AUTHOR PLEASE NOTE: The following is a list of names generated from your manuscript text that will be included in the encyclopedia Name Index.

Please check this list carefully and if forenames or initials are missing, please add to the list.

Also please check that the forenames and surnames are in the correct order, if not please mark on the proof.

If no names now follow on this page, no action is required from you on this matter.

Y. Xie[1]
[1] University of Michigan, Ann Arbor, Michigan, USA

Abstract

Demographic models can have two meanings, one broad and one narrow. In its broad meaning, demographic models refer to all mathematical, statistical, forecast, and microsimulation models that are applied to studies of demographic phenomena. In its narrow meaning, demographic models refer to empirical regularities in age patterns of demographic events. This article is concerned with demographic models in the narrow definition. Demographic models are widely use (a) to improve data quality and (b) to compare demographic outcomes and processes across populations or subpopulations. Both parametric and semiparametric specifications have been proposed for modeling age patterns of demographic events, giving rise to parametric models and semiparametric models. Successful applications of both types of models are found in research on mortality, nuptiality, and fertility. As an integral part of formal demography, demographic models have been linked closely to mathematical demography. In recent decades, however, statistical demography has played an increasingly important role in demographic models.

Proofreader Please Note:
Margin marks (in the left margin as the example below) indicate a link from the text reference to an entry in the bibliography. These must be checked against bibliography, and where wrong they must be corrected.

B13   morality (Austin 1954, p. 112). The second sense of morality is found to be indicative of changes of state and not part of the basis of law.

Demographic Models

The term 'demographic models' can have two meanings, one broad and one narrow. In its broad meaning, demographic models refer to all mathematical, statistical, forecast, and microsimulation models that are applied to studies of demographic phenomena. In its narrow meaning, demographic models refer to empirical regularities in age patterns of demographic events. Demographic models in the broad definition can be found in various entries related to demography (see, among others, *Demographic analysis*, *a probabilistic approach to*; *Multi−state transition models*; *Event history analysis*: *applications*; Population Dynamics, *Population forecasts*; *Microsimulation in demographic research*). This article is concerned with demographic models in the narrow definition.

1.   Age Patterns of Demographic Events

In a classic statement, Hauser and Duncan (1959) defined demography as 'the study of the size, territorial distribution, and composition of population, changes therein, and the components of such changes' (p. 2). In the Hauser–Duncan definition of demography, the study of population changes goes hand in hand with the study of the components of population changes. This is necessitated by the need to decompose a population into components and then study changes in the components before arriving at an overall understanding of the changes in the population. The most elementary, and also the most important, form of population decomposition is by sex and age. Since men are only indirectly involved in reproduction, demographic analysis is often simplified by focussing on women. Such simplification is called 'one-sex' modeling (see *Population dynamics*: *two−sex demographic models*), as is commonly seen in fertility and nuptiality models. In mortality models, however, men and women are always kept distinct, and there are substantial mortality differentials in favor of women.

The treatment of age is of significant concern in demographic research. Without exception, the occurrence of all demographic events is age-dependent. Here, the correct interpretation of age-dependency is one of life-course, that is, the likelihood of the occurrence of an event changes as a person (or a cohort) ages. This is true even though most demographic methods and models use cross-sectional data, capitalizing on age-gradients of vital rates in any given population. The use of period-based data usually is necessitated by the lack of cohort-based data.

Demographic models of age schedules are developed on the observation that age patterns of demographic events often show some regularity. Two cautionary notes are in order. First, age regularity is not universal, either across space or over time. Second, all that is assumed is empirical regularity; theoretical

reasons behind such regularities are typically neither well established nor well understood. Nonetheless, demographic models capitalizing on empirical regularities are very useful in practice and may provide the basis for theoretical work. A brief discussion of the main uses of demographic models is given as follows.

2. Use of Demographic Models

Demographic models are intended to summarize empirical regularities in age patterns of demographic events, ideally in simple mathematical formulas. Such models can prove very useful in demographic research. One type of use of demographic models concerns data quality. For example, demographic models may be used to detect and correct faulty data, impute missing data, and allow researchers to infer from partial data. Not surprisingly, this use of demographic models is often found in historical demography and research in less developed countries, where quality data are scarce.

sx0030

The second type of use is substantive and commonly it is found in research with a comparative focus, be it over time (trends) or across societies, regions, or subpopulations. Age-schedule models are also useful for actuarial calculations of life-insurance premiums and premium reserves and for the computation of national or regional population forecasts (e.g., Lee and Tuljapurkar 1994). In demographic applications, cross-group variations in age patterns are typically parameterized as functions of two components: a shape component to capture age effects and a modification component reflecting group membership. Ideally, we would like the second component to be as parsimonious as possible. When demographic models are fully parametric, the two components are integrated, with the functional form being the shape component and the parameter values being the modification component. However, parametric models are not the norm in demography. Parametric models and semi-parametric models will be discussed separately below, with concrete examples taken from research on mortality, nuptiality, and fertility. Coale and Trussell (1996) give a more detailed account of some of the models.

