.035	-0.077	-0.046	0.057	-0.041	0.027	-0.062	0.021	1.021
Above complusory	0.082	0.234	0.091	0.060	0.102	0.087	0.051	-0.083	0.920
Occupation									
Farm (reference)	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	1.000
Non-farm	-0.003	0.129	0.003	-0.021	-0.001	-0.020	-0.024	0.001	1.001
Age at Marriage									
Before 20 years (reference)	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	1.000
20-21	-0.258***	-0.381***	-0.270***	-0.280***	-0.261**	0.257**	-0.240***	0.298	1.347***
22-23	-0.321***	-0.252***	-0.320***	-0.340***	-0.321***	-0.340***	-0.333***	0.456	1.578***
24 years and above	-0.432***	-0.537***	-0.440***	-0.420***	-0.435***	-0.420***	-0.420***	0.528	1.695***
Marriage cohort									
Before 1970 (reference)	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	1.000
1970-79	-0.135*	-0.097	-0.128	-0.179***	-0.130	-0.166	-0.132*	0.168	1.183
1980-84	-0.315***	-0.294***	-0.300***	-0.450**	-0.301**	-0.410***	-0.260***	0.422	1.525***
1985-00	-0.376***	-0.270***	-0.350***	-0.530***	-0.344***	-0.480***	-0.320***	0.573	1.774***
Age difference with spouse									
Wife older (reference)	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	1.000
Same age	-0.101	-0.107	-0.082	-0.370***	-0.088	-0.250**	-0.041	0.228	1.256*
Husband 2-4 yrs older	-0.167**	-0.184**	-0.154*	-0.380***	-0.159	-0.280**	-0.098	0.254	1.290**
Husband 5+years older	-0.307***	-0.253***	-0.290***	-0.570***	-0.295**	-0.460***	-0.240***	0.456	1.578***
Introduction									
Self (reference)	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	1.000
Parents/ friends	0.080	0.028	0.770	0.072	0.076	0.078	0.073	-0.104	0.901
Match maker	0.080	0.039	0.081	0.081	0.080	0.077	0.078*	-0.099	0.906
Log Likelihood	-1020	-1198	-1023	-1124	-1145	-1232	-971		
Chi-square (df)	-	355.97	6.13	207.61	249.83	423.29	-		
p-value	-	0.000	0.013	0.000	0.000	0.000	-		

Notes: Number of women=1080; Number of First Births=1050 (97%); Censored cased=30 (3%)

Notes: *indicates the corresponding effect is statistically significant at 10%, ** indicates significance at 5%; *** indicates significance at 1%.

Table 4: Estimated Effects of Covariates on Log Duration Between First and Second Birth and on the Log Hazard of Second Birth

