



Demographic Research a free, expedited, online journal
of peer-reviewed research and commentary
in the population sciences published by the
Max Planck Institute for Demographic Research
Konrad-Zuse Str. 1, D-18057 Rostock · GERMANY
www.demographic-research.org

, citation and similar papers at core.ac.uk

brought to

provided by Research P

DEMOGRAPHIC RESEARCH

VOLUME 9, ARTICLE 10, PAGES 223-262

PUBLISHED 13 NOVEMBER 2003

www.demographic-research.org/Volumes/Vol9/10/

DOI: 10.4054/DemRes.2003.9.10

Research Article

**Elaboration of the Coale-McNeil
Nuptiality Model as the Generalized
Log Gamma Distribution:
A New Identity and Empirical
Enhancements**

Ryuichi Kaneko

© 2003 Max-Planck-Gesellschaft.

Table of Contents

1	Introduction	224
2	Coale-McNeil Model and the Generalized Log Gamma Distribution	225
2.1	Coale-McNeil Model	225
2.2	Coale-McNeil Model as The Generalized Log Gamma Distribution	228
3	The GLG Model as an Analytic Tool for First Marriage	231
3.1	Development of Country Specific Standard Schedules	231
3.2	Estimation of Covariate Effects on First Marriage Timing with/without Competing Risk Framework	236
4	Empirical Enhancement	239
4.1	Empirical Adjustment of the GLG Model	239
4.2	Method of Parameter Estimation	244
4.3	Censoring Effects on Parameter Estimation	245
5	Application of the Adjusted GLG Model	246
5.1	Estimation and Projection of First Marriage	246
5.2	Application for Fertility Projection	250
6	Summary and Conclusion	255
7	Acknowledgements	256
	Notes	257
	References	259
	Appendix	261

Research Article

Elaboration of the Coale-McNeil Nuptiality Model as the Generalized Log Gamma Distribution: A New Identity and Empirical Enhancements

Ryuichi Kaneko¹

Abstract

The Coale-McNeil nuptiality model is a particular case of the generalized log gamma distribution model. In this paper, we demonstrate that recognition of this connection allows an expansion of the possible applications of the Coale-McNeil model. As examples, we propose a procedure to develop country specific standard schedules, and illustrate the utility in regression analysis (directly and via the competing risk framework). In addition, we employ this identification to enhance the ability of the models with empirical adjustment to trace the trajectory of the lifetime schedule for cohorts which have not completed the process. We illustrate an application to Japanese female cohorts. We also propose an application to fertility projection by modeling the fertility schedule by birth order.

¹ National Institute of Population and Social Security Research, Hibiya-kokusai Bldg. 6F, 2-2-3, Uchisaiwai-cho, Chiyoda-ku, Tokyo 100-0011, Japan. E-mail: r-kaneko@ipss.go.jp

1. Introduction

The Coale-McNeil (CM) nuptiality model is a mathematical expression of regularity in age patterns of first marriage. It is a standard demographic tool for the estimation and projection of age schedules of first marriage and birth by birth order. However, it is not generally recognized by researchers that the CM model without a prevalence parameter is precisely the log version of the generalized gamma distribution with limited parameter space (Kaneko 1991a). Clear recognition of this connection is useful because it enables the utilization of the rich body of knowledge about the statistical properties of the generalized gamma distribution when pursuing demographic applications. Conversely, some of the analysis of the structure of the CM model (such as the interpretation of the convolution structure) can be applied to understanding the generalized log gamma distribution. The first purpose of this article is to demonstrate some of the demographic applications that benefit from this new description of the CM framework. We present algorithms for formalized development of country-specific standard schedules. In addition, we provide an analysis of the effects of covariates on marriage timing, both with and without application of the competing risk framework for different types of marriages.

The second purpose of this article is to present enhancements to the ability of the model to trace the trajectories of lifetime marriage and fertility schedules by incorporating an empirical model of the residual pattern. This provides more precise estimation even for cohort experiences of the processes that have not been completed. Period measures of nuptiality and fertility are subject to compositional and distributional "distortions" such as those from flux in marital and parity composition and tempo effects. Although some effective remedies have been proposed to correct for these distortions (Bongaarts and Feeney 1998, Kohler and Philipov 2001, Kohler and Ortega 2002, Ryder 1964, 1980), cohort nuptiality and fertility measures which are free from those effects are of primary importance in understanding what is taking place in people's life course in the demographic sense. The only drawback of cohort measures is that they cannot be evaluated until the life course processes of the events are completed, and therefore they do not provide information on the current situation of uncompleted phenomena. It is impossible to "measure" cohort experiences that are not completed (Ryder 1964, van Imhoff 2001). However, a model embodying lifetime regularities of the events (i.e. the "law" of nuptiality and fertility) may provide useful predictions of the current situation. Our empirical enhancement of the GLG model is a very practical effort in this line. We demonstrate its usage in estimation and prediction of first marriage schedules by providing long-term estimations of lifetime measures for marriage behavior relevant to recent marriage and fertility developments in Japan. We also discuss fertility projection as an application of the empirically enhanced model.

2. Coale-McNeil Model and the Generalized Log Gamma Distribution

2.1 Coale-McNeil Model

Following the finding by Coale (1971) that age specific rates of first marriages for female cohorts from different countries showed virtually identical patterns if location, scale, and eventual proportion of ever marrying are adjusted, Coale and McNeil derived a statistical distribution that described observed distributions of age at first marriage (Coale and McNeil 1972). The closed form description of the probability density function (PDF) for this distribution is:

$$g(x) = \frac{\beta}{\Gamma(\alpha/\beta)} \exp\left[-\alpha(x - \mu) - \exp\{-\beta(x - \mu)\}\right] \quad (1)$$

where Γ denotes the gamma function ($\Gamma(x) = \int_0^{\infty} t^{x-1} e^{-t} dt$), $\alpha(> 0)$, $\beta(> 0)$, and $\mu(-\infty < \mu < \infty)$ are three parameters (Coale and McNeil 1972). For practical application, they provided a standard marriage schedule model from a location-scale family of this distribution by fixing the shape according to the experiences of Swedish female cohorts. The following is an adjusted version of the standard model with mean zero, and variance unity by Rodriguez and Trussell (1980), obtained by setting $\alpha = 1.145$, $\beta = 1.896$, and $\mu = -0.805$ in equation (1):

$$g_s(z) = 1.2813 \exp\left[-1.145(z + .805) - \exp\{-1.896(z + .805)\}\right]. \quad (2)$$

Let $g(x)$ denote a distribution of age at first marriage of any female cohort with its observed mean, u , standard deviation, b . Using the standard model above, it is given by:

$$g(x; u, b) \cong \frac{1}{b} g_s\left(\frac{x - u}{b}\right) \quad (3)$$

The marriage schedule that embodies the probability density of marrying at exact age x for all members of the cohort (Note 1), $f(x)$, is represented by:

$$f(x) = C g(x) \quad (4)$$

where C denotes proportion eventually marrying in the cohort. Thus, the age distribution underlies the age schedule (the age specific rate) with parameter of prevalence measure, C .

Clear definition of terms is essential. In this paper, the probability density $f(x)$ given by (4) is called first marriage schedule, the underlying distribution $g(x)$ in the same equation is called distribution of age at first marriage, and the normalized distribution with fixed shape $g_s(x)$ given by (2) is the (global) standard distribution of age at first marriage. Thus, $C g_s(x)$ is the (global) standard schedule.

An interesting feature of the CM distribution is that it is a limiting probability distribution of convolution of infinite set of mean-related exponential distributions. Thus, it is regarded as a convolution of distribution of its own form and some numbers of related exponential distributions as well. This structure provides a mathematical model for the multistage process, by which we mean a process that consists of multiple processes required for the target event to happen. In fact, Coale and McNeil (1972), inspired by Feeney (1972), viewed first marriage as a multistage process in which entry into marriageable state, meeting of the eventual spouse, and engagement are required to take place prior to the marriage.

Suppose we form the convolution of the m exponential distribution with parameter α , $\alpha + \beta$, $\alpha + 2\beta$, ..., $\alpha + m\beta$, where α and β are two parameters with positive real values, and let $h_T(t; m)$ denote PDF of the resulting distribution, then the CM distribution given in equation (1) is the convolution of two distributions whose PDFs are:

$$g_x(x; m) = \frac{\beta}{\Gamma(\alpha/\beta + m)} \exp\left[-(\alpha + m\beta)(x - \mu) - \exp\{-\beta(x - \mu)\}\right] \quad (5)$$

$$h_T(t; m) = \frac{\beta\Gamma(\alpha/\beta + m)}{\Gamma(\alpha/\beta)(m-1)!} \{1 - \exp(-\beta t)\}^{m-1} \exp(-\alpha t) \quad (6)$$

where α , β , and μ are three parameters of the CM distribution found in (1) (Coale and McNeil, 1972). Here $g_x(x; m)$ represents the distribution of times of entering a

stage from when the process starts, and $h_T(t; m)$ is the distribution of the waiting time that is composed of m exponentially distributed waiting times. The mean and variance of distribution $g_X(x; m)$ are respectively $\mu - \frac{1}{\beta} \psi\left(\frac{\alpha}{\beta} + m\right)$, and $\frac{1}{\beta^2} \psi'\left(\frac{\alpha}{\beta} + m\right)$, where ψ and ψ' are the digamma and trigamma functions. Those

of distribution $h_T(x; m)$ are respectively $\sum_{j=1}^m \{\alpha + (m-1)\beta\}^{-1}$, and $\sum_{j=1}^m \{\alpha + (m-1)\beta\}^{-2}$.

The exponential distribution with the three largest means convoluted in distribution $h_T(t; m)$ have the parameters α , $\alpha + \beta$, and $\alpha + 2\beta$. For the first marriage process, Coale and McNeil supposed that these are distributions of duration from entry into the marriage market to the meeting of future husband, dating duration, and engagement duration. According to parameter values of the CM standard age distribution (2), which are derived from the experiences of Swedish female cohorts, the mean duration from entry into the marriage market to the meeting of future husband is estimated as $(1/0.174)$ or 5.75 years. Similarly, means of the second and third waiting durations are 2.16 years $(1/(0.174 + 0.2881))$ and 1.33 years $(1/(0.174 + 2 \times 0.2881))$ respectively (Coale and McNeil, 1972). However, empirical evidence on female first marriage process in Japan indicates that age at the meeting and durations between meeting and marriage are highly correlated (Kaneko 1991a) (Note 2). This implies that the assumption of independence of the convoluted distributions is violated in actual processes, and the estimated mean durations above should be biased. Nonetheless, the convolution structure of the CM distribution may provide a structured approximation of the complicated multistage model. Moreover, it should serve as an important prototype in developing models with process dependences which make parameters of sub-processes dependent on the timing of the outcomes of previous stages.

