Insurance: Mathematics and Economics 52 (2013) 275–285

Contents lists available at SciVerse ScienceDirect

Insurance: Mathematics and Economics

journal homepage: www.elsevier.com/locate/ime

Modeling and forecasting mortality rates

Daniel Mitchell *, Patrick Brockett, Rafael Mendoza-Arriaga, Kumar Muthuraman

McCombs School of Business, University of Texas at Austin, Austin, TX 78712, USA

ARTICLE INFO

Article history:
Received November 2011
Received in revised form
January 2013
Accepted 4 January 2013

Keywords:
Mortality rates
Statistics
Time series
Mortality forecasting

ABSTRACT

We show that by modeling the time series of mortality rate changes rather than mortality rate levels we can better model human mortality. Leveraging on this, we propose a model that expresses log mortality rate changes as an age group dependent linear transformation of a mortality index. The mortality index is modeled as a Normal Inverse Gaussian. We demonstrate, with an exhaustive set of experiments and data sets spanning 11 countries over 100 years, that the proposed model significantly outperforms existing models. We further investigate the ability of multiple principal components, rather than just the first component, to capture differentiating features of different age groups and find that a two component NIG model for log mortality change best fits existing mortality rate data.

1. Introduction

Modeling and forecasting mortality rates has been an active area of research since Graunt (1662) examined mortality in London to create a warning system related to the onset and spread and decline of the bubonic plague. Gaunt's work showed that while individual life length was uncertain, there was a more predictable pattern of longevity and mortality in groups. Halley (1693) showed how to actually construct a non-deficient mortality table from empirical birth–death data and showed how to perform a life annuity calculation based on this table. Such early tables were empirical and calculation was time consuming. Theoretical mortality modeling began with de Moivre (1725) who postulated a uniform distribution of deaths model, and showed simplified annuity calculation methods. Taking a biological approach to mathematical modeling, Gompertz (1825) assumed that the mortality rate represents the body's propensity to succumb to death and that its inverse was the body's ability to withstand death. Assuming that the change in the body's ability to withstand death is proportional to the ability it has to withstand death to begin with led him to a differential equation whose solution is exponential. Solving this and then for the survival function that corresponds to this the solution yields the double exponential survival curve known as the Gompertz curve, which Gavrilov and Gavrilova (2011) shows to fit data to approximately ages 102–105.

The above mortality models are static, however actual mortality is stochastic and evolves over time. Thus, while the mortality models described above fit data at a fixed point in time, the parameters must be re-fit periodically to accommodate changes in mortality patterns. Moreover, the forecasting of future mortality rates is important and not easily accomplished using static mortality models. Future death rates are important to national governments, corporations, and insurance companies. National governments use forecasts of mortality rates to plan social security and health care programs. Couzin-Frankel (2011) estimate that every additional year of life expectancy in the United States costs the U.S. Social Security Administration $50 billion. Corporations offering defined benefit pension plans must assure proper funding of future liabilities, however these future liabilities depend on the yet to be observed future mortality rates. A 2006 study, by Pension Capital Strategies and Jardine Lloyd Thompson (2006) in the UK, found that recognizing the underestimation of expected lifetimes in FTSE100 index companies would cause the aggregate deficit in pension reserves to more than double from £46 billion to £100 billion. In 2010 improved life expectancy added £5 billion to corporate pension obligations in the U.K. as seen in Reuters (2010). In the U.S. the level of pension contributions needed for adequate reserving will increase pension liabilities by 5%–10%, as seen in Halonen (2007). Similarly, insurance companies must use mortality forecasts for pricing annuity contracts and to decide on required future cash reserves. In order to identify, elucidate and quantify these trends we must have a model that adequately captures the temporal as well as age specific dynamics of mortality rates.

1.1. Lee–Carter model and inter-temporal evolution of mortality rates

One of the first papers to model the separate effects of current age and year was Lee and Carter (1992). These authors propose a

* Corresponding author. Tel.: +1 512 471 3322.
E-mail address: dan.mitchell@utexas.edu (D. Mitchell).

0167-6687/$ – see front matter © 2013 Elsevier B.V. All rights reserved.
doi:10.1016/j.insmatheco.2013.01.002

log-bilinear model for mortality rates incorporating both age and year effects. Lee and Carter (1992) has been heavily cited and has been recommended for use by two U.S. Social Security Technical Advisory Panels, it is also used by the U.S. Bureau of the Census as a benchmark model, see for example Hollmann et al. (2000). Moreover since this paper, most other models attempting to assess both time and age evolution of mortality have started with the Lee–Carter framework.

Lee and Carter (1992) model mortality rates for different ages over time by extracting an unobserved state variable from the historical data on mortality rates. This state variable is interpreted as a single temporal mortality evolution index applicable for each age group in the entire population. Since each age group is allowed to respond to the temporal mortality index in different ways, the mortality rate of each age group is some linear function of the temporal mortality index. The mortality index itself is modeled as a Brownian motion with a drift so future predictions of mortality rates can be made by extrapolating the index. Specifically, Lee and Carter (1992) model $m(x, t)$, the central mortality rate of age group x at time t, as a bilinear model for $\ln[m(x, t)]$,

$$m(x, t) = e^{a_x + b_x \kappa_t + \epsilon_{x,t}}. \tag{1.1}$$

Here a_x describes the general shape of the mortality curve, κ_t is the temporal mortality index that captures the evolution of rates over time, and b_x describes each age group's response or susceptibility to the temporal mortality index. If $b_x = 0$ or κ_t is constant, then one returns to static mortality table construction. To estimate the parameters of the model the authors use the singular value decomposition of the matrix of age specific log mortality rates through time to find the matrix of rank one that best approximates the actual log mortality rates. This is numerically equivalent to performing principal component analysis on the covariance matrix of log mortality rate levels.

1.2. Problems with the Lee–Carter model and its variants

Since mortality rates have been trending downwards over at least the last 100 years for all age groups, the Lee–Carter estimation process confounds the first principal component with the time trend. The fact that mortality rates are trending downwards means the covariance matrix of mortality rates vastly overestimates dependence. For example, the covariance over 100 years between the log mortality rates of people aged 5–14 and people aged 65–74 is necessarily very high because early in the time series the mortality rate for both age groups was relatively high compared to their respective means, and later in the time series the mortality rate for both age groups was relatively low. This however does not necessarily mean that if we observe a better than average change in mortality for 65–74 year olds we should expect a better than average change for 5–14 year olds. A cure for cancer would certainly have a large impact on older people and a relatively lesser effect on children, due to the variability in the causes of death for different age groups, but when the cure is found both mortality rates may still decline due to their long term trends. As a result of this phenomenon the fit of the Lee–Carter model can be explained by an exogenous variable.

To illustrate this problem, consider, for example, the consumer price index in Argentina, the miles driven in a particular car, the population of the earth and the GDP of China. These four variables have little relation to each other, but a principal component analysis would certainly show much was "explained" by the first component because all these variables are highly correlated with calendar time. As a simple experiment to illustrate this phenomenon we generated 11 *independent* Brownian motions each with a negative drift and sampled 100 points along each path. The drifts and volatilities of each Brownian motion were randomly chosen from uniform distributions. We then "demeaned" the data and performed a singular value decomposition on the data matrix as would be done in estimating the temporal trend κ_t in Lee–Carter. After repeating the experiment 1000 times we find that according to this matrix decomposition the first singular value, on average, accounted for 99.2% of the variability in the data! Reminding ourselves that this implies that a model with one source of randomness explains 99.2% of the variability, it clearly does not align with the fact that the 11 Brownian motions were generated independently. The only thing these Brownian motions have in common is they all have negative drift, however, following the reasoning of Lee and Carter (1992) we might be led to infer that the first singular value is very informative and we can model all the data by simply modeling this first singular value.

