“Impact of different distributional assumptions in forecasting Italian mortality rates”

Rosella Giacometti (Italy), Sergio Ortobelli (Italy), Maria Ida Bertocchi (Italy)

Impact of different distributional assumptions in forecasting Italian mortality rates

Abstract

In this paper we value the impact of different distributional assumptions relative to Lee-Carter innovations in forecasting age-specific mortality in Italy. We fit the matrix with Italy death rates from 1960 to 2004, and we observe that the innovation series presents significant kurtosis. We implement the model approximating the innovations with a symmetric Normal Inverse Gaussian (NIG) distribution for different groups of ages. We value the impact of Gaussian and NIG approximations on the distributional hypotheses considering an ex post analysis of the distributional approximation. We observe that for some age groups the NIG distributional assumption on the residuals of the Lee-Carter model produces dominant results compared to the Gaussian one.

Keywords: mortality rates, Normal Inverse Gaussian distribution, AR innovations.
JEL Classification: C53, G22.

Introduction

The decline in the mortality level of populations has induced the social security systems of the most developed countries to reconsider their mortality tables taking into account Longevity Risk. The Lee-Carter model (see Lee and Carter (1992)) represents probably the first model that considers the increased life expectancy trends in mortality rates. Originally applied to USA mortality data, now it is applied to all-cause and cause-specific mortality data from many countries (see Tuljapurkar et al. (2000)). Moreover, many of the recent approaches are consistent with it (see Lifemetrics (2007) and the reference therein) and the literature of the last decade considers it the leading statistical model for forecasting mortality (see, among others, Lee and Miller (2000), Lee (2000), Deaton and Paxson (2004)). The Lee-Carter method combines a demographic model with a statistical model of time series to forecast mortality rates. Referring to Girosi and King (2007), it can be seen as a special type of multivariate process in which the covariance matrix depends on the drift vector and the innovations are intertemporally correlated. With this model, we can define a complete set of death probabilities k_t for a given value of the time index t, the estimated parameters depending on age α_x, β_x remaining constant and invariant through time.

In this paper we relax the Gaussian distributional hypothesis on the stochastic innovations of mortality rates in Lee-Carter model and with an ex post analysis we evaluate the impact of a different distributional assumption, considering the Italian mortality rates of the last eight years. Since we observe that the innovation process is essentially symmetric but presents semi-heavy tails we propose to estimate them using a Normal Inverse Gaussian distribution (NIG) (see Barndorff Nielsen (1995)). Using the analysis proposed by Girosi and King (2007) the error in the mortality rates can be easily simulated from both Gaussian and NIG distributions. Therefore, to evaluate the impact of the new assumptions, we propose an ex post analysis of eight years of Italian mortality tables. From this empirical comparison we observe that generally the absolute errors we get with NIG scenarios are "stochastically smaller" than the Gaussian one.

In section 1, we give the main characteristics of the Lee-Carter model. The second section proposes an empirical comparison of the different distributional assumptions. Finally, we briefly summarize the results.

1. Further insights regarding the Lee-Carter model

Let $m_{x,t}$ be the central death rate for age x in year t. Lee and Carter (1992) suggested a log bilinear form for the force of mortality $\mu_{x,t}$, that is

$$\mu_{x,t} = \ln\left(m_{x,t}\right) = \alpha_x + \beta_x k_t + \varepsilon_{x,t}$$

$$x=1,\dots,A;\ t=1,\dots,T, \qquad (1)$$

where α_x, β_x, k_t are the parameters of the model and $\varepsilon_{x,t}$ is a set of i.i.d. Gaussian errors $N\left(0,\sigma^2\right)$. The random term $\varepsilon_{x,t}$ reflects a particular age-specific historical influence not captured in the model. The coefficients α_x are age specific constants that describe the general shape of the age mortality profile. The index k_t serves to capture the main temporal level of mortality. Since $\partial\mu_{x,t}/\partial t \approx \beta_x\,\partial k_t/\partial t$ (without considering random noises) then the coefficients β_x indicate changes in mortality rates at age x in response to changes in the general level of mortality k_t. So if β_x is large for some *x* (as for infant mortality), then the death rate at age *x* varies significantly when the general level of mortality k_t changes. Conversely, if β_x is small (as for the mortality at older ages), the death rate at that age varies little when the general level of mortality k_t changes. Since the parameterization in (1) is invariant with respect to the trasformations:

