Five different distributions for the Lee–Carter model of mortality forecasting: a comparison using GAS models

César Neves[1] Cristiano Fernandes[2] Henrique Hoeltgebaum[3]

Abstract

This paper extends the well-known Lee-Carter model used for forecasting mortality rates by utilizing a new class of time series models, known as Generalized Autoregressive Score (GAS) or Dynamic Conditional Score (DCS) models. This framework can be used to derive a wide range of non-Gaussian time series models with time varying coefficients and has shown to be very successful in financial applications. In this paper we propose five probability models (Poisson, binomial, negative binomial, Gaussian and beta) based on the GAS framework to estimate the Lee-Carter parameters and dynamically forecast the mortality rates using a single unified step. The models are applied to the mortality rates time series for the male population of the United States, Sweden, Japan and the UK. Diagnostic tests are performed on quantile residuals, model selection is made via AIC and predictive accuracy of the models is compared using the Diebold-Mariano test. We conclude that, amongst the proposed models, the negative binomial extension of the Lee-Carter model is the most appropriate for forecasting mortality rates.

JEL: C22

Keywords: GAS models, mortality rates, Lee-Carter model, forecasting, observation-driven time series models

[1] Centre for Research on Insurance Economics (CPES), Brazilian School of Insurance; Department of Statistics and Actuarial Sciences, Rio de Janeiro State University (UERJ); and Brazilian Insurance Supervisor (SUSEP), Rio de Janeiro, Brazil. cesar.neves@susep.gov.br.

[2] Department of Electrical Engineering, Pontifical Catholic University of Rio de Janeiro (PUC-Rio), Rio de Janeiro, Brazil. cris@ele.puc-rio.br.

[3] Department of Electrical Engineering, Pontifical Catholic University of Rio de Janeiro (PUC-Rio), Rio de Janeiro, Brazil. hhhelfer@ele.puc-rio.br.

1. Introduction

The choice of a suitable model for forecasting mortality rates is essential when evaluating the solvency of life insurers. Mortality forecasting is used to form the best estimate of future commitments to policyholders and assess the required level of risk-based capital. Given the observed downward trend in mortality rates over time for many industrialized countries, it is important to adopt statistical models that can accurately and robustly predict the longevity gains. The Lee-Carter (1992) model is one of the most well-known models for forecasting mortality rates. With this model, the time series of the log mortality rates of each age is described by an age-specific intercept plus a common trend for all age groups multiplied by an age-specific coefficient. The model employs singular value decomposition (SVD) and least squares (OLS) to extract both the common trend and all age-specific parameters. ARIMA models are typically used to extrapolate the common trend, making it possible to forecast mortality rates for any age group.

Many extensions of the Lee-Carter model have been proposed, as shown in Pitacco et al. (2009). Brouhns et al. (2002) presented an improvement on the Lee-Carter method, considering that the main weakness of the OLS estimated by SVD is that the errors are assumed to be homoscedastic. They adapted the Lee–Carter model by supposing that the number of deaths follows a Poisson distribution (which is intrinsically heteroscedastic), with parameters estimated using an iterative method for log-linear models with bilinear terms. The authors also use ARIMA models to forecast the mortality rates. Renshaw and Haberman (2006) also assumed that the number of deaths is Poisson distributed and incorporated cohort effects into the Lee-Carter methodology. In contrast, Cossette et al. (2007) and Haberman and Renshaw (2008) explored a binomial version of the Lee-Carter model. Delward et al. (2007) considered the over-dispersion present in the mortality data and assumed that the number of deaths follows a negative binomial distribution, extending the Lee-Carter model. De Jong and Tickle (2006) used the Kalman filter to estimate the Lee-Carter model, assuming that the disturbance terms are normally distributed. Chen et al.(2014) presented a dynamic multi-population mortality model based on a two-factor copula whose parameters are assumed time-varying via the Generalized Autoregressive Score (GAS) updating mechanism.

The main contribution of this paper is to extend the Lee-Carter model, keeping the common trend structure adopted in this framework, but considering several competing conditional distributions for different outcome variables, namely, the mortality rates and the number of deaths. Using a new class of observation-driven time series models, known as Generalized Autoregressive Score (GAS, by Creal, Koopman et al. (2008, 2013)) or Dynamic Conditional Score model (DCS, following Harvey (2013)), we estimate, forecast and simulate mortality rates trends for different age groups. In GAS models, the mechanism for updating the parameters that change over time uses the scaled score of the likelihood function. Creal et al. (2013) argue that the use of the score for updating time-varying parameters is intuitive given that it defines the steepest ascent direction for improving the model's local fit in terms of the likelihood or density at time t given the current position of that parameter. Blasques et al. (2015) have justified the GAS updating mechanism using optimality arguments based on Kullback-Leibler distance. Analogous to the Generalized Linear Model (GLM), in the GAS framework it is also necessary to tailor appropriate links functions, so that parameters are constrained to appropriate subsets of the real line.

Since GAS models are applicable to a wide class of non-Gaussian conditional distributions, an advantage of our Lee-Carter extension is that we can adopt any likelihood function to estimate the parameters of the Lee-Carter model from a chosen variable of interest (outcome variable), such as number of deaths, log mortality rates or mortality rates. Consequently, in this paper, the Poisson, binomial, negative binomial, Gaussian and beta distributions are tested, resulting in different likelihood functions to estimate the Lee-Carter parameters and to dynamically forecast the mortality rates. Using the proposed framework and several competing distributions, we can identify the best variable of interest and the best probability model to be used for forecasting mortality rates using the Lee-Carter model.

Another potential advantage of our proposed framework in relation to the original Lee-Carter model is that parameter estimation is accomplished in just one-step. Also mortality rate forecasting is derived from the model assumptions, via Monte Carlo simulation, without the need to assume an auxiliary model for forecasting (usually ARIMA(0,1,1)), as it is usually the case for the Lee-Carter model. Thus, parameter estimation, signal extraction, and forecasting are obtained from a single model (Creal at al. (2013)), differentiating our proposed model from the Lee-Carter model and its extensions, such as those by Brouhns et al. (2002) and Renshaw and Haberman (2006). Consequently, our approach preserves validity of inference that is lost in the original multi-step model of Lee-Carter and some of its extensions

The remainder of the paper is organized as follows. The second section summarizes the Lee-Carter model. The third section presents the GAS models framework. In the fourth section, the proposed models for mortality rate forecasting are presented in details. In the fifth section, we apply the proposed models to the time series of mortality rates. The sixth section contains the conclusions.

2. The original Lee-Carter model

Lee and Carter (1992) proposed a single and efficient model for forecasting the central mortality rates:

$$log(m_{xt}) = \alpha_x + \beta_x \kappa_t + \varepsilon_{xt} \tag{1}$$

for $x = 1, \dots, N$ and $t = 1, \dots, T$;

where

m_{xt} is the central mortality rate of age x at time t;

κ_t is time-varying parameter, which represents the common trend of the log of the mortality rates for all ages;

β_x is an age-specific parameter, representing the sensitivity of the log of the mortality rates at age x to the time trend represented by κ_t;

α_x is an age-specific intercept ; and ε_{xt} represents the effects not captured by the model (errors), assumed to be i.i.d. $\mathrm{N}(0, \sigma^2)$. Also, one needs to impose the constraints: $\sum_{x=1}^{N} \beta_x = 1$ and $\sum_{t=1}^{T} \kappa_t = 0$, added to identify the model. Using these constraints one obtains the least squares estimator for α_x, given by $\widehat{\alpha_x} = \frac{\sum_{x=1}^{N} \log(m_{xt})}{N}$.

In the first step of the estimation, the unknown parameters β_x and κ_t are estimated via singular value decomposition (SVD) of the matrix of centered age profiles $(log(m_{xt}) - \alpha_x)$. At the second step, OLS is used to improve the fitting of κ_t, the common trend, by minimizing the errors in the estimated number of deaths, and such adjustment, by construction, gives more weight to ages at which deaths are higher. To forecast the log mortality rates, the model needs a third step, which is obtained by maintaining the estimated values of α_x and β_x and forecasting the estimated time-varying parameter κ_t , usually via ARIMA models. In Lee and Carter (1992), it was found that a random walk with drift is the most appropriate model for the evolution of κ_t - the common trend on log mortality rates for all ages.