2.2 Coale-McNeil Model as The Generalized Log Gamma Distribution

Some authors have discussed relationships of the Coale-McNeil distribution to other well-known probability distributions. Coale and McNeil themselves pointed out that the CM distribution is the extreme value distribution (of type I, or equivalently the Gompertz distribution for non-negative random variables) when $\alpha = \beta$ (Coale and McNeil, 1972). Rodriguez and Trussell discussed its relationships with the gamma distribution (Rodriguez and Trussell, 1980). Liang (2000) discussed its relationships to the extreme value, log-gamma, and the normal distribution based on discussion of the log-gamma distribution by Johnson et al. (1994). The conclusive observation on these relationships is that the CM distribution is precisely equivalent to the *generalized* log gamma (GLG) distribution with a somewhat different parameter space (Kaneko 1991a). This explains the relationships of the CM distribution to the extreme, the log gamma, and the normal distributions, since these arise as special cases of the generalized log gamma distribution.

The generalized gamma (GG) distribution was defined by Stacy (1962), introducing an additional parameter into the gamma distribution (Note 3). If a random variable follows the GG distribution, then the log-transform of the random variable follows the GLG distribution (some authors such as Johnson et al. 1994, call it the log generalized gamma distribution), which is a mirror image of the CM distribution reflected at the origin ($x = 0$). Prentice (1974) proposed an alternative parameterization of the GLG distribution which extends the parameter space so as to express both mirror images of the distribution corresponding to random variable X and $-X$ by one model. Hence, it includes the CM distribution as a constrained version with half of the extended parameter space. Here we refer to Prentice's extended version simply as the GLG distribution.

The PDF and the cumulative distribution function (CDF) of the GLG distribution are given by:

$$g(x) = \frac{|\lambda|}{b\Gamma(\lambda^{-2})} (\lambda^{-2})^{\lambda^{-2}} \exp \left[\lambda^{-1} \left(\frac{x-u}{b} \right) - \lambda^{-2} \exp \left\{ \lambda \left(\frac{x-u}{b} \right) \right\} \right] \quad (7)$$

$$G(x) = 1 - I \left(\lambda^{-2}, \lambda^{-2} \exp \left(\lambda \frac{x-u}{b} \right) \right) \quad (8)$$

where λ ($-\infty < \lambda < \infty, \neq 0$), u ($-\infty < u < \infty$), b (> 0) are three parameters, and Γ and I denote the gamma function and the incomplete gamma function respectively (Note 4). Parameter λ determines the shape of the distribution; if it is positive, the distribution is skewed to the left, and if it is negative, it is skewed to the right. The distribution is not defined for $\lambda = 0$, but as $\lambda \rightarrow 0$, the distribution approaches the normal distribution. u is a location parameter which determines the location of the mode of the distribution. b is the scale parameter of the distribution.

The following alternative parameterization of the CM distribution allows representation of the full range of the parameter space as the GLG distribution:

$$g(x) = \frac{|\beta|}{\Gamma(k)} \exp\left[-k\beta(x-\mu) - \exp\{-\beta(x-\mu)\}\right] \quad (9)$$

where k (> 0) is a new parameter, and β is now allowed to take a negative value (Note 5).

Since the original CM distribution corresponds to the GLG distribution with a negative value of λ , and we consider only situations in which λ takes on a negative value, we can regard the GLG distribution as equivalent to the CM distribution throughout the paper.

One of the advantages of the GLG formulation is that it has only one shape parameter, i.e. λ . Describing the shape of the distribution by a single value is quite useful in applications. Since it identifies a distribution once location and scale are controlled for, it can be regarded as an index of schedule shape specific to cohorts of a country or region, for instance. In fact, the essential nature of the CM global standard nuptiality schedule is constancy across population groups of the shape parameter at a value of $\lambda -1.287$. Substantial implication of the shape value is discussed later. In the next section, we describe a simple procedure to develop country specific standard schedules which takes advantage of the single shape parameter. The single shape parameter is quite effective in developing fertility schedules by birth order as well, since the shape value varies by birth order.

The mean and variance of the GLG distribution are respectively given as:

$$u + (b/\lambda)\{\psi(\lambda^2) + \ln \lambda^2\} \quad (10)$$

$$(b/\lambda)^2 \psi'(\lambda^{-2}) \quad (11)$$

The mode is simply u , with the maximum of PDF given by:

$$\hat{g} = \frac{|\lambda|}{b\Gamma(\lambda^{-2})} (\lambda^{-2})^{\lambda^{-2}} e^{-\lambda^{-2}}.$$

As mentioned earlier, the GLG distribution includes many of the fundamental distributions as special cases. Specifically, it specializes to the extreme value distribution when $\lambda = 1$, the (standard) log-gamma distributions when $\lambda = b$, $u = -2 \ln \lambda$, and even the normal distribution as a limiting case when $\lambda \rightarrow 0$. These relations are a reflection of relationships between the GG distribution and the exponential, the Weibull, the gamma, and the log normal distributions, and it guarantees that the GLG distribution describes age distribution of first marriages better than those specialized distributions.

The correspondence between the parameters of the CM and GLG distributions can be expressed as:

$$\alpha = -\frac{1}{b\lambda}, \quad \beta = -\frac{\lambda}{b}, \quad \mu = u - \frac{b}{\lambda} \ln \lambda^2 \quad (12)$$

or equivalently,

$$\lambda = -\left(\frac{\alpha}{\beta}\right)^{\frac{1}{2}}, \quad b = (\alpha\beta)^{\frac{1}{2}}, \quad u = \mu - \frac{1}{\beta} \ln\left(\frac{\alpha}{\beta}\right), \quad (13)$$

where α , β , and μ are parameters of the CM formulation in (1) (Kaneko 1991a). The revised version of the CM standard distribution of age at first marriage by Rodriguez and Trussell given by (2) is expressed by the GLG with parameters $\lambda = -1.287$, $u = -0.5390$, $b = 0.6787$.

The identification of the CM distribution as the GLG distribution allows many important properties previously explored separately for each of the distributions to be unified. For instance, since the GLG distribution includes as special cases some important distributions as noted above, so does the CM distribution. Conversely, the characterization of the CM distribution as a convolution of an infinite number of mean related exponential distributions applies to the GLG distribution as well. In other words, the convolution representation of the CM distribution as expressed in formulations (5) and (6) holds for the GLG distribution as:

$$g_x(x; m) = \frac{|\lambda|}{b\Gamma(\lambda^{-2} + m)} (\lambda^{-2})^{\lambda^{-2} + m} \exp \left[(\lambda^{-2} + m) \lambda \left(\frac{x-u}{b} \right) - \lambda^{-2} \exp \left\{ \lambda \left(\frac{x-u}{b} \right) \right\} \right] \quad (14)$$

$$h_T(t; m) = \frac{|\lambda| \Gamma(\lambda^{-2} + m)}{b\Gamma(\lambda^{-2})(m-1)!} \left\{ 1 - \exp \left(\frac{\lambda t}{b} \right) \right\}^{m-1} \exp \left(\frac{t}{b\lambda} \right) \quad (15)$$

where the parameters are the same as given above and m is the number of the mean-related exponential distributions that compose the waiting time distribution $h_T(x; m)$.

Theories, application frameworks, and computer software packages developed for either distribution can be applied to the other. This is particularly useful in view of the wide array of utilities for working with the GG and GLG distributions. In the following section, we first demonstrate the usefulness of the GLG formulation as an analytic tool for studying first marriage behavior, and then we conduct empirical enhancement of the applicability of the model to predict trajectories of marriage and fertility schedules.

3. The GLG Model as an Analytic Tool for First Marriage

3.1 Development of Country Specific Standard Schedules

The first demonstration of the usefulness of the new identification of the CM model exploits the feature that it has only one shape parameter (λ). This enables us to make country specific standard schedules through undemanding procedure. Country specific standards are sometimes required and often desirable, since the global standard schedule derived from Swedish experiences might be inappropriate for some populations. In addition, identifying the specific shape value to apply for a schedule is crucial for predicting the fertility by birth order, since the value for each order varies from that of the nuptiality standard. In the following, we illustrate the development of a country specific schedule using Japanese female cohorts.

Some authors have reported that the CM standard marriage schedule does not fit Japanese experiences as well as those of Western countries (Takahashi 1978, Kojima 1985, Kaneko 1991b). Kaneko (1991b) examined shape parameter values of the GLG model fitted to cohort and period marriage schedules of Japanese females, and found substantial deviations from the shape value λ of the CM standard schedule, i.e. -1.287.

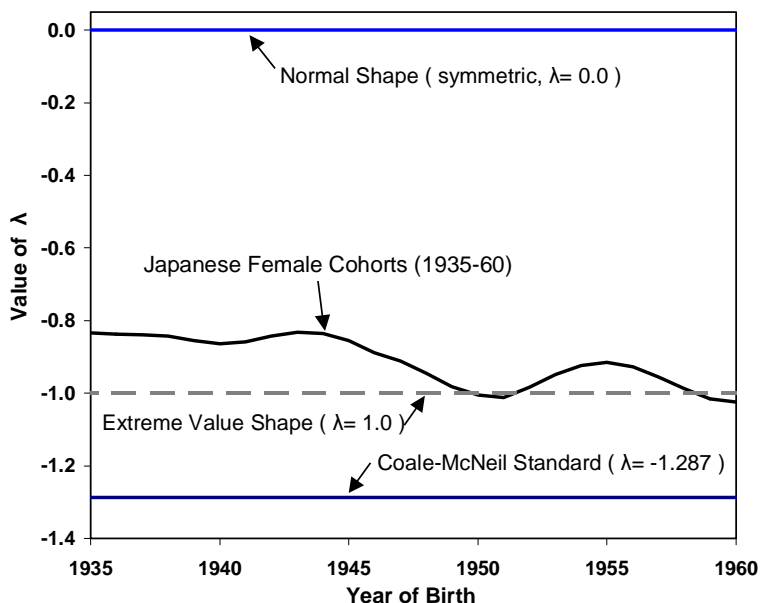


Figure 1: *Estimated Shape Parameter Values (λ) of The GLG Model For Japanese Female Cohorts born in 1935-60*

Figure 1 shows the trend of parameter λ estimated for Japanese female cohorts born in 1935-1960 who have attained at least age 40 at the time of evaluation. The figure indicates deviations in the shape values of Japanese cohorts from the value of the CM standard schedule, which is derived from Swedish experiences. The shape values of Japanese cohorts fall in the range from -1.0 to -0.8 , while it is at -1.287 for the CM standard. Other particular values correspond to well-known underlying distributions. The value zero corresponds to the normal distribution, and the value unity to the extreme value distribution. The shape values of Japanese cohorts are located in the middle of the CM standard and the normal distribution near the extreme value model, implying that the Japanese schedule is more symmetric than the CM standard. It seems a little more feasible to use the extreme value distribution to describe Japanese cohorts. We consider some of reasons for the symmetry seen in Japanese cohorts later. A mild decline in the shape value over cohorts is also identified in the figure.

