The Lee-Carter Model

ECO 572/ SOC 532 • SPRING 2009

We review the Lee-Carter approach to forecasting mortality. This subject is not discussed in the textbook, but their 1992 JASA article is very clear. The most distinctive feature of their approach is the use of a stochastic process to model uncertainty about the future.

The Mortality Surface

Lee and Carter seek to summarize an age-period surface of log-mortality rates $\log m_{xt}$ in terms of vectors $\boldsymbol{a}$ and $\boldsymbol{b}$ along the age dimension and $\boldsymbol{k}$ along the time dimension such that

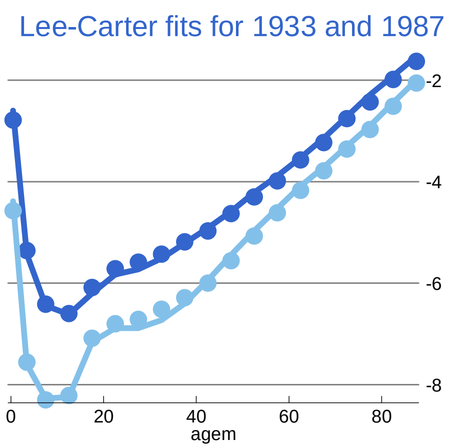
$$\log m_{xt} = a_x + b_x k_t + e_{xt}$$

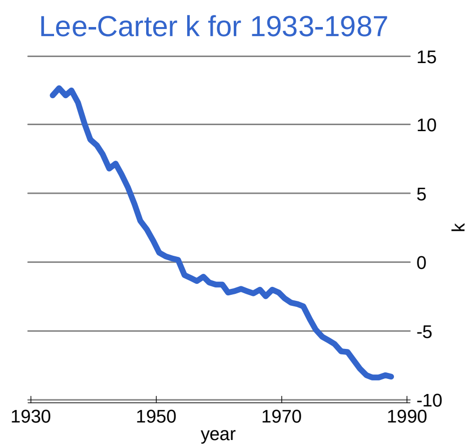
with restrictions such that the b's are normalized to sum to one and the k's sum to zero, so the a's are average log rates.

The vector $\boldsymbol{a}$ can be interpreted as an average age profile, the vector $\boldsymbol{k}$ tracks mortality changes over time, and the vector $\boldsymbol{b}$ determines how much each age group changes when k_t changes. When k_t is linear on time each age group changes at its own exponential rate, but this is not a requirement of the model. The error term reflects age-period effects not captured by the model.

Estimation using SVD

Lee and Carter estimated the a's, b's and k's with U.S. mortality data from 1933 to 1987 using least squares. Specifically, they estimate $\boldsymbol{a}$ by averaging log-rates over time and $\boldsymbol{b}$ and $\boldsymbol{k}$ via a singular value decomposition of the residuals, essentially a method for approximating a matrix as the product of two vectors. In a second step they adjust the k's so they predict the correct total number of deaths each year, but this step is not essential and I have skipped it. The basic data used in the original paper consisted of rates up to ages 80-84 and 85+.





The companion Stata handout shows that one can reproduce their calculations quite closely using data from the Human Mortality Database. The figure on the left shows how well the model fits U.S. mortality in 1933 and 1987, showing the familiar shape of mortality by age and larger relative declines at younger ages. The figure on the right shows the trajectory of k, with a steady decline over time.

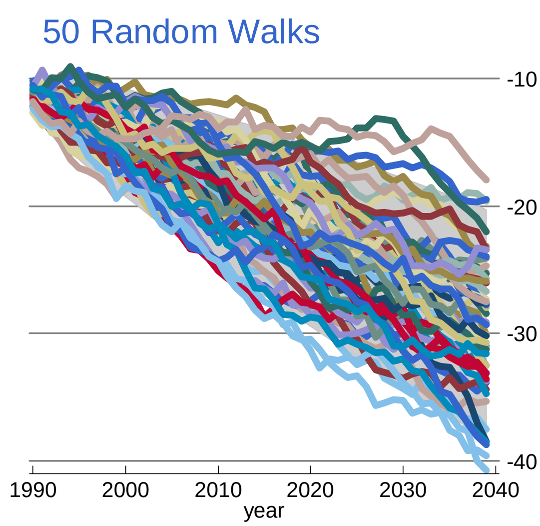
Because a large fraction of the U.S. population survives to age 80 (30% in 1987) Lee and Carter extended the model to older ages on the basis of work by Coale and Guo in 1989 and Coale and Kisker in 1990, showing that after age 80 mortality increases at a linearly declining rate, rather than the constant rate assumed in a Gompertz model. Through this extension the published $\boldsymbol{a}$ schedule goes up to age 105+.

The Time Series Model

The second distinguishing feature of the Lee-Carter approach is that, having reduced the time dimension of mortality to a single index k_t, they use statistical time series methods to model and forecast this index. In their application to U.S. mortality they discovered that, except for the flu epidemic of 1918, the index behaves like a simple random walk with drift, where

$$k_t = k_{t-1} + d + e_t$$

where d is the drift, estimated as -0.365, and the e_t are independent error terms with variance v, estimated as 0.652^2. Note that the k's are not independent; it is successive differences (or innovations) that are independent.