3. Basic GAS models specification

One of the advantages of using the GAS model framework when employing Lee-Carter model is that parameter estimation, signal extraction, and forecasting occur in a single unified step. Therefore, our approach conserves inference results that are lost in a multi-step model, such as that of Lee-Carter. In addition, the Lee-Carter model is distribution free, while GAS models require one to choose, at the outset, a statistically sound conditional density for the variable under investigation. As such, a properly fitted GAS model will never simulate values outside the chosen density's support.

A general description of GAS models is given in the sequel.

Following Creal et al. (2013), the basic GAS model is defined in terms of a scalar y_t, the dependent variable of interest (outcome variable), f_t as the time-varying parameter vector, all at time t, and θ as the vector of static parameters. At time$t$, the available information set is $\{f_t, \mathcal{F}_{t-1}\}$ where:

$$\mathcal{F}_{t-1} = \{Y^{t-1}, F^{t-1}\}, \quad t = 1, 2, \ldots, T,$$

where $Y_t = \{y_1, \dots, y_t\}$ and $F_t = \{f_1, \dots, f_t\}$.

In GAS models y_t is a time series with known conditional probability model:

$$y_t \sim p(y_t | f_t, \mathcal{F}_{t-1}; \theta) \tag{2}$$

The time-varying parameter vector f_t, is updated according to a GAS(p,q) model:

$$f_{t+1} = \omega + \sum_{i=1}^{p} A_i s_{t-i+1} + \sum_{j=1}^{q} B_j f_{t-j+1} \tag{3}$$

where:

ω is a vector of constants;

A_i and B_j are matrices that have appropriate dimensions for $i = 1, \dots, p$ and $j = 1, \dots, q$, respectively; and

s_t is the scaled score.

The unknown elements of A_i, B_j, ω and any fixed parameters of the distribution $p(y_t | f_t, \mathcal{F}_{t-1}; \theta)$ are combined into a vector of static parameters, θ. The scaled score s_t is a function of the past observations $\left(s_t(y_t, f_t, \mathcal{F}_{t-1}, \theta)\right)$ and is given by:

$$s_t = S_t \nabla_t \tag{4}$$

where

$\nabla_t = \frac{\partial \, log \, p(y_t | f_t, \mathcal{F}_{t-1}; \theta)}{\partial f_t}$ is the score vector; (5)

$S_t = I_t^{-d}$, $I_t = E_{t-1}[\nabla_t \nabla_t{}']$ $d = 1/2, 1.$ (6)

where, I_t is the conditional information matrix, and E_{t-1} denotes expectation with respect to the past information $\mathcal{F}_{t-1}$.

The scaling matrix S_t introduces additional flexibility to the model. As stated by Creal et al. (2013), different choices of d leads to different GAS models.

In summary, in GAS models, when a new observation y_t comes in, the time-varying parameter vector f_t is updated for the next period $t + 1$ following the recursion given by equation (3).

As shown by Creal et al. (2013), GAS models encompass many well-known observation-driven models, such as the GARCH model of Engle (1982), Bollerslev (1986), the ACD model of Engle and Russell (1998), and the ACI model of Russell (2001), and also most of the Poisson count data models considered by Davis et al. (2003).

Usually, some elements of the time-varying parameters vector have natural constraints. To overcome this situation and ensure that f_t will remain in its appropriate domain (e.g. positive for variance), it is common to adopt a suitable parameterization. For example, if $y_t \sim Poisson\ (\lambda_t), \lambda_t > 0$, then it is natural to take $\widetilde{f_t} = \log(\lambda_t)$.

4. GAS models for mortality rates

Consider y_{xt} a generic outcome variable in the context of mortality forecasting, where x is the age, and t, time. By assumption y_{xt} has conditional density/probability mass function given by $p_x(y_{xt}|f_t, \mathcal{F}_{t-1}; \theta)$. Here following Lee- Carter, and adapting the ideas of Creal et al. (2013), assume a factor model structure in which the $y_{xt}'s$ at time t are cross-sectionally independent. Then conditional on the time-varying parameter f_t and on the information set $\mathcal{F}_{t-1}$, it follows that the conditional distribution $p(y_t|f_t, \mathcal{F}_{t-1}; \theta)$ will be given by:

$$p(y_t|f_t, \mathcal{F}_{t-1}; \theta) = \prod_{x=1}^{N} p_x(y_{xt}|f_t, \mathcal{F}_{t-1}; \theta) \tag{7}$$

In order to adapt the Lee-Carter model to the fully parametric GAS framework, start with the general expression for the evolution of the log of the mortality rates as given by eq (1). Assume that the term κ_t in this equation, which represents the common trend for all age groups, is the time-varying parameter of our proposed GAS model, that is $\kappa_t = f_t$. More specifically, adopt a GAS (1,1) mechanism for κ_t (see eq. (3)), given by:

$$\kappa_{t+1} = \omega + As_t + B\kappa_t \tag{8}$$

where s_t is the scaled score of the likelihood, and ω, A and B are static unknown parameters. Note that by making $B = 1$ on equation (8) turns our updating mechanism very similar to that originally assumed by Lee-Carter (1992).

Given an assumed probability model for the outcome variable y_{xt} given by $p_x(y_{xt}|f_t, \mathcal{F}_{t-1}; \theta)$ and using eqs. (4), (5) and (6) it is easy to see that the appropriate expressions for the scaled score s_t and the information matrix I_t are:

$$s_t = S_t \ \sum_{x=1}^{N} \nabla_{xt} \tag{9}$$

where

$$\nabla_{xt} = \frac{\partial \log p_x(y_{xt}|\kappa_t, \mathcal{F}_{t-1}; \theta)}{\partial \kappa_t}; \tag{10}$$

$$S_t = I_t^{-1/2}, \tag{11}$$

where

$$I_t = \sum_{x=1}^{N} I_{xt} \tag{12}$$

with

$$I_{xt} = E_{t-1}[\nabla_{xt}\nabla_{xt}{}'].\tag{13}$$

For each age x, the derivation of the log of the probability model given in eq. (10), results in the “partial” score, from which it is possible to obtain the expected value of $\nabla_{xt}\nabla_{xt}{}'$, using eq. (13), resulting in the “partial” information matrix. With such expressions available, the scaled score matrix S_t (eq. 11) is readily obtained

Finally, for a given choice of $p_x(y_{xt}|\kappa_t, \mathcal{F}_{t-1}; \theta)$, in view of the hypothesis embodied in eq. (7), the likelihood function will be given by

$$L(\theta) = \prod_{t=1}^{T} p(y_t|f_t, \mathcal{F}_{t-1}; \theta) = \prod_{t=1}^{T}\prod_{x=1}^{N} p_x(y_{xt}|f_t, \mathcal{F}_{t-1}; \theta)$$

From this, it follows that the log of the likelihood is given by:

$$l(\theta) = \textstyle\sum_{t=1}^{T}\sum_{x=1}^{N} log\, p_x(y_{xt}|f_t, \mathcal{F}_{t-1}; \theta)\tag{14}$$

In practice, to find maximum likelihood estimators appropriate non-linear optimization algorithms can be used, such as Broyden–Fletcher–Goldfarb–Shanno (BFGS, Broyden (1970), Fletcher (1970), Goldfarb (1970) and Shanno (1970)) or Berndt–Hall–Hall–Hausman (BHHH, Berndt et al. (1974)).