Coale’s original finding about first marriage schedules is, in our translation, that the shape parameter of the age distribution is common over countries and periods (Coale 1971). It has been validated by the wide use of the shape-fixed CM standard schedule as a practical tool in various demographic applications. However, the stability of shape is, of course, an approximation. Our close examination indicates that it varies across countries (Japan is not the same as Sweden), and undergoes changes over time to some extent as well. Thus, a schedule with a shape value specific to Japanese women is expected to be more accurate. On the grounds that the shape characterizes the first marriage process of a nation, developing country specific standard schedule with an appropriate shape value is quite beneficial. Such a specific schedule offers more accuracy with the same general form. We now describe the procedure for producing a country specific shape-fixed standard schedule with mean zero and standard deviation of unity.

Let λ_s denote shape parameter of the GLG model which is specific to cohorts of a country. The following parameters are to be calculated for the new standard:

$$k_s = \lambda_s^{-2},$$

$$\alpha_s = k_s \sqrt{\psi'(k_s)}, \quad \beta_s = \sqrt{\psi'(k_s)}, \quad \mu_s = \frac{\psi(k_s)}{\sqrt{\psi'(k_s)}}, \quad (16)$$

where ψ and ψ' denote the digamma and trigamma function. Using these parameters, the underlying distribution of age at first marriage in the new standard schedule is given by:

$$g_s(z) = \frac{\beta_s}{\Gamma(\alpha_s/\beta_s)} \exp\left[-\alpha_s(z - \mu_s) - \exp\{-\beta_s(z - \mu_s)\}\right] \quad (17)$$

The country specific shape value λ_s can be obtained by averaging values of λ of the GLG model fitted to source schedules. Then any first marriage or fertility distribution $g(x)$ and corresponding schedule $f(x)$ among the location-scale family of the standard are given by:

$$g(x) = \frac{1}{s_x} g_s\left(\frac{x - \bar{x}}{s_x}\right), \quad (18)$$

$$f(x) = C g(x), \quad (19)$$

where \bar{x} and s_x are the mean and standard deviation of the distribution, and C is the proportion eventually marrying, of a target cohort.

For Japanese female cohorts born between 1935-60, the average of the estimated $\hat{\lambda}$ of the GLG model is -0.9123. With this value for $\hat{\lambda}_s$, and following the formulas (16) and (17), our new Japanese female standard marriage schedule $g_j(z)$ is given by:

$$g_j(z) = 1.226 \exp\left[-1.351(z + 0.2553) - \exp\{-1.125(z + 0.2553)\}\right]. \quad (20)$$

In Figure 2 the Japanese standard schedule developed here is compared with the CM standard. Substantial difference in shape is observed. The Japanese standard is more symmetrical than the CM standard. In fact, the shape value indicates that the Japanese standard is only a little more symmetrical than the extreme distribution. Besides Japanese cohorts, Liang (2000) reported that the first marriage distribution of a Chinese female cohort was close to the normal distribution rather than the CM standard.

Why are the distributions from these countries more symmetrical than the CM standard? What determines the shape of the distribution in the first place? Kaneko (1991a), using national representative survey results, examined the shape of first marriage schedule by type of marriage in Japan, and concluded that a rather symmetric shape in the Japanese schedule was formed by presence of arranged marriages. Namely, when the competing risk models were applied to age distributions of marriage by the types of marriage (arranged marriage and others), the shape of each distribution by disaggregated by type was less symmetric than the aggregate, and each shape value was closer to that of the CM standard (arranged -1.065, non-arranged -0.965, while the aggregate -0.644). This implies that the major part of the deviation of the shape in Japan from that of the CM standard is caused by mixture of types of marriage whose timing is distinctively different. In addition, the shape values of schedules of non-arranged and all marriages become even closer to the CM global standard if socioeconomic covariates are controlled for. Therefore, the skewed shape exhibited by the CM global standard may represent schedules of homogeneous marriage behaviors. Comparison of the shape values associated to different marriage types with covariates controlled through the competing risk framework is given in the following section.

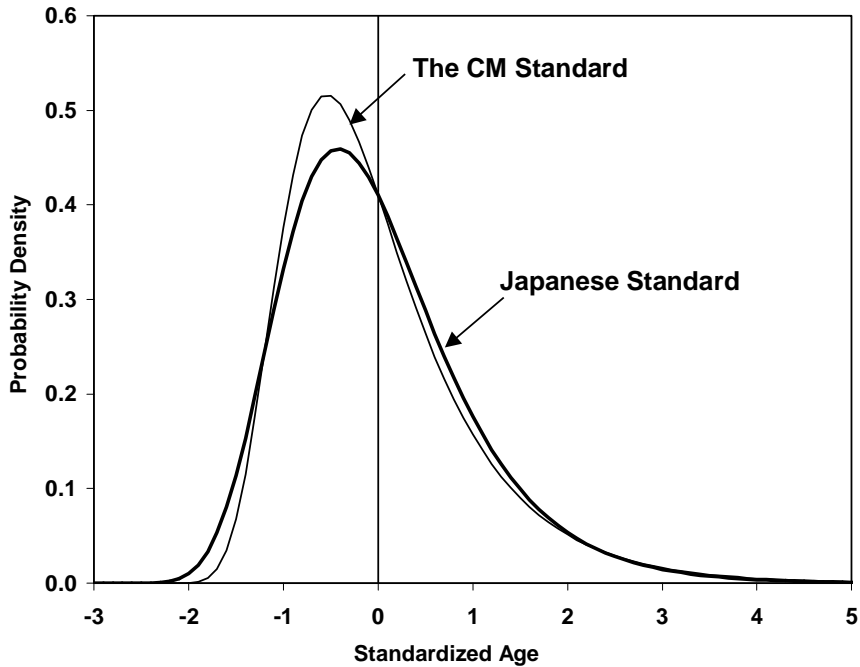


Figure 2: Comparison of Japanese Standard Schedule with The Coale-McNeil Standard for First Marriage of Female Cohort

Searching for the determinants of the shape is essential for many applications of the model. As shown later, it is crucial to identify the shape in predicting the schedules of young cohorts that have yet to complete their process. As seen in Figure 1, the shape value changes over time. For example, if presence of arranged marriage makes the shape symmetric and proportion of arranged marriage decreases over time, the shape value is expected to approach that of the CM standard.

3.2 Estimation of Covariate Effects on First Marriage Timing with/without Competing Risk Framework

Though extensions to incorporate covariates into the CM model have been conducted by several authors (Trussell and Bloom 1983, Sørensen and Sørensen 1986, Liang, 2000), the GLG specification has some advantages for this purpose in both theoretical and practical developments since it is one of the standard parametric regression models in survival analysis (Lawless 1982, Johnson et al. 1994 1995, Klein and Moeschberger 1997). Here, we demonstrate the effectiveness of the model in analysis of covariate effects on age at first marriage, and the effect of heterogeneity on shape parameter value, which is significant in predicting the parameter required for nuptiality and fertility projection described later in this paper.

In the standard specification of the GLG regression model, a vector of covariates for individual i , \mathbf{X}_i , are incorporated into the model in linear form with regression parameters $\boldsymbol{\theta}$, so that the parameter u of equation (7) and (8) for individual i should be $u_i = \mathbf{X}_i' \boldsymbol{\theta}$ where \mathbf{X}_i' denotes the transpose of vector \mathbf{X}_i . Since the parameter u determines the location of the schedules (the mode), the specification implies that individuals have underlying probabilities that differ in marriage or reproduction timing depending on their characteristics. This formulation is particularly useful for analysis of the current fertility decline to below replacement level in many countries, since the decline is firmly connected with delay in timing of marriage and childbearing.

We conduct the GLG regression for age at first marriage with some demographic and socio-economic characteristics using survey data for illustrative purposes (Note 6) (The source is the national representative sample in the Ninth National Fertility Survey conducted in 1987 in Japan). The results are presented in Table 1 in the far left two columns (“All Marriage”), where estimated parameter values and regression coefficients for two different model specifications (model 1 and 2) are shown.

The coefficients listed indicate the amount of marriage delay (year) in relation to timing in reference category (marked with #) if variables are categorical, or to unit change of covariates if they are quantitative (here only “Number of Sibling”). Model 1 incorporates only “cohort” as covariates, and shows no significant difference in marriage timing among them, though slight delay is observed in younger cohorts. However, Model 2, in which all other covariates at hand are incorporated, reveals that the delay in younger cohorts is fully attributable to the effects of other covariates than “cohort”, mainly due to the expansion of higher educational groups, since coefficient values of cohorts are reversed when those effects are controlled in this model (Note 7). It is worth noting that the value of λ tends to decrease when more covariates are introduced, which supports the view that the symmetric shape of schedule is caused by heterogeneity in marriage timing.

Table 1: *Effects of Covariates on Age at First Marriage of Japanese Women: The GLG Regression by Type of Marriage with Competing Risk Model*

Covariates	All Marriage		Non-Arranged	Arranged
	Model 1 (N=4682)	Model 2 (N=4682)	(N=4682) (n=2878)	(N=4682) (n=1804)
Intercept	23.34	22.43	23.86	23.33
Cohort (Birth Year)			****	****
# 1938-39	0.00	0.00	0.00	0.00
1940-44	0.04	-0.05	-0.36	0.30
1945-49	0.17	-0.11	-0.72 ***	0.75 ***
1950-54	0.20	-0.18	-1.02 ****	1.29 ****
Educational Background		****	****	****
# Junior College		0.00	0.00	0.00
High School		0.87 ****	0.82 ****	0.89 ****
Junior College		1.49 ****	1.69 ****	1.00 ****
University		2.48 ****	2.61 ****	2.05 ****
Father's Occupation		**	**	***
# Agriculture		0.00	0.00	0.00
Self-employed		0.13	-0.10	0.50 **
White-collar		0.17	-0.16	0.70 ****
Blue-collar		-0.13	-0.49 **	0.48 *
Not working/ temporary		-0.44 *	-0.69 **	0.02
Area of Residence		****		****
# Rural		0.00	0.00	0.00
Urban		0.42 ****	0.13	0.98 ****
Co-residence with			***	****
# Living seperately		0.00	0.00	0.00
Living together		0.09		
Heiress				
# Not heiress		0.00	0.00	0.00
heiress		-0.16	-0.06	-0.22
Number of Sibling		-0.07 **	-0.12	0.04
Scale Parameter (<i>b</i>)	2.614	2.460	3.082	3.453
Shape Parameter (<i>λ</i>)	-0.673	-0.761	-1.161	-1.054

N : Sample Size n : Number of Samples without Censor # : Reference Category

* P<0.05 ** P<0.01 ***P<0.001 ****P<0.0001

Data Source: The Ninth National Fertility Survey in Japan, Female cohorts born in 1938-54.