sx0035

B15

B8

3. Parametric Models

Gompertz (1825) is accredited for the discovery that mortality rates in human populations increase nearly as an exponential function of age. This regularity holds true only after early adulthood (ages 25–30). According to 'Gompertz's law,' the force of mortality (or hazard) is a parsimonious loglinear function of age

$$\ln(\mu_x) = A + Bx \qquad (1)$$

where μ_x denotes the force of mortality at age x. Naturally, A is the parameter characterizing the level of mortality at early ages and B is the parameter

sx0040

B9

characterizing the rate of increase in mortality with age. While very parsimonious, the Gompertz model does not always fit empirical data well. Other researchers have modified the model either by adding additional terms or by changing the functional form. For example, Makeham (1867) altered the simple loglinear relationship of Equation 1 by adding a time-invariant term intended to capture cause-specific 'partial forces of mortality.' Heligman and Pollard's (1980) three-component model is another such extension.

sx0045 In studying age patterns of first marriages among women, Coale (1971) found a basic form that is common in different populations. Through trial-and-error, Coale was able to characterize the typical age pattern of first marriage by a double-exponential curve

$$r_x = (0.174) \exp[-4.411 \exp(-0.309x)] \qquad (2)$$

where r_x denotes the hazard rate (or risk) of first marriage at age x. This mathematical function was shown later to be approximated closely by the density function corresponding to the sum of distributions representing demographic processes (Coale and McNeil 1972).

sx0050 To fit observed age patterns of fertility, Brass (1960) proposed a model using polynomial functions. Although Brass's parametric model is very flexible, it requires the estimation of four unknown parameters. Since fertility rates typically are given in a limited number of 5-year intervals, this method leaves very few degrees of freedom for evaluating goodness-of-fit. An evaluation of various parametric models based on empirical data is given by Hoem et al. (1981).

sx0055 There are three notable problems with parametric models. First, parametric models often do not fit observed phenomena. Demographers have dealt with such empirical deviations through (a) restricting the applicable age range and (b) allowing for further parameterization. Both solutions have surfaced in research using the Gompertz model.

sx0060 The second major disadvantage associated with parametric models is the lack of behavioral interpretations for key parameters. This problem arises from the fact that almost all parametric models have resulted from exercises of curve-fitting. Parametric models may reproduce observed age patterns of demographic events, but theoretical interpretation of involved parameters is often unclear.

sx0065 Finally, parametric models are not always convenient to use for comparing populations or subpopulations even though this was one of the motivations for developing them in the first place. This problem is apparent, for example, in the case of a polynomial model. When several parameters in a polynomial function differ between two populations, it is difficult to characterize one population as having higher or lower rates than the other population.

4. Semiparametric Models

In response to these problems with parametric models, semiparametric models have been developed. Semiparametric models are similar to parametric models in specifying parsimonious mathematical functions but differ from parametric models in allowing age-dependency to be unconstrained and subject to empirical estimation, that is, semiparametric models do not impose any global constraint limiting the age pattern to the rigidity of a parametric mathematical function. Instead, the age pattern is estimated freely and empirically from observed data or calculated from external sources. One manifestation of the semiparametric approach is the use of model schedules or model tables. While allowing for the flexibility in a common age function, model tables place constraints on the variations in the age pattern across populations or subpopulations. Such constraints are often motivated by substantive knowledge of demographic phenomena.

In the area of mortality studies, for example, model life tables have been in wide use, primarily as a tool for correcting faulty data and estimating missing data. In essence, a model life table allows for a typical age schedule of mortality shared by a set of populations that differ mainly in their levels of mortality. The age pattern is flexible and empirically determined over all ages but constrained across populations; and cross-population variability lies in the overall level of mortality. The earliest model life table was developed by the United Nations (1955) for all national populations. Subsequently, Coale and Demeny (1966) added more flexibility by identifying four regional model life tables based on distinctive age patterns of mortality and refined the technique for constructing model life tables.

As an alternative, Brass et al. (1968) developed a system of relational life tables, in which a standard mortality schedule is modified by two parameters through a logit equation

$$\text{logit}(p_{xj}) = \alpha_j + \beta_j \, \text{logit}(p_x^*) \qquad (3)$$

Here p_{xj} is the probability of survival to age x in population j, p_x^* is the probability of survival to age x in a selected standard table, and α_j and β_j are parameters characterizing population j. In the Brass system, α determines the level of survival, whereas β changes the shape of survival parametrically.

