Session 2B: Mortality Modeling I—Modified Lee-Carter Methods

Discussant: Ward Kingkade

Presented at the Living to 100 Symposium
Orlando, Fla.

January 5-7, 2011

Liu and Yu, Assessing and extending the Lee_Carter model for long-term mortality prediction

The paper by Liu and Yu documents a Herculean effort by the authors. Perhaps it should be several papers. A great deal of research went into it, and it makes several important contributions. It contains assessments by rolling backtesting of the fit of Lee-Carter time series models real data. It also includes Monte Carlo simulation of the fit of the Lee-Carter model in the presence or absence of non-homogenous disturbances underlying data. Part of the analysis addresses the concept of drift uncertainty, which I suppose can be viewed as assessing the fit of the Lee-Carter model when the drift coefficient in the time-dependent K parameter has a disturbance to it. This is a potential contributor to the standard error of estimate that is not incorporated in the original Lee-Carter model. The analysis employs robust statistical methodology in the form of Quantile Regression which is really a big help in dealing with some of these problems. I should also point out that the statistical methods in the analysis involve some pretty advanced mathematical statistics: Skorokhod spaces, Brownian bridges, and the Kolmogorov-Smirnov statistic, to mention a few. Thus, this paper can be viewed as a very good example of a practical application in applied mathematics.

The research comes up with several overall noteworthy findings. The one I want to stress is the one that is usually missed in most other analyses, namely that the authors find evidence of real structural change in the process that drives mortality dynamics. With the Lee-Carter methodology, we're assuming that the parameters A, B and K are constant over the period for which they are estimated, as well as the ensuing projection period[1]. Originally Lee and Carter (1992) fit their model to roughly a century of US data. That might be fine if what you want to do is take a mechanical approach to population projection and you need something simple, but it's a rather strong, and therefore questionable, assumption to envision that the process that drove mortality change in the 1980s is the same process that was driving mortality change at the beginning of the 20th century[2].

The authors find that for males, the K parameter, which is the component that represents the trend in mortality over time, is curved for males. That means basically that the model that's holding in one period is not holding in another. The curved trajectory means there's a structural change such that k departs

---

[1] As originally introduced, the Lee-Carter model takes the form

$$\ln(m_{x,t}) = a_x + b_x k_t \ ,$$

where x indexes age, t indexes time, a is a vector indicating the benchmark level of the mortality schedule, b is a vector indicating the pattern of relative change at the various ages, and k is a stochastic process involving an error term. Lee and Carter advanced a model in which k is a random walk with drift, that is

$$k_{t+n} = k_t + cn + e_{t+n},$$

in which the drift coefficient, c, is constant for equally spaced t, and e is a vector of serially independent, identically distributed, random disturbances with a mean of zero. In other words, k is taken to be a straight line with a slope of n subject to random disturbances whose average is zero.

It should be mentioned that Lee and various associates have introduced a number of modifications of the model in the original (1992) article.

[2] Reductions in mortality from infectious diseases, due in part to the development of vaccines and modern antibiotics, had much to do with the reduction in US mortality around the beginning of the 20th Century. Towards the end of the Century the major causes of death were degenerative diseases such as heart disease and cancer.

from a straight line.

The authors find that males are harder to predict than women, which I suppose is not really that original. They also find that real data are much harder to fit, especially among males, than are simulated misbehaving data. They find evidence of drift uncertainty. Most of the examples in the exposition feature Swedish data from the human mortality database, but the authors in the paper looked at several other countries and the results are similar. The authors make the very well taken point that when you're dealing with structural changes, models estimated over shorter base periods are going to perform better because they're going to correspond to the process that's going on currently and they lead to better forecasts in the short run, but it makes long-run predictions riskier. I'm not sure it necessarily obviates them because you can still use them as a "what if" scenario in which the current process goes on indefinitely, but it's a valid point that models estimated over a shorter period are very often used in practical applications and to extrapolate them too far is risky.

In the analysis, the authors are using life expectancy *at birth* as a measure of the overall mortality level. It could be argued that this overemphasizes infant mortality, but what I like about it is that it covers the whole age range, which is the same as the Lee-Carter model. It should be noted that at late age, Lee and Carter in their original article used a model life table, which many researchers would hesitate to do[3]. In any

[3] This symposium features many analyses and models of the age-pattern of mortality at late age, in the tradition of the late Roger Thatcher.

case, it’s good to have something that covers the whole age range.