Many papers since Lee and Carter (1992) have tried to improve upon their model by adding more principal components, or a cohort effect, or any range of similar statistical quantities, but they all model the level, and dependence between age groups is modeled using a downward trending temporal trend κ_t. Booth et al. (2006) modify the Lee–Carter model by optimally choosing the time period over which to fit the model and adjust the state variable, κ_t to fit the total number of deaths in each year. DeJong and Tickle (2006) reduce the number of parameters in Lee–Carter to model mortality rates as a smoothed state space model. Yang et al. (2010) use multiple principal components to expand the Lee–Carter model. Chen and Cox (2009) introduce jumps into modeling the state variable, found in Lee and Carter (1992), to increase goodness of fit measures and price insurance linked securities. Deng et al. (2012) use a more advanced jump diffusion model to fit the temporal state variable and Li et al. (2011) identify non-linearities in the temporal state variable. A cohort effect, which incorporates the year of birth into the model, is added to the Lee–Carter model in Renshaw and Haberman (2006). In Booth et al. (2006) the authors compare five variants of the Lee–Carter model with data from several countries. Each of these models are interesting, however all suffer from the same design vulnerability as Lee and Carter (1992) described below, in that they all model the level of log mortality rates and hence misrepresent the temporal dependence structure of mortality rates by age group.

1.3. Outline and contribution

In this paper we build upon the idea of bilinear modeling of age and time from Lee and Carter (1992), however we propose a model that looks at mortality data from a different perspective. We show that by first performing a simple transformation of the data prior to modeling, the subsequent modeling vastly improves our ability to replicate the dynamics of mortality rates through time. This improvement applies not only to the original Lee–Carter model, but also can be used to improve each of the extensions and variants described previously. For forecasting mortality rates we also propose a Normal Inverse Gaussian based mortality index model that is extremely easy to calibrate, has relatively few parameters and performs extremely well. We document this by comparing our model to several other models using several metrics found in literature.

This new model we propose avoids the common problem of modeling log mortality levels. Our model is similar to the Lee–Carter model, however we model changes in log mortality rates rather than levels of log mortality rates. By considering the changes we are able to more accurately capture the dependence structure between ages of mortality and use this to construct a more encompassing model. Referring back to the independent Brownian motions experiment described above, we performed a singular value decomposition on the matrix of differences through

time of the independent Brownian motions and find that now on average the first singular value only accounts for 27.2% of the variability of the data. Theoretically this value should be about 9%, because each singular value should explain exactly the same amount of information and there are 11 singular values.[1] This indicates that by detrending the data we can learn much more about dependence in the data than if we model trending data. By detrending the data in our simple experiment we learned that the data does not exhibit much dependence, a result which we would expect since the components were generated independently. In this paper we show that mortality rates across age groups are indeed dependent upon each other, however by detrending the actual data we can model this dependence better than if we consider the trending data. We show this by considering actual mortality data from 11 countries and by comparing our model to several models from literature.

The rest of this paper is organized as follows. In Section 2, we formulate our model in more detail and describe how to fit the model and provide an overview of the data that used. In Section 3 we compare our model to five other models using in-sample statistical measures of goodness of fit. In Section 4 we discuss forecasting future mortality rates using our model and propose a few different ways to extrapolate into the future, and in Section 5 we conclude.

2. Model formulation and estimation

Following up on this insights described in Section 1.2, we construct a new model that is based on log mortality rate changes rather than levels. We model the mortality rate for each age group as

$$m(x, t+1) = m(x, t)e^{\alpha_x + \beta_x k_t + \epsilon_{x,t}}, \tag{2.1}$$

where α_x describes the average change in log mortality rate for each age group, k_t describes a mortality change index that is the same for all age groups, β_x tells us how each age group responds to the mortality index, and $\epsilon_{x,t}$ is the age and year specific error, which has expected value zero and which we assume to be uncorrelated across time and age group. This is analogous to how the log mortality level is a linear function of the mortality index in Lee and Carter (1992). Now however the trend in the data comes from previous years' mortality rates rather than a trend in the hidden mortality index. In this way we are able to condition on the mortality rates for the current year to make predictions about the future years' mortality. This leads to very high quality in-sample measures of fit and out-of-sample predictions.

We can also work with the equivalent

$$\ln[m(x, t+1)] - \ln[m(x, t)] = \alpha_x + \beta_x k_t + \epsilon_{x,t}. \tag{2.2}$$

This means that the change in log mortality rate for each age group, x, from year t to year $t+1$ is a linear transformation of a time indexed variable k_t plus some error. The way we interpret this model is that in each year we observe a random vector of log mortality rate changes and the log mortality rate the next year for each age group is the log mortality rate from this year plus the random vector. In this way we escape the main drawback of the Lee–Carter model because we have detrended the data by considering changes rather than levels.

The values here of α_x, β_x and k_t are not unique in this representation of the model. For example, if we were to multiply every value of k_t by some constant we could then also divide every value of β_x by that same constant and the model would not change. We could also add a constant to every value of α_x and then adjust k_t in a similar way so that the model would not change. To address this identifiability problem we need to impose restrictions on the model to guarantee uniqueness. We force α_x to be the mean change in log mortality rates, which forces k_t to have expected value zero. We also force the sum of b_x across age groups to be equal to 1, as in Lee and Carter (1992).

2.1. Estimating the model

In order to fit this model we employ methods similar to the Lee–Carter method. First we construct the matrix, $M_{x,t}$, of log mortality rate changes. To do this take the matrix of log mortality rates from the second year to last year under consideration and then subtract from that matrix the matrix of log mortality rates from the first year to the penultimate year under consideration, call this matrix of differences M. In this way we see that $M_{x,t} = \ln[m(x, t+1)] - \ln[m(x, t)]$.

Then from this matrix we must find the mean change over time for each age group, call that α_x, and create a demeaned matrix $\hat{M}$ such that $\hat{M}_{x,t} = M_{x,t} - \alpha_x$. From this matrix we can finally apply the singular value decomposition to obtain β_x and k_t. The singular value decomposition of $\hat{M}$ finds matrices U, S and V such that $\hat{M} = USV'$, where U and V are orthogonal matrices and S is a non-negative diagonal matrix, with the same dimension as $\hat{M}$, and $S_{1,1}$ is the largest element of S. This means k_t is a scalar multiple times the first column of U times $S_{1,1}$ and β_x is the first column of V divided by that same scalar. Then the model prediction of $\ln[m(x, t+1)]$ is equal to $\ln[m(x, t)] + \alpha_x + \beta_x k_t$.

As we mentioned before this is exactly equivalent to performing principal component analysis on the covariance matrix of $\hat{M}$. For example, if we were to find the eigendecomposition of the covariance matrix of $\hat{M}$ and then multiply the first eigenvector by $\hat{M}$ we would find the first principal component. This would be exactly the same as k_t found using the singular value decomposition, up to a scalar multiple. For details on singular value decomposition and principal component analysis see, for example, Jolliffe (2002).

2.2. Data description

In order to fit our model and compare it to other models we use historical data from two different sources. For U.S. mortality rates we use data from the HIST290 National Center for Health Statistics,[2] the same data used in Chen and Cox (2009) and Deng et al. (2012). The data lists death rates per 100,000 people and separates the population into 11 different age groups: (<1), (1–4), (5–14), (15–24), then every 10 years until (75–84) and finally (>85). We combine both sex and race categories to consider all people in the United States in each age group. We use the time frame from 1900–2004.