While the effects of heterogeneity of individual characteristics in relation to first marriage timing are measured above, we next view the effects of heterogeneity of characteristics of marriage itself. There occur several different types of marriage such as non-arranged and arranged marriages, registered marriage and cohabitation, or inter-racial and intra-racial marriages, and so forth. For instance, consider a situation in which the marriage processes of non-arranged and arranged marriages are to be compared. One plausible supposition here is that the same person goes through different processes simultaneously and ends up in either of these different types of marriage, whichever comes first. According to the survival analysis framework, this type of situation can be dealt with by the competing risk model, in which several different events have their own mutually independent probabilities of taking place at a given time.

We illustrate the use of the competing risk framework by applying it to analysis on determinants of first marriage timing in Japan taking into account the type of marriage, i.e. non-arranged and arranged marriage. The results are presented in the right two columns of Table 1. Here, some interesting tendencies hidden in the analysis of all over marriage appear. First, age at first marriage decreased by cohort for non-arranged marriages, while it increased for arranged marriages. These changes in opposite directions are both statistically significant and substantial in amount. On the other hand, as described before, the trend as a whole for all marriages indicates no significant change by cohort. The analysis by type of marriage here revealed active changes behind the seeming stability over the cohorts. Similar opposite effects by type of marriage are seen for some other covariates. Co-residence with parent(s) before marriage significantly affects marriage timing of each type in opposite directions (delay in the non-arranged, and accelerated in the arranged marriage) while that of over all marriage appears to be unaffected. Residence in urban areas delays only arranged marriage. Only non-arranged marriage is accelerated by the presence of siblings. The analysis illustrates that examination by type of marriage with the competing risk framework provides us with information about detailed features of the process which are otherwise not observable.

Again, it should be noted that the values of shape parameter λ for each type of marriage are substantially smaller in absolute value than that of overall marriages, approaching the value of the CM global standard schedule derived from the Swedish experience. This confirms the view that a mixture of different processes such as type of marriage makes the shape of the age distribution of overall marriages more symmetric than the CM standard, while shape of each underlying process tends to follow the standard.

4. Empirical Enhancement

4.1 Empirical Adjustment of the GLG Model

No model fits actual data perfectly. Discrepancies consist of two types of errors; one is random error induced by exogenous factors such as measurement error, and the other is systematic error derived from simplification or insufficiency in model specification. The latter may be corrected by exploiting regularity in the pattern of error. Here we introduce empirical adjustments of the GLG model, seeking a better fit to actual experiences in first marriage of Japanese female cohorts.

The GLG model does not satisfactorily describe the first marriage experiences of Japanese female cohorts. This issue is partly discussed above, where the shape of the standard schedule is inappropriate and therefore is to be set to a specific value to create a country-specific schedule. But even allowing the shape parameter to take value specific to a target cohort, the model schedule deviates noticeably from the observed data. Figure 3 shows observed (dots) and modeled (broken line) first marriage rates for Japanese female cohorts born in 1950. Although the model is best-fitted by optimizing all parameter values including the shape value, the discrepancy is sizable. A similar error pattern is found for every cohort that completed the marriage process in our data set, and therefore the errors can be regarded as systematic. The discrepancy causes serious distortion in estimated parameter values especially when the model is applied to censored cohorts that have not completed their marriage processes. Therefore, seeking better fit for the model is critical in predicting eventual schedule of nuptiality and fertility for cohorts that have not completed the process.

To improve the predictive power of the model in this circumstance, we should capture regularity in the error pattern to be modeled. Difference in the cumulative first marriage rates by age between actual and fitted experiences for 16 cohorts (born in 1935 through 1950) that completed the marriage process are examined. We adjust the cumulative rate function instead of the first marriage rate function because the former is used in parameter estimation, as describe later.

Figure 4 shows the errors for the cohorts. In the figure, the horizontal coordinate is calibrated by standardized age z in terms of parameter u and b , i.e. with usual age x : $z = (x - u) / b$. The origin (0) of the axis indicates the location of mode, since parameter u designates the mode of the GLG schedule. Let $\xi(z)$ denote the error as: $\xi(z) = F(u + bz) - \hat{F}(u + bz; C, \lambda, u, b)$, where $F(x)$ and $\hat{F}(x; \theta)$,

$\theta = (C, \lambda, u, b)$ are the cumulative function of the first marriage rate of observed and model (the latter is alternatively represented by $\hat{F}(z; \theta), \theta = (C, \lambda, 0, 1)$.)

As mentioned above, a highly systematic age pattern of error exists. It is reasonable to assume that there is a particular cause for the very persistent age pattern of discrepancy seen in Figure 4. However, here we just model the pattern empirically. We return to a discussion of underlying causes of the error pattern later.

The straightforward way to incorporate the error pattern into the model is to add an average error pattern to the model. The resulting model $\bar{F}(x; \theta), \theta = (C, \lambda, u, b)$ is expressed as:

$$\bar{F}(x; C, \lambda, u, b) = \hat{F}(x; C, \lambda, u, b) + \hat{\xi}\left(\frac{x-u}{b}\right), \quad (21)$$

where $\hat{F}(x; \theta)$ is the GLG model, and $\hat{\xi}(z)$ is the average error at standardized age z , called the adjustment function (Note 8). We call this model the empirically adjusted GLG model. Although incorporating empirical residual pattern into the mathematical model is not an elegant solution, the simple way out is of practical efficacy.

Here, $\hat{\xi}(z)$ is obtained by averaging the errors of the model applied for Japanese cohorts (born in 1935-50) described above, and is presented in numerical form in Table A-1 in the Appendix. The function is also shown in Figure 4 in a continuous curve along with error dots. To obtain the average error pattern on standardized ages, and to evaluate $\hat{\xi}(z)$ in the new model $\bar{F}(x; \theta)$, some interpolation method is required. Although here the cubic spline interpolation technique is employed, linear interpolation may be adequate for most purposes. There are some constraints on the adjustment function $\hat{\xi}(z)$. First, it is to be zero as z goes to plus or minus infinity to keep parameter C intact as is in the original GLG model. Secondly, integration of $\hat{\xi}(z)$ over the full domain of z should be zero to keep the mean age of the schedule intact. Therefore, we slightly adjust the average error pattern to derive $\hat{\xi}(z)$ so that these properties of schedule are preserved.

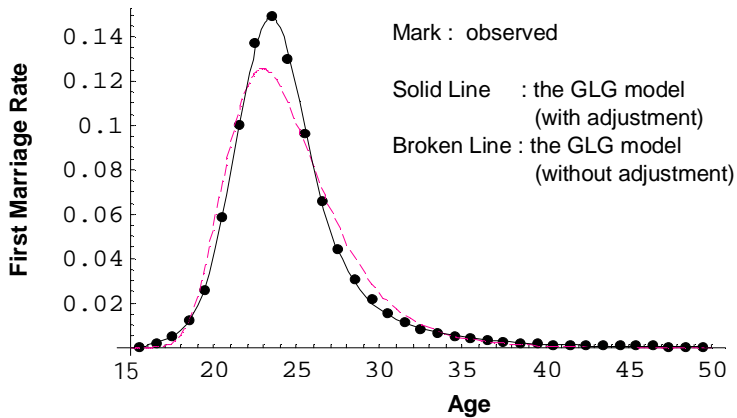


Figure 3: *Observed Age Specific First Marriage Rates and Fitted GLG Model (with and without Adjustment): Japanese Female Cohort born in 1950*

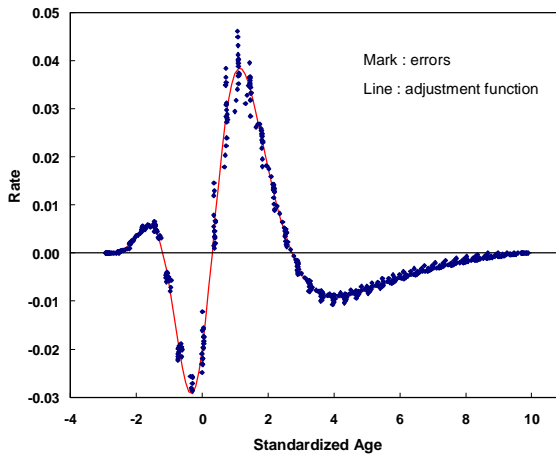
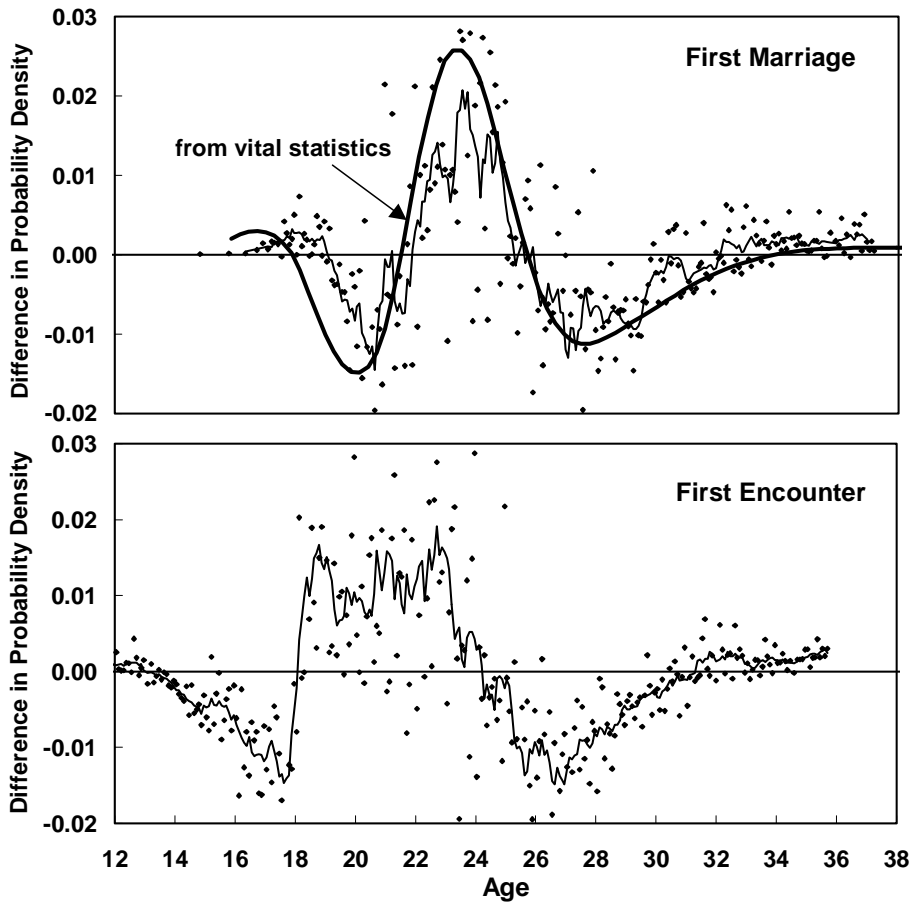


Figure 4: *Errors of the GLG Model in Cumulative First Marriage Rate for Japanese Female Cohort (1935-50) and Adjustment Function*

In Figure 3, we see an improvement in the results produced by the adjusted model (solid line). The curve produced by the adjusted GLG model traces almost exactly the observed rates, while, as already mentioned, the GLG model without adjustment (broken line) does not (Note 9).