It should be noticed that every choice of a particular distribution $p_x(y_{xt}|\kappa_t, \mathcal{F}_{t-1}; \theta)$ results in a different updating equation for the common trend (eq. 8) (given that the expression for scaled score will change) and also on a different likelihood function$l(\theta)$, as it can be seen through eq. (14). This will be made explicit in sub–sections 4.1 to 4.5 when particular forms for $p_x(y_{xt}|\kappa_t, \mathcal{F}_{t-1}; \theta)$ are assumed. It is also important to notice that the resulting updating mechanisms for k_t is constructed in such a way that the age-specific parameters $\beta_x{}'s$ weight the unique time-varying parameter and the elements used to obtain the scaled score at time t.

The vector of static unknown parameters θ is estimated by maximizing the log-likelihood function with respect to θ (eq. (14)). Thus, in the GAS extension of Lee-Carter, parameter estimation is obtained in a single step. Multi step ahead forecasting of both the time-varying parameters and future observations are obtained by Monte Carlo simulation using the recursion on eq. (8). Consequently, our approach retains valid inference results that are lost in the original Lee-Carter and its extensions, ensuring that the extracted factors are related to the outcome variables of interest through the estimation and forecasting process (Creal at al., 2014).

To model mortality rates using versions of the Lee-Carter model, it is assumed, like in its original formulation, that for any age group x at time t the force of mortality rates obeys the following relation:

$$\mu_{(x+c)t} = \mu_{xt} \quad \text{for } 0 \leq c < g\tag{15}$$

where g is the width of the age group. It follows that the force of mortality rates is equal to the central mortality rates ($\mu_{xt} = m_{xt}$).

Since the GAS likelihood function has a closed form, we can adopt different probability distributions for the variable of interest y_{xt}, extending the Lee-Carter model to distributions other than the lognormal. In the sequence, we propose and develop five different GAS models for forecasting mortality rates and related variables.

4.1. Poisson GAS model for the number of deaths

In the first proposed model the variable of interest, is the number of deaths (d_{xt}) of age x at time t, assumed to be independent realizations of a Poisson random variable, conditional on the number of people exposed to risk (L_{xt}), which will be known in real data applications. Thus

$$p(d_{xt}|L_{xt}, \mathcal{F}_{t-1}) \sim Poisson(\lambda_{xt}) \tag{16}$$

where $\lambda_{xt} = L_{xt} exp(\alpha_x + \beta_x \kappa_t)$. Now, assuming that the mortality rate is $m_{xt} = \frac{d_{xt}}{L_{xt}}$, it follows that, conditional on the knowledge of the number of people exposed to risk (L_{xt}), the mean and variance of the mortality rate will be given, respectively, by:

$$E(m_{xt}|L_{xt}, \mathcal{F}_{t-1}) = exp(\alpha_x + \beta_x \kappa_t);$$

$$Var(m_{xt}|L_{xt}, \mathcal{F}_{t-1}) = \frac{1}{L_{xt}}\ exp(\alpha_x + \beta_x \kappa_t).$$

Using the Poisson assumption, it is easy to show that the corresponding "partial" score and information used in the GAS (1,1) updating mechanism (see eq 8) will be given by $\nabla_{xt} = \frac{\partial\, log\, p_x(d_{xt}|L_{xt}, \mathcal{F}_{t-1})}{\partial \kappa_t} = \beta_x(d_{xt} - \lambda_{xt})$ and $I_{xt} = E_{t-1}[\nabla_{xt} \nabla_{xt}{}'] = \beta_x^2 \lambda_{xt}$, respectively.

4.2. Binomial GAS model for the number of deaths

The second GAS model still considers the number of deaths (d_{xt}) as the variable of interest, but now, conditional on the population size on the first day of the year (l_{xt}), d_{xt} is assumed to have a binomial distribution :

$$p\left(d_{xt}\middle|l_{xt,}\mathcal{F}_{t-1}\right) \sim Bin(l_{xt}, q_{xt}), \quad 0 < q_{xt} \leq 1 \tag{17}$$

where q_{xt} is the probability of death for age x at time t and is linked to the common trend κ_t through a logistic function $q_{xt} = \frac{1}{1+e^{-exp(\alpha_x + \beta_x \kappa_t)}}$. Given the population size on the first day of the year (l_{xt}) and the binomial assumption, the mean and the variance for the mortality rate are given by:

$$E\left(m_{xt}|l_{xt,}\mathcal{F}_{t-1}\right) = \frac{l_{xt}}{L_{xt}}\frac{exp(\alpha_x+\beta_x\kappa_t)}{1+exp(\alpha_x+\beta_x\kappa_t)};$$

$$Var\left(m_{xt}|l_{xt,}\mathcal{F}_{t-1}\right) = \frac{l_{xt}}{{L_{xt}}^2}\frac{exp(\alpha_x+\beta_x\kappa_t)}{\left(1+exp(\alpha_x+\beta_x\kappa_t)\right)^2}.$$

Similar to the Poisson case, in order to implement the GAS (1,1) updating mechanism for k_t , it is necessary to evaluate both the "partial" score and the "partial" information, which are given by $\nabla_{xt} = \beta_x(\, d_{xt} - l_{xt}q_{xt})$ and $I_{xt} = \beta_x^2 l_{xt}q_{xt}(1-q_{xt})$, respectively.

4.3. Negative binomial GAS model for the number of deaths

According to Delward et al. (2007), the over–dispersion observed in much mortality data can be appropriately tackled by assuming that the number of deaths has a negative binomial distribution conditioned on the number of people exposed to risk (L_{xt}). Using this result a negative binomial GAS model for the number of deaths is proposed:

$$p(d_{xt}|L_{xt},\mathcal{F}_{t-1}) \sim NB(r_x, h_{xt}) \tag{18}$$

It is hoped that the extra parameter to be estimated for each age (r_x), albeit fixed in time, may add flexibility, improving data fitting when compared to the Poisson GAS model. As in the Poisson distribution, here it is also assumed that $E(d_{xt}|L_{xt},\mathcal{F}_{t-1}) = L_{xt}exp(\alpha_x + \beta_x\kappa_t)$. Then it follows that, conditional on the number of people exposed to risk (L_{xt}), the mean and variance of the mortality rate are given by:

$$E(m_{xt}|L_{xt},\mathcal{F}_{t-1}) = exp(\alpha_x + \beta_x\kappa_t) = \frac{r_x(1-h_{xt})}{L_{xt}h_{xt}};$$

$$Var(m_{xt}|L_{xt},\mathcal{F}_{t-1}) = \frac{r_x(1-h_{xt})}{(L_{xt}h_{xt})^2} = \frac{exp(\alpha_x+\beta_x\kappa_t)}{L_{xt}h_{xt}}.$$

Given the negative binomial assumption, it is not difficult to show that $\nabla_{xt} = \beta_x[(d_{xt}h_{\mathrm{xt}}) - (1-h_{xt})r_{\mathrm{x}}]$ and $I_{xt} = \beta_x^2(1-h_{xt})r_{\mathrm{x}}$ the necessary expressions to derive the GAS(1,1) mechanism.

4.4. Gaussian GAS model for the log of mortality rate

As in the Lee-Carter model, here it is assumed that the log of mortality rates follows a Gaussian distribution, but now the errors ($\varepsilon_{\mathrm{xt}}$'s) are heteroscedastic with respect to the ages x, i.e., ε_{xt} ~ i.i.d. $N(0,\sigma_x^2)$. It then follows that the proposed model for the log mortality rates is given by:

$$log(m_{xt}|\mathcal{F}_{t-1}) \sim N(\mu_{xt},\sigma_x^2) \tag{19}$$

Where $\mu_{xt} = E(log(m_{xt})|\mathcal{F}_{t-1}) = \alpha_x + \beta_x\kappa_t$. From this it follows that the mortality rate will be log normally distributed, that is, $p(m_{xt}|\mathcal{F}_{t-1}) \sim lnN(\mu_{xt},\sigma_x^2)$, from which it follows that:

$E(m_{xt}|\mathcal{F}_{t-1}) = exp\left(\alpha_x + \beta_x\kappa_t + \frac{\sigma_x^2}{2}\right);$

$Var(m_{xt}|\mathcal{F}_{t-1}) = [exp(\sigma_x^2) - 1]\left[exp\left(\sigma_x^2 + 2(\alpha_x + \beta_x\kappa_t)\right)\right].$

As before, we need to derive expressions to obtain the full updating equation for k_t. It can be shown that under Gaussianity, these take the form $\nabla_{xt} = \beta_x \frac{(log(m_{xt}) - \mu_{xt})}{2\sigma_x^2}$ and $I_{xt} = 0.5\beta_x^2\sigma_x^{-2}$.

