Alternative Time-Series Approaches to Mortality Projection

Ward Kingkade
Population Projections Branch
Bureau of the Census

Extended Abstract

Lee and Carter (1992) analyzed a historical time series of annual age-specific mortality rates for both sexes and all races combined,1900-1987. The data were analyzed in conventional aggregate age detail (ages 0, 1-4, 5-9, …). They considered models of the form:

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

where x indexes age and k_t is a stochastic process in time. In their (1992) analysis the authors estimated the model parameters by applying the Singular Value Decomposition to their data matrix, in which they subtracted the age-specific mean logarithms of the death rates from each annual log age-specific death rate, so that a_x becomes the vector of mean logarithms, b_x is the principal eigenvector, and k_t is the first principal component. The authors argued that k_t could be adequately described as a simple random walk with drift, although some results from their time series analysis indicated that a moving average term should be incorporated.

In our case, we distinguish death rates by sex and race/origin in 3 categories: Hispanic, Black Non-Hispanic, and Non-Black Non-Hispanic. Our data consist of the annual deaths in single-year age detail by sex and the race/origin categories from the NCHS annual mortality databases, together with US Census Bureau estimates of the population in the same race/sex detail, by single year of age thru 100. Because US states did not begin to code Hispanic origin from death certificates prior to 1984, this left us with the years 1984-1999. For ages over 80 we eschewed the Lee-Carter formulation, in which b_x is the same in every age group, in favor of Kannisto's (1998) 2-parameter model of the force of mortality. For ages under 80, we investigated the Lee-Carter model.

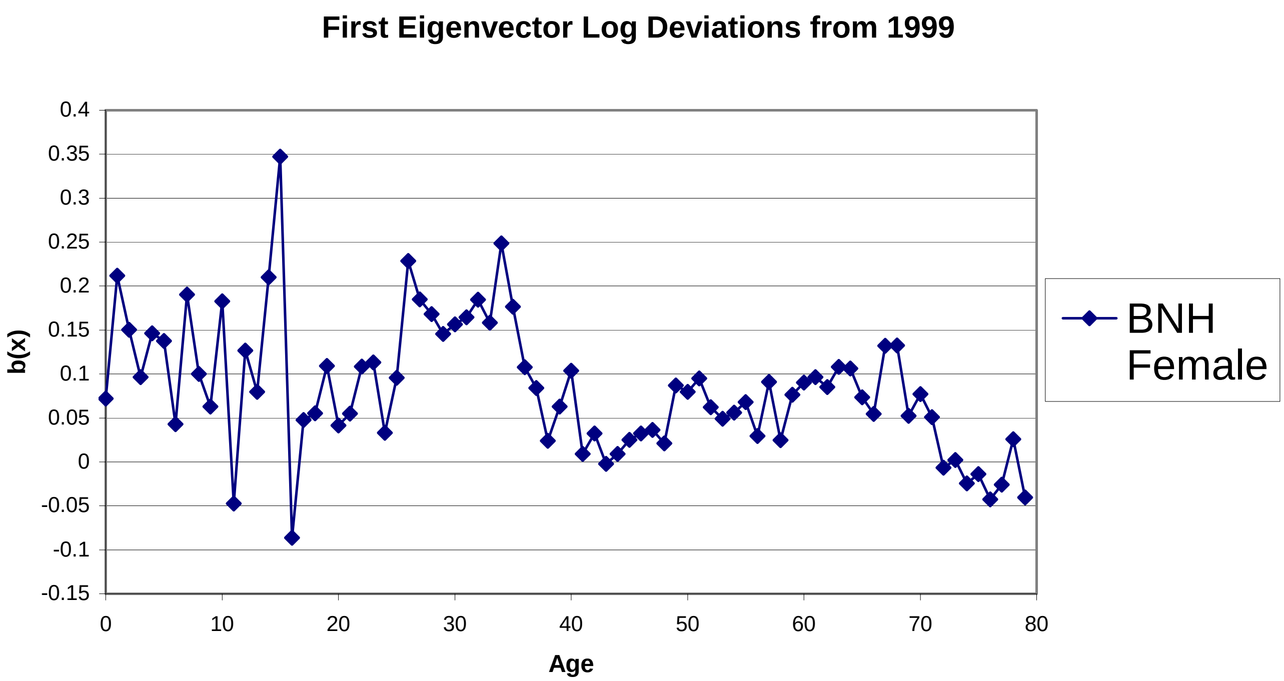
According to Lee/Carter, k_t indicates the level of mortality at a given time, while b_x indicates how rapidly or slowly the respective age-specific rates change with respect to k_t. Thus b_x serves as a "standard pattern" of mortality change associated with changes in the level parameter.