Also, the variance of k_t increases with the forecast horizon t, as you might expect. Using the law of iterated expectations it is easy to show that starting from a fixed value at time t_0, the variance of k_t is

$$\mathrm{var}(\mathrm{k_t}) = (\mathrm{t} - \mathrm{t_0})\mathrm{v}.$$

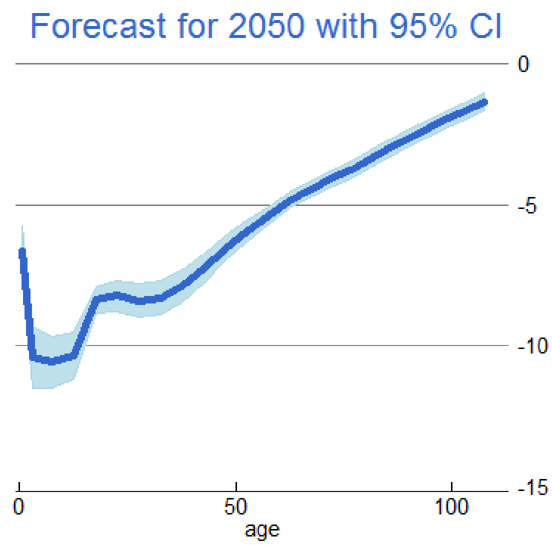
This is important because it gives us the standard error of a forecast.

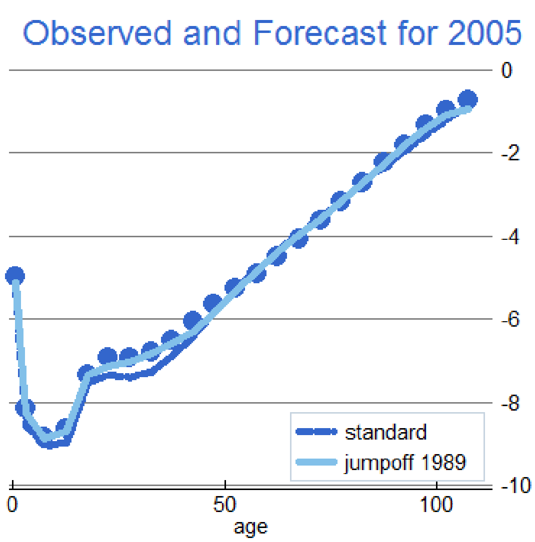
Simulating the Random Walk

Perhaps the best way to understand the stochastic nature of the projection is to do a bit of simulation. In the Stata log we set a 50-year horizon and generate 50 random trajectories, starting with a value of $k = -11.05$, which is my estimate of k_{1989}. The results are shown in the figure on the right. The key thing to note is how our uncertainty regarding the level of mortality increases as the projection horizon gets longer. The shaded area barely visible under the lines represents the 95% confidence region $k \pm 1.96\sqrt{(t - t_0)}0.652$

Forecasting Age-Specific Mortality

Once we have a forecast for k we combine it with the vectors $\boldsymbol{a}$ and $\boldsymbol{b}$ to produce a forecast of age-specific mortality. The next figure shows a forecast for 2050 using the published values of $\boldsymbol{a}$ and $\boldsymbol{b}$ as well as $k = -33.3$, which has a standard deviation of 5.08. (Can you reproduce these values?). The figure also shows a region bounded by the upper and lower 95% confidence bands. In this forecast I tried to reproduce the results in the Lee-Carter paper as closely a possible. Apparently they use the $\boldsymbol{a}$ schedule only up to age 80-84, and close the life table following a variation of the Coale-Guo procedure where the rate at 105-109 is about 0.72 and the slope declines linearly between 80-85 and 105-109. In reality the rates at 105-109 are not that high, and one gets simpler and better forecasts using the published $\boldsymbol{a}$ schedule for all ages.





If you were to try forecasting for a recent year and then compared predicted and observed rates you would discover that the forecast is quite good but a bit too optimistic at ages between 20 and 50. The reason is that the age pattern of mortality at the jump off in 1989 was a bit different from the average pattern in the 1900s. The solution is to use the 1989 rates instead of the vector $\boldsymbol{a}$ to reflect the new age pattern and reset k to zero keeping $\boldsymbol{b}$ unchanged. This produces much better results, as you can see from the figure on the left, which compares the rates observed in 2005 with forecasts made as of 1989 with the average $\boldsymbol{a}$ and the actual 1989 rates. At first Lee and Carter worried about giving too much weight to a single year, but eventually concluded that it was better to update the age-pattern at the jump off.

The next step in the forecast is to construct a full life table from the age-specific mortality rates. This can be done using standard techniques, so I'll skip the details. I find that the simple assumption of a constant risk in each age interval works well enough for most purposes. The forecast for 2050 has an expectation of life of 84.3 years with 95% confidence limits of 80.7 and 87.7 (in close agreement with Table 4 in the paper). The Social Security Administration (SSA) has their own data and they produce forecasts that are usually more pessimistic than the Lee-Carter estimates. For 2050 the SSA predicts 80.2 (with lower and upper bounds of 77.9 and 83.8). These differences were already apparent at the time the original paper was published and Lee and Carter comment on them. They express concern that the SSA will be unprepared for the high dependency ratios that will accompany life expectancy that substantially exceeds their forecasts.