In this data we have the much discussed flu pandemic of 1918 which resulted in a large jump in mortality for this flu season. Several authors, such as Chen and Cox (2009), have discussed this one year because it seems like an outlier. Some authors, such as Lee and Carter (1992), suggest ignoring it because it is an outlier, while other authors, such as Deng et al. (2012), suggest leaving it in because it has valuable information about rare events. We choose to leave it in because we also believe outliers have valuable information. If one's goal is to forecast future mortality, leaving out arguably the most significant mortality jump in recent history could render the data set more smooth than it actually is.

[1] In our experiment, however, we only sampled the Brownian motions at 100 points in time, leading to large standard errors in estimations of covariance.

[2] Source: www.cdc.gov/nchs/datawh/statab/unpubd/mortabs.htm.

For international mortality rates we use data from the Human Mortality Database (2011). From this database we select the same countries considered in Booth et al. (2006). The data from the Human Mortality Database is richer than the HIST290 data, with data available for every age from infants up to 105. It is also available separated into five year age groups. We use the grouped data because it makes comparison between models simpler and some data from the Human Mortality Database is smoothed over age groups, which we wish to avoid. When available we consider data going back to 1900 but we do not use data before that for issues of consistency. In this international data there are more outliers than in the U.S. data. This is because there have been more extreme events in Europe than in the U.S. over the time frames considered, such as the flu pandemic and World Wars I and II. This means that in order to create a unified model for all countries we must consider a model with heavy tailed distributions.

The Human Mortality Database also has mortality data for the U.S. but the data there only goes back to 1933, whereas the HIST290 data goes back to 1900. In order to stay consistent we also consider U.S. data from the Human Mortality Database, but in later sections when we discuss forecasting we mainly use the HIST290 data so that we capture the flu pandemic of 1918.

3. Performance evaluation

Now that we know how to fit the model to data, we would like to see how well our model performs relative to other models found in literature. We will compare our model to the others using data from 11 different countries. For fairness and direct comparison to existing literature, such as in Booth et al. (2006), we judge the quality of the model using two different in-sample measures of goodness of fit. In Cairns et al. (2011) the authors also use a qualitative comparison of model forecasts. We will also discuss this in Section 4. Most of the models we compare in this section are two factor models while our model is a single factor model. Even with the extra factor considered in these competing models, our model outperforms them in most countries using both measures of fit. We then see that if we include a second time component our model's performance is improved such that we have less error than *every* other two factor model for *all* 11 countries, and when we include a third component our model also performs better than a four factor model in every country.

3.1. Other models

In our comparisons we only consider models of the log mortality rate for ease of comparison to our model. There are other models that use the logit transformation of mortality rates rather than the log; these models are examined extensively in literature and do not seem to perform any better or worse than log models. The models we consider in this section are the Lee–Carter (LC) model, the model of Renshaw and Haberman (2006), which includes a cohort effect (RH), two sub-models of this,[3] denoted H_1 and M3, and the model of Plat (2009) which we denote as P. We choose the M3 model because in Cairns et al. (2011) the authors find this model performs best qualitatively out of sample. We also consider the RH, H_1 and P model because in Haberman and Renshaw (2011) the authors find that these models perform best in sample and make biologically suitable predictions. There are other models that we could also use for comparison but with the models considered here we are able to get a good feel for the performance of our model relative to the others existing in literature. We will find that with the same or fewer factors than the other models, our model will perform the best in all 12 data sets described above.

The RH model states that

$$\ln[m(x,t)] = a_x + b_x^{(1)}k_t + b_x^{(2)}\gamma_{t-x} + \epsilon_{x,t}, \tag{3.1}$$

where γ_{t-x} is the cohort effect. In the model H_1 we set $b_x^{(2)} = 1$ for all x and in the model M3 we set $b_x^{(1)} = 1$ and $b_x^{(2)} = 1$ for all x. We also observe that this is the LC model when $b_x^{(2)} = 0$. The Plat model is a four factor model that has three time factors and the cohort factor and states that

$$\ln[m(x,t)] = a_x + k_t^1 + k_t^2(\bar{x} - x) + k_t^3(\bar{x} - x)^+ + \gamma_{t-x} + \epsilon_{x,t}. \tag{3.2}$$

Here, even though there are four factors the coefficients for each factor are predetermined.

Each of these models can be fit using numerical optimization, where we try to maximize the likelihood function of the data. In Renshaw and Haberman (2006) the authors describe how this can be done using Newton's method. The authors present a table with formulas to update each variable iteratively in order to find the maximum for the RH model. In order to fit the other models we can use these same formulas and add the applicable constraints to the optimization routine by assigning specific values to the appropriate parameters.

Fig. 1 shows the actual log U.S. mortality rate for two different age groups and the Lee–Carter prediction of the log mortality rate and our model's prediction of the log mortality rate. We see that our model predicts next year's mortality rate much better than the Lee–Carter model. We show age group 15–24 and >85 because they are among the best and worst levels of accuracy respectively for the Lee–Carter model. We show similar plots for France and Canada in the age group 25–29 in Fig. 2. In the French data we can see three large spikes in mortality. First we see a spike around 1914 which coincides with the beginning of World War I, then there is a spike almost immediately after that in 1918 highlighting the flu pandemic and finally there is a spike during World War II. Canada on the other hand has no large jumps in mortality as their mortality data starts in 1921, after the flu pandemic, and there were no major wars fought within their borders during the time observed.

The comparison of our model versus the Lee–Carter model presented in Figs. 1 and 2 may seem almost too good to be true. This is due to the strict interpretation of the Lee–Carter model as it is explicitly stated in their paper as compared to our model. In their model the mortality rates for next year do not explicitly depend on the mortality rates this year, whereas in our model we calculate the log mortality rates next year as this year's log mortality rates plus a random vector, so that

$$\ln[m(x,t+1)] = \ln[m(x,t)] + \alpha_x + \beta_x k_t + \epsilon_{x,t}.$$

We intentionally display Figs. 1 and 2 in this way to highlight the large disparity in methodologies. In Girosi and King (2007) the authors reinterpret the Lee–Carter model so that the mortality rates next year are explicitly dependent on the mortality rates this year and we will examine this model and others in Section 3.2.

3.2. In-sample fit

With the data described above we can now compare our model to the others using in-sample measures of fit. We first consider in-sample tests to stay consistent with existing literature. The first measure of fit we use is the root sum of squared errors, or RSSE. This measures the square root of the sum of squared differences

[3] Labeled as in Cairns et al. (2011) and Haberman and Renshaw (2011).

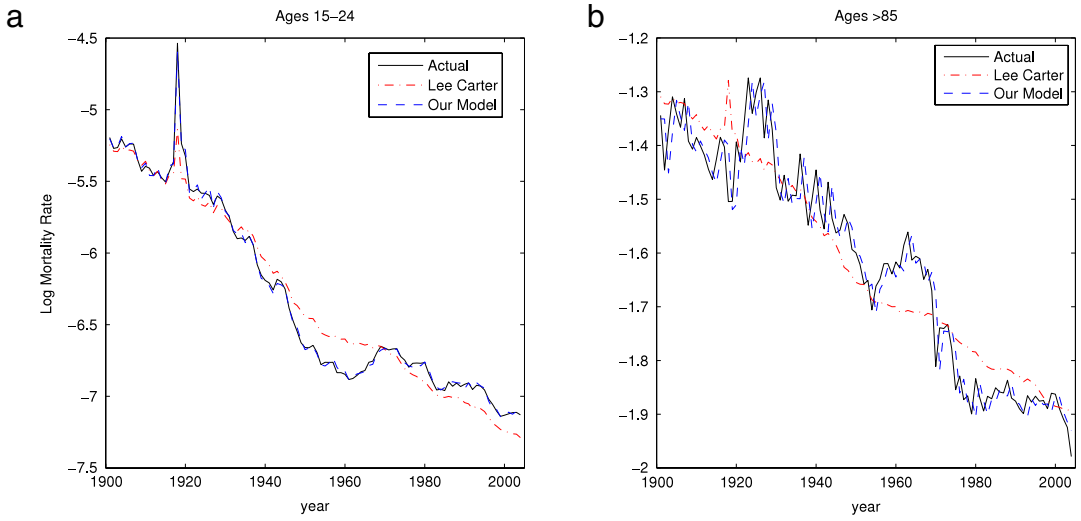

Fig. 1. Comparison of log mortality rate predictions in U.S. data.

