Forecasting death rates using exogenous determinants

1. Introduction

This paper addresses three research questions: can the factors found in stochastic mortality forecasting models be associated with real-world trends in health-related variables; does inclusion of health-related factors in models improve forecasts? ; do resulting models give better forecasts than existing stochastic mortality models?

Stochastic mortality models exploit patterns of common variation in deaths data across ages over time. We argue in this paper that taking account of real-world trends in the factors shown to explain mortality decline such as income, health expenditure and lifestyle leads to improved forecasts. The Lee and Carter (1992) model provided the seminal approach to mortality modelling using a principal components analysis of mortality data with one common factor. Subsequent innovations include modelling the cohort effect (Renshaw & Haberman, 2006; Currie, 2006), adding a second period effect (Cairns et al., 2006), using functional principal components analysis (Hyndman and Ullah, 2007), adding additional factors for varying mortality improvement rates across ages (Plat, 2009). Girosi and King (2008) use Bayesian methods to smooth over time, age and country and this approach is extended further by King and Soneji (2011) to incorporate lagged exogenous variables in a Bayesian hierarchical model of mortality rates. The stochastic mortality model approach, which makes use of the regularities found in the age and time profile of mortality data has been the most successful method to date but fails to explain the drivers of mortality improvements and assumes that trends seen in the past will be continued into the future (Booth and Tickle, 2008).

This paper can be divided into three parts. In the first part we use a principal components approach to identify the factor structure of the mortality data for the U.S., U.K., Japan, Finland, Netherlands and Sweden. The second stage takes the latent factors and explains these factors by observed, exogenous factors (GDP, health expenditure, smoking levels, alcohol consumption and diet) using appropriate statistical techniques and using stopping rules to prevent the model become over-parameterised. Finally, having identified the most appropriate exogenous determinants we build the forecasted exogenous variables into a model using the King and Soneji (2011) approach.

There are many possible explanations for recent changes in mortality rates. The health production function approach where health is proxied by mortality provides a framework for understanding the determinants of mortality. Auster et al. (1969) used the following health production model:

$$m_i = c_i + \alpha Z_i + \beta X_i + \gamma HC_i + \delta E_i + u_i$$

where m_i are logged (standardised) mortality rates by US state, Z_i socio-economic status (income, education), X_i lifestyle inputs (alcohol, tobacco), HC_i are healthcare inputs (drugs, doctors, hospital capital stock), E_i captures environmental variables (urbanization, industrialization) and u_i is a random element. Higher incomes allow people to spend more on health inputs. As average incomes rise, people can purchase more non-healthcare inputs that benefit health such as better housing, more nutritious food and gym membership. Where healthcare coverage must be privately paid for, higher incomes also allow people to spend more on better doctors and better hospital care.[1] The choices that individuals make in relation to their

[1] Although higher incomes also permit increased consumption of goods injurious to health such as alcohol and tobacco. In addition, Ruhm (2004) argues that there are less motor vehicle accidents and people adopt healthier lifestyles in economic downturns.

health also affect mortality. Lifestyle factors such as smoking (Leon, 2011; Thornton et al. 2002), obesity (Cutler et al., 2009) and alcohol consumption (Miller and Frech, 2000) are all recognised as significant risk factors. In studying secular trends in mortality, the role of advances in medical technology must be also considered. Cutler and Meara (2004) attributed much of the decline in US adult mortality in the second half of the twentieth century to cardiovascular disease treatment (new drugs, new surgical procedures and specialised equipment). Other factors considered are economic instability (Bethune, 1997; Iversen et al, 1987), environmental air pollution (Schwarz and Dockery, 1992), pharmaceutical expenditure (Miller and Frech, 2000) and crime (Thornton et al., 2002).

The remainder of this paper is laid out as follows. In section 2 we discuss the data that has been used in this study. The methodology is discussed in section 3. The results are presented and analysed in section 4. Section 5 concludes.

2. Data

The mortality data used is taken from the Human Mortality Database collated by the Department of Demography at the University of California, Berkeley and the Max Planck Institute for Demographic Research in Rostock, Germany. Death rates, a ratio of the death count by single age and year divided by an estimate of the exposure-to-risk in the same interval, for males in the US, UK, Japan, Finland, Netherlands and Sweden over the period 1970-2009 were selected. This was the largest set of countries possible given the available mortality and health data. Models were estimated over the period 1970-2000 and mortality rates for the remaining 9 years, 2001-2009, were retained for comparison with forecasts. Due to the exponential nature of mortality rates we model the logarithmically transformed mortality rates.

Data on possible determinants of health were taken from OECD Health data 2009. The following candidate variables were chosen: Alcohol consumption (for those aged 15+), Tobacco consumption (15+), Total fat intake, Fruit and vegetable consumption, Gross domestic product per capita (in 2000 prices) and Total expenditure on health per capita (in 2000 prices). Definitions and descriptive statistics are given in Table 1. GDP and Health expenditure have been both logged in the statistical analysis. Other variables were excluded. The obesity time series are short and patchy and to some extent this information is captured by the food measures included. Data on pharmaceutical expenditure and medical technology capital stock (CT and PET scanners, MRI units, radiation therapy etc) are insufficient and are captured crudely by aggregate health expenditure. Air quality emissions data (SOx,NOx and CO) are inadequate.

Not all determinants of mortality are contemporaneous. Barker (1992) provided evidence that insults to foetal health had life-long consequences based on an analysis of the risk factors for cardiovascular diseases found in adults who were born at the time of the WW2 Dutch famine. The short time series data considered in our study precludes the inclusion of variables of large lag length. Several authors anyhow indicate that these effects may be relatively minor. Murphy (2010) argues that exposure to a health shock has two opposing consequences: selection (excess mortality in the relevant period perhaps leading to the survival of a more robust cohort than average) and scarring (a weakened cohort more susceptible to illness going forward) and that the resultant effect is ambiguous. In a study of twins, Herskind et al. (1996) found no evidence that family environment had an impact on longevity whereas current environmental influences were influential. Similarly, Cutler et al. (2006) indicate cardiovascular risk factors experienced in adulthood are much more significant for mortality than early life exposures. Smoking is probably the one exception. Tobacco consumption influences lung cancer through a lifetime's

exposure which requires information on age of initiation and periods of cessation as well as historic information on cigarettes per day, tar levels of cigarettes and degree of inhalation (Pampel, 2005). This level of data is unavailable.