$(\beta_x, k_t) \to (c\beta_x, k_t/c)$ or

$(\alpha_x, k_t) \to (\alpha_x - c\beta_x, k_t + c)$

for some $c \in R \setminus \{0\}$, (2)

then the parameters β_x, k_t should satisfy the constraints:

$$\sum_{x=1}^{A} \beta_x = 1; \ \sum_{t=1}^{T} k_t = 0, \tag{3}$$

in order to ensure the identifiability of the model. The constraint $\sum_{t=1}^{T} k_t = 0$ implies that the estimates of parameters α_x are given by the averages of the force of mortality over the time period, i.e., $\hat{\alpha}_x = \frac{1}{T}\sum_{t=1}^{T} \mu_{x,t}$. Considering that $\mu_{x,t} - \hat{\alpha}_x = \beta_x k_t + \varepsilon_{x,t} \approx N\left(\beta_x k_t, \sigma^2\right)$ are Gaussian distributed with mean $\beta_x k_t$ and variance σ^2, then the parameters β_x and k_t can be estimated via maximum likelihood. In particular, as remarked by Lee and Carter (1992), the optimal solution can be found using the Singular Value Decomposition (SVD) of the matrix of centered age profiles $z_{x,t} = \mu_{x,t} - \hat{\alpha}_x$. Given the matrix $\mathbf{Z} = [z_{x,t}]_{\substack{x=1,\dots,A \\ t=1,\dots,T}}$ we can compute the normalized eigenvector $\mathbf{u_1} = [u_{1,1}, \dots, u_{1,T}]'$ (respectively $\mathbf{v_1} = [v_{1,1}, \dots, v_{1,A}]'$) of the matrix $\mathbf{Z'Z}$ (respectively $\mathbf{ZZ'}$) corresponding to the largest eigenvalue λ_1. Then the optimal estimates satisfying the constraints (3) imposed on the parameters, are given by the vectors:

$$\hat{\boldsymbol{\beta}} = [\hat{\beta}_1, \dots, \hat{\beta}_A]' = \frac{\mathbf{v_1}}{\sum_{j=1}^{A} v_{1,j}},$$

and $\hat{\mathbf{k}} = [\hat{k}_1, \dots, \hat{k}_T]' = \lambda_1 \left(\sum_{j=1}^{A} v_{1,j}\right)\mathbf{u_1}$.

Typically for low-mortality populations the approximation $\mathbf{Z} \approx \lambda_1 \mathbf{v_1}\mathbf{u'_1}$ accounts for more than 90% of the variance of $\ln\left(m_{x,t}\right)$. We need a further reestimation step for the parameters k_t since with the above procedure the number of fitted deaths does not equal the number of observed deaths. The parameters $\hat{k}_t$ are adjusted (taking estimates $\hat{\alpha}_x$, $\hat{\beta}_x$ as given) such that the new estimates $\overline{\overline{k}}_t$ solve the equations

$$D_t = \sum_{x=1}^{A} N_{x,t} \exp\left(\hat{\alpha}_x + \hat{\beta}_x \overline{\overline{k}}_t\right) \ t=1,\dots,T,$$

where D_t and $N_{x,t}$ are respectively the total number of deaths in year *t* and the total population with age *x* in year *t*. In order to satisfy the Lee-Carter constraints (3) we should also consider the admissible transformation of type (2): $\beta_x^* = \hat{\beta}_x$; $\alpha_x^* = \hat{\alpha}_x + \hat{\beta}_x \frac{1}{T}\sum_{t=1}^{T} \overline{\overline{k}}_t$; $k_t^* = \overline{\overline{k}}_t - \frac{1}{T}\sum_{t=1}^{T} \overline{\overline{k}}_t$. This reestimation step does not always have a unique solution and some researchers skip this re-estimation stage altogether. That is why we define the estimates of the Lee-Carter model with $\hat{\alpha}_x$, $\hat{\beta}_x$, and $\hat{k}_t$ and we do not consider the above reestimation stage.

In order to forecast future mortality rates Lee and Carter assume that α_x and β_x remain constant over time and the time factor k_t is intrinsically viewed as a stochastic process. They suggest using the following random walk with drift model for k_t:

$$\hat{k}_t = \hat{k}_{t-1} + \theta + \xi_t, \tag{4}$$

where $\xi_t \approx N\left(0, \sigma_{rw}^2\right)$ are i.i.d. Gaussian distributed with null mean and variance σ_{rw}^2. The maximum likelihood estimate of the drift parameter θ is given by $\hat{\theta} = \left(\hat{k}_T - \hat{k}_1\right)/(T-1)$ and the variance estimates $\hat{\sigma}_{rw}^2 = \frac{1}{T-1}\sum_{t=1}^{T-1}\left(\hat{k}_{t+1} - \hat{k}_t - \hat{\theta}\right)^2$. To estimate $\hat{k}_{T+\Delta t}$ at time $T + \Delta t$ we get $\hat{k}_{T+\Delta t} = \hat{k}_T + \left(\Delta t\right)\hat{\theta} + \sqrt{\Delta t}\tilde{\xi}$ where $\tilde{\xi} \approx N\left(0, \sigma_{rw}^2\right)$ and the expected log-mortality can be approximated as follows:

$$\hat{\mu}_{x,T+\Delta t} = \hat{\alpha}_x + \hat{\beta}_x\left(\hat{k}_T + (\Delta t)\hat{\theta}\right) = \hat{\alpha}_x + \hat{\beta}_x\left(\hat{k}_T + \Delta t\frac{\left(\hat{k}_T - \hat{k}_1\right)}{(T-1)}\right).$$

Formulas (1) and (4) can be written in a vectorial way as follows:

$$\boldsymbol{\mu}_t = \bar{\boldsymbol{\mu}} + \boldsymbol{\beta}k_t + \boldsymbol{\varepsilon}_t,$$

$$\hat{k}_t = \hat{k}_{t-1} + \theta + \xi_t, \qquad (5)$$

where $\boldsymbol{\mu}_t = [\mu_{1,t},\ldots,\mu_{A,t}]'$, $\bar{\boldsymbol{\mu}} = \frac{1}{T}\sum_{t=1}^{T}\boldsymbol{\mu}_t$, $\boldsymbol{\beta} = [\hat{\beta}_1,\ldots,\hat{\beta}_A]'$ while vector $\boldsymbol{\varepsilon}_t$ and the scalar ξ_t are independent Gaussian errors with variances $\sigma^2 I$ (here I is the $A\times A$ identity matrix) and σ^2_{rw} respectively. Referring to the discussion in Girosi and King (2007), we propose viewing the Lee-Carter model as follows:

$$\boldsymbol{\mu}_t = \boldsymbol{\mu}_{t-1} + \boldsymbol{\beta}\theta + \left(\boldsymbol{\beta}\xi_t + \boldsymbol{\varepsilon}_t - \boldsymbol{\varepsilon}_{t-1}\right) = \boldsymbol{\mu}_{t-1} + \boldsymbol{\psi} + \left(\frac{\boldsymbol{\psi}}{|\boldsymbol{\psi}|}\xi_t + \boldsymbol{\varepsilon}_t - \boldsymbol{\varepsilon}_{t-1}\right), \quad (6)$$

where $\boldsymbol{\psi} = \boldsymbol{\beta}\theta = [\psi_1,\ldots,\psi_A]'$, $|\boldsymbol{\psi}| = \sum_{i=1}^{A}\psi_i = \theta$, and the variance-covariance matrix

$$\sum\nolimits_{LC} = \sigma^2_{rw}\frac{\boldsymbol{\psi}\boldsymbol{\psi}'}{|\boldsymbol{\psi}|^2} + 2\sigma^2 I \qquad (7)$$

of the noise is a function of the drift vector $\boldsymbol{\psi}$. The first term in (7) describes the shocks that are perfectly correlated across age groups and the second term those that are uncorrelated across age groups. Girosi and King define (6) as a random walk with drift model, while, the innovations in (6) are not i.i.d. since there is a non negative autocorrelation of the first order.

Notice that in the classic Lee-Carter model, σ is the same for the considered groups of ages. In case we restrict the analysis to a single age, σ will depend on age.

1.1. The Lee-Carter model with different distributional assumptions. Let us assume that the i.i.d. error terms $\varepsilon_{x,t}$, ξ_t follow infinitely divisible distributions different from the Gaussian one. This assumption takes into account the skewness and the semi-heavy tails often observed in the innovation distributions. Using infinitely divisible distributions the sum of i.i.d. components belongs to the same family of infinitely divisible distributions. The characteristic function of infinitely divisible distributions is univocally determined by the triplet $[\gamma,\sigma^2,\nu]$ that identifies the so-called Lévy-Khintchine characteristic exponent $\varphi(u) = \log\phi(u)$ given by:

$$\varphi(u) = i\gamma u - \frac{1}{2}\sigma^2 u^2 + \int_{-\infty}^{+\infty}(\exp(iux) - 1 - iux1_{\{|x|<1\}})\nu(dx),$$