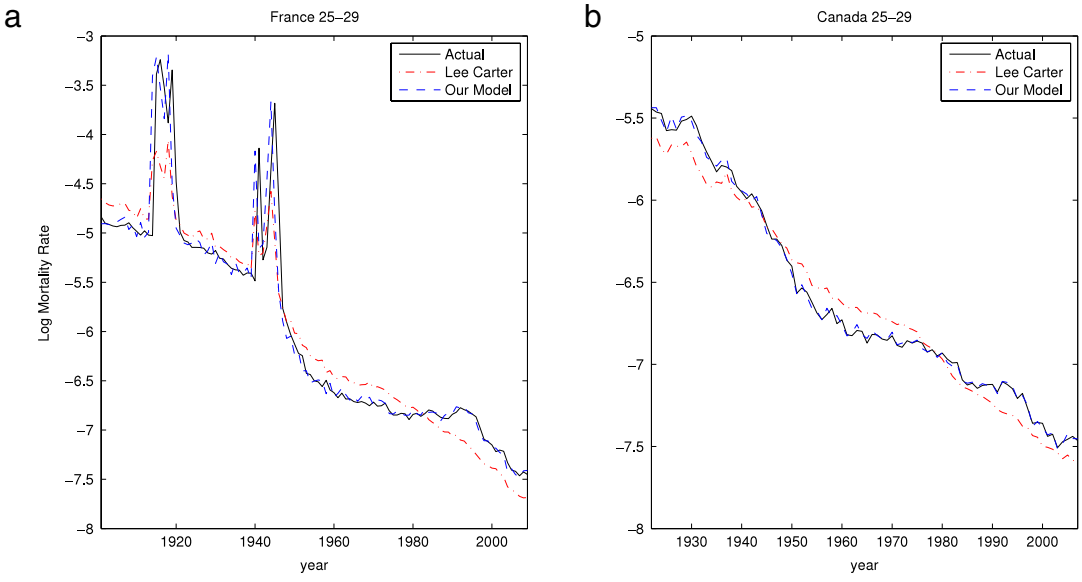

Fig. 2. Comparison of log mortality rate predictions in France and Canada.

between the actual log mortality rate and the model prediction. We define the RSSE as

$$\text{RSSE} = \sqrt{\sum_{x,t} \epsilon_{x,t}^2}. \tag{3.3}$$

In Table 1 we show the results of all 5 models in each country. As each country has different amounts of data we include columns to describe the years considered and the number of age groups for each country, this can explain some of the difference between errors in different countries. In all tables, for each country if our model has the least error then we highlight it, otherwise we highlight each model that has less error than ours. In Table 1 there are two rows for U.S. data, denoted U.S.A.[1] and U.S.A.[2] representing, respectively, the Human Mortality Database and HIST290 data sources.

We see that only in Denmark and Norway does the RH model have less error than our model and in Denmark the P model also has less error than our model. For all other countries our model has the least error of all models considered. The RH model is a two factor model, because it includes a factor for time and cohort,

Table 1
RSSE for several models and countries.

Country	Years	# Ages	Ours	LC	RH	H_1	M3	P
# Factors	–	–	1	1	2	2	2	4
Australia	1921–2007	22	2.20	5.98	3.25	3.96	7.39	3.84
Canada	1921–2007	22	2.25	6.06	2.49	3.00	8.38	3.23
Denmark	1900–2009	21	4.02	6.54	3.77	4.68	9.85	3.97
Finland	1900–2009	22	5.23	10.39	6.23	6.98	13.19	6.85
France	1900–2009	22	3.79	8.92	6.35	7.11	12.24	6.82
Italy	1900–2008	22	4.29	9.04	5.69	6.93	13.58	7.90
Norway	1900–2009	22	4.50	9.26	3.85	5.66	11.41	5.13
Sweden	1900–2010	21	3.41	8.66	3.73	4.69	10.23	4.35
Switzerland	1900–2009	21	3.47	5.94	3.54	4.10	7.90	4.03
U.K.	1922–2009	22	2.77	7.65	2.95	3.55	9.08	3.47
U.S.A.[1]	1933–2009	22	0.97	3.09	1.95	2.39	5.77	2.35
U.S.A.[2]	1900–2004	11	1.10	3.99	2.04	2.54	6.39	3.51

Table 2
Unexplained variance of models using U.S. data.

Model	<1	1–4	5–14	15–24	25–34	35–44	45–54	55–64	65–74	75–84	>85
LC	0.025	0.0028	0.0013	0.041	0.040	0.010	0.039	0.075	0.062	0.086	0.12
Ours	0.0023	0.0013	0.0008	0.0009	0.0014	0.0012	0.0027	0.0063	0.0094	0.016	0.054
RH	0.0093	0.0015	0.0022	0.015	0.010	0.013	0.0089	0.018	0.016	0.018	0.083
H_1	0.017	0.0029	0.0029	0.018	0.019	0.011	0.041	0.030	0.013	0.025	0.069
M3	0.0096	0.077	0.011	0.081	0.046	0.027	0.094	0.14	0.099	0.24	0.54
P	0.038	0.029	0.0022	0.040	0.014	0.011	0.0093	0.021	0.019	0.010	0.012

while our model is only a single factor model. As we will show in the next section, if we include a second or third time factor, such as in Yang et al. (2010), then the error of our model is greatly reduced and no competing model (including the RH and P models) outperforms our model in any country. It is interesting to note that in Janssen and Kunst (2005) the authors find that Denmark and Norway exhibit an especially strong non-linear cohort effect which is picked up in the RH model and not our model. This together with the fact that during World War II the Germans launched a joint operation against these two countries, in Operation Weserübung, might explain why the RH model performs best in these two countries.

The next measure of fit we consider is the one considered in Lee and Carter (1992), which measures the unexplained variance (UV_x) of each age group in the model. For each age group we find the variance of the time series of log mortality rates and the variance of the error between the model predicted log mortality rate and the actual log mortality rate. The ratio of these two variances is then called the unexplained variance, as in

$$UV_x = \frac{\mathrm{Var}(\epsilon_{x,t})}{\mathrm{Var}(\log[m(x,t)])}, \tag{3.4}$$

where the variance is taken through time. Table 2 lists the unexplained variance for our model and the other models in each age group, using only HIST290 U.S. data. Using this measure it is not possible to get an overall measure of the model because there is no way to combine the variance of one age group with the variance of another age group.

Here all of the models do relatively poorly in explaining the variance of older age groups, with the 85 and older age group performing the worst. This is best explained by the fact that in the current state of the world people routinely live to be much older than 85 and thus grouping everyone older than 85 together fails to differentiate people that live to be quite old. Neither our model nor the other models are limited to these age groups, however we selected this data for ease of comparison. We will later see that this is also due to the fact that a single temporal component may not adequately represent the older age groups, and we will discuss a way to correct this. The Plat model, which has 4 factors, performs moderately better than our model in the older age groups for this reason.