4.5. Beta GAS model for mortality rate

In our final proposed GAS model, the variable of interest is also the mortality rate for age x at time t, m_{xt}, assumed to follow a GAS beta model, with conditional mean given by $\mathrm{E}(m_{xt}|\mathcal{F}_{t-1}) = exp(\alpha_x + \beta_x\kappa_t)$. From this it follows that the first parameter of each beta distribution is time-varying, being dependent on κ_t:

$$p(m_{xt}|\mathcal{F}_{t-1}) \sim beta(\gamma_{xt}, \xi_x) \tag{20}$$

where:

$E(m_{xt}|\mathcal{F}_{t-1}) = \frac{\gamma_{xt}}{\gamma_{xt} + \xi_x} = exp(\alpha_x + \beta_x\kappa_t)$ with $\gamma_{xt} = \xi_x\left(\frac{exp(\alpha_x + \beta_x\kappa_t)}{1 - exp(\alpha_x + \beta_x\kappa_t)}\right);$

$Var(m_{xt}|\mathcal{F}_{t-1}) = \frac{\gamma_{xt}\xi_x}{(\gamma_{xt} + \xi_x)^2(\gamma_{xt} + \xi_x + 1)}$.

In this case, it can be shown that the associated quantities needed to fully specify the GAS (1,1) mechanism are:

$\nabla_{xt} = \beta_x\left[\left(\frac{\gamma_{xt}(\gamma_{xt} + \xi_x)}{\xi_x}\right)\right][log(m_{xt}) + \Psi(\gamma_{xt} + \xi_x, 1) - \Psi(\gamma_{xt}, 1)];$
$I_{xt} = \beta_x^2\left[\left(\frac{{\gamma_{xt}}^2(\gamma_{xt} + \xi_x)^2}{{\xi_x}^2}\right)\right][-\Psi(\gamma_{xt} + \xi_x, 2) + \Psi(\gamma_{xt}, 2)],$

where $\Psi(x, k) = \frac{\partial^k log\Gamma(x)}{\partial x^k}$, $k = 1,2, \ldots$, $\Gamma(x)$ being the gamma function.

4.6. Summary of the GAS updating mechanisms

All our proposed GAS models share a common statistical feature, which brings extra flexibility in data fitting: both the conditional mean and conditional variance of the distribution of the mortality rates of each age are time-varying. Also, in practice one does not need an extra source of data, other than the mortality rates time series themselves when fitting either the log Normal or beta models. Furthermore, to implement the proposed GAS models we do not impose the constraints used by Lee and Carter (1992) (eq.(1)). Table 1 summarizes the different proposed GAS models.

Notice that the discrete distribution models here proposed, namely, Poisson, binomial and negative binomial are originally specified for the number of deaths for age x at time t, d_{xt}, as was shown in sub–sessions 4.1, 4.2 and 4.3. Nevertheless, it is possible to obtain the distribution for the mortality rates m_{xt} from these discrete models, conditional on the number of people exposed to risk (L_{xt} - for Poisson and negative binomial models) and on the population size on the first day of the year ($l_{xt} -$ for binomial model). From any of the proposed GAS models, based either on discrete or continuous distributions, it is then possible to derive predictions for death rates for different ages.

5. Applications

The proposed GAS models are applied to time series of mortality rates of the male population of the US, UK, Sweden and Japan. The data is from the Human Mortality Database[2] covering the period from 1960 to 2010, considering the following age groups: 30-34 years, 35-39 years, 40-44 years, 45-49 years, 50-54 years, 55-59 years, 60-64 years, 65-69 years, 70-74 years, 75-79 years, 80-84 years, 85-89 years and 90-94 years.

[2] 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 (data downloaded on February 26, 2012).

Table 1 - Summary of the proposed GAS models

Probability model	Variable of interest (y_{xt})	Partial Score (∇_{xt})	Partial Information matrix (I_{xt})
Poisson (λ_{xt}) $\frac{e^{-\lambda_{xt}}\lambda_{xt}{}^{y_{xt}}}{y_{xt}!}$	d_{xt}	$\beta_x(d_{xt} - \lambda_{xt})$	$\beta_x^2 \lambda_{xt}$
binomial (l_{xt}, q_{xt}) $\binom{l_{xt}}{y_{xt}} q_{xt}^{y_{xt}} \quad (1-q)^{l-y_{xt}}$	d_{xt}	$\beta_x(d_{xt} - l_{xt}q_{xt})$	$\beta_x^2 l_{xt} q_{xt}(1-q_{xt})$
negative binomial (r_x, h_{xt}) $\binom{y_{xt}+r_x-1}{y_{xt}} h_{xt}^{r_x} \quad (1-h_{xt})^{y_{xt}}$	d_{xt}	$\beta_x[(d_{xt}h_{\mathrm{xt}}) - (1-h_{xt})r_{\mathrm{x}}]$	$\beta_x^2(1-h_{xt})r_{\mathrm{x}}$
Gaussian (μ_{xt}, σ_x^2) $\frac{exp\left(-(y_{xt}-\mu_{xt})^2 / 2\sigma_x^2\right)}{\sqrt{2\pi\sigma_x^2}}$	$log(m_{xt})$	$\beta_x \frac{(log(m_{xt}) - \mu_{xt})}{2\sigma_x^2}$	$0.5\beta_x^2\sigma_x^{-2}$
Beta (γ_{xt}, ξ_x) $\frac{\Gamma(\gamma_{xt}+\xi_x)}{\Gamma(\gamma_{xt})\Gamma(\xi_x)}.\left[y_{xt}{}^{\gamma_{xt}-1}(1-y_{xt})^{\xi_x-1}\right]$	m_{xt}	$\beta_x\left[\left(\frac{\gamma_{xt}(\gamma_{xt}+\xi_x)}{\xi_x}\right)\right][log(m_{xt}) + \Psi(\gamma_{xt}+\xi_x,1) - \Psi(\gamma_{xt},1)]$	$\beta_x^2\left[\left(\frac{\gamma_{xt}{}^2(\gamma_{xt}+\xi_x)^2}{\xi_x{}^2}\right)\right][-\Psi(\gamma_{xt}+\xi_x,2) + \Psi(\gamma_{xt},2)]$,

Note: m_{xt} is mortality rate for age x at time t; q_{xt} is the probability of death for age x at time t; and d_{xt} is the number of deaths for age x at time t.

We chose to use these thirteen 5-year age groups because these cover the public that participates in pension plans and life insurance. By restricting the number of age groups the number of parameters to be estimated non linearly is reduced. The last 5 years of data has been omitted for out-of-sample validation. In the sequel the results for the fitting of the US data are presented. The analysis for the remaining countries (the UK, Sweden and Japan) can be found in the Appendix. Figure 1 presents the time series of the observed US mortality rates for four age groups, namely: 40-44 years, 50-54 years, 60-64 years and 70-74 years. The observed downward trends for all these series confirm the well-established fact of the steady decline of US mortality rates in the last decades.

