# Model Life Tables

No single mathematical formula fits the wide range of observed age patterns of mortality, so several authors have developed empirical standards. We review the Coale-Demeny-Vaughan model life tables (a.k.a. the Princeton Regional Model Life Tables), the Brass Relational Logit model and a modification, and the log-quadratic model of Wilmoth and collaborators. The Coale-Demeny and original Brass models are discussed in Section 9.1 of the textbook. Later we will dicuss the Lee-Carter model used in forecasting mortality.

Coale-Demeny

Coale and Demeny examined a large number of life tables from countries with reliable data, mostly in Europe, and used regression methods to build four families of model life tables, labeled "West", "North", "East" and "South" because they corresponded roughly to regions of Europe, with variants for males and females. Each family is indexed by a single parameter representing the level of mortality.

The figure below shows the West female model life tables when life expectancy is 50 and 75 years, to give you an idea of the shape of the models.
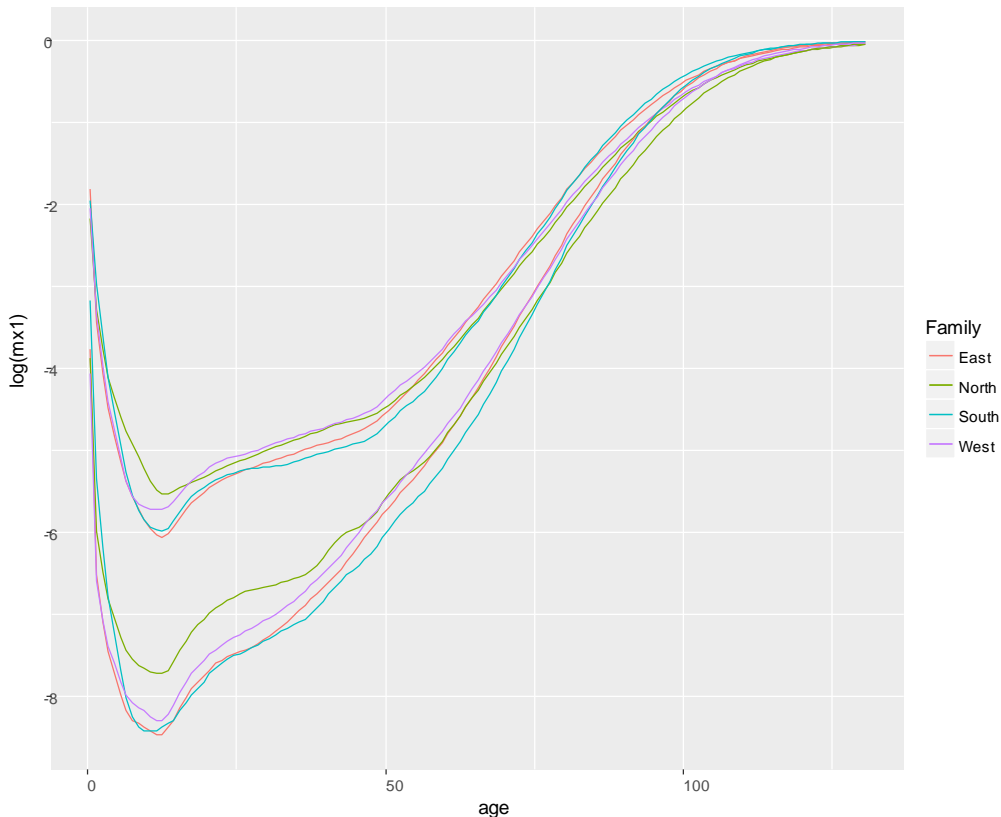
The method of construction is not without interest. For each age group 0, 1-4, 5-9, ..., 75-79 in standard abridged life tables they regressed $_nq_x$ on e_{10}. They used a linear regression and a logarithmic one with $\log_{10} 10000\,_nqx$ as the outcome. Table XI in their book has the coefficients. Neither model was fully adequate, but the two regressions always crossed twice within the range of e_{10}. They decided to use the linear model to the left of the first crossing, the logarithmic to the right of the second, and the average of the two in between, as shown on the right
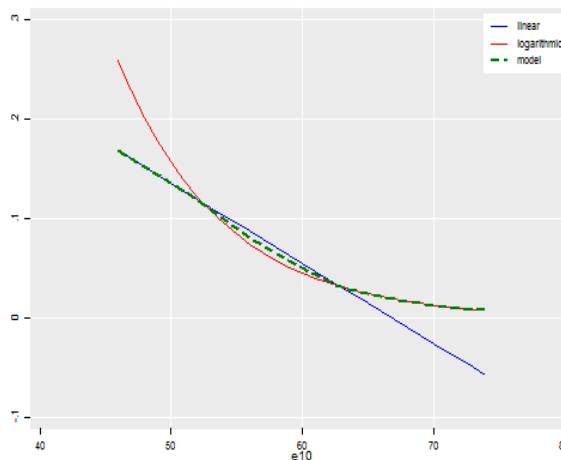
To proceed beyond age 80 they used a Gompertz extension. Briefly they use $_5q_{75}$ to estimate $\mu(77.5)$ and a linear regression to predict $\mu(105)$, use these two values to estimate the Gompertz slope, and then use that to estimate death and survival probabilities for ages 80-84 to 90-94 and 95+. Time lived in 0-1 and 1-5 was computed using the regression equations we already encountered, see Table 3.3 on page 48. Time lived after age 80 was computed by numerical integration of the Gompertz survival function.

