Working Paper Series (E)

No.16

Elaboration of the Coale-McNeil Nuptiality Model as The Generalized Log Gamma Distribution: Empirically Adjusted Coale-McNeil Model

Ryuichi Kaneko

December 2002

National Institute of Population and Social Security Research
2-2-3 Uchisaiwai-cho, Chiyoda-ku,
Tokyo, Japan

No.16

Elaboration of the Coale–McNeil Nuptiality Model as The Generalized Log Gamma Distribution: Empirically Adjusted Coale–McNeil Model

Ryuichi Kaneko

December 2002

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

Ryuichi Kaneko*

Abstract

The first purpose of this paper is to show that recognition of the identity of the Coale-McNeil nuptiality model as the generalized log gamma distribution model should expand its possibility of application with illustrative examples, including development of country specific first marriage standard schedule, and incorporation of covariates with and without competing risk framework for types of marriage. Another purpose of the study is to enhance the model ability to trace trajectories of the lifetime first marriage schedule by incorporating empirical model of residual pattern so as to ensure precise estimates for cohort experiences that have not been completed. As an application, a long-term estimation of cohort lifetime measures of first marriages including those relevant to the recent drastic reduction in nuptiality and fertility in Japan is conducted, finding emerging phase of marriage behavior with rising proportion never-marrying without timing delay. An application for fertility projection system applying the model for fertility schedule by birth order is also briefly described.

Introduction

The Coale-McNeil (CM) nuptiality model is a mathematical expression of regularity in age patterns of first marriages. It is a standard demographic tool for estimation and projection of age schedules of first marriages and even birth by birth order. However, it is not recognized by researchers that the CM model without a prevalence parameter is identical to the log version of the well-known generalized gamma distribution with limited parameter space (Kaneko 1991). Clear recognition of the identity is beneficial because a rich body of knowledge about statistical properties of generalized gamma distribution can be utilized for demographic applications for nuptiality and fertility on one hand, and some interesting feature such as interpretable convolution structure found in the CM model is to apply to the generalized log gamma distribution on the other. The first purpose of this article is to demonstrate some of the demographic applications that afford an unobstructed view on account of the new identity: Formalization of country-specific standard schedule development, and analysis of the covariate effects on marriage timing with or without application of competing risk framework for different types of marriages.

The second purpose of the study is to enhance the model ability to trace trajectories of lifetime marriage and fertility schedules by incorporating the empirical model of residual pattern so as to ensure precise estimation results for cohort experiences that have not been completed. We demonstrate its usage in estimation and prediction of first marriage schedules by illustrating long-term

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

estimations of lifetime measures for marriage behavior relevant to the recent marriage and fertility decline in Japan. We also illustrate fertility projection for an application of the enhanced model.

Coale-McNeil Model and the Generalized Log Gamma Distribution

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 and scale, and eventual proportion ever marrying are adjusted, Coale and McNeil derived a statistical distribution that fitted observed distribution of age at first marriage (Coale and McNeil 1972). It has a closed form of the probability density function (PDF) given as:

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

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 provide a standard marriage schedule model as a location-scale family of this distribution by fixing 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), by setting $\alpha = 1.145, \beta = 1.896$, and $\mu = -0.805$:

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

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

$$g(x; u, b) = \frac{1}{b} g_s\left(\frac{x-u}{b}\right). \qquad (3)$$

Then, the age specific first marriage rate $f(x)$, a marriage schedule that embodies a rate of marrying at each age for all members of the cohort is represented by:

$$f(x) = C\, g(x) \qquad (4)$$

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

Clear distinction in terms is required. In this paper, the age specific first marriage rate $f(x)$ given by (4) is called first marriage schedule, the underline distribution $g(x)$ given by (3) is called distribution of age at first marriage, and the shape fixed version of the distribution $g_s(x)$ given by (2) is the (global) standard distribution of age at first marriage. Thus, $C\ g_s(x)$ is the (global) standard schedule.

The interesting feature of the CM distribution is that it is a limiting probability distribution of convolution of infinite numbers of mean-related exponential distributions. In other words, it is regarded as a convolution of distribution of its own form and some numbers of related exponential distributions. This structure provides a mathematical mold 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 to eventual spouse, and engagement are required to take place prior to the marriage.

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

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

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

where α, β, and μ are three parameters of the CM distribution used in (1) (Coale and McNeil, 1972). Here $g_X(x;m)$ represents a distribution of time at entering a stage from which the process starts, and $h_T(t;m)$ is the distribution of the waiting time that is composed of *m* exponentially distributed waiting time. Mean and variance of distribution $g_X(x;m)$ are respectively $\mu-\frac{1}{\beta}\psi\left(\frac{\alpha}{\beta}+m\right)$, and $\frac{1}{\beta^2}\psi'\left(\frac{\alpha}{\beta}+m\right)$. 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 mean convoluted in distribution $h_T(t;m)$ has 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 ($1/(0.174+2\times0.2881)$) respectively (Coale and McNeil, 1972). Since Kaneko (1991b) found in empirical examination on the female first marriage process in Japan that age at the meeting and durations between meeting and marriage are highly dependent, which is a violation of the convolution assumption of the independence among the sub process, the estimated mean durations above should be biased. Nonetheless, the convolution structure of the CM distribution may provide approximated model of the complicated multistage model, and an important mold in developing models with process dependences in such a way to make parameter of sub process dependent on the outcome of previous stages.

Coale-McNeil Model as The Generalized Log Gamma Distribution

The Coale-McNeil distribution is mathematically identical to the generalized log gamma (GLG) distribution with a somewhat different parameter space (Kaneko, 1991b). The generalized gamma (GG) distribution was defined by Stacy (1962), introducing an additional parameter into the gamma distribution. 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 regarded as a mirror image of the CM distribution by the origin. Prentice (1974) proposed an alternative parameterization of the GLG distribution to extend 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 call the extended version by Prentice simply the GLG distribution.

The PDF and CDF of the GLG distribution are given by:

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

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

where $\lambda\ (-\infty<\lambda<\infty,\neq 0), u\ (-\infty<u<\infty), b\ (>0)$ are three parameters, Γ and I denote the gamma function and the incomplete gamma function respectively ($\Gamma(y)=\int_0^\infty u^{y-1}e^{-u}du$, $I(y,t)=\frac{1}{\Gamma(y)}\int_0^t u^{y-1}e^{-u}du$).

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

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

where $k\ (>0)$ is a new parameter replacement for α , and β is now allowed to take a negative value. Since the CM distribution corresponds to the GLG distribution being with negative value of λ , we regard the GLG distribution as an equivalent of the CM distribution throughout the paper considering only the negative value of λ. One of the advantages of the GLG formulation is that it has only a one-shape parameter, which is λ. Describing the shape of distribution by a single value is crucial in, for instance, making country specific standard schedules and obtaining better parameter estimates by fixing the shape, which are illustrated later.

