BEYOND STATIC FACTOR MODELING

COLIN O'HARE[†], DECLAN FRENCH[†]

ABSTRACT. Using techniques from macroeconomic forecasting we propose a dynamic factor model of mortality that fits and forecasts mortality rates in a parsimonious fashion. We focus on the forecasting quality of this model and of existing models (Lee Carter model, the Cairns, Blake and Dowd model and the 4 factor model of Plat), and show that our model is superior. In particular over the longer term the dynamic factor model provides more accurate forecasts than existing approaches.

JEL Classification: C51, C52, C53, G22, G23, J11

Keywords and Phrases: Mortality, dynamic factor models, forecasting.

1. INTRODUCTION

In recent years there has been an increasing amount of attention put on the modeling of mortality risk as a significant risk that pension providers and insurance firms are exposed to. These developments have been driven in part by the introduction of more stringent regulation and historically low rates of interest and inflation. The latter has exposed longevity risk as being a significant risk in its own right and the development of innovative hedging products has allowed risk holders to unbundle longevity risk from the interest and inflation risks.

The introduction of Solvency 2 will lead to a change in the regulatory required solvency capital for insurers. Under Solvency 2 the so-called Solvency Capital Requirement (SCR) will be risk-based, and market values of assets and liabilities will be the basis for these calculations. This has further impacted on mortality research with papers from Plat 2010 for example identifying a mortality model that can capture one year value at risk (VAR).

The implications of tightening regulation and increasing longevity for insurers and pension providers are that modeling mortality rates more accurately, and in particular being able to

---

Date: Latest version: April 1, 2011.

[†]School of Management, Queen's University of Belfast, BT7 1NN, Belfast, United Kingdom. Tel.: +44 28 9097 2522; Fax: +44 28 9097 5156. Emails: c.ohare@qub.ac.uk and declan.french@qub.ac.uk.

Acknowledgements: We are indebted to Youwei Li, Donal McKillop and Michael Moore for many stimulating discussions. Thanks are also given to Dermot O'Reilly of the Centre of Excellence for Public Health and Shane Whelan of UCD, for their helpful comments.

forecast mortality rates with some level of confidence has become a matter of extreme importance. Accurate modeling and forecasting will enable advisors and decision makers to assess the financial implications of ageing populations more confidently.

Mortality rates show a stochastic characteristic and there is a significant amount of literature on modeling of mortality rates using stochastic methods. The Lee and Carter (1992) model was the first to identify a trend in the logarithm of mortality rates and modeled US male data using a one factor time series approach. Many innovations of the Lee-Carter model have been developed since including, Brounhs *et al.* (2002), Girosi and King (2002), Renshaw and Haberman (2003, 2006), Cairns *et al.* (2006), Currie *et al.* (2004), Currie (2006), Hári *et al.* (2008), Plat (2009). Each of these innovations focuses on one or more weaknesses of the Lee Carter model and improves the fitting quality as a result.

In many of these models the innovation has been to add factors to explain newly identified aspects of the mortality profile or to improve the flexibility of the model to fit wider age range. Examples of these innovations include modeling the cohort effect, as in Renshaw & Haberman (2003, 2006), adding a second period effect, as in the Cairns *et al* (2006), widening of the model to fit ages 20-89, Plat (2009). Many of these models focus on capturing additional variation in the data through the addition of period or cohort factors. Adding new factors in this way may improve the fit of the model but it also adds more complexity to the structure of the model which may or may not continue into the future. In this paper we generalise the basic one static factor model as in Lee Carter not by adding more factors but by using a dynamic factor approach. This approach has the advantage of exploiting the dynamics of the data when extracting factors. The most recent version of the technique Forni *et al.* (2005) also allows considerable flexibility to capture that part of the data not related to any common factors and provides the most efficient parameter estimates. We develop a parsimonious model which has sufficient factors to describe mortality variation adequately and which outperforms the existing models in a forecasting respect. We find even the most parsimonious facto model to be consistently superior across a range of developed countries especially when forecasting over a longer period.

The remainder of the paper is organized as follows. Firstly, in Section 2 the background to stochastic mortality modeling is reviewed, including an outline of the framework of Cairns *et. al.* (2008) used to assess the quality of a stochastic model. In section 3 we discuss the mortality data that we are using to carry out our analysis. Section 4 discusses the theory of dynamic factor modeling in the context of modeling mortality rates and develops the proposed dynamic factor model for mortality rates. In Section 5 we fit the model to mortality data from 6 different countries including U.K. and the U.S. and show the empirical performance of the dynamic factor model both from a fitting and forecasting perspective. We focus on the forecasting performance and analyse a range of dynamic factor models with differing numbers of dynamic and static factors to identify the optimal number of factors to explain the mortality forecast. We compare the results against other mortality models showing the superior performance of the dynamic factor model approach. Conclusions are drawn in Section 6.

2. BACKGROUND TO STOCHASTIC MORTALITY MODELING

The use of stochastic modeling methods for modeling mortality rates began in 1992 with the Lee Carter model. This model was developed to fit to U.S. male data between 1933 - 1987 and was used to forecast mortality rates up to 2065. The Lee-Carter model essentially describes the logarithmically transformed age-specific central rate of death $(y_{x,t})$ as a sum of an age-specific component that is independent of time (a_x), and the product of a time-varying parameter (κ_t, also known as the mortality index) that summarizes the general level of mortality and an additional age-specific parameter (b_x) that represents how rapidly or slowly (relative to the general mortality index) age-specific mortality varies. Mathematically, the Lee Carter model is given by:

$$y_{x,t} = ln(m_{x,t}) = a_x + b_x\kappa_t + \epsilon_{x,t}. \tag{1}$$

The final term, $\epsilon_{x,t}$, is the error term, which reflects the age-specific influences not captured by the model. Mortality forecasting is carried out using the model of the mortality index time series κ_t. The equation underpinning the Lee-Carter model is over-parameterized and requires some identifiability constraints to ensure a unique solution. The standard constraints in the case of the Lee Carter model are to take a_x as the arithmetic mean of the $ln(m_{x,t})$ over time, while the sums of b_x and κ_t are normalized to unity and zero, respectively. As all parameters on the right-hand side of equation 1 are unobservable, fitting the model by the ordinary least squares method is impossible. To overcome the situation, Lee and Carter employed a two-stage estimation procedure, which gives an exact solution. In the first stage, singular value decomposition (SVD) is applied to the matrix of $ln(m_{x,t})a_x$ to obtain estimates of b_x and κ_t. In the second stage, the time series of κ_t is re-estimated by solving for κ_t such that

$$\sum_x D_{x,t} = \sum_x E_{x,t} exp(a_x + b_x \kappa_t), \quad (2)$$

where $D_{x,t}$ is the number of deaths of lives aged x in year t, and $E_{x,t}$ is the exposure to risk of lives aged x in time t. The second stage ensures that the mortality rates fitted over the sample years reconcile the total number of deaths and the population age distributions. Finally, for forecasting purposes an autoregressive integrated moving-average (ARIMA) model is fitted to the κ_t parameter to model the dynamics of κ_t. The Box and Jenkins (1976) approach is employed to obtain a fitted ARIMA model from the empirical κ_t data. This model can be seen as a 1 factor model as the only time varying factor present in the model is κ_t.