The authors come up with some curious findings, one of which is that in their simulations they consistently get a negative estimation error. What’s most interesting about that is that they structured the simulations themselves, so this shows that there’s something in the way the authors are thinking about what’s going on in the mortality data and fitting models in such a fashion that embodies some sort of a constraining view and I would encourage the authors to pursue this.

The analysis demonstrates that “robust” statistical methodology can really help resolve a lot of problems. They have their Case 2 which is where the underlying data have a certain disturbance that’s subject to episodic shocks like that 1918 flu epidemic or the SARS outbreak. Quantile regression, which is a robust method, really helps with that and it’s been pointed out that when you lose the properties required for certain statistical tests, under this approach you can still go on with the Kolmogorov-Smirnov statistic and retrieve meaningful findings out of the data. This is definitely a contribution to this research area. Unfortunately, quantile regression, along with any other statistical methodology I am familiar with, can’t really fix the problem of structural change in the model, but that’s the nature of the beast[4].

<u>Yang, Yue, and Yeh, Coherent Mortality Modeling for a Group of</u>

[4] To deal with structural change in the process, at a very minimum, nonlinear trends and/or piecewise discontinuities in the k parameter need to be entertained in the model.

Populations

Turning to the coherent modeling paper by Yang, Yue and Yeh, we have an analysis that addresses the problem of making Lee-Carter projections for a group of populations, and in projection work, you are usually in this situation. The coherent group could be a group of countries, for example, and places like the U.N. and the Census Bureau do make multi-country projections, but it also arises inevitably when you're modeling mortality with two sexes. You can easily end up if you're not careful in situations where your female projected death rates are higher than the male ones or an equally awkward situation, and constraining the Lee-Carter temporal change parameter and the B parameter while letting the sexes have a separate age profile has a real advantage there; it's very workable. I should point out that when you're able to pool data it often leads to more stable results, so there's a lot to be said for this approach.

Now most of the exposition takes the U.S. and Canada as an example, but the authors have also presented data for Taiwan and Japan. Their results seem to show that pooling countries for the same gender often works out better than pooling genders within countries and that's not surprising because the male/female differential is one of the most universal regularities in human morality. The authors point out that they get this result especially for the conventional Lee-Carter model with only age and period effects in it. They use a variety of measures for assessing the goodness of fit. They have mentioned some, but they also use the Akaike information criterion for time series and they examine log likelihoods.

The authors also examine the Lee-Carter model with a cohort effect, and in making this assumption they seem to be assuming a random walk with drift for the cohort effect. I should point out that unlike Lee and Carter (1992), the authors fit their models by maximum likelihood, which is not robust to misspecification. This methodology requires you to specify the exact form of the model. As implemented by the authors, the model's time-varying parameter is constrained to be a random walk with drift, whereas in the original Lee-Carter analysis of US mortality, this property was an empirical finding. It isn't clear how the authors specified the cohort parameter, but it seems to me that it's a random walk with drift as well.

In the analysis, the cohort model often gives contradictory results. Before getting into them, I should point out that in a lot of demographic applications that I've seen and been involved with, the identification of models with an age and a period and a cohort effect is not a simple thing because the cohort parameters are not independent when you have age and period effects in play at the same time. A cohort is defined as being a conjunction of age and period, so you're by definition not dealing with independent effects. With maximum likelihood, you can always put more terms into the model, and if the fitting algorithm converges, you will always get a better fit with more parameters in the kind of hierarchical model the authors are estimating. The authors compare performance with and without cohort terms and, unsurprisingly, they do find that you get a better fit with the cohort parameters in there.

What is often done to determine the importance of parameters estimated by maximum likelihood is to compare the likelihoods of the model with and without the parameters in question, keeping in mind the degrees of freedom lost in incorporating these parameters. The difference in log-likelihoods leads to a Chi square test statistic for the statistical significance of the parameters in question. Such tests would be of interest here.

The cohort parameters are difficult to interpret and no explanation is offered in the paper to indicate what it is about people born in the given year that causes them to exhibit a particular mortality pattern beyond what is attributed to time and age. This may not be needed for the purpose of simple projection, but it leaves me wondering what the cohort effect is. Most of the effects judging from the graphs are characterized by sort of a gradual trend and some of them have jerks in them, so something is certainly being captured by the cohort parameters. What I find disquieting is that the cohort models also yield odd effects. For example, it's noted in the paper that the Lee-Carter B, that is the amount by age at which mortality changes as the level changes, the age pattern of mortality changes. This rises with increasing age in the cohort model and that's seems unusual to me. It may have an explanation, and I would be interested to know more about it.

