Journal of Data Science **12**(2014), 1-18

Two Factor Stochastic Mortality Modeling with Generalized Hyperbolic Distribution

Seyed Saeed Ahmadi* and Patrice Gaillardetz
Concordia University

Abstract: In this paper, we reconsider the two-factor stochastic mortality model introduced by Cairns, Blake and Dowd (2006) (CBD). The error terms in the CBD model are assumed to form a two-dimensional random walk. We first use the Doornik and Hansen (2008) multivariate normality test to show that the underlying normality assumption does not hold for the considered data set. Ainou (2011) proposed independent univariate normal inverse Gaussian Lévy processes to model the error terms in the CBD model. We generalize this idea by introducing a possible dependency between the 2-dimensional random variables, using a bivariate Generalized Hyperbolic distribution. We propose four non-Gaussian, fat-tailed distributions: Student's t, normal inverse Gaussian, hyperbolic and generalized hyperbolic distributions. Our empirical analysis shows some preferences for using the new suggested model, based on Akaike's information criterion, the Bayesian information criterion and likelihood ratio test, as our in-sample model selection criteria, as well as mean absolute percentage error for our out-of-sample projection errors.

Key words: Generalized hyperbolic distribution, Doornik-Hansen test, stochastic mortality model.

1. Introduction

It is now well established that mortality should be modeled as a stochastic process and that longevity has improved over the past century. We should consider that this improvement is also random and cannot be easily predicted. Many authors have tried to find reasonable mortality models in recent years. The obvious reason behind all this research is that the improved longevity costs money to the insurance industry. To explain it further, consider a life annuity business or even a pension plan. If longevity improves and annuitants or pensioners live

*Corresponding author.

longer than were expected, then the benefits will be paid out over a longer period of time. The consequences can result in much more adverse outcomes if we consider a large portfolio of annuitants.

Different approaches can be considered when forecasting mortality rates. Specifically, age-period-cohort models as well as the model proposed in the prominent paper by Lee and Carter (1992) and its generalizations are among the most successful methods. Detailed explanations of these models can be found in Lee and Miller (2001), Renshaw and Haberman (2003), Jong and Tickle (2006).

As mentioned in Cairns *et al.* (2006), there are 3 different mortality risks for insurance companies that offer annuities and life insurance products. We summarize them here:

- Mortality risk: that means fluctuations of mortality rates over time.
- Longevity risk: which can be considered as any randomness in the long-term trend of mortality rates.
- Short-term, catastrophic mortality risk: that can be explained by any sudden phenomena like the influenza pandemic in 1918, the tsunami of December 2004 in Indonesia and of 2011 in Japan.

Some authors try to add jump processes to model mortality, mostly incorporated in the Lee-Carter framework. See for example, Wang *et al.* (2011), Giacometti *et al.* (2009) and Hainaut and Devolder (2008).

Over relatively short period of time, longevity has become a considerable risk in countries like Japan and Taiwan. This brings a significant attention to longevity projections. In this paper, longevity risk is modeled according to Cairns *et al.* (2006) that includes two stochastic factors. They proposed a bivariate normal distribution to model the dynamic of the stochastic factors. However, the error terms of the CBD model seem to have tails thicker than those of a normal distribution, as we show in our empirical analysis. Therefore, we consider a family of bivariate generalized hyperbolic (GH) distributions to model the error terms in the CBD model. This family of distributions have semi-heavy tails with a dependence structure. We use this desirable property to model longevity risk. Thus, we suggest using the bivariate generalized hyperbolic (GH) distribution in the CBD model.

We consider four non-Gaussian distributions within the GH class, which include Student's t, normal inverse Gaussian, hyperbolic and generalized hyperbolic distributions. Next, in order to compare our model with the CBD model, we use likelihood ratio test, Akaike's information criterion (AIC) and the Bayesian information criterion (BIC) as our in-sample model selection criteria. In addition, for the out-of-sample performance, we project mortality rates and apply the mean

absolute percentage error to the proposed model in order to indicate some preferences.

The paper is organized as follows: In Section 2, a brief review of the CBD model is given together with the data set that we used to fit our model. Then, we test the assumptions given in the CBD model by applying Doornik-Hansen's multivariate normality test in Section 3. Next, generalized hyperbolic distributions are defined in Section 4. We propose and fit our model in Section 5. Section 6 is devoted to model comparisons. In Section 7, we summarize the empirical results for three more data sets including mortality data for Russia, Spain and the U.S. Conclusions are given in Section 8.

2. Review of the CBD Model

In this section, we summarize the two-factor stochastic mortality model proposed by Cairns *et al.* (2006). The realization of the one-year survival probabilities for the cohort aged x and still alive at time t is denoted by $\tilde{p}(t,x)$. Furthermore, the realized mortality rate is defined by $\tilde{q}(t,x) = 1 - \tilde{p}(t,x)$. For their empirical analysis, they choose the following model for the mortality curve:

$$\tilde{q}(t,x) = \frac{e^{A_1(t+1)+A_2(t+1)(x)}}{1+e^{A_1(t+1)+A_2(t+1)(x)}}, \tag{1}$$

where they assume that $\boldsymbol{A}(t) = (A_1(t), A_2(t))'$ is a two-dimensional random walk with drift. The first factor influences the changes in mortality at all ages equivalently, while the second factor alters mortality rates at higher ages much more than at lower ages as mentioned in Cairns *et al.* (2006). To make forecasts of the future distribution for $\boldsymbol{A}(t)$, they propose to model the factors in $\boldsymbol{A}(t)$ according to,

$$\boldsymbol{A}(t+1) = \boldsymbol{A}(t) + \boldsymbol{\mu} + C\boldsymbol{Z}(t+1), \tag{2}$$

where $\boldsymbol{\mu}$ is a constant 2×1 vector, C is a constant 2×2 upper triangular matrix and $\boldsymbol{Z}(t)$ is a two-dimensional standard normal random variable.