Among many discussions of the Lee-Carter model, Cairns *et al.* (2006, 2008, and 2009) summarized the main disadvantages of the model. The model has one factor, resulting in mortality improvements at all ages being perfectly correlated (trivial correlation structure). For countries where a cohort effect[1] is observed in the past, the model gives a poor fit to historical data. The uncertainty in future death rates is proportional to the average improvement rate b_x which for high ages can lead to uncertainty being too low, since historical improvement rates have often been lower at high ages. Also, the model can result in a lack of smoothness in the estimated age effect b_x.

Despite the weaknesses of the Lee-Carter model its simple structure and the ease with which it can be fitted to mortality data has led to it being taken as a benchmark for future stochastic mortality models. There is a significant amount of literature developing additions to or modifications of the Lee Carter model aimed primarily at addressing the weakness of the model as outlined above. For example Brouhns *et al.* (2002), Lee and Miller (2001), Booth *et al.* (2002), Girosi and King (2005), De Jong and Tickle (2006), Delwarde *et al.* (2007) and Renshaw and Haberman (2003, 2006), Plat (2009). Many of these developments address one or more of the weaknesses of the Lee Carter model and taken as a collection of work they have progressively led to models with more and more factors which fit the data better but which have given little

[1]The cohort effect was identified in reports published by the Government Actuary Department (1995, 2001, 2002). These reports highlighted the existence of a cohort of the United Kingdom population who had experienced rapid mortality improvement. The generations born between 1925 and 1945 (centered on the generation born in 1931) experienced more rapid improvement than earlier and later generations. This feature had been noted for both males and females in the UK.

attention to the forecasting ability of the model. Before introducing our alternative approach we discuss some of these adaptations below.

Renshaw and Haberman (2006) modified the Lee-Carter model to capture the cohort effect. A factor κ_{t-x} was added to capture effects that could be attributed to the year of birth $(t-x)$.

$$y_{x,t} = \ln(m_{x,t}) = a_x + b_x^1 \kappa_t + b_x^2 \gamma_{t-x} + \epsilon_{x,t}, \tag{3}$$

where κ_t is defined as before and γ_{t-x} is a random cohort effect that is a function of the year of birth $(t-x)$.

The Renshaw-Haberman model does have a much better fit for countries such as the UK where a cohort effect has been identified, however it has been shown to suffer from a lack of robustness perhaps due to the presence of more than one local maximum in the likelihood function. Currie (2006) noted that if the model was fitted using data from 1961-2000 then the parameters showed qualitatively different characteristics to those obtained when fitting to data from 1981-2000. Furthermore, although the model incorporates the cohort effect, for most of the simulated mortality rates the correlation structure is still trivial with the simulated cohort parameters only being relevant for the higher ages at the far end of the projection.

Currie (2006). Following this analysis Currie (2006) introduced a simplified model that removed the robustness problem but at the expense of the fitting quality:

$$y_{x,t} = \ln(m_{x,t}) = a_x + \kappa_t + \gamma_{t-x} + \epsilon_{x,t}. \tag{4}$$

Cairns *et. al. 2009* observed that for England & Wales and United States data, the fitted cohort effect appeared to have a trend in the year of birth. This suggested that the cohort effect was compensating for the lack of a second age-period effect, as well as trying to capture the cohort effect in the data. This led them to introduce a two factor model of mortality,

$$\text{logit}(q_{x,t}) = \kappa_t^1 + \kappa_t^2 (x - \bar{x}) + \epsilon_{x,t}, \tag{5}$$

where $\bar{x}$ is the mean age in the sample range and (κ_t^1, κ_t^2) are assumed to be a bivariate random walk with drift. The two factors in this model were both period factors with no cohort effect allowed for. This was rectified in Cairns *et al.* (2009), namely capturing the cohort effect as an additional effect on top of the two age-period effects. All these models have multiple factors resulting in a non-trivial correlation structure which mirrors the reality that improvements in mortality rates are different for different age ranges. A further adaptation was also created allowing for the cohort effect to diminish over time. The main problem with these models arises from the fact that they were designed for higher ages and so ignored the modeling of mortality at the lower ages (for example the accident hump). Cairns *et al.* (2009) argue that the significant cost associated with mortality is at the older ages and thus their modeling focused on those ages. When using these models for full age ranges, the fit quality is relatively poor and the projections are biologically unreasonable.

In developing these models Cairns *et. al.* had increased the models complexity by adding up to 3 factors whilst at the same time reducing the flexibility by focusing on ages 50 and above.

An important criterion picked up by the Plat (2009), and following on from the models of Cairns *et. al.* was that the model should be applicable for a full age range. The Plat model addressed this weakness with a 4 factor model which maintained the good aspects of the existing models whilst leaving out the weaker features. The model took its beginnings from the Lee-Carter model and added factors to capture the second age-period effect, as per the Cairns *et al.* (2006) model and the cohort effect, as per the Renshaw Haberman (2006) model. The innovation in the Plat model was to then add a further period factor affecting only the lower ages and designed to allow the model to fit to the whole age range. The model specification is given by:

$$y_{x,t} = \ln(m_{x,t}) = a_x + \kappa_t^1 + \kappa_t^2(\bar{x} - x) + \kappa_t^3(\bar{x} - x)^+ + \gamma_{t-x} + \epsilon_{x,t}, \tag{6}$$

where the a_x is similar to that of the Lee-Carter model and makes sure that the overall shape of the mortality curve by age is reasonable, the κ_t^1 and κ_t^2 model the mortality rates as in the Cairns *et al.* (2006) and the κ_t^3 models the effects specific to the lower ages only, and the γ_{t-x} models the cohort effect.$(\bar{x} - x)^+$ takes the value $(\bar{x} - x)$ when this is positive and zero otherwise.

In 2008 Cairns *et al*, in the wake of a range of mortality models which all claimed to explain mortality data adequately, set out a range of criteria against which we could compare the quality of each model (from both a fitting and a forecasting perspective). The criteria set out the good qualities that you might expect a mortality model to have including:

- In a mortality model, mortality rates should be positive,*CRITERION A*
- The model should be consistent with historical data, *CRITERION B*

- Long-term dynamics under the model should be biologically reasonable, *CRITERION C*
- Parameter estimates and model forecasts should be robust relative to the period of data and range of ages employed, *CRITERION D*
- Forecast levels of uncertainty and central trajectories should be plausible and consistent with historical trends and variability in mortality data,*CRITERION E*
- The model should be straightforward to implement using analytical methods or fast numerical algorithms,*CRITERION F*
- The model should be relatively parsimonious,*CRITERION G*
- It should be possible to use the model to generate sample paths and calculate prediction intervals,*CRITERION H*
- The structure of the model should make it possible to incorporate parameter uncertainty in simulations, *CRITERION I*
- At least for some countries, the model should incorporate a stochastic cohort effect,*CRITERION J* and
- The model should have a non-trivial correlation structure. *CRITERION K*

Using these criteria we can determine how good a particular model is at fitting and forecasting mortality. The criteria set out the characteristics that you would want a good model of mortality rates to have and the range of models discussed above meet most of these to varying extents. The Plat model meet all of the criteria as they are set out above. However, it is debatable whether criterion G has been considered in the model of Plat (2009) (4 time series factors) and Cairns *et. al.* (3 time series factors) to fit to an age range from 50-89. A particular weakness in the existing literature in our view is the analysis of the forecasting quality of stochastic models. Specifically the analysis of medium and longer term forecasting ability of stochastic models has not been extensively reported. This characteristic is arguably the most important aspect of a good stochastic mortality model given the long term nature of longevity risk. In our view of the criteria above we believe that the most significant criteria that a mortality model should fulfill therefore is that of producing plausible, accurate forecasts. This is covered in part by criterion C-E and so has been addressed in part for the existing models but extensive empirical analysis has not been carried out to our knowledge. The success of a model is primarily in its ability to forecast and we focus on this aspect of our model when we compare results.

