Forecasting US Mortality

Neal Marquez

Modeling Mortality

Modeling human mortality has had a long history of attempts and revisions since Gompertz first made his claim of human mortality patterns in the 1880s[1]. The ability to accurately model and describe rates of mortality can have numerous effects on how social institutions and policies can be structured in order to meet the needs of its populations. For example knowing the age pattern of mortality allows a society to build the medical infrastructure to cope with the needs of different aged individuals. Knowing the root causes of mortality allow for the change of policy in order to reduce specific drivers of mortality.

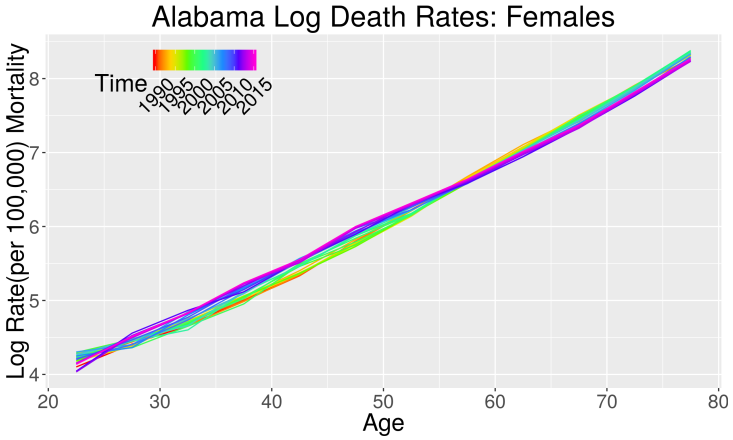
Another aspect of mortality that is often of concern is the future rates and how they differ from the past. Over the past century there has been dramatic declines in mortality across all locations and ages[1]. While that progress has stagnated to some degree we still see dramatic decreases in mortality among developing countries where medical infrastructure is expanding and improving. As this happens we would expect there to be a direct effects on the population, production, and the needs of a society.

History of Descriptive Models

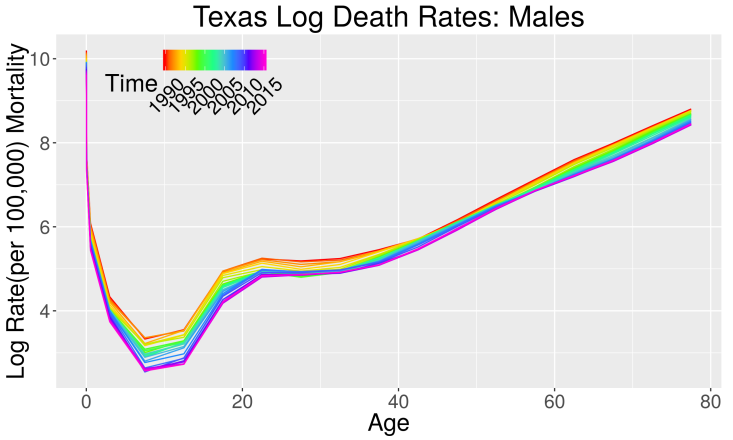
Descriptive models of mortality have an inherent age structure to them that allow for a simple descriptive explanation of how death effects individuals across their life-span. Gompertz first claim about human mortality was that there is a linear increase in log rate mortality as one aged[1,3]. This model follows the form:

$$ln(m(x)) = ax + b$$

where m_x is the mortality rate for an individual of age x, a is the rate of increase in log rate mortality by increasing one unit of age, usually years, and b is the baseline mortality rate that you would expect to see at birth. Below is an example of this phenomenon where we graph the log rate mortality of women in Alabama between the ages of 22 and 77.



While this model holds well for older ages, henceforth referred to as the senescence group or period, younger individuals experience quiet different mortality patterns as they age. From birth the rate of mortality decreases exponentially until the senescence period is started at which time the mortality rates will then again increase. Below is the log rate mortality for males in Texas which shows this pattern that is visible not just within the United States but nearly all countries since written records of mortality have allowed for demographic analysis.