The main reason our model performs better than the other models considered here is that we explicitly use this year's mortality rates to predict next year's, whereas the other models only use an aggregate index of next year's mortality rates. This is exactly the point we are trying to make, however one can also examine how the other models perform if they also explicitly use this year's rates in future predictions. This is the topic of Girosi and King (2007) where the authors reinterpret the Lee–Carter model, and is similar to the model proposed in Haberman and Renshaw (2012). If we call $\hat{m}(x,t)$ the models prediction of the mortality rate then Girosi and King (2007) explicitly use this year's rate to say

$$\begin{aligned}\ln[m(x,t+1)] &= \ln[m(x,t)] + \ln[\hat{m}(x,t+1)] \\ &\quad - \ln[\hat{m}(x,t)] + \epsilon_{x,t}.\end{aligned} \tag{3.5}$$

Table 3
RSSE of Girosi and King (2007) type transformations.

Country	Ours	LC	RH	H_1	M3	P
Australia	2.20	2.56	2.28	2.29	2.25	1.75
Canada	2.25	3.25	1.82	1.79	1.91	1.53
Denmark	4.02	4.29	3.94	3.91	4.16	3.40
Finland	5.23	6.46	5.72	5.97	6.05	4.99
France	3.79	5.67	4.67	5.33	5.73	4.79
Italy	4.29	4.85	4.00	4.37	4.51	3.51
Norway	4.50	4.85	3.80	4.25	4.55	3.88
Sweden	3.41	3.78	3.30	3.39	3.60	2.94
Switzerland	3.47	4.02	3.75	3.71	3.91	3.22
U.K.	2.77	4.81	1.96	1.98	1.97	1.36
U.S.A.[1]	0.97	1.58	0.94	0.96	1.08	0.69
U.S.A.[2]	1.10	1.66	1.40	1.46	1.62	1.15

In their paper the authors only perform this transformation on the Lee–Carter model but we can easily extend this to the other models considered here, and we report the RSSE of this in Table 3. We see that even though some models do not predict mortality rates well, they do a relatively better job at predicting rates when this year's rate is considered. In the next section we show that, as before, if we include a second or third time component to our model we have less error than every other model in every country.

3.3. Multiple factors

In the model described above we have restricted ourselves to only include one time component for predicting mortality rates for all age groups. The response of each age group to this time component is completely determined by β_x. Fig. 3 shows the value for β_x, labeled $\beta^{(1)}$ in the figure, for each age group in the HIST290 U.S. data and we can see that for the older age groups β_x gets very close to zero. This tells us that for older age groups, the age groups most important for pension plans, the k_t index has little effect on the change in mortality rate. This suggests that it may be prudent to add a second time component, as in Yang et al. (2010), in order to account for the declining importance of k_t. Also, when we consider the eigenvalues of the covariance matrix, we observe that the first singular value accounts for 80% of the variation in the HIST290 U.S. data and the first two account for 90% of the variation in the data. If we were to include a third singular value we could explain 95% of the variation. In the international data, the first singular value accounts for 50.5% of variation on average, the first and second account for 65.5% on average and the first three singular values account for 75.1% of variation on average. The first three singular values account for less variation in international data than in the U.S. data because there are more age groups in the international data than in the HIST290 U.S. data. There are as many singular values as there are age groups so we cannot expect the fit to be as good for this many age groups. Despite this however, with only two factors we can outperform every other two factor model

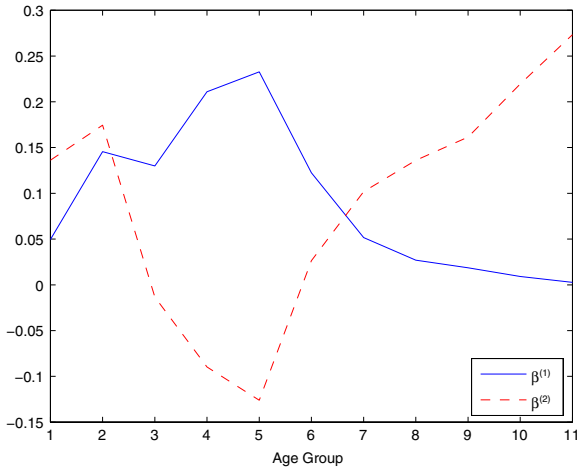

Fig. 3. $\beta_x^{(1)}$ and $\beta_x^{(2)}$ for each age group in the US.

in every country, and with three factors we can even outperform the Plat four factor model.

In order to add a second time component we propose a second model where

$$\ln[m(x,t+1)] - \ln[m(x,t)] = \alpha_x + \beta_x^{(1)}k_t^{(1)} + \beta_x^{(2)}k_t^{(2)} + \epsilon_{x,t}. \tag{3.6}$$

This model is calibrated exactly the same way that the first model, except that we find $\beta_x^{(2)}$ and $k_t^{(2)}$ from the second rows and columns of the singular value decomposition of the data matrix, similarly for a third factor.

In Fig. 3 we plot the values of $\beta_x^{(1)}$ and $\beta_x^{(2)}$, using U.S. data, and see that for the older age groups the second time component is much more important than the first. Also, for the middle age groups the second component is negative, indicating that perhaps the first component overstates correlation between these ages and the others, because in years in which $k_t^{(2)}$ is positive the predicted log mortality change for the middle age groups decreases, whereas it increases for the younger and older age groups, and vice versa for years in which $k_t^{(2)}$ is negative. The third time component mostly adjusts the model to account for younger age groups.

With the addition of this second time component we also improve the goodness of fit of the model. When we used only one time component there were some cases when the two factor RH model had less error than our one factor model. This, however, is not the case when we expand our model to include two time factors. In this case our model has less error than every other two factor model considered in every country. We also see that when we extend to three factors our model has less error than the four factor Plat model in every country. In Table 4 we report the RSSE of the different models with respect to changes. This is essentially a copy of Table 3 except we add the column for the two factor and three factor models. In this table we see that no other two factor model, in any country, has less error than our two factor model, and our three factor model has less error than the four factor Plat model in every country. Using two factors also decreases the error substantially for older age groups. When we consider the RSSE calculated over just the oldest age group in the U.S. using our one factor model the error is 0.22, but when we add the second factor the error is reduced to 0.063, which is a 72% reduction in error for this age group.

Including a second or third factor to our model is similar, in spirit, to how Yang et al. (2010) added multiple time components to the Lee–Carter model. With this in mind we can see how it would be possible that any of the innovative modifications to the

Table 4
RSSE of Girosi and King (2007) type transformations using our two and three factor models.

Country	1 Factor	2 Factor	LC	RH	H_1	M3	3 Factor	P
Australia	2.20	1.88	2.56	2.28	2.29	2.25	1.66	1.75
Canada	2.25	1.63	3.25	1.82	1.79	1.91	1.42	1.53
Denmark	4.02	3.41	4.29	3.94	3.91	4.16	2.94	3.40
Finland	5.23	4.39	6.46	5.72	5.97	6.05	3.64	4.99
France	3.79	3.05	5.67	4.67	5.33	5.73	2.50	4.79
Italy	4.29	3.38	4.85	4.00	4.37	4.51	2.51	3.51
Norway	4.50	3.76	4.85	3.80	4.25	4.55	3.26	3.88
Sweden	3.41	2.88	3.78	3.30	3.39	3.60	2.49	2.94
Switzerland	3.47	3.03	4.02	3.75	3.71	3.91	2.62	3.22
U.K.	2.77	1.49	4.81	1.96	1.98	1.97	1.28	1.36
U.S.A.[1]	0.97	0.74	1.58	0.94	0.96	1.08	0.65	0.69
U.S.A.[2]	1.10	0.82	1.66	1.40	1.46	1.62	0.58	1.15

Lee–Carter model could also be applied to our model; several of these modifications are referenced in Section 1.2. For example, it would be possible to add a cohort effect to Eq. (2.2), rather than just a time component, as Renshaw and Haberman (2006) did to the Lee–Carter model. With this modification it may be possible to find an improved fit over the original RH model. However, since we have found that with two or three time components our model has less error than every other model considered we do not expand this idea further. In addition, for forecasting it is difficult to extrapolate a cohort effect as there is no general consensus on the model that this factor follows.