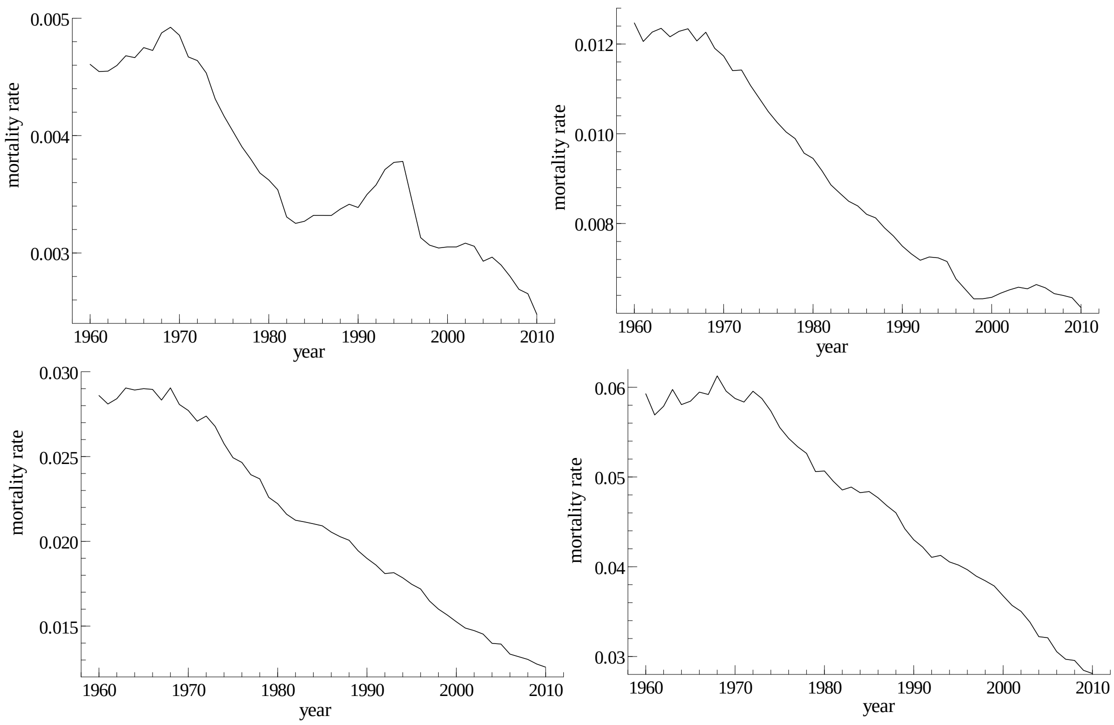

Figure 1**.** Observed mortality rates. The top-left graph is for the 40-44 age group, the top-right graph is for the 50-54 age group, the bottom–left graph is for the 60-64 age groups, and the bottom-right graph is for the 70-74 age group.

Using the AIC reported for each of the models fitted to the number of deaths, as shown in Table 2, it can be concluded that among the models with discrete distribution (Poisson, binomial and negative binomial), the negative binomial is the best choice, since it minimizes the AIC. This may be explained by the extra static parameter r_x, whose improvement brought in model fitting outperforms its contribution to increase model complexity. Among the models with continuous distribution, the AIC is -2,059.27 for the Gaussian model and and -6,709.63 for the beta model. Nevertheless, these values are not directly comparable given that the Gaussian model is fitted to log mortality rates while the beta model to mortality rates.

Table 2. AIC for the different GAS models with discrete distributions fitted to the numbers of deaths

Model	AIC
Poisson	75,495.15
binomial	79,652.75
negative binomial	11,102.11

Estimated values of the static parameters for all GAS models considered in our study are shown in the Table A1 in the Appendix. Using Student's *t*-test, we can reject the null hypotheses that the static parameters are zero (p-value $\leq 10^{-9}$ for all estimated parameters) at 1% level or less. Quantile residuals for the different GAS models considered in this paper are tested for normality, homoscedasticity and absence of serial correlation using the Jarque-Bera test, the Box-Ljung test in the squared residuals, and the Box-Ljung test, respectively. Adopting a significance level of 1%, the hypothesis of uncorrelated residuals, for all estimated models is rejected. This unsatisfactory behavior may be explained by the fact that the common trend used to explain the variation on mortality time series for all age groups, adapted from the Lee-Carter model, in all our GAS models, seems insufficient to capture all linear dependence of these time series.

On the other hand, it can be seen from Table 3, diagnostic tests on the quantile residuals reject neither normality nor homoscedasticity for the majority of age groups. Nevertheless, for the first four age groups (see Figure 1, for the 40-44 years age group), the squared residuals still show some dependence. When all the proposed GAS models are compared using residual diagnostics, the negative binomial is the best model for the number of deaths and the beta model for mortality rates.

Table 3
Diagnostic tests of normality and homoscedasticity based on quantile residuals: p-values

Age group (years)	**Poisson**		**Binomial**		**Negative binomial**		**Gaussian**		**Beta**	
	norm.	homos.	norm.	homos.	norm.	homos.	norm.	homos.	norm.	homos.
30-34	0.181	0.000	0.276	0.000	0.196	0.000	0.168	0.000	0.175	0.000
35-39	0.119	0.000	0.117	0.000	0.245	0.000	0.262	0.000	0.229	0.000
40-44	0.073	0.000	0.241	0.000	0.904	0.000	0.895	0.000	0.898	0.000
45-49	0.108	0.000	0.133	0.000	0.311	0.000	0.309	0.000	0.302	0.000

50-54	0.135	0.006	0.088	0.000	0.000	0.986	0.000	0.984	0.000	0.992
55-59	0.085	0.116	0.132	0.000	0.527	0.012	0.512	0.014	0.363	0.181
60-64	0.109	0.002	0.096	0.000	0.274	0.475	0.268	0.481	0.210	0.437
65-69	0.048	0.000	0.000	0.000	0.265	0.024	0.255	0.022	0.190	0.006
70-74	0.001	0.350	0.387	0.000	0.135	0.350	0.116	0.380	0.211	0.101
75-79	0.030	0.109	0.253	0.000	0.483	0.115	0.464	0.111	0.598	0.136
80-84	0.000	0.161	0.339	0.000	0.089	0.608	0.085	0.586	0.129	0.312
85-89	0.024	0.508	0.339	0.000	0.432	0.067	0.465	0.065	0.504	0.056
90-95	0.203	0.173	0.241	0.000	0.376	0.260	0.386	0.186	0.371	0.095

Table 4 reports the mean absolute percentage error (MAPE), both in sample and out-of-sample, for the GAS fitted models. The beta and binomial GAS models are the most accurate, in sample and out-of-sample, respectively. In addition, forecasting performance amongst the competing models is also compared using the Diebold-Mariano (DM) test (Diebold and Mariano, 2002) via the MAPE loss function. In DM test the null hypothesis of no difference between forecasts cannot be rejected in the period in sample. Nevertheless, in the out-of-sample period, the test would not reject that the binomial model is more accurate than the other competing models.

Table 4
MAPE values (%) for the proposed GAS models, in sample and out-of-sample.

Age groups (years)	**Poisson**		**Binomial**		**Negative binomial**		**Gaussian**		**Beta**	
	in sample	out-of-sample	in sample	out-of-sample	in sample	out-of-sample	in sample	out-of-sample	in sample	out-of-sample
30-34	9.68%	12.59%	10.01%	11.98%	9.46%	12.07%	9.44%	10.29%	9.46%	12.42%
35-39	8.53	14.06	8.38	15.64	8.37	13.83	8.33	12.71	8.32	13.97
40-44	5.79	4.12	5.34	5.91	5.59	4.19	5.59	3.98	5.33	4.00
45-49	4.64	7.85	4.60	5.47	4.44	7.55	4.43	7.76	3.98	8.07
50-54	2.98	13.64	3.18	12.39	2.72	12.73	2.71	12.91	2.28	12.57
55-59	2.07	8.03	1.84	6.80	2.04	6.57	2.06	6.60	2.20	5.21
60-64	2.11	1.59	1.95	0.31	2.09	3.17	2.10	3.09	2.22	4.60
65-69	2.19	6.09	2.14	4.68	2.20	7.29	2.22	7.18	2.38	8.85
70-74	2.29	10.43	2.41	10.12	2.36	11.64	2.35	11.54	2.49	13.57

75-79	2.04	7.87	2.13	9.08	2.03	9.22	2.04	9.16	2.13	10.56
80-84	2.07	11.59	2.24	12.89	2.08	12.90	2.08	12.81	2.16	13.76
85-89	3.12	14.53	3.24	13.90	3.01	14.81	3.02	14.75	2.76	13.68
90-95	3.83	9.95	3.81	8.50	3.58	9.21	3.59	9.15	3.48	8.92
Total	3.95	9.41	3.94	9.05	3.84	9.63	3.84	9.38	3.78	10.00

Since the negative binomial and beta GAS models produced the best results using AIC and diagnostic tests, the equality of their out-of-sample MAPE is tested via DM. The results suggest that for the US data, the negative binomial model outperforms the beta model. To provide further evidence of these findings we also applied our proposed framework to forecast central mortality rates for others countries than the US, namely, Japan, Sweden and UK. The data is also obtained from the Human Mortality Database, considering the same period and the 5-year age groups used in the US example. Tables A2, A3 and A4 in the Appendix depict out of sample MAPE values for these countries. The findings are similar to those encountered when analyzing Table 4 for the US mortality data: the negative binomial GAS model produces better forecasting than the beta model in two among the three countries.