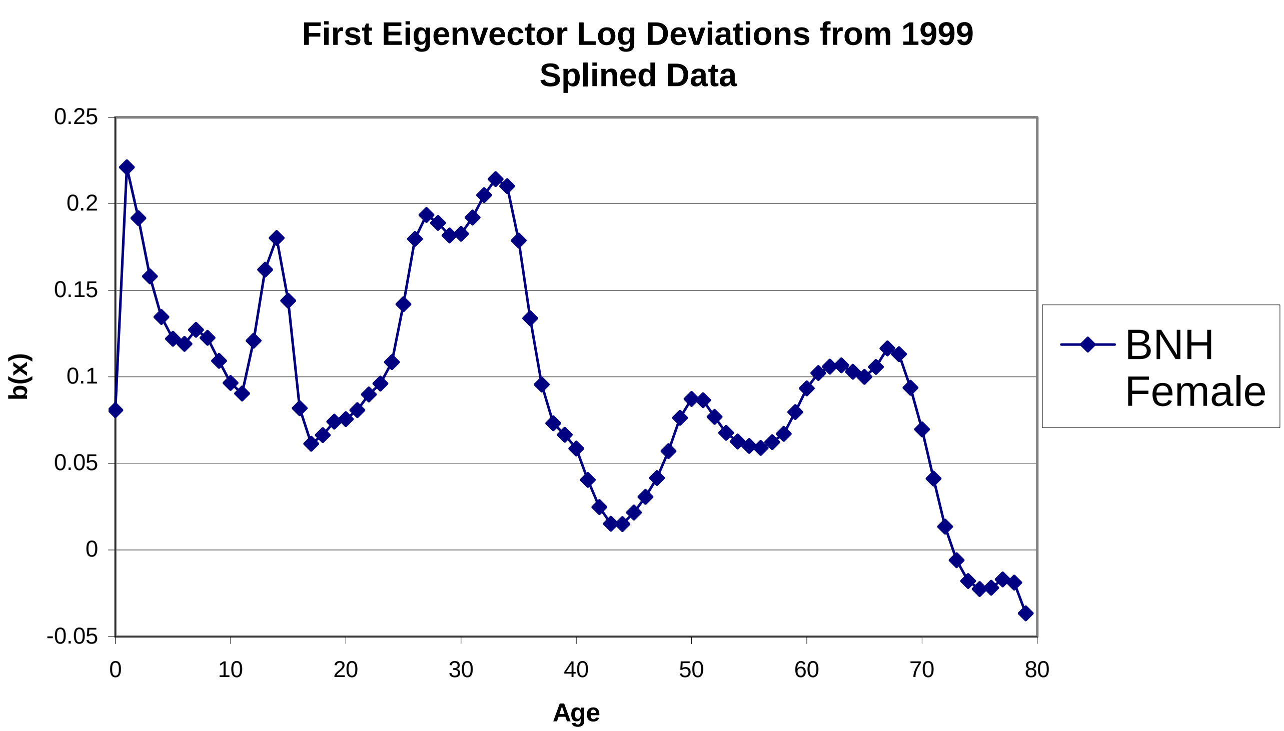
When our data are put through the SVD, the first principal components exhibit clear trends which differ radically by race/origin but are very similar across the two sexes within race/origin categories. Among Non-Black Non-Hispanics, who comprise the majority of the population, the trend in k_t is dominated by a clear decline. Among Hispanic men and women, a decline of this nature doesn't become apparent until the 1990s. Among Non-Hispanic Blacks there is a graphic increase in k_t thru the early 1990s, followed by a precipitous decline. It should be noted that, according to NCHS publications, life expectancy did, in fact, decline in the Black Population of the United States in the 1980s, after which a steady increase set in.