To its credit, this analysis includes an annuity product and it makes the claim that the cohort model gives you narrower confidence bands around its values. I hadn't noticed this that until I saw the authors' slide, where it is clear.

Another comment I have is that the focus on ages 65 and over, makes a kind of sense in a symposium having to do with living to 100 and beyond, but in a real projection you need the whole age range, and so there always remains the question of how do you connect the mortality projection for these ages to mortality at younger ages. Lee and Carter (1992) did not model late separately. They had one A vector covering the whole age range.

Also, I'm wondering why the authors use data in 5-year age intervals from the human mortality database for Canada and the U.S. when single-year age data are available from the National Statistics offices of both countries. What's more alarming is that earlier in the day of this session Leonid Gavrilov said in a presentation that he and his associates had used the human mortality database data for the US and they got results that seemed too smooth, so they called back to the people that maintain the human mortality database asking what accounts for this smoothness and found out that the data they had obtained from the HMD were graduated using a logistic curve. Dr. Gavrilov made the point that in approaching the HMD, you have to actually make sure from them that you're getting the raw data with all the kinks and the fur in it. I don't know this first hand. It's something I heard in an earlier presentation. I was surprised at the time I heard it.

Finally, the authors make the point that a lot depends on which countries you select for a coherent group and that's definitely the case. For some groupings of countries pooling countries within genders is better than pooling genders within countries; for other groupings the reverse is true. The authors make the very well taken point that in many instances geographic proximity is not paralleled by similarity in mortality patterns and trends. I quite agree with the authors that selecting coherent groups is a delicate exercise.

Doray and Tang, Projection of mortality rates at advanced ages in Canada with a new Lee-Carter type model

The paper by Doray and Tang on projecting mortality rates at advanced age with a new Lee-Carter type model starts with a description of old age mortality models more or less like what you have in the Kannisto-Thatcher papers, and they come forward favoring the Kannisto logistic model. They state that in Canada you don't get sex convergence with the Lee-Carter models and they maintain that's a good thing. They fit their models to males ages 70 to 99 and females ages 80 to 105, and they use the Canadian National database with single year data. They note some problems in the original Lee-Carter model, including the question of the temporal and variability of the Lee-Carter parameters (the A's and the B's). The authors point out that in subnational projections under the model one has to incorporate widening differences in mortality between sexes, or else crossovers in mortality among regions.

The authors' methodology rather strictly follows the approach presented in the original Lee-Carter article (1992). They do a singular value decomposition to come up with A, B, and K parameters, then they fit time series models to the K parameters by the Box-Jenkins methodology. They come up with ARIMA models that are not a random walk with drift. They have a model for males with two levels of integration, which means you're dealing with second differences. Models of this kind can be problematic when used in projection of mortality[5]. The authors model the force of mortality estimated three ways: 1) as the logarithm of the life table probability of dying; 2) as the logit of the life table probability of dying; and 3) as the logit of minus the logarithm of life table survival probability, which is an estimate of the force of mortality at the age in question . I suppose the reason for using the logarithms of q_x is an effort to be similar to the original Lee-Carter formulation, which was applied to logarithms of m_x. The logits of the probability of dying offer the advantage of ensuring that q_x remains bounded by zero and one. The model for the logit of the force of mortality is not used to represent the age-pattern of mortality at advanced age in the tradition of Kannisto and Thatcher. Rather, what the authors have in mind is that the the age-specific forces of mortality follow logistic trajectories over time. The authors end up preferring the latter of the three models.

The authors find some odd results. They find that with males at late age their projections are giving you increases in mortality over time and they attribute this to small populations, presumably implying unreliable data, at those ages. The authors also find that among women the mortality rate and the probability of dying at age 100 rises over time. They find that different ARIMA models fit men and women. They find no support for convergence in the mortality patterns of the two sexes, and suggest a coherent group approach to address the

[5] Perhaps for this reason Lee and Carter in their original (1992) article preferred the random walk with drift to more elaborate models for their temporal parameter (K), which they stated they had found some support for.

problem of divergence of the sexes.

The other point that I should make is that again this is another analysis that doesn’t show you how mortality at late age connects to mortality at younger ages, but that’s not necessarily a fault in a seminar concerned with mortality at late age.

Citation

Lee, R.D. and L.R. Carter, “Modeling and Forecasting U.S. Mortality”, *Journal of the American Statistical Association*, 87(419), September 1992, pp. 659-671.