where $y \in R$, $\sigma^2 > 0$ and ν is a measure on $R\setminus\{0\}$ with $\int_{-\infty}^{+\infty}(1\wedge x^2)\nu(dx) < \infty$. In particular the Lévy triplet $[\gamma,\sigma^2,\nu]$ identifies the three main components of any Lévy process: the deterministic component (γ), the Brownian component (σ^2) and the pure jump component (ν). For further details on the theoretical aspects we refer to Sato (1999). Thus, if $\varepsilon_{x,t}$, ξ_t follow two alternative infinitely divisible distributions with characteristic exponents respectively $\varphi_1(u)$, $\varphi_2(u)$, then according to the random walk model with drift (6), an individual with age x should present the force of mortality:

$$\mu_{x,t} = \mu_{x,t-1} + \beta_x\theta + \left(\beta_x\xi_t + \varepsilon_{x,t} - \varepsilon_{x,t-1}\right)$$

x=1,…, A; t=1,…,T, (8)

where the error term $\beta_x\xi_t + \varepsilon_{x,t} - \varepsilon_{x,t-1}$ is uniquely determined by its characteristic function. Therefore, considering that the terms $\beta_x\xi_t$, $\varepsilon_{x,t}$, $\varepsilon_{x,t-1}$ are independent, then the characteristic function of the global error term is given by:

$$\phi_{\beta_x\xi_t+\varepsilon_{x,t}-\varepsilon_{x,t-1}}(u) = \exp\left(\varphi_2(\beta_x u) + \varphi_1(u) + \varphi_1(-u)\right). \quad (9)$$

Clearly, the variance covariance of the vector of errors is still given by formula (7) but the singular variance terms σ^2 and σ^2_{rw} depend on the parameters characterizing the different infinitely divisible distributions of $\varepsilon_{x,t}$ and ξ_t. In particular, next, we consider the Normal Inverse Gaussian distribution (NIG) (see Barndorff-Nielsen (1995)) that, differently from Gaussian distribution, presents skewness and semi heavy tails. An NIG distribution $NIG(\alpha,\beta,\delta,q)$ depends on the parameters $\alpha > 0$, $\beta \in (-\alpha,\alpha)$, $\delta > 0$, $q \in R$. The characteristic function of an NIG is given by:

$$\phi_{NIG}(u;\alpha,\beta,\delta,q) =$$
$$= \exp\left(-\delta\left(\sqrt{\alpha^2 - (\beta+iu)^2} - \sqrt{\alpha^2-\beta^2}\right) + iuq\right)$$

and the density is given by:

$$f_{NIG}(x;\alpha,\beta,\delta,q) = \frac{\delta\alpha \mathbf{K}_1(\alpha\sqrt{\delta^2+(x-q)^2})}{\pi\sqrt{\delta^2+(x-q)^2}} \times \times \exp\left(\delta\sqrt{\alpha^2-\beta^2}+\beta(x-q)\right),$$

where $\mathbf{K}_\lambda(x)$ denotes the modified Bessel function of the third kind with index λ. To estimate the parameters we can use either maximum likelihood estimation or semi-parametric methods (see, among others, Caviezel et al. (2009)). Observe that if a random variable X follows a $NIG(\alpha,\beta,\delta,q)$ then the opposite $-X$ follows a $NIG(\alpha,-\beta,\delta,-q)$. In particular, q represents a location parameter, while β is a parameter of symmetry and when it is equal to zero the distribution is symmetric around the mean q (when $\beta = 0$). This can be easily derived considering that the mean, the variance, the skewness and the kurtosis of a NIG variable are simply given by:

$$Mean = \frac{\delta\beta}{\sqrt[2]{\alpha^2-\delta^2}} + q;$$

$$Variance = \frac{\alpha^2\beta}{\sqrt[2]{\left(\alpha^2-\delta^2\right)^3}};$$

$$Skewness = \frac{3\beta}{\alpha\sqrt{\beta}\cdot\sqrt[4]{\alpha^2-\beta^2}};$$

$$Kurtosis = 3\left(1+\frac{\alpha^2+4\beta^2}{\delta\alpha^2\sqrt{\alpha^2-\beta^2}}\right). \qquad (10)$$

Next, we characterize the error terms of Lee-Carter model for the Italian mortality rate.

2. An empirical analysis based on the Italian mortality rate

In this section we implement the Lee-Carter model on Italian dead/exposure to death data taken from the "Human Mortality Database" (see the references), available from 1922 to 2004. We choose an opportune range of data (from 1960 to 2004) in order to have a reliable and complete data set for ages 50 to 90. We propose the following empirical analysis: first, we discuss how to model the time factor k_t and we propose, following Girosi and King (2007), the random walk model as in (4). Further, we propose an ex-post comparison considering simulated data from both a Gaussian and a NIG distributions.