In the following analysis, to ease the notation we will use the naming convention established by Cairns *et al.* (2009). Table 1 sets out the names we will use for each of the models.

Table 1: The names of stochastic mortality models

Name	Model and Name
M1	Lee and Carter (1992)
	$\ln(m_{x,t}) = a_x + b_x\kappa_t + \epsilon_{x,t}$
M2	Renshaw and Haberman (2006)
	$\ln(m_{x,t}) = a_x + b_x^1\kappa_t + b_x^2\gamma_{t-x} + \epsilon_{x,t}$
M3	Currie (2006)
	$\ln(m_{x,t}) = a_x + \kappa_t + \gamma_{t-x} + \epsilon_{x,t}$
M5	Cairns *et al.* (2006)
	$\text{logit}(q_{x,t}) = \kappa_t^1 + \kappa_t^2(x-\bar{x}) + \epsilon_{x,t}$
M6	Cairns *et al.* (2009) with cohort effect
	$\text{logit}(q_{x,t}) = \kappa_t^1 + \kappa_t^2(x-\bar{x}) + \gamma_{t-x} + \epsilon_{x,t}$
M7	Cairns *et al.* (2009) with cohort and quadratic age effect
	$\text{logit}(q_{x,t}) = \kappa_t^1 + \kappa_t^2(x-\bar{x}) + \kappa_t^3((x-\bar{x})^2 - \sigma_x^2) + \gamma_{t-x} + \epsilon_{x,t}$
M9	Plat (2009)
	$\ln(m_{x,t}) = a_x + \kappa_t^1 + \kappa_t^2(\bar{x}-x) + \kappa_t^3(\bar{x}-x)^+ + \gamma_{t-x} + \epsilon_{x,t}$
M10	dynamic factor model
	$\ln(m_{x,t}) =$

Note: The we focus on comparisons with models M1, M5 and M9 to show the performance of our model relative to a range of existing stochastic factor models. The M10 is the model that we propose in this paper.

3. Mortality Data

The data that we use in this paper comes from the Human Mortality Database.[2] The data available for each country includes number of deaths $D_{x,t}$ and exposure to death $E_{x,t}$ for lives aged x last birthday during year t. We can use this to gain a proxy for the central mortality rate for lives aged x during year t as:

$$m_{x,t} = \frac{D_{x,t}}{E_{x,t}} \tag{7}$$

Due to the exponential nature of mortality rates we model the logarithmically transformed central mortality rates. We defined this as $y_{x,t} = \ln m_{x,t}$.

Data is available going back to the mid nineteenth century in some cases but we have restricted this study to data from 1950-2006 and to six countries representing a geographical spread around the globe. Specifically, we report performance on U.K., U.S., Netherlands,

[2]This can be found at http://www.mortality.org/. The Human Mortality Database (HMD) was created to provide detailed mortality and population data to researchers, students, journalists, policy analysts, and others interested in the history of human longevity. The database is maintained in the Department of Demography at the University of California, Berkeley, USA, and at the Max Planck Institute for Demographic Research in Rostock, Germany.

France, Japan, and Australia. These countries show similar characteristics with regard to economic development and so should be expected to demonstrate similar mortality characteristics.

Before looking at the theory of dynamic factor modeling we take a look at the mortality profiles to gain some information regarding the shape of mortality rates over time and over ages. We do this looking specifically at the U.K. mortality profile and the U.S. mortality profile.

The plots in figure 1 give the logarithmically transformed mortality rates for U.S. and U.K. lives aged 30, 50 and 70 over the years 1950 - 2000. The general trend in all cases is a downward one showing that mortality rates have decreased at these ages over time. The rate of improvement seems to differ for different ages and the levels of variation around the improvement differs for different ages as well.

If we look at the U.K. and U.S. mortality profiles by age for certain years in the past we can see how mortality changes over age have changed in past years. Figure 2 shows the U.K. and U.S. plots for the years 1950, 1970 and 1990 respectively. From this plot we can see that mortality experience over a lifetime has not changed significantly, the accident hump remains present and beyond that the logarithmically transformed mortality follows an approximately linear trajectory.

4. Proposed Dynamic Factor model

The Lee-Carter model can be treated as a Principal Component model with one component and other researchers have used this understanding to generalise the model by extracting a larger number of components (Yang et al., 2010). We propose to generalise the model instead by taking account of the dynamic structure of mortality data. The approach has been traditionally used in forecasting macroeconomic indicators and is increasingly also used in macroeconomic analysis (Bretung and Eickmeier, 2005).

Traditional factor analysis (of which Principal Component Analysis is a variant) is characterised by extraction of contemporaneous co-movements in a vector of standardised data $\mathbf{y_t} = [\mathbf{y_{1t}}, \ldots, \mathbf{y_{Nt}}]'$ and is of the form $\mathbf{y_t} = \mathbf{\Gamma f_t} + \mathbf{u_t}$ where $\mathbf{f_t} = [\mathbf{f_{1t}}, \ldots, \mathbf{f_{rt}}]'$,$\mathbf{u_t} = [\mathbf{u_{1t}}, \ldots, \mathbf{u_{Nt}}]'$ and $\mathbf{\Gamma} = [\lambda_\mathbf{1}, \ldots, \lambda_\mathbf{N}]$ for $\lambda_\mathbf{i} = [\lambda_\mathbf{i1}, \ldots, \lambda_\mathbf{iN}]$. The estimator for Γ is consistent for a fixed number of variables, N and $T \to \infty$ as long as $E(u_t u_t') = \Sigma = \sigma^2 I$.