Now we briefly discuss the cause of the error pattern. The upper graph of Figure 5 shows the average error pattern in the first marriage rate of the Japanese female cohorts from vital statistics and from a national representative sample. Both patterns indicate that first marriages concentrate on the mode (age 23-24) more than is predicted by the GLG model. A similar error pattern is reported in attempts to fit the Coale-McNeil model to cohort experiences in other countries (for the U.S., Bloom and Bennett 1990, for Swedish male, Ewbank 1974). If the model should represent the “natural” course of first marriage schedule, people should exert a certain kind of regulation on age at marriage resulting in the error pattern. Since in the US, the actual rate exceeds the prediction of the model in the late teens, where the mode locates, Bloom and Bennett speculate that there is a threshold age of 18 before which marriage is hindered by laws or cultural norms. In our case in Japan, however, excess marriages concentrate on age 23-24. Inquiring as to the cause of this residual pattern, we might ask if age at marriage is regulated directly by couples or if the pattern is formed spontaneously in course of marriage process. We observe the error pattern of distribution of age at first encounter with eventual spouse through a national representative survey (the National Fertility Survey) in Japan. The lower graph of Figure 5 indicates that there is a similar deviation pattern in distribution of age at first encounter from the GLG model, which suggests that the regulation is exerted largely on the timing of first encounter, although a difference in the error pattern between first encounter and marriage, especially in their dispersion, indicates that duration from encounter to marriage is partly regulated as well. A sharp rise in deviation of the actual rates of first encounter around age 18 from the model prediction seen in the lower graph of Figure 5 suggests that graduation from high school may be a threshold of behavioral change in first meeting, which supports the view of Bloom and Bennett(1990) that the residual pattern is formed by interference of some social activities.

Hereafter we exclusively use the empirically adjusted version of the model for the first marriage schedule. Notation $F(x)$ instead of $\bar{F}(x)$ is used for simplicity.



Note: Dots stand for residual that are obtained as difference between the Kaplan-Meyer estimates and the GLG model prediction. Thin lines represent their moving average. Thick lines represent the residual pattern from vital statistics. Data is from the National Fertility Survey, round 9, 10, and 11, for married cohorts born during 1937-1959, from the vital statistics for cohorts born in 1935-1950.

Figure 5: *PDF Residual Pattern of the GLG Model of First Marriage and First Encounter with Spouse by Age*

4.2 Method of Parameter Estimation

The Parameter Estimation Method for the adjusted GLG model is no different from the standard method as long as the proper interpolation technique is used for the adjustment term. In a simple situation where age at first marriage of the married people and age at survey of the never married are measured, the likelihood function $L(\boldsymbol{\theta})$ is constructed as:

$$L(\boldsymbol{\theta}) = \prod_{i \in P} f(x_i; \boldsymbol{\theta})^{\delta_i} [1 - F(x_i; \boldsymbol{\theta})]^{1 - \delta_i} \quad (22)$$

where $f(x; \boldsymbol{\theta})$ and $F(x; \boldsymbol{\theta})$ are respectively the density function (age specific first marriage rate) and the cumulative function of first marriage schedule at age x with parameter set $\boldsymbol{\theta}$, which includes C, λ, u, b in our model (21), x_i is age at marriage or age at survey (consor) of individual i depending on whether i is married or never married, δ_i is a indicator variable that takes value one if individual i is married at age x_i and zero otherwise, and P denotes the sample set as a whole. We estimate a set of parameters $\boldsymbol{\theta}$ so as to maximize $L(\boldsymbol{\theta})$, although its logarithm is to be maximized in practice for the sake of ease of calculation.

In the situation above, x_i , age at marriage or at survey is to be exact age. If only aggregated information, such as numbers of marriage classified by age group or even by completed age of single year, is available, the maximum likelihood method with interval censoring is appropriate. Most data of the national level is available only in this form. Suppose that a female cohort of size N at exact age x had m_a marriages in each completed age a ($a < x$), and n_x is left as never married, i.e. $N = \sum_{a=a_0}^{x-1} m_a + n_x$, where

a_0 is age at onset of first marriages. Assuming marriages take place independently, the probability of having such a sample follows the multinomial distribution with $x - a_0 + 1$ parameters (m_a ($a = a_0, a_0 + 1, \dots, x - 1$), n_x). Let $F(x; \boldsymbol{\theta})$ denote the cumulative first marriage rate function. Then the probability (L) is given by:

$$L(\boldsymbol{\theta}) = \frac{N!}{m_{a_0}! m_{a_0+1}! \dots m_{x-1}! n_x!} \left[\prod_{a=a_0}^{x-1} (F(a+1; \boldsymbol{\theta}) - F(a; \boldsymbol{\theta}))^{m_a} \right] (1 - F(x; \boldsymbol{\theta}))^{n_x}. \quad (23)$$

Eliminating the constant factors from the log-transform of L , we maximize the following function to obtain an estimate of θ :

$$\sum_{a=a_0}^{x-1} m_a \ln (F(a+1; \theta) - F(a; \theta)) + n_x \ln (1 - F(x; \theta)) \quad (24)$$

The estimation procedure described above requires number of marriages and population never married as inputs. But in most applications with aggregate data, it is desirable to input rates rather than numbers for the estimation, since numbers are subject to direct influences of death and migration. Here we use the age specific first marriage rate in completed age a as input for m_a , and the proportion never married at exact age x for n_x so as to focus on behavioral aspects of first marriage free from influences of death and migration (Note 10).

4.3 Censoring Effects on Parameter Estimation

Parameter estimation is affected by censoring. This takes place in our research for cohorts that have not completed the marriage process (right censoring). The extent of censoring effects on parameter estimation depends both on the exactness of model specification and data adequacy. Here, we conduct some experiments in which censoring is artificially performed on non-censored cohorts to assess the effects of censoring at various ages on estimated value of parameters.

Examination of estimated values of parameters with artificial censoring shows that the values are quite stable and close to the “true” values that are estimated without censoring when the censoring takes place after standardized age 5.0, which approximately corresponds to normal age 36-40 in the case of Japanese females. It is suggested, therefore, that estimates with censoring after standardized age 5.0 are mostly trustworthy. Examination of estimates of C indicates that the differences between estimated and the true values are within a range of -1.5% to 1.0% for those censored around and after standardized age 2.0, which corresponds to normal age 28-32 in Japan. Therefore, we may expect that we can estimate the proportion eventually marrying for the cohort that has completed the marriage process up to around age 30 with error of less than $\pm 2\%$.

If the values of some parameters are known a priori, it is observed that the prediction of other parameters for young cohorts are more accurate, and with the same accuracy the target range can be extended to younger cohorts. Since parameter λ is expected to be stable in value, it is reasonable to fix it at a certain value such as the

global standard (-1.287) or a country specific value in order to obtain a better prediction for younger cohorts in first marriage schedule. According to our examination, differences of C between estimated and the true values are within a range of -0.4% to 0.2% with censor at standardized age 2.0, if true value of λ is known. In this case we may reasonably expect to be able to predict the proportion never married for cohorts who are above age 30 with an error of less than $\pm 1\%$. In the same condition, parameter u , the location parameter that designates location of the mode, is estimated within a range of -0.015 to 0.01 of the target when censored at standardized age 2.0, and parameter b is estimated within range of -0.05 to 0.01 around the target value. This is adequate accuracy for most demographic applications. Since u and b are only determinants of the mean and standard deviation of age at first marriage if λ is fixed, similar stabilities are expected for those moments.

5. Application of the Adjusted GLG Model

5.1 Estimation and Projection of First Marriage

Now we apply the empirically adjusted GLG model described above to estimate and predict first marriage schedules for female birth cohorts including those that have yet to complete the marriage process. Annual first marriage rates derived from the vital statistics with correction of delayed registration are used as the source data so that the results represent overall Japan (the correction procedure is described elsewhere, Kaneko 2002).

From the estimated annual first marriage rates through the ages and years of 1950-2000, the full lifetime first marriage experiences over ages 15-49 can be extracted only for 16 single year cohorts born during 1935-1950. However, the relevant cohorts to the unprecedented nuptiality and fertility decline in Japan since the mid 1970s are mostly those born after the 1950s. Hence, some reliable predictive tool is required to identify the changes seen in the contemporary nuptiality and fertility reduction. We employ the GLG model adjusted for Japanese females described in the previous chapter for this purpose. We apply it to the cohort first marriage processes to estimate lifetime behavioral measures such as mean age at first marriage, or proportion never married at age 50.

The model schedule is fitted to each cohort experience by estimating model parameter values specific to the cohort through the maximum likelihood method described in the previous section. First, parameter estimation is performed without constraint on parameter for cohorts that have fully and substantially completed their

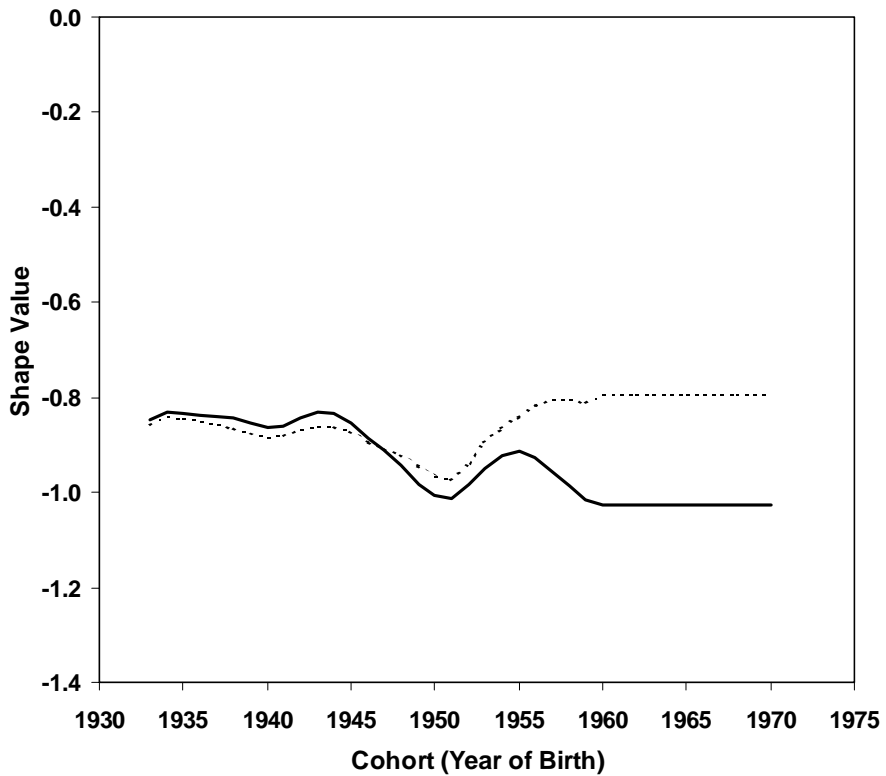
lifetime first marriage schedules. Then, we extend the estimation to younger cohorts that are undergoing various stages in the process, by keeping the shape parameter constant at feasible values as described in the following.