The principal eigenvectors emerging from the SVD of our rates also differ by race/origin. To some extent these differences are obscured by a substantial noise component reminiscent of such phenomena as age misreporting and differences in recording and imputation procedures between statistical agencies (e.g. the Bureau of the Census and NCHS). However real these phenomena may be, this is not the type of stochastic uncertainty that we propose to incorporate in our projections.

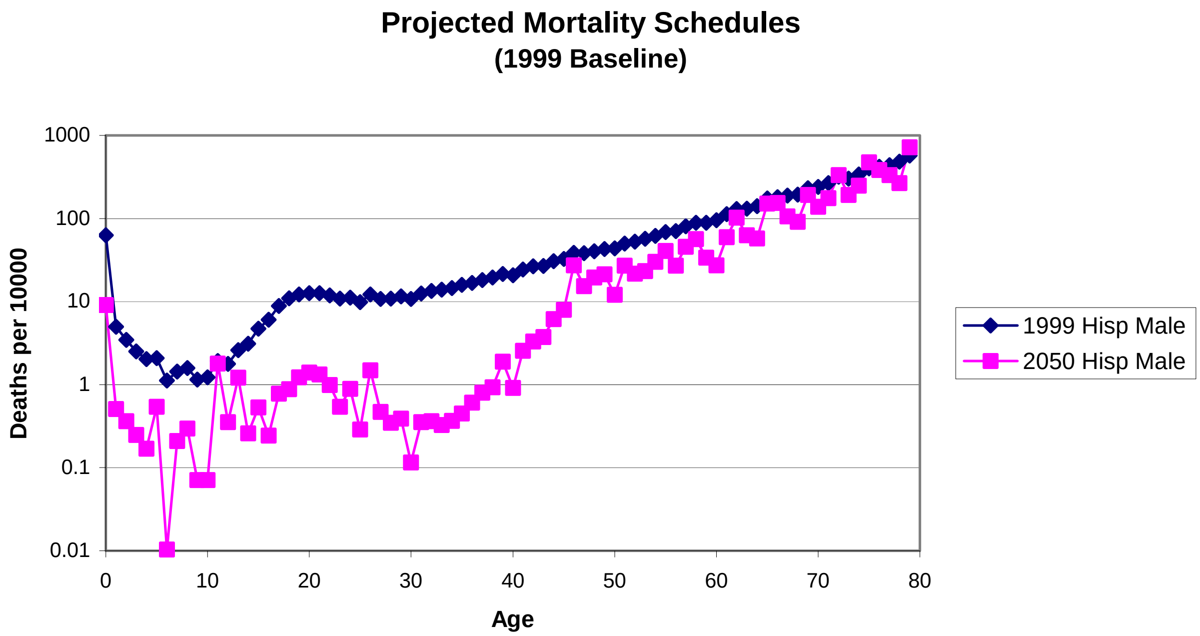
This kind of irregularity in annual mortality statistics has given rise to a multiplicity of actuarial graduation procedures, many of them under the heading of “Osculatory Interpolation” (Shryock and Siegel, 1975). Essentially, these procedures effect a piecewise polynomial smoothing of an array of rates with continuity conditions on derivatives at the points where the pieces join. Life tables are constructed not from the raw annual age-specific death rates, but using their smoothed values instead, on the notion that there is an underlying general mortality function smoother than any observed schedule of annual death rates. The modern successor to these graduation procedures is the smoothing spline (de Boor, 2001).

