國立政治大學風險管理與保險學系

博士論文

指導教授：黃泓智　博士

厚尾分配在財務與精算領域之應用

Applications of Heavy-Tailed Distributions in Finance and Actuarial Science

博士生：劉議謙　撰

中　華　民　國　一　〇　一　年　九　月

謝　誌

論文終於完成了！非常感謝我的指導教授黃泓智教授，高雄第一科技大學的王昭文教授，以及加拿大滑鐵盧大學的 Ken Seng Tan 教授，沒有這三位教授的大力指導，我相信我是無法完成這本論文的，議謙在此致上最高的感謝之意。另外，還要感謝論文口試委員：俞明德教授、張傳章教授、王儷玲教授以及戴天時教授，口試時給予議謙許多建議及指導，使本論文更加縝密與完整，議謙感恩在心。還有所有教過我的授課老師們，因為有您們的教學，啟發了我許多想法並增強了我許多觀念，謝謝您們。還要感謝中央大學的楊曉文教授，也就是我碩士班的指導老師，感謝楊老師持續願意與我討論許多議題並指導我撰寫投稿文章。

在博士班求學期間，幸運獲得 2010-2011 年度「中華扶輪獎學金」，感謝扶輪社的栽培，並謝謝嘉義北區社的照顧。此外，感謝國科會讓我有機會得到「補助博士生赴國外研究」，得以至加拿大滑鐵盧大學向 Ken Seng Tan 教授學習，議謙非常感謝。

於風險管理與保險學系就讀的四年間，感謝博士班的永琮學長、芳書學長、瑞益學長、雅文學姐、尚穎學長、麗如學姐、文彬學長、俞沛學長、士弘學長、志峰、昶翔、耿郎學長、玉鳳學姐、紓婷、可倫、恕銘、柏翰、芳文學姐、筱昀學姐等人，以及碩士班的芠綺、珮娟、慧婷、德全、振格等人，在生活上以及課業上相互扶持，豐富的我博士班苦悶的求學歷程，這是一段非常美好的回憶。

最後，感謝我的家人，爸爸、媽媽及姊姊總是全力支持我並給我正面的力量，每每讓沮喪的我馬上又恢復了活力。感謝一路上支持我的親戚朋友們，您們的鼓勵，議謙都會銘記在心，在未來的人生中，議謙會更加努力來回饋社會。

摘　要

本篇論文將厚尾分配（Heavy-Tailed Distribution）應用在財務及保險精算上。本研究主要有三個部分：第一部份是用厚尾分配來重新建構Lee-Carter模型(1992)，發現改良後的 Lee-Carter 模型其配適與預測效果都較準確。第二部分是將厚尾分配建構於具有世代因子（Cohort Factor）的 Renshaw and Haberman 模型（2006）中，其配適及預測效果皆有顯著改善，此外，針對英格蘭及威爾斯（England and Wales）訂價長壽交換（Longevity Swaps），結果顯示此模型可以支付較少的長壽交換之保費以及避免低估損失準備金。第三部分是財務上的應用，利用 Schmidt 等人（2006）提出的多元仿射廣義雙曲線分配（Multivariate Affine Generalized Hyperbolic Distributions; MAGH）於 Boyle 等人(2003)提出的低偏差網狀法(Low Discrepancy Mesh; LDM）來定價多維度的百慕達選擇權。理論上，LDM 法的數值會高於 Longstaff and Schwartz(2001)提出的最小平方法(Least Square Method; LSM）的數值，而數值分析結果皆一致顯示此性質，藉由此特性，我們可知道多維度之百慕達選擇權的真值落於此範圍之間。

關鍵字：隨機死亡率模型；厚尾分配；長壽交換；百慕達選擇權；多元 Lévy 分配；低偏差網狀法。

Abstract

The thesis focus on the application of heavy-tailed distributions in finance and actuarial science. We provide three applications in this thesis. The first application is that we refine the Lee-Carter model (1992) with heavy-tailed distributions. The results show that the Lee-Carter model with heavy-tailed distributions provide better fitting and prediction. The second application is that we also model the error term of Renshaw and Haberman model (2006) using heavy-tailed distributions and provide an iterative fitting algorithm to generate maximum likelihood estimates under the Cox regression model. Using the RH model with non-Gaussian innovations can pay lower premiums of longevity swaps and avoid the underestimation of loss reserves for England and Wales. The third application is that we use multivariate affine generalized hyperbolic (MAGH) distributions introduced by Schmidt et al. (2006) and low discrepancy mesh (LDM) method introduced by Boyle et al. (2003), to show how to price multidimensional Bermudan derivatives. In addition, the LDM estimates are higher than the corresponding estimates from the Least Square Method (LSM) of Longstaff and Schwartz (2001). This is consistent with the property that the LDM estimate is high bias while the LSM estimate is low bias. This property also ensures that the true option value will lie between these two bounds.

Keywords: Stochastic Mortality Models; Heavy-Tailed Distributions; Longevity Swaps; Bermudan Options; Multivariate Lévy Distributions; Low Discrepancy Mesh.

Contests

List of Figures

List of Tables

Chapter 1

Introduction

The heavy-tailed distributions have been used in financial field for many years. The scholars of insurance field use these distributions to discuss some problems step by step. First we want to refine existing two mortality models by some heavy-tailed distributions, and we then price high-dimensional Bermudan options using the low discrepancy mesh (LDM) with multivariate affine generalized hyperbolic (MAGH).

Longevity represents an increasingly important risk for defined benefit pension plans and annuity providers, because life expectancy is dramatically increasing in developed countries. In 2007, exposures to improved life expectancy amounted to $400 billion for pension funds and insurance companies in the United Kingdom and United States (see Loeys et al., 2007). Stochastic mortality models quantify mortality and longevity risks, which makes mortality risk management possible and provides the foundation for pricing and reserving. Among all stochastic mortality models, the Lee-Carter model (LC), proposed in 1992, is one of the most popular choices because of its ease of implementation and acceptable prediction errors in empirical studies.

Various modifications of the LC model have been extended by Brouhns et al. (2002), Renshaw and Haberman (2003, 2006), Cairns et al. (2006), Li and Chan (2007), Biffis et al. (2010), and Hainaut (2012) to attain a broader interpretation. Cairns et al. (2006) propose a two-factor stochastic mortality model, the CBD model, in which a first factor affects mortality at all ages, whereas a second factor affects mortality at older ages much more than at younger ages. Modeling the number of deaths with the Poisson model, Cairns et al. (2009) classify and compare eight stochastic mortality models, including an extension of the CBD model, with mortality data from England and Wales and the United States. They find that an extension of the CBD model that incorporates the cohort effect fits the English and Welsh data best, whereas for the U.S. data, the Renshaw and Haberman (2006) model (RH), which also allows for a cohort effect, provides the best fit (Cairns et al., 2009). In addition to the cohort effect, short-term catastrophic mortality events, such as the influenza pandemic in 1918 and the Tsunami in December 2004, may lead to much higher mortality rates. Using empirical data from 1900 to 1983, we find that the residuals in the RH model for England and Wales, France, and Italy exhibit leptokurticity. It is crucial to address such mortality jumps in age–period–cohort mortality models.

To the best of our knowledge, Hainaut and Devolder (2008) were the first to apply α-stable subordinators (infinite-activity, strictly positive, Lévy processes) to model mortality hazard rates. However, in the Lee-Carter model, the first difference of mortality indices may be negative, to reflect mortality improvements. Giacometti et al. (2009) consider both the error distributions of the Lee-Carter model and its mortality index, using the NIG distribution to model mortality for different age groups. They observe that the NIG distributional assumption for the residuals of the Lee-Carter model is better than the Gaussian one for some age groups. To take

non-Gaussian distributions into account in stochastic mortality models, Milidonis et al. (2011) use a Markov regime-switching model to analyze the U.S. mortality data and price mortality securities. In addition, some research use diffusion processes with jump components, one of the finite-activity Lévy processes, to describe the dynamics of mortality rates. Biffis (2005) uses affine jump diffusions and models asset prices and mortality dynamics, in the context of risk analysis and market valuation in life insurance contracts. Luciano and Vigna (2005) find, with Italian mortality data, that introducing a jump component provides a better fit than does a diffusion component for stochastic mortality processes. Cox, Lin, and Wang (2006) combine geometric Brownian motion with a compound Poisson process to model the age-adjusted mortality rates for the United States and United Kingdom, using an evaluation of the first pure mortality security, the Swiss Re Vita bond. In addition, Lin and Cox (2008) combine Brownian motion with a discrete Markov chain and log-normal jump size distribution to price mortality-based securities in an incomplete market framework. Chen and Cox (2009) incorporate a jump process into the Lee-Carter model and use it to forecast mortality rates and analyze mortality securitization. These studies all use diffusion processes with jump components (JD) and finite-activity Lévy processes to describe the dynamics of morality rates.