The published set of tables has 25 levels, designed to give life expectancy at birth from 20 to 80 in steps of 2.5 years. An oddity of the tables is that the value of e_{10} is not the same as the input value used to generate the death probabilities, which is not surprising in a non-linear system; it is best to think of that as just a seed. These tables have been widely used in demographic applications for many years.

The UN has published a set of "Coale-Demeny" tables which differ from the original published set because they were extended beyond age 90 (all the way to 130) and because they smoothed some of the rates. The UN also has its own set of model life tables, with families called "General", "Far Eastern", "South Asian", "Latin American" and "Chilean", but they haven't quite overcome the popularity of the Coale-Demeny system. The whole set of nine model schedules with life expectancy from 20 to 100 is available from the UN in a spreadsheet.

Brass Relational Logits

Brass proposed a model where he first transforms the survival function by taking logits, which he defines as (assuming a radix of one)

$$Y(l_x) = \frac{1}{2}\log\frac{1 - l(x)}{l(x)}$$

Note the arcane use of 1/2, which was common among British statisticians, and the fact that he works with $(1 - p)/p$, which changes the sign.

He then writes the transformed schedule as a linear function of a standard logit schedule $Y(l_x^s)$, so that

$$Y(l_x) = \alpha + \beta Y(l_x^s)$$

The essential idea is that a simple transformation of the l_x function allows relating any survival curve to another, including the standard. It is very easy to check visually the fit of the model because a plot of the observed logits versus the model logits should yield a straight line.

The plots below show how the parameter α reflects the level of mortality and β the shape of the schedule, or balance between child and adult mortality. (I find the plots of the force of mortality more informative than the plots of the survival functionsdepende shown in the textbook.)
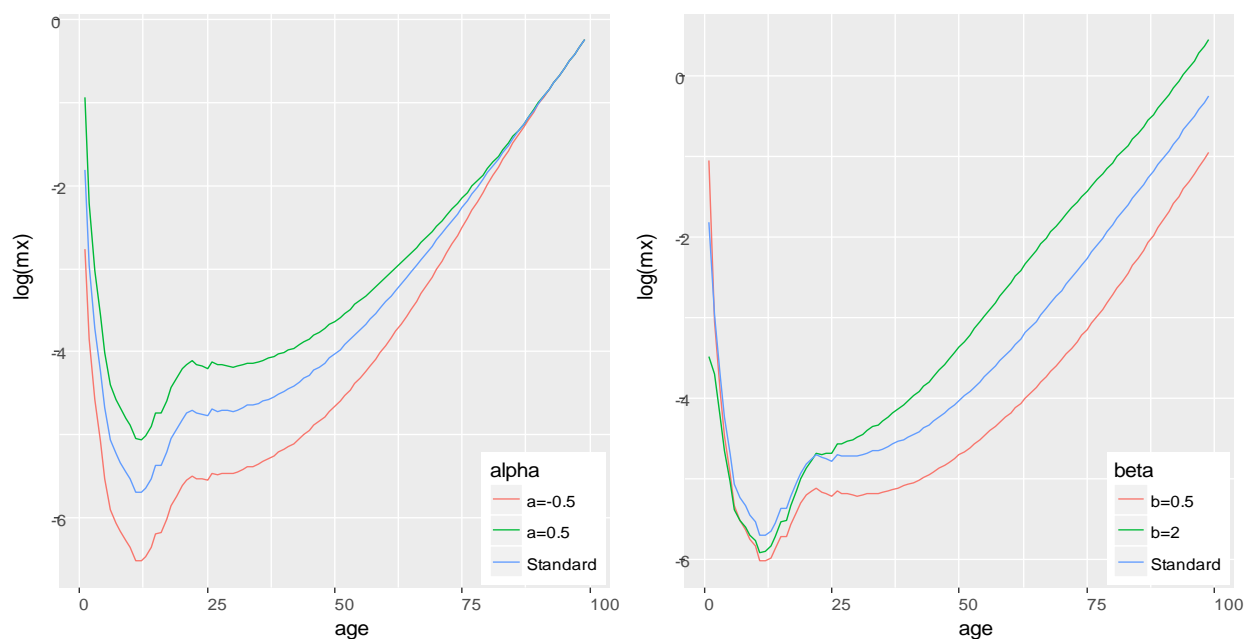
The relational logit model found many applications in countries with limited data, particularly in Africa. The website shows an application of the model to the mortality of Seychelles males in 1971-75, with an excellent fit. The website also has the original single and five-year standards. (The latter is just for convenience, as obviously the survival at five-year age intervals is a subset of the single-year standard.)