3. Methodology

3.1 Identifying exogenous factors

Many of the approaches to mortality modelling used in the demographic and actuarial literatures are based on a principal component analysis (PCA) of time series of mortality data by single age. The Lee-Carter model is a one–PC model and other multifactorial derivatives of this model add further cohort terms or additional factors to capture younger or older age mortality. For example, Yang et al. (2010) building on previous PCA studies of mortality (Bell, 1997; Hyndman and Ullah, 2007) considers a two-PC model.

An econometric literature on factor analysis is well-developed. Factor analysis has been used extensively in economic forecasting, modelling business cycles and analysing contagion effects of economic crises. In order to put an economic interpretation on latent factors extracted in these cases, Bai and Ng (2006) developed a statistical test for large cross section (N) and large time dimension (T) datasets to test the adequacy of observed variables as proxies for the unobserved factors. These tests take into account that latent factors are not known but must be estimated.

Assuming that a set of N age-specific death rates, m_{xt}, can be described by a weighted linear combination of r (smaller than N) factors, F_t, we can apply Factor Analysis to the datasets.

This statistical technique accounts for the maximum amount of data variance with a small number of factors while best reproducing the observed correlations between the variables.

$$m_{xt} = \lambda_x^{'} \mathrm{F_t} + \mathrm{e_{xt}} \quad \mathrm{x} = 1, \ldots . \mathrm{N}, \mathrm{t} = 1, \ldots \ldots, \mathrm{T} \quad (1)$$

In classical factor analysis the error terms $\mathrm{e_{xt}}$ are presumed to be independent across x and t. In approximate factor analysis this condition is relaxed.

Using principal components as estimates for the factors, the matrix of factor estimates $\tilde{\mathrm{F}} = \left(\tilde{\mathrm{F}}_1, \ldots . , \tilde{\mathrm{F}}_\mathrm{T}\right)'$ is given by the r eigenvectors associated with the largest eigenvalues of the matrix $MM^{'}/(NT)$ where M is the NxT matrix of age-specific death rates. The factor loadings $\Lambda = (\lambda_1, \ldots . , \lambda_\mathrm{N})'$ are given as $\Lambda = M\tilde{F}/T$. Many of the stopping rules for terminating extraction of principal components were developed in psychometric literature and are not used by econometricians as they require the time dimension to be much larger than the number of variables (Breitung and Eickmeier, 2006). In order to determine r, we use the stopping rule for principal component analysis of the approximate factor model developed by Bai and Ng (2002). The number of factors r for which the information criterion is minimised gives the estimated number of factors $\hat{r}$ .A number of variants of the information criteria are given with the most popular statistic being:

$$IC_p(r) = log\tilde{\sigma}^2(r) + r.\frac{N+T}{NT}\ln[\min(N,T)] \qquad (2)$$

where $\tilde{\sigma}^2(k) = \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\tilde{e}_{it}^2$ and the tilda(~) indicates estimation by PCA.

Given a matrix $\mathrm{G_t}$ of m observed variables, we want to know if they are a linear combination of the r latent variables $\mathrm{F_t}$. Tests have been developed for testing each variable of

G_t singly and for testing G_t as a group. Considering the single tests, each variable of G_t may be an exact factor i.e. $G_{jt} = \delta_j' F_t \; \forall \, t$ or an approximate factor $G_{jt} = \delta_j' F_t + \varepsilon_{jt} \; \forall \, t$. Let $\hat{\delta}_j$ be the least squares estimate of δ_j . Two tests have been developed for the exact case. Letting $\widehat{G}_{jt} = \hat{\delta}_j' \tilde{F}_t$ and $\tau_t(j) = \frac{\widehat{G}_{jt} - G_{jt}}{(\text{var}\widehat{G}_{jt})^{1/2}}$, we count the proportion of the time series for which $\widehat{G}_{jt}$ deviates from G_{jt} by more than ϕ_α, the α percent critical value of the limiting distribution of $\tau_t(j)$.

This gives the statistic

$$A(j) = \frac{1}{T}\sum_1^T 1(\widehat{\tau_t}(j) > \phi_\alpha) \qquad (3)$$

We also test how far $\widehat{G}_{jt}$ is from G_{jt} using the statistic

$$M(j) = \max_{1 \le t \le T} |\widehat{\tau_t}(j)| \qquad (4)$$

This is a more stringent test as it demands that $\widehat{G}_{jt}$ be close to G_{jt} at every point in time. Here, e_{it} must be serially uncorrelated for the limiting distribution of $\tau_t(j)$ to be asymptotically normal.

In the approximate case, we use two goodness of fit statistics:

(i) the noise to signal ratio

$$NS(j) = \frac{\widehat{\text{var}}(\hat{\varepsilon}(j))}{\widehat{\text{var}}(\widehat{G}(j))}. \qquad (5)$$

(ii) the coefficient of determination

$$R^2(j) = \frac{\widehat{\text{var}}(\widehat{G}(j))}{\widehat{\text{var}}(G(j))}. \qquad (6)$$

Testing the group G_t as a set, the canonical correlations between G_t and F_t are considered. The first canonical correlation, ρ_1, is the largest correlation that can be found for linear combinations of G_t and F_t. The second canonical correlation, ρ_2, is the largest correlation that can be found from linear combinations of G_t and F_t uncorrelated with those giving the first canonical correlation, and so on. Having to estimate F_t has no effect on the sampling distribution of the canonical correlations. For k=1, ...,min[m,r] and $(F'_t, G'_t)'$ identically independently normally distributed,

$$(\rho_k^{2-}, \rho_k^{2+}) = \left(\tilde{\rho}_k^2 - 2\phi_\alpha \frac{\tilde{\rho}_k(1-\tilde{\rho}_k^2)}{\sqrt{T}}, \tilde{\rho}_k^2 + 2\phi_\alpha \frac{\tilde{\rho}_k(1-\tilde{\rho}_k^2)}{\sqrt{T}}\right) \qquad (7)$$

where $\tilde{\rho}_k$is the k[th] canonical correlation between G_t and $\tilde{F}_t$. If all the m variables in G_t are exact factors then the canonical correlations will all be unity. If the m variables are linearly dependent then the number of non-zero canonical correlations will be less than m. Any single variable in G_t may be found to be exact or approximate factors from the single tests but may be a linear combination of other observed variables as indicated by the group tests.