4. Forecasting

After we have fit this model we would also like to make forecasts of mortality rates in the future and evaluate the quality of these predictions. In order to do this we must model the future distribution of mortality rates. In this section we propose a few different models to fit the k_t series and use them with a normal approximation of $\epsilon_{x,t}$ to make forecasts of future mortality rates. We compare these different models for k_t using the Bayesian Information Criterion (BIC). We first consider forecasts using our one factor model and then expand to two factors.

Fig. 4 shows the time series of k_t for our model using the HIST290 U.S. data. We see that this does not seem to follow any particular process, and this data series passes a runs test therefore we cannot reject the hypothesis that samples of k_t are independent of each other. We interpret this to mean that k_t are independent observations of a random variable. This does however pose a problem around 1918 because right after a big jump up we observe a big jump down. This could be indicative of a state dependent process, but this problem is also observed in the Lee–Carter model when calibrating $\kappa_{t+1} - \kappa_t$ as independent observations of a Gaussian random variable. We ignore this problem and simply interpret k_t as independent observations of a random variable.

4.1. Modeling k_t

There are several different ways we can model k_t, each having its own merits. The first model to consider is that the k_t are independent observations of a Gaussian random variable with a certain mean and variance. This model is the easiest to calibrate as the maximum likelihood estimation of the parameters are the usual average and population variance, requiring no numerical optimization. This distribution is overly simple and events like the

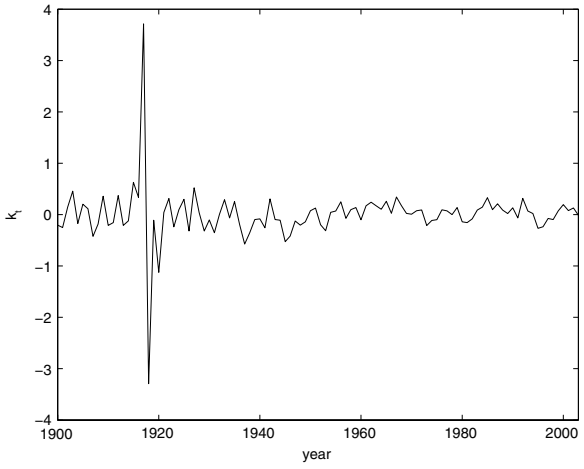

Fig. 4. The k_t time series.

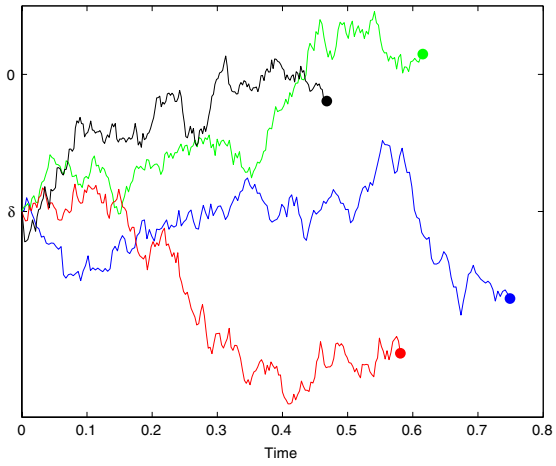

Fig. 5. Four Brownian motions with drift evaluated at random times.

1918 flu and World War II are extremely unlikely under this model and therefore the fit to actual data for most countries is quite bad.

The next model we consider is the Double Exponential Jump Diffusion (DEJD) found in Kou and Wang (2004) and applied to mortality forecasting in Deng et al. (2012). Under this model k_t is modeled as a Gaussian random variable plus a random sum of asymmetric double exponential random variables, where the number of summands is a Poisson random variable. This model requires 6 parameters and in order to fit these parameters we must use numerical optimization to find the maximum likelihood approximation to the parameters. The parameters represent the mean and variance of the normal, the arrival rate of the Poisson random variable, the probability of each summand being positive or negative, and the intensities of each exponential random variable. This model fits the data well, however there is no closed form for the distribution, rather it is in terms of an infinite sum of double integrals. This means that to calibrate the parameters we must approximate the distribution using a finite sum and numerical integration, which is computationally expensive. The details of the calibration and the parameter explanation can be found in Deng et al. (2012).

We propose a third model to fit the k_t series, called the Normal Inverse Gaussian (NIG). This distribution comes from the area of stochastic processes called subordination and is seen applied to mortality modeling in Wang et al. (2011). The distribution describes the value of a Brownian motion, with some drift and unit variance, evaluated at a random time described by an Inverse Gaussian random variable. An Inverse Gaussian random variable describes the distribution of time it takes for a second, independent, Brownian motion with drift to reach a fixed level. This leads to a very rich distribution that can incorporate skewness and excess kurtosis, both of which are seen in the k_t series. The pdf of a random variable, X, that is distributed as a NIG is

$$f_X(x) = e^{\lambda/\theta+\mu(x-\delta)}\sqrt{\frac{\lambda(\lambda+\mu^2\theta^2)}{\pi^2\theta^2(\lambda+(x-\delta)^2)}} \times K_1\left(\frac{1}{\theta}\sqrt{(\lambda+\mu^2\theta^2)(\lambda+(x-\delta)^2)}\right), \tag{4.1}$$

where $\mu \in \mathbb{R}$ represents the drift of the Brownian motion, $\delta \in \mathbb{R}$ represents the starting value of the Brownian Motion, $\theta > 0$ is the mean time at which the Brownian motion is evaluated and $\lambda > 0$ is related to the volatility of the second Brownian motion used to construct the Inverse Gaussian. Also, K_ν is the modified Bessel function of the third kind with $\nu = 1$ in our setting. Although the modified Bessel function is a special function, most numerical packages, including R and Matlab, contain built in routines to approximate it numerically. Using this pdf we can easily find a closed form for the log likelihood function.

Table 5
BIC for different models of k_t.

Country	Gaussian	DEJD	NIG
Australia	191.5	192.1	188.2
Canada	152.8	167.6	159.3
Denmark	202.2	183.3	177.8
Finland	446.1	296.8	296.1
France	487.3	282.0	280.6
Italy	460.0	364.1	355.4
Norway	331.1	305.7	300.4
Sweden	340.1	234.1	233.5
Switzerland	362.9	197.4	211.0
U.K.	255.3	238.7	233.7
U.S.A.[1]	116.4	133.2	124.7
U.S.A.[2]	178.5	53.6	63.4

Fig. 5 shows an example of four Brownian motions starting at δ. Here each Brownian motion runs for a different amount of time, determined by an Inverse Gaussian random variable. At the end of each Brownian motion there is a dot, and the y coordinate here represents an observation of an NIG random variable.

We must use numerical optimization to find the optimal values of $(\mu, \delta, \theta, \lambda)$ but since we have the likelihood function in closed form we can analytically calculate the first and second derivatives to facilitate optimization. Using standard optimization routines included in Matlab finding the optimal $(\mu, \delta, \theta, \lambda)$, from a random initial guess, takes less than one second.

Table 5 reports the Bayesian Information Criterion (BIC) of the k_t process for each of the three models using the international data and both U.S. data sources. The BIC is used in statistics for model selection and it discounts the likelihood function by the number of parameters in the model and the number of data points; smaller values of the BIC indicate a better fit. If L is the optimal value of the log likelihood function, n is the number of parameters in the model and q is the number of data points then the BIC is $-2L + n\ln(q)$. The BIC is just one measure of goodness of fit of a probability distribution function to data.