COVARIATES	PARAMETRIC MODELS (Effects on Log Duration)							Cox PH Model (Effects on Log hazard)	
	Extended Generalized Gamma	Reciprocal Weibull	Log Normal	Weibull	Gamma	Exponential	Log logistic	Coefficient	Hazard Ratio
Constant	3.375***	3.248***	3.436***	3.675***	3.398***	3.555***	3.433****	.	.
Scale parameter	0.555 (.01)	0.565(.01)	0.560(.01)	0.699(.01)	1,000	1,000	0.300(.00)	.	.
Shape parameter	-0.317 (.07)	-1,000	0,000	1,000	-0.310(.08)	1,000	-	.	.
Ethncity									
Han (reference)	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	1,000
Yi	0,051	0,064	0,044	0,021	0,055	0,031	0,049	-0,044	0,957
Other	0,136	0,130	0,154	0.312**	0,164	0,263	0,095	-0,284	0,753
Religion									
No Religion (reference)	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	1,000
Budha	0,071	0.150**	0,046	-0,003	0,059	0,019	0,020	-0,035	0,966
Don't know	0,103	0.1933**	0,065	-0,024	0,046	-0,063	0,068	-0,054	0,947
Education									
None (reference)	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	1,000
Compulsory	0,072	0,442	0.092*	0.302***	0,085	0.199**	0,060	-0,179	0.836*
Above complusory	0,010	-0,103	0,078	0.532***	0,034	0,352	0,001	-0,187	0,829
Occupation									
Farm (reference)	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	1,000
Non-farm	-0,075	-0,032	-0,098	-0.246**	-0,078	-0,165	-0,032	0,157	1,170
Age at Marriage									
Before 20 years (reference)	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	1,000
20-21	-0,002	-0,006	0,006	-0,010	0,008	0,022	0,005	-0,064	0,938
22-23	0.105*	0.104*	0.119**	0.192***	0,126	0.198*	0.094*	-0,237	0.789**
24 years and above	0.141**	0.112*	0.162**	0.243***	0.196*	0.286**	0.153***	-0,327	0.721***
First child birth year									
Before 1970 (reference)	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	1,000
1970-79	-0.117*	-0.172**	-0,960	-0,061	-0,110	-0,060	-0.117*	0,178	1,195
1980-84	-0.250***	-0.233***	-0.262***	-0.375***	-0.267*	-0.335**	-0.248***	0,480	1.616***
1985-00	0.251***	0.234***	0.246***	0,020	0.338***	0.294**	0.303**	-0,457	0.633***
Sex of First child									
Boy (reference)	0,000	-0,008	0,000	0,000	0,000	0,000	0,000	0,000	1,000
Girl	0,011	-0,107	0,016	0,024	0,019	0,032	0,014	-0,065	0,937
Survival Status of first child									
Alive (reference)	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	1,000
Dead	-0,141	-0.204*	-0,114	-0,146	-0,170	-0,169	-0,043	0,104	1,110
Log Likelihood	-768	-808	-779	-944	-959	-1090	745		
Chi-square (df)	-	79,91	21,46	352,06	383	644	-		
p-value	-	0,000	0,013	0,000	0,000	0,000	-		

Notes: Number of women=1050; Number of Second Births=792 (75%); Censored cases=262 (25%)

Notes: *indicates the corresponding effect is statistically significant at 10%, ** indicates significance at 5%; *** indicates significance at

Table 5: Estimated Effects of Covariates on Log Duration Between Second and Third Birth and on the Log Hazard of Third Birth

COVARIATES	PARAMETRIC MODELS (Effects on Log Duration)							Cox PH Model (Effects on Log hazard)	
	Extended Generalized Gamma	Reciprocal Weibull	Log Normal	Weibull	Gamma	Exponential	Log logistic	Coefficient	Hazard Ratio
Constant	2.780***	2.816***	3.124***	3.493***	2.763***	3.533***	3.212***	.	.
Scale parameter	1.02(.05)	1.03(.05)	1.11(.05)	1.08(.05)	1,000	1,000	0.59(.03)	.	.
Shape parameter	-1.10(.09)	-1,000	0,000	1,000	-1.14(.09)	1,000	-	.	.
Ethncity									
Han (reference)	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	1,000
Yi	-0,290	-0,302	-0,359	-0,283	-0,285	-0,268	-0.443*	0,472	1.603*
Other	0,249	0,211	-0,154	-0,538	0,267	-0,547	-0,147	0,185	1,203
Religion									
No Religion (reference)	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	1,000
Budha	0,271	0.264*	0,099	-0,097	0.273*	-0,102	0,062	0,008	1,008
Don't know	0,175	0,155	-0,042	-0,220	0,184	-0,227	-0,098	0,162	1,176
Education									
None (reference)	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	1,000
Compulsory	0.409***	0.402***	0.335**	0.342**	0.410***	0.339**	0.229*	-0,187	0,829
Above complusory	0.630*	0.612*	0,502	0,493	0.635*	0,497	0,297	-0,420	0,657
Occupation									
Farm (reference)	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	1,000
Non-farm	-0,221	-0,237	-0,491	-0.894***	-0,213	-0.903***	-0.482*	0,755	2.128**
Age at Marriage									
Before 20 years (reference)	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	1,000
20-21	0.196*	0.199*	0,214	0,206	0.194*	0,194	0,117	-0,125	0,882
22-23	0,007	0,010	0,045	0,002	0,007	-0,005	0,091	-0,050	0,951
24 years and above	0.406**	0.404**	0,362	0,181	0.405**	0,153	0,316	-0,208	0,812
Second child birth year									
Before 1970 (reference)	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	1,000
1970-79	0,100	0,105	0,216	0.369**	0,098	0.364**	0,152	-0,265	0.767*
1980-84	0.510***	0.568***	1.233***	1.723***	0.479***	1.658***	1.452***	-1,390	0.249***
1985-00	2.183***	2.235***	2.926***	4.382***	2.144***	4.067***	3.129***	-4,240	0.014***
Sex of 1st & 2nd child									
Two girls (reference)	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	1,000
One each	0.409***	0.405***	0.418***	0.523***	0.411***	0.510***	0.305**	-0,307	0.7356**
Two boys	0.451***	0.436***	0.394**	0.435**	0.456***	0.421**	0.306**	-0,327	0.721*
Survival Status of 2nd child									
Alive (reference)	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	1,000
Dead	-1.583***	-1.532***	-0.913**	-0,691	-1.602***	-0.691*	-0,566	0,585	1,795
Log Likelihood	-483	-483	-494	-531	-483	533	-484		
Chi-square (df)	-	0,218	22,77	97	0,162	100			
p-value	-	0,614	0,000	0,000	0,687	0,000			

