Age at Marriage

We consider models for age at first marriage proposed by Coale (1971) and by Hernes (1972), with an application to predict the "future" of first unions in Colombia. The textbook discusses the first of these models in Section 9.2. We start by introducing some common notation treating age as continuous.

Notation

Let $F(a)$ denote the proportion ever married by exact age a, so $F(\infty)$ is the proportion who ever marry. The derivative $f(a) = F'(a)$ can be described as the marriage density (or frequency of first marriages) at exact age a. The complement $1 - F(a)$ is the proportion who remain never married by exact age a, and is analogous to the survival function $l(x)/l(0)$. Dividing first marriage frequencies by the proportion single we obtain the hazard of first marriage $\mu(a) = f(a)/(1 - F(a))$.

Coale-McNeil

Coale (1971) examined first marriage frequencies in a number of countries and discovered that if he adjusted them for the proportion who eventually marry and for the location and scale of age at marriage they all had almost exactly the same shape (see his Figures 3 and 4). He used reliable data from Sweden to represent this common shape, leading to the model

Here c is the proportion who ever marry and $G_s()$ is the Swedish schedule of proportions ever married by age a among women who eventually marry. To map age in the population of interest to age in the Swedish standard you subtract a_0 and divide by k. The parameter a_0 is described as the age by which a "consequential" number of marriages first occur, and k represents the "pace" of marriage relatively to the Swedish standard. For example if $k = 1$ people marry just as fast as in Sweden, but if $k = 2$ they marry more slowly, taking two years to achieve the same proportions married that Sweden achieves in just one year.

$$F(a) = c\, G_s\left(\frac{a - a_0}{k}\right)$$

In later work Coale and McNeil (1972) found an analytic expression that fits the Swedish standard very well and turned out to be related to the gamma distribution. They also showed that to a very close approximation the model could be described as consisting of the sum of a normally distributed random variable and three exponential waiting times. The first component could represent the age at which one enters the marriage market and the delays could represent the time needed to find a suitable partner, the length of the courtship, and the length of the engagement period. Data from a 1959 survey in France provided some partial support for this behavioral interpretation.

In work I did with Trussell developing methods to fit this model to survey data by maximum likelihood we used the mean μ and standard deviation σ instead of a_0 and k, so that

$$F(a) = c\, G_0\left(\frac{a - \mu}{\sigma}\right)$$

where $G_0(z)$ is a standardized schedule with mean zero and variance one. This yields parameters that are easier to interpret. If $z = (a - \mu)/\sigma$ denotes standardized age the schedule $G_0(z)$ can be computed in terms of the c.d.f. $\Gamma()$ of the gamma distribution as $1 - \Gamma(\exp(-1.896\, z + 0.805), 0.604)$ in R, Stata or Excel.

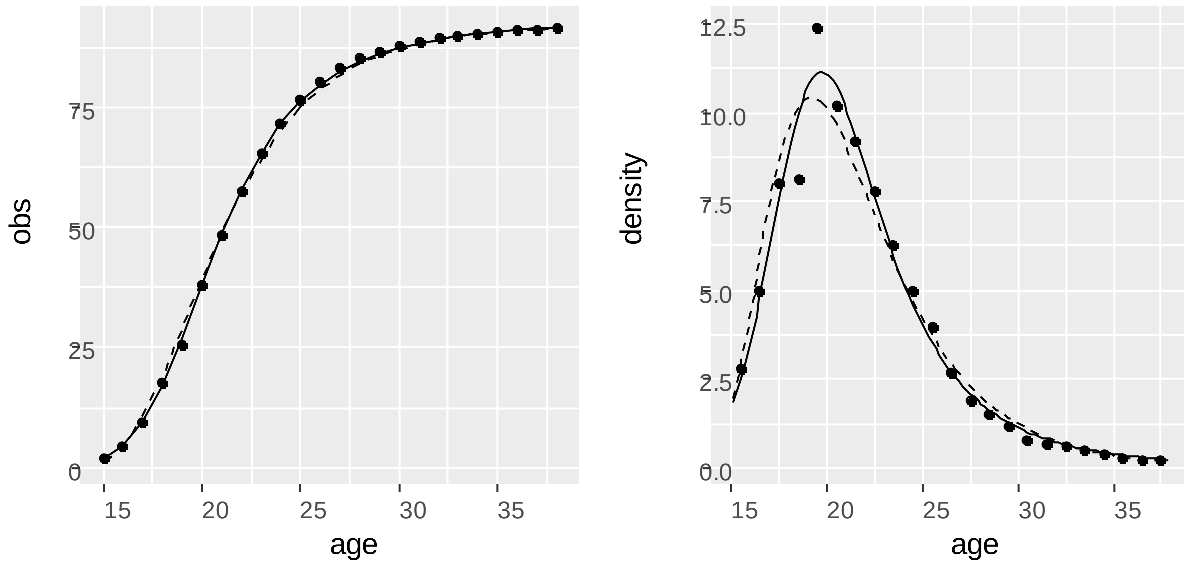

Figure 1 Coale-McNeil and Hernes Nuptiality Models Fit to U.S. Women born in 1920-24

Figure 1 shows the results of fitting the Coale-McNeil model as well as Hernes's model (discussed below) to the 1920-24 cohort of U.S. white women, the data used in Hernes's original paper. The fits to the cumulative schedule are extremely close, but the fits to the first marriage frequencies differ around age 20. We estimate that 92.2% of the cohort ever marry, and that age at marriage has a mean of 21.64 and a standard deviation of 4.56 among those who marry.