Similarly to Cairns *et al.* (2006), we use an ordinary least square method to estimate $A(t)$. Then, we change the distributional assumption in (2); next we apply the maximum likelihood method to estimate the dynamic properties of the two factors.

The first contribution of this paper is to show that the normality assumption in (2) cannot describe well the historical data and should be tested before making any further inference. We also, propose GH distribution to use in the CBD model and show how it can provide a better fit. First we explain the data set that we use for our empirical illustration.

We rely on mortality data for Males in Italy, 1969-2008 and ages 60-90. The data set was obtained from the Human Mortality Database (HMD[1]). We deliberately consider more recent data, since this period of time exhibit less levels of uncertainty compared to the mortality data for first half or two-thirds of 20th century. This allows us to have the data with less volatility and develop our model accordingly.

3. Doornik-Hansen's Multivariate Normality Test

In order to estimate the mean and the variance-covariance matrix in (2), we first need to estimate $\boldsymbol{A}(t)$. To do so, the ungraduated mortality rates for each t are transformed from $\tilde{q}(t,x)$ to

$$\log\left(\frac{\tilde{q}(t,x)}{\tilde{p}(t,x)}\right) = A_1(t+1) + A_2(t+1)(x). \tag{3}$$

Then, the linear regression is applied in (3) to estimate $\boldsymbol{A}(t)$. It is clear from (2) that

$$\begin{aligned} E[\boldsymbol{A}(t+1) - \boldsymbol{A}(t)] &= \boldsymbol{\mu}, \\ \mathrm{Var}[\boldsymbol{A}(t+1) - \boldsymbol{A}(t)] &= CC^{'}. \end{aligned} \tag{4}$$

Equation (4) shows that the mean and the variance of the first consecutive differences, $\boldsymbol{A}(t+1) - \boldsymbol{A}(t)$, can be used to estimate $\boldsymbol{\mu}$ and $V = CC^{'}$, respectively. We use R as our software environment.[2] The estimations are given in Table 1. Generally, the negative value for μ_1 shows mortality improvement. At the same time, the positive value for μ_2 indicates that mortality rates at higher ages are improving at a slower rate. These results are consistent with those of Cairns *et al.* (2006) in the original CBD model.

Table 1: Estimated mean and variance matrices for the CBD model in (4)

Mortality data	$\hat{\boldsymbol{\mu}}$	$\hat{V}$
Italy	$\begin{bmatrix} -0.056037813 \\ 0.000529807 \end{bmatrix}$	$\begin{bmatrix} 0.0524230293 & -0.0006519287 \\ -0.0001527592 & 0.000002196178 \end{bmatrix}$

Before explaining the Doornik and Hansen normality test, we need to check our data (i.e., the increments $\boldsymbol{A}(t+1) - \boldsymbol{A}(t)$) for any serial autocorrelation, dependency as well as conditional Heteroscedasticity (non-constant variance). These are requirements for the hypothesis tests that we use in this paper.

[1]The data are available online at www.mortality.org.

[2]R is a share-ware for statistical computing and graphics, http://www.r-project.org/

In order to test the serial autocorrelation, we obtain the sample autocorrelation function (ACF), r_k at lag k defined by:

$$r_k = \frac{\sum_{t=k+1}^{n}(Y_t - \bar{Y})(Y_{t-k} - \bar{Y})}{\sum_{t=k+1}^{n}(Y_t - \bar{Y})^2}, \quad k = 1, 2, \cdots, \tag{5}$$

where $Y_t = A_i(t+1) - A_i(t),\ i = 1, 2$. Figure 1 shows the sample ACF of $A_1(t+1) - A_1(t)$ and $A_2(t+1) - A_2(t)$ with the 95 % confidence limits (dotted lines). Looking at this figure, we can see that the sample ACF's are not statistically significant. Therefore, there is no indication of serially autocorrelated increments.

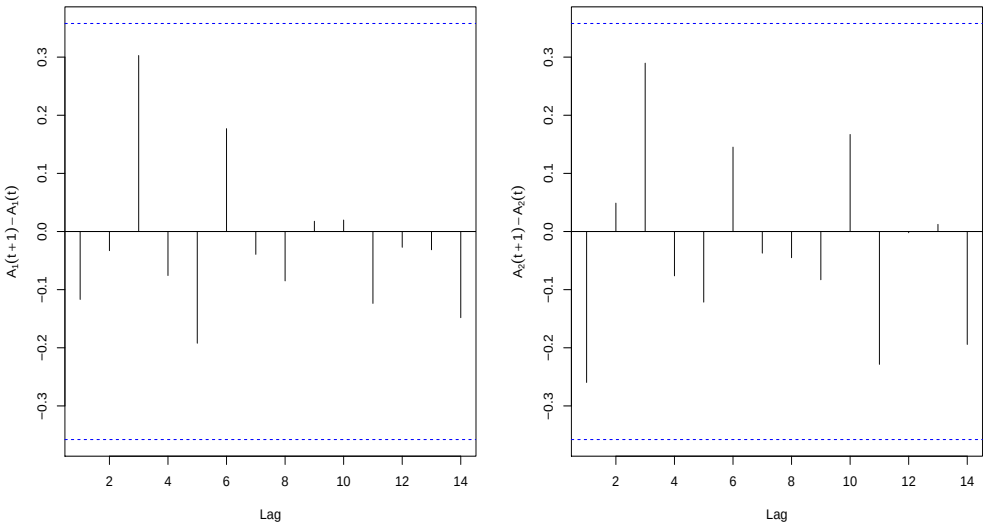

Figure 1: Sample ACF plot for Italy-Data: 1969-1999

The Ljung-Box test (Ljung and Box, 1978) can be applied to test for independence of $A(t+1) - A(t)$. The test statistics is defined as:

$$Q(k) = n(n+2)\sum_{m=1}^{k}\frac{\hat{r}_k^2}{n-m}, \tag{6}$$

where n is the number of observation and $\hat{r}_k$ is the estimated sample ACF defined in (5). The null hypothesis is the linear independence in $A(t+1) - A(t)$. Under the null hypothesis, $Q(k)$ has an asymptotic chi-squared distribution with k degrees of freedom. The null hypothesis is rejected when the value of $Q(k)$ is greater than the selected critical value of chi-squared distribution with k degrees of freedom. Table 2 shows the value of the test statistic, $Q(k)$, the degrees of freedom, *df* and the p-value of the Ljung-Box test. Therefore, we cannot reject the null hypothesis of linear independence at a significance level of 0.05.

Table 2: Ljung-Box test of independency

Data	$Q(k)$	df	p-value
$A_1(t+1) - A_1(t)$	15.1886	24	0.9151
$A_2(t+1) - A_2(t)$	19.9168	24	0.7015

We also check the increments for the assumption of identical independent distribution (iid) by applying the test proposed by McLeod and Li (1983). The null hypothesis assumes that the data are iid. The test statistic is similar to Ljung-Box test (6), except that for the sample ACF, $\hat{r}_k$, which is replaced by the sample autocorrelations of the squared data. This is due to the fact that if the data are iid then square of the data are iid as well. The McLeod-Li test can also be used to assess the conditional Heteroscedasticity of the data. The test statistic is chi-squared distributed with k degrees of freedom under the null hypothesis of iid. Figure 2 shows the p-values of the McLeod-Li test, evaluated up to lag 14. The red dashed line represents the 5% confidence level. Based on this test, we cannot reject the null hypothesis of identical independent distribution.

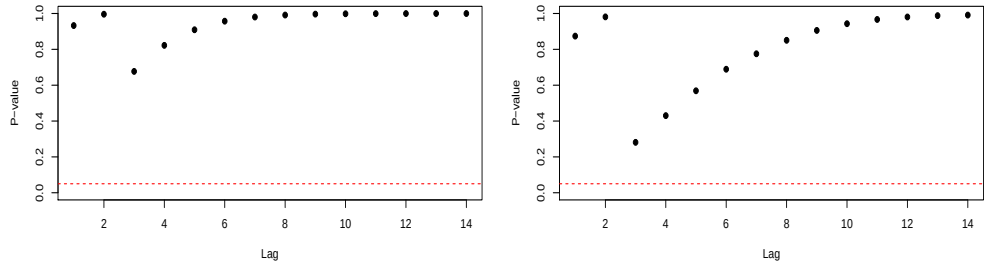

Figure 2: p-values of the McLeod-Li test in Italy for $A_1(t+1) - A_1(t)$ (left) and $A_2(t+1) - A_2(t)$ (right)

We can now test the bivariate normality assumption in Cairns *et al.* (2006) by using the method, proposed by Doornik and Hansen (2008). The latter compares their suggested test with four other tests for multivariate normality and conclude that it has the best size and power properties over the other tests considered. This test is relatively simple, it controls most sizes very well and can be applied for samples as low as 10 observations. The Doornik-Hansen's test for multivariate normality is based on the skewness and kurtosis of multivariate data that is transformed to insure independence, as explained in Appendix A. The package and corresponding function that we used to perform Doornik-Hansen's test are *asbio* and *DH.test*, respectively.

The results are reported in Table 3. The multivariate section of Table 3 indicates that the test statistic is significant and based on the p-value of the test, the bivariate normality assumption is rejected at a significance level of 0.05. The

table also contains the univariate tests for normality. The univariate normality assumption is rejected at the 5% significance level. Consequently, the multivariate normality assumption does not hold.

Table 3: Doornik and Hansen (2008) normality test

Mortality data for Italy	Test statistics	df	p-value
Multivariate	25.20	4	4.6×10^{-4}
Univariate	19.74	2	5.2×10^{-4}
	5.46	2	0.065

Moreover, we applied a Shapiro-Wilk test for multivariate data sets using the package *mvnormtest*. Details of Shapiro-Wilk's test can be found in Royston (1982a, 1982b, 1995). Table 4 summarizes the results. The normality assumption is rejected based on this test as well.

Table 4: Multivariate Shapiro-Wilk normality test

Mortality data	Test statistic	p-value
Italy	0.7629	1.48×10^{-4}

Here, we address the importance of the measurement error, as explained in Cannon (2010). It originates from the fact that the CBD methodology, first estimates the factors, then it analyses their dynamic properties. This may affect the statistical results given in this section.

To propose an appropriate model, we need a more flexible class of distributions namely generalized hyperbolic distributions. We briefly review this class in the next section.