3.2 Forecasting

Having identified the most appropriate exogenous factors to build into our model of mortality we take the models of Girosi and King (2008) and its extension allowing for exogenous variables (King and Soneji, 2011) as a starting point to build our covariate-driven model of mortality. Girosi and King (2008) developed a method of modelling mortality rates across ages, years and countries which uses a Bayesian hierarchical approach to information pooling. Their objective in doing this was to make use of beliefs that data across neighbouring ages, years or countries should show similar characteristics. For example, we might expect that the mortality

rate experienced by a 20 year old in a given year should be similar to that experienced by the 21 year old or the 19 year old in the same year. Similarly, the mortality rate in say 2000, for a given age should be similar to the mortality rate for that same age in 1999 or in 2001. The hierarchical approach allows the smoothing of mortality rates for a single country across ages and time and so produces realistic forecasts of mortality that do not break norms in terms of age and time going forward (for example, mortality rates increasing with age and improving in time). Considering the logarithmically transformed mortality rate during year t for life aged x as $m_{x,t}$ and a matrix of covariates $\boldsymbol{Z}$ they set out the following model specification:

$$m_{x,t} \sim N\left(\mu_{x,t}, \frac{\sigma_x^2}{b_{x,t}}\right), x = 1, \dots, N, t = 1, \dots, T \qquad (8)$$

$$\mu_{x,t} = \boldsymbol{Z_{x,t}}\beta_x$$

This specification only differs from a standard linear regression model in the $b_{x,t}$ weighting that is applied to the variance and in the approach to defining the parameters β_x and σ_x^2. The specification above provides the basic building block of the Bayesian hierarchical approach in which the coefficients β_x and standard deviations σ_x^2 are random variables with their own prior distributions. The prior for the variance random variable σ is denoted $P(\sigma)$. The prior on the coefficients β_x which depends on its own "hyper-parameter" θ is denoted $P(\beta \mid \theta)$ is chosen to reflect the "similarity" belief across cross sections. This is formalised by introducing a density function for the prior defined as:

$$P(\beta|\theta) \propto exp\left(-\frac{1}{2}H^{\mu}[\beta, \theta]\right) \quad (9)$$

where

$$H^{\mu}[\beta,\theta] \equiv \frac{1}{2}\sum s_{i,j}\left\|\beta_i - \beta_j\right\|_{\theta}^{2} \quad (10)$$

where the notation $\left\|\beta_i - \beta_j\right\|_{\theta}^{2}$ denotes a weighted Euclidean norm and where the symmetric matrix s is called the adjacency matrix. Its entries reflect the "proximity" of cross section i to cross section j and hence the weight put on the relationship between the coefficients of cross section i and cross section j. Using this approach the fitted model shows forecasts that are smooth in the age and time dimension and that do not violate the smoothness beliefs across age and time that may be violated by using multiple regression methods.

The method used to develop their model with exogenous covariates was to identify links between mortality rates and lagged covariates, specifically smoking habits and obesity. They argue against using contemporaneous relationships in favour of lagged relationships and from the literature determined the optimal lag period to be 25 years in the case of smoking. Although it may be appropriate to use current data to determine future mortality in the case of smoking rates this approach does not facilitate the inclusion of variables which affect mortality contemporaneously such as GDP, health expenditure, alcohol consumption or diet. Therefore we attempt to forecast these variables here while acknowledging better forecasts could be obtained using more adequately specified structural models or more sophisticated statistical techniques. In our model we forecast the identified exogenous variables using ARIMA methods.

Two benchmark models are used to assess the results of our approach. Lee and Carter (1992) takes the first principal component, $\mathrm{k_t}$, of the log-mortality matrix, m_{xt}, and then uses ARIMA time series models to forecast $\mathrm{k_t}$.

$$m_{xt} = a_x + b_x \mathrm{k_t} + \mathrm{e_{xt}} \qquad (11)$$

Hyndman and Ullah (2007) first smooth the data to give a set of curves $f_t(x)$ and then use functional principal component analysis to decompose $\{f_t(x)\}$ in terms of a set of orthonormal basis functions $\{\varphi_k(x)\}$. Mortality forecasts are then based on forecasts of the coefficients β_{tk}multiplied by the basis functions.

$$f_t(x) = \mu(x) + \sum_{k=1}^{K} \beta_{tk}\varphi_k(x) + \mathrm{e_t}(x) \qquad (12)$$

The number of basis functions or *order*, K*,* is found by minimising squared errors on rolling forecasts over the fitting period. This model also includes a bias-adjustment to bring forecasts into line with the fitting period data by adjusting forecasts by the difference between the fit and the last year of observed data.

4. Results

4.1 Identifying exogenous factors

We first of all try to get a sense of the latent factor structure of the mortality data for males in each country over the fitting period 1970-2000. The number of factors is first determined and these factors are analysed to check their association with younger or older age mortality variation.

Applying the stopping rule (equation 2) we find the following factor structure: the estimated number of factors is $\hat{r}$=4 for the US and Japan ; $\hat{r}$=2 for the UK, Finland and Sweden ; and $\hat{r}$=1 for the Netherlands. It would appear that for larger countries data are less noisy and the multifactor structure of mortality data is more easily identified. A Lee-Carter model with one factor or other early derivatives of this model would be therefore inadequate to capture all the

common variation in the US and Japanese data while more recent multifactorial models such as Plat (2009) or Cairns et al. (2006) would provide a factor structure of a more suitable dimension for model fitting.