In order to describe this phenomenon in 1983 Siler proposed a model to capture this switch in mortality patterns[3] by decomposing rate of mortality into three terms, an infant mortality term, a constant risk term and a senescence term. Siler had originally formulated the model in terms of a survivorship $l(x)$, that is the equation estimated the percentage of the population that would live to some age x

$$l(x) = exp\bigg(-\frac{a_1}{b_1}\big(1 - exp(-b_1 x)\big)\bigg) exp(-a_2 x) exp\bigg(\frac{a_3}{b_3}\big(1 - exp(b_3 x)\big)\bigg)$$

If we assume that x represent discrete age groups then with $l(x)$ we may calculate $S(x)$ which is the survival rate or the rate that individuals survive to age group $x+1$ given that they have lived to age group x as well as $m(x)$ which is the mortality rate.

$$S(x) = \frac{l(x+1)}{l(x)}$$

if $S(x)$ is per person then it is true that

$$S(x) = 1 - m(x)$$

and it can be found that $m(x)$ is

$$m(x) = a_1 exp(-b_1 x) + a_2 + a_3 exp(b_3 x)$$

In this form the Siler model can be seen to have three independent components which dictate the rate of mortality for three distinct stages in life. The first component dictates the rate of mortality that exponentially decreases from birth and is tuned by the parameters (a_1, b_1), the second is the constant mortality threat that is faced by individuals and most experienced in adulthood, parameter a_2, and last is the exponentially increasing rate of mortality that is experienced late in life, parameters (a_3, b_3).

The benefit of the Siler model of mortality is that it captures the change most of the phenomenon that we observe across age well while still being a fairly straightforward and interpretable model. Others have proposed more complex models in order to catch other nuances observed in mortality such as the higher than expected mortality rates for young adults however, as more terms have been added the generalizability of the model tends to suffer and the amount of data needed to fit the model greatly increases[1,4]. Even so, models that solely focus on the age structure of mortality are ill suited at making projections, either into the future or geographically, and do not take into account how these relationships can effect the estimates of the parameters in the model.

Forecasting and Lee Carter

These models all offer a descriptive frame work for mortality over age however they do not offer a good solution for mortality over time or across regions. In 1992 Lee & Carter developed a model that is still this day widely used for forecasting mortality at the all cause and cause specific level[1,10]. Abandoning the traditional framework of looking at age patterns the model argues that better forecasts can be made by assuming that ages are largely independent in their level and only similar in their rate of change. In this model we estimate $m_{x,t}$

which is the rate of mortality for age x at time t. $m_{x,t}$ is estimated using the following equation

$$log(m_{x,t}) = a_x + b_x k_t + e_{x,t}$$

The model has terms that are both age group specific (a_x and b_x), a set of terms that are specific to time (k_t) and an error term that follows a normal iid distribution

$$e_{x,t} \sim \mathcal{N}(0, \sigma_e)$$

Lee and Carter outline a least squares estimate to these specifications in their original paper. a_x is calculated as the mean of age specific log mortality rates over time or

$$a_x = \sum_{t=1}^{T} log(m_{x,t})/T$$

where T is the ordinal time points in the analysis indexed from 1, such as years. b_x and k_t are calculated simultaneously by taking the singular value decomposition of the matrix $log(m_{x,t} - a_x)$ such that the equation follows $log(m_{x,t} - a_x) = USV^\intercal$ from which b_x and k_t can be obtained from $U[, 1]$ and $V[, 1]$ respectively. These values then can generate estimates for any in sample time point. In order to forecast Lee Carter states that the k_t parameters can be forecasted forward using a forecasting method from the ARIMA family. In the paper they use a Random Walk model that follows the specification

$$k_t = k_{t-1} + d + \epsilon_t$$
$$\epsilon_t \sim \mathcal{N}(0, \sigma_\epsilon)$$

This model produces sensible results in terms of short term forecasts and has performed well in datasets in the US from 1970 to 2005[1,10]. Additionally, the Social Security Administration and the US Census Bureau have reported using variants of the Lee Carter model for their projections and social security planning[1].

While this model has preformed well when tested on US mortality data it has performed lackluster in other environments. Because of the lack of age structure, the model produces nonsensical results where adjacent age groups have differing and sometimes opposite rates of change log rate mortality. This is seldom a problem for short term forecasting, such as 5 years, but with more long term forecasting creates patterns of mortality that does not resemble the standard age curve that Siler captured in his descriptive model. In 2006 Girosi and King

wrote a response to the Lee-Carter model showing where the model works well, the many times that it does not, and criticizing the approach for not pooling information across age and geography[2].