Tools for Demographic Estimation (TDE) relies heavily on relational models. Instead of using the Brass standard, however, they propose taking as the standard one of the model life tables in the Coale-Demeny system or in the UN family, usually the Princeton West or the UN general family, with life expectancy 60. Their online materials have extensive notes on the choice of a standard and they provide a spreadsheet with the nine standards.

The Modified Logit System

Murray and collaborators noted in 2003 that the original Brass system could be made to fit much better if instead of the logit function they used a different transformation, while maintaining the idea that after transformation the schedule should be a linear function of a standard, so

$$Z(l_x) = \alpha + \beta Z(l_x^s)$$

After examining a large number of life tables they decided to use a modified logit transformation which incorporates corrections based on mortality at ages 5 and 60. Specifically,

$$Z(l_x) = Y(l_x) + \gamma_x \left(1 - \frac{Y(l_5)}{Y(l_5^s)}\right) + \theta_x \left(1 - \frac{Y(l_{60})}{Y(l_{60}^s)}\right)$$

where γ_x and θ_x are constants and l_x^s is a standard survival function, defined for ages 1,5(5)80, with different values for males and females, all chosen to improve the fit of the model to a large training set of life tables.

The γ_x and θ_x coefficients are zero at ages 5 and 60, which makes this transformation identical to the Brass logit at those ages. At other ages the transformation differs from the logit, with age-specific corrections depending on departures from the standard at ages 5 and 60. Note that the transformation of the standard *is* the Brass logit, as the correction terms vanish. Putting everything together the model is

$$Y(l_x) + \gamma_x \left(1 - \frac{Y(l_5)}{Y(l_5^s)}\right) + \theta_x \left(1 - \frac{Y(l_{60})}{Y(l_{60}^s)}\right) = \alpha + \beta Y(l_x^s)$$

Because the model is linear on the parameters it can be fitted by OLS, regressing $Z(l_x)$ on $Y(l_x^s)$. This yields fitted values $\hat{Z}(l_x)$, which can be converted to survival probabilities solving for $Y(\hat{l}_x)$ in the above equation.

(My presentation follows closely the paper. TDE write the model with the Brass logit $Y(l_x)$ on the left-hand-side and everything else on the right, which I think is confusing because the outcome also appears on the right in the guise of $Y(l_5)$ and $Y(l_{60})$. Because they do this, they note that the signs of their coefficients are reversed relative to the paper. In the end the results are, of course, exactly the same.)

The Log-Quadratic System

In 2012 Wilmoth and collaborators published a new 2-parameter system of model life tables. The model is based on a very large set of reliable life tables from the Human Mortality database, which Wilmoth had started in his Berkeley days.

The basic idea is that the log of the mortality rate in the usual abridged life table is a quadratic function of the log of $_5q_0$ called h for short, and a second parameter k which controls the shape of the schedule.