Before associating the factors extracted from the data with real-world trends, the *communality* (the percentage of the variation explained) at each age is estimated and graphed in figure 1. In US and Japan male mortality data, the four principal components extracted explain almost all the variation in the data at every age. In the other smaller countries, the principal components do not explain variation in younger and older age mortality well probably due to smaller numbers of deaths at these ages.

The common factors extracted for each country tend to be associated with particular ages. From the rotated factor loadings graphed in figures 2-6, we see that factors are either associated with younger or older age mortality (for the Netherlands, $\hat{r}=1$ and the solution cannot therefore be rotated). Male mortality over 45 years of age is explained by US, UK and Finland factor 1, while the other factor(s) for these countries explain younger age mortality. This would indicate that we need at least two types of exogenous factors to explain the variation in mortality rates: perhaps lifestyle-related factors to explain younger age mortality (e.g. alcohol consumption) and factors related to health treatment improvements to explain older age mortality. Japanese and to a lesser extent Swedish mortality rates behave differently with most factors not particularly associated with any particular ages. This would indicate that we require exogenous factors associated with mortality improvements at every age (e.g. income) or alternatively a large set of exogenous factors which together explain each principal component extracted.

The proposed exogenous factors are graphed for each country in figures 7-12. Alcohol consumption has peaked and declined in most countries except noticeably the UK where it has been progressively rising (figure 7). Smoking has declined everywhere except in Japan (figure 8). Diets have generally getting worse in Japan with fat intake increasing to Western levels (figure 9) and fruit and vegetable consumption declining slightly (figure 10). In contrast, people in the US and UK have been consuming increasing amounts of fruit and vegetables. Improving economic growth (figure 11) and steep increases in health expenditure (figure 12) should also

factors. The goodness of fit statistic (R^2(j)) and the noise to signal ratio (NS(j)) indicate how far the proxies are from the true factors. Bai and Ng suggest that if NS(j) > 0.5 and/or R^2(j) < 0.95 then errors in the linear relationship between the proposed factors and the latent factors are non-negligible and the proposed factors are not strong proxies for the latent factors. According to these measures, Alcohol consumption, Tobacco consumption, Health expenditure per capita and GDP per capita are particularly strong proxies. Of course, numerous studies have found cointegration between national income and health expenditure (Freeman, 2003; Westerlund, 2007; Moscone and Tosetti, 2010) and using the two variables may provide little extra information than simply using one.

The squared canonical correlations, $\hat{\rho}(k)^2$, are given in the final column. The first value indicates that there is a linear combination of the proposed proxies and a linear combination of the four latent factors that are highly correlated ($\hat{\rho}(k)^2 = 0.999$). The second value indicates that there is a second linear combination orthogonal to that already found which is also highly correlated with the four latent factors ($\hat{\rho}(k)^2 = 0.951$).). The set of six proposed factors does not span the latent factor space as there are only two well-defined relations (subsequent canonical correlations have a lower bound close to zero). The squared canonical correlations between the latent factors and a set of just the two variables Alcohol consumption and Health expenditure per capita are 0.997 and 0.934 (not in table) suggesting that little is gained by adding the extra four variables. The variables Alcohol consumption and Health expenditure per capita therefore underlie the two non-zero canonical correlations between latent and observed factors. For the purposes of forecasting, these two variables - one a lifestyle variable and the other a medical care variable - appear to be strongly associated with mortality trends and are sufficiently different conceptually to provide distinct forecasting power.

From the discussion above, the steps to identify which observed factors best explain the factor structure of the data can be summarised as:

- Extract $\hat{r}$ latent factors (the set F_t).
- Determine the number of non-zero squared canonical correlations $\hat{\rho}(k)^2$ between F_t and the matrix G_t of observed variables
- Find the minimal subset of G_t which replicates the canonical correlations. Choose candidate variables based on the NS(j) < 0.5 and $R^2(j) > 0.95$ criterion where possible.

In the case of Japan (Table 4), almost all the variables except Tobacco consumption and Fruit and vegetable consumption are strong proxies for the four latent factors using Bai and Ng's criterion. This finding is in keeping with figure 4 where a more complicated latent factor structure was observed. Nevertheless, the set of factors considered does not encompass the latent factor structure. There are two well-defined relations between proxies and latent factors with canonical correlations approaching unity although the third canonical correlation is also large. As health expenditure and GDP cointegrate, the three variables - alcohol consumption, fat intake and health expenditure - provide an appropriate basis for forecasting models. This set has squared canonical correlations of 0.995, 0.851 and 0.416 with the latent factors which when compared to column 6 of Table 4 indicates a little information is lost by focusing on this smaller subset.

The results for the other countries are given in Table 3, 5-7. For all countries there is only one non-zero canonical correlation (the lower bound for the second canonical correlation is close to zero). In each case there is at least one variable with high $R^2(j)$ suggesting one variable

is sufficient in each case to proxy for the latent factor and other variables add little extra information. Variables with highest $R^2(j)$ were selected and are emboldened in these tables.

As a general summary, this analysis would indicate that the factor structure of mortality data in these countries is explained first of all by GDP or Health expenditure per capita. These factors are generally highly correlated with the factors extracted. Although GDP and Health care expenditure per capita are not good predictors of population health across high-income countries as observed by Leon (2011) among others trends in these variables are generally the most closely associated with the variation in mortality rates within country over time. Where the factor structure of mortality has many factors (the larger countries *viz*. USA, Japan), the factors associated with smoking and drinking explain the additional factors extracted. Fat intake is also important in Japan due to changing dietary patterns indicated in figure 9. The primacy of tobacco consumption in the Netherlands male mortality data is unusual although idiosyncratic patterns in coronary heart disease mortality in the Netherlands have been observed (Vaartjes et al, 2011) and the importance of the pattern of smoking decline in explaining secular trends in the life expectancy of Dutch women has been noted (Leon, 2011).