Mean and variance of this distribution are respectively given as:

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

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

Mode is simply u, with the maximum of PDF given by: $\hat{g}=\dfrac{|\lambda|}{b\Gamma(\lambda^{-2})}(\lambda^{-2})^{\lambda^{-2}}e^{-\lambda^{-2}}$.

The GLG distribution includes as special cases some of the fundamental distributions. The extreme value ($\lambda=1$), the (standard) log-gamma ($\lambda=b, u=-2\ln\lambda$) distributions, and even the normal distribution as a limiting case when $\lambda\to 0$. These relations are, of course, a reflection of relationships between the GG distribution with the exponential, the Weibull, the gamma; and the log normal distributions. This generalized feature of the GLG distribution mathematically guarantees that it describes age distribution of first marriages better than those fundamental distributions.

Correspondence of parameters between the CM and GLG distributions are given as:

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

or equivalently,

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

where α, β, and μ are parameters of the CM formulation in (1) (Kaneko, 1991b). 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 is imperative to ensure that many important features explored separately for each of the distributions are to be unified. For instance, since the GLG distribution includes some important distributions as noted above, so does the CM distribution. Conversely, characterization of the CM distribution as a convolution of the infinite number of mean related exponential distribution applies to the GLG distribution as well. In other words, the convolution structure of the CM distribution expressed as formulations (5) and (6) are deservedly held in the GLG distribution as:

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

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

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

A number of theories, application frameworks, and computer software developed for either distribution, especially for the GG and GLG distribution, should be applied to the other. In the following, first we demonstrate the usefulness of the GLG formulation as an analytic tool for first marriage behavior through illustrations, and then we conduct empirical enhancement of applicability of the model to predict trajectories of marriage and fertility schedules.

The GLG Model as an Analytic Tool for First Marriage

Development of Country Specific Standard Schedules

The first demonstration of usefulness from the new identification of the CM model is derived from the feature that it has only one shape parameter (λ). It enables us to make country specific standard schedules quite easily instead of the global standard derived from Swedish experiences that might be inapt for particular countries. In the following, we employ the Japanese case as an illustration.

Some author reported that the CM standard marriage schedule did not fit Japanese experiences equally well to 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, finding substantial difference in the value from the CM standard schedule, i.e. -1.287.

[Figure 1]

Figure 1 shows the trend of parameter λ estimated for Japanese female cohorts who completed a substantial part of the first marriage process (or attained age 40), i.e. cohorts born in 1935-1960. The figure designates differences in the shape values of Japanese cohorts from that 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. Two particular shape values correspond to well-known underlying distributions. The value zero corresponds to the normal distribution, and the value one to the extreme value distribution. The shapes of Japanese cohorts are located in the middle of the CM standard and the normal near the extreme value model, indicating 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 fist marriage schedules is translated in our terms that the shape of the age distribution is common over countries and periods (Coale, 1971). Our close examination, however, indicates that it varies over countries at least for Japan, and undergoes changes along with time as well, although the scale of variation is not large enough to amend Coale's assertion. This approximate stability has been proven by the wide use of the shape-fixed CM standard schedule as a practical tool in many demographic applications. On the grounds that the shape is approximately stable over time in a certain nation, it is advantageous to fix the shape at an appropriate value because of the robustness of parameter estimation of the resulting schedule, since only location and dispersion are to be identified. The following is the procedure to produce a country specific shape-fixed standard model with mean zero and standard deviation of unity.

Denoting shape parameter(s) by $\bar{\lambda}$ for the GLG model at hand, we need the following new parameters to be calculated for the standard:

$$k_s = \bar{\lambda}^{-2},$$

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

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

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

Any distribution $g(x)$ and corresponding schedule $f(x)$ in the family that the standard represents are expressed as:

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

$$f(x) = C\, g(x), \tag{19}$$

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

For Japanese female cohorts born between 1935-60, the average of estimated λ of the GLG model is -0.9123. Following the formula (16) and (17) with this value, our new Japanese standard marriage schedule is given by:

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

In Figure 2 the Japanese standard schedule developed here is compared with the CM standard. Substantial difference in shapes is acknowledged. As noted above, the Japanese standard is more symmetrical than the CM standard, locating a little closer to the standard normal distribution near the standard extreme distribution as noted before. Besides Japanese cohorts, Liang (2000) reported that first marriage distribution of a Chinese female cohort was close to the normal distribution rather than the CM standard.

[Figure 2]

It is reasonable to think that the difference in shape reflects some behavioral features specific to population other than timing and pace of marriage. Therefore, differences in the shape are of interest in its own right. λ is a promising index that conveys the specific feature of an aspect of marriage behavior. Kaneko (1991a) examined the shape of first marriage schedule in Japan, and concluded that a rather symmetric shape in the Japanese schedule was produced by a mixture of different types of marriage, i.e. arranged marriage and others, since each type of marriage tends to occur in a distinctive timing in life. We found that the shape value of each type of marriage approaches the level of the global standard if the GLG model is fitted separately with the competing risk framework, which is described in detail in the following section. When the model is applied to a sample of female cohorts from a national representative survey (the Ninth National Fertility Survey), the shape values of the schedules of non-arranged and arranged marriages are respectively −0.965 and −1.065, while that of all marriages is −0.644. It implies that the dissimilar shape from the CM standard in Japan is formed by a mixture of more than one type in marriage, each of which has a somewhat similar shape to the standard. Furthermore, as we present in the following section, shape values of schedules of non-arranged and all marriages become even closer to the CM global standard if socioeconomic covariates are controlled. Therefore, the skewed shape exhibited by the CM global standard may represent schedules of homogeneous marriage behaviors.

As shown later, it is critical to know the shape in predicting the schedules of young cohorts that have yet to complete their process. Therefore, the search for determinants of the shape is of importance for many applications of the model.

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 standard specification of the GLG regression model, 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 $u_i = \mathbf{X}_i^t \theta$ where $\mathbf{X}_i^t$ denotes transpose of vector $\mathbf{X}_i$. Since the parameter u determines location of schedules (mode), the specification implies that individuals have underlying probabilities that differ in marriage timing depending on their characteristics. With this specification, it applies to age at first marriage for only those who have married.

We conduct the GLG regression for age at first marriage with some demographic and socio-economic characteristics using survey data for illustrative purposes (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 at 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.

[Table 1]

The coefficients listed indicate the amount of marriage delay in 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 some delay is recognized 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.

Here, 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 formed by heterogeneity in marriage timing as pointed out in the previous section.

While the effects of heterogeneity in individual characteristics in relation to first marriage timing are measured above, we next view the effects of heterogeneity in characteristics of marriage itself. There occur several different types of marriage such as non-arranged and arranged marriages, or inter-racial and intra-racial marriages, and so forth. For instance, suppose 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 at the same time to end up in either of these different types of marriage that comes first. According to the survival analysis framework, this type of situation is dealt by the competing risk model, in which several different events have their own probability to take place at a certain point of time mutually independently.

We illustrate competing risk framework by applying it to analysis on determinants of first marriage timing in Japan taking account of 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 on 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 all the more by the presence of siblings. The analysis illustrates that examination by type of marriage with the competing risk framework provides us detailed features of the process, which are otherwise not observable.