Dynamic factor analysis as developed by Forni et al (2005) (hereafter referred to as FHLR) uses a *dynamic approximate* factor model by which we mean (i) an approximate factor model

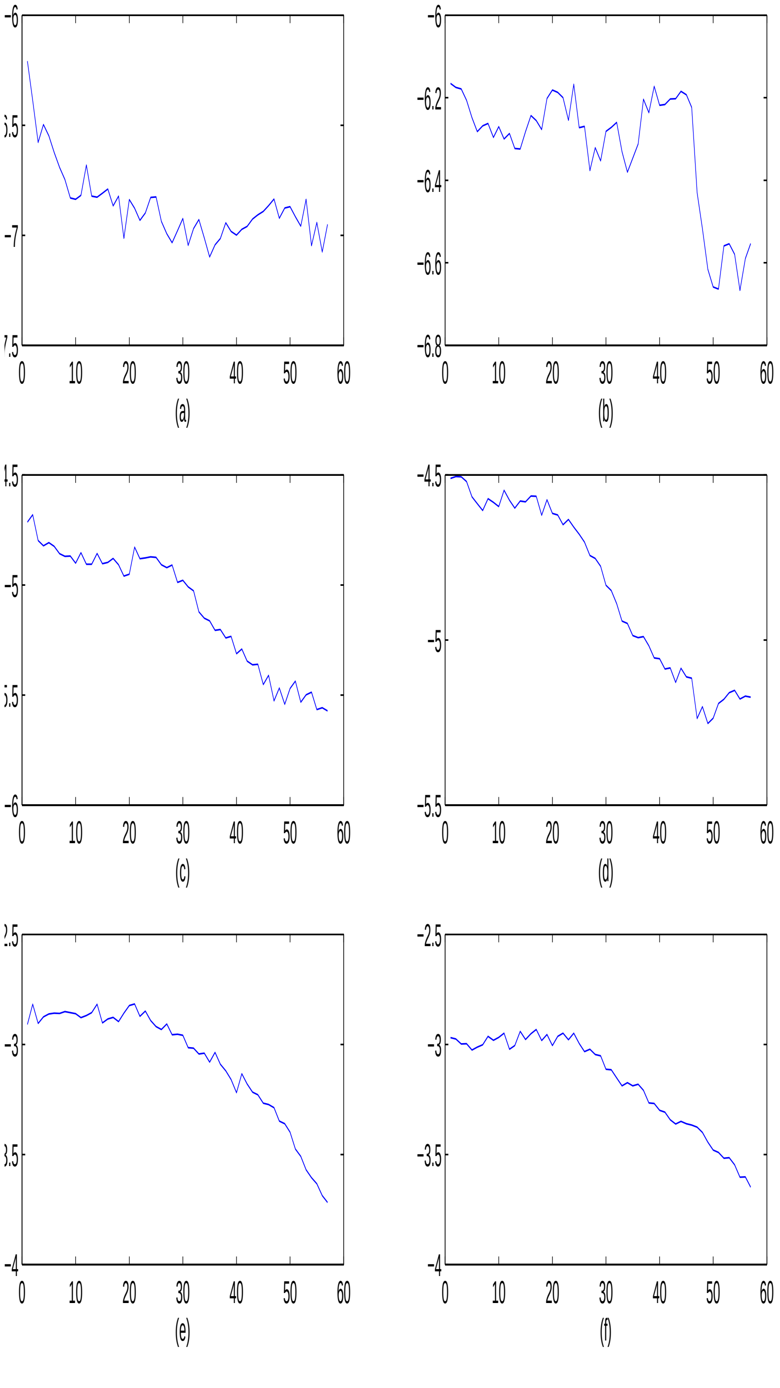

Figure 1: Logarithm of mortality by age for U.K. males aged (a) 30, (c) 50, and (e) 70, and U.S. males aged (b) 30, (d) 50, and (f) 70.

where the idiosyncratic errors are allowed to be weakly correlated and (ii) a dynamic factor model where the common factors are loaded onto the variables through a lag structure which is assumed finite. Their model represents a generalisation of the traditional static factor model

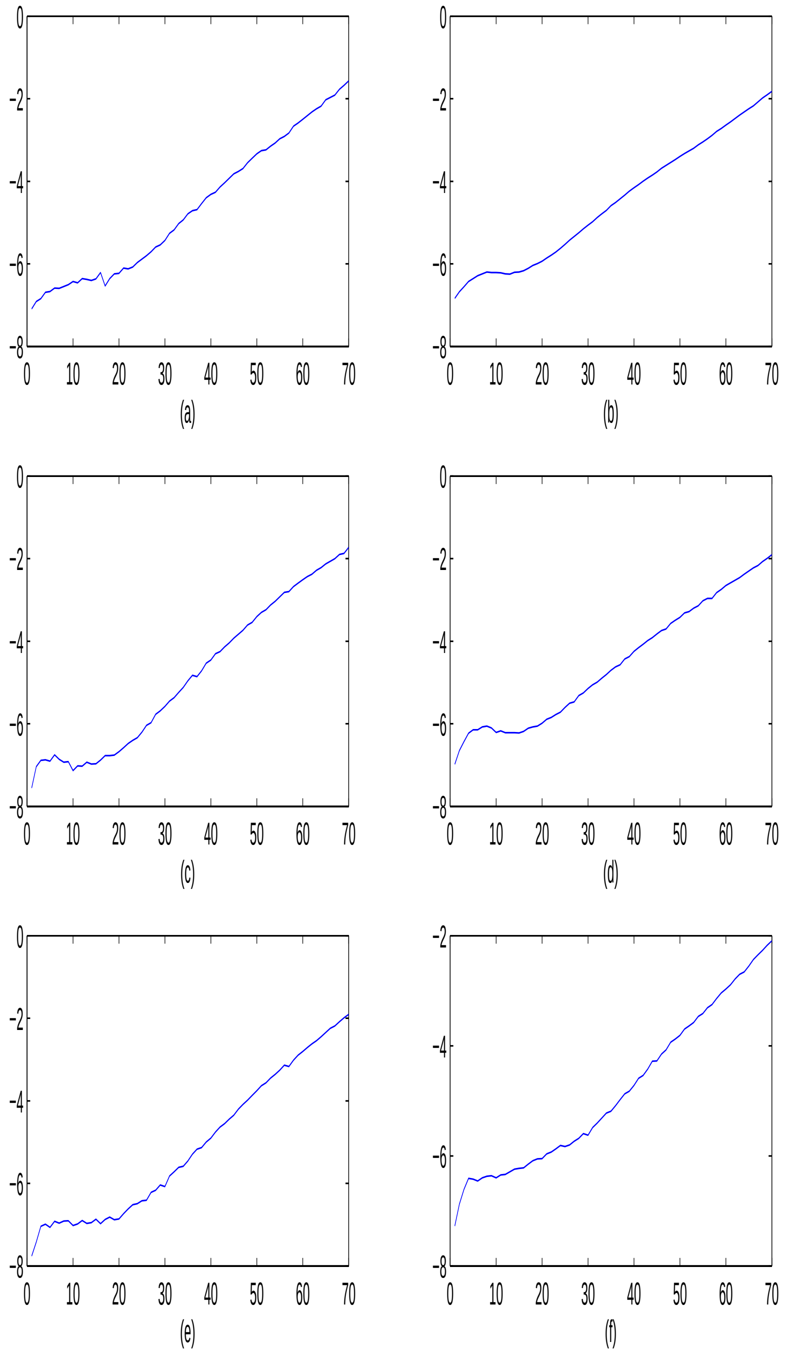

Figure 2: Logarithm of mortality by age for UK males during the year (a) 1950, (c) 1970, and (e) 1990, and US males during the year (b) 1950, (d) 1970, and (f) 1990.

building on the dynamic exact factor models of Geweke (1977) and Sargent and Sims (1976); the static approximate factor model of Chamberlain and Rothschild (1983) and the dynamic approximate factor model of Stock and Watson (2002). The latter bases projections on standard principal components extracted from the data. Hence, unlike FHLR, it thus does not fully exploit the dynamics of the data in calculating projections and does not downweight noisy data in calculating the common component. The FHLR procedure is now explained.