4.2 Forecasting results

For the purposes of forecasting these variables, the Schwarz information criterion was used to decide between models with various number of auto-regressive (AR) and moving average parameters (MA). As GDP is generally found to be non-stationary (e.g. Westerlund, 2007), this variable was first-differenced and consequently health expenditure per capita also. Assessing prediction errors post-hoc, these models did not necessarily provide the best forecasts but reflect the level of uncertainty encountered in practice. This approach is not dissimilar to

forecasting the common factor with ARIMA in Lee and Carter (1992) and making mortality forecasts conditional on these forecasts.

Taking the predicted exogenous factors in the analysis above we apply the King and Soneji (2011) approach to forecast mortality rates using the YourCast software[2]. We present the fitting and forecasting results of our model in tables 8-10. We also present results using the models of Lee Carter (1992), Hyndman and Ullah (2007) and the King and Soneji model with no exogenous variables (Girosi and King, 2008) for comparison. For a given mortality rate at time t and for age x ($m_{x,t}$), we measure the fitting and forecasting quality using the root mean square error (RMS)

$$RMS = \sqrt{\frac{1}{(X_1 - X_2 + 1)T}\sum_{x=X_1}^{X_2}\sum_{t=1}^{T}\left(projected\ m_{x,t} - actual\ m_{x,t}\right)^2} \quad (13)$$

The most important point to note is that using forecast exogenous variables in a structured model (row labelled *King and Soneji* in Table 8) improves forecasts compared to the model without these variables (labelled *Girosi and King)* in all countries except the UK. For Finnish mortality, the RMS is approximately one-third of the value after inclusion of health-related variables while for Japan and Sweden it is approximately halved. The degree of improvement appears unrelated to the number of exogenous variables added. When compared to the *Lee Carter* and *Hyndman Ullah* models, the model containing information on the determinants of health surpasses the benchmark models for Japan, Finland and the Netherlands whereas for US data the *King and Soneji* model outperforms the *Hyndman Ullah* model and gives results only marginally worse

[2] For more details on the YourCast software used in this study and developed by King and Soneji go to http://www.gking.harvard.edu/yourcast.

than *Lee Carter*. This result is encouraging bearing in mind the simplicity of the included variables which vary in the time dimension but not across the age dimension.

Results for the one-year ahead forecast and the nine-year forecast only are given in Table 9. Results when the exogenous variables are added are better than the statistical Girosi and King model for the one-step ahead forecast for almost every country (practically the same for Japan). The picture is not as clear for the end of the forecasting period where adding the variables is seen to decrease the RMS in only three out of the six countries (Japan, Finland and Sweden). There is no clear best approach for the short-run forecast where *Lee Carter* and *King and Soneji* are best each for two countries. For the longer run forecasts the Bayesian hierarchical models are best with the *Girosi and King* version (i.e. without the exogenous variables) having the lowest RMS for three countries and *King and Soneji* best for two countries.

The actual and forecast age-specific mortality rates for 2009 are plotted in figures 13-18 for the *King and Soneji* model. Forecasts are very accurate for Japan at most ages except for young adults. For all other countries, improvements in old-age mortality are typically underpredicted whereas the opposite is the case for younger ages. This would indicate that the general level has been predicted correctly by using exogenous variables reflecting the general trend but age-specific exogenous variables should be used to provide greater accuracy at younger and older ages. Mortality rates for younger ages in the smaller countries fluctuate significantly over time making accurate forecasting prone to error for all models. The default parameters used in the Bayesian hierarchical models to smooth such irregularities could therefore be reduced to produce more reasonable forecasts for noisier data.

4.3 Fitting results

Root mean square errors for the fitting results are given in Table 10. Adding exogenous variables improves the fit of the Bayesian hierarchical models comparing *King and Soneji* results with *Girosi and King*. That the fit improves with more variables is of course not surprising although RMS is reduced substantially even when one variable is added. RMS is reduced by two-thirds for the US (two exogenous variables added) which is comparable to the reduction for the Netherlands and Sweden (one variable added). The *King and Soneji* model provides the best fit for the US and Japan with RMS figures elsewhere comparable to the benchmark models. The Hyndman Ullah model gives the best fit where a large number of basis functions are fitted to the data (see footnote to Table 8 – UK order = 3, Finland 6, Netherlands 7).

In figures 13-18, fitted results are also given by age for 1970 for the King and Soneji model. The fit across ages is generally very close except where there are smaller numbers of deaths i.e. in smaller countries at younger ages (figures 16-18 males aged 20-40).

5. Conclusion

In this paper, we have identified the statistical factors used in conventional mortality forecasting models with exogenous factors deemed plausible by health economics and epidemiological literature. The incorporation of the identified exogenous factors into mortality models improves forecasts and gives results comparable or better than conventional models. Although this was done in a simple way the potential for improving forecasting by using an explanatory-based approach in contrast to conventional extrapolative models has been demonstrated.

The factor structure is easy to identify for larger countries where common variation at all ages can be identified. For smaller countries there are less factors identified and common variation at younger and older ages is less well described. Regardless of country size, the variable that best describes the factor structure is generally either income per capita or health expenditure per capita. Variables related to smoking and drinking behaviour are next most important. Only for Japanese mortality data were changing patterns in diet closely related to the factors extracted.

The exogenous determinants of health related to the latent factor structure were then added to the Girosi and King Bayesian hierarchical model and were seen to improve forecasts in almost all countries. Even though these variables reflected general trends and were not age-specific, forecast results for this model were best in 3 out of the 6 countries studied. An analysis of forecast errors by age would indicate that age-specific variables would improve forecasts by correcting for an underestimation of mortality decline at old ages and overestimation of young adult mortality.

Further work could also consider the inclusion of lagged variables where appropriate. There are other approaches which would also permit a more-explanatory based approach. For example, the application of the semi-parametric estimation approach of Connor et al. (2011) in this context would allow us to unite the identification of exogenous variables and the estimation of the latent factor model in one step.