However, non-normal innovations can be generated by heavy-tailed distributions. An alternative set of distributions thus involves infinite-activity, or pure jump, Lévy processes, such as the normal inverse Gaussian (NIG) distributions that appear repeatedly in financial applications as unconditional return distributions (Bølviken and Benth, 2000; Eberlein and Keller, 1995; Lillestøl, 2000; Prause, 1997; Rydberg, 1997) or the variance gamma (VG) distributions of Madan and Seneta (1987, 1990). Another method relies on Student's *t*-distribution (T) and its skew extensions,

such as the generalized hyperbolic skew Student's *t*-distribution (GHST), as described by Prause (1999), Barndorff-Nielsen and Shepard (2001), Jones and Faddy (2003), Mencia and Sentana (2004), Demarta and McNeil (2004), and Aas and Haff (2006). Therefore, this study aims to examine whether mortality indices can be described by jump models, such as JD, NIG, and VG, or by the Student's *t* family, including the T and GHST distributions.

Therefore, in line with their work, the Chapter 3 incorporates the normal, *t*, JD, NIG, VG, and GHST distributions into the original Lee-Carter model, in an attempt to fit and forecast mortality rates. We rely on mortality data from six countries—Finland, France, the Netherlands, Sweden, Switzerland, and the United States—from 1900 to 2007. We fit the model to mortality rates from 1900 to 1999 using the normal, *t*, JD, VG, NIG, and GHST distributions, then forecast the development of the mortality curve for the subsequent eight years. According to the Jarque-Bera (JB) test statistics, we must largely reject the assumptions of normality for the residuals of the Lee-Carter model and the mortality indices. The results of the Kolmogorov-Smirnov (KS), Anderson-Darling, and Cramér-von-Mises tests provide powerful evidence to support the rationality of using heavy-tailed distributions for the residuals of the Lee-Carter model and the first difference of mortality indices. Finally, according to the mean absolute percentage errors (MAPE) in the mortality projection, our empirical results indicate that the GHST distribution is the most appropriate choice for modeling long-term mortality indices for most countries.

As proposed by Pitacco (2004), various disadvantages arise in connection with the LC model. To improve the LC model, it is possible to model the number of deaths as a Poisson model, as commonly employed in literature on mortality modeling (e.g., Wilmoth, 1993; Brouhns et al., 2002; Renshaw and Haberman, 2006; Cairns et al.,

2009; Haberman and Renshaw, 2009). However, with the Poisson error structure, the intensity at age *x* and time *t* is determined by the death rate at age x and time t, which is broadly described by stochastic mortality models. Consequently, instead of using a Poisson model with a deterministic intensity function, an alternative means of fitting the number of deaths is to specify a doubly stochastic Poisson process, or Cox process (Cox, 1955), to capture the stochastic intensity. Biffis et al. (2010) first implement a doubly stochastic setup in the LC model, introducing a class of equivalent probability measures for pricing life insurance liabilities and mortality-indexed securities. Following the double stochastic setup proposed by Biffis et al. (2010), the second goal of this article is to provide an iterative fitting algorithm for estimating the Cox regression model in which mortality rates adhere to the RH model with non-Gaussian innovations.

In Chapter 4, we use three mortality data sets—England and Wales, France, and Italy—from 1900 to 2008 as the observed data. We first fit the model to the mortality rates from 1900 to 1983 using the normal, JD, variance gamma (VG), and NIG distributions, and then we forecast the development of the mortality curve for the subsequent 25 years. According to the Jarque-Bera statistical test, the assumption of normality must be rejected for the logarithm of mortality rates. Finally, according to the mean absolute percentage errors (MAPEs) of the mortality projections, our empirical results indicate that the RH model with non-Gaussian innovations is the most appropriate choice for modeling long-term mortality data. In addition, as an application for England and Wales, we provide the fair values of longevity swaps and their value at risk (VaR) and conditional tail expectations (CTE). According to the RH model with non-Gaussian innovations, the swap premiums are lower, but the VaR and CTE are higher, which means that using the RH model with non-Gaussian

innovations can reduce the costs of the longevity risk hedger and avoid the underestimation of loss reserves.

In finance field, the past three decades has sparked the development of a large body of theory concerning multivariate probability distributions. Most studies about multivariate asset models are based on Brownian motions due to their simple structure. However, since the work of Mandelbrot and Taylor (1967) and Clark (1973), it has been widely recognized the presence of significant skewness and excess kurtosis in empirical asset return distributions; that is, the returns are non-normally distributed. To allow for both kurtosis and skewness for the multivariate probability distribution of assets returns, multivariate Lévy processes are also used as a tractable model for asset returns.

There are essentially two multidimensional models for financial asset pricing: one is multivariate normal mixtures based on a common mixing distribution and the other is multivariate time-changed Brownian motions based on a common time change. Barndorff-Nielsen (2001), Cont and Tankov (2004), Luciano and Schoutens (2006) and Eberlein and Madan (2009) provide a multivariate time changed Brownian motion by a common subordinator. As noted in Luciano and Semeraro (2010), however, the common subordinator exerts a strict restriction on the joint process, which in turn leads to the lack of independence. Semeraro (2008) and Luciano and Semeraro (2007) propose a similar model with idiosyncratic and systematic subordinators to capture idiosyncratic and systematic jump shocks simultaneously. In this line, Luciano and Semeraro (2010) consider correlated Brownian motions with idiosyncratic and systematic subordinators to increase the range of dependence using correlated Brownian motions. However, as noted in Luciano and Semeraro (2010), this approach is still less flexible in terms of high correlation.

Since the introduction of the generalized hyperbolic distributions (GH) by Barndorff-Nielsen (1977, 1978), it has widely been used in many applications because it provides a flexible tool for modeling the empirical distribution of financial data exhibiting skewness, leptokurtosis and fat-tails. The GH distribution encompasses many other distributions as special case, which includes the well-known normal inverse Gaussian (NIG) distributions of Barndorff-Nielsen (1995) and the Variance Gamma (VG) distributions of Madan and Seneta (1987, 1990). Multivariate GH (MGH) distributions were introduced and investigated by Barndorff-Nielsen (1978) and Blasild and Jensen (1981) according to a variance–mean mixture of a multivariate normal distribution. Prause (1999) first uses the MGH distributions to fit financial market. However, as Schmidt et al. (2006) show, it is computational burdensome to estimate the MGH distribution parameters since all parameters must be estimated simultaneously. Also, the MGH distributions do not allow independent margins, and they are not able to model tail-dependence.

Schmidt et al. (2006) introduced multivariate affine generalized hyperbolic (MAGH) distributions. Because MAGH distributions are defined as an affine transformation of independent GH margins, these distributions possess four desirable features: easier for estimation and simulation algorithms, the existence of characteristic functions in closed form, better goodness-of-fit that MGH distributions, and the ability to capture a wide range of dependence structure. More recently, Fajardo and Farias (2010) price multidimensional European derivatives by obtaining the density resulting from the convolution of MAGH distributions. Distinct from Schmidt et al. (2006) and Fajardo and Farias (2010) who use the univariate GH distributions with zero location and unit scaling to construct MAGH distributions, we use standard GH margins to construct MAGH distributions.

The demand on more sophistical multivariate distributions on modeling asset prices poses implies that in many cases, closed form solutions does not exist on the price of the derivative securities. It is numerically even more challenging to price Bermudan/American options. While there exists many numerical methods for pricing Bermudan/American options when there is only one underlying asset price, many of these methods are breakdown when the option depends on more than one asset. An example is the multinomial tree of Këllezi and Webber (2004) and Maller et al. (2006) which is known to be very efficient for pricing Bermudan/American options when there is one underlying asset in Lévy process models. However, if the option depends on more than one asset, the multinomial tree becomes computational infeasible. For derivative pricing under multivariate Brownian motion, a possible method is the stochastic mesh (Monte Carlo mesh; for short MCM) method of Broadie and Glasserman (2004) and low discrepancy mesh (LDM) method of Boyle, et al. (2003). In Boyle et al. (2003), they have shown that the LDM method can be a competitive method to price multivariate Bermudan/American options. Their studies, however, have confined to assuming that the returns of the assets are multivariate normally distributed.

In Chapter 5, we are concerned with an efficient algorithm of pricing multivariate Bermudan options assuming that the underlying asset returns follow MAGH distributions. To the best of our knowledge, there is no other numerical works that have used these distributions to price Bermudan derivatives. Here we demonstrate that LDM method can be extended to price Bermudan options even when the underlying asset returns follow MAGH distributions. In addition, in consistent with the results of Boyle et al. (2003), the LDM estimates are higher bias while the estimates from the Least Square Method (LSM) of Longstaff and Schwartz (2001) are

low bias. This property also ensures that the true value will lie between these two bounds.