4

$$\log {}_n m_x = a_x + b_x h + c_x h^2 + v_k k$$

where a_x, b_x, c_x and v_k are constants, which were estimated from a large training set of reliable life tables.

An interesting feature of the model is that all the coefficients for the age group 1-4 are zero, so the equation does not produce an estimate of $\log {}_4 m_1$. The reason is that one of the parameters is the log of ${}_5 q_0$ and the authors obviously care about consistency. So they use the fitted value of ${}_1 m_0$ to produce ${}_1 q_0$, then estimate ${}_4 q_1 = 1 - (1 - {}_5 q_0)/(1 - {}_1 q_0)$ and finally convert this to a rate. These steps use the Coale-Demeny ${}_n a_x$ values that we have encountered already. Because of this extra work, the fitted life table reproduces exacly the ${}_5 q_0$ value used as an input parameter.

I think this model represents the current state of the art and appears to fit well a wide variety of mortality schedules, but the modified logit system is a strong competitor.

All of these models can be used to estimate mortality from limited data. TDE has an application estimating a complete life table for Kenya using only measures of child and adult mortality, namely ${}_5 q_0$ and ${}_{45} q_{15}$.

Austrian Males in 1992

The figure below shows the log-mortality rates in the life table we estimated for Austrian males in 1992, which is Box 3.1 in the textbook, and fitted values based on the modified logit and the log-quadratic models.

5
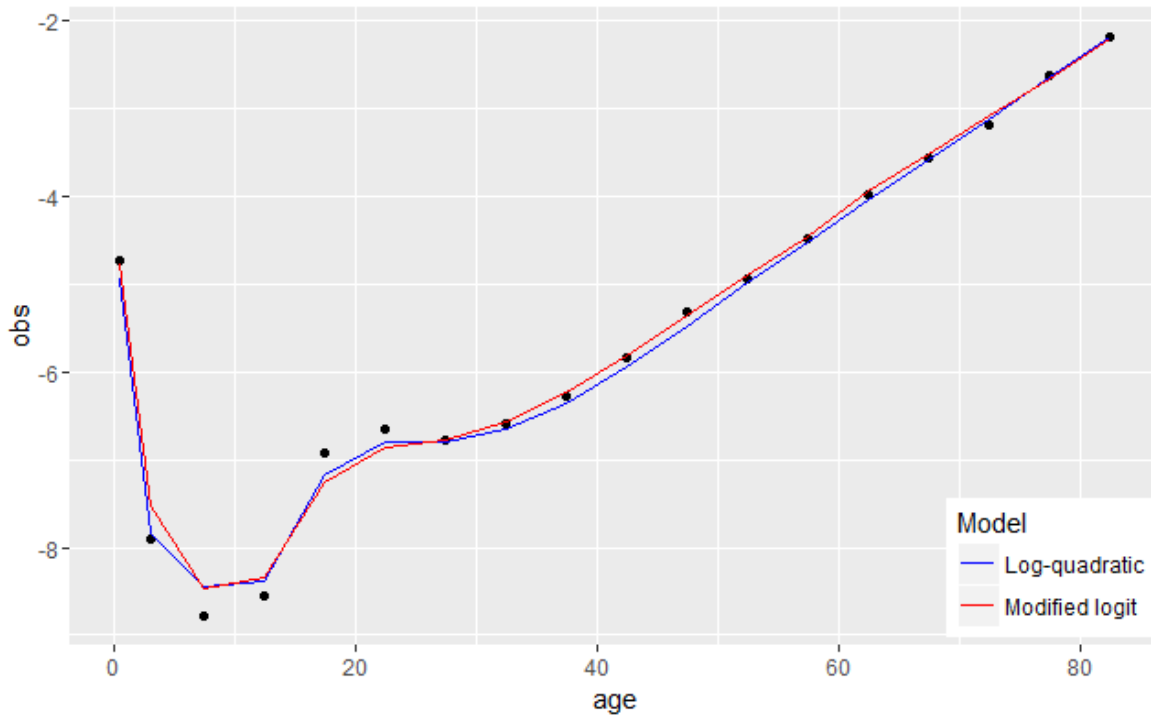
As you can see, the models have some difficulty following the rates for teenagers and young adults, but do pretty well at the very young ages, and even better after age 25. In my opinion, it would be very hard to get a better fit without adding a third parameter.

The code used to produce this graph is available in the course website, including functions to evaluate and fit the models.