For the sake of completeness the standard Lee-Carter model (using three steps to estimate and forecast the mortality rates) has also been fitted to the different age groups of the US male mortality data. Averaging the MAPEs of the different age groups result in 3.19% for the in-sample period and 7.19% for the out-of-sample period. Figure 2 shows the time series of the common trend κ_t (the time-varying parameter) for the US data, estimated both by the negative binomial GAS and the Lee-Carter model. It can be seen that the negative binomial GAS model produces a smoother trend than the LC model, which may be explained by the fact that in the latter the common trend is re-estimated several times, in order to minimize the error associated with the number of deaths for each age group.

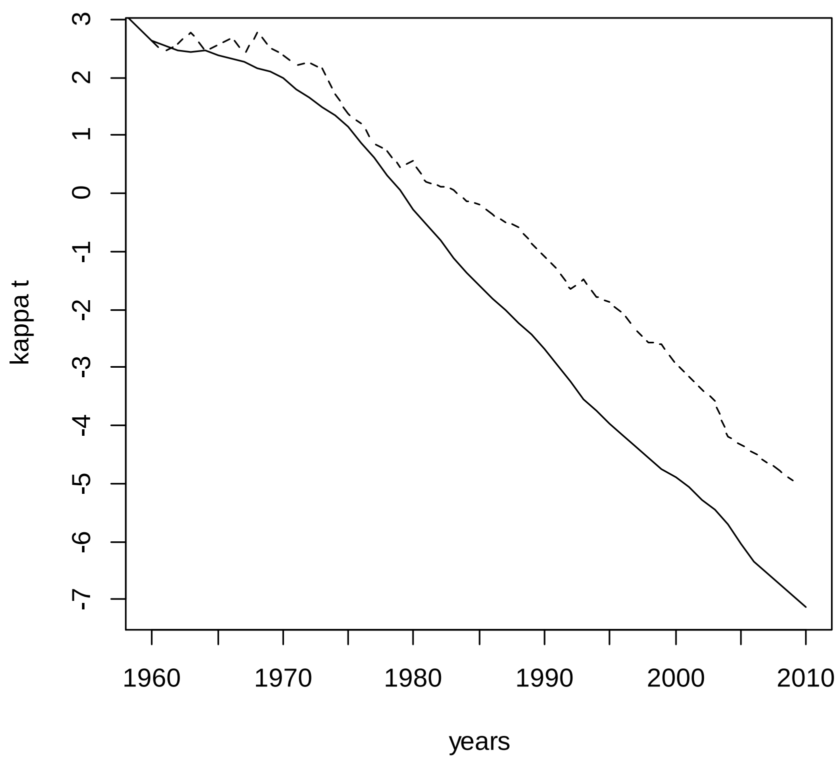

Figure 2. Time series of the mortality common trend for the US data (κ_t) estimated by the negative binomial GAS model (solid line) and by the standard Lee-Carter model (dashed line).

In forecasting mortality time series, it is important to verify whether the model is able to capture the volatility of the time series, given that the distribution of forecasted mortality rates is used to measure the risk–based capital, which is evaluated using the tail of the loss distribution. In our application the models that better capture the volatility of the mortality rates time series are the Gaussian and the negative binomial GAS models. The first model estimates constant parameters for each log mortality time series (σ_x), while the negative binomial model, as stated by Deward et al. (2007), takes into account the over–dispersion of the mortality data. To illustrate this behavior, Figures 3 and 4 display the observed mortality rates and the predicted values for the out-of-sample period (from 2006 to 2010) and their 95% confidence interval for two representative age groups, 40-44 and 60-64 years. The confidence intervals are larger for the Gaussian and negative binomial models and the majority of their observed mortality rates fall within these intervals, contrary to what happens to the others competing models.

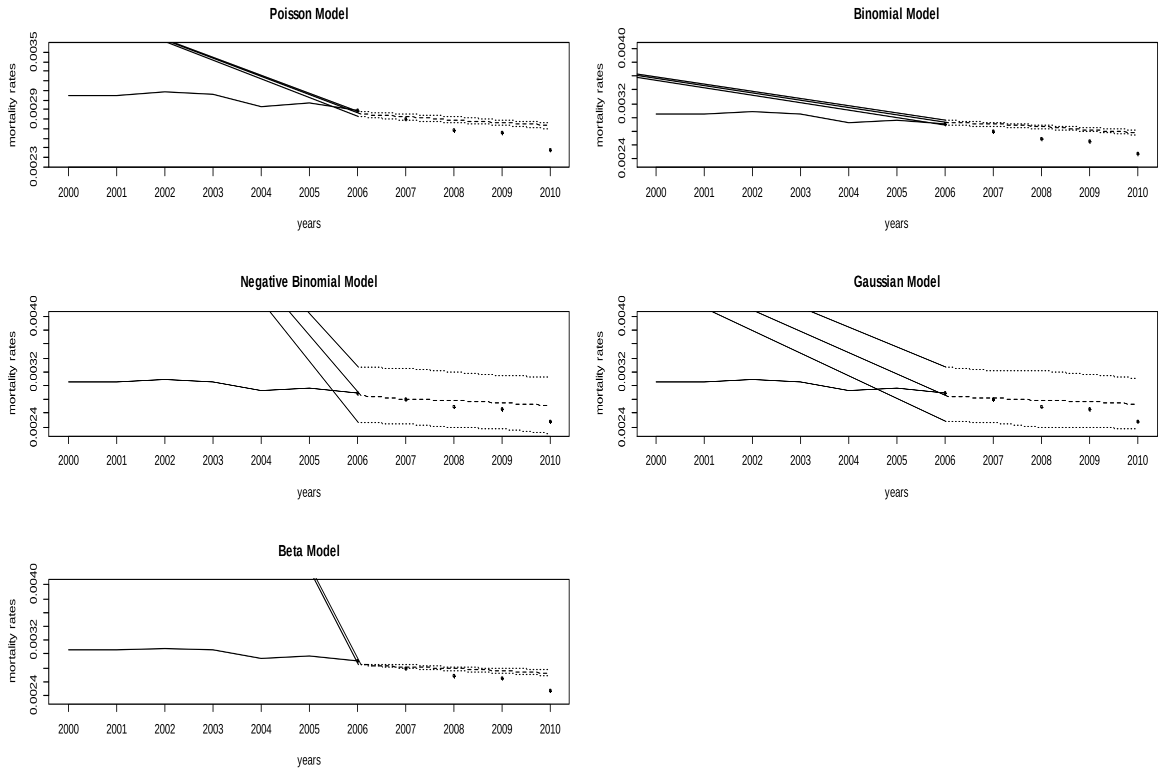

Figure 3. Observed and forecasted US male mortality rates for 40-44 years age group. Observed mortality rates up to 2005 (solid lines); predicted value of mortality rates for 2006 to 2010 (dashed lines); and its 95% confidence interval (dotted lines); and observed surrender rates in 2006 to 2010 (circles).