The remainder of this thesis is organized as follows. In Chapter 2, we introduce heavy-tailed distributions. We then illustrate the Lee-Carter model with *t*, JD, VG, NIG, and GHST innovations, as well as provide the dynamics of the mortality indices in Chapter 3. In Chapter 4, we provide a an iterative fitting algorithm to generate the maximum likelihood estimates of the Cox regression model under which the residuals of the RH model, the mortality indices and the cohort effects adhere to heavy-tailed distributions. And we employ the RH model with non-Gaussian innovations to price a longevity swap and calculate its VaR and CTE using England and Wales mortality data. The final application is in Chapter 5. We present the MAGH distributions as well as the estimation algorithm, and we employ the MAGH processes for asset returns, providing the multivariate Esscher transform for the MAGH asset model. However, we demonstrates the convergence of our proposed method by using some high dimensional Bermudan options when the underlying assets follow a MAGH distribution. In the end of thesis, we have some conclusions.

國立政治大學
National Chengchi University

Chapter 2

Heavy-Tailed Distributions

2.1 Introductions of Heavy-Tailed Distributions

If a random variable X adheres to a JD distribution, then

$$X = a + \sigma Z + \sum_{i=1}^{N} Y_i , \tag{2-1}$$

where *N* follows the Poisson distribution with intensity λ_N; Z is a standard normal random variable; and each Y_i, independent of *z* and *N*, is a normal distribution with mean μ_Y and variance δ_Y^2. Therefore, the probability density function takes the form:

$$f_{JD}\left(x \middle| a, \sigma, \lambda_N, \mu_Y, \delta_Y\right) = \sum_{i=0}^{\infty} \Phi\left(x; a + N\mu_Y, \sigma^2 + N\delta_Y^2 \middle| N = i\right) \text{Prob}\left(N = i\right),$$

$$= \sum_{i=0}^{\infty} \frac{\lambda_N^i e^{-\lambda_N}}{i!} \Phi\left(x; a + i\mu_Y, \sigma^2 + i\delta_Y^2\right), \tag{2-2}$$

where $\Phi(x; \tilde{\mu}, \tilde{\sigma}^2)$ is a normal probability density function with mean $\tilde{\mu}$ and

variance $\tilde{\sigma}^2$. The characteristic function of the JD distribution is of the form:

$$\phi_{JD}\left(\omega\middle|a,\sigma,\lambda_N,\mu_Y,\delta_Y\right)=\exp\left[ia\omega-0.5\omega^2\sigma^2+\lambda_N\left(e^{i\mu_Y\omega-0.5\omega^2\delta_Y^2}-1\right)\right]. \tag{2-3}$$

However, the first two moments of the JD distribution are

$$E\left(X\right)=a+\lambda_N\mu_Y, \tag{2-4}$$

$$Var\left(X\right)=\sigma^2+\lambda_N\left(\mu_Y^2+\delta_Y^2\right). \tag{2-5}$$

The moment-generating function is $\phi_{JD}\left(-i\omega\middle|a,\sigma,\lambda_N,\mu_Y,\delta_Y\right)$.

The *t*, VG, NIG, and GHST distributions are the special cases of the generalized hyperbolic (GH) model proposed by Barndorff-Nielsen (1977, 1978) and offer flexible tools for modeling the empirical distribution of financial data that exhibit skewness, leptokurtosis, and fat tails.[1] The generalized hyperbolic probability density function, following Prause (1999), is

$$f_{GH}\left(x\middle|\alpha,\beta,\lambda,\delta,\mu\right)=\frac{\left(\frac{\sqrt{\alpha^2-\beta^2}}{\delta}\right)^{\lambda}}{\sqrt{2\pi}\left(K_\lambda\left(\delta\sqrt{\alpha^2-\beta^2}\right)\right)}e^{\beta(x-\mu)}\frac{K_{\lambda-\frac{1}{2}}\left(\alpha\sqrt{\delta^2+\left(x-\mu\right)^2}\right)}{\left(\frac{\sqrt{\delta^2+\left(x-\mu\right)^2}}{\alpha}\right)^{\frac{1}{2}-\lambda}}, \tag{2-6}$$

where K_λ is the modified Bessel function of the second kind with index λ; δ is the scale parameter; μ is the shift parameter; and λ, α and β determine the shape of the GH distribution. The parameters must fulfill the following constraints:

$$\begin{aligned}&\delta\geq 0,\ \alpha>\left|\beta\right| \text{ if } \lambda>0.\\ &\delta>0,\ \alpha>\left|\beta\right| \text{ if } \lambda=0.\\ &\delta>0,\ \alpha\geq\left|\beta\right| \text{ if } \lambda<0.\end{aligned} \tag{2-7}$$

[1] In the empirical analyses, we also use the GH distribution to fit the mortality data of our six countries. However, the calibration results of the GH distribution always reduce to those of the GHST distribution, so we focus on this special case instead of the broader GH distribution.

The characteristic function of the GH distribution is of the form:

$$\phi_{GH}\left(\omega|\alpha,\beta,\lambda,\delta,\mu\right)=e^{i\mu\omega}\left(\frac{\alpha^2-\beta^2}{\alpha^2-\left(\beta+i\omega\right)^2}\right)^{\frac{\lambda}{2}}\frac{K_\lambda\left(\delta\sqrt{\alpha^2-\left(\beta+i\omega\right)^2}\right)}{K_\lambda\left(\delta\sqrt{\alpha^2-\beta^2}\right)}. \tag{2-8}$$

However, the first two moments of the GH distribution are

$$E\left(X\right)=\mu+\frac{\beta\delta}{\sqrt{\alpha^2-\beta^2}}\frac{K_{\lambda+1}\left(\delta\sqrt{\alpha^2-\beta^2}\right)}{K_\lambda\left(\delta\sqrt{\alpha^2-\beta^2}\right)}, \tag{2-9}$$

$$Var\left(X\right)=\delta^2\left[\frac{K_{\lambda+1}\left(\delta\sqrt{\alpha^2-\beta^2}\right)}{\delta\sqrt{\alpha^2-\beta^2}K_\lambda\left(\delta\sqrt{\alpha^2-\beta^2}\right)}+\frac{\beta^2}{\alpha^2-\beta^2}\left(\frac{K_{\lambda+2}\left(\delta\sqrt{\alpha^2-\beta^2}\right)}{K_\lambda\left(\delta\sqrt{\alpha^2-\beta^2}\right)}\right)-\left(\frac{K_{\lambda+1}\left(\delta\sqrt{\alpha^2-\beta^2}\right)}{K_\lambda\left(\delta\sqrt{\alpha^2-\beta^2}\right)}\right)^2\right]. \tag{2-10}$$

The moment-generating function is $\phi_{GH}\left(-i\omega|\alpha,\beta,\lambda,\mu,\delta\right)$.

If we let $\delta=0,\ \lambda>0$ in Equation (2-6), using $K_\lambda(x)\sim\Gamma(\lambda)2^{\lambda-1}x^{-\lambda}$ as $\lambda>0$ and $x\to 0$, we can obtain the VG distribution, with the following density function:

$$f_{VG}\left(x|\alpha,\beta,\lambda,\mu\right)=\frac{\left(\alpha^2-\beta^2\right)^\lambda\left|x-\mu\right|^{\lambda-0.5}K_{\lambda-0.5}\left(\alpha\left|x-\mu\right|\right)}{\sqrt{\pi}\left(2\alpha\right)^{\lambda-0.5}\Gamma\left(\lambda\right)}\exp\left(\beta\left(x-\mu\right)\right). \tag{2-11}$$

Note that when $\alpha=\left(G+M\right)/2$, $\beta=\left(G-M\right)/2$, and $\lambda=C$, we obtain the VG distribution, which is a special case of the CGMY distribution defined by Carr et al (2002). The characteristic function of the VG distribution is of the form:

$$\phi_{VG}\left(\omega|\alpha,\beta,\lambda,\mu\right)=e^{i\mu\omega}\left(\frac{\alpha^2-\beta^2}{\alpha^2-\left(\beta+i\omega\right)^2}\right)^\lambda. \tag{2-12}$$

However, the first two moments of the VG distribution are

$$E\left(X\right)=\mu+\frac{2\beta\lambda}{\alpha^2-\beta^2}, \tag{2-13}$$

$$Var(X)=\frac{2\lambda}{\alpha^2-\beta^2}\left(1+\frac{2\beta^2}{\alpha^2-\beta^2}\right). \tag{2-14}$$

The moment-generating function is $\phi_{VG}\left(-i\omega\middle|\alpha,\beta,\lambda,\mu\right)$. The another popular representation of the VG distribution is provided in Appendix A.