4. Generalized Hyperbolic Distributions

In this section, we define the family of generalized hyperbolic (GH) distributions based on Chapter 3 of McNeil *et al.* (2005).

The random vector $\boldsymbol{X}$ is said to have a d-dimensional GH distribution with parameters $(\lambda, \chi, \psi, \boldsymbol{\mu}, \Sigma, \boldsymbol{\gamma})$, denoted as $\boldsymbol{X} \sim GH(\lambda, \chi, \psi, \boldsymbol{\mu}, \Sigma, \boldsymbol{\gamma})$ if

$$\boldsymbol{X} \stackrel{d}{=} \boldsymbol{\mu} + W\boldsymbol{\gamma} + \sqrt{W} A \boldsymbol{Z}, \tag{7}$$

where $\stackrel{d}{=}$ denotes equality in distribution and

(i) $\boldsymbol{Z} \sim N_k(\mathbf{0}, I_k)$,

(ii) $A \in R^{d \times k}$,

(iii) $\boldsymbol{\mu}, \boldsymbol{\gamma} \in R^d$,

(iv) $W \geq 0$ is a scalar-valued random variable which is independent of $\boldsymbol{Z}$ and has a generalized inverse Gaussian distribution, denoted GIG(λ, χ, ψ). (See Appendix B for details.)

The joint density function of the GH distribution in non-singular case (Σ has rank d) is

$$\begin{aligned} f_{\boldsymbol{X}}(\boldsymbol{x}) &= \int_0^\infty f_{\boldsymbol{X}|W}(\boldsymbol{x}|w) f_W(w) dw \\ &= \int_0^\infty \frac{e^{(\boldsymbol{x}-\boldsymbol{\mu})'\Sigma^{-1}\gamma}}{(2\pi)^{\frac{d}{2}}|\Sigma|^{\frac{1}{2}}} \exp\left\{-\frac{Q(\boldsymbol{x})}{2w} - \frac{\gamma'\Sigma^{-1}\gamma}{2/w}\right\} f_W(w) dw \\ &= c \times \frac{K_{\lambda-\frac{d}{2}}\left(\sqrt{(\chi+Q(\boldsymbol{x}))(\psi+\gamma'\Sigma^{-1}\gamma)}\right) e^{(\boldsymbol{x}-\boldsymbol{\mu})'\Sigma^{-1}\gamma}}{\left(\sqrt{(\chi+Q(\boldsymbol{x}))(\psi+\gamma'\Sigma^{-1}\gamma)}\right)^{\frac{d}{2}-1}}, \quad -\infty < x < \infty, \end{aligned} \tag{8}$$

where the normalizing constant is

$$c = \frac{(\sqrt{\psi/\chi})^\lambda (\psi+\gamma'\Sigma^{-1}\gamma)^{\frac{d}{2}-\lambda}}{(2\pi)^{\frac{d}{2}}|\Sigma|^{\frac{1}{2}} K_\lambda(\sqrt{\chi\psi})},$$

and f_W is the density function of the GIG random variable W. Here $K_\lambda(\cdot)$ is the modified Bessel function of the third kind and $Q(\boldsymbol{x})$ denotes the Mahalanobis distance $(\boldsymbol{x}-\boldsymbol{\mu})'\Sigma^{-1}(\boldsymbol{x}-\boldsymbol{\mu})$. These parameters admit the following interpretations:

- λ, χ, ψ specify the shape of the distribution and how much weight is assigned to the tails compared with the center. The larger those parameters are the closer the distribution is to the normal distribution.
- $\boldsymbol{\mu}$ is the location parameter.
- $\Sigma = AA'$ is the dispersion-matrix. It controls inter-correlations between components of $\boldsymbol{X}$ and has to fulfill the usual conditions for covariance matrices, i.e., symmetry and positive definiteness as well as a full rank property.
- $\boldsymbol{\gamma}$ is the skewness parameter. If $\boldsymbol{\gamma} = \boldsymbol{0}$, then the distribution is symmetric.

The characteristic function of the GH distribution can be expressed as:

$$\phi_{\boldsymbol{X}}(\boldsymbol{u}) = E\left(e^{i\boldsymbol{u}'\boldsymbol{X}}\right) = e^{i\boldsymbol{u}'\mu}\hat{H}\left(\frac{1}{2}\boldsymbol{u}'\Sigma\boldsymbol{u} - i\boldsymbol{u}'\boldsymbol{\gamma}\right), \tag{9}$$

where $\hat{H}(\theta) = \int_0^\infty e^{-\theta\nu} dF(\nu)$ is the Laplace-Stieltjes transform of the distribution function F of the GIG random variable W. See McNeil *et al.* (2005) for more details.

The GH distribution family includes some special cases under different names, listed as follows:

- If $\lambda = (d+1)/2$, we have a multivariate hyperbolic (hyp) distribution.
- If $\lambda = -1/2$, a normal inverse Gaussian (NIG) distribution is obtained.
- If $\chi = 0$, $\lambda > 0$, we have a variance gamma (VG) distribution.
- If $\psi = 0$, $\lambda < 0$, one gets a generalized hyperbolic Student's t distribution. The shape parameter for this particular case is $\nu = -2\lambda$, which determines the degrees of freedom.

In the next section, we use the GH distributions to model the error terms in the CBD model.