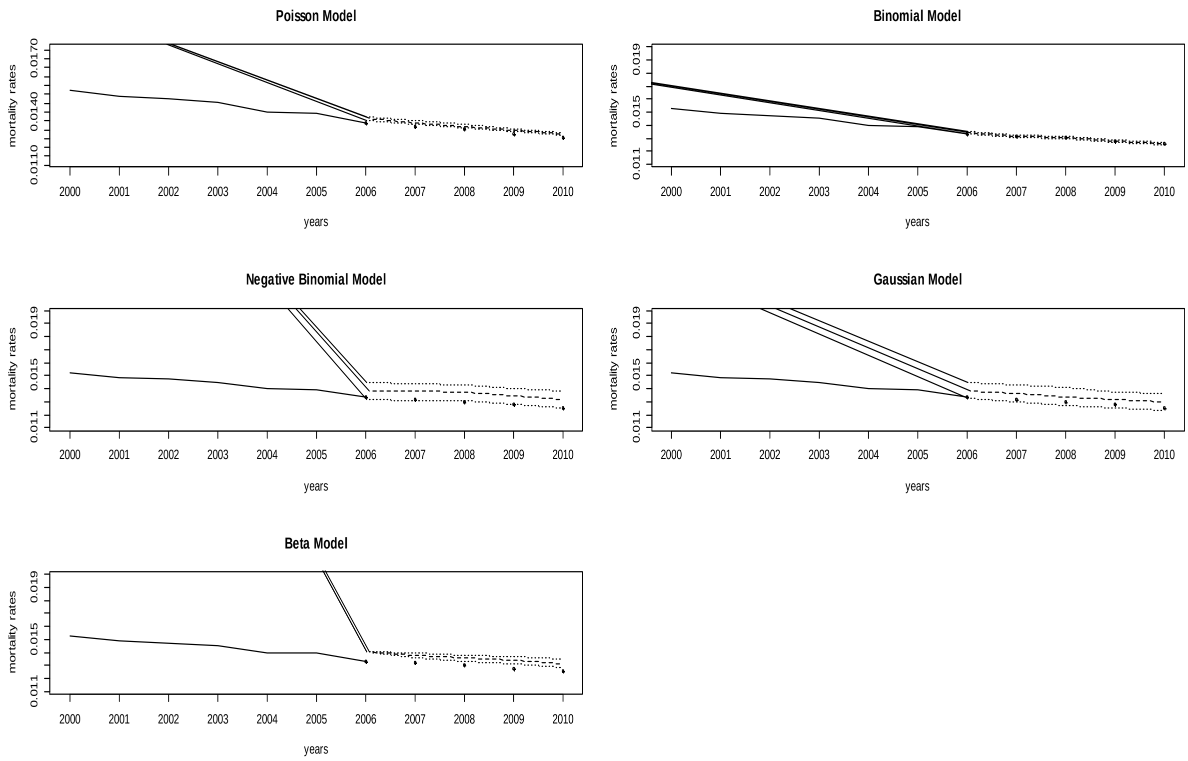

Figure 4. Observed and forecasted US male mortality rates for 60-64 years age group. Observed mortality rates up to 2005 (solid lines); predicted value of mortality rates for 2006 to 2010 (dashed lines); and its 95% confidence interval (dotted lines); and observed surrender rates in 2006 to 2010 (circles).

Given the results presented in this section, it can be concluded that the negative binomial model is the most appropriate for forecasting mortality rates amongst the GAS extensions of the Lee-Carter model.

6. Conclusion

In this paper the framework of the recently developed Generalized Autoregressive Score models has been applied to forecast mortality rates for several countries. The proposed GAS models extend the Lee-Carter model by considering flexible distribution assumptions, and present the advantage of producing forecasts in a single step, while Lee-Carter is a three-step model.

Using the GAS framework a wide class of non-Gaussian distributions can be chosen to model an appropriate variable in the context of mortality rate forecasting, resulting in different likelihood functions to estimate the parameters of the Lee-Carter model. In this paper five different distributions to forecast mortality rates via the Lee-Carter model have been proposed: Poisson, binomial, negative binomial, Gaussian and beta.

The proposed GAS models were applied to the time series of mortality rates for the male population of the United States, U.K, Sweden and Japan in the period from 1960 to 2010. Using AIC, diagnostic tests and measures of forecast accuracy, the negative binomial (conditional on the number of deaths for each age group) was chosen as the most appropriate model to forecast mortality rates for those countries.

Due to the flexibility of the GAS framework, the proposed models can be extended in several directions. For example, a multivariate distribution for the mortality data can be assumed and the common trend for mortality rates can be extended by including an extra parameter in order to capture extra linear dependence present in the time series of mortality rates. We believe that GAS models have a huge potential for the successful modeling of actuarial time series, which require models with time varying parameters and non-Gaussian distributions.

Acknowledgments

The authors would like to thank the Editor, Rob Kaas, and two anonymous referees for helpful and constructive comments. Henrique Helfer and Cristiano Fernandes acknowledge support from CNPq/Brazil. César Neves acknowledges support from Centre for Research on Insurance Economics (CPES) of the Brazilian School of Insurance.

References

Berndt, E. R., Hall, B. H., Hall, R. E., & Hausman, J. A. (1974). Estimation and inference in nonlinear structural models. Annals of Economic and Social Measurement, 3 (4), 653-665.

Blasques, F., Koopman, S. J., & Lucas, A. (2015). Information Theoretic Optimality of Observation Driven Time Series Models for Continuous Responses. Biometrika, forthcoming.

Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. Journal of Econometrics, 31 (3), 307-327.

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

Broyden, C. G. (1970). The convergence of a class of double-rank minimization algorithms. Journal of the Institute of Mathematics and Its Applications, 6(1), 76–90.

Chen, H., MacMinn, R. D., & Sun, T. (2014). Mortality Dependence and Longevity Bond Pricing: A Dynamic Factor Copula Mortality Model with the GAS Structure - http://www.gasmodel.com/gaspapers.htm.

Cossette, H., Delwarde, A., Denuit, M., Guillot, F., & Marceau, É. (2007). Pension plan valuation and mortality projection: a case study with mortality data. North American Actuarial Journal, 11(2), 1-34.

Creal, D., Koopman, S. J., & Lucas, A. (2008). A General Framework for Observation Driven Time-Varying Parameter Models. Tinbergen Institute Discussion, 108/4.

Creal, D., Koopman, S. J., & Lucas, A. (2013). Generalized autoregressive score models with applications. Journal of Applied Econometrics, 28(5), 777-795.

Creal, D., Schwaab, B., Koopman, S. J., & Lucas, A. (2014). Observation-Driven Mixed-Measurement Dynamic Factor Models with an Application to Credit Risk. Review of Economics and Statistics, 96(5), 898-915.

Davis, R. A., Dunsmuir, W. T. M., & Streett, S. (2003). Observation driven models for Poisson counts. Biometrika 90 (4), 777–790.

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

Delwarde, A., Denuit, M., & Partrat, C. (2007). Negative binomial version of the Lee–Carter model for mortality forecasting. Applied Stochastic Models in Business and Industry, 23(5), 385-401.

Diebold, F. X., & Mariano, R. S. (2002). Comparing predictive accuracy. Journal of Business & economic statistics, 20(1).

Dunn, P. K., & Smyth, G. K. (1996). Randomized quantile residuals. Journal of Computational and Graphical Statistics, 5(3), 236-244.

Engle, R. F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. Econometrica, 50 (4), 987–1007.

Engle, R. F., & Russell, J. R. (1998). Autoregressive conditional duration: a new model for irregularly spaced transaction data. Econometrica 66 (5), 1127–1162.

Fletcher, R. (1970). A New Approach to Variable Metric Algorithms. Computer Journal, 13(3), 317–322.

Goldfarb, D. (1970). A Family of Variable Metric Updates Derived by Variational Means. Mathematics of Computation, 24 (109), 23–26.