When $\lambda=-0.5$, realizing that $K_{0.5}(x)=\sqrt{\pi/2}x^{-0.5}e^{-x}$ and $K_{-\lambda}(x)=K_{\lambda}(x)$, we obtain the NIG distribution with the following density function:

$$f_{NIG}\left(x\middle|\alpha,\beta,\delta,\mu\right)=\frac{\alpha\delta}{\pi}\exp\left(\delta\sqrt{\alpha^2-\beta^2}+\beta(x-\mu)\right)\frac{K_1\left(\alpha\sqrt{\delta^2+(x-\mu)^2}\right)}{\sqrt{\delta^2+(x-\mu)^2}}. \tag{2-15}$$

The characteristic function of the NIG distribution is of the form:

$$\phi_{NIG}\left(\omega\middle|\alpha,\beta,\delta,\mu\right)=\exp\left(i\mu\omega+\delta\left(\sqrt{\alpha^2-\beta^2}-\sqrt{\alpha^2-(\beta+i\omega)^2}\right)\right). \tag{2-16}$$

However, the first two moments of the NIG distribution are

$$E(X)=\mu+\frac{\beta\delta}{\sqrt{\alpha^2-\beta^2}}, \tag{2-17}$$

$$Var(X)=\frac{\delta\alpha^2}{\left(\alpha^2-\beta^2\right)^{1.5}}. \tag{2-18}$$

The moment-generating function is $\phi_{NIG}\left(-i\omega\middle|\alpha,\beta,\delta,\mu\right)$. This distribution is one of the most promising versions of the GH distribution for asset returns, because it possesses several attractive theoretical properties and analytical tractability. It therefore appears frequently in financial applications as an unconditional return distribution (Bølviken and Benth, 2000; Eberlein and Keller, 1995; Lillestøl, 2000; Prause, 1997; Rydberg, 1997) and for stochastic mortality modeling (Giacometti et al., 2009). The another representation of the NIG distribution is provided in Appendix B.

If instead we let $\lambda=-\nu/2$ and $\alpha\to|\beta|$ in Equation (2-6), and we realize that

$K_\lambda(x) \sim \Gamma(-\lambda)2^{-\lambda-1}x^\lambda$ as $\lambda < 0$ and $x \to 0$. In addition, $K_\lambda(x) = K_{-\lambda}(x)$. We can obtain the density of the GH skew Student's *t*-distribution (GHST), proposed by Aas and Haff (2006), as follows:

$$f_{GHST}\left(x|\beta,\nu,\delta,\mu\right) = \frac{\delta^\nu |\beta|^{\frac{\nu+1}{2}} \left(\sqrt{\delta^2+(x-\mu)^2}\right)^{-\frac{\nu+1}{2}} e^{\beta(x-\mu)}}{\sqrt{\pi}\,2^{\frac{\nu-1}{2}}\Gamma(\nu/2)} \times K_{\frac{\nu+1}{2}}\left(|\beta|\sqrt{\delta^2+(x-\mu)^2}\right),\ \beta \neq 0. \tag{2-19}$$

The characteristic function of the GHST distribution is of the form:

$$\phi_{GHST}\left(\omega|\beta,\nu,\delta,\mu\right) = e^{i\mu\omega}\frac{\left(\delta\sqrt{\omega^2-2\beta i\omega}\right)^{\frac{\nu}{2}} K_{\frac{\nu}{2}}\left(\delta\sqrt{\omega^2-2\beta i\omega}\right)}{2^{\frac{\nu}{2}-1}\Gamma(\nu/2)},\ \beta \neq 0. \tag{2-20}$$

However, the first two moments of the GHST distribution are

$$E(X) = \mu + \frac{\beta\delta^2}{\nu-2}, \tag{2-21}$$

$$Var(X) = \frac{2\beta^2\delta^4}{(\nu-2)^2(\nu-4)} + \frac{\delta^2}{\nu-2}. \tag{2-22}$$

The moment-generating function of the GHST distribution is undefined.

The GHST distribution is one of the skew extensions of Student's *t*-distribution. Letting $\delta = \sqrt{\nu}$ and $\beta = 0$ in Equation (2-19), we obtain the non-central Student's *t*-distribution with ν degrees of freedom, as follows:

$$f_{t\text{-distribution}}\left(x|\nu,\mu\right) = \frac{\Gamma(\frac{\nu+1}{2})}{\sqrt{\pi\nu}\Gamma(\frac{\nu}{2})}\left[1+\frac{(x-\mu)^2}{\nu}\right]^{\frac{-(\nu+1)}{2}}. \tag{2-23}$$

The characteristic function of the non-central Student's *t*-distribution is of the form:

$$\phi_{t\text{-distribution}}\left(\omega\middle|v,\mu\right)=e^{i\mu\omega}\frac{\left(\sqrt{v\omega^2}\right)^{\frac{v}{2}}K_{\frac{v}{2}}\left(\sqrt{v\omega^2}\right)}{2^{\frac{v}{2}-1}\Gamma(v/2)}, \tag{2-24}$$

However, the first two moments of the non-central Student's *t*-distribution are

$$E\left(X\right)=\mu,\ v>1, \tag{2-25}$$

$$Var\left(X\right)=\frac{v}{v-2},\ v>2. \tag{2-26}$$

In order to standardization, we first introduce the three parameters of Student's *t*-distribution (Bishop, 2006). We illustrate how to implement the Student's *t*-distribution in next subsection. The probability density function of three parameters of Student's *t*-distribution is

$$f_{t\text{-distribution}}\left(x\middle|v,\lambda,\mu\right)=\frac{\Gamma(\frac{v+1}{2})}{\sqrt{\frac{\pi v}{\lambda}}\Gamma(\frac{v}{2})}\left[1+\frac{\lambda\left(x-\mu\right)^2}{v}\right]^{\frac{-(v+1)}{2}}. \tag{2-27}$$

where λ is called the precision of the Student's *t*-distribution. The characteristic function of the three parameters of Student's *t*-distribution is of the form:

$$\phi_{t\text{-distribution}}\left(\omega\middle|v,\lambda,\mu\right)=e^{i\mu\omega}\frac{\left(\sqrt{\frac{v\omega^2}{\lambda}}\right)^{\frac{v}{2}}K_{\frac{v}{2}}\left(\sqrt{\frac{v\omega^2}{\lambda}}\right)}{2^{\frac{v}{2}-1}\Gamma(v/2)}, \tag{2-28}$$

However, the first two moments of the three parameters of Student's *t*-distribution are

$$E\left(X\right)=\mu,\ v>1, \tag{2-29}$$

$$Var\left(X\right)=\frac{v}{\lambda\left(v-2\right)},\ v>2. \tag{2-30}$$

The moment-generating function of the Student's *t*-distribution is undefined.

2.2 The Standardization Approaches for Heavy-Tailed Distributions

In the subsection, we introduce the standardizations of the JD, GH, VG, NIG, GHST and Student's *t*-distribution . We let

$$X = a + b \cdot \varepsilon , \tag{2-31}$$

where ε are standardized random variables (i.e. $E(\varepsilon) = 0$ and $Var(\varepsilon) = 1$, imply $E(X) = a$ and $Var(X) = b^2$). ε can be one of the standardized Lévy Processes. We can choose a finite-activity jump process, the jump-diffusion model of Merton (1976), and an infinite-activity jump process, the GH model.

From Equation (2-4) and (2-5), if ε follows a standardized JD distribution, then

$$\varepsilon = -\lambda_N \mu_Y + \sqrt{1 - \lambda_N \left(\mu_Y^2 + \delta_Y^2 \right)} \cdot Z + \sum_{i=1}^{N} Y_i , \tag{2-32}$$

where N is the Poisson distribution with intensity λ_N; Z is a standard normal random variable; each Y_i, independent with Z and N, is a normal distribution with mean μ_Y and variance δ_Y^2. The ε setting satisfies $E(\varepsilon) = 0$ and $Var(\varepsilon) = 1$. The probability density function of ε is of the form

$$f_{JD}\left(y \middle| \lambda_N,\ \mu_Y,\ \delta_Y \right)$$

$$= \sum_{i=0}^{\infty} \frac{e^{-\lambda_N} \lambda_N^i}{i!} \Phi\left(y \middle| -\lambda_N \mu_Y + i\mu_Y,\ \left(1 - \lambda_N \left(\mu_Y^2 + \delta_Y^2 \right)\right) + i\delta_Y^2 \right), \tag{2-33}$$

where $\Phi\left(y \middle| \tilde{\mu},\ \tilde{\sigma}^2 \right)$ is the probability density function with mean $\tilde{\mu}$ and variance $\tilde{\sigma}^2$ for y.