It is assumed that each of the stochastic processes $\mathbf{y_{it}}$ are stationary and that $\mathbf{y_t} = [\mathbf{y_{1t}}, \ldots, \mathbf{y_{Nt}}]'$ can be decomposed as the sum of an idiosyncratic component $\xi_\mathbf{t}$ and a common component $\chi_\mathbf{t}$. The common factors and idiosyncratic terms are assumed orthogonal.

$$\mathbf{y_t} = \chi_\mathbf{t} + \xi_\mathbf{t} \tag{8}$$

The common component $\chi_\mathbf{t}$ is driven by q factors with possibly different lags. Using the lag operator to indicate a linear combination of various lags of a factor f_{jt} we have

$$\chi_{it} = b_{i1}(L)f_{it} + b_{i2}(L) + \ldots + b_{iq}(L)f_{qt} or \chi_\mathbf{t} = \mathbf{B}(\mathbf{L})\mathbf{f_t} \tag{9}$$

then

$$\mathbf{y_t} = \chi_\mathbf{t} + \xi_\mathbf{t} = \mathbf{B}(\mathbf{L})\mathbf{f_t} + \xi_\mathbf{t}. \tag{10}$$

Estimation of the dynamic factors $\chi_\mathbf{t}$ is performed in the frequency domain. Any stationary variable y_t can be written in the frequency domain as a weighted sum of periodic functions

$$y_t = \mu + \int_0^\pi [\alpha(\omega)\cos(\omega t)d\omega + \delta(\omega)\sin(\omega t)]d\omega \tag{11}$$

The spectral density of y_t is then given by

$$\Sigma^y(\omega) = \frac{1}{2\pi}\Sigma_{j\to-\infty}^{\infty}\gamma_j e^{-i\omega j} \tag{12}$$

where γ_j is the j^{th} autocovariance (Hamilton, 1994). The frequency domain representation of the factor model is given by $\Sigma^y(\omega) = \Sigma^\chi(\omega) + \Sigma^\xi(\omega)$. The first step of the process provides an estimate of $\Sigma^\chi(\omega)$ and consequently $\Sigma^\xi(\omega)$ by extracting the first q principal components from an estimate of $\Sigma^y(\omega)$. This is known as *dynamic principal components analysis* (Brillinger, 1981). By taking inverse Fourier transforms of the respective spectral densities we can determine the autocovariance matrices Γ_k^χ and Γ_k^ξ and this information is used in the second step to generate a more efficient representation of the factor space than standard PCA. A linear combination of the y's is chosen so as to approximate the factor space i.e. since $\mathbf{a}\mathbf{y_t} = \mathbf{a}\chi_\mathbf{t} + \mathbf{a}\xi_\mathbf{t}$

we choose $\mathbf{a}$ such that $var(\mathbf{a}\xi_\mathbf{t})$ is minimised subject to the constraint $var(\mathbf{a}\mathbf{y_t}) = \mathbf{1}$. Using the covariance matrices $\hat{\Gamma}_0^\chi$ and $\hat{\Gamma}_0^\xi$ estimated in the first step this problem reduces to a generalised principal components problem of the form

$$\hat{\mathbf{Z}}_\mathbf{j}\hat{\mathbf{\Gamma}}_\mathbf{0}^\chi = \hat{\mathbf{v}}_\mathbf{j}\hat{\mathbf{Z}}_\mathbf{j}\hat{\mathbf{\Gamma}}_\mathbf{0}^\xi \tag{13}$$

. such that $\hat{\mathbf{Z}}_\mathbf{j}\hat{\mathbf{\Gamma}}_\mathbf{0}^\xi\hat{\mathbf{Z}}_\mathbf{j} = \mathbf{1}$ and $\hat{\mathbf{Z}}_\mathbf{i}\hat{\mathbf{\Gamma}}_\mathbf{0}^\xi\hat{\mathbf{Z}}_\mathbf{j} = \mathbf{1}$ for $i \neq j$. It can be shown that the generalised principal components of $\mathbf{y_t}$ are equivalent to the standard principal components of the transformed vector $\tilde{\mathbf{y_t}} = \left(\hat{\mathbf{\Gamma}}_\mathbf{0}^\xi\right)^{\mathbf{1/2}}\mathbf{y_t}$. Therefore this approach provides estimates of the common factors that weights the data in such a way as to suppress information from those variables with the highest idiosyncratic errors. This should provide more efficient estimates. Forecasts are then obtained from the generalised principal components

$$\hat{\mathbf{Z}} = \left(\hat{\mathbf{Z}}_\mathbf{1}, \ldots, \hat{\mathbf{Z}}_\mathbf{r}\right)' \tag{14}$$

giving

$$\chi_{T\hat{+}h|T} = \left[\hat{\mathbf{\Gamma}}_\mathbf{h}^\chi\hat{\mathbf{Z}}\prime\left(\hat{\mathbf{Z}}\hat{\mathbf{\Gamma}}_\mathbf{0}\hat{\mathbf{Z}}\prime\right)^{-\mathbf{1}}\right]\left[\hat{\mathbf{Z}}\mathbf{x_t}\right]. \tag{15}$$

To apply this approach to the context of mortality modelling, let $D_{x,t}$ denote the number of deaths in a population at age x and period t and $E_{x,t}$ the corresponding exposure. We define mortality rate

$$y_{x,t} = \ln(\frac{D_{x,t}}{E_{x,t}}). \tag{16}$$

To make the data stationary we take the first difference $\Delta y_{x,t}$ and normalise. Forecasts are then generated by cumulatively summing the estimated change in (log) mortality as

$$\hat{y}_{x,T+h|T} = y_{x,T} + \hat{\Delta} y_{x,T+1|T} + \ldots + \hat{\Delta} y_{x,T+h|T} \tag{17}$$

The prediction accuracy at horizon h is evaluated using the mean absolute percentage error (MAPE) over all ages from 20 to 89 inclusive

$$MAPE_y^h = \frac{1}{70 * h}\Sigma_{x=20}^{x=89}\Sigma_{k=1}^{h}\frac{|\hat{y}_{x,T+h|T} - y_{x,T+h}|}{y_{x,T+h}} \tag{18}$$

In the context of mortality modelling, the FHLR dynamic approximate factor model represents an improvement over simply extracting static factors from the data as

- (i) By being approximate it allows the component of age-specific mortality rates not explained by common factors to be correlated across a subset of ages (cross-correlation) and across time (serial correlation). This is reasonable under the plausible assumption that there are unidentified drivers affecting mortality only for subsets of the population e.g. improvements in geriatric care for the very elderly or alcohol consumption among young men.
- (ii) By being dynamic it allows for a small number of core factors that are driving mortality changes for every age but with possibly different lags. This would capture cohort effects due to early life effects, educational improvements or the young being quicker to adapt behaviour.
- (iii) By projecting on an estimate of the common component it fully exploits the dynamics of the data by using all contemporaneous and lagged covariances in death rates across the ages considered.
- (iv) By using generalised principal components to extract the common factor it down-weights data which is not associated with the common trend and this should provide more efficient estimates.