For cohorts that have completed the marriage process, i.e. those born in the years up to 1950, predicted measures by the model agree almost exactly to the observed, since model schedules fit the actual experiences quite well. However, censoring effects on estimates are apparent in younger cohorts born after the mid 1960s, causing estimation results to be increasingly implausible. According to our criterion of reliability in the estimated value of C assessed in the censoring experiments described above, we employ free estimation for cohorts with censoring at standardized age 5.0, which corresponds to cohorts born in 1960 in our data set. For cohorts born after 1960, the value of λ is to be fixed while the other parameters are freely estimated. The criteria for reliable estimation with fixed λ described in the previous section suggests that the border of feasible estimation is around the cohort of 1970. Hence, we limit our observation up to cohorts born in 1970.

Which value should we fix λ to for cohorts born from 1961 to 1970? According to the free estimation, the value of λ shows upward development during 1961 to 1970. It is not certain if the trend is actually happening or is just an artifact due to the censoring effect. Previously we found that the shape value becomes larger (smaller in absolute value) when marriages are a mixture of non-arranged and arranged marriages. Since arranged marriages have been diminishing through the postwar period, the value of λ is expected to decrease instead of increase as seen in the results of free estimation. Thus, here we fix λ at the level of 1960 so as not to let λ increase.

Estimated and fixed values of λ are shown in Figure 6. In the figure, we added the graph of estimated λ for the model without empirical adjustment in a broken line to see effect of the adjustment. Its values after 1960 are fixed at the level of 1960 again. Shape values of the non-adjustment model deviate from that of the adjusted particularly in later cohorts with earlier truncation of the marriage process, tending to be larger which implies a more symmetric shape.

Predicted marriage schedules for the cohort of 1970 are contrasted with those observed in Figure 7. The model schedule follows the actual experiences quite well, even though the cohort is the youngest and its exactitude of the fit is supposed to be the weakest in our data set.



Notes: Solid line: Estimates with empirical adjustment, Broken line: Estimates without empirical adjustment.

Figure 6: Trends of Estimated Value of Parameter λ (Shape Value): Japanese Female Cohorts born in 1933-70

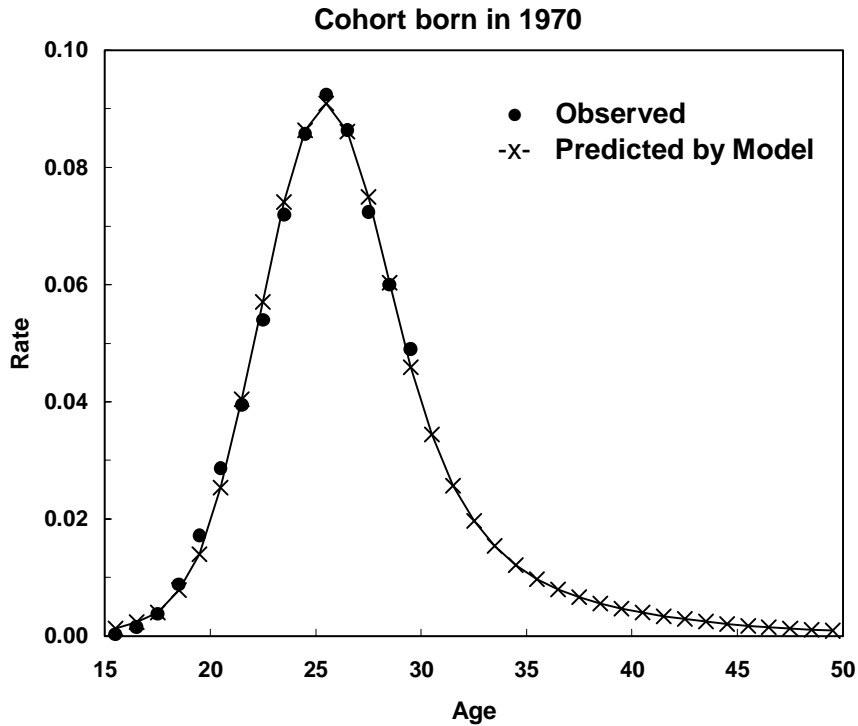


Figure 7: *Observed and Predicted Age Specific First Marriage Rate: Japanese Female Cohort born in 1970*

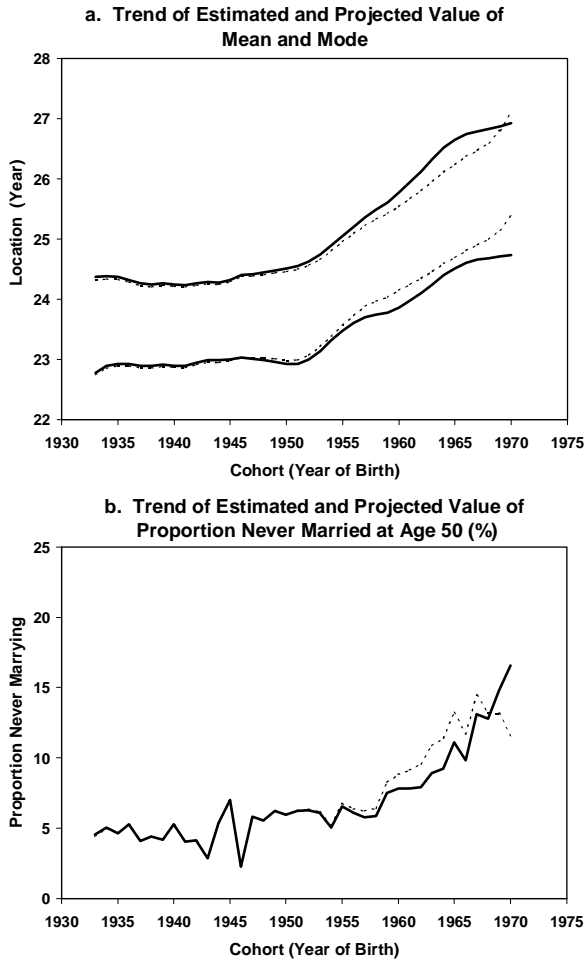
The results of estimation for the mean and the mode of age at first marriage, and the proportion never married at age 50 (γ) are portrayed (solid lines) in Figure 8-a and -b along with estimates for the non-adjustment model (broken lines) again. The trends show a smooth continuous transition from cohort to cohort except the relatively large fluctuation in C for cohorts born at the end of the World War II, probably caused by a flaw in raw statistics. For these indices, the original GLG model without adjustment yields similar estimates to those from the adjusted model for older cohorts. But the results from the former show somewhat different paths from the latter for younger truncated cohorts. These are expected from differences in the abilities of the models to trace age schedules of first marriage (see Figure 3, for instance).

What are the findings from the estimated trends of lifetime measures of first marriage by the empirically adjusted GLG model? The results for cohorts born in 1935-1970 indicate that there are five phases of behavioral change, of which the last three are relevant to the recent unprecedented nuptiality and fertility decline. The change was initiated with a delay in marriage by the cohort born in 1952, followed by a diffusion of never-marrying in cohorts born after 1959 along with prolonged delaying. Then there is an emerging new phase in which the timing shift of marriage is gradually ending in cohorts born after 1965, while the diffusion of never-marrying is rather accelerated. Close examination of hazard rates revealed that the diffusion of never-marrying in the second phase is related to the delaying behavior since marriage propensity in later ages seems to have a bound on increase, and some of postponed marriage have been foregone. On the contrary the diffusion of never-marrying in the third phase is caused by a decline in the propensity to marry even in higher ages as well as early ages. The results suggest that a new phase of marriage behavior is emerging among Japanese women born in and after 1965, which will result in steep increase in lifetime proportion never-marrying (Kaneko 2002).

Note that observation of the trends over cohorts born in from 1952 to 1970 is possible only via the application of some model, and a high level of accuracy in model is required to draw substantive conclusion. The original GLG model (CM model) seems not sufficient in the Japanese case for the recent period described above.

5.2 Application for Fertility Projection

As mentioned before, a model of first marriage schedules also serves for modeling fertility schedules by birth order. Those processes of first marriage and birth by birth order share common structures. The application of the GLG model to birth by order is theoretically expected because of the convolution structure of the GLG model described



Note: Solid line : Estimates with empirical adjustment Broken line : Estimates without empirical adjustment

Figure 8: Trends of Estimated and Projected Lifetime Measures of First Marriage: Japanese Female Cohorts born in 1933-70

in this paper (Note 11). We briefly illustrate an immediate application of the GLG model to fertility in a system of fertility projection, following Kaneko (1993).

Let $F_n(x; C_n, \theta_n)$ be a function of age specific cumulative fertility rate of the n -th child at age x with proportion eventually having n -th child C_n and a set of other parameters θ_n , then:

$$F_n(x; C_n, \theta_n) = C_n G(x; \theta_n) \tag{25}$$

where G denotes the distribution function of the GLG distribution. The function of age specific fertility rate of the n -th birth $f_n(x; C_n, \theta_n)$ is given by:

$$f_n(x; C_n, \theta_n) = \frac{dF_n(x; C_n, \theta_n)}{dx} = C_n g(x; \theta_n) \tag{26}$$

where g denotes PDF of the GLG distribution. However, the observed age specific fertility rate in completed age a should be given by $F_n(a + 1) - F_n(a)$.

The estimation scheme is also identical to that for first marriages except substituting observed frequencies of n -th birth for those of first marriages. If schedules for all birth order are estimated, then the overall age specific cumulative birth rate $F(x)$ is given simply by summing them up to the highest birth order as:

$$F(x) = \sum_{n=1}^L F_n(x; C_n, \theta_n) \tag{27}$$

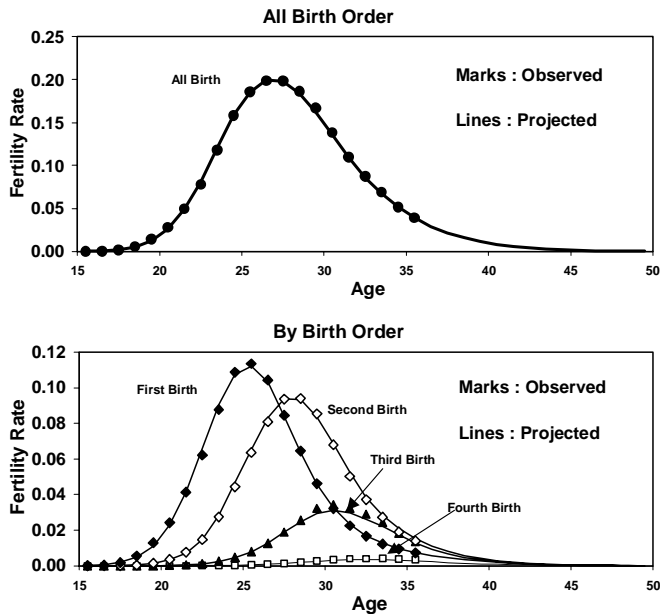
where L denotes the highest birth order. In practice, the class of highest birth order may include certain order of births (e.g. 5-th birth) and higher together so that the summation in (27) includes all births.