Barndorff-Nielsen (1977, 1978) proposes the generalized hyperbolic (GH) model and the *t*, VG, NIG and GHST distributions are the special cases of the GH model. From Equation (2-9) and (2-10), if ε follows a standardized GH distribution, then

$$f_{GH}\left(y\middle|\alpha,\beta,\lambda\right)=\frac{\left(\frac{\sqrt{\alpha^2-\beta^2}}{\delta}\right)^{\lambda}}{\sqrt{2\pi}\left(K_{\lambda}\left(\delta\sqrt{\alpha^2-\beta^2}\right)\right)}e^{\beta(y-\mu)}\frac{K_{\lambda-\frac{1}{2}}\left(\alpha\sqrt{\delta^2+\left(y-\mu\right)^2}\right)}{\left(\frac{\sqrt{\delta^2+\left(y-\mu\right)^2}}{\alpha}\right)^{\frac{1}{2}-\lambda}},\quad (2\text{-}34)$$

where μ and δ must satisfy the first two moments:

$$E(\varepsilon)=\mu+\frac{\beta\delta}{\sqrt{\alpha^2-\beta^2}}\frac{K_{\lambda+1}\left(\xi\right)}{K_{\lambda}\left(\xi\right)}=0,\quad (2\text{-}35)$$

$$V(\varepsilon)=\frac{\delta^2}{\xi}\frac{K_{\lambda+1}\left(\xi\right)}{K_{\lambda}\left(\xi\right)}+\left(\frac{\beta\delta^2}{\xi}\right)^2\left(\frac{K_{\lambda+2}\left(\xi\right)}{K_{\lambda}\left(\xi\right)}-\left(\frac{K_{\lambda+1}\left(\xi\right)}{K_{\lambda}\left(\xi\right)}\right)^2\right)=1,\quad (2\text{-}36)$$

where K_{λ} is the modified Bessel function of the second kind with index λ; $\xi=\delta\sqrt{\alpha^2-\beta^2}$; μ is the shift parameter and δ is the scale parameter. In accordance with λ, α and β we can describe the shape of the GH distribution. When $\beta=0$, the shape is symmetric. In addition, these parameters obey the following constraints: $\alpha>\left|\beta\right|\geq 0$ and $\lambda\in R$.

From Equation (2-13) and (2-14), we can obtain ε follows a standardized VG distribution:

$$f_{VG}\left(y\middle|\alpha,\beta\right)=\frac{\left(\alpha^2-\beta^2\right)^{\lambda}\left|y-\mu\right|^{\lambda-0.5}K_{\lambda-0.5}\left(\alpha\left|y-\mu\right|\right)}{\sqrt{\pi}\left(2\alpha\right)^{\lambda-0.5}\Gamma\left(\lambda\right)}e^{\beta(y-\mu)},\quad (2\text{-}37)$$

where $\mu=-\frac{2\beta\lambda}{\alpha^2-\beta^2}$ and $\lambda=\frac{\left(\alpha^2-\beta^2\right)^2}{2\left(\alpha^2+\beta^2\right)}$ such that $E\left(\varepsilon_t\right)=0$ and $Var\left(\varepsilon_t\right)=1$.

From Equation (2-17) and (2-18), we also obtain that ε_t follows a standardized NIG distribution:

$$f_{NIG}\left(y|\alpha,\beta\right)=\frac{\alpha\delta}{\pi}\exp\left(\delta\sqrt{\alpha^2-\beta^2}+\beta\left(y-\mu\right)\right)\frac{K_1\left(\alpha\sqrt{\delta^2+\left(y-\mu\right)^2}\right)}{\sqrt{\delta^2+\left(y-\mu\right)^2}}, \quad (2\text{-}38)$$

where $\mu=-\frac{\beta\delta}{\sqrt{\alpha^2-\beta^2}}$ and $\delta=\frac{\left(\alpha^2-\beta^2\right)^{1.5}}{\alpha^2}$ such that $E\left(\varepsilon\right)=0$ and $Var\left(\varepsilon\right)=1$.

From Equation (2-21) and (2-22), we have standardized GHST distribution, as follows:

$$\begin{aligned} f_{GHST}\left(y|\beta,v\right)=&\frac{\delta^v|\beta|^{\frac{v+1}{2}}\left(\sqrt{\delta^2+\left(y-\mu\right)^2}\right)^{-\frac{v+1}{2}}\mathrm{e}^{\beta(y-\mu)}}{\sqrt{\pi}2^{\frac{v-1}{2}}\Gamma(v/2)}\\ &\times K_{\frac{v+1}{2}}\left(|\beta|\sqrt{\delta^2+\left(y-\mu\right)^2}\right),\ \beta\neq 0. \end{aligned} \quad (2\text{-}39)$$

where μ and δ must satisfy the first two moments:

$$E\left(\varepsilon\right)=\mu+\frac{\beta\delta^2}{v-2}=0, \quad (2\text{-}40)$$

$$Var\left(\varepsilon\right)=\frac{2\beta^2\delta^4}{\left(v-2\right)^2\left(v-4\right)}+\frac{\delta^2}{v-2}=1. \quad (2\text{-}41)$$

For standardized Student's *t*-distribution, we employ the three parameters of Student's *t*-distribution to transfer. From Equation (2-29) and (2-30), we obtain that ε_t follows a standardized Student's *t*-distribution:

$$f_{t\text{-distribution}}\left(y|v\right)=\frac{\Gamma(\frac{v+1}{2})}{\sqrt{\frac{\pi v}{\lambda}}\Gamma(\frac{v}{2})}\left[1+\frac{\lambda\left(y-\mu\right)^2}{v}\right]^{\frac{-(v+1)}{2}}, \quad (2\text{-}42)$$

where $\mu=0,\ v>1$ and $\lambda=\frac{v-2}{v},\ v>2$ such that $E\left(\varepsilon\right)=0$ and $Var\left(\varepsilon\right)=1$.

2.3 Estimation Scheme with Standardization

To estimate parameters of these models, we use the maximum likelihood method. Let a time series $\{X_t\}_{t=1}^{n}$, and they are independent and identically distributed (i.i.d.), as follows:

$$X_t \overset{i.i.d.}{\sim} f_X\left(a, b^2 \middle| \Theta\right), \tag{2-43}$$

where X is a random variable with parameters Θ; a represents the mean of X and b represents the standard derivation of X. Or equivalently, we have

$$X_t = a + b \cdot \varepsilon_t, \tag{2-44}$$

where ε_t are standardized random variables (i.e. $E(\varepsilon_t) = 0$ and $Var(\varepsilon_t) = 1$, imply $E(R_t) = a$ and $Var(R_t) = b^2$) and independent and identically distributed (i.i.d.). The log-likelihood function with respect to θ is

$$LLF = \sum_{t=1}^{n} \ln\left(f_X\left(\theta \middle| X_t\right)\right) \tag{2-45}$$

where $\theta \in \mathbb{R}^d$; d is the number of the unknown parameters. From Equation (2-44), we have

$$\varepsilon_t = \frac{X_t - a}{b}. \tag{2-46}$$

Using transformation of random variable, if the probability density function of random variable U is given by $f_U(x)$ and h is a monotonic function and we know the probability density function of $V = h(U)$ is

$$f_V(v) = f_U\left(h^{-1}(v)\right) \cdot \left| \frac{d\left(h^{-1}(v)\right)}{dv} \right| \tag{2-47}$$

where h^{-1} denotes the inverse function. So equivalently, we let $h^{-1}(X_t) = \frac{X_t - a}{b}$ and rewrite Equation (2-47) as follows:

$$f_X(x) = f_\varepsilon\left(\frac{X_t - a}{b}\right) \cdot \frac{1}{b} \tag{2-48}$$

Considering Equation (2-48), substituting into Equation (2-45) we obtain

$$LLF = \sum_{t=1}^{n}\left[\ln\left(f_\varepsilon\left(\theta\middle|\frac{X_t - a}{b}\right)\right) - \ln(b)\right] \tag{2-49}$$

Originally we estimate parameters with respect to random variable R_t, using the method of standardization, we change to random variable ε_t when estimating. It is an advantage for reducing the number of estimated parameters. For example, assume the logarithm returns follow GH distribution, we need estimate five parameters α, β, λ, δ, μ if we do not use standardization approach. But we just estimate three parameters α, β, λ when we employ standardization approach.

國立政治大學
National Chengchi University

Chapter 3

A Quantitative Comparison of the Lee-Carter Model under Different Types of Non-Gaussian Innovations

3.1 The Lee-Carter Model with Heavy-Tailed Innovations

In this Chapter, we first review the classical Lee-Carter model, under which the mortality index follows an ARIMA model with normal innovations. Using the mortality data of six countries, we find that all the residuals of the Lee-Carter model and the mortality indices exhibit non-zero skewness and excess kurtosis. Therefore, we use the Lee-Carter model with five non-Gaussian distributions—*t*, JD, VG, NIG, and GHST—to model both the residuals and the dynamics of the mortality indices.