Figure 1 gives the estimation of the Lee-Carter temporal level $\hat{k}_t$ of mortality and its adjusted re-estimation for the Italian population (1960-1996).

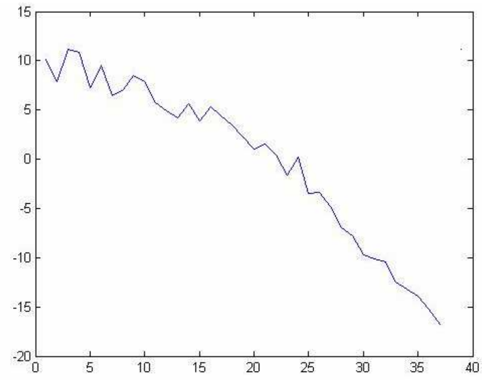


Fig. 1A. Italian $\hat{k}_t$ estimation (1960-1996)

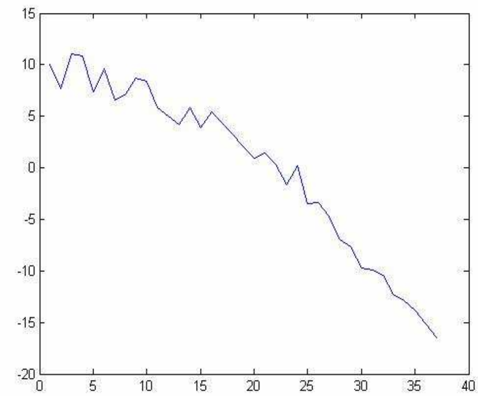


Fig. 1B. Adjusted Italian k_t^* estimation (1960-1996)

We consider Italian mortality from 1960 until 1996 for the population of age 50 to 90. In particular, Figure 1 plots the estimated $\hat{k}_t$ and its adjusted estimation k_t^* so that the number of fitted deaths equals the number of observed deaths. Since there is not a big difference between the two alternative estimates and we want to value the distributions of the statistical approximations, we operate with the first estimated $\hat{k}_t$ that is not adjusted. In order to value the forecasting power of the Lee-Carter model we propose to model $\hat{k}_t$ with a random walk with drift, see equation (4).

From a first analysis of the residuals of the model, we observe that the residuals of the AR(1) model with drift are substantially i.i.d. Gaussian (no autocorrelation is revealed, the residual mean is 0, standard deviation is 1.566 and the kurtosis is 3.3). We tested for the presence of the unit root through the ADF (Adjusted Dickey-Fuller) test and it turned out that we could not reject a null hypothesis for the presence of the unit root for different lags and different significance levels (from 0.01 to 0.05). These results and the scarcity of data prevented us from testing an ARMA-GARCH model on $\hat{k}_t$. Then, we tested ex-post the capacity of the model to forecast future Italian mortality rates. Starting from the last year of observation used to value the model, we forecast 8 years $\hat{k}_t$ and we inserted them into the Lee-Carter model (1). We computed the difference between the forecasted mortality rates and the real ones for the last eight years and we observed that the Lee-Carter forecasting for Italian mortality rates could be further improved.

2.1. Lee-Carter model with NIG distributions. As further empirical analysis we valued the impact of a NIG distributional assumptions on the residuals of the Lee-Carter model. From a preliminary analysis of the empirical residuals:

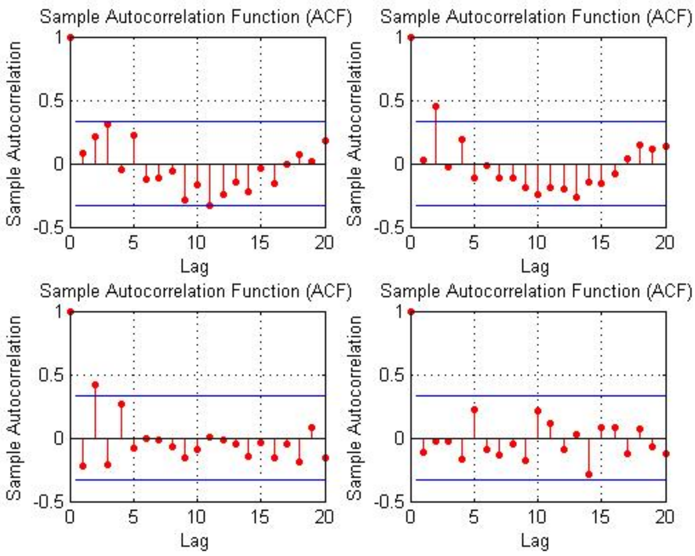
$$\hat{\varepsilon}_{x,t} = \mu_{x,t} - \hat{\alpha}_x - \hat{\beta}_x \hat{k}_t$$

we verified that they are not autocorrelated (see Figure 2), they are not Gaussian distributed and present semi-heavy tails.