In fertility research, Henry (1961) posited the existence of natural fertility with a peculiar age pattern. According to him, for physiological and social factors, different populations not exercising fertility control may have different levels of fertility but exhibit the same age pattern. Coale and Trussell (1974) later formalized this idea into the Mm method for studying age patterns of marital fertility. Let the marital fertility rate for the xth age of the jth population be denoted by r_{xj}. The Coale–Trussell Mm method parsimoniously

sx0070

sx0075

sx0080

B17

B5

B2

B12

B7

describes the cross-population variation in age patterns of marital fertility with

$$r_{xj} = n_x M_j \exp(m_j v_x) \tag{4}$$

Here n_x is the standard age pattern of natural fertility; v_x is a typical age-specific deviation of controlled fertility from natural fertility, and M_j and m_j measure the jth population's fertility level and control, respectively. Equation 4 then states that marital fertility can be modeled as a product of natural fertility and a factor reflecting fertility control. The former is represented by $n_x M_j$; the latter by $\exp(m_j v_x)$. Coale and Trussell estimated n and v and suggested that these age schedules be taken as known in all applications. This approach has been superseded by Xie (1990) and Xie and Pimentel (1992), who suggested that n and v be regarded as parameters to be estimated along with M and m from any set of empirical fertility rates r_{xj}.

B18
B19

5. Statistical Demography vs. Mathematical Demography

The construction and derivation of parametric and semiparametric demographic model age schedules have been accomplished through mostly mathematical and sometimes graphical methods. From the outset demographic modeling has been regarded as part of mathematical demography. Since the 1980s, however, a mainline statistical approach has played an increasingly important role in the development of demographic models (e.g., Clogg and Eliason 1988, Hoem 1987, Xie 1990, Xie and Pimentel 1992; see also *Demographic analysis, a probabilistic approach to*) for a number of good reasons.

sx0085

B3
B13 B18 B19

303052

First, the advancement of demography has brought with it more, richer, and better data in the form of sample data, and the use of sample data requires statistical tools. Treating sample data as exactly known quantities induces the danger of contamination by sampling errors. Second, while the method of disaggregation commonly used in mathematical demography may easily lead to inaccurate estimation due to small group sizes, statistical methods more efficiently utilize covariates to examine group differences. Third, because observed data may sometimes be irregular or simply missing, statistical models can help smooth or impute data. Conversely, the strength of empirical techniques developed around demographic models (e.g., indirect estimation, age patterns of fertility, model life tables, etc.) is that they provide descriptions of age patterns that can be utilized to improve statistical analysis.

sx0090

Bibliography

b1
Brass W 1960 The graduation of fertility distributions by polynomial functions. *Population Studies* **14**: 148–62

b2
Brass W et al. 1968 *The Demography of Tropical Africa.* Princeton University Press, Princeton, NJ

!QA:1 Names of all authors please for bib ref 2

Clogg C C, Eliason S R 1988 A flexible procedure for adjusting rates and proportions, including statistical methods for group comparisons. *American Sociological Review* **53**: 267–83

b4 Coale A J 1971 Age patterns of marriage. *Population Studies* **25**: 193–214

b5 Coale A J, Demeny P 1966 *Regional Model Life Tables and Stable Populations*. Princeton University Press, Princeton, NJ

b6 Coale A J, McNeil D R 1972 The distribution of age of the frequency of first marriage in a female cohort. *Journal of the American Statistical Association* **67**: 743–49

b7 Coale A J, Trussell T J 1974 Model fertility schedules: Variations in the age structure of childbearing in human populations. *Population Index* **40**: 185–258

b8 Coale A J, Trussell T J 1996 The development and use of demographic models. *Population Studies* **50**: 469–84

b9 Gompertz B 1825 On the nature of the function expressive of the law of human mortality and on a new mode of determining the value of life contingencies. *Philosophical Transactions of the Royal Society of London* **115**(Part II): 513–85

b10 Hauser P M, Duncan O D 1959 *The Study of Population: An Inventory and Appraisal*. University of Chicago Press, Chicago

b11 Heligman L, Pollard J H 1980 The age pattern of mortality. *Journal of the Institute of Actuaries* **107**: 49–80

b12 Henry L 1961 Some data on natural fertility. *Eugenics Quarterly* **8**: 81–91

b13 Hoem J M 1987 Statistical analysis of a multiplicative model and its application to the standardization of vital rates: a review. *International Statistical Review* **55**: 119–52

b14 Hoem J M, Madsen D, Nielsen J L, Ohlsen E, Hansen H O, Rennermalm B 1981 Experiments in modelling recent Danish fertility curves. *Demography* **18**: 231–44

b15 Lee R D, Tuljapurkar S 1994 Stochastic population forecasts for the United States: Beyond high, medium, and low. *Journal of the American Statistical Association* **89**: 1175–89

b16 Makeham W M 1867 On the law of mortality. *Journal of the Institute of Actuaries* **13**: 325–67

b17 United Nations 1955 *Age and Sex Patterns of Mortality: Model Life Tables for Under-developed Countries*. United Nations, New York

b18 Xie Y 1990 What is natural fertility? The remodelling of a concept. *Population Index* **56**: 656–63

b19 Xie Y, Pimentel E E 1992 Age patterns of marital fertility: Revising the Coale–Trussell method. *Journal of the American Statistical Association* **87**: 977–84

Y. Xie