Haberman, S., & Renshaw, A. (2008). On Simulation-Based Approaches to Risk Measurement in Mortality with Specific Reference to Binomial Lee-Carter Modelling. In Society of Actuaries Living to 100 Symposium.

Harvey, A. C. (2013). Dynamic models for volatility and heavy tails: with applications to financial and economic time series (No. 52). Cambridge University Press.

Hatzopoulos, P., & Haberman, S. (2009). A parameterized approach to modeling and forecasting mortality. Insurance: Mathematics and Economics, 44(1), 103-123.

Lee, R. D., & Carter, L. (1992). Modeling and Forecasting the Time Series of US Mortality. Journal of the American Statistical Association, 87(659), 71.

Pitacco, E., Denuit, M., & Haberman, S. (2009). Modelling longevity dynamics for pensions and annuity business. Oxford University Press.

Renshaw, A. E., & Haberman, S. (2006). A cohort-based extension to the Lee–Carter model for mortality reduction factors. Insurance: Mathematics and Economics, 38(3), 556-570.

Russell, J. R. (2001). Econometric modeling of multivariate irregularly-spaced high-frequency data. Unpublished manuscript, University of Chicago, Graduate School of Business.Shanno, D. F. (1970). Conditioning of quasi-Newton methods for function minimization. Mathematics of computation, 24(111), 647-656.

Appendix

Table A1: Estimated values for the static parameters (vector θ) –US data.

Parameter	Poisson	Binom.	NB	Gaussian	Beta	**Parameter**	Poisson	Binom.	NB	Gaussian	Beta
α_1	-6.236	-6.170	-6.197	-6.196	-6.148	β_{10}	0.050	0.088	0.059	0.109	0.029
α_2	-5.965	-5.882	-5.905	-5.902	-5.838	β_{11}	0.038	0.068	0.044	0.082	0.022
α_3	-5.606	-5.491	-5.521	-5.514	-5.424	β_{12}	0.020	0.037	0.024	0.044	0.013
α_4	-5.190	-5.040	-5.078	-5.067	-4.949	β_{13}	0.006	0.013	0.010	0.018	0.005
α_5	-4.765	-4.583	-4.628	-4.614	-4.478	ω	-0.235	-0.144	-0.195	-0.106	-0.394
α_6	-4.336	-4.146	-4.195	-4.180	-4.048	A	0.009	0.004	0.065	0.050	0.151
α_7	-3.894	-3.698	-3.757	-3.741	-3.611	B	1.000	1.000	1.000	1.000	1.000
α_8	-3.492	-3.298	-3.365	-3.352	-3.233	κ_1	5.139	1.736	2.627	1.334	1.709
α_9	-3.081	-2.885	-2.969	-2.957	-2.855	r_1, σ_1, ξ_1	-	-	78.00	0.112	40,687.636
α_{10}	-2.680	-2.471	-2.581	-2.571	-2.480	r_2, σ_2, ξ_2	-	-	102.011	0.099	41,003.754
α_{11}	-2.233	-2.006	-2.158	-2.151	-2.082	r_3, σ_3, ξ_3	-	-	201.643	0.071	56,870.628
α_{12}	-1.800	-1.548	-1.758	-1.755	-1.711	r_4, σ_4, ξ_4	-	-	386.618	0.051	84,911.538
α_{13}	-1.408	-1.084	-1.383	-1.383	-1.366	r_5, σ_5, ξ_5	-	-	869.858	0.034	149,642.041
β_1	0.023	0.040	0.028	0.055	0.014	r_6, σ_6, ξ_6	-	-	1,716.415	0.025	101,229.239
β_2	0.031	0.051	0.038	0.072	0.019	r_7, σ_7, ξ_7	-	-	1,811.139	0.024	75,196.545
β_3	0.044	0.071	0.053	0.099	0.027	r_8, σ_8, ξ_8	-	-	1,445.833	0.027	39,388.554
β_4	0.059	0.094	0.070	0.130	0.036	r_9, σ_9, ξ_9	-	-	1,094.950	0.031	18,642.523
β_5	0.070	0.113	0.084	0.155	0.042	$r_{10}, \sigma_{10}, \xi_{10}$	-	-	1,604.502	0.025	17,072.619
β_6	0.071	0.116	0.085	0.156	0.042	$r_{11}, \sigma_{11}, \xi_{11}$	-	-	1,562.979	0.026	10,792.517
β_7	0.070	0.116	0.083	0.153	0.041	$r_{12}, \sigma_{12}, \xi_{12}$	-	-	811.938	0.035	3,739.627
β_8	0.064	0.109	0.077	0.142	0.038	$r_{13}, \sigma_{13}, \xi_{13}$	-	-	422.383	0.049	920.788
β_9	0.057	0.098	0.067	0.124	0.033						

Note: r parameters belong to negative binomial GAS model, σ parameters belong to Gaussian GAS model and ξ parameters belong to beta GAS model.

Table A2: Out of sample MAPE values (%) for the proposed GAS models applied to Japan mortality rates.

Age groups (years)	Poisson	binomial	negative binomial	Gaussian	beta
30-34	21.66%	15.72%	12.33%	14.50%	15.59%
35-39	17.09	11.55	8.13	10.49	10.19
40-44	10.11	5.07	2.57	6.42	1.89
45-49	2.27	3.07	4.13	1.84	5.05
50-54	1.78	6.30	7.06	4.43	6.19
55-59	4.00	1.89	1.34	1.01	1.26
60-64	3.58	0.63	1.95	0.84	1.10
65-69	1.62	4.05	5.73	5.38	4.55
70-74	0.81	5.64	6.85	4.12	7.52
75-79	2.46	4.36	4.79	1.85	5.83
80-84	1.81	6.75	7.06	5.11	8.30
85-89	3.70	6.76	7.05	5.65	7.74
90-95	5.76	4.41	3.96	3.26	5.43
Total	5.90	5.58	5.62	4.99	6.20

Table A3: Out of sample MAPE values (%) for the proposed GAS models applied to Sweden mortality rates.

Age groups (years)	Poisson	binomial	negative binomial	Gaussian	beta
30-34	12.12%	12.22%	13.69%	12.74%	12.59%
35-39	6.03	5.25	5.12	5.71	5.97
40-44	2.18	2.06	1.91	2.48	2.34
45-49	1.47	1.73	2.51	1.37	1.36
50-54	8.33	8.82	10.03	8.87	9.46
55-59	4.75	5.12	7.04	6.23	5.87
60-64	4.14	4.32	6.37	5.43	5.25
65-69	2.65	1.77	1.76	2.58	2.37
70-74	4.17	3.13	2.71	3.97	3.65
75-79	2.17	1.19	1.14	1.93	1.77
80-84	1.52	1.78	1.30	1.74	1.67
85-89	10.51	10.61	17.46	17.41	18.98
90-95	6.73	7.01	20.60	16.99	14.08
Total	5.14	5.53	7.07	6.72	6.57

Table A4: Out of sample MAPE values (%) for the proposed GAS models applied to the UK mortality rates.

Age groups (years)	Poisson	binomial	negative binomial	Gaussian	beta
30-34	4.72%	4.89%	4.83%	4.86%	4.84%
35-39	12.96	12.86	12.73	12.67	13.38
40-44	17.29	15.78	15.30	15.37	16.72
45-49	15.11	12.30	12.16	12.41	13.67
50-54	14.14	11.82	11.70	11.70	12.25
55-59	11.64	9.23	9.58	9.56	9.08
60-64	2.41	1.59	1.43	1.46	1.71
65-69	2.96	3.34	4.62	4.70	6.27
70-74	9.72	10.18	11.36	11.42	13.10
75-79	9.41	9.49	11.15	11.18	12.54
80-84	7.37	7.73	9.01	9.07	10.14
85-89	14.55	14.86	19.64	18.46	20.81
90-95	8.34	8.80	11.41	10.60	11.66
Total	10.05	10.29	10.38	10.27	11.23