Table 1. Variance, skewness and kurtosis of Lee-Carter residuals

AGE	50	51	52	53	54	55	56	57	58	59	60	61	62	63
Variance	.002	.0016	.0018	.0015	.0009	.0011	.0010	.0013	.0014	.0010	.0011	.0009	.0012	.0009
Skewness	-0.04	0.189	0.154	0.451	0.365	-0.08	-0.46	0.01	-0.235	-0.369	-0.29	-0.05	-0.22	0.042
Kurtosis	2.989	2.67	2.99	2.831	2.849	4.238	3.745	3.997	3.634	3.201	3.465	3.157	4.522	3.088
AGE	64	65	66	67	68	69	70	71	72	73	74	75	76	77
Variance	.0010	.0012	.0011	.0011	.0010	.0009	.0011	.0012	.0011	.0010	.0009	.0008	.0009	.0008
Skewness	-0.183	0.025	-0.527	-0.31	-0.03	-0.99	-0.25	-0.124	0.328	0.493	0.401	0.722	0.932	1.115
Kurtosis	3.376	4.625	3.668	3.944	5.054	7.374	4.278	3.808	4.985	4.6	4.473	4.738	5.966	6.178
AGE		78	79	80	81	82	83	84	85	86	87	88	89	90
Variance		.0008	.0010	.0005	.0004	.0003	.0004	.0006	.0004	.0004	.0007	.0004	.0008	.0010
Skewnless		1.304	0.573	-0.34	-0.28	-0.6	-0.36	-0.542	-0.81	-0.177	-0.18	-0.69	-1.23	-0.161
Kurtosis		7.412	5.356	3.017	3.598	4.485	3.607	4.224	5.432	3.246	3.851	4.496	7.683	3.51

Notes: This table gives the variance, skewness and kurtosis of Lee-Carter residuals for each age from 50 till 90 for Italian mortality between 1960-1996.



Notes: This figure shows the absence of autocorrelation in the residuals for ages: 50, 60, 70, 80 in the period of 1960-1999.

Fig. 2. Autocorrelation of Lee-Carter residuals

This fact is confirmed by Table 1, which gives the variance, the skewness and the kurtosis of empirical residuals $\hat{\varepsilon}_{x,t}$ for each age x from 50 to 90. The Kolmogorov-Smirnov, Jarque-Bera and Lilliefors tests on the normality of residuals reject the null hypothesis of normality for the whole dataset (1887 data) with a kurtosis equal to 5.95. Table 1 tells us that residuals are non-Gaussian distributed for the oldest ages, since they present skewness and heavier tails than a Gaussian distribution. Even Figure 3 shows that the empirical residuals present several peaks for the population with ages 50, 60, 70, 80.

Since there was empirical evidence of a non-Gaussian distribution for the Lee-Carter residuals we estimated the residuals $\hat{\varepsilon}_{x,t}$ for Italian individuals with age x varying from 50 to 90 with a $NIG(\alpha,\beta,\delta,0)$. The NIG maximum likelihood parameters for the residuals were $\alpha = 32.2327$, $\beta = 0$, $\delta = 0.0458$. Thus, we deduced that the NIG estimates for residuals were symmetric around the null mean, with variance δ / α and kurtosis $3\dfrac{\delta\alpha+1}{\delta\alpha}$. These results partially confirm those of Table 1. Table 1 shows that the skewness is significant only in 5 cases out of 40 (the 5% critical value for skewness is 0.745 for n=40) whereas kurtosis is significant in 15 cases out of 40 (the 5% critical value for kurtosis is 4.37 for n=40 given a one-side hypothesis). This justifies our choice for a NIG with $\beta = 0$.

In order to value the impact of this distributional assumption we simulated different scenarios of the exposure to death for the last eight years, and we used the random walk model with Gaussian residuals ξ_t to model $\hat{k}_t$.