5. Model Justification and empirical results of fitting and forecasting

In this section we begin by discussing the characteristics of data that we would expect a dynamic factor model to be able to capture, we subsequently demonstrate the presence of these characteristics in our mortality data. We follow that with an empirical comparison of the existing models against the dynamic factor model that we propose. We check the robustness of our findings across forecasting different forecasting periods. Using a range of country data we firstly fit the models to the years 1950 - 2000 and ages 20-89 and forecast for the 6 year period from 2001 to 2006. Secondly, we repeat the fitting and forecasting analysis but this time fitting to 1950-1990 and forecasting from 1991 to 2006. The data sets used are male mortality data during 1950-2006 for United Kingdom (U.K.), United States (U.S.), Australia (AUS), France (FRA), The Netherlands (NTH) and Japan (JPN).

When assessing the forecasting quality we compare the in sample forecasts with actual mortality for 2001-2006 and 1991-2006 respectively. The model forecasting ability is compared using the mean absolute percentage error (MAPE) as defined in the previous section.

5.1. **Justification of methodology.** In section 3 we described the criteria we felt were important for a mortality forecasting model. We next describe why we think dynamic factor modelling is a suitable approach. D'Agostino and Giannone (2007) state that a dynamic factor model would give a reasonable representation of the data if the data displayed the following characteristics, (i) strong co-movements in the data,(ii) a rich dynamic structure, and (iii) variation in the amount of idiosyncratic variance.

In the first instance therefore, we will examine U.S. mortality data for those aged 20-89 over the period 1950-2000 for these patterns taking each one in turn. All of the mortality models reviewed in section 2 contain a common factor component (e.g.κ_t in Lee-Carter (M1) or κ_t^1,κ_t^2,and κ_t^3 in Plat(M9)) reflecting the consistent finding that death rates for different ages tend to follow common secular trends. It should therefore be clear that there should be co-movements in the data over time. This can also be seen in the third row of table 2 where a standard principal components analysis (PCA) has been performed on the data which has first been transformed to be stationary[3]. In PCA, the data is decomposed into a small number, K of orthogonal common factors $\kappa_t(j)$ as follows:

$$ln(m_{x,t}) = \alpha_x + \sum_{j=1}^{K} \beta_x(j)\kappa_t(j) + \epsilon_{x,t} \tag{19}$$

If the data are contemporaneously cross-correlated then extracted factors should capture much of the variation in the data. PCA is carried out by calculating the eigenvalues and eigenvectors of the covariance matrix of the death rates. The percentage of variation explained by the model is then given by the sum of eigenvalues divided by the number of variables. This figure is given in the third row of Table 2 for models with one through to ten factors and in all cases the amount of variance explained by the common factors is substantial. The first principal component explains 29% of the variation in the data while ten factors explain 72% of the variation. Mortality at each age is therefore strongly associated with trends over time that are common to all ages.

[3]The natural logarithm of the data was first-differenced and standardised. A full account of PCA can be found in Joliffe, I.T. (1986), Principle components Analysis, Springer, New York

To check if there is a rich dynamic structure we also give the percentage variation explained by one to ten dynamic factors in the fourth row of table 2.[4]. It can be seen, first of all, that a much smaller number of dynamic factors are required to capture the variation in the data e.g. three dynamic factors capture approximately the same amount of variance as nine static factors. If we extracted factors until 95% of the total variance of the data was explained[5] we would require ten dynamic factors and twenty-eight static factors. This indicates that it may be possible to parsimoniously represent large numbers of static factors by a linear combination of lags of a small number of dynamic factors justifying the representation in equation 10.

Table 2: Table 1: Percentage variance explained by factors

Number of Factors	1	2	3	4	5	6	7	8	9	10
Static	29.3%	42.2%	47.7%	52.4%	56.5%	60.3%	63.8%	67.1%	69.8%	72.3%
Dynamic	41.3%	58.6%	69.3%	77.6%	83.6%	88.2%	91.4%	93.4%	94.8%	95.8%

We will now show that mortality rates for each age have different levels of association with these common factors. Allowing for one static factor as in the Lee-Carter model, the percentage of the variation in mortality rates explained by the common factor or communality is calculated for each age. Results are graphed below in Figure 3. The variation is clear: while 52% of the variation in mortality rates at age 33 over time can be explained by this one common factor only 11% of the variation in mortality rates at age 27 is captured. The variation in signal to noise observed indicates that the FHLR dynamic factor approach of correcting estimates of common factors by suppressing information from those variables with the highest idiosyncratic errors will provide more efficient results. This is analogous to corrections for heteroscedasticity in ordinary least squares by using weights inversely proportional to the error variance[6].

5.2. **Results.** From the analysis above, it would appear then that dynamic factor modelling provides a reasonable approach to mortality forecasting. In the first instance we use a model with a minimal amount of factors: one dynamic factor (q=1 in equation 8) and one static factor

[4]As in D'Agostino and Giannone (2007), we use $\frac{trace(\hat{\Gamma}_0^{\chi})}{trace(\hat{\Gamma}_0)}$ as our measure of the percentage of variance explained where $\hat{\Gamma}_0^{\chi}$ is estimated by the first q dynamic factors

[5]This rule is advocated by Joliffe (1986) to determine when to stop extracting factors.

[6]See Gujarati, D. N. and Porter, D.C. (2009) Basic Econometrics 5^{th} edition, McGraw Hill pp 371-373 or any other standard econometrics text

Figure 3: Communality (i.e. percentage of variation in mortality rate explained by the common factor) by age

(r=1 in equation 8). Using 1950-2000 U.S. males mortality data to forecast death rates for 2001-2006, we get projected mortality rates for those aged 20-89 and compare them to the actual mortality rates. Generally the model provides a good fit to the transformed data. The correlation coefficient between the actual and estimated first differences of log mortality rates over all ages is 0.52. However, as we have argued above, an exact fit to the data is not necessarily helpful in improving the accuracy of predictions.

Forecasting results are given in Table 3 below for those aged 70 (column 3) and show forecasts are reasonably close to the actual values (column 2) initially but overshoot slightly as projections get further into the future. Predictions can be improved by increasing the number of static and dynamic factors. Using five static and two dynamic factors (column 4), forecasts capture the decline in mortality rates for 70-year old men marginally better than in the simpler one factor model. Comparing these results to other existing models we see that despite their shortcomings the dynamic factor models perform best. Predictions from all three existing models overestimate mortality rates with the Cairns model providing particularly poor predictions. The Plat and Lee-Carter models forecast a decline in mortality but not of the correct magnitude. Nevertheless they are not dramatically worse than the dynamic model for this particular age.