5. Proposed Model

We first define the generalized hyperbolic Lévy process based on the GH distribution. Then we use this process to model the increments $\boldsymbol{A}(t+1) - \boldsymbol{A}(t)$. The generalized hyperbolic Lévy process is defined by

$$\boldsymbol{X}^{\mathrm{GH}} = \{\boldsymbol{X}(t),\ t > 0\}, \tag{10}$$

where $\boldsymbol{X}(0) = 0$, with stationary and independent increments and $\boldsymbol{X}(t)$ has characteristic function,

$$E\left[\exp(i\boldsymbol{u}\boldsymbol{X}(t))\right] = \left(\phi_{\boldsymbol{X}}(\boldsymbol{u})\right)^t, \tag{11}$$

where $\phi_{\boldsymbol{X}}(\boldsymbol{u})$ is the characteristic function of the GH distribution, defined in (9).

Similarly to Cairns *et al.* (2006), we adopt the following mortality curve:

$$\log\left(\frac{\tilde{q}(t,x)}{\tilde{p}(t,x)}\right) = A_1(t+1) + A_2(t+1)(x), \tag{12}$$

and assume that for $t > 0$

$$\boldsymbol{A}(t+1) - \boldsymbol{A}(t) = \boldsymbol{X}(1), \tag{13}$$

where $\boldsymbol{X}(1)$ is a bivariate GH Lévy process with the unit time scale. Here $\boldsymbol{A}(t) = (A_1(t), A_2(t))'$ are two stochastic factors. In other words, we change the normality assumption in the CBD model with a bivariate GH Lévy process, while we keep

the same structure as the CBD model for the evolution of survival probabilities. It is worth mentioning that the iid assumption of the increments for the GH Lévy process has already been tested for the selected data set in Section 3. Therefore, we can now fit the proposed model and compare it with the CBD model.

6. Model Comparisons

This section compares the CBD model with the proposed model. On the other hand, we emphasize the fact that the results given here regarding the CBD model and the normality assumption have been statistically rejected as shown in Section 3. For the sake of comparison, we use the log-likelihood function (LLF), Akaike information criterion (AIC), Bayesian information criterion (BIC) and the likelihood ratio test. The AIC is defined as

$$\text{AIC} = -2\ \text{LLF} + 2\text{NPS}, \tag{14}$$

where NPS is the effective number of parameters being estimated. The BIC is defined as

$$\text{BIC} = -2\ \text{LLF} + \text{NPS} \times \log(\text{NOS}), \tag{15}$$

where NOS is the number of observations. Hence, both the AIC and the BIC not only reward goodness of fit by considering the log-likelihood function, but also include a penalty that is an increasing function of the number of estimated parameters. This penalty discourages over-fitting. Higher values of the LLF and smaller values of the AIC and the BIC, show an improved goodness of fit for the considered mortality model.

The likelihood-ratio test can be used to check whether a special case of the GH distribution is the true underlying distribution. The LRT test statistic is defined as:

$$\Lambda = \frac{\sup\{L(\theta|\boldsymbol{Y}) : \theta \in \Theta_0\}}{\sup\{L(\theta|\boldsymbol{Y}) : \theta \in \Theta\}}, \tag{16}$$

where L denotes the likelihood function with respect to the parameter θ and data $\boldsymbol{Y}$, and Θ_0 is a subset of the parameter space Θ. The null hypothesis H_0 states that $\theta \in \Theta_0$ and the alternative hypothesis H_1 states that $\theta \in \Theta_0^c$, where Θ_0^c is the complement of Θ_0. Under the null hypothesis and certain regularity conditions, it can be shown that $-2\log\Lambda$ is asymptotically chi-squared distributed with ν degrees of freedom. Here ν is the number of free parameters specified in Θ minus the number of free parameters specified in Θ_0. The null hypothesis is rejected if $-2\log\Lambda$ exceeds the confidence level-quantile of the chi-squared distribution with ν degrees of freedom. In this study, H_0 is the bivariate normal distribution in the CBD model and H_1 is the special case of the GH distribution.

6.1 Empirical Analysis

For the purpose of in-sample model performance, we use the mortality data from 1969-1999. Then, to assess the out-of-sample model performance, we forecast the development of the mortality rates for the 9 subsequent years. We first, estimate $\boldsymbol{A}(t)$ in (13) by using the least square technique, then we fit eight GH distributions with the maximum likelihood method. The considered distributions are: Student's t, NIG, hyp and generalized hyperbolic distributions (ghyp) with density function defined in (8), both in symmetric and asymmetric cases. We use the package *ghyp* in order to fit above distributions for $\boldsymbol{X}(1)$ defined by (13).

Table 5 provides in-sample goodness of fit measures based on the LLF, the AIC and the BIC statistics, together with their corresponding ranks. A commonly used rule of thumb consists in considering that two models are significantly different if the difference in the AIC is larger than 10, as discussed in Burnham and Anderson (2002). Raftery (1995) suggests that a model significantly outperforms a competitor if the difference in their respective BIC values exceeds 5. Therefore, all three criteria show a preference for the GH distributions when comparing to the normal distribution with the lowest rank. The symmetric ghyp distribution is the best distribution based on the BIC and the AIC. According to the LLF, the asymmetric ghyp distribution offers the best fit for our mortality data set.