Tables

Table 1: Descriptive statistics, 1970-2000 : Mean (standard deviation)

Exogenous factor	US	UK	Japan	Finland	Netherlands	Sweden
Alcohol -Annual consumption of pure alcohol in litres, per person, aged 15 years and over	9.5 (0.8)	9.3 (0.7)	7.8 (1.0)	8.2 (0.8)	10.5 (1.1)	6.6 (0.5)
Tobacco -Annual consumption of tobacco items (e.g. cigarettes, cigars) in grams per person aged 15 years and over	2645 (675)	2349 (511)	3227 (164)	1410 (242)	2368 (713)	1899 (205)
Fat -Total fat (grams per capita per day)	133.5 (10.1)	138.9 (3.1)	73.1 (9.4)	128.0 (3.7)	121.6 (5.8)	122.9 (4.6)
Fruit & Veg - All fruit and vegetable consumption (except wine) in kilos per capita	219.7 (18.6)	151.8 (14.3)	173.1 (8.7)	119.3 (22.3)	204.0 (50.0)	141.4 (16.2)
GDP -Gross domestic product per capita in national currency units at 2000 price levels	25420 (4712)	11896 (2275)	2,994,819 (707,685)	18325 (3520)	25337 (2157)	194137 (27,101)
Health exp - Total health expenditure (private and public) per capita in national currency units at 2000 price levels	2787 (1098)	712 (222)	191,957 (63,732)	1299 (372)	1633 (324)	15870 (2693)

Table 2: Testing the factors in US age-specific mortality rates by single age 20-89, 1970-2000

G_j	A(j)	M(j)	R^2(j)	NS(j)	$\hat{\rho}(k)^2$
G_1, Alcohol	**0.323**	**5.94**	**0.976 (0.960, 0.993)**	**0.024**	0.999 (0.998, 1.000)
G_2, Tobacco	**0.710**	**7.53**	**0.991 (0.984, 0.997)**	**0.009**	0.951 (0.918, 0.985)
G_3, Fat	0.806	25.82	0.911 (0.850, 0.971)	0.098	0.366 (0.096, 0.636)
G_4, Fruit & Veg	0.935	44.99	0.878 (0.798, 0.959)	0.139	0.130 (-0.091, 0.352)
G_5, GDP	0.806	14.56	0.975 (0.958, 0.992)	0.025	-
G_6, Health exp	**0.419**	**5.20**	**0.997 (0.994, 0.999)**	**0.003**	-

A(j) is the frequency that $|\hat{\tau}_t(j)|$ exceeds the 5% asymptotic critical value. M(j) is the value of the test. R^2 is defined in 6, NS(j) defined in 5 and $\hat{\rho}(k)^2$ is the vector of canonical correlations G_t with respect to F_t.

Table 3: Testing the factors in UK male age-specific mortality rates by single age 20-89, 1970-2000

G_j	A(j)	M(j)	R^2(j)	NS(j)	$\hat{\rho}(k)^2$
G_1, Alcohol	0.839	49.14	0.546 (0.310, 0.782)	0.832	0.991 (0.987, 0.997)
G_2, Tobacco	0.871	24.62	0.809 (0.688, 0.930)	0.236	0.323 (0.052, 0.594)
G_3, Fat	0.645	26.43	0.313 (0.042, 0.584)	2.195	-
G_4, Fruit & Veg	0.871	28.02	0.815 (0.698, 0.933)	0.227	-
G_5, GDP	**0.645**	**10.43**	**0.967 (0.944, 0.990)**	**0.035**	-
G_6, Health exp	0.484	14.80	0.970 (0.949, 0.991)	0.031	-

A(j) is the frequency that $|\hat{\tau}_t(j)|$ exceeds the 5% asymptotic critical value. M(j) is the value of the test. R^2 is defined in 6, NS(j) defined in 5 and $\hat{\rho}(k)^2$ is the vector of canonical correlations G_t with respect to F_t.

Table 4: Testing the factors in Japan age-specific mortality rates by single age 20-89, 1970-2000

G_j	A(j)	M(j)	R^2(j)	NS(j)	$\hat{\rho}(k)^2$
G_1, Alcohol	**0.484**	**9.494**	**0.978 (0.962, 0.993)**	**0.023**	0.995 (0.992, 0.999)
G_2, Tobacco	0.484	6.893	0.804 (0.680, 0.928)	0.244	0.938 (0.895, 0.980)
G_3, Fat	**0.581**	**8.961**	**0.992 (0.986, 0.998)**	**0.008**	0.786 (0.652, 0.919)
G_4, Fruit & Veg	0.839	39.868	0.795 (0.666, 0.924)	0.258	0.310 (0.039, 0.580)
G_5, GDP	0.484	8.338	0.993 (0.988, 0.998)	0.007	-
G_6, Health exp	**0.548**	**7.130**	**0.992 (0.987, 0.998)**	**0.008**	-

A(j) is the frequency that $|\hat{\tau}_t(j)|$ exceeds the 5% asymptotic critical value. M(j) is the value of the test. R^2 is defined in 6, NS(j) defined in 5 and $\hat{\rho}(k)^2$ is the vector of canonical correlations G_t with respect to F_t.

Table 5: Testing the factors in Finland age-specific mortality rates by single age 20-89, 1970-2000

G_j	A(j)	M(j)	R^2(j)	NS(j)	$\hat{\rho}(k)^2$
G_1, Alcohol	0.774	22.746	0.433 (0.170, 0.696)	1.311	0.979 (0.964, 0.993)
G_2, Tobacco	0.839	11.285	0.897 (0.829, 0.966)	0.114	0.135 (-0.089, 0.359)
G_3, Fat	0.903	25.608	0.137 (-0.088, 0.361)	6.319	-
G_4, Fruit & Veg	0.871	19.263	0.792 (0.662, 0.922)	0.262	-
G_5, GDP	**0.548**	**10.386**	**0.904 (0.839, 0.968)**	**0.107**	-
G_6, Health exp	0.710	7.314	0.900 (0.834, 0.967)	0.111	-

A(j) is the frequency that $|\hat{\tau}_t(j)|$ exceeds the 5% asymptotic critical value. M(j) is the value of the test. R^2 is defined in 6, NS(j) defined in 5 and $\hat{\rho}(k)^2$ is the vector of canonical correlations G_t with respect to F_t.