The Lee-Carter Model

We analyze the changes in mortality as a function of both age *x* and time *t*. The

mortality forecast relies on the classical Lee-Carter model, namely,

$$ln\left(m_{x,t}\right)=\alpha_x+\beta_x k_t+e_{x,t}, \tag{3-1}$$

where $m_{x,t}$ is the central death rate for age *x* in calendar year *t*, defined as the span from time t to time $t+1$. This structure is designed to capture age–period effects; α_x describes the average pattern of mortality for the age group; β_x represents the age-specific patterns of mortality change, indicating the sensitivity of the logarithm of the force of mortality at age *x* to variations in the time index k_t; k_t explains the time trend of the general mortality level; and $e_{x,t}$ represents the deviation of the model from the observed log-central death rates, which should be a normal distribution with zero mean and a relatively small variance (Lee, 2000).

We use approximation to fit the three parameters. According to two constraint conditions, $\sum_t k_t=0$ and $\sum_x \beta_x=1$, $\hat{\alpha}_x$ is simply the average value over time of $ln\left(m_{x,t}\right)$, and $\hat{k}_t$ is the sum over various ages of $ln\left(m_{x,t}\right)-\hat{\alpha}_x$. Using $ln\left(m_{x,t}\right)-\hat{\alpha}_x$ as the dependent variable and $\hat{k}_t$ as the explanatory variable, we can obtain $\hat{\beta}_x$ by using a simple regression model without an intercept parameter. Finally, we re-estimate the $\hat{k}_t$ by iteration, using actual number of deaths, population, $\hat{\alpha}_x$, and $\hat{\beta}_x$, such that the actual number of deaths is close to the estimated number of deaths, and the adjusted $\hat{k}_t$ is denoted as $\hat{k}_t^*$.

To forecast future mortality dynamics, Lee and Carter (1992) assume that α_x and β_x remain constant over time and therefore forecast the dynamics of adjusting the mortality index k_t^* using an ARIMA(0,1,0) model, as follows:

$$k_t^*-k_{t-1}^*=\gamma+\varepsilon_t, \tag{3-2}$$

where γ is a drift term, and ε_t is a sequence of independent and identically Gaussian random variables with mean 0 and variance σ^2.

Normality Tests for the Residuals and Mortality Indices

In this subsection, we apply the JB (Jarque and Bera, 1980) test to determine empirically the normality of the mortality data of six countries (Finland, France, Netherlands, Sweden, Switzerland, U.S.) from 1900 to 2007. The mortality data for the first five countries come from the Human Mortality Database (HMD) website,[2] whereas the U.S. data come from the National Center Health Statistics (NCHS) website.[3]

First, we examine the normality test for the residuals in Equation (3-1). Figure 3-1 depicts the probability density function of the standardized residuals. Clearly, the empirical residuals peak around the mean and fatter tails; that is, the residuals are non-normally distributed. Second, Figure 3-2 reveals the patterns of mortality indices, offering evidence of mortality improvements. We also find a lot of jump points. Chen and Cox (2009) attribute jump points in the U.S. mortality rate to influenza epidemics and argue against the naïve belief that a pandemic is a one-time event that cannot happen again. We thus cannot just ignore such extreme events. In addition, as we show in Figure 3-3, the probability density functions of the first differences in the mortality indices exhibit higher central peaks and larger tails than does a normal distribution. Therefore, we can fit the mortality indices to the non-Gaussian distributions.

[2] http://www.mortality.org/.
[3] http://www.cdc.gov/nchs/nvss/mortality_tables.htm. Death rate files: HIST290 and GMWK290R. Death files: HIST290A and GMWK23F.