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

Empirical Enhancement

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 taking advantage of regularity recognized in the error pattern. Here we introduce empirical adjustment of the GLG model in seeking further goodness of fit to actual experiences in first marriage of Japanese female cohorts.

Performance of the GLG model is not satisfactory for first marriage experiences of Japanese female cohorts. The issue is partly discussed above, where shape of the schedule is inapt and therefore is set to a specific value to create an own standard schedule. But even allowing shape parameter to take specific value to a target cohort, the model schedule deviates somewhat noticeably from the observed. 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 with optimized parameter values including shape, the discrepancy is sizable. This same pattern is found for every cohort who completed the marriage process in our data set, and therefore the errors are regarded as systematic. They may cause serious distortion in estimated parameter values especially when the model is applied to censored cohorts that have not completed their marriage processes. Therefore, seeking further goodness of fit of the model is essential especially in use for demographic projection of nuptiality and fertility.

[Figure 3]

To improve the performance, 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) who completed the marriage process are examined. We adjust the cumulative rate function instead of the first marriage rate itself 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. for normal age x: $z = (x-u)/b$. The origin of the axis (0) 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 first marriage rate of observed and model (the latter is alternatively represented by $\hat{F}(z;\theta), \theta = (C,\lambda,0,1)$).

[Figure 4]

As mentioned above, a highly systematic age pattern of error exists. It is reasonable to assume that there is a particular cause to generate this very persistent age pattern of discrepancy. However, here we just model the pattern only empirically. We return to discuss causes of the pattern later.

The simplest way to incorporate the error pattern into the model is to add an average pattern to it. 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, and called the adjustment function. We call this model the empirically adjusted GLG model. It seems possible to introduce an additional parameter as a coefficient of $\hat{\xi}(z)$ to seek further flexibility. However, it may distort the estimation of the other parameters due to over identification.

Here, $\hat{\xi}(z)$ is obtained by averaging the errors of the model applied for Japanese cohorts described above, and is presented in numerical form in Table 2. The function is also shown in Figure 4 in a solid line along with dots. To obtain the average error pattern on standardized ages, and to evaluate $\hat{\xi}(z)$ in the new model $\bar{F}(x;\theta)$, an interpolation method is required. Although here the cubic spline interpolation technique is employed, the 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 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 $\hat{\xi}(z)$ in Table 2 so that these properties of schedule are kept.

[Table 2]

In Figure 3, we see an improvement 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.

Now we briefly discuss the cause of the error pattern. 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. The both patterns indicate that first marriages concentrate on the mode (age 23-24) at the cost of those on neighboring ages. A similar error pattern is reported in the attempt to fit the Coale-McNeil model to cohort experiences in other countries (for the U.S., Bloom and Bennett 1990, for Sweden, Ewbank 1974). If the model represents 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 the actual rate exceeds model prediction in 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 marriage concentrate on age 23-24. To inquire the cause of this residual pattern, we here ask if age at marriage is regulated directly by couples or the pattern is formed spontaneously in course of marriage process. We observee the error pattern of distribution of age at first encounter with eventual spouse through the National Fertility Survey in Japan. 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 between first encounter and marriage in the error pattern, especially in their dispersion, indicates that age at marriage itself is also partly regulated through regulation of dating duration from first encounter to marriage. A sharp rise in deviation of actual rates from the course that the model predicts around age 18 suggests that graduation from high school is a threshold of behavioral change (first encounter) leading to first marriage, which supports the view of Bloom and Bennett that the residual pattern is formed by interference of some social activities.

Method of Parameter Estimation

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

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

where $f(x;\theta)$ and $F(x;\theta)$ are respectively the density function (age specific first marriage rate) and the cumulative function of first marriage schedule at age x with parameter set θ, x_i is age at marriage or age at survey 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 and zero otherwise, and P denotes the sample set as a whole. We estimate a set of parameters θ so as to maximize $L(\theta)$, although its logarithm is to be maximized in practice for the sake of handiness in calculation.

In the situation above, age at marriage or at survey x_i is supposed to be exact. If only aggregated information, such as numbers of marriage classified by age group even by single year of age, is available, the maximum likelihood method with the interval censoring is appropriate for parameter estimation. Most data at national level falls into this condition. Suppose that a female cohort of size N at exact age x had number of marriage m_a in completed age a, and with number of never married n_x, i.e. $N = \sum_{a=a_0}^{x-1} m_a + n_x$, where a_0 is age at onset of first marriage. Then, assuming marriages are independent of each other, the probability of having such a sample follows multinomial distribution with $x - a_0 + 1$ parameters (m_a $(a = a_0, a_0 + 1, \cdots, x - 1)$, n_x). Letting $F(x;\theta)$ denote the cumulative first marriage rate function, the probability (L) is given by:

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

Since it is equivalent to maximize log transformed of L eliminating constant factors, we use the following function to maximize to obtain estimate of θ:

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

Since the marriage rate is regarded as number of marriages that are standardized for population size, it is sometimes preferable as "raw" data for schedule parameter estimation for the purpose of eliminating influences from death and migration. Hence, we replace m_a with observed first marriage rate at age a, and n_x with the complement of cumulative first marriage rate up to exact age x in the following estimations.

Censoring Effects on Parameter Estimation

If specification of the model to data is not perfect, the result of parameter estimation is affected by censoring, which takes place in our research for cohorts who 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 during parameter estimation using data of 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 true values that are estimated without censor when 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 censor after standardized age 5.0 are mostly trustworthy. Examination of estimates of C indicates that the differences between estimated and true value 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 proportion eventually marrying (consequently proportion never marrying) for the cohort that has completed the marriage process up to around age 30 with error of less than $\pm 2\%$.

If 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 is to be extended to younger. 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 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 true value are within a range of −0.4% to 0.2% with censor at standardized age 2.0 and older, 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 older, and parameter b is estimated within range of −0.05 to 0.01 around the target value. These are 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 these moments. Hence, to identify value of λ and its determinants is of particular significance for predicting precise demographic measures of marriage behavior of young cohorts.

Application of the Adjusted GLG Model

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 lifetime first marriage experiences over ages 15-49 of 16 single year cohorts of 1935-1950 are fully constructed just by rearranging rates. 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 needed to identify the behavioral causes of changes in the contemporary nuptiality and fertility reduction. We make use of the adjusted GLG model for Japanese females described in the previous chapter for this purpose. It is fitted to 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 above. First, parameter estimations are performed without any constraint on parameter values in order to obtain estimated and predicted marriage trajectories for cohorts who have fully and substantially completed their first marriage schedules. Then, we try to extend the estimation to younger cohorts who are at ongoing stages in the process by fixing the shape parameter at feasible values as described in the following.

For cohorts who 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, making estimation results increasingly implausible afterward.