Table 6: Testing the factors in Netherlands age-specific mortality rates by single age 20-89, 1970-2000

G_j	A(j)	M(j)	R^2(j)	NS(j)	$\hat{\rho}(k)^2$
G_1, Alcohol	0.871	42.800	0.568 (0.339, 0.797)	0.761	0.945 (0.907, 0.983)
G_2, Tobacco	**0.806**	**9.667**	**0.882 (0.804, 0.960)**	**0.134**	-
G_3, Fat	0.839	79.899	0.237 (-0.025, 0.498)	3.225	-
G_4, Fruit & Veg	0.742	30.208	0.699 (0.522, 0.876)	0.431	-
G_5, GDP	0.774	15.855	0.796 (0.668, 0.924)	0.256	-
G_6, Health exp	0.774	20.155	0.812 (0.693, 0.931)	0.232	-

A(j) is the frequency that $|\hat{\tau}_t(j)|$ exceeds the 5% asymptotic critical value. M(j) is the value of the test. R^2 is defined in 6, NS(j) defined in 5 and $\hat{\rho}(k)^2$ is the vector of canonical correlations G_t with respect to F_t.

Table 7: Testing the factors in Sweden age-specific mortality rates by single age 20-89, 1970-2000

G_j	A(j)	M(j)	R^2(j)	NS(j)	$\hat{\rho}(k)^2$
G_1, Alcohol	0.742	43.455	0.602 (0.384, 0.819)	0.662	0.953 (0.920, 0.985)
G_2, Tobacco	0.839	24.029	0.766 (0.622, 0.910)	0.306	0.340 (0.069, 0.611)
G_3, Fat	0.903	11.160	0.569 (0.340, 0.798)	0.759	-
G_4, Fruit & Veg	0.710	19.601	0.822 (0.709, 0.936)	0.216	-
G_5, GDP	**0.710**	**17.009**	**0.856 (0.762, 0.950)**	**0.169**	-
G_6, Health exp	0.903	46.286	0.608 (0.393, 0.823)	0.645	-

A(j) is the frequency that $|\hat{\tau}_t(j)|$ exceeds the 5% asymptotic critical value. M(j) is the value of the test. R^2 is defined in 6, NS(j) defined in 5 and $\hat{\rho}(k)^2$ is the vector of canonical correlations G_t with respect to F_t.

Table 8: Root mean square error for forecasting results, 2001-2009

	US	UK	Japan	Finland	Netherlands	Sweden
Lee Carter	**0.0043**	0.0051	0.0022	0.0056	0.0079	0.0028
Hyndman Ullah	0.0046	0.0058	0.0033	0.0076	0.0087	**0.0025**
Girosi and King	0.0053	**0.0044**	0.0023	0.0140	0.0087	0.0093
King and Soneji	0.0044	0.0077	**0.0014**	**0.0053**	**0.0075**	0.0044

Lowest RMSE for each column is emboldened. The order (K in equation 12) for Hyndman Ullah models are US 1, UK 3, Japan 1, Finland 6, Netherlands 7, Sweden 2. All forecasts for Hyndman Ullah are bias-adjusted. The smoothness parameters for the prior distributions in the Bayesian hierarchical models were $\sigma_a = \sigma_t = 0.30$ and $\sigma_{at} = 0.20$ across all countries.

Table 9: Root mean square error for forecasting results, 2001 and 2009

	US		UK		Japan		Finland		Netherlands		Sweden	
	2001	2009	2001	2009	2001	2009	2001	2009	2001	2009	2001	2009
Lee Carter	0.0011	0.0071	0.0029	0.0058	0.0036	0.0011	0.0028	0.0057	**0.0030**	0.0125	**0.0018**	0.0038
Hyndman Ullah	0.0009	0.0076	**0.0023**	0.0071	0.0020	0.0045	0.0045	0.0076	0.0031	0.0135	0.0025	**0.0030**
Girosi and King	0.0064	**0.0037**	0.0039	**0.0052**	**0.0016**	0.0028	0.0150	0.0155	0.0096	**0.0061**	0.0083	0.0093
King and Soneji	**0.0008**	0.0073	0.0038	0.0091	0.0016	**0.0009**	**0.0027**	**0.0053**	0.0032	0.0116	0.0035	0.0059

Lowest RMSE for each column is emboldened

Table 10: Root mean square error for fitting results, 1970-2000

	US	UK	Japan	Finland	Netherlands	Sweden
Lee Carter	0.0014	0.0017	0.0022	0.0051	0.0026	**0.0020**
Hyndman Ullah	0.0014	**0.0016**	0.0024	**0.0040**	**0.0019**	0.0021
Girosi and King	0.0042	0.0051	0.0029	0.0117	0.0073	0.0077
King and Soneji	**0.0012**	0.0024	**0.0017**	0.0049	0.0023	0.0026

Lowest RMSE for each column is emboldened

References

Auster, R., Leveson, I. & Sarachek, D. 1969, "The Production of Health, an Exploratory Study", *The Journal of Human Resources,* vol. 4, no. 4, pp. pp. 411-436.

Bai, J. & Ng, S. 2002, "Determining the Number of Factors in Approximate Factor Models", *Econometrica,* vol. 70, no. 1, pp. 191-221.

Bai, J. & Ng, S. 2006, "Evaluating latent and observed factors in macroeconomics and finance", *Journal of Econometrics,* vol. 131, no. 1-2, pp. 507-537.

Barker, D.J. 1992, *Fetal and infant origins of adult disease,* British Medical Journal.