In general, the higher the birth order is, the more the shape of the schedule becomes symmetric. There is difficulty for the GLG model to describe the schedule whose shape approaches perfect symmetry. The distribution underlies the perfect symmetric GLG model is the normal distribution. Therefore, the normal distribution model (with extra parameter for prevalence level, C) is to be used as an approximation for the case in which the shape is highly symmetric, or value of parameter λ is near zero.

The model (27) contains $4 \times L$ parameters, which seem to be many more than required to describe overall fertility schedules. Parameters for subsequent birth orders should be correlated and the relationships might be modeled so that we could reduce the number of parameters for parsimony. However, the maximum precision is attained in the original form as long as fertility rate by birth order are available, which is mostly the case with national data.

The empirical adjustment technique employed for first marriage schedule developed in the previous section is applicable to the model of fertility as well. Kaneko (1993) examined the error pattern of the model for each birth order with regard to Japanese female cohorts, and presented the adjustment functions in table form (see Table A-2 in the Appendix).

We now provide an illustration of the application of the model to cohort fertility. In Figure 9, the observed and predicted age specific fertility rates by birth order for Japanese female cohorts born in 1955 with data up to age 35 are plotted together. The model schedules follow the observed rates quite well for all birth orders.



Note: Fifth and higher birth order is not shown

Figure 9: Observed (as of 1991) and projected Cohort Fertility Rates: Japanese Female Cohort born in 1955

The model projects the schedule of this cohort beyond age 35 (the point after which data was not available) to conclude the processes. Applying this projection procedure to every relevant cohort with some assumptions of future fertility behavior for very young cohorts, we obtain a prediction of the period fertility schedule. Using fertility data of cohorts born in 1935-75, the period age specific fertility rates for the year 1985 through 1990 are reconstructed by the model system. The fits are visually presented in Figure 10, which indicates that the system is capable of generating period fertility schedules with adequate precision for most practical purposes (Note 12).

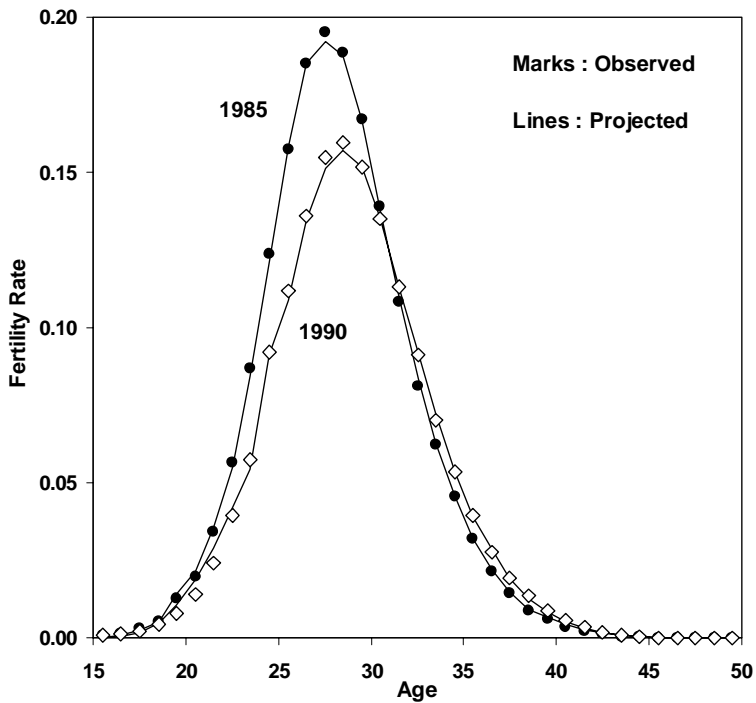


Figure 10: Observed and Projected Period Fertility Rates: Japanese Female, 1985, 1990

6. Summary and Conclusion

The first purpose of the present paper is to show that recognition that the Coale-McNeil (CM) nuptiality model is equivalent to the generalized log gamma (GLG) distribution model allows an expansion of possible application of the model. Some of these applications are illustrated. First, by taking advantage of single parameter representation for the shape of the GLG model, a simple method to derive a country specific standard schedule is proposed. In this course, the significance of the shape specific to a country or region, represented by single value by the GLG model, is indicated. Second, we demonstrated a regression analysis of the effects of covariates on marriage timing. Immediate application of the theories, techniques and software packages of the GLG statistical model to analyze first marriage is main advantage of the new recognition. Here the effects of individual characteristics on first marriage timing were measured with the GLG regression technique, taking account of types of marriage, such as arranged and non-arranged marriages, with the competing risk framework. In our illustration of analysis on Japanese female experiences, we found interesting hidden effects of covariates that would not be found otherwise. These applications revealed also some mechanisms that determine the shape of distribution underlying the first marriage schedule. Heterogeneities of the marriage processes depending both on individual characters and types of marriage (presence of arranged marriage) in Japanese case promote symmetry in shape, which is significantly different from the shape of the Coale-McNeil global standard derived from Swedish experiences. When both types of heterogeneity are controlled, the shape of the schedule of each underlying process tends to follow the global standard.

The second purpose of the present study is to enhance the ability of the model to trace trajectories of the lifetime marriage schedule by incorporating an empirical model of residual error so as to ensure precise estimation results for cohort processes that have not been completed. Employing our findings about stability of residual error patterns for Japanese female cohorts, we successfully incorporated the empirical residual pattern, adjusting its location and scale into the GLG model. The behavioral foundation of the residual pattern was examined, and confirmed to be mainly caused by adjustment in the time of first encounter with future spouse, though there seems an adjustment of dating duration. We conducted a long-term estimation of cohort lifetime measures of first marriages including cohort behavior relevant to the recent drastic reduction in nuptiality and fertility observed in Japan, finding that a new phase of marriage behavior where the proportion never-marrying will drastically increase is emerging. The predictive power of the GLG model allows estimation for behavior of young cohorts, and the empirical adjustment ensures the precision, otherwise quite different schema is drawn, as illustrated.

It should be noted that every aspect of arguments on the GLG model for the first marriage schedule can be directly applied for the fertility schedule by birth order, because of the formal equivalence in structures of those processes. Finally, we demonstrated an application of the enhanced model to the fertility projection system. The performance of the system to predict cohort and period age specific fertility rates seems satisfactory so that it is utilized for country specific precise fertility projection.

7. Acknowledgements

The author thanks Samuel Preston, Shiro Horiuchi, and the anonymous reviewers for valuable comments on the preceding drafts of this paper. He also thanks the late Ansley Coale, who gave essential comments on this study in its early stages.

Notes

1. This may be regarded as a continuous version of the age specific first marriage rate that is in the strict sense defined by a definite integral of $f(x)$ over the relevant age range.
2. Pearson's correlation coefficient between meeting and waiting time to first marriage is -0.48 (that of meeting and time to engagement is -0.45, and of meeting and engagement period is -0.22) for Japanese women born in 1938-54. Partial correlation coefficient between the age at meeting and the waiting time to first marriage with cohort effect controlled is virtually not affected (-0.45). The analysis was carried out on the data from the Ninth National Fertility Survey in 1987 conducted by National Institute of Population and Social Security Research.
3. PDF of the Gamma distribution with two parameters, k and δ , is:

$$f(t) = \frac{\delta(\delta t)^{k-1}}{\Gamma(k)} \exp[-\delta t] \quad t > 0$$
 while that of the GG distribution with additional parameter, η , is;

$$f(t) = \frac{\delta\eta(\delta t)^{k\eta-1}}{\Gamma(k)} \exp[-(\delta t)^\eta] \quad t > 0$$
4. The gamma function and the incomplete gamma function are here defined as:

$$\Gamma(y) = \int_0^\infty u^{y-1} e^{-u} du \quad \text{and} \quad I(y, t) = \frac{1}{\Gamma(y)} \int_0^t u^{y-1} e^{-u} du$$
 respectively.
5. k is corresponding to $\alpha/|\beta|$ in equation (1).
6. Some of statistical packages include regression application with the generalized gamma distribution. We here utilized LIFEREG procedure in SAS/STAT. Constant term of u , b , and λ are respectively correspond to INTECEPT, SCALE, and SHAPE in the SAS output of the procedure with option NOLOG.
7. We here present only a naïve analysis for examination of covariates effects on marriage processes, because of its illustrative purpose for the use of GLG model. Raymo (2003) closely examined the effects of individual characteristics, educational attainment in particular, on transition probability to first marriage among Japanese women, finding higher educational attainment is increasingly associated with later and less marriage by cohort.

8. It seems possible to introduce an additional parameter as a coefficient of ξ to seek further flexibility. However, it may distort the estimation of the other parameters due to identification problem.
9. The empirical adjustment proposed here is primarily aimed to improve model's ability to trace the age schedules in macro level (demographic) applications, in which the improved accuracy is crucial for predictive use. Though it improves accuracy of estimates in the regression analysis as well, the adjustment is mostly surplus for the regression analysis in the cost of giving up application of the techniques and prevailing software packages. For instance, the cohort effects on first marriage timing presented in Table 1 (Model 1) revised with the adjustment are respectively 0.00, 0.09, 0.16, 0.24 with intercept 23.14 (log-likelihood - 11475.3), while original estimates are 0.00, 0.04, 0.17, 0.20 with intercept 23.34 (log-likelihood -11491.0). As seen in this example, changes in regression coefficients are usually not significant, though likelihood is slightly improved.
10. For the parameter estimation of the empirically adjusted GLG model in this study, specific software (written in C and C++) is developed. To obtain the software, contact the author.
11. If age at (n-1)-th birth (or first marriage if n=1) follows the GLG model, then age at n-th birth that is expressed as a convolution of age at (n-1)-th birth (or marriage) and birth interval to n-th birth follows the GLG distribution. This convolution structure, however, holds only approximately in practice, since it is valid only if age at (n-1)-th birth (or marriage) and birth interval to n-th birth are independent of each other.
12. This system of fertility projection with some modifications has been employed in the official population projection in Japan conducted in 1992, 1997, and 2002.