According to criterion of reliability in the estimated value of C assessed in the censoring experiments, we employ free estimation up to cohorts with a censor at standardized age 5.0, which corresponds to cohorts born in around 1960 in our data set. As 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 also 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 a pretense due to the censoring effect. Previously we found that the shape value becomes larger 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, first we fix λ at the level of 1960 so as not to let λ increase. However, we cannot fully exclude the possibility of λ rising for the younger cohorts of 1961-68. Hence, we provide an alternative prediction in which free estimation is employed for cohorts of 1961-68, and then λ is fixed at the level of 1968 for cohorts of 1969 and 1970.

Estimated and fixed values of λ are shown in Figure 6 with the alternative path in a broken line.

[Figure 6]

Predicted marriage schedules for the cohort of 1970 from the results of parameter estimation are contrasted with those observed in Figure 7. The model schedule follows the actual experiences quite well, even though the cohort is the youngest in our examination and exactitude of fit is the weakest in our examination. Alternate schedule with λ fixed at the level 1968 fits slightly better, since the value is closer to that from the free estimation.

[Figure 7]

The results of estimation for the mean and mode of age at first marriage, and the proportion never married at age 50 (γ) are portrayed (solid lines) in Figure 8 with alternative estimates for cohorts born after 1961 (branching broken lines). 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 World War II, probably caused by some inconsistency in original statistics.

[Figure 8]

Estimated trends of lifetime measures of first marriage for cohorts born in 1935-1970 indicates 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 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 with the delaying behavior since marriage propensity in later ages seems to have a bound to increase, and some of postponed marriage have foregone. On the contrary the diffusion of never-marrying in the third phase is caused by decline in the propensity even in higher ages as well as early ages. The results suggest that new phase of marriage behavior is emerging among Japanese women born in and after 1965, which would result in steep increase in proportion never-marrying (Kaneko 2002).

Note that observation of the trends over cohorts born up to 1970 with a certain reliability is made possible only through application of the GLG model with the adjustment we devised here.

Application for Fertility Projection

Models of first marriage schedule serve for generating fertility schedules by birth order making use of the structural resemblance in the process. The application of the model to birth by birth order is rather theoretically expected because of the convolution structure of the GLG model described in this paper(If age at (n-1)-th birth (or first marriage if n=1) follows the GLG model, 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). Here, we briefly illustrate a system of fertility projection following Kaneko (1993).

Let $F_n(x;C_n,\theta)$ be a function of age specific cumulative birth 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:

$$F_n(x;C_n,\theta_n) = C_n G(x;\theta_n) \quad ,$$

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

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

where g denotes PDF of the GLG distribution. However, the observed age specific birth 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 up them as:

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

where L denotes the last birth order. The last birth order may be made up by putting together certain order and higher births that are not significant in frequency.

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 when shape approaches perfect symmetry, which is the schedule of the normal distribution. Therefore, the normal distribution model (with extra parameter for overall level, C) may be used as an approximation for the case in which the shape is highly symmetric, or value of parameter λ is near zero.

It contains many parameters ($4 \times L$), which is apparently more than required to describe overall fertility schedules. Relationships among parameters for subsequent birth orders may be modeled so that we can reduce the number of parameters in the model over all fertility rates for the sake of parsimony. However, the maximum precision is attained in the original form as long as birth data by order are obtained.

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 error pattern of the model for each birth order with regard to Japanese female cohorts, and presented the adjustment functions in table form, which are presented in Table 3.

[Table 3]

Here we now provide an illustration of the application of the model for estimation and projection of age schedules with Japanese female cohort fertility. In Figure 9, the observed and predicted age specific fertility rates by birth order for 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 except the third, for which the model slightly deviates at the modal area. For overall fertility, however, the discrepancies are almost invisible on the graph.

[Figure 9]

The models project the schedules of this cohort beyond age 35, up to which births are observed, to conclude the processes. Applying this projection procedure to every relevant cohort with some assumptions for young cohorts, we obtain a 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, and are visually evaluated in goodness of fits in Figure 10, indicating that the system is capable of generating period fertility schedules with high precision. This system of fertility projection with some modifications has been employed in the official population projection of Japan conducted in 1993, 1994, 1997, and 2002.

[Figure 10]

Recently, minor deviations in the fertility schedule of marginal ages have been observed in Japan: Firstly a small bump seen in first birth in the early 20s due to an increase in premarital pregnancy, and secondly somewhat lower rates in the 40s than predicted by the model. These incidents suggest that the schedule that the model generates may be regarded as demand for fertility at corresponding ages, which may not be realized as it is, due to forces from environmental factors.

Conclusion

The first purpose of the present paper is to show that recognition of the identity of the Coale-McNeil nuptiality model as the generalized log gamma distribution model should expand the possibility of application of the model. Some of these applications are illustrated.

Firstly, taking advantage of the single shape parameter of the GLG model, a simple method to derive a country specific schedule, whose shape is specific to the country, is described. Secondly, two types of heterogeneity relevant to the marriage process are managed for identifying their effect on timing of occurrence with a standard survival analysis framework. One is heterogeneity dependent on individual characteristics, and the other is heterogeneity of the marriage process itself, such as arranged and non-arranged marriages. As for the former, individual characteristics as covariates are incorporated to the model in regression form with a standard survival analysis technique. For the latter heterogeneity, competing risk framework seems appropriate to apply since the same person goes through different processes at a same time to end up as one of these different types of marriages, such as non-arranged and arranged marriages. In our illustration of analysis on those heterogeneities, we found interesting hidden effects of covariates that would not be found otherwise. These applications reveal also some mechanisms that determine the shape of the distribution. Heterogeneities of the marriage processes depending both on individual characters and types of marriage promote symmetry in shape getting remote from the shape of the Coale-McNeil global standard derived from Swedish experiences.

The second purpose of the present study is to enhance the model ability to trace trajectories of the lifetime marriage schedule by incorporating the empirical model of residual pattern so as to ensure precise estimation results for cohort experiences that have not been completed. According to our finding that the residual pattern is stable for Japanese female cohorts with the process completed, we successfully incorporate the empirical residual pattern that follows location and scale of model into the GLG model. As for the cause of the residual pattern, it is confirmed that the pattern is mainly formed at the time of first encounter with future spouse, though there seems an adjustment of dating duration.

We demonstrated a long-term estimation of cohort lifetime measures of first marriages including cohort behavior relevant to the drastic reduction in nuptiality and fertility prolonged to date in Japan, finding that new phase of marriage behavior where proportion never-marrying would drastically increase is emerging. Finally, we demonstrate an application of the enhanced model to the fertility projection system taking advantage of a similar structure of birth process by birth order to first marriage process. The performance of the system to predict cohort and period age specific fertility rates seems satisfactory so that it is utilized for country specific precision-demanded fertility projection, as has been done for official population projections over the last ten years in Japan.

References

Bloom, D. E. and Bennett, N. G., 1990, "Modeling American Marriage Patterns," *Journal of American Statistical Association*, Vol. 85: 1009-1012.

Coale, A. J., 1971, "Age Patterns of Marriage," *Population Studies*, Vol. 25(2): 193-214.