We can see that the simple Gaussian model performs best for Canada and the U.S.A.[1] data. Both of these datasets start after the

flu of 1918 and World War II was not fought within their borders so these two countries do not have major jumps in mortality rates during the years of available data. The DEJD model performs best in two data sets, and the NIG performs best in the remaining 8 countries. We therefore only consider the NIG model going forward due to its ability to adequately handle spikes in mortality data and its computational simplicity. In a recent paper Haberman and Renshaw (2012) consider a similar model for forecasting but they use a time series model and assume everything to be Gaussian.

4.2. Forecasting mortality rates

Now that we have decided on a model for k_t we would like to make forecasts of the log mortality rates for each age group. To do this we use Monte Carlo simulation because we need to include the error term in future predictions. We assume that the error for each age group, $\epsilon_{x,t}$, is Gaussian with zero mean and standard deviation σ_x. This means the distribution of future log mortality rate is the sum of an NIG and a Normal, which would need to be computed using numerical convolution. But we can easily generate NIG random variables and Gaussian random variables so we can combine them to simulate the distribution of future log mortality rate changes. Then using the Monte Carlo results we can obtain the expected value of future mortality rates along with error boundaries. For example, we can get a 95% predictive interval for future log mortality rates by considering the 2.5% and 97.5% values from the Monte Carlo.

Although the NIG may seem like a complicated distribution, because of the simple way that it is constructed it is easy to simulate random variables from the distribution. As explained before, an NIG is simply a Brownian motion, starting at δ, evaluated at a random time. This means that if T is distributed like an Inverse Gaussian with parameters (θ, λ) and $(X|T)$ is distributed like a normal with mean $\mu T + \delta$ and variance T, then the marginal distribution of X is a NIG with parameters $(\mu, \delta, \theta, \lambda)$. So to generate a NIG first generate an Inverse Gaussian random variable, and then generate a normal with the mean and variance determined by T, repeating these two steps in each of Monte Carlo draw. Each normal random variable will have a different mean and variance and this collection of normal random variables, each with random means and variances, are NIG random variables. See Michael et al. (1976) for information on generating Inverse Gaussian random variables.

A property of the NIG distribution that is useful for forecasting is that if $\{X_i\}_{i=1}^n$ are independent random variables distributed like NIG's with parameters $(\mu, \delta, \theta, \lambda)$ then $\sum_{i=1}^n X_i$ is also a NIG, but with parameters $(\mu, n\delta, n\theta, n^2\lambda)$. So if the change in the log mortality index each year is an NIG then the change over n years is the sum of n changes. This means we can easily forecast log mortality rates n years into the future without having to rely on sums between years in the Monte Carlo simulation.

We have considered two sample periods to forecast future log mortality rates for U.S. data. First we considered 1900–1980 as the training set and 1981–2004 as the testing set, then we considered 1900–1960 as the training set and 1961–2004 as the testing set. Using the training data we calculated α_x, β_x, k_t and the variance of each $\epsilon_{x,t}$, and then fit an NIG distribution to k_t. For the forecast, in each future year we simulated 100,000 random draws from the appropriate distribution, using the parameters found in the training set, and found the average value along with the 95% predictive interval. The results for 1981–2004 are presented in Fig. 6 for a few age groups and the results for 1961–2004 are presented in Fig. 7. In these two figures we compare the 95% predictive interval for our method and the Lee–Carter model; we do not consider forecasts of the other models because there is no general consensus on how to best model the cohort effect. The predictive intervals for the Lee–Carter model are the red dashed lines, our model's predictive interval is shown in the blue line and the green line represents the mean predicted log mortality rate. In general the predictive interval in our model is tighter than that of the Lee–Carter model. This is true in all but three cases; in all but two of the 22 cases our model's predictive interval is inside that of the Lee–Carter model and in one the predictive intervals overlap. In each age group and in each year the future log mortality rate never left the 95% predictive interval for our model or the Lee–Carter model.

We have also run these experiments for all the other countries we considered. We use different training and testing periods than the U.S. data because data availability is different for each country. This results in many graphs similar to those presented in Figs. 6 and 7 and therefore we do not display them in this paper. In the international data we have considered many age groups, many countries and several years of forecasting leading to about 4000 predictions and find that for our model the data escapes the 95% predictive intervals about 3% of the time. For the LC model the data never leaves the 95% predictive intervals. The fact that the data never escapes the 95% predictive interval for the LC model over this many trials leads us to believe that their model over predicts variation in future mortality rates.

4.3. Multiple factors

Making forecasts of future mortality rates using the two factor model is very similar to using the one factor model. In the case of two factors we need a model to fit $k_t^{(1)}$ and $k_t^{(2)}$. Since $k_t^{(1)}$ and $k_t^{(2)}$ are obtained using the singular value decomposition they are necessarily uncorrelated, therefore we assume that $k_t^{(1)}$ and $k_t^{(2)}$ are independent. This means that they may be fit independently of each other. Here we assume that $k_t^{(1)}$ is an NIG random variable and so is $k_t^{(2)}$. Each k must be fit using the methods described above, then to forecast we follow the same procedure as the one factor model, except we also include $k_t^{(2)}$ in the Monte Carlo simulation of future mortality rates, which changes the variance of the error term from the one factor model. For the younger age groups the forecasts of future mortality rates do not change much using the two factor model, but for older age groups the 95% predictive interval is quite different for some.

In Table 6 we have reported a selection of predictive intervals for different age groups from different countries in the last year of available data, using different models. In each test the training set was selected to include every year of available data, except the last 20 years, and we compute the predictive interval of the log mortality rate in the last year of available data. We see that mostly the Lee–Carter model has much wider predictive intervals than our one factor model and two factor model, and the two factor model is wider than the one factor model. This is because we include the variation in the second time component, which is not included in the one factor model. In each of the tests presented in Table 6 the true value of the mortality rate lies inside each predictive interval. It is also easy to add the third component to forecasts by modeling the third factor as an independent NIG. This however does not add much to the predictions of future mortality and so we do not consider it here.

5. Concluding remarks

The central idea that has been demonstrated in this paper is that by using a bilinear model to model mortality rate changes rather than the mortality rates directly, we can fit and forecast mortality rates significantly better. Apart from describing the intuitive reasoning for such a model, we have also shown its

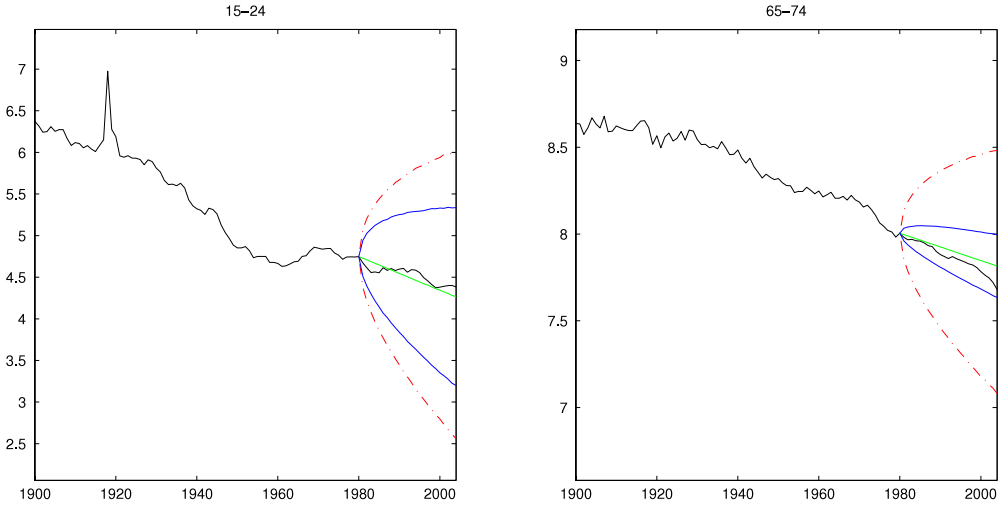

Fig. 6. Forecasts of log mortality rates 24 years into the future along with 95% predictive intervals.