Bell, W. 1997, "Comparing and assessing time series methods for forecasting age-specific fertility and mortality rates.", *Journal of Official Statistics,* vol. 13, pp. 279-303.

Bethune, A. 1997, "Unemployment and mortality" in *Health inequalities*, eds. F. Drever & M. Whitehead, HMSO, London.

Breitung, J. & Eickmeier, S. 2006, "Dynamic factor models", *Allgemeines Statistisches Archiv,* vol. 90, no. 1, pp. 27-42.

Booth, H., & Tickle, L., 2008, "Mortality modeling and forecasting: A review of methods", *The Australian Demographic & Social Research Institute.*

Cairns, A.J.G., Blake, D. & Dowd, K. 2006, "A Two-Factor Model for Stochastic Mortality with Parameter Uncertainty: Theory and Calibration", *Journal of Risk and Insurance,* vol. 73, no. 4, pp. 687-718.

Connor, G.,Hagmann, M. & Linton, O. 2011, " Efficient Semiparametric Estimation of the Fama-French Model and Extensions", forthcoming in *Econometrica*.

Currie, I.D. (2006), " Smoothing and forecasting mortality rates with P-splines", Talk given at the Institute of Actuaries, June 2006, available at: http:www.ma.hw.ac.uk/~iain/research/talks.html

Cutler, D., Deaton, A. & Lleras-Muney, A. 2006, "The Determinants of Mortality", *The Journal of Economic Perspectives,* vol. 20, no. 3, pp. 97-120.

Cutler, D.,M., Glaeser, E.L. & Rosen, A.B. 2009, "Is the U.S. Population Behaving Healthier?" in *Social Security Policy in a Changing Environment*, eds. J.R. Brown, J.B. Liebman & D.A. Wise, National Bureau of Economic Research, University of Chicago Press, , pp. 423-442.

Cutler, D.M. & Meara, E. 2004, "Changes in the Age Distribution of Mortality over the Twentieth Century" in *Perspectives on the Economics of Aging* University of Chicago Press, , pp. 333-366.

Freeman, D.G. 2003, "Is health care a necessity or a luxury? Pooled estimates of income elasticity from US state-level data", *Applied Economics,* vol. 35, no. 5, pp. 495-502.

Girosi, F. and G. King 2008, *Demographic Forecasting*, Princeton: Princeton University Press.

Herskind, A., McGue, M., Holm, N., Sørensen, T., Harvald, B. & Vaupel, J. 1996, "The heritability of human longevity: A population-based study of 2872 Danish twin pairs born 1870–1900", *Human genetics*, vol. 97, no. 3, pp. 319-323.

Hyndman, R.J. & Ullah, S. 2007, "Robust forecasting of mortality and fertility rates: a functional data approach", *Computational Statistics & Data Analysis*, vol. 51, no. 10, pp. 4942-4956.

Human Mortality Database [University of California, Berkeley (USA), and Max Planck Institute for Demographic Research (Germany).], Available at www.mortality.org or www.humanmortality.de [Accessed 2011, 10th May].

Iversen, L., Andersen, O., Andersen, P.K., Christoffersen, K. & Keiding, N. 1987, "Unemployment and mortality in Denmark, 1970-80.", *British Medical Journal (Clinical research ed.),* vol. 295, no. 6603, pp. 879-884.

King, G., and Soneji, S., 2011 "The Future of Death in America." Demographic Research 25: 1-38.

Lee, R.D. & Carter, L.R. 1992, "Modeling and Forecasting U. S. Mortality", *Journal of the American Statistical Association,* vol. 87, no. 419, pp. 659-671.

Leon, D.A. 2011, *Trends in European life expectancy: a salutary view*. International Journal of Epidemiology, vol. 40, no.2, pp. 271-277.

Miller, R.D.J. & Frech, H.E.I. 2000, "Is There a Link Between Pharmaceutical Consumption and Improved Health in OECD Countries?", *PharmacoEconomics,* vol. 18, no. 3, pp. 33-45.

Moscone, F. & Tosetti, E. 2010, "Health expenditure and income in the United States", *Health Economics,* vol. 19, pp. 1385-1403.

Murphy, Michael J. 2010 “Reexamining the dominance of birth cohort effects on mortality”. *Population and development review*, vol. 36, no.2. pp. 365-390.

Organization for Economic Cooperation and Development 2009, *OECD Health Data 2009: Statistics and Indicators for 30 Countries*, Paris.

Pampel, F. 2005, “Forecasting sex differences in mortality in high income nations: The contribution of smoking.”, *Demographic Research*, vol. 13, pp. 455–484.

Plat, R. 2009, "On stochastic mortality modeling", *Insurance: Mathematics and Economics,* vol. 45, no. 3, pp. 393-404.

Renshaw, A.E. & Haberman, S. 2006, "A cohort-based extension to the Lee–Carter model for mortality reduction factors", *Insurance: Mathematics and Economics,* vol. 38, no. 3, pp. 556-570.

Ruhm, C.J. 2004, *Macroeconomic Conditions, Health and Mortality*, National Bureau of Economic Research.

Schwartz, J. & Dockery, D.W. 1992, "Increased mortality in Philadelphia associated with daily air pollution concentrations", *American Review of Respiratory Disease,* vol. 145, pp. 600-604.

Thornton, J. 2002, "Estimating a health production function for the US: some new evidence", *Applied Economics*, vol. 34, pp. 59-62(4).

Vaartjes, I., O'Flaherty, M., Grobbee, D.E., Bots, M.L. & Capewell, S. 2011, "Coronary heart disease mortality trends in the Netherlands 1972–2007", *Heart*, vol. 97, no. 7, pp. 569-573.

Westerlund, J. 2007, "Testing for Error Correction in Panel Data", *Oxford Bulletin of Economics & Statistics,* vol. 69, no. 6, pp. 709-748.

Yang, S.S., Yue, J.C. & Huang, H. 2010, "Modeling longevity risks using a principal component approach: A comparison with existing stochastic mortality models", *Insurance: Mathematics and Economics,* vol. 46, no. 1, pp. 254-270.