Coale, A.J., McNeil, D. R., 1972, "The Distribution by Age of the Frequency of First Marriage in Female Cohort," *Journal of American Statistical Association*, Vol. 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., Kotz, S., and Balakrishnan, N., 1994, *Continuous Univariate Distributions*, Vol. 1 (2nd ed.), New York: John Wiley and Sons, Inc.

Johnson, N. L., Kotz, S., and Balakrishnan, N., 1995, *Continuous Univariate Distributions*, Vol. 2 (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, "Syokon kaei no jinkougakuteki bunseki (A Demographic Analysis of the First Marriage Process)," *The Journal of Population Problems*, Vol.47, No. 3: 3-27.

Kaneko, R., 1991b, "Jinkougaku moderu ni yoru syussyou-kekkon doukou no bunseki (Analysis of marriage and birth with demographic models)," *Journal of Health and Welfare Statistics*, vol.38, No.15, pp.12-19

Kaneko, R., 1993, "Nenreibetu syussyouritu no syourai suikei sisutem (A Projection System for Future Age-specific Fertility Rates)," *The Journal of Population Problems*, Vol.49, No. 1: 17-38.

Kaneko, R., 2002, "Transformations in first marriage behavior of Japanese female cohorts: estimation and projection of lifetime measures via empirically adjusted Coale-McNeil model," National Institute of Population and Social Security Research, Working Paper Series (E), No.17: Tokyo.

Kojima, H., 1985, "Coale-McNeil no Kekkon Model ni Tuite (About the Coale-McNeil Nuptiality Model)," *The Journal of Population Problems*, No. 175: 36-47.

Klein, J. P. and Moeschberger, M. L., 1997, *Survivla Analysis, Techniques for Censored and Truncated Data*, Splinger-Verlag: New York.

Prentice, R.L., 1974, "A Log-Gamma Model and Its Maximum Likelihood Estimation," *Biometrika*, Vol. 61: 539-544.

Rodriguez, G., Trussell, J., 1980, *Maximum Likelihood Estimation of the Parameters of Coale's Model Nuptiality Schedule from Survey Data*, World Fertility Survey, Technical Bulletins, No.7.

Sörensen, Annemette, and Sörensen, Aage B., 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: 53-71 pp.

Stacy, E. W., 1962, "A Generalization of the Gamma Distribution," *The Annals of Mathematical Statistics*, Vol. 33: 1187-1192.

Takahashi, S., 1978, "Kekkon model schedule no Nihon eno Tekiyouni Tuite (Application of the Nuptiality Model Schedule to Japan)," *Kokumin Keizai Zasshi*, Vol. 138, No. 6: 80-95.

Trussell, J., and Bloom, E. D., 1983, "Estimating the Co-variates of Age at Marriage and First Birth", *Population Studies*, Vol. 37: 403-416.

Table 1 Effects of Covariates on Age at First Marriage: The GLG Regression by Type of Marriage without/with Competing Risk Model

Covariates	All Marriage Model 1 (N=4682)	All Marriage Model 2 (N=4682)	Non-Arranged (N=4682) (n=2878)	Arranged (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 High School		0.00	0.00	0.00
High School		0.87 ****	0.82 ****	0.89 ****
Junior College		1.49 ****	1.69 ****	1.00 ****
College and Higher		2.48 ****	2.61 ****	2.05 ****
Fathers 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 *
None or Temporary		-0.44 *	-0.69 **	0.02
Urban Residence		0.42 ****	0.13	0.98 ****
Co-residence of Parent(s)		0.09	0.42 ***	-0.60 ****
Heiress		-0.16	-0.06	-0.22
Number of Sibling		-0.07 **	-0.12 ****	0.04
Scale Parameter (*b*)	2.614	2.460	3.082	3.453
Shape Parameter (λ)	-0.673	-0.761	-1.161	-1.054

N : Sample Size   n : Number of Samples without Censor
* P<0.05   ** P<0.01   *** P<0.001   **** P<0.0001

Note: See note on Table 3-1.
Data Source: The Ninth National Fertility Survey in Japan.

Table 2 Adjustment Function of LGG model for Japanese Female Cohorts: ξ(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	0.03543	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

Table 3 Adjustment Function of The GLG Fertility Model
by Birth Order for Japanese Femeale Cohort: $\xi_n(z)$

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	0.00297	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	-0.00267	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: Figures are adjustments for Cumulative Function of theGLG model at standardized ages for fertility schedules by birth order prepared for Japanese female cohort.

Figure 1 Estimated Shape Parameter Values of LGG Model (λ)
For Japanese Female Cohort born in 1935-60

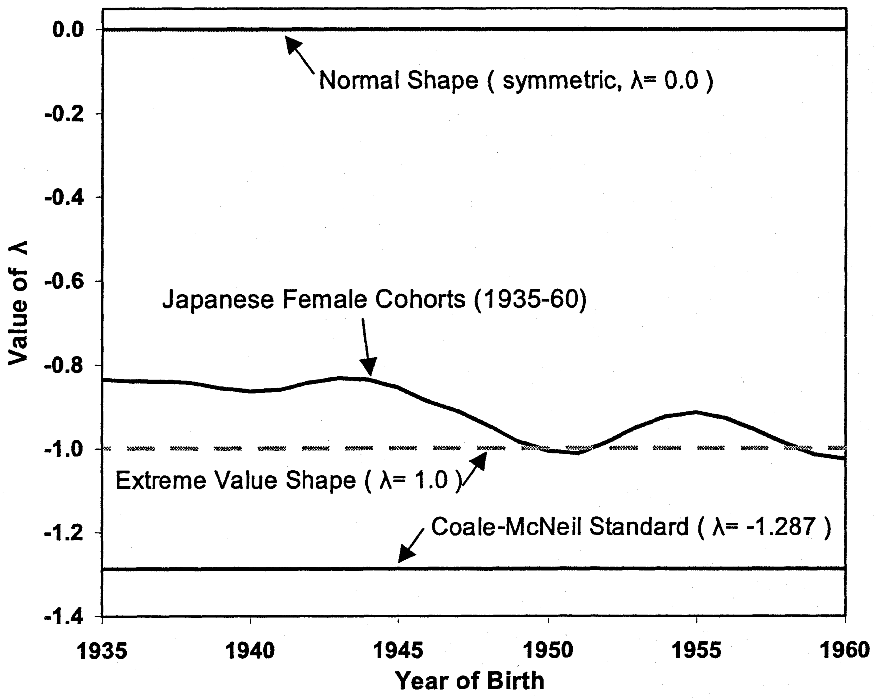


Figure 2 Comparison of Japanese Standard Schedule with The Coale-McNeil Standard for First Marriage of Female Cohort