Table 5: In-sample goodness of fit measures

GH Distribution	Symm.	LLF	AIC	BIC	LLF Rank	AIC Rank	BIC Rank
ghyp	T	233.39	-454.79	-446.38	2	1	1
NIG	T	231.84	-451.68	-443.27	4	2	3
ghyp	F	233.66	-451.32	-440.12	1	3	4
t	T	230.53	-451.06	-444.06	6	4	2
hyp	T	229.96	-447.93	-439.52	8	5	5
NIG	F	231.91	-447.81	-436.60	3	6	7
t	F	230.73	-447.46	-437.65	5	7	6
hyp	F	230.03	-444.07	-432.86	7	8	8
Normal	T	221.29	-432.57	-425.56	9	9	9

We use the function *lik.ratio.test* to perform likelihood ratio tests. Table 6 provides the test statistics, degrees of freedom and the p-values. The likelihood ratio tests are statistically significant for the selected GH distributions. This table indicates that the considered mortality data set is more likely to come from the GH distribution than a bivariate normal distribution.

Table 6: Likelihood-ratio test

Model	Symm.	L-statistic	df	p-value
ghyp	T	5.52×10^{-06}	1	8.62×10^{-07}
NIG	T	2.61×10^{-06}	1	4.30×10^{-06}
ghyp	F	4.20×10^{-06}	2	4.20×10^{-06}
hyp	T	1.7×10^{-04}	1	3.1×10^{-05}
NIG	F	2.40×10^{-05}	2	2.40×10^{-05}
t	F	7.91×10^{-05}	1	1.38×10^{-05}
hyp	F	1.59×10^{-04}	2	1.59×10^{-04}

Overall, Tables 5 and 6 provide some evidence to support the use of the GH distributions for modeling $\boldsymbol{X}(1)$ in (13).

6.2 Mortality Projection

In this section, the out-of-sample performance of the proposed model is investigated. The reference cohort is the set of males aged 65 in 1999. We first, explain how to use (12) in order to project the mortality rates for nine years corresponding to $t = 2000, \cdots, 2008$. We generate nine iid copies of $\boldsymbol{X}(1)$ from the fitted symmetric ghyp distribution based on the mortality data over the period of 1969-1999. Then, we apply (13) together with the estimated value of $\boldsymbol{A}(1999)$ in order to obtain $\boldsymbol{A}(t)$ for $t = 2000, \cdots, 2008$. Next, we use (12) to project the mortality rates for the considered reference cohort, denoted by $\hat{q}(1999+i, 65+i)$, $i = 1, 2, \cdots, 9$. Finally, to evaluate the out-of-sample performance, we repeat the above procedure 20,000 times and record the projected mortality rates. Similarly to Wang *et al.* (2011), we find the mean absolute percentage error (MAPE) for each replication $j = 1, 2, \cdots, 20,000$, defined as follows:

$$\text{MAPE}_j = 100\% \times \frac{1}{9} \sum_{i=1}^{9} \left| \frac{\tilde{q}(1999+i, 65+i) - \hat{q}(1999+i, 65+i)}{\tilde{q}(1999+i, 65+i)} \right|,$$

where $\tilde{q}(t, x)$ is the realized mortality rate at time t for the cohort aged x. Table 7 illustrates the differences in mortality projection between the symmetric ghyp distribution and normal distribution, based on the mean, 90^{th} percentile, and the 95^{th} percentile of the MAPE. In this table, the model with a better predictive power will have lower mean and percentiles. We find that the symmetric ghyp distribution provides better mortality projection performance based on the mean, 90^{th} percentile, and the 95^{th} percentile of MAPE.

Table 7: Percentile of MAPE of mortality projection (Unit: %)

Model	Symm.	Mean	90%	95%
ghyp	T	8.76	14.48	16.55
Normal	T	9.74	16.13	18.32

7. Additional Data Sets

In this section, we consider three additional mortality data sets including males in Russia, Spain and the U.S. obtained from HMD as given in Table 8. We summarize our empirical results here. Similarly to the Italian data set explained in Section 3, we first obtain the increments $\boldsymbol{A}(t+1) - \boldsymbol{A}(t)$ using (3). Next, the data sets are tested against the linear independence (Ljung-Box test), the iid assumption (McLeod-Li test) and the serial autocorrelation using the ACF plot. The null hypothesis of linear independence cannot be rejected at a significance level of 0.05 based on Ljung-Box test. The bivariate normality assumption is then tested using Doornik-Hansen test and the results are reported in Table 9. The periods of our in-sample analysis for each data set are denoted in the parentheses. The bivariate normality assumption is rejected at 5% significance level for the U.S. and Russia. However, we cannot reject the normality assumption for the Spain.

Table 8: Mortality data used for Russia, Spain and the U.S.

Country	Age	Year
Russia	60-90	1960-2010
Spain	55-85	1970-2008
The U.S.	60-91	1985-2007

Table 9: Multivariate Doornik and Hansen (2008) normality test

Mortality data	Test statistics	df	p-value
Russia (1960-1993)	28.99	4	7.9×10^{-6}
Spain (1970-1999)	2.58	4	0.63
The U.S. (1985-2000)	10.99	4	0.026

We then fit the GH distributions and compare it with the bivariate normal distribution based on the AIC, the BIC criteria and LLF. We provide the results for the preferred GH distribution and the normal distribution together with their corresponding ranks in Table 10. According to this table, the GH distribution fits better than normal distribution even for Spain in which the normality assumption cannot be rejected.

Table 10: In-sample goodness of fit measures (Russia, Spain & the U.S.)