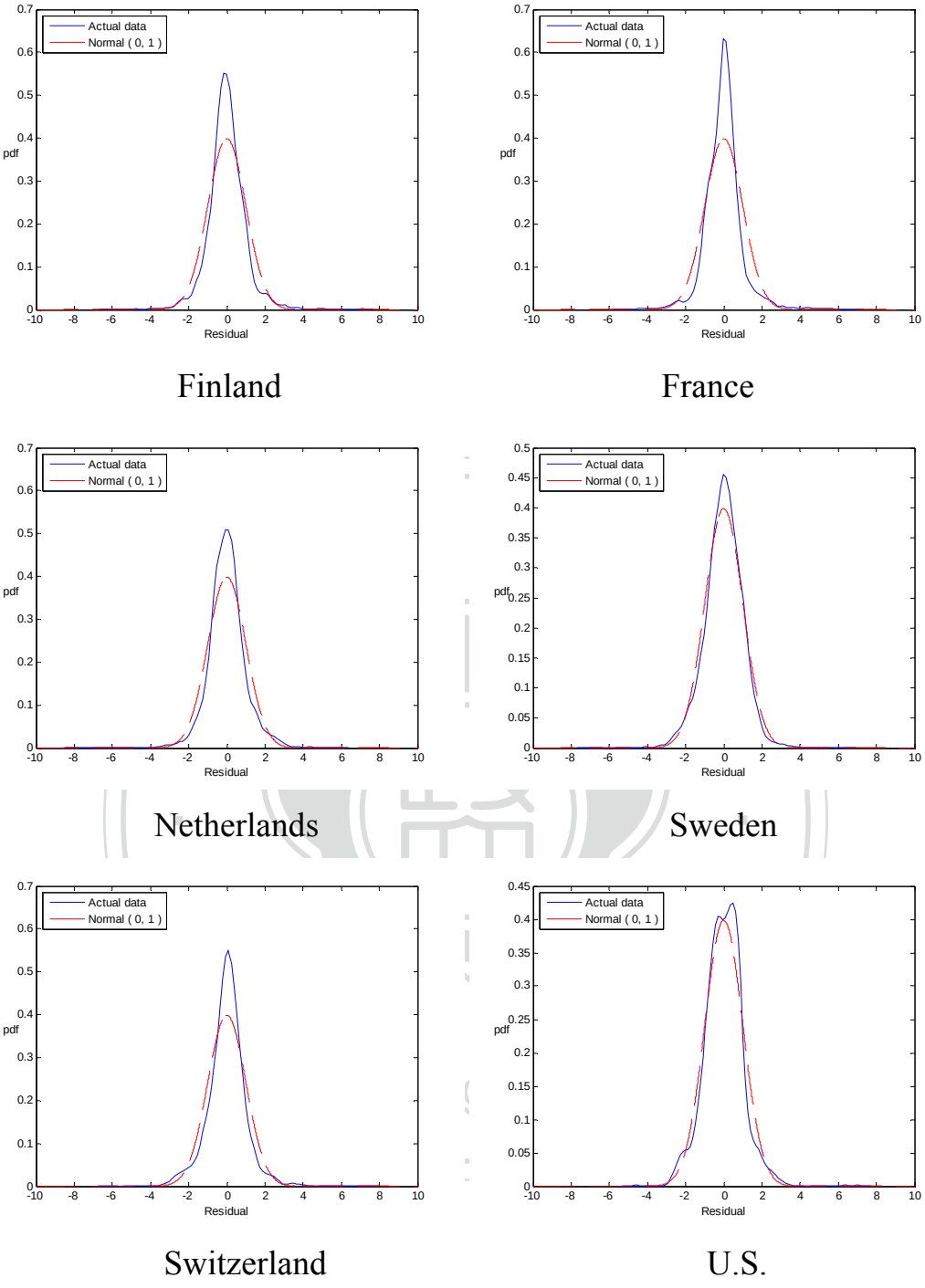


Figure 3-1. The Probability Density Functions of Standardized Residuals

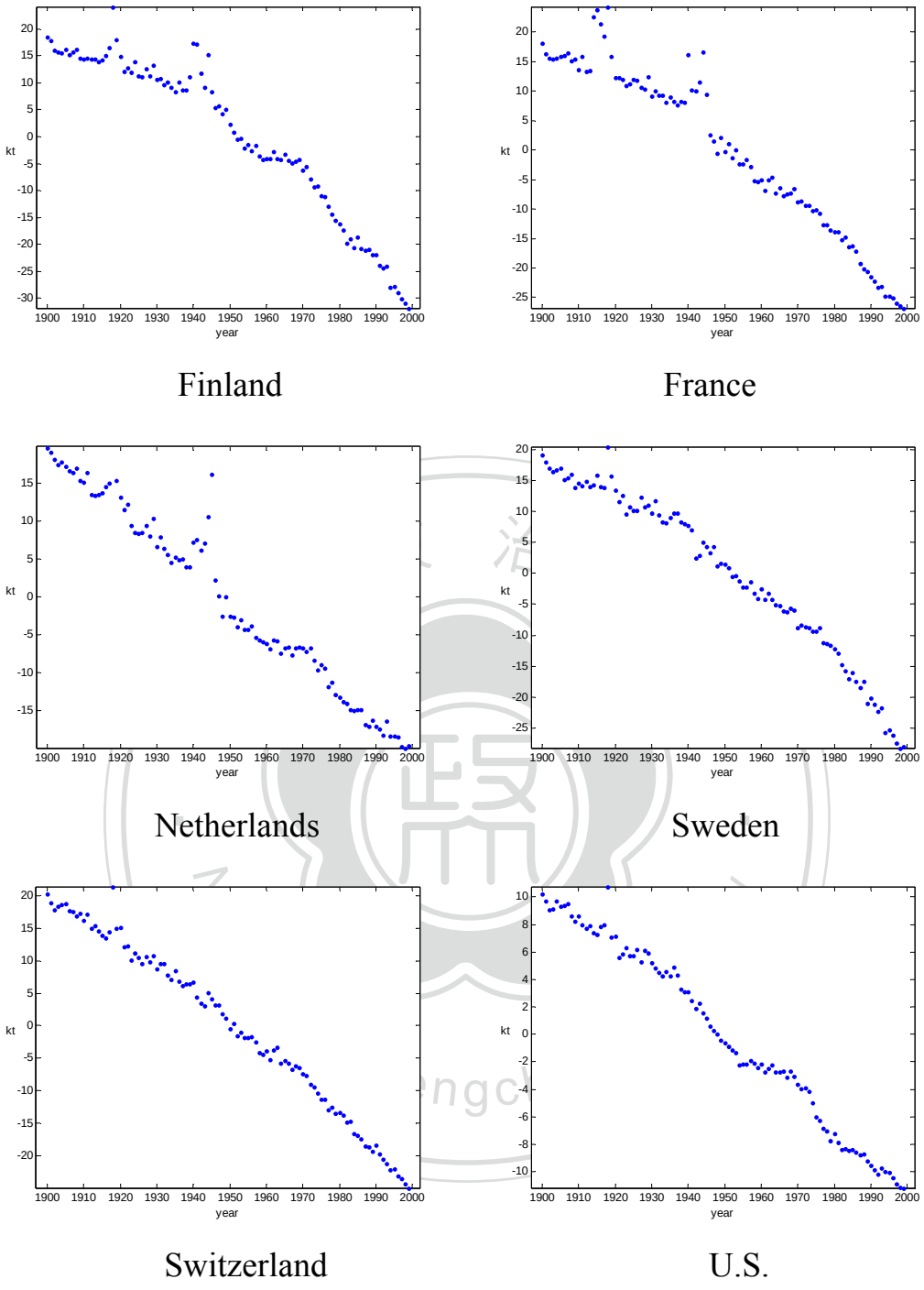
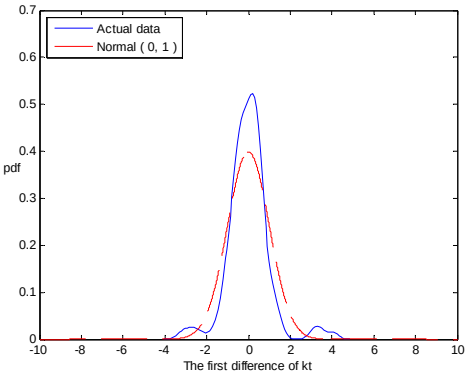
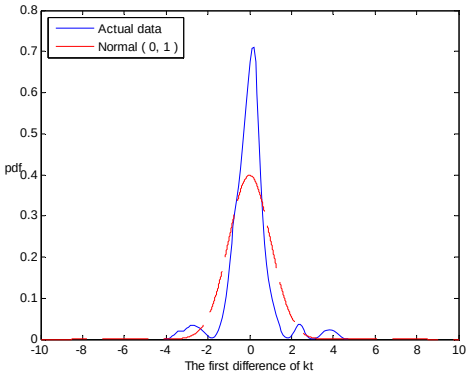


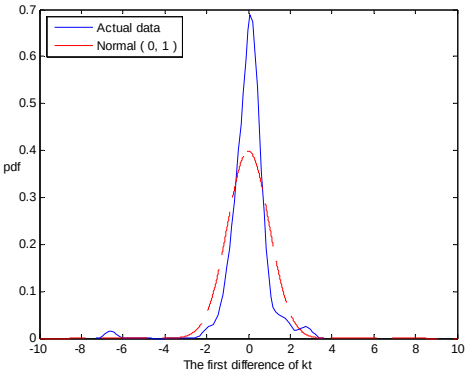
Figure 3-2. The Pattern of Mortality Indices



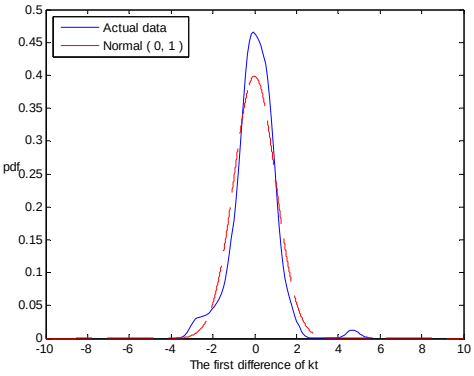
Finland



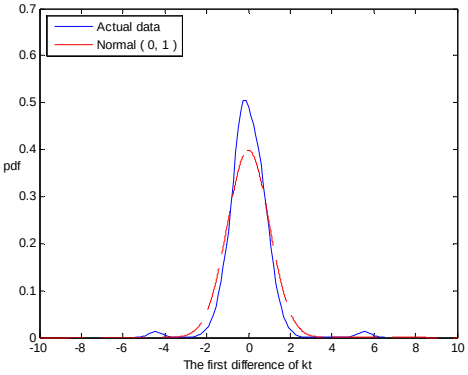
France



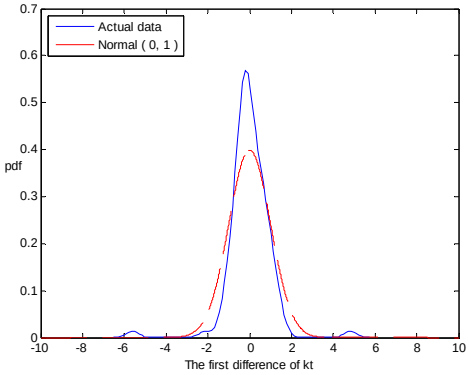
Netherlands



Sweden



Switzerland



U.S.

Figure 3-3. The Probability Density Functions of the First Difference in Mortality Indices

To examine the assumption of normality of mortality rates in the Lee-Carter model, we also use the JB test statistic, a goodness-of-fit measure of the departure from normality:

$$JB = n\left[\frac{s^2}{6} + \frac{(k-3)^2}{24}\right], \tag{3-3}$$

where n is the sample size, s is sample skewness, and k is sample kurtosis. Table 3-1 provides the results of the JB test, together with the skewness and excess kurtosis values, for the residual of the Lee-Carter model and the first difference of the six countries' mortality indices from 1900 to 1999. The skewness is significantly different from zero, and the excess kurtosis is large. The JB statistics also are significantly large, which means we must reject the assumption of normality. In turn, we use the heavy-tailed distributions—*t*, JD, VG, NIG, and GHST—to model the non-Gaussian property of the error terms in Equations (3-1) and (3-2).

Table 3-1. Skewness, Excess Kurtosis and the Jarque-Bera Test

Panel A: the Residuals of the Lee-Carter Model

Country	Finland	France	Netherlands	Sweden	Switzerland	U.S.
Skewness	0.394	0.950	-0.298	-0.100	0.273	0.444
	(0.053)	(0.053)	(0.053)	(0.053)	(0.053)	(0.074)
Excess Kurtosis	6.813	7.168	6.058	3.071	4.925	4.018
	(0.107)	(0.107)	(0.107)	(0.107)	(0.107)	(0.148)
JB Test	4116	4811	3242	829	2148	776
	[< 0.001]	[< 0.001]	[< 0.001]	[< 0.001]	[< 0.001]	[< 0.001]

The table presents the skewness and excess kurtosis of the standardized residuals of the Lee-Carter model. Standard errors of the skewness and excess kurtosis given in the parentheses are calculated as $\sqrt{6/n}$ and $\sqrt{24/n}$, respectively. n denotes the number of observations. The p-values of Jarque-Bera (JB) test are given in bracket.

Panel B: the First Difference in Mortality Indices

Country	Finland	France	Netherlands	Sweden	Switzerland	U.S.
Skewness	0.665	0.441	-2.525	0.470	0.827	-0.661
	(0.246)	(0.246)	(0.246)	(0.246)	(0.246)	(0.246)
Excess Kurtosis	4.439	5.339	18.396	4.601	11.446	13.476
	(0.492)	(0.492)	(0.492)	(0.492)	(0.492)	(0.492)
JB Test	89	121	1501	91	552	756
	[< 0.001]	[< 0.001]	[< 0.001]	[< 0.001]	[< 0.001]	[< 0.001]

The table presents the skewness and excess kurtosis of the first difference in mortality indices. Standard errors of the skewness and excess kurtosis given in the parentheses are calculated as $\sqrt{6/n}$ and $\sqrt{24/n}$, respectively. n denotes the number of observations. The p-values of Jarque-Bera (JB) test are given in bracket.