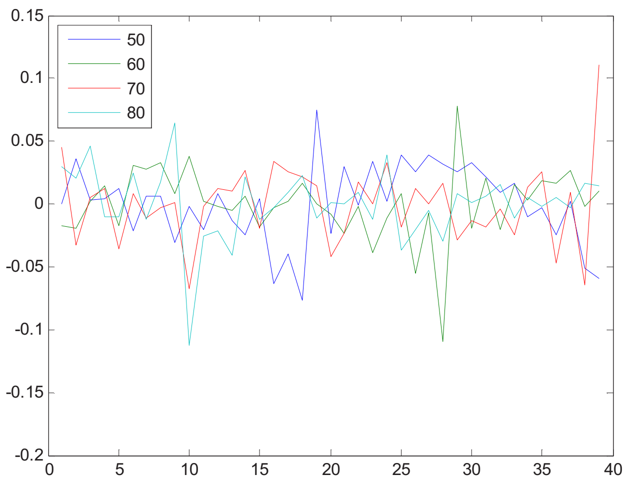
We compared ex-post the performance we obtained modeling the Lee-Carter residuals $\hat{\varepsilon}_{x,t}$ either with a NIG or with Gaussian distributions. We used the last 8 years (1996-2004) of Italian mortality data. For each scenario we computed the average over the 8 years for the individual absolute difference between the forecasted and the historical ones. Since these are absolute errors they represent the occurrences of positive random variables that we can suppose are uniquely determined by their mean and standard deviation. Notice that, due to the estimate parameter β being statistically not different from zero in the NIG distribution, the NIG depends only on two parameters.

Thus, we can determine if types of stochastic dominance orderings among these positive errors exist. In particular, we can use the stochastic dominance rules proposed by Ortobelli (2001) to compare positive random variables belonging to a translation and scalar invariant family of distribution determined by only two parameters: the mean *m* and the standard deviation *σ*. In particular, we summarized three of these stochastic dominance rules:

1) $\dfrac{m_X}{\sigma_X} \geq \dfrac{m_Y}{\sigma_Y}$ and $\sigma_X \geq \sigma_Y$ (with at least one inequality strict) implies X dominates at the first stochastic order (FSD) Y.

2) $\dfrac{m_X}{\sigma_X} \geq \dfrac{m_Y}{\sigma_Y}$ and $m_X \geq m_Y$ (with at least one inequality strict) implies X dominates at the second stochastic order (SSD) Y.

3) $m_X \geq m_Y$ and $\sigma_X \leq \sigma_Y$ (with at least one inequality strict) implies $X\ SSD\ Y$.



Notes: This figure shows the empirical Lee-Carter residuals time series plot for Italians of ages: 50, 60, 70, 80 in the period of 1960-1999.

Fig. 3. Time series plot of Lee-Carter residuals

Table 2. NIG parameters and orderings between absolute errors

Parameters NIG									
AGE	**0-11**	**11-21**	**21-31**	**31-41**	**41-51**	**51-61**	**61-71**	**71-81**	**81-91**
Alpha	23.0275	91.5494	20.1377	14.7931	34.7884	29.4403	30.8112	30.6179	105.058
beta	0	0	0	0	0	0	0	0	0
delta	0.1609	0.3621	0.0839	0.0877	0.0639	0.03	0.0237	0.0215	0.0231

Table 2 (cont.). NIG parameters and orderings between absolute errors

AGE	0-11	11-21	21-31	31-41	41-51	51-61	61-71	71-81	81-91
Mean and standard deviation of the absolute error we get simulating NIG residuals									
Mean	0.2471	0.2025	0.2144	0.318	0.0679	0.0775	0.0712	0.0735	0.0829
STD	0.0691	0.0575	0.0604	0.0536	0.0123	0.0295	0.0325	0.0322	0.0373
mean/STD	3.57598	3.52174	3.54966887	5.932836	5.52033	2.627119	2.19077	2.282609	2.22252
Mean and standard deviation of the absolute error we get simulating Gaussian residuals									
Mean	0.2484	0.2022	0.2147	0.317	0.0681	0.0775	0.0716	0.0736	0.083
STD	0.0693	0.0575	0.0606	0.054	0.0121	0.0293	0.0323	0.0321	0.0373
mean/STD	3.584416	3.51652	3.54290429	5.87037	5.6281	2.645051	2.21672	2.292835	2.2252
Stochastic	Gaussian	NIG SSD	NON	NIG SSD	Gaussian	Gaussian	Gaussian	Gaussian	Gaussian
dominance	FSD NIG	Gaussian	comparable	Gaussian	SSD NIG	SSD NIG	SSD NIG	SSD NIG	SSD NIG
Rule applied	1	2-3	//	2-3	2-3	2-3	2-3	2-3	2-3

Notes: This table gives: 1. The NIG (MLE) estimated parameters for Lee-Carter residuals of group ages. 2. The mean and standard deviation of the absolute error we get simulating either Gaussian or NIG residuals. 3. The stochastic dominance relationships (based on mean and standard deviation) between the absolute errors we get simulating either Gaussian or NIG residuals.