Modified GeoTemporal Siler

In order to fit all the dimensions of concern while still maintaining a coherent age structure this project attempts to use a modification of the Siler model which accounts for deviations away from the expected value due to relatedness across multiple dimensions while also including temporal change. The model can be broken down into three familiar components. The first component is a modification of the Siler model specified above and follows the form

$$S_a = (N0exp(\lambda a) + c) \times (1 - pr_a) + pr_a \times (ma + b)$$

where

$$pr_a = 1/(1 + exp((\kappa - a)))$$

The first term in the first equation corresponds to infant mortality term in the Siler model and the second term is the senescence component. The term $N0$ and c dictate the drop in infant mortality from birth as the constant and intercept in log space. The terms m and b account for the linear growth in log space of mortality in older ages. The pr_a term accounts for the transition from experiencing mortality at child levels to adult levels as the proportion of mortality threats that you experience. This component uses a logit transfer to model the parameters such that they scale from 0 to 1 as age increases such that as you get older you experience less and less of the infant mortality effects. The model also has a temporal trend added to it which accounts for some technological innovation over time.

$$temporal = \beta \times time$$

This component has the ability incorporate within it the effects of covariates relating to medical and technological innovations that effect mortality however we will simply use time in this exercise as a proxy for their effect on mortality. That is to say that we expect that as time passes mortality will decrease as innovations happen. The first two components will constitute what we will consider the deterministic skeleton. This is how we perceive that mortality is structured and how we believe it progresses over time independent of any observations. Observations that we do see are assumed be centered around this deterministic skeleton with some error.

The last component is the structured random error component which captures relatedness across three dimensions of error: age, space and time. In this way we acknowledge that the deterministic skeleton captures the baseline estimate

and that our errors are not independent but rather are correlated across the dimensions that we have stated above. In order to capture the relatedness of these errors we use the following specification

$$\phi_{l,a,t} \sim \mathcal{N}(0, Q^{-1})$$

ϕ can either be thought of as a vector of random variables which follow the distribution as shown above on the right or a 3-dimensional array for the dimensions location, age, and time which we show above on the left for the convenience of notation.

Q acts as the precision matrix, which is the inverse of the variance-covariance matrix. This matrix is a square matrix of length equal to the product of the numbers of locations, ages, and times and is composed of three structured precision matrices for the dimensions of location, age, and time. These precision matrices can be combined for joint precision by using the Kronecker product as shown below.

$$Q = Q_{loc} \otimes Q_{age} \otimes Q_{time}$$

The age and the time precision components follow an AR1 process where the precision matrix is as follows

$$Q_{i,j}^{AR} = \begin{cases} \frac{1}{\sigma^2}, & \text{if } i = j = 0 | i = j = max(i) \\ \frac{1+\rho^2}{\sigma^2}, & \text{else if } i = j \\ \frac{-\rho}{\sigma^2}, & \text{else if } i \sim j \\ 0, & \text{otherwise} \end{cases}$$

For modeling purposes we will be using discrete units for both age and time where the ages reflect Global Burden of Disease age Groups and years are single year groups. This allows for a trivial application for the AR1 precision matrix to be applied to each dimension.

The precision matrix for a the geospatial portion follows a conditional autoregressive form where elements are autoregressive if they are considered neighbors. Because we are using an areal approach,using two dimensional units in space rather than a single point in space, we determine neighbors as those geographical units that share a border with one another. The matrix is defined as follows

$$Q_{i,j}^{CAR} = \begin{cases} \frac{1}{\sigma^2}, & \text{if } i = j \\ \frac{-\rho}{\sigma}, & \text{else if } i \sim j \\ 0, & \text{otherwise} \end{cases}$$

Restrictions to this precision matrix require that σ be greater than zero and the matrix be diagonal dominant. That is that any diagonal element for any row must be greater than the sum of the absolute values of the off diagonal for that row. This may be achieved, by making the maximum value of ρ be slightly less than 1 over the maximum numbers of neighbors any one location has. This may be achieved by a logit transform with a scale.