The Lee-Carter Model with Non-Gaussian Distributions

Because the residuals of the Lee-Carter model and the mortality indices are non-normally distributed, we model the error term, $e_{x,t}$ and ε_t, using the five heavy-tailed distributions: *t*, JD, VG, NIG, and GHST. These distributions are refered to Chapter 2.

For mortality data at age $x = 1,...,g$ and time period $t = 1,...,T$, the calibrated parameters of the Lee-Carter model can be obtained by maximizing the sample log-likelihood function (LLF),

$$LLF = \sum_{x=1}^{g}\sum_{t=1}^{T} ln\left(f\left(e_{xt}\middle|\Theta\right)\right), \tag{3-4}$$

with respect to Θ, which satisfies two constraint conditions, $\sum_t k_t = 0$ and $\sum_x \beta_x = 1$.[4] As suggested by Lee and Carter (1992), we re-estimate the k_t factors by iteration, given the values of α_x and β_x we obtained in the maximum likelihood

[4] Let $a = -\lambda\mu$ in the JD model and $\theta = -\beta\delta K_{\lambda+1}\left(\delta\sqrt{\alpha^2-\beta^2}\right)\Big/\left(\sqrt{\alpha^2-\beta^2}K_\lambda\left(\delta\sqrt{\alpha^2-\beta^2}\right)\right)$ in the special cases of the GH model. Thus, we ensure that the mean of error terms equals 0.

estimation, such that the implied number of deaths equals the actual number of deaths, or

$$D_t = \sum_x N_{x,t} \exp\left(\alpha_x + \beta_x k_t\right), \quad t = 1, ..., T, \tag{3-5}$$

where D_t is the total number of deaths in year t, and $N_{x,t}$ is the total population of age group x at time t. Using the re-estimated mortality indices, we can calculate the parameters of Equation (3-2) by maximizing the log-likelihood function, as follows:

$$\sum_{t=1}^{T} ln\left(f\left(\varepsilon_t\right)\right). \tag{3-6}$$

3.2 Empirical Analysis

In this section, we illustrate the mortality data and investigate the goodness-of-fit distributions for the residuals of the Lee-Carter model and the first difference of mortality indices. Using the mortality data from 1900 to 1999, we first fit the residuals of the Lee-Carter model with our six distributions: normal, *t*, JD, VG, NIG, and GHST. We then fit the first difference of k_t from the best goodness-of-fit model, according to the Bayesian information criterion (BIC), to the same six distributions and project the subsequent eight-year mortality rates.

Model Comparison

For the sake of comparison, we use the log-likelihood function (LLF), Akaike information criterion (AIC; Akaike, 1974), BIC (Schwarz, 1978), KS test (Kolmogorov, 1933), Anderson-Darling (AD) test (Stephens, 1974), and Cramér-von-Mises (CvM) test (Anderson, 1962) as goodness-of-fit measures. The AIC is defined as

$$AIC = -LLF + NPS, \tag{3-7}$$

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

$$BIC = -LLF + 0.5 \times NPS \times \log(NOS), \tag{3-8}$$

where NOS is the number of observations. For these criteria, a higher value of LLF and a smaller value of AIC and BIC indicate a better goodness of fit for the mortality model.

For the KS test, the null hypothesis is $H_0 : G(x) = F(x;\Theta)$ for all sample data x and the parameters Θ of the distribution, where $G(x)$ represents the empirical distribution function of the sample mortality index, and $F(x;\Theta)$ is the hypothesized cumulative density distribution (CDF). The test statistic is defined as

$$KS = \sup_{\{x\}} \left| F(x;\Theta) - G(x) \right|. \tag{3-9}$$

Thus a higher *p*-value in the KS test means a better goodness of fit for the mortality model.

The AD test is a modification of the KS test, which also determines whether a sample of data come from a specific distribution. However, unlike the KS test, the AD test focuses on the weight of the tail. Its null hypothesis is that the data follow a specific distribution. The AD test statistic is defined as

$$AD^2 = -NOS - S, \tag{3-10}$$

where

$$S = \sum_{i=1}^{NOS} \frac{(2i-1)}{NOS} \left[\ln F(y_i;\Theta) + \ln\left(1 - F(y_{NOS+1-i};\Theta)\right) \right]; \tag{3-11}$$

F is a cumulative distribution function of the specified distribution; and y_i are the observed values in increasing order. A lower value of the test statistic indicates a

higher possibility that the mortality data come from the distribution F.

The CvM test, an alternative to the KS test, is a criterion used to judge the goodness of fit of a probability distribution, compared with a given empirical distribution function. The test statistic CvM is defined as

$$CvM = \frac{1}{12NOS} + \sum_{i=1}^{NOS}\left[\frac{2i-1}{2NOS} - F(y_i;\Theta)\right]^2. \tag{3-12}$$

A lower value of this test statistic indicates a higher possibility that the mortality data come from the distribution F. Each test offers some benefits. That is, the KS test is known for the independence of its critical values from the tested distribution. Compared with the KS test, the main advantage of the AD test is that it assigns more weight to the tails of the distribution. Similarly, the CvM test incorporates information about the total sample and is insensitive to a slight dislocation of the empirical CDF. However, a major disadvantage of the CvM and AD tests is that the critical values depend on the analyzed distribution.[5]

In-Sample Goodness of Fit

Using mortality data from Finland, France, the Netherlands, Sweden, Switzerland, and the United States, Table 3-2 provides the LLF, AIC, and BIC results, together with their corresponding ranks. All three criteria indicate that the normal distribution is the worst model for all our mortality data. However, the JD model is the best model for the French mortality data; the NIG model is the best for the mortality data of Finland, the Netherlands, and Switzerland; and the VG model is the best option for Sweden. For the U.S. mortality data, the JD model offers the best fit according to the LLF and AIC values, but the t model is the best according to the BIC.

In Table 3-3 we report the results for the KS, AD, and CvM tests, together with

[5] We obtain the critical values through a Monte Carlo simulation with the estimated parameters (see Chernobai et al., 2007, p. 219).

their critical values for all six countries. For each country, the three test statistics are greater than the 5% critical value, so the empirical distribution of the residuals does not follow a normal distribution. For the Netherlands and Sweden, no test results reject the null hypothesis that the residuals come from non-Gaussian distributions. In addition, except for the U.S. mortality data, the results of the three tests support the null hypothesis that the residuals come from the best BIC models. For the U.S. mortality data though, the KS statistic rejects the *t* model, which is the best model according to the BIC, at a 1% significance level. Thus the difference between the theoretical and empirical CDF appears significant. However, if we ignore the dislocation of the empirical CDF, the *t* model offers better goodness of fit for the U.S. residuals, from the standpoint of the AD and CvM tests. Because the AD and CvM test results do not reject the claim that the error terms in Equation (3-1) come from the best models, according to the BIC, we use the mortality indices obtained from the best BIC model to investigate the pattern of innovations in Equation (3-2).

Table 3-2. Goodness-of-fit Measures for the Residuals of the Lee-Carter Model

Panel A: the Finland Mortality Data

Model	LLF	AIC	BIC	LLF Rank	AIC Rank	BIC Rank
Normal	789.78	-646.78	-242.82	6	6	6
t	1088.73	-944.72	-537.95	4	4	3
JD	1094.23	-948.23	-535.80	3	3	4
VG	1073.33	-928.33	-518.73	5	5	5
NIG	1105.54	-960.54	-550.94	1	1	1
GHST	1097.20	-952.20	-542.59	2	2	2

Panel B: the France Mortality Data

Model	LLF	AIC	BIC	LLF Rank	AIC Rank	BIC Rank
Normal	826.85	-683.85	-279.90	6	6	6
t	1279.79	-1135.79	-729.02	4	4	4
JD	1344.31	-1198.31	-785.88	1	1	1
VG	1277.39	-1132.39	-722.78	5	5	5
NIG	1336.77	-1191.77	-782.17	2	2	2
GHST	1296.69	-1151.69	-742.08	3	3	3

Panel C. the Netherlands Mortality Data

Model	LLF	AIC	BIC	LLF Rank	AIC Rank	BIC Rank
Normal	1878.08	-1735.08	-1331.13	6	6	6
t	2094.57	-1950.57	-1543.79	5	5	4
JD	2103.27	-1957.27	-1544.84	2	3	3
VG	2102.66	-1957.66	-1548.06	3	2	2
NIG	2118.45	-1973.45	-1563.85	1	1	1
GHST	2095.81	-1950.81	-1541.21	4	4	5

Panel D: the Sweden Mortality Data

Model	LLF	AIC	BIC	LLF Rank	AIC