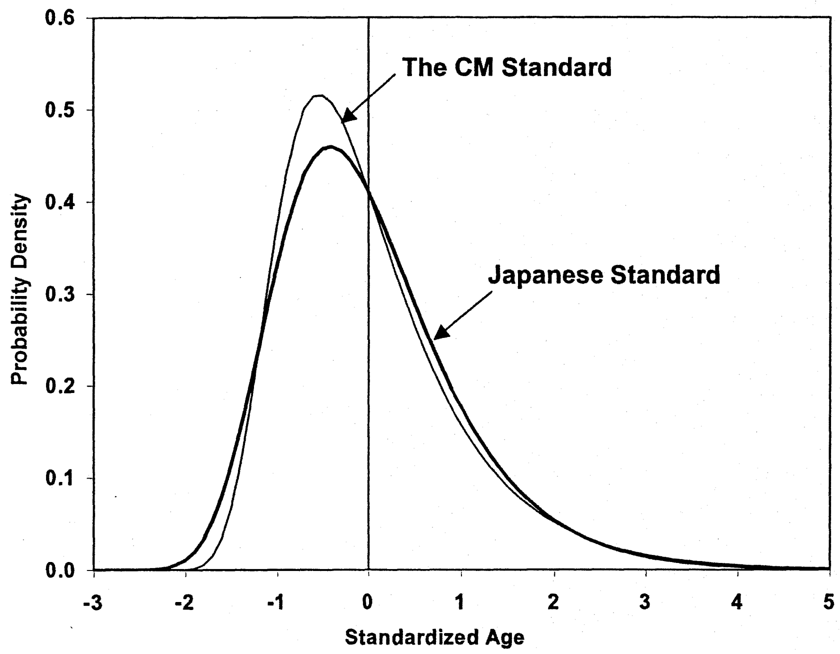


Figure 3 Observed Age Specific First Marriage Rates and Fitted aGLG Model (with Adjustment): Japanese Female Cohort born in 1950

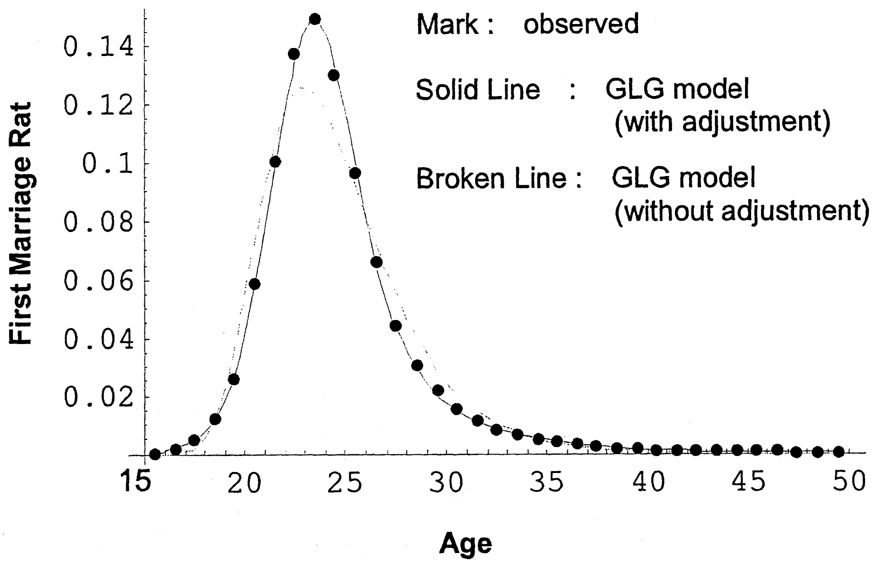


Figure 4 Errors of GLG Model in Fist Marriage Rate for Japanese Female Cohort (1935-50) and Adjustment Function

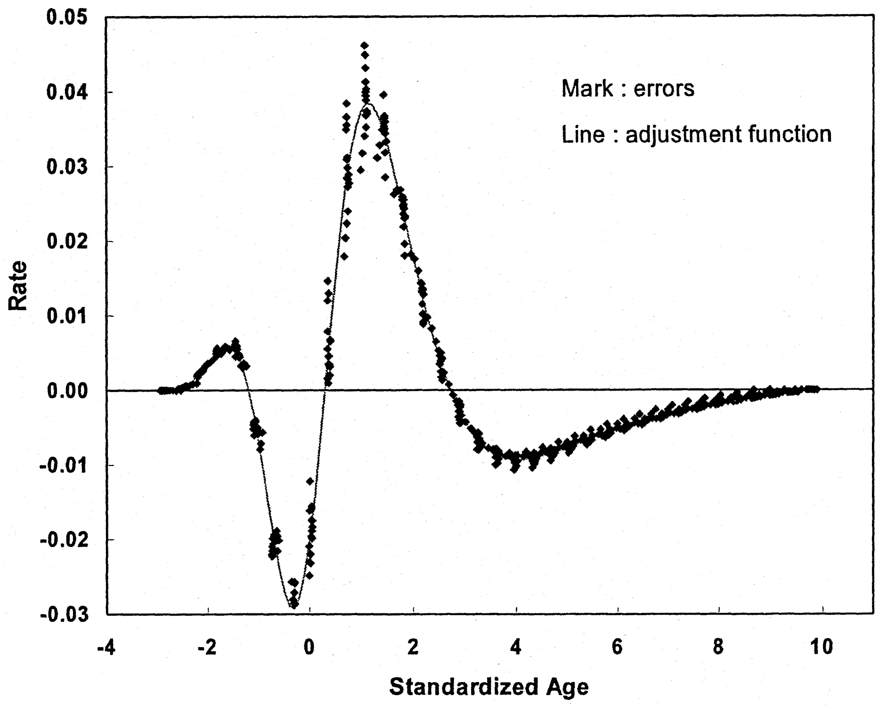
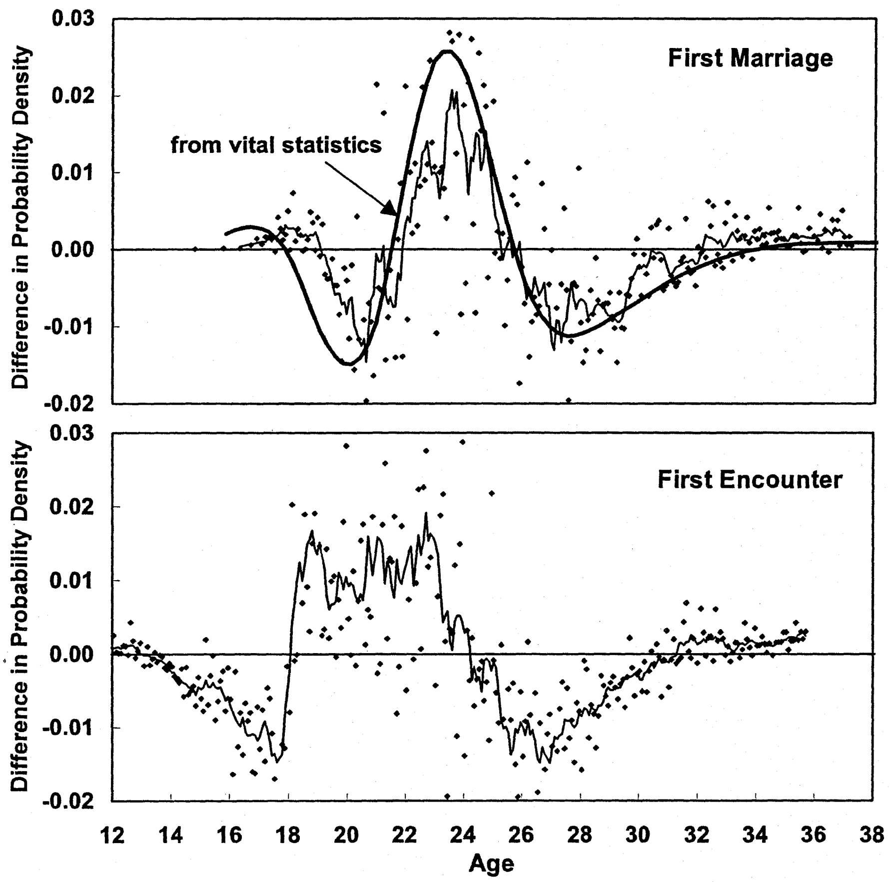