We found that applying a cubic smoothing spline with a resolution of 5 years to the age-specific death rates prior to taking their logarithms yielded more appealing results. The eigenvectors are more readily interpreted and there seems to be less risk of projecting measurement errors into the future (the procedure had little effect on the principal components, other than to amplify the magnitudes of their elements, apparently at the expense of the magnitudes of eigenvector elements). The two following figures exemplify our findings. It should be noted that instead of subtracting the age-specific means of the logarithms of each age-specific death rate, we subtract the logarithms of the 1999 death rates (unsmoothed in the first example, smoothed as described in the second).

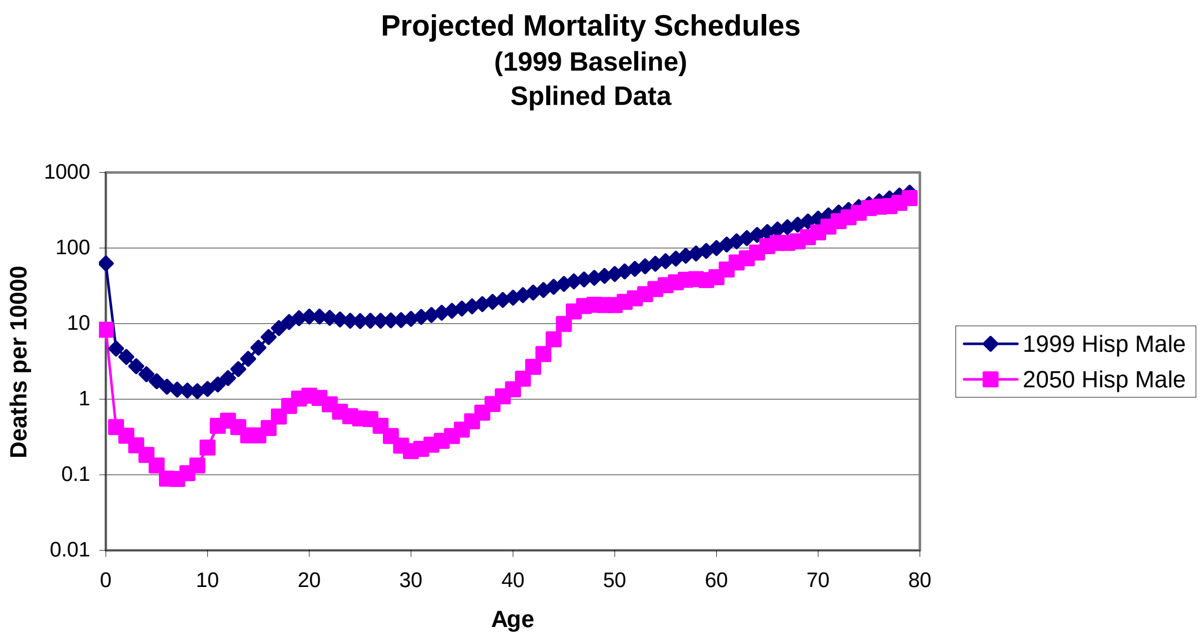

First Eigenvector Log Deviations from 1999
b(x)
Age
0.4
0.35
0.3
0.25
0.2
0.15
0.1
0.05
0
-0.05
-0.1
-0.15
0
10
20
30
40
50
60
70
80
BNH Female


First Eigenvector Log Deviations from 1999
Splined Data
b(x)
Age
0.25
0.2
0.15
0.1
0.05
0
-0.05
0
10
20
30
40
50
60
70
80
BNH Female