Notes: Number of women=788; Number of Third Births=534 (68%); Censored cases=254 (32%)

Notes: *indicates the corresponding effect is statistically significant at 10%, ** indicates significance at 5%; *** indicates significance at 1%.

Table 6: Hypotheses and Corresponding Likelihood Ratio Statistics for Testing Special Case Model (Null Hypotheses) Against the More General (Alternative Hypothesis) Within the Parametric Family of Models for Time Marriage, First Birth, and Third Birth In

	Null Hypothesis	Model under Null Hypothesis	Model under Alternative Hypothesis	Likelihood Ratio Statistic (p values)
Time to Marriage	Shape = -1	Reciprocal Weibull	Extended Generalized Gamma	549.22 (0.000)
	Shape = 0	Log-normal	Extended Generalized Gamma	51.78 (0.000)
	Shape = 1	Weibull	Extended Generalized Gamma	102.78 (0.000)
	Scale=1	Ordinary Gamma	Extended Generalized Gamma	1114 (0.000)
	Scale = 1 and Shape =1	Exponential	Extended Generalized Gamma	1335.32 (0.000)
	Scale =1, given Shape =1	Exponential	Weibull	1232.53 (0.000)
First Birth interval	Shape = -1	Reciprocal Weibull	Extended Generalized Gamma	355.97 (0.000)
	Shape = 0	Log-normal	Extended Generalized Gamma	6.13 (0.013)
	Shape = 1	Weibull	Extended Generalized Gamma	207.61 (0.000)
	Scale=1	Ordinary Gamma	Extended Generalized Gamma	249.84 (0.000)
	Scale = 1 and Shape =1	Exponential	Extended Generalized Gamma	423.29 (0.000)
	Scale =1, given Shape =1	Exponential	Weibull	215.68 (0.000)
Second Birth interval	Shape = -1	Reciprocal Weibull	Extended Generalized Gamma	79.90 (0.000)
	Shape = 0	Log-normal	Extended Generalized Gamma	21.46 (0.000)
	Shape = 1	Weibull	Extended Generalized Gamma	352.06 (0.000)
	Scale=1	Ordinary Gamma	Extended Generalized Gamma	383.09 (0.000)
	Scale = 1 and Shape =1	Exponential	Extended Generalized Gamma	644.72 (0.000)
	Scale =1, given Shape =1	Exponential	Weibull	292.66 (0.000)
Third Birth interval	Shape = -1	Reciprocal Weibull	Extended Generalized Gamma	0.218 (0.641)
	Shape = 0	Log-normal	Extended Generalized Gamma	22.77(0.000)
	Shape = 1	Weibull	Extended Generalized Gamma	97.97 (0.000)
	Scale=1	Ordinary Gamma	Extended Generalized Gamma	0.162 (0.687)
	Scale = 1 and Shape =1	Exponential	Extended Generalized Gamma	100.88 (0.000)
	Scale =1, given Shape =1	Exponential	Weibull	2.91 (0.000)