Figure 5 PDF Residual Pattern of GLG Model of First Marriage and First Encounter with Spouse by Age



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 line represents the residual pattern from vital statistics. Data is from National Fertility Survey, round 9,10, and 11, for married cohort born during 1937-1959, and from the vital statistics for cohorts born in 1935-50.

Figure 6 Trends of Estimated Value of Parameter λ (Shape Value)

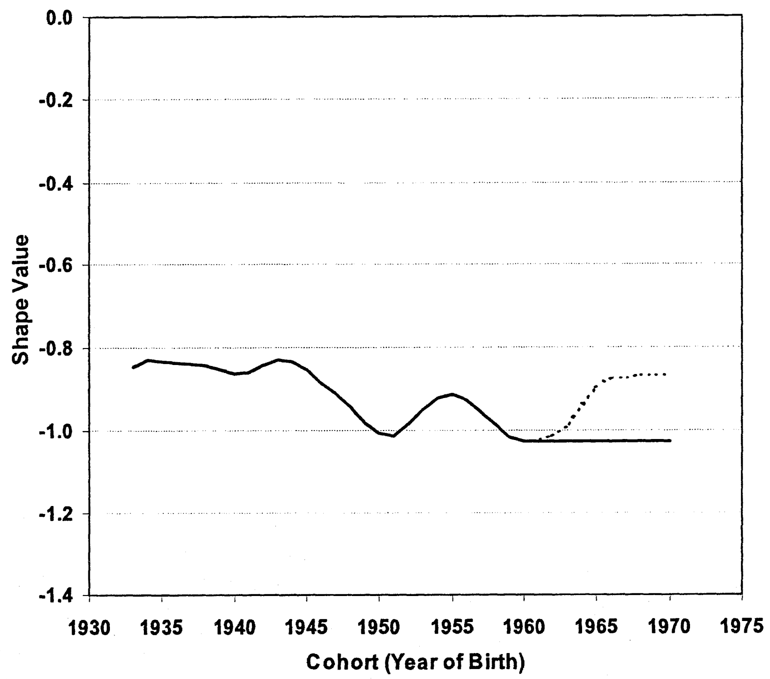


Figure 7 Observed and Predicted Age Specific First Marriage Rate

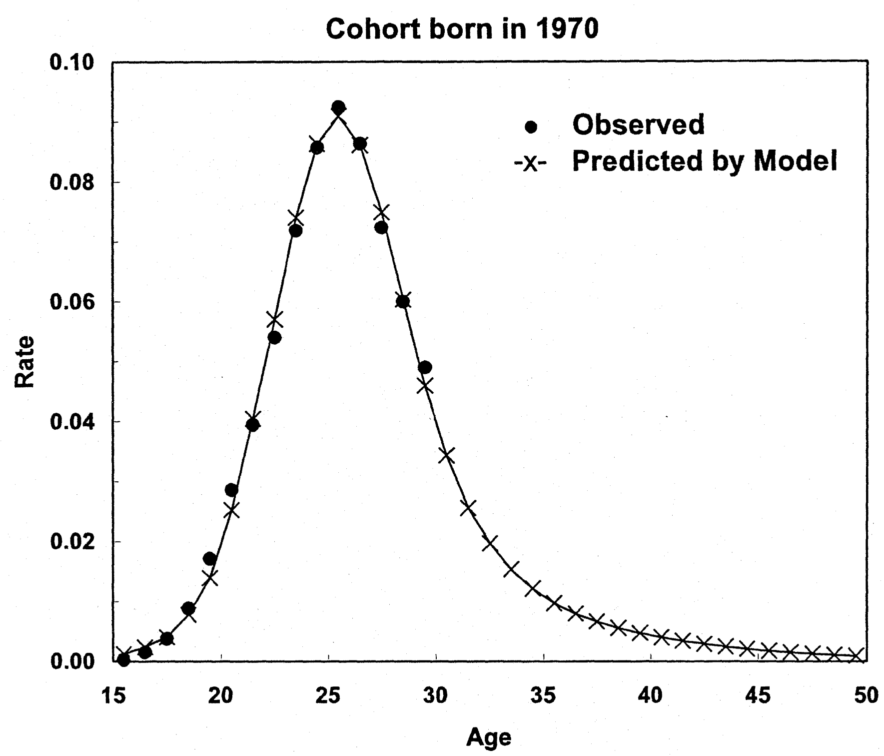
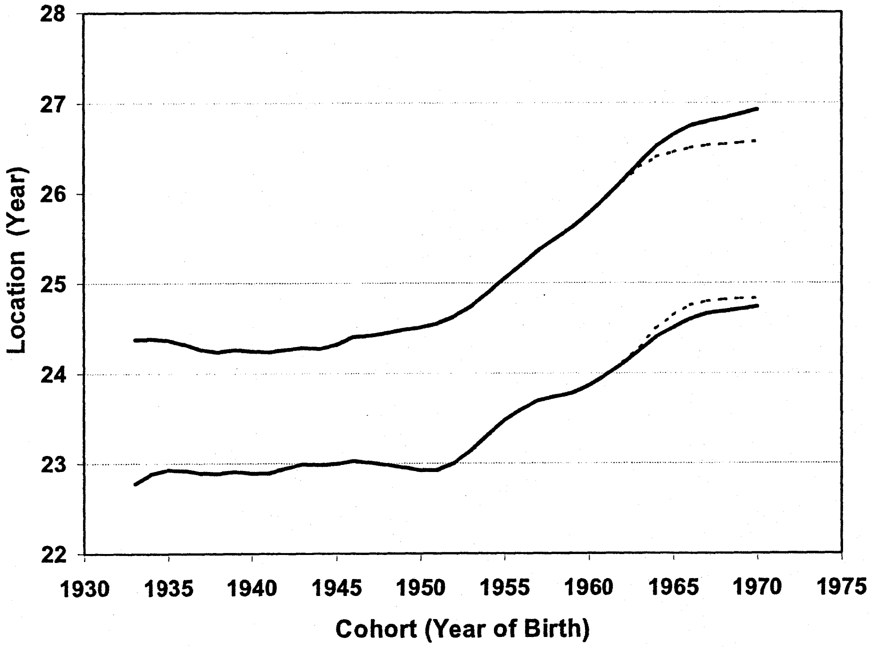