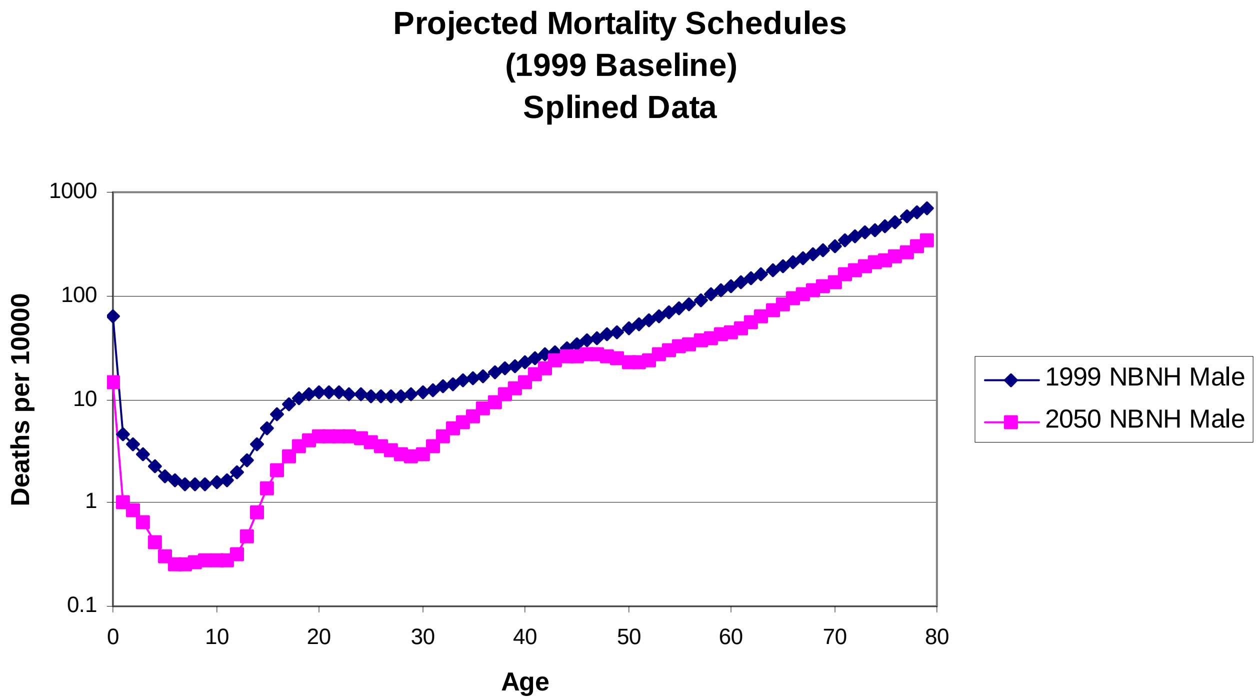
The following graphs illustrate the impact the smoothing has on projected mortality schedules. For variety's sake we examine Hispanic Males. For this simple exercise, we extrapolated k_t by fitting a trend line to the portion of the curve in the 1990s. As the graph below reveals, the noise in the unsplined data leads to much greater noise in the projected series than in its 1999 baseline.



The alternating pattern typical of age misreporting errors is unacceptable in a population projection. This feature is absent from the results obtained by smoothing the death rates prior to taking their logarithms. Nonetheless, more undulations are present in the projected series than the 1999 baseline. The magnitudes of the projected mortality declines also defy credibility.



The following graph illustrates the "best case" population, namely Non-Black Non Hispanics, among whom measurement errors are least pronounced and for whom k_t exhibits an almost unbroken linear decline. Unfortunately, the results obtained from the smoothed data leave much to be desired. The trough projected at age 30 seems exaggerated for a population in which homicide is not a mass phenomenon. The decline in projected age-specific mortality rates from ages 45 to 50 is baffling, as is the lack of progress around age 45. Finally, the radical declines in mortality in youth are entirely implausible. This kind of progress can be sustained over a 15-year period, and is reflected to a substantial degree in infant mortality statistics in NCHS publications for the period since 1984. But declines at such dramatic rates are seldom (if ever) sustained over durations such as 50 years.