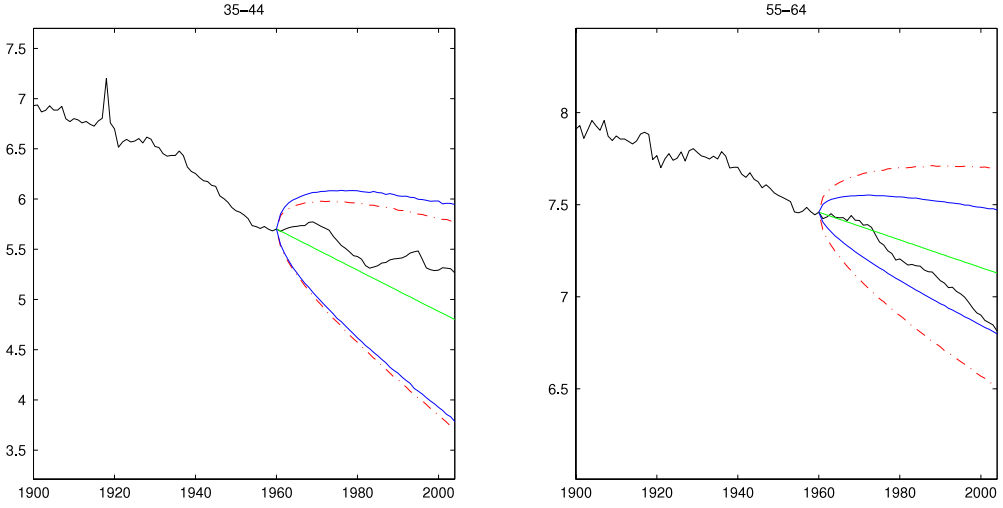

Fig. 7. Forecasts of log mortality rates 44 years into the future along with 95% predictive intervals.

Table 6
A selection of predictive intervals for different age groups in different countries.

Country	Age group	Lee–Carter	One factor	Two factor
Australia	25–29	(−8.94, −5.78)	(−7.99, −6.90)	(−8.04, −6.84)
Canada	≥100	(−1.74, −0.14)	(−1.59, −0.29)	(−2.04, 0.15)
Denmark	55–59	(−5.34, −3.98)	(−4.98, −4.40)	(−5.04, −4.35)
Italy	85–89	(−2.87, −1.25)	(−2.60, −1.48)	(−2.76, −1.32)
Sweden	65–69	(−5.19, −3.29)	(−4.44, −4.00)	(−4.53, −3.91)

benefits with an exhaustive array of computational experiments. Almost all variants of the bilinear Lee–Carter model perform better when used on mortality rate changes rather than the mortality rate data. We have also demonstrated that a Normal Inverse Gaussian (NIG) model for the mortality index component of the model, significantly out performs simple random walk models. The NIG model has only four parameters and is extremely easy to calibrate as compared to other stochastic models like the double exponential jump diffusion model. We have also studied the first two and three principal components of the rate change data and their differentiating abilities in capturing features.

Our exhaustive computational results show that our double component NIG model for log mortality change stands out in its ability to provide much better performance measures for all data sets and is still extremely easy to calibrate.

References

Booth, H., Hyndman, R., Tickle, L., Jong, P.D., 2006. Lee–Carter mortality forecasting: a multi-country comparison of variants and extensions. Demographic Research 15, 289–310.

Cairns, A., Blake, D., Dowd, K., Coughlan, G., Epstein, D., Khalaf-Allah, M., 2011. Mortality density forecasts: an analysis of six stochastic mortality models. Insurance: Mathematics & Economics 48, 355–367.

Chen, H., Cox, S., 2009. Modeling mortality with jumps: applications to mortality securitization. The Journal of Risk and Insurance 76 (3), 727–751.

Couzin-Frankel, J., 2011. A pitched battle over life span. Science 29, 549–550.

DeJong, P., Tickle, L., 2006. Extending Lee–Carter mortality forecasting. Mathematical Population Studies 13 (1), 1–18.

de Moivre, A., 1725. Annuities upon Lives. W.P. Fayram.

Deng, Y., Brockett, P., MacMinn, R., 2012. Longevity/mortality risk modeling and securities pricing. Journal of Risk and Insurance 79 (3), 697–721.

Gavrilov, L., Gavrilova, N., 2011. Mortality measurement at advanced ages: a study of the social security administration death master file. North American Actuarial Journal 15, 432–447.

Girosi, F., King, G., 2007. Understanding the Lee–Carter mortality forecasting method. URL gking.harvard.edu/files/lc.pdf.

Gompertz, B., 1825. On the nature of the function expressive of the law of human mortality, and on a new mode of determining the value of life contingencies. Philosophical Transactions of the Royal Society of London 115, 513–585.

Graunt, J., 1662. Natural and political observations made upon the bills of mortality.
Haberman, S., Renshaw, A., 2011. A comparative study of parametric mortality projection models. Insurance: Mathematics & Economics 48, 35–55.
Haberman, S., Renshaw, A., 2012. Parametric mortality improvement rate modeling and projecting. Insurance: Mathematics & Economics 50, 309–333.
Halley, E., 1693. An estimate of the degrees of the mortality of mankind, drawn from curious tables of the births and funerals at the city of Breslaw; with an attempt to ascertain the price of annuities upon lives. Philosophical Transactions of the Royal Society of London 17, 596–610.
Halonen, D., January 22, 2007. IRS sets costly table. Pension and Investments.
Hollmann, F., Mulder, T., Kallan, J., 2000. Methodology and assumptions for the population projections of the United States: 1999 to 2100. Population Division, U.S. Bureau of Census 38.
Human Mortality Database, 2011. University of California, Berkeley (USA), and Max Plank Institute for Demographic Research (Germany). URL www.mortality.org.
Janssen, F., Kunst, A., 2005. Cohort patterns in mortality trends among the elderly in seven European countries, 1950–99. International Journal of Epidemiology 34, 1149–1159.
Jolliffe, I., 2002. Principal Component Analysis. Springer, New York.
Kou, S., Wang, H., 2004. Option pricing under a double exponential jump diffusion model. Management Science 50 (9), 1178–1192.
Lee, R., Carter, L., 1992. Modeling and forecasting U.S. mortality. Journal of the American Statistical Association 87 (419), 659–671.
Li, J., Chan, W., Cheung, S., 2011. Structural changes in the Lee–Carter mortality indexes: detection and implications. North American Actuarial Journal 15 (1), 13–31.
Michael, J., Schucany, W., Haas, Roy, 1976. Generating random variates using transformations with multiple roots. The American Statistician 30, 88–90.
Pension Capital Strategies, Jardine Lloyd Thompson, 2006. Pension capital strategies releases report on FTSE100 pension disclosures. URL www.jltgroup.com/content/UK/employee_benefits/pressreleases/070117PCSQuaReport.pdf.
Plat, R., 2009. On stochastic mortality modeling. Insurance: Mathematics & Economics 45, 393–404.
Renshaw, A., Haberman, S., 2006. A cohort-based extension to the Lee–Carter model for mortality reduction factors. Insurance: Mathematics & Economics 38, 556–570.
Reuters News Service, November 23, 2010. UK pension liabilities spike as people live longer. URL www.reuters.com/article/idUSLNE6AO01120101125.
Wang, C., Huang, H., Liu, I., 2011. A quantitative comparison of the Lee–Carter model under different types of non-Gaussian innovations. The Geneva Papers 36, 675–696.
Yang, S., Yue, J., Huang, H., 2010. Modeling longevity risks using a principal component approach: a comparison with existing stochastic mortality models. Insurance: Mathematics & Economics 46, 254–270.