Figure 8 Trends of Estimated and Projected Lifetime Measures of First Marriage

a. Trend of Estimated and Projected Value of Mean and Mode



b. Trend of Estimated and Projected Value of Proportion Never Married at Age 50 (%)

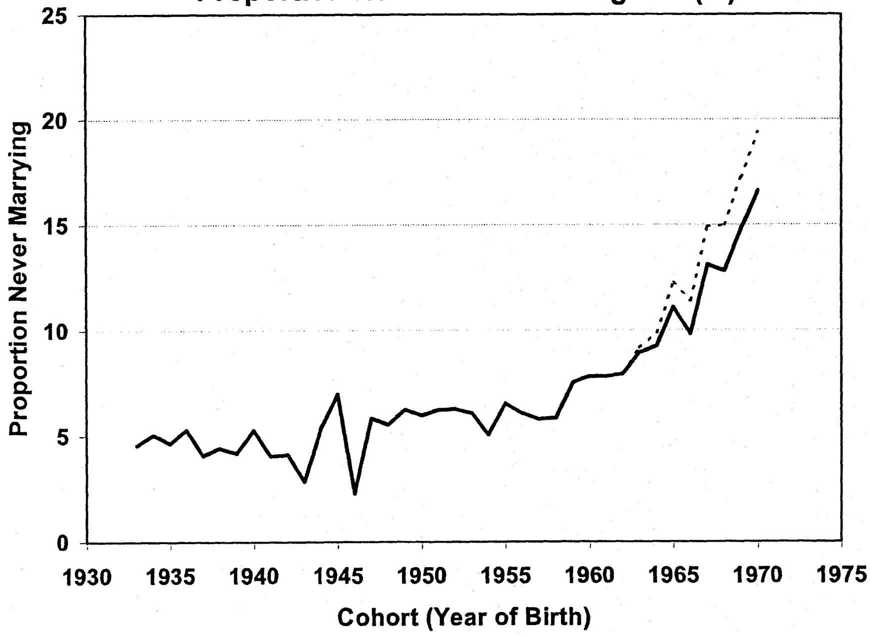
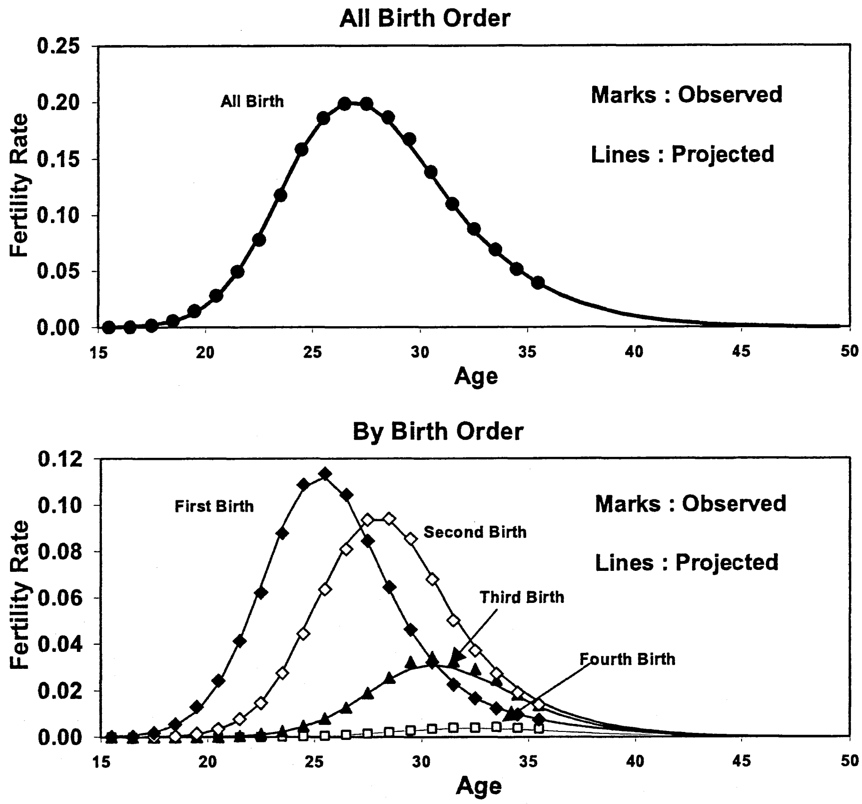


Figure 9 Observed and Projected Cohort Feritility Rates:
Cohort born in 1955 as of 1991



Note: Fifth and higher birth order is not shown.

Working Paper Series(E)

＊ *Working Paper Series are provided for discussions and critical comments, and their circulation is limited.*
Do not quote and cite without a written consent from the authors.

1 The Pattern of Intergenerational Mobility Among Men and Woman
(Sawako Shirahase)
July 1997

2 Pigou' s Welfare Economics and Ethical Belief
(Yuichi Shionoya)
February 1998

3 Health expenditure of deceased elderly in Japan
(Tetsuo Fukawa)
February 1998

4 Population Development and Social Problems in the Inner City and Suburbs of the Tokyo Metropolitan Area
(Satoshi Nakagawa)
April 1998

5 A Social Procedure for Choosing Fair Allocation Rules
－Formalizing the Rawlsian Principles of Justice－
(Reiko Gotoh and Naoki Yoshihara)
April 1998

6 Economy and Morality:A Conceptual Framework
(Yuichi Shionoya)
May 1998

7 Schumpeter on the Relationship between Economics and Sociology from the Perspective of Doctrinal History
(Yuichi Shionoya)
May 1998

8 Women and the Class Structure in Japan and Great Britain
(Sawako Shirahase)
March 1999

9 The Transformation of Partnerships of Japanese Women in the 1990s:
(Miho Iwasawa)
March 2000

10 Effects of Ethnic Concentration on migration in peninsular Malaysia
(Yoshimi Chitose)
June 2000

11 Japanese Income Inequality by Household Types in Comparative Persepective
(Sawako Shirahase)
September 2001

12 Effects of Government Policy on Internal Migration in Peninsular Malaysia: A Comparison between Malays and Non-Malays
(Yoshimi Chitose)
March 2002

13 What Determines Earning Differentials in Urban China?
(Yanfei Zhou)
April 2002

14 Women's Working Pattern and the Support to Working Mothers in Contemporary Japan
(Sawako Shirahase)
September 2002

15 Leaving Home in Japan:
Its Trends, Gender Differences, and Determinants
(Toru Suzuki)
November 2002

16 Elaboration of the Coale-McNeil Nuptiality Model as The Generalized Log Gamma Distribution: New Identity and Empirical Enhancement

(Ryuichi Kaneko)

December 2002