We tested the stochastic dominance using means and standard deviations obtained by a large number of simulated scenarios from the NIG and Gaussian residuals' distributions. In particular, computing the mean and the standard deviation of absolute errors on twenty thousand scenarios generated by NIG and Gaussian residuals, we obtained, for the NIG scenarios, a mean m_{NIG} =0.0735 and a standard deviation σ_{NIG} =0.0176 and, for the Gaussian scenarios, a mean $m_{Gaussian}$ =0.0738 and standard deviation $\sigma_{Gaussian}$ =0.0174. Since

$$\frac{m_{Gaussian}}{\sigma_{Gaussian}} = 4.24138 \geq \frac{m_{NIG}}{\sigma_{NIG}} = 4.1761,$$

$m_{Gaussian} \geq m_{Gaussian}$ then the absolute errors obtained with the Gaussian scenarios of residuals dominate at the second stochastic order over the errors obtained with NIG residuals. In some sense the "NIG errors" are stochastically smaller than the Gaussian ones. We also extended this ex-post comparison considering different age groups. In particular, in Table 2 we give the maximum likelihood estimates of residuals NIG parameters and we see that even for groups of age we estimate symmetric residuals around the null mean. Then we generated a consistent number of scenarios and Table 2 shows that the absolute errors we get with Gaussian scenarios generally dominate at the second stochastic order over the same errors we get with NIG scenarios except for age groups 11-21 and 31-41 where the NIG errors dominate over the Gaussian ones. Therefore, this empirical analysis confirms that it makes sense to use distributions with semi heavy tails for the residuals of the mortality rates.

Concluding remarks

This paper compares alternative distributional assumptions and modeling for forecasting the Italian mortality rate. First, we analyzed the historical series considered in the Italian mortality rate. The observed skewness and kurtosis together with other empirical tests indicate that the residuals of Lee Carter model are not Gaussian distributed.

Secondly, we proposed an expost comparison with simulated scenarios based on either Normal Inverse Gaussian residuals or simulated Gaussian ones. This analysis shows that the absolute errors we create using NIG distributions are generally stochastically smaller than those we get with Gaussian residuals.

Acknowledgments: Research has been partially supported by PRIN 2008 grant "Financial Innovation and demographic changes: new products and pricing instruments with respect to the stochastic factor 'ageing'" (Bergamo unit coordinated by Bertocchi) within the national project "The impact of population ageing on financial markets, intermediaries and financial instability", (Modena central unit coordinated by Torricelli) and by Fondi Ateneo 2008 grant "Un modello stocastico per il tasso di mortalità per la valutazione di Guaranteed Annuity Options".

References

1. Barndorff-Nielsen, O.E. Normal Inverse Gaussian Distributions and the modeling of stock returns // Research Report, 1995, N. 300, Department of Theoretical Statistics, Aarhus University.

2. Caviezel, V., S., Ortobelli, S.T. Rachev. Semi-parametric estimators for heavy tailed distributions. Technical Report, 2009, Institute of Statistics and Mathematical Economic Theory, University of Karlsruhe (to appear in Journal of Concrete and Applicable Mathematics).
3. Deaton, A., C. Paxson. Mortality, Income, and Income Inequality Over Time in the Britain and the United States // Technical Report, 2004, N. 8534, National Bureau of Economic Research Cambridge, MA:. http://www.nber.org/papers/w8534.
4. Girosi F., King G. Understanding the Lee-Carter Mortality Forecasting Method // Technical Report, 2007, Rand Corporation.
5. Human Mortality Database. University of California, Berkeley (USA), available at www.mortality.org
6. Lee, R.D., L.R., Carter. Modelling and Forecasting U.S. Mortality // Journal of the American Statistical Association, 1992, N. 87, pp. 659-671.
7. Lee, R.D. The Lee-Carter Method for Foresting Mortality, with Various Extensions and Applications // North American Actuarial Journal, 2000, N. 4, pp. 80-93.
8. Lee, R.D., T. Miller. Assessing the Performance of the Lee-Carter Approach to Modeling and Forecasting Mortality // 2000, Available at www.demog.berkeley.edu
9. LifeMetrics. A toolkit for measuring and managing longevity and mortality risks // Technical Report, 2007.
10. Ortobelli S. The classification of parametric choices under uncertainty: analysis of the portfolio choice problem // Theory and Decision, 2001, N. 51, pp. 297-327.
11. Sato, K. *Lévy Processes and Infinitely Divisible Distributions* // Cambridge University Press: Cambridge, UK, 1999.
12. Tuljapurkar, S., N. Li, C. Boe. A universal pattern of mortality decline in the G-7 countries // Nature, 2000, N. 405, pp. 789-792.