Hernes

Hernes (1972) proposed a model that has an interesting behavioral basis in terms of a diffusion process. Specifically, he postulates that people marry as a result of (1) social pressure, which increases as more and more people in a cohort have married, and (2) a person's own attractiveness, which regrettably declines exponentially with age. He develops the model in terms of a differential equation, but I find it easier to think in terms of the hazard of marrying, which is simply the product of the two influences:

$$\mu(a) = A\, e^{-ra} F(a)$$

Actually he writes b^a where I write e^{-ra}, but this is the same thing with $b = e^{-r}$ and $r = -\log(b)$. My notation is intended to remind you of the population growth equation and the exponential survival curve. Basically the model assumes that we lose attractiveness at a constant rate r per year. We often measure age from 15 so A represents attractiveness at that age.

A minor drawback of the model is that you must start the process with someone already married, otherwise there's no social pressure to marry and nothing ever happens. To see this point note that if

$F(a)$ is zero then the hazard is zero and nobody marries. This is not a serious issue in practice; the model usually predicts a small but non-zero probability of being married at very young ages.

The hazard is the ratio of the density $f(a) = F'(a)$ to the survival $1 - F(a)$, so we can write the model as (his equation 7)

$$F'(a)\ Ae^{-ra}F(a)(1 - F(a))$$

Solving this differential equation requires a boundary condition that can best be expressed in terms of the proportion who eventually marry, $F(\infty)$, which for consistency with the previous model I will denote c. The resulting solution looks rather complicated (see his equation 10 or the somewhat simpler forms 12 and 13), but can be simplified drastically by taking logits to obtain:

$$\text{logit}\big(F(a)\big) = \text{logit}(c) - \frac{A}{r}e^{-ra}$$

Hernes illustrates his model by fitting it to data from the U.S. for white women born in 1920-24, with the results shown in Figure 1. He gets an attractiveness parameter of 1.046 at age 15 with a decay rate of 15.7% per year (so his $b = 0.855$) and 92.99% eventually marrying. (I get a slightly better OLS fit with $A = 0.980$, $r = 0.148$ and $c = 0.936$, and a maximum likelihood fit with $A = 1.01$, $r = 0.152$ and $c = 0.931$.)

Age at Marriage in Colombia

The next application comes from a paper I wrote with Trussell years ago, using data from the Colombian World Fertility Survey conducted in 1976. We were particularly interested in checking how well models would predict future marriages. We focused on the cohort aged 35-39 at interview, which had already gone through the prime marrying ages and fitted the Coale-McNeil model. We then repeated the fit using the data we would have had if we had interviewed those cohorts 15 years earlier, when they were only 20-24. Later I fitted the Hernes model to the same data. Figure 2 depicts the fits

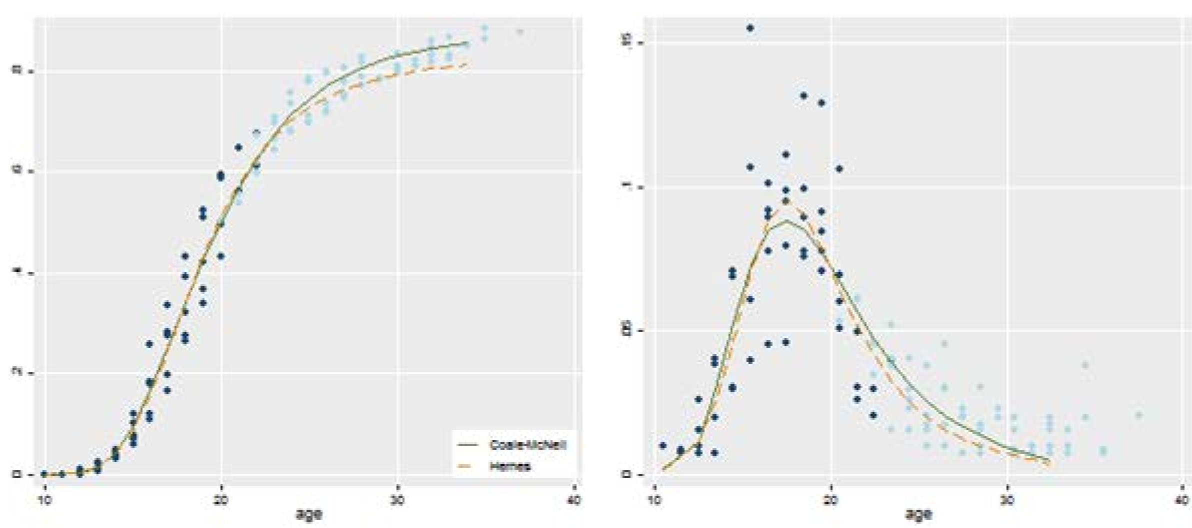

Figure 2. Proportions Ever Married and Marriage Frequencies in Colombia Fitted at Ages 20-24

The bottom line is that the Coale-McNeil did a pretty good job predicting the course of marriage for these cohorts. The Hernes model adapted better to the younger ages, but didn't predict the future quite as well.

The table that follows shows the parameter estimates used in the figure:

Model	Parameter	Cohort at 35-39	15 Years Earlier
Coale-McNeil	Mean	20.44	20.15
	St. Dev	5.38	5.10
	Pem	0.885	0.874
Hernes	Attractiveness	0.652	0.732
	Decay	0.151	0.188
	Pem	0.887	0.830

Please refer to the computing logs in the course website for code showing how to reproduce the graphs. The functions used to compute the models are available in Stata and R in a package called nuptfer that includes a few classic nuptiality and fertility models.

In Stata type `net from http://data.princeton.edu/eco572/stata` and follow the instructions to install the nuptfer package. The documentation is at http://data.princeton.edu/eco572/nuptfer.html.

In R you need to install Hadley Wickman's devtools package using `install.packages("devtools")`, and then install nuptfer from GitHub using `install_github("grodri/nuptfer")`. The package includes documentation.