The full model is then seen as

$$log(\mu_{lat}) = S_a + temporal_t + \phi_{lat}$$

with a probability distribution

$$log(m_{lat}) \sim \mathcal{N}(\mu_{lat}, \sigma_{obs})$$

where m_{lat} is the observed data and μ_{lat} is the predicted.

In structuring the model in this way the deterministic skeleton insures that estimates will follow a coherent age structure that fits with our past observations on how mortality operates.

Fitting the model

Data

In order to test the model we will evaluate the model as fitted on US state mortality data between the years of 1990 to 2015. US data has the luxury of being from a near complete and comprehensive vital registration system that tracks most deaths at the state level for any given year well while documenting age. This data also has the benefit of being tested on many times such that comprehensive benchmarks exist for forecasting. Data was collected from the Institute of Health Metrics and Evaluations reporting values for the states within the United States for every year between 1990-2015 and for the age groups reported in the Global Burden of Disease(GBD) project. Data on the number of deaths that occur in a state along with the population for each year and age group were obtained so that a rate of death could be obtained for each.

The model uses discrete age groups in line with what the GBD has produced for their studies and attempts to model mortality for ages between 0 and 80 in 5 year age groups save for younger ages which use age groups 0-7 days, 8-28 days, 28-365 days, and 1 year to 5 years as the first age groups.

Evaluation

To evaluate the model performance the model will be fit on the years spanning between 1990 and 2005 and then error metrics, RMSE for varying time points, will be calculated using holdout data from 2006 to 2015. After this full forecasts will be made out to 2030 using the entire data set. Separate models will be run for males and females. In order to compare the model to a benchmark we will compare our results against the Lee-Carter model described above. Comparisons will be made by taking the out of sample root mean squared error for the holdout model in two time points, 2009 to 2011, and 2013 to 2015. These time points were chosen to test how changes deeper in forecasting periods can reflect differences in model preference. Results will be reported separately for males and females.

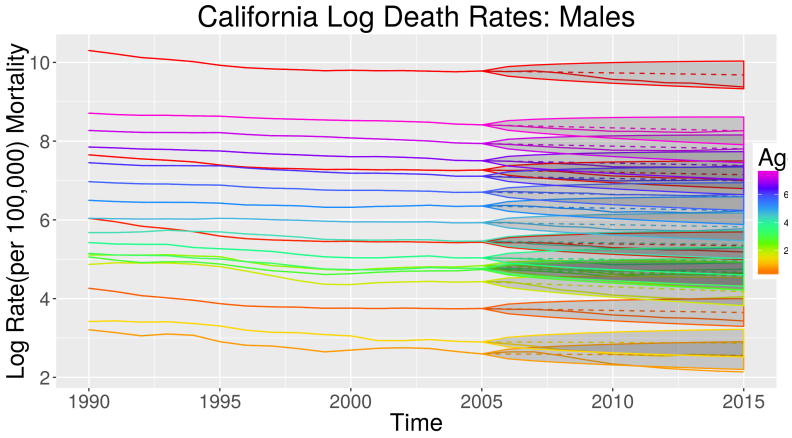
Statistical Approach

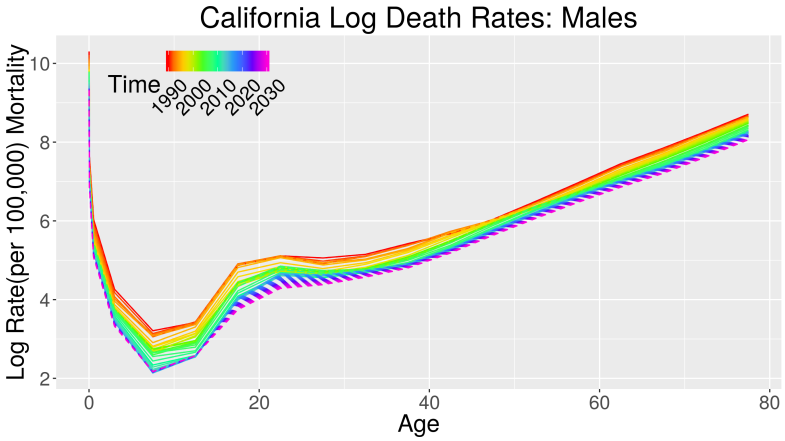
Both models were fit using R version 3.3.2 on a unix operating system. The modified Siler model used the package Template Model Builder (TMB) which uses a C++ template file to construct a model specified by the user. TMB allows users to specify which parameters to be considered as fixed effects vs random effects which dictate which part of the optimization process, either inner or outer, the parameter is a part of. All parameters in the deterministic skeleton were considered fixed effects for this analysis while all the others were considered random effects. The likelihood of the model is determined by the probability distribution specified above. The model is fit using a maximum likelihood estimate to minimize the negative log likelihood of the data given the random effects structure of our ϕ parameter.