References

- Bloom, D. E. and N. G. Bennett. 1990. "Modeling American Marriage Patterns." *Journal of American Statistical Association* 85: 1009-1012.
- Bongaarts, J. and G. Feeney. 1998. "On the Quantum and Tempo of Fertility." *Population and Development Review* 24: 271-292.
- Coale, A. J. 1971. "Age Patterns of Marriage." *Population Studies* 25(2): 193-214.
- Coale, A.J. and D. R. McNeil. 1972. "The Distribution by Age of the Frequency of First Marriage in Female Cohort." *Journal of American Statistical Association* 67: 743-749.
- Ewbank, D. 1974. "An Examination of Several Applications of the Standard Pattern of Age at First Marriage." unpublished Ph.D. dissertation, Princeton University, Dept. of Economics.
- Feeney, G. 1972. *A Model for the Age Distribution of First Marriage*. Working Paper No. 23, East-West Population Institute.
- Johnson, N. L., S. Kotz, and N. Balakrishnan. 1994. *Continuous Univariate Distributions*, Vol. 1 (2nd ed.). New York: John Wiley and Sons, Inc.
- Lawless, J. F. 1982. *Statistical Models and Methods for Lifetime Data*. New York: John Wiley and Sons, Inc.
- Liang, Z. 2000. *The Coale-McNeil model. Theory, generalisation and application*. Amsterdam, The Netherlands: Thela Thesis.
- Kaneko, R. 1991a. "A Demographic Analysis of the First Marriage Process." *The Journal of Population Problems* 47(3): 3-27 (in Japanese).
- Kaneko, R. 1991b. "Analysis of marriage and birth with demographic models." *Journal of Health and Welfare Statistics* 38(15): 12-19 (in Japanese)
- Kaneko, R. 1993. "A Projection System for Future Age-specific Fertility Rates." *The Journal of Population Problems* 49(1): 17-38 (in Japanese).
- Kaneko, R. 2003. "Transformations in first marriage behavior of Japanese female cohorts: estimation and projection of lifetime measures via empirically adjusted Coale-McNeil model." Paper presented at the 2003 Annual Meeting of the Population Association of America, May 1-3, Minneapolis, MN.
- Kohler, H.-P. and D. Philipov. 2001. "Variance Effects in the Bongaarts-Feeney Formula." *Demography* 38(1): 1-16.

- Kohler, H.-P. and J. A. Ortega. 2002. “Tempo-Adjusted Period Parity Progression Measures, Fertility Postponement and Completed Cohort Fertility.” *Demographic Research* 6(6): 91-144.
<http://www.demographic-research.org/Volumes/Vol6/6/6-6.pdf>
- Kojima, H., M. Atoh, T. Itoh and M. Ikenoue. 1985. “On the Coale-McNeil Nuptiality Model.” *The Journal of Population Problems* No. 175: 36-47 (in Japanese).
- Klein, J. P. and M. L. Moeschberger. 1997. *Survival Analysis, Techniques for Censored and Truncated Data*. Springer-Verlag: New York.
- Prentice, R.L. 1974. “A Log-Gamma Model and Its Maximum Likelihood Estimation.” *Biometrika* 61: 539-544.
- Rodriguez, G. and J. Trussell. 1980. *Maximum Likelihood Estimation of the Parameters of Coale’s Model Nuptiality Schedule from Survey Data*. World Fertility Survey, Technical Bulletins, No.7.
- Raymo, J. M. 2003. “Education Attainment and the Transition to First Marriage among Japanese Women.” *Demography* 40(1): 83-103.
- Ryder, N. B. 1964. “The Process of Demographic Translation.” *Demography* 1(1): 74-82.
- Ryder, N. B. 1980. “Components of temporal variations in American fertility.” In R. W. Hiorns (ed.), *Demographic Patterns in Developed Societies*. pp. 15-54. Taylor and Francis: London.
- Sørensen, A. and A. B. Sørensen. 1986. An event history analysis of the process of entry into first marriage, in David, I. Kertzer(ed.), *Current Perspectives on Aging and the Life Cycle: A Research Annual : Family Relations in Life Course Perspective*, Vol. 2. pp.53-71.
- Stacy, E. W. 1962. “A Generalization of the Gamma Distribution.” *The Annals of Mathematical Statistics* 33: 1187-1192.
- Takahashi, S. 1978. “Application of the Nuptiality Model Schedule to Japan.” *Kokumin Keizai Zasshi* 138(6): 80-95 (in Japanese).
- Trussell, J., and E.D. Bloom. 1983. “Estimating the Co-variables of Age at Marriage and First Birth.” *Population Studies* 37: 403-416.
- van Imhoff, E. 2001. “On the Impossibility of Inferring Cohort Fertility Measures from Period Fertility Measures.” *Demographic Research* 5(2): 23-64.
<http://www.demographic-research.org/Volumes/Vol5/2/5-2.pdf>

Appendix

Table A-1: Adjustment Function of the GLG Model for First Marriage of Japanese Female Cohorts: $\xi(z)$

Standardized Age (z)	Adjustment Function	Standardized Age (z)	Adjustment Function
-3.0	0.00000	3.6	-0.00859
-2.8	0.00000	3.8	-0.00912
-2.6	0.00011	4.0	-0.00931
-2.4	0.00069	4.2	-0.00926
-2.2	0.00188	4.4	-0.00901
-2.0	0.00358	4.6	-0.00864
-1.8	0.00513	4.8	-0.00821
-1.6	0.00600	5.0	-0.00774
-1.4	0.00478	5.2	-0.00724
-1.2	0.00006	5.4	-0.00673
-1.0	-0.00713	5.6	-0.00623
-0.8	-0.01573	5.8	-0.00573
-0.6	-0.02372	6.0	-0.00524
-0.4	-0.02885	6.2	-0.00478
-0.2	-0.02761	6.4	-0.00436
0.0	-0.02014	6.6	-0.00395
0.2	-0.00728	6.8	-0.00356
0.4	0.00756	7.0	-0.00319
0.6	0.02134	7.2	-0.00284
0.8	0.03183	7.4	-0.00252
1.0	0.03737	7.6	-0.00222
1.2	0.03830	7.8	-0.00192
1.4	0.03542	8.0	-0.00164
1.6	0.03027	8.2	-0.00138
1.8	0.02393	8.4	-0.00115
2.0	0.01766	8.6	-0.00092
2.2	0.01178	8.8	-0.00071
2.4	0.00669	9.0	-0.00051
2.6	0.00234	9.2	-0.00033
2.8	-0.00127	9.4	-0.00016
3.0	-0.00408	9.6	-0.00004
3.2	-0.00616	9.8	-0.00001
3.4	-0.00763	10.0	0.00000

Note: These are the adjustment values for the cumulative function of the GLG model for first marriage schedule by standardized age prepared for Japanese female cohorts.

Table A-2: Adjustment Function of the GLG Model for Fertility of Japanese Female Cohorts by Birth Order: $\xi(z)$

Standardized Age (z)	Birth Order (n)				
	1	2	3	4	5 and over
-3.6	0.00000	0.00001	-0.00001	-0.00001	-0.00004
-3.4	0.00000	0.00002	-0.00001	-0.00003	-0.00009
-3.2	0.00000	0.00006	0.00001	-0.00008	-0.00012
-3.0	0.00000	0.00012	0.00007	-0.00012	-0.00009
-2.8	0.00011	0.00027	0.00024	-0.00010	-0.00023
-2.6	0.00041	0.00057	0.00062	0.00007	-0.00075
-2.4	0.00097	0.00110	0.00117	0.00043	-0.00131
-2.2	0.00185	0.00188	0.00171	0.00082	-0.00187
-2.0	0.00291	0.00260	0.00192	0.00100	-0.00198
-1.8	0.00386	0.00280	0.00162	0.00054	-0.00171
-1.6	0.00381	0.00199	0.00058	-0.00045	-0.00173
-1.4	0.00213	-0.00015	-0.00156	-0.00150	-0.00147
-1.2	-0.00142	-0.00321	-0.00459	-0.00289	-0.00070
-1.0	-0.00667	-0.00626	-0.00740	-0.00394	0.00158
-0.8	-0.01246	-0.00913	-0.00905	-0.00414	0.00565
-0.6	-0.01713	-0.01163	-0.00886	-0.00310	0.00829
-0.4	-0.01836	-0.01164	-0.00649	-0.00064	0.00888
-0.2	-0.01562	-0.00854	-0.00240	0.00256	0.00953
0.0	-0.00982	-0.00323	0.00254	0.00423	0.00840
0.2	-0.00128	0.00317	0.00707	0.00481	0.00534
0.4	0.00845	0.00906	0.00943	0.00605	-0.00010
0.6	0.01640	0.01321	0.00989	0.00744	-0.00558
0.8	0.02127	0.01503	0.00952	0.00694	-0.00925
1.0	0.02286	0.01437	0.00861	0.00412	-0.01156
1.2	0.02157	0.01162	0.00701	0.00108	-0.01133
1.4	0.01817	0.00772	0.00457	-0.00101	-0.00855
1.6	0.01364	0.00386	0.00175	-0.00292	-0.00586
1.8	0.00890	0.00075	-0.00065	-0.00406	-0.00334
2.0	0.00449	-0.00154	-0.00228	-0.00394	-0.00048
2.2	0.00064	-0.00314	-0.00326	-0.00378	0.00203
2.4	-0.00248	-0.00410	-0.00369	-0.00337	0.00386
2.6	-0.00474	-0.00446	-0.00377	-0.00367	0.00411
2.8	-0.00617	-0.00438	-0.00350	-0.00189	0.00346
3.0	-0.00689	-0.00404	-0.00295	-0.00106	0.00269
3.2	-0.00708	-0.00354	-0.00235	-0.00039	0.00185
3.4	-0.00689	-0.00298	-0.00182	0.00006	0.00123
3.6	-0.00645	-0.00242	-0.00135	0.00032	0.00076
3.8	-0.00581	-0.00188	-0.00095	0.00042	0.00040
4.0	-0.00506	-0.00139	-0.00063	0.00040	0.00010
4.2	-0.00428	-0.00099	-0.00039	0.00030	0.00000
4.4	-0.00352	-0.00068	-0.00021	0.00021	0.00000
4.6	-0.00285	-0.00044	-0.00010	0.00015	0.00000
4.8	-0.00225	-0.00026	-0.00004	0.00010	0.00000
5.0	-0.00172	-0.00015	-0.00001	0.00005	0.00000
5.2	-0.00126	-0.00008	0.00000	0.00002	0.00000
5.4	-0.00090	-0.00003	0.00001	0.00000	0.00000
5.6	-0.00062	0.00000	0.00000	0.00000	0.00000
5.8	-0.00041	0.00001	0.00000	0.00000	0.00000
6.0	-0.00025	0.00001	0.00000	0.00000	0.00000
6.2	-0.00013	0.00001	0.00000	0.00000	0.00000
6.4	-0.00005	0.00001	0.00000	0.00000	0.00000
6.6	0.00000	0.00001	0.00000	0.00000	0.00000
6.8	0.00002	0.00001	0.00000	0.00000	0.00000
7.0	0.00003	0.00000	0.00000	0.00000	0.00000

Note: These are the adjustment values for the cumulative function of the GLG model for fertility schedules by birth order by standardized age prepared for Japanese female cohorts.