Data	GH Dist.	Symm.	LLF	AIC	BIC	LLF Rank	AIC Rank	BIC Rank
Russia	t	T	235.53	-459.07	-450.09	5	1	1
	Normal	T	226.26	-442.51	-435.03	9	9	9
Spain	ghyp	F	246.08	-474.16	-461.85	1	1	1
	Normal	T	219.32	-428.65	-421.81	9	6	5
U.S.	ghyp	T	175.44	-336.87	-331.92	2	1	1
	Normal	T	145.72	-281.44	-277.90	9	8	6

Also, the preferred GH distributions given in Table 10 are compared to the normal distribution using likelihood ratio test and results are reported in Table 11. All three tests are statistically significant at 5% confidence level. Therefore, according to the LRT, we can conclude that the considered data sets are more likely to come from the GH distribution than a bivariate normal distribution.

Table 11: Likelihood-ratio test, Russia, Spain and the U.S.

Mortality data	Model	Symm.	L-statistic	df	p-value
Russia	t	T	9.34×10^{-05}	1	1.65×10^{-05}
Spain	ghyp	F	2.40×10^{-12}	3	1.43×10^{-12}
U.S.	ghyp	T	1.24×10^{-13}	2	1.24×10^{-13}

Moreover, the out-of-sample performance of the preferred models in this section is tested against normal distribution using MAPE. Table 12 gives the reference cohort and the projected years for each data set. We perform 20,000 simulations and obtain the mean, 90^{th} percentile, and the 95^{th} percentile of the MAPE as explained in Section 6.2. The results are summarized in Table 13. Based on the results given in this table, we find that the symmetric t and asymmetric ghyp distribution can lead to a more reliable mortality projections in Russia and Spain, respectively. For the U.S., the symmetric ghyp distribution can slightly perform better than normal distribution.

Table 12: Reference cohort used for projections in Russia, Spain the U.S.

Mortality data	Cohort age	Reference year	Projected years
Russia	65	1993	1994-2010
Spain	65	1999	2000-2008
U.S.	65	2000	2001-2007

Table 13: Percentile of MAPE of mortality projection for Russia, Spain the U.S. (Unit: %)

Mortality data	Model	Symm.	Mean	90%	95%
Russia	t	T	17.06	30.19	36.48
	Normal	T	19.54	33.30	37.99
Spain	ghyp	F	13.58	21.45	24.12
	Normal	T	13.90	22.87	25.91
U.S.	ghyp	T	4.19	6.03	6.92
	Normal	T	4.58	6.35	6.91

8. Conclusions

In this paper, we show that the bivariate normality assumption in the CBD model is sometimes not appropriate for some mortality data. We test normality using the multivariate normality test of Doornik and Hansen (2008) and the multivariate Shapiro-Wilk normality test. We reject the bivariate normality assumption for Italy, Russia and the U.S.

Generalized hyperbolic distributions were proposed to model the increments of $\boldsymbol{A}(t)$. We estimate the parameters by maximum likelihood. Four GH distributions, in the symmetric and asymmetric cases, are compared with the CBD model based on the AIC, BIC, LLF and likelihood ratio test. We find that the symmetric ghyp distribution provides the best fit for Italy and the U.S. and also gives better mortality projections according to the MAPE. The symmetric t and asymmetric ghyp models were the best ones for the Russia and Spain, respectively.

To summarize the paper, we suggest testing the bivariate normality assumption in the CBD model before using it for projections. In addition, the GH distributions can be considered as an alternative to the normal distribution in the CBD model.

Appendix A: Doornik-Hansen's Test

In this section, we briefly explain the multivariate normality test proposed by Doornik and Hansen (2008). Denote a $p \times n$ matrix of n observations on a p-dimensional vector by $X' = (\boldsymbol{x_1}, \cdots, \boldsymbol{x_n})$ with sample mean $\bar{\boldsymbol{x}} = n^{-1} \sum_{i=1}^{n} \boldsymbol{x_i}$ and sample covariance matrix $S = n^{-1} \sum_{i=1}^{n} (\boldsymbol{x_i} - \bar{\boldsymbol{x}})(\boldsymbol{x_i} - \bar{\boldsymbol{x}})'$. Let $V = \text{diag}(\hat{\sigma_1}, \cdots, \hat{\sigma_p})$ be the diagonal matrix which has the variances on the diagonal and obtain the correlation matrix $C = V^{-1/2} S V^{-1/2}$. The eigenvalues of C can then be used to define a diagonal matrix $\Lambda = \text{diag}(\lambda_1, \cdots, \lambda_p)$. Next, each observation is transformed according to $\boldsymbol{y_i} = H \Lambda^{-1/2} H' V^{-1/2} (\boldsymbol{x_i} - \bar{\boldsymbol{x}})$ to obtain a $p \times n$ transformed matrix $Y' = (\boldsymbol{y_1}, \cdots, \boldsymbol{y_n})$. Here, the columns of H are the

corresponding eigenvectors of C, such that $H'H = I_p$ and $\Lambda = H'CH$. Denote the univariate skewness and kurtosis for each of the p-transformed vectors of n observations by $\boldsymbol{B}_1' = (\sqrt{b_{11}}, \cdots, \sqrt{b_{1p}})$ and $\boldsymbol{B}_2' = (b_{11}, \cdots, b_{1p})$, respectively. The multivariate Doornik-Hansen's test statistic is:

$$E_p = \boldsymbol{Z}_1'\boldsymbol{Z}_1 + \boldsymbol{Z}_2'\boldsymbol{Z}_2,$$

that has approximately a chi-squared distribution with $2p$ degrees of freedom, where $\boldsymbol{Z}_1' = (z_{11}, \cdots, z_{1p})$ and $\boldsymbol{Z}_2' = (z_{21}, \cdots, z_{2p})$ are determined by (A.1) and (A.2) as follow:

$$\begin{aligned}
\beta &= \frac{3(n^2+27n-70)(n+1)(n+3)}{(n-2)(n+5)(n+7)(n+9)},\\
w^2 &= -1 + \{2(\beta-1)\}^{\frac{1}{2}},\\
\delta &= \frac{1}{\{\log(\sqrt{w^2})\}^{\frac{1}{2}}},\\
y &= \sqrt{b_1}\left\{\frac{w^2-1}{2}\frac{(n+1)(n+3)}{6(n-2)}\right\}^{\frac{1}{2}},\\
z_1 &= \delta \log\{y + (y^2+1)^{\frac{1}{2}}\},
\end{aligned} \tag{A.1}$$

$$\begin{aligned}
\delta &= (n-3)(n+1)(n^2+15n-4),\\
a &= \frac{(n-2)(n+5)(n+7)(n^2+27n-70)}{6\delta},\\
c &= \frac{(n-7)(n+5)(n+7)(n^2+2n-5)}{6\delta},\\
k &= \frac{(n+7)(n+5)(n^3+37n^2+11n-313)}{12\delta},\\
\alpha &= a + b_1 c,\\
\chi &= (b2-1-b1)2k,\\
z_2 &= \left\{\left(\frac{\chi}{2\alpha}\right)^{\frac{1}{3}} - 1 + \frac{1}{9\alpha}\right\}(9\alpha)^{\frac{1}{2}}.
\end{aligned} \tag{A.2}$$

Appendix B: The Generalized Inverse Gaussian Distribution

The GIG(λ, χ, ψ), is defined by

$$f_{\text{GIG}}(\omega) = \left(\frac{\psi}{\chi}\right)^{\frac{\lambda}{2}} \frac{\omega^{\lambda-1}}{2K_\lambda\left(\sqrt{\chi\psi}\right)} \exp\left\{-\frac{1}{2}\left(\frac{\chi}{\omega} + \psi\omega\right)\right\}, \quad \omega > 0, \tag{B.1}$$

with parameters satisfying

$$(\chi > 0, \psi \geq 0, \lambda < 0), \text{ or}$$
$$(\chi > 0, \psi > 0, \lambda = 0), \text{ or else}$$
$$(\chi \geq 0, \psi > 0, \lambda > 0).$$

Special cases of the GIG distribution are the gamma distribution, when $\chi = 0$ and $\lambda > 0$, as well as the inverse gamma distribution, with $\psi = 0$ and $\lambda < 0$.

References

Ainou, V. (2011). A two factor Lévy model for stochastic mortality. Working Paper. http://iriaf.univ-poitiers.fr/colloque2011/article/v1s1a3.pdf

Burnham, K. P. and Anderson, D. R. (2002). *Model Selection and Multi-Model Inference*: *A Practical Information-Theoretic Approach.* Springer, New York.

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.

Cannon, E. (2010). Estimation and pricing with the Cairns-Blake-Dowd model of mortality. Discussion Paper PI-1003, Pensions Institute.

Doornik, J. A. and Hansen, H. (2008). An omnibus test for univariate and multivariate normality. *Oxford Bulletin of Economics and Statistics* **70**, 927-939.

Giacometti, R., Ortobelli, S. and Bertocchi, M. I. (2009). Impact of different distributional assumptions in forecasting Italian mortality rates. *Investment Management and Financial Innovations* **3**, 186-193.

Hainaut, D. and Devolder, P. (2008). Mortality modelling with Lévy processes. *Insurance*: *Mathematics and Economics* **42**, 409-418.

Jong, P. D. and Tickle, L. (2006). Extending Lee-Carter mortality forecasting. *Mathematical Population Studies* **13**, 1-18.

Lee, R. D. and Carter, L. R. (1992). Modeling and forecasting U. S. mortality. *Journal of the American Statistical Association* **87**, 659-671.

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

Ljung, G. M. and Box, G. E. P. (1978). On a measure of lack of fit in time series models. *Biometrika* **65**, 297-303.

McLeod, A. I. and Li, W. K. (1983). Diagnostic checking ARMA time series models using squared residual autocorrelations. *Journal of Time Series Analysis* **4**, 269-273.

McNeil, A. J., Frey, R. and Embrechts, P. (2005). *Quantitative Risk Management*: *Concepts, Techniques and Tools.* Princeton University Press, New Jersey.

Raftery, A. E. (1995). Bayesian model selection in social research. *Sociological Methodology* **25**, 111-163.

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

Royston, J. P. (1982a). Algorithm AS 181: the W test for normality. *Applied Statistics* **31**, 176-180.

Royston, J. P. (1982b). An extension of Shapiro and Wilk's W test for normality to large samples. *Applied Statistics* **31**, 115-124.

Royston, J. P. (1995). A remark on algorithm AS 181: the W-test for normality. *Applied Statistics* **44**, 547-551.

Wang, C. W., Huang, H. C. and Liu, I. C. (2011). A quantitative comparison of the Lee-Carter model under different types of non-Gaussian innovations. *Geneva Papers on Risk and Insurance* **36**, 675-696.

Received March 20, 2013; accepted June 14, 2013.

Seyed Saeed Ahmadi
Department of Mathematics and Statistics
Concordia University
Montreal, Quebec H3G 1M8, Canada
saeed.ahmadi@yahoo.com

Patrice Gaillardetz
Department of Mathematics and Statistics
Concordia University
Montreal, Quebec H3G 1M8, Canada
gaillard@mathstat.concordia.ca