The full set of code used to run the analysis can be found here
https://github.com/nmmarquez/2016_Spatio-temporal_models/tree/master/mort_project

Results

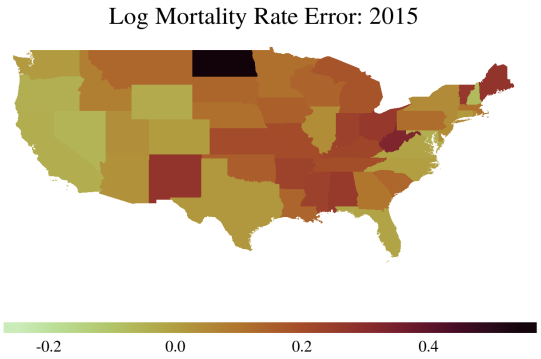
Below is the fit of data between the years 1990 and 2005 predicting from then on forward for the state of California. The second plot switches the age and time dimensions to show that as the model forecasts it maintains the Gompertz age structure. The second plot is fitted using the full data set and forecasted out to 2030.



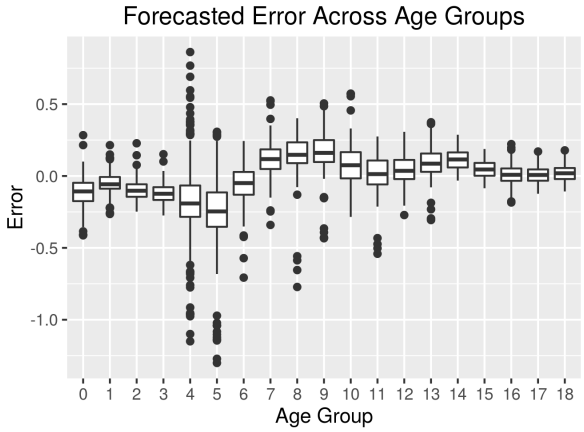


An important take away from these results is the difference in the rate of declining mortality for all age groups. Because the temporal component of the model varies by age group much of the explanation of temporal change is attributed to the correlation in the error with only a minimal portion attributed to the temporal component of the model. Because of this we see big drops in mortality in the future at the age groups around 20 year olds and smaller drops in mortality in other age groups.

Below is the error map shown with model for out of sample results between 2013 and 2015,



as well as a diagnostic plot for error by age for out of sample estimates.



The age distribution of the plot shows that their is systematic bias in the age groups in terms of the directionality and the variation of the error of the residuals. Younger age groups tend to have an underestimated prediction shown by the negative error in the age plot while the age groups in the middle ages tend to have a systematic over estimation in the mortality rate. In addition, looking at age group 4, which is the age group of individuals from 5 years to 10 years old, there is a large amount of variation in the residuals when compared to the adjacent age groups suggesting that our model may not be well capturing the transition from declining mortality rates at young ages to increasing rates at older ages.

The geographic error plot also shows that their is still some systematic bias to overestimate mortality in the Midwest states compared to states on the west coast.

Comparison to Lee Carter Model

Model RMSE Results for Males

YEARS EVALUATED	SILER	LEE CARTER
2009-2011	.1329	.1470
2013-2015	.1822	.1844

Model RMSE Results for Females

YEARS EVALUATED	SILER	LEE CARTER
2009-2011	.1629	.1627
2013-2015	.1910	.1875

Despite the faults of the model stated above, when compared to the Lee-Carter model there is an even split on how well the model performs on tests of out of sample predictive validity. The modified Siler model performs better in measures of RMSE for both sets of years evaluated for males while the reverse is true for females. This result indicates that the Siler model is more robust to the shifts in mortality for males than for females. Both models performed worse for females than males indicating that there is more variation in females that is harder to capture with either model. Case and Deaton have highlighted in a recent study that within the United States there has been a reversal of mortality trends over time for middle age females in recent years[5] that, perhaps, both models have struggled to capture however this has not been tested in this analysis.