Table 7: Estimated Effects of Covariates on Log Duration to and Log Hazard of Various Events

	Extended Generalized Gamma Models				Cox PH Model (exp b)			
	Marriage	First Birth	Second Birth	Third Birth	Marriage	First Birth	Second Birth	Third Birth
Ethncity								
Han	0,000	0,000	0,000	0,000	1,000	1,000	1,000	1,000
Yi	0,042	-0,030	0,051	-0,290	0,853	1,105	0,957	1.603*
Other	-0,047	0,125	0,136	0,249	1,044	0,889	0,753	1,203
Religion								
No Religion	0,000	0,000	0,000	0,000	1,000	1,000	1,000	1,000
Budha	-0.079*	0,045	0,071	0,271	1,214	0,969	0,966	1,008
Don't know	-0,083	-0,079	0,103	0,175	1.362*	1,067	0,947	1,176
Education								
None	0,000	0,000	0,000	0,000	1,000	1,000	1,000	1,000
Compulsory	0.117***	-0,035	0,072	0.409***	0.761***	1,021	0.836*	0,829
Above complusory	0.362***	0,082	0,010	0.630*	0.394***	0,920	0,829	0,657
Occupation								
Farm	0,000	0,000	0,000	0,000	1,000	1,000	1,000	1,000
Non-farm	0,018	-0,003	-0,075	-0,221	0,953	1,001	1,170	2.128**
Birth cohort								
1936-40	0,000	-	-	-	1,000	-	-	-
1941-45	-0.197**	-	-	-	2.637***	-	-	-
1946-50	-0,070	-	-	-	1.747***	-	-	-
1951-55	0.134*	-	-	-	0,796	-	-	-
1956-60	0.308***	-	-	-	0.579**	-	-	-
1961-65	0.283***	-	-	-	0.618**	-	-	-
1966-70	0.376***	-	-	-	0.487***	-	-	-
1971-75	0.344***	-	-	-	0.509***	-	-	-
1976-80	0.332***	-	-	-	0.545***	-	-	-
1981-85	0.526***	-	-	-	0.203**	-	-	-
Age at Marriage								
Before 20 years	-	0,000	0,000	0,000	-	1,000	1,000	1,000
20-21	-	-0.258***	-0,002	0.196*	-	1.347***	0,938	0,882
22-23	-	-0.321***	0.105*	0,007	-	1.578***	0.789**	0,951
24 years and above	-	-0.432***	0.141**	0.406**	-	1.695***	0.721***	0,812
Marriage/child cohort					-			
Before 1970	-	0,000	0,000	0,000	-	1,000	1,000	1,000
1970-79	-	-0.135*	-0.117*	0,100	-	1,183	1,195	0.767*
1980-84	-	-0.315***	-0.250***	0.510***	-	1.525***	1.616***	0.249***
1985-00	-	-0.376***	0.251***	2.183***	-	1.774***	0.633***	0.014***
Age difference with spouse								
Wife older	-	0,000	-	-	-	1,000	-	-
Same age	-	-0,101	-	-	-	1.256*	-	-
Husband 2-4 yrs older	-	-0.167**	-	-	-	1.290**	-	-
Husband 5+years older	-	-0.307***	-	-	-	1.578***	-	-
Introduction							-	-
Self	-	0,000	-	-	-	1,000	-	-
Parents/ friends	-	0,080	-	-	-	0,901	-	-
Match maker	-	0,080	-	-	-	0,901		
Sex of First child								
Boy	-	-	0,000	-	-	-	1,000	-
Girl	-	-	0,011	-	-	-	0,937	-
Sex of 1st & 2nd child	-	-			-			
Two girls	-	-		0,000	-	-	-	1,000
One each	-	-		0.409***	-	-	-	0.736**
Two boys	-	-		0.451***	-	-	-	0.721*
Survival Status of1st child						-		
Alive	-	-	0,000	0,000	-	-	1,000	1,000
Dead	-	-	-0,141	-1.583***	-	-	1,110	1,795