The accuracy of forecasts from the various models varies by each age. Looking at mortality rates for all ages together, the mean absolute percentage error (MAPE) described in section ?? has been calculated in the last row. Predictions of dynamic factor models can be improved by increasing the number of static and dynamic factors. A discussion of ex-ante choices will be left until later in subsection ??. An ex-post examination of the MAPE for the number of dynamic factors $q = 1, \ldots, 10$ and the number of static factors $r = 1, \ldots, 10$ indicates that a combination of $q = 2$ and $r = 5$ gives the lowest MAPE and hence the most accurate forecasts. We can see that both dynamic models have lower forecast error than existing models. The Cairns model which was not designed for the full age range, but which was based on extensions of the Lee Carter approach is by far the most inaccurate of the models that we have tested in this paper. The improvement in accuracy of the dynamic models over the other two models can be seen to be substantial and of the order of 2-3%.

Table 3: US Projected mortality rates for males aged 70 and forecast error (MAPE) for all ages, 2001-2006.

	Actual	M10 (q=1,r=1)	M10a (q=2,r=5)	M1 (1992)	M5 (2009)	M9 (2009)
2001	0.0297	0.0302	0.0301	0.0309	0.0315	0.0308
2002	0.0297	0.0298	0.0296	0.0302	0.0320	0.0299
2003	0.0288	0.0295	0.0294	0.0307	0.0324	0.0295
2004	0.0272	0.0292	0.0290	0.0294	0.0328	0.0294
2005	0.0273	0.0289	0.0286	0.0296	0.0333	0.0291
2006	0.0260	0.0286	0.0284	0.0288	0.0337	0.0291
All ages MAPE		5.12%	4.84%	8.04%	16.39%	7.39%

In order to verify that dynamic factor modelling is an improvement over existing methods two robust checks are carried out: the first across country and the second over a longer-term. The error for forecasts for 2001-2006 based on data on male mortality rates for those aged 20-89 over the period 1950-2000 from a broad range of large developed countries is given in Table 4. The combination of static and dynamic factors giving the lowest MAPE was determined for each country separately and is labelled 'Dynamic best'. Results for the U.S. from Table 3 have also been included as a reference. Of the existing models, the Plat model is the most accurate (except for Australia) perhaps due to its greater number of factors and that it contains an adjustment to specifically fit to the full age range. On the other hand, the Cairns model is the poorest by far. The dynamic factor model even with the simplest specification of one static and one dynamic factor gives more accurate forecasts for all countries except the Netherlands and France. The improvement is generally of the order of 1-3% over the best of the other models. A straight column sum of the MAPE for each method indicates that if one method were to be chosen to apply in all instances the dynamic factor models would give the most accurate results consistently. Also it can be seen that using a more complicated dynamic factor specification over the default of the simplest model gives a relatively small improvement in accuracy.

Table 4: Forecast error (MAPE) for all ages 20-89, 2001-2006.

	M10 (q=1,r=1)	M10a (q=2,r=5)	M1 (1992)	M5 (2009)	M9 (2009)
USA	5.12%	4.84%	8.04%	16.39%	7.39%
		(q=2 r=5)			
UK	7.91%	7.54%	9.26%	24.03%	9.56%
		(q=3 r=6)			
NTH	12.24%	11.00%	10.62%	22.74%	10.00%
		(q=3 r=9)			
FRA	7.51%	7.32%	14.10%	19.80%	6.87%
		(q=3 r=3)			
AUS	10.94%	10.83%	11.46%	31.02%	13.22%
		(q=2 r=2)			
JPN	6.44%	5.89%	12.35%	21.25%	8.82%
		(q=2 r=10)			
Average MAPE	8.35%	7.90%	10.97%	22.53%	9.31%

Repeating this analysis over a longer time frame provides even more convincing evidence of the superiority of the dynamic factor approach. The error for forecasts for 1990-2006 based on the period 1950-1990 for the same countries is given in Table 5. In contrast to the situation in short-run forecasts above, we see here that multi-factorial models become a liability in the longer-run. The Cairns model as before produces completely wayward predictions. Except in the case of the UK, the Lee-Carter model is more accurate than the Plat model with a very substantial reduction in forecast error of 6-7% for the U.S., Australia and France. This highlights the problem of overfitting models to data. Whereas many papers examine the fit of a model as an indicator of suitability (refs ????) we see that the application of minor adjustments to models may ensure a better data fit and provide reasonable short run forecasts but in the longer-run introduce unsustained relationships and compromise forecastibility. The simple one factor dynamic model generally improves on the Lee-Carter with an almost 4% reduction in the MAPE for the UK, Netherlands, France and Australia. The improvement in the case of Japan is less substantial (0.38%) while the dynamic factor forecast error for the US is a minor

0.36% worse than the Lee-Carter forecast. We can improve the dynamic factor predictions further by incorporating a larger number of factors. Once again the improvement is not very large (0.22%-1.87%) considering the loss of parsimony although now the dynamic model has a lower forecast error for all countries. Also, comparing the number of static (r) and dynamic factors (q) in the dynamic model with the lowest MAPE here with the equivalent in Table 5, it can be seen that the best forecasts over longer periods generally require more factors. This point is explored further in the next section.

Table 5: Forecast error (MAPE) for all ages 20-89, 1991-2006.

	M10 (q=1,r=1)	M10a	M1 (1992)	M5 (2009)	M9 (2009)
USA	10.82%	9.44% (q=6, r=8)	10.46%	27.38%	17.10%
UK	12.18%	10.98% (q=1, r=5)	17.05%	48.48%	15.97%
NTH	14.94%	13.07% (q=2,r=10)	18.83%	41.35%	20.28%
FRA	13.40%	13.18% (q=7,r=5)	16.91%	46.57%	22.61%
AUS	16.40%	15.98% (q=1, r=10)	20.15%	62.27%	26.89%
JPN	14.42%	14.01% (q=6,r=4)	14.80%	50.31%	16.16%

5.3. **Choosing the number of static and dynamic factors.** In the previous section, an examination of forecasts for six large developed countries showed that a simple dynamic factor with one dynamic and one static factor surpassed existing models in four out of six cases over a six-year period and in five out of six cases over a sixteen period. As a first step, it therefore seems reasonable to advocate using dynamic factor modelling with a default of q=1,r=1 as generally the most accurate approach to forecasting. Increasing the number of factors, q and r, decreases the MAPE by 0.11%- 1.23% over the short run and 0.22%-1.87% over a longer timeframe although this has only been assessed ex-post. This section examines whether there is a simple ex-ante rule-of-thumb for choosing q and r to improve on the forecasting power of the simplest dynamic factor model.

Calculating the MAPE for the number of dynamic factors $q = 1, \ldots, 10$ and the number of static factors $r = 1, \ldots, 10$ and ranking models on the size of the forecast error (Table 6) we see that the simplest model is certainly nowhere near the most accurate except in the case of Australia. To determine the combination of dynamic and static factors that was generally most accurate in all situations, we used the ersatz approach of first ranking the MAPE for each combination of q and r for each country and then summing the ranks across country (q=Q, r=R for some Q,R fixed for every country) . The results which are not shown indicate that generally for short range forecasts a model with $3 \leq r \leq 6$ and $q \leq 1$ would give best results while for longer-run forecasts r should be at least 4 and q should be either small ($q \leq 2$) or large ($q \geq 7$). The simple model $(q = 1, r = 1)$ would be surpassed by almost 90% of other combinations according to this measure.