Discussion

In this analysis we presented an alternative model for forecasting mortality which uses a structured age format in is deterministic skeleton based off of the Siler model of mortality along with a temporal for development and structured random effects to be able to pool data across regions and forecast mortality results for separate locations and ages. This model differed from other models in that it incorporates the age structure that is inherent in descriptive models of mortality and leverages relationships in other dimensions of concern, namely time and space, in order to make forecasts. In doing so the model is able to take into account rates of changes observed in the past for a particular age group, such as in the Lee-Carter model, while simultaneously keeping a coherent age structure, a criticism that was posited by Girosi-King in their critique of the model.

In addition to addressing the theoretical concerns that are often addressed when modeling mortality this model was shown to produce comparable results in terms of RMSE when compared to the Lee-Carter model, a model that is used heavily in demographic forecasting for mortality. In addition the model retains sensible patterns of mortality for ages even in years forecasted far into the future unlike the Lee Carter model.

In this test case scenario using data from the US our Siler model performed at least as well as the Lee Carter in the demographic break downs that we analyzed. As is the model could be considered for use in other endeavors of demographic forecasting given that it performs well in an out of sample validity test. That being said the model could be improved upon.

Alterations to the Deterministic Skeleton

The backbone of the modified Siler model is in the deterministic skeleton which closely resembles the original Siler model with an added component for temporal technological advancement. The advantage of this model is that it constrains estimates to be centered around a pattern of mortality relating to age that we have empirically observed to be true. This being said there have been several used in the field of descriptive demography which could have been use as the basis for our model. Testing how different models could be swapped in as the age governing portion of the deterministic skeleton. For example replacing the parameters related to the Silder portion of the model with a functional form such as the Heligman-Pollard model or the the multi-exponential model, both of which are outlined in a review of mortality models and forecasting by Boothe and Tickle[1,4].

In addition to replacing the deterministic skeleton with a similar model for mortality which largely imposes our assumptions about how mortality operates over age, we could also use a less structured model and allow the data to govern the shape of the base model. One way that this could be done is by the use of a spline model. Spline models have the advantage of being well studied outside of the context of demographic modeling while still having a history of being used in mortality modeling as well. Both Currie and Shyamalkumar[7,8] outline how they have used P-splines and cubic splines respectively in order to make estimates for mortality with Currie detailing methods for how to make projections. One advantage of using a spline model is that formulations exist which reduce the amount of correlation among parameters such as a basis spline. This is a peril that most of the descriptive models face as outliers have an extreme effect on parameters due to the high correlated nature of those parameters.

One other component of the model that may be changed that is not directly related to the deterministic skeleton is the likelihood evaluation portion of the model. Currently the likelihood of the model is being assessed in log rate space with a normal distribution centered around our estimate. The data generation

process however is one that is observed in counts and thus may not reflect a normal distribution in log space. Rather than evaluate the models likelihood in this fashion

$$log(m_{lat}) \sim \mathcal{N}(\mu_{lat}, \sigma_{obs})$$

we could alternatively evaluate the models counts using a Poisson distribution.

$$deaths_{lat} \sim Poiss(\mu_{lat} * pop_{lat})$$

where pop_{lat} is the population for a particular observed location, age, and time, μ_{lat} is our predicted rate from the model and $deaths_{lat}$ is the actual observed counts of death in that demographic. While this model was not feasible for our case because of the size of units of our location, with smaller areas of estimation, such as at the county or zip code level, this model could act as an alternative.

Alterations to the Random Effects Structure

Another area of improvement comes from the structure of the random effects. As a recap each location, age, year for a modeled sex had its own random effect which was correlated in each dimension by using either a CAR or AR precision structure from which the Kronecker product was taken to get the joint change across these demographic dimensions which was captured in our parameter ϕ. This structure captures the joint correlation in the residuals of these dimensions however it can be argued that each of these dimensions has an independent effect as well.