Figure 1: Density functions for Marriage, 1st birth, 2nd birth, and 3rd birth

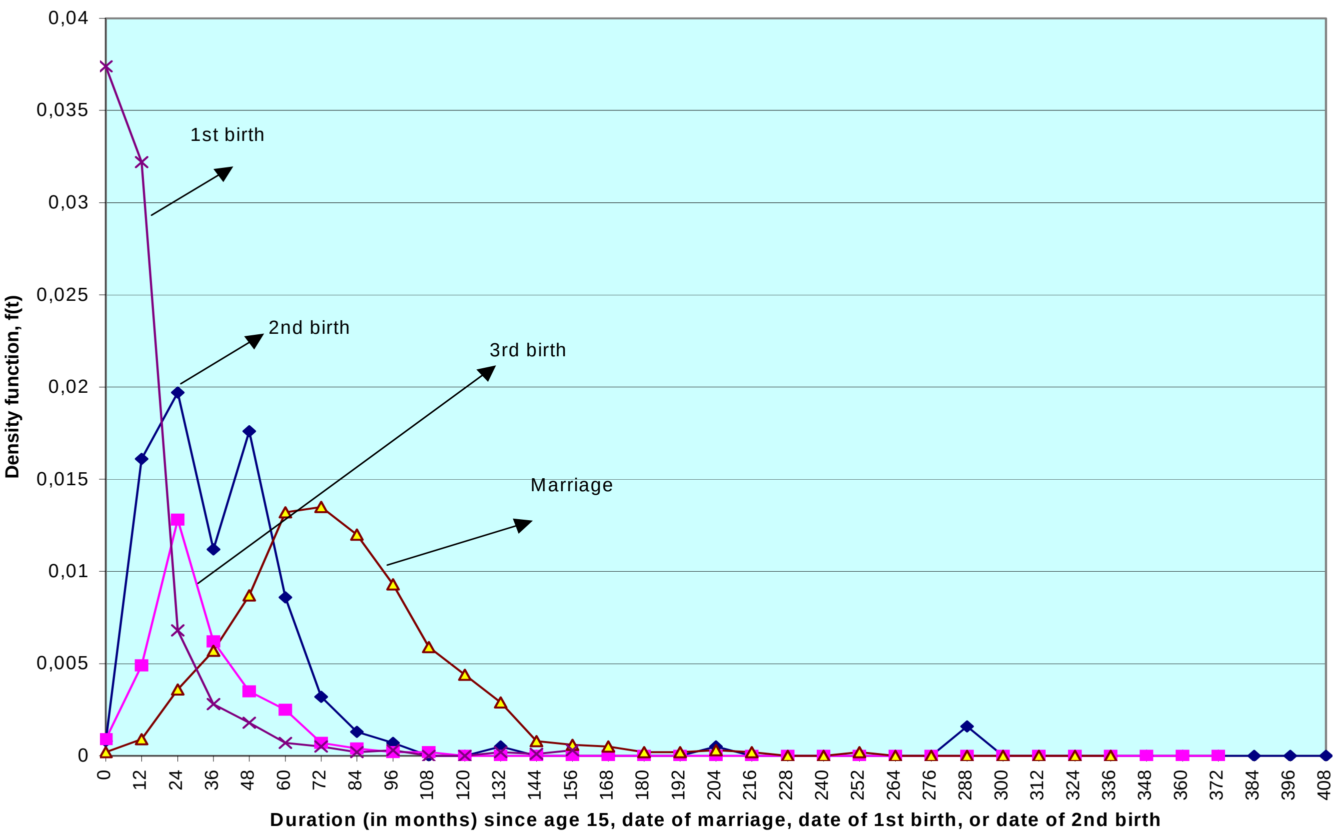

Figure 2: Hazard functions of marriage, first birth, second birth, and third birth

