Mortality forecasts are an essential input to any population projection, which in turn, form an essential component of long term planning with respect to pension programs, health care systems, and some long-term environmental applications. Mortality forecasts in themselves form an important part of many life-insurance and annuity systems. For these reasons, both research in and implementation of mortality forecast methods is highly significant.

Traditional mortality forecasts, usually either governmental or quasi governmental (for example the United Nations), have been based on expert opinion. However, such forecasts have systematically under-projected decreases in mortality, leading to concomitant under-projection in the elderly population (National Research Council 2000:132). In order to address both the possible biases inherent in expert opinion, and at the same time find a simpler approach (at least once the mathematical computer libraries are in place), Lee and Carter (1992) proposed a new method using standard methods for forecasting stochastic time series and a simple model for the age-time surface of the log of mortality. This method uses the singular value decomposition (SVD) to produce a forecast of the of the probability of distribution of future age specific death rates, which can be used generate probability distributions for any life table functional of interest. The important outputs from an application of the model to an empirical population include a_x, b_x, and k_t -- vectors which represent, respectively, the average (over time) of the log of the age-specific death rates, the magnitude and direction of the age specific change in death rates, and the magnitude of a change (away from a_x) at any given moment in time. Standard time series methods are used estimate k_t parameters; using these parameters, multiple k_t traces are simulated, from which projected death rates and life expectancies are derived.

This method originally was used to forecast age specific death rates (ASDRs) for a single population, but has since been extended in several directions. Lee (2000) presents a modification to forecast age specific death rates from male and females of the same population, using a k_t and b_x derived from the population-weighted combined death rates but a separately derived ax. Li, Lee and Tuljapukar (2004) modified the method to project mortality from populations with missing data, using a modification of the standard random walk with drift time series model (RDW) to estimate the drift and variance of k_t. Additionally, Li and Lee (2005) extended the method to include several new features for a coherent forecast -- deriving the b_x and k_t from a population weighted average of the ASDRs, applying the SVD to the residuals of each separate population's LC model to fit short-term variations in particular populations, and deriving several measures of the relative improvement in fit gained by application of these additional LC processes.

Although Girosi and King (forthcoming) write that LC is "the best available and widely used method forecasting method in the literature", certain modifications have been proposed, choosing historical periods with more consistent trends (Booth et al 2002), using recent ASDRs as the starting point for projections (Bell 1997), and others. Some of these criticisms have been incorporated into newer LC methods. In spite of these criticisms, and partly due to their incorporation into the LC framework, LC has generally performed better than official forecasts (Bell, 1997; Lee and Miller 2000). LC is now used by the US Census Bureau, and is also used by the U.S. Social Security office of the Actuary to complement their expert based forecasts. The United Nations has expressed interest in incorporating LC into their forecasts, and the coherent and missing data extensions described above are particularly useful to their application.

However, the LC remains somewhat esoteric. For any given application, it uses multiple inputs, generates multiple outputs, and requires several sub-processes that are typically coded as functions. Code that implements LC typically uses relatively complex sets of intermediate variables. A Matlab(tm) program that runs single population LC will contain over 100 lines of code, take as input a file with ASDRs, and require the specification of several parameters, typically by editing the code directly -- hardly an intuitive method. Because of its relative complexity, compared say to standard life-table methods (which can be run on a simple spreadsheet), as well as its lack of wide availability in an easy to run format, the LC method can be hard to learn or teach. Furthermore, some of the available programs which purport to do LC are poorly documented, hard to understand, and may not have been thoroughly tested.

In order to facilitate the application of LC by interested researchers and provide a reference implementation, we have implemented a web-based version of LC implemented using forms. This application provides the four methods outlined above -- single population LC, two-sex LC, coherent LC, and missing data LC. To run the application requires only a web browser, an internet connection, and the ability to cut-and-paste ASDRs into forms.

It is probably easiest to give an impression of the system by describing a typical session. The paradigm for working with the system is Input - List - View: first, the user inputs ASDRs into a web-form, clicking "Run" when finished; then the user lists all of the previous LC runs performed, with the most recent on top; finally, the user views the run just performed by clicking on its link. A view of a run displays the most salient statistics, including the a_x, b_x, and k_t vectors. Four plots are also generated: one, the a_x, b_x, and k_t for the empirical data; two, plots of the forecasted k_t's and life-expectancies; three, time-series of the log empirical mortality rates; and four, log mortality profiles of the initial empirical ASDRs, the end empirical ASDRs, and the ASDRs at the end of the forecast. The user navigates between the various screens -- input, list, view -- by accessing a set of options that remains in the navigation bar at the left-hand side. Additionally, the user is assigned a specific username - password combination and can only see her own runs.

The software is built using a combination of free software, including the Linux operating system, the Apache web-server, and the Postgresql relational database system. It is written in Python, using the scipy scientific library. Interestingly, we have relied heavily on the object-oriented features of Python for organizing and storing the multiple variables involved in each LC run. All of the information in a run -- typically at least twenty five separate variables -- is "pickled" and inserted into the relational database with information to connect it to a specific run identifier and a specific user; later, when the user wants to display that information, the application "un-pickles" the object (found by its identifier, an integer) and retrieves the data for display. Without the object oriented capabilities, organizing and storing a run would have been much more difficult. In general, the free software used to build this application has been completely suitable.

LAR-E has been tested by using it to duplicate the data in the various papers cited with respect to their methodologies. It will soon be beta-tested by students and researchers within the Department of Demography at UC Berkeley and other affiliated researchers at Mountain View Research and the United Nations. In the future we hope to expand its functionality and open it to an open source process.

References

1. BELL, W. R. 1997. Comparing and Assessin Time Series Methods for Forecasting Age Specific Demographic Rates. Journal of Official Statistics. 13 pp 279-303.

2. BOOTH, H., J. MAINDONALD, and L. SMITH. 2002. Applying Lee-Carter Under Conditions of Variable Mortality Decline. Population Studies. 56(3).

3. GIROSI, Frederico and Gary KING. Forthcoming. Demographic Forecasting. Princeton, New Jersey: Princeton University Press.

4. HORIUCHI, S. and J. WILMOTH. The Aging of Mortality Decline.

5. LEE, R. D. 2000. The Lee-Carter Method for Forecasting Mortality, with various Extensions and Applications. North American Actuarial Journal. 4(1), pp 80-91.

6. LEE, R. D. and L. CARTER. 1992. Modeling and Forecasting the Time Series of U.S. Mortality. Journal of the American Statistical Association. 87(419), pp 659-671.

7. LEE, R. and T. MILLER. 2001. Evaluating the Performance of the Lee-Carter Method for Forecasting Mortality. Demography. 38 pp 537-549.

8. LI, N. and R. D. LEE. 2005. Coherent Mortality Forecasts for a Group of Populations: An Extension of the Lee-Carter Method. Demography. 42(3), pp 575-594.

9. LI, N., R. D. LEE, and S. TULJAPURKAR. 2004. Using the Lee-Carter Method to Forecast Mortality for Populations with Limited Data. International Statistical Review. 72(1), pp 19-36.

10. National Research Council. 2000. Beyond Six Billion: Forecasting the World's Population. J. Bongaarts and R.A. Butatao. Washington, D.C.: National Academy Press.