To highlight this we will describe three hypothetical parameters that may be included in the model, what including them means for our estimates, and how their inclusion can be used as a diagnostic tool for the model. Each of these vectors of parameters will be described as marginal effects, as they pertain to the effect of correlation across a single dimension and can be conceptualized as the partial derivative with respect to that dimension. In addition each marginal effect will have the same covariance structure as the corresponding precision matrix in the parameter ϕ. That is to say that the marginal time, age, and location parameters will have a covariance structure of AR, AR, and CAR respectively.

The first marginal effect we will examine is the age marginal effect. This is potentially the most interesting because our model is strongly based on defining a sensible age structure. Nonetheless, the inclusion of a marginal age effect would reduce the variation in age that is placed on the joint parameter ϕ while individually highlighting the amount of variation explained simply by changes in age. This effect has the potential of acting as a diagnostic function for the age structure of our deterministic skeleton because we would expect that as the skeleton more accurately describes the general pattern of mortality across

age that the amount of variability explained by the age marginal effect would decrease.

A marginal effect on time serves a different theoretical purpose than the marginal effect on age. Rather than acting as an indicator of the goodness of fit for the deterministic skeleton the time marginal effect acts as a safe guard to shocks that effect a single year across all geographies. Perhaps the most exemplary example of this kind of temporal shock in the US is the flu epidemic that occurred in 1918[6]. What is often considered as one of the deadliest natural disasters in recent human history, the 1918 influenza epidemic caused life expectancy to drop in males from 49.6 to 36.6 years of life and 54.3 and 42.2 years of life in females. While the time series that we covered in this analysis did not cover any mortality epidemics of the magnitude of this incident the effects that the 2008 financial fallout had on mortality rates could be a factor that differentially effects the time point just proceeding it.

A marginal geographic effect would offer the same kind of protection to geographic outliers in our model. This has been shown to be true for Southern states in particular[9] and the combined effect in ϕ may miss the marginal geographic effects and at the very least attribute these effects to combined spatial-temporal-age effects.

Conclusion

In this paper we present an alternative method for forecasting all cause mortality by combing a traditionally descriptive model for mortality, a modification of the Silder model, with a temporal component for medical and technological advancement as well as a correlated random effects structure that takes into account relatedness over multiple dimensions. While this model is not meant to take the place of any existing models that exist for forecasting mortality it has been shown in this analysis to provide comparable results to well established models, i.e. the Lee-Carter model, and should be considered when modeling mortality and pooling data across geographies and time.

References

[1] Booth, Tickle (2008). Mortality Modeling and Forecasting. A Review of Methods.

[2] Girosi, F., King, G. (2006). Demographic Forecasting. Cambridge University Press, Cambridge.

[3] Siler, W. (1983). Parameters of mortality in human populations with widely varying life spans. Statistics in Medicine, 2, 373-380. Lee,R.D., Carter, L.R. (1992). Modeling and Forecasting U.S. Mortality. Journal of the American Statistical Association, 87(419), 659-671.

[4] Heligman, L., Pollard, J.H. (1980). The age pattern of mortality. Journal of the Institute of Actuaries, 107(1, No 434), 49-80

[5] Case, A., Deaton, A. (2015). Rising morbidity and mortality in midlife among white non-Hispanic Americans in the 21st century. PNAS 112(No 49), 15078–15083.

[6] Noymer, A. Garenne, M. (2000). The 1918 Influenza Epidemic's Effects on Sex Differentials in Mortality in the United States. Population and Development Review 26(3):565–581.

[7] Shyamal-Kumar, N.D. (2006). Analysis of Mortality Data using Smoothing Spline Poisson Regression. Actuarial Research Clearing House.

[8] Currie I. D., Durban, M., Eilers, P. H. C. (2004). Smoothing and Forecasting Mortality Rates. Statistical Modeling, 4, 279-298.

[9] State-Specific Healthy Life Expectancy at Age 65 Years — United States, 2007–2009. CDC Weekly Report. July 19, 2013 / 62(28);561-566.

[10] Lee, R. Carter, L. (1992). Modeling and Forecasting the Time Series of U.S. Mortality. Journal of the American Statistical Association Vol. 87, No. 419.