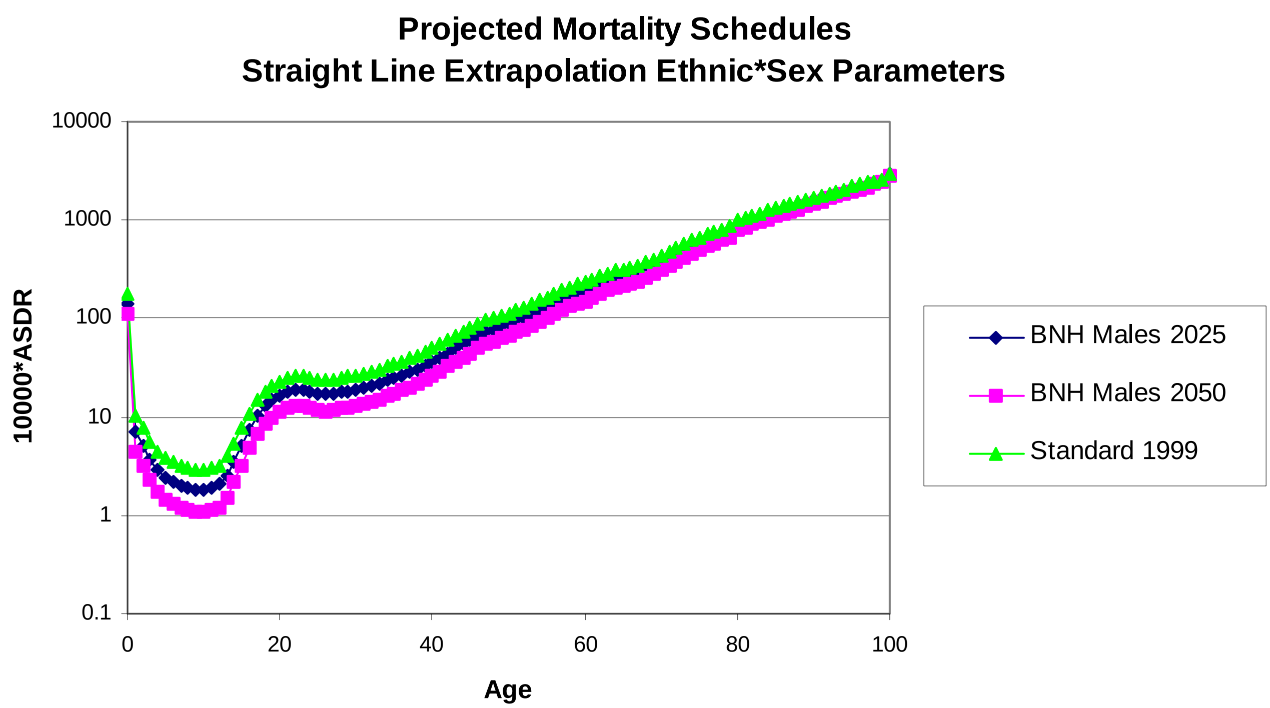


Given these highly aberrant results, we must conclude that a direct adoption of the Lee-Carter methodology, with or without smoothing, is not appropriate for population projections based on the available data for these race/ethnic categories. Although more aggressive smoothing might remove the undulations appearing in the projected mortality schedules, there is no apparent reason to expect the undue quantitative declines to diminish to plausible levels. Moreover, a successful, readily automated smoother would have to be so drastic as to have little or no advantage over the familiar and straightforward relational models long and well known in Demography.