Less heuristically, we can use statistical approaches to terminating factor extraction. Standard stopping rules for determining the number of static principal components,r, in data which were developed in psychological or sociological studies are not applicable in datasets with large number of variables (Breitung and Eickmeier, 2005). Bai and Ng (2002) have developed a statistical test based on information criteria with appropriately chosen penalties to determine the number of static factors in this situation. In later work (Bai and Ng, 2007), they also develop a rule for estimating the number of dynamic factors,q , based on a VAR in r static factors. Applying these rules to our data, leads to the choice of $q = 1, r = 1$ (UK,NTH, AUS) or $q = 2, r = 2$ (USA, FRA,JPN). These combinations do not correspond to the most accurate models and using the ersatz rule above their accuracy would be surpassed by almost one-quarter of models with q and r fixed for every country. Perhaps this is due to the fact that these statistical rules are based on the fit of factor models to the data over which the model is estimated and do not necessarily relate directly to predictive power. Nevertheless, these rules have a statistical basis and offer a slight improvement on blindly applying the simplest model. This topic has only been preliminarily investigated here and requires more substantial further research.

6. Conclusions and further discussion

In this paper we have provided an alternative approach to modeling and forecasting mortality rates. We have chosen to focus on the forecast quality of the model using actual data and as a result conclude that existing multifactorial models give a good forecasting performance in the

Table 6: Rank of simplest dynamic factor model (q=1, r=1) compared to multifactorial dynamic factor models (100 combinations of $q = 1, \ldots, 10$, $r = 1, \ldots, 10$).

1950-2000	1950-1990	
USA	83	92
UK	76	76
NTH	68	83
FRA	67	37
AUS	4	8
JPN	99	41
Sum	397	457

short run but not in the long run. Our dynamic factor approach provides a superior forecast, especially so over the longer term. Even with the simplest of specifications the dynamic model outperforms the the existing models of mortality.

Further work may be carried out to examine the forecasting power of the approach over specific age ranges rather than the full age range, and over a longer forecast horizon. Identifying the number of static and dynamic factors giving optimal forecast needs to be further examined. In keeping with the current direction of research in this area incorporating exogenous determinants also appears to be a a worthwhile piece of work that we would like to carry out.

REFERENCES

[1] Brouhns, N., Denuit, M., and Vermunt, J.K., 2002. *A Poisson log-bilinear approach to the construction of projected lifetables*. Insurance: Mathematics and Economics 31(3), 373-393.

[2] Booth, H., Maindonald, J., Smith, L., 2002. *Applying Lee-Carter under conditions of variable mortality decline.* Population Studies 56, 325-336

[3] Cairns, A.J.G., 2000. *A discussion of parameter and model uncertainty in insurance*. Insurance: Mathematics and Economics 27(3), 313-330.

[4] Cairns, A.J.G., Blake, D., and Dowd, K., 2006. *A two-factor model for stochastic mortality with parameter uncertainty: Theory and calibration*. Journal of Risk and Insurance 73, 687-718.

[5] Cairns, A.J.G., Blake, D., Dowd, K., Coughlan, G.D., Epstein, D., and Khalaf-Allah, M., 2008. *Mortality density forecasts: An analysis of six stochastic mortality models*. Pensions Institute Discussion Paper PI-0801, April.

[6] Cairns, A.J.G., Blake, D., Dowd, K., Coughlan, G.D., Epstein, D., Ong, A., and Balevich, I., 2009. *A quantitative comparison of stochastic mortality models using data from England & Wales and the United States*. North American Actuarial Journal, 13(1), 1-35.

[7] Continuous Mortality Investigation (CMI), 2006. *Stochastic projection methodologies: Further progress and P-Spline model features, example results and implications*. Working Paper 20.

[8] CMI, 2007. *Stochastic projection methodologies: Lee- Carter model features, example results and implications*. Working Paper 25.

[9] CMI, 1990. *Continuous Mortality Investigation Reports,1990*. Institute of Actuaries and Faculty of Actuaries.

[10] CMI, 1999. *Continuous Mortality Investigation Reports,1999*. Institute of Actuaries and Faculty of Actuaries.

[11] Coughlan, G.D., Epstein, D., Ong, A., Sinha, A., Balevich, I., Hevia Portocarrera, J., Gingrich, E., Khalaf-Allah, M., and Joseph, P., 2007. *LifeMetrics: A toolkit for measuring and managing longevity and mortality risks*. Technical Document (JPMorgan, London, 13 March). Available at www.lifemetrics.com.

[12] Currie, I.D., Durban, M., and Eilers, P.H.C., 2004. *Smoothing and forecasting mortality rates*. Statistical Modelling 4, 279-298.

[13] Currie, I.D., 2006. *Smoothing and forecasting mortality rates with P-splines*. Presentation to the Institute of Actuaries. Available at: http://www.ma.hw.ac.uk/ iain/research.talks.html.

[14] Delwarde, A., Denuit, M., and Eilers, P., 2007. *Smoothing the Lee & Carter and Poisson log-bilinear models for mortality forecasting: A penalized log-likelihood approach.* Statistical Modelling 7, 29-48.

[15] De Jong, P., and Tickle, L., 2006. *Extending the Lee & Carter model of mortality projection.* Mathematical Population Studies 13, 1-18.

[16] Dowd, K., Cairns, A.J.G., Blake, D., Coughlan, G.D., Epstein, D., and Khalaf-Allah, M., 2008. *Backtesting stochastic mortality models: An ex-post evaluation of multi-period-ahead density forecasts*. Pensions Institute Discussion Paper PI-0803,March.

[17] Girosi, F., and King, G., 2005, *A reassessment of the Lee-Carter mortality forecasting method*, Working Paper, Harvard University.

[18] Hári, N., Waegenaere, A., Melenberg, B. and Nijman, T., 2008. *Estimating the term structure of mortality*. Insurance: Mathematics and Economics 42, 492-504.

[19] Human Mortality Database (HMD), 2004, University of California, Berkeley (USA), and Max Planck Institute for Demographic Research, Rostock (Germany). http://www.mortality.org or http://www.humanmortality.de.

[20] Lee, R.D., and Carter, L.R., 1992. *Modeling and forecasting U.S. mortality*. Journal of the American Statistical Association 87, 659-675.

[21] Lee, R. D., and T. Miller (2001) *Evaluating the performance of the Lee-Carter method for forecasting mortality*. Demography 38: 537-549.

[22] Plat, R., 2009. *On stochastic mortality Modeling*. Insurance: Mathematics and Economics 45(3), 393-404.

[23] Renshaw, A.E., Haberman, S., 2003. *Lee-Carter mortality forecasting with age-specific enhancement*. Insurance: Mathematics and Economics 33, 255-272.

[24] Renshaw, A.E., Haberman, S., 2006. *A cohort-based extension to the Lee-Carter model for mortality reduction factors*. Insurance: Mathematics and Economics 38, 556-70.