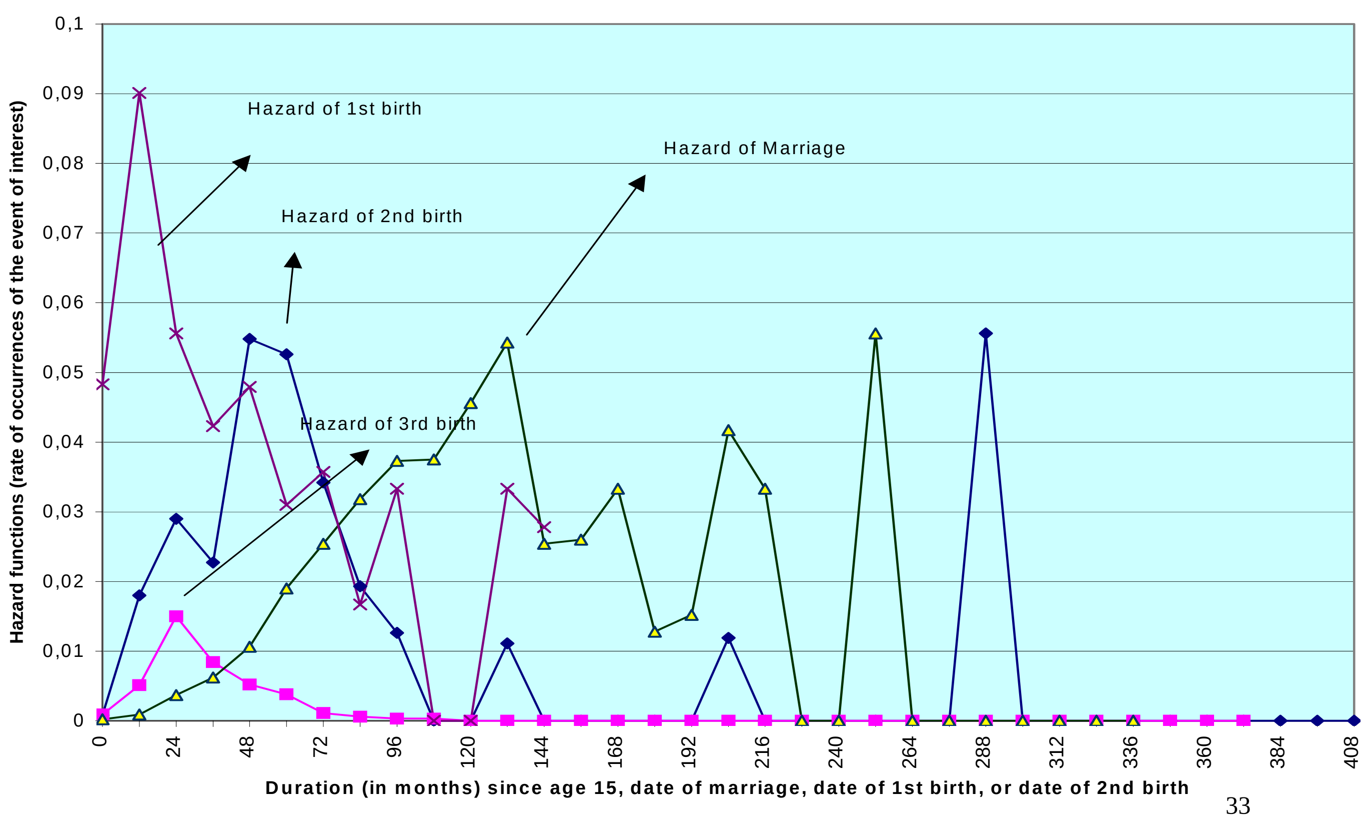

Figure 3: Survival functions for marriage, first birth, 2nd birth, and 3rd birth

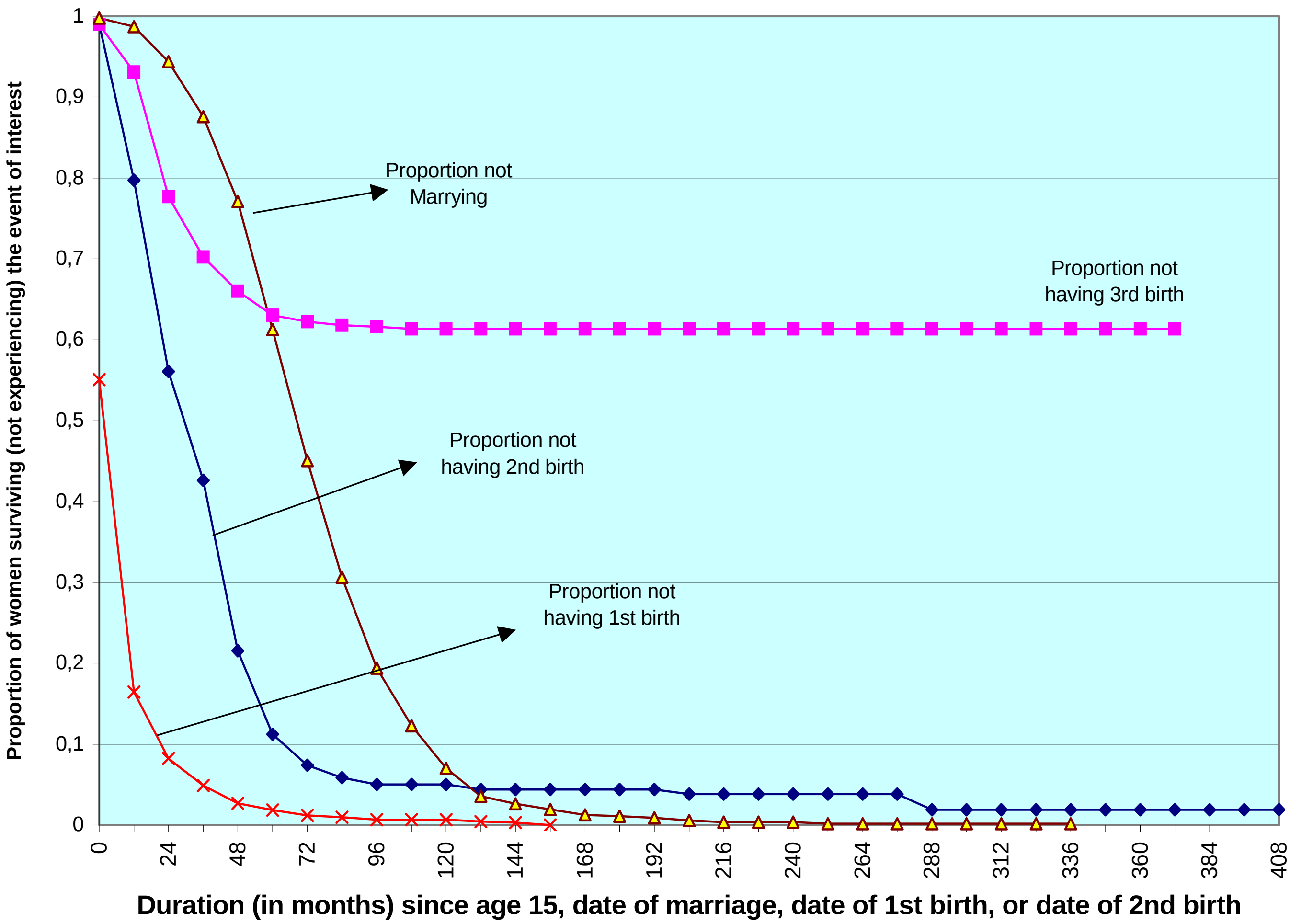

Figure 4: Hazard function by educational level (assuming proportional hazards for the three educational levels)

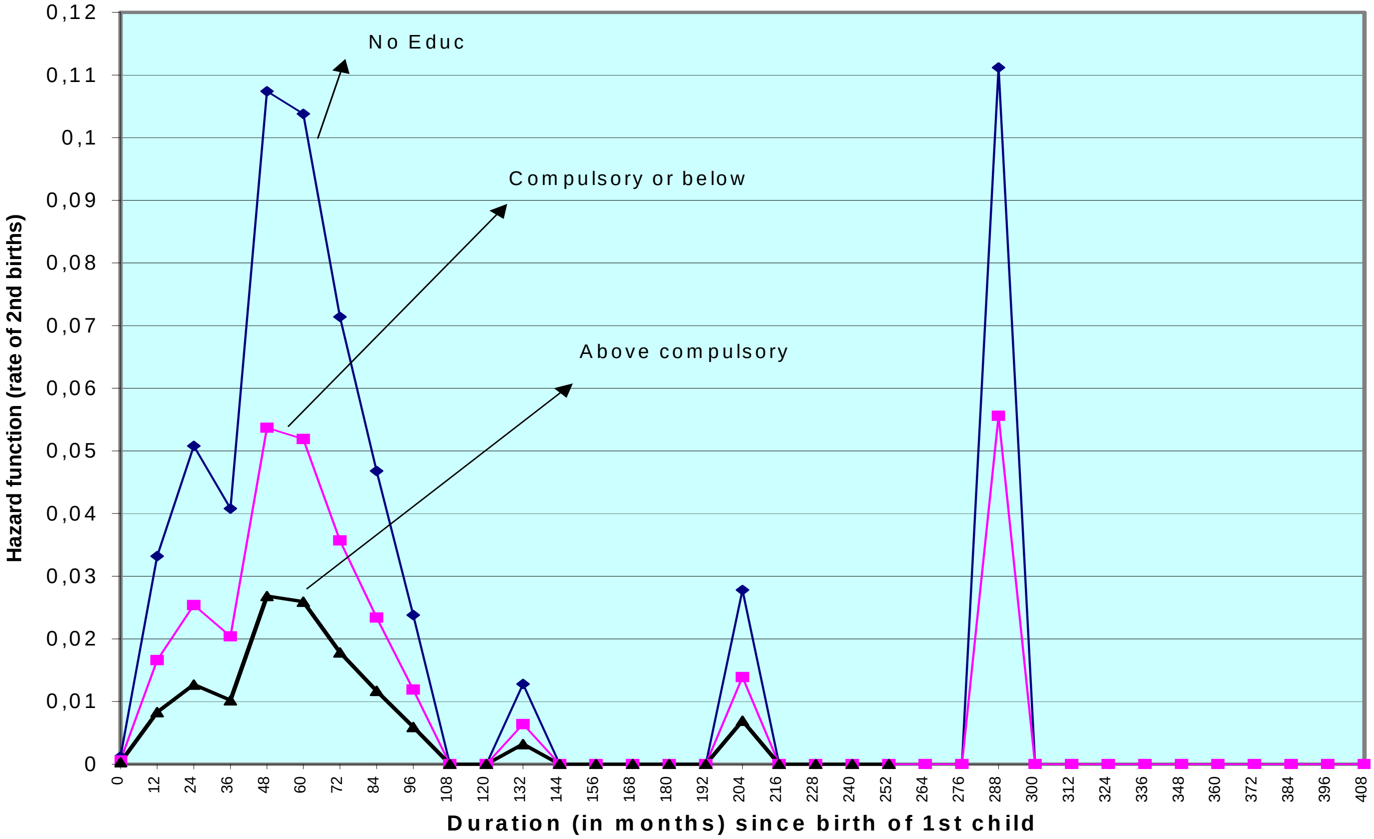

Figure 5: Hazard function by educational level (as obtained from the data)