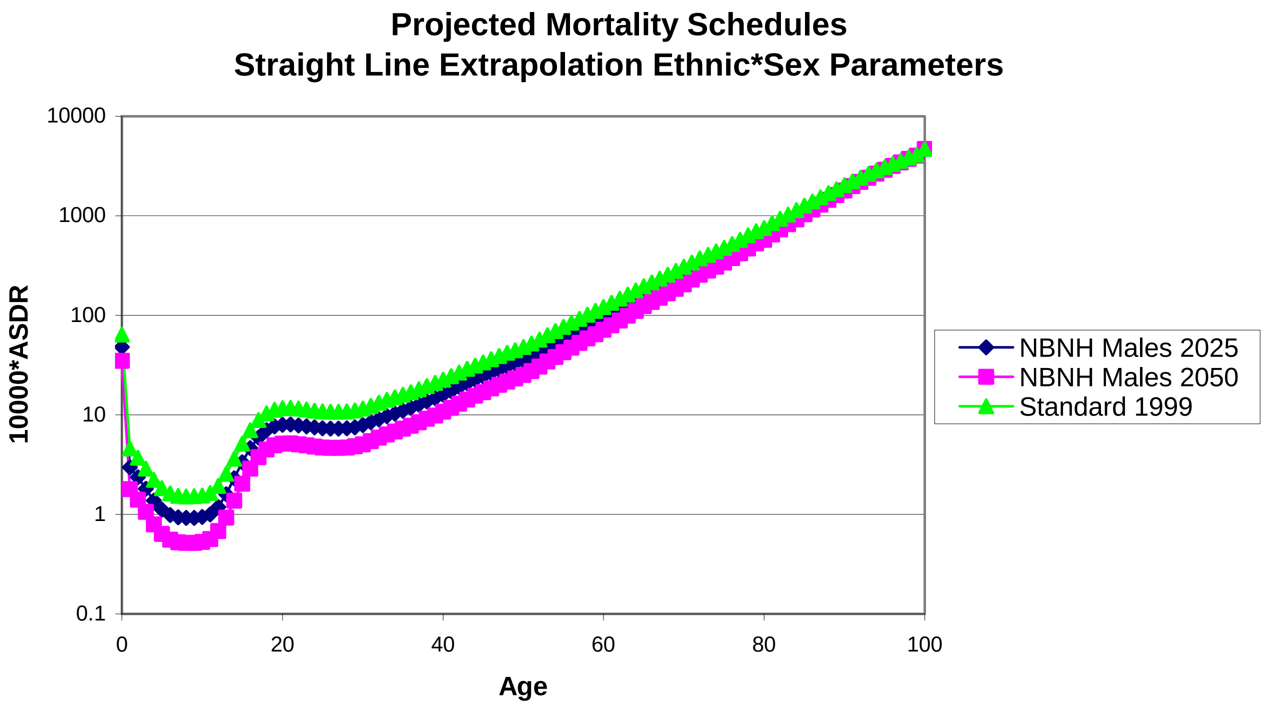
The most familiar relational model in mortality analysis is Brass' logit formulation

$$\lambda(x) = \ln(\frac{m_x}{1-m_x}) = \alpha + \beta \cdot \lambda_s(x) \quad ,$$

where λ_s refers to the logit of a standard mortality schedule. In our analysis, taking as standards the spline-smoothed 1999 mortality schedules of each ethnic/sex category, we obtain much more appealing results with this approach. The following figures were based on linear extrapolations of the 1984-1999 time series of the two parameters α and β. The subpopulations include an "awkward" case, namely Black Non-Hispanic Males, and an "easy" case, their Non-Black Non-Hispanic counterparts. The schedules for Black Non-Hispanic Males feature the hump in the



early 20s associated with accidents and injuries (including violent death), but contain no further pronounced undulations. Moreover, the projected decline in death rates is much more modest than those in the projections obtained from the Singular Value Decomposition. The same is true of the results for Non-Black Non-Hispanic Males, as well as the other subpopulations not featured here.


Projected Mortality Schedules
Straight Line Extrapolation Ethnic*Sex Parameters
10000*ASDR
10000
1000
100
10
1
0.1
0
20
40
60
80
100
Age
NBNH Males 2025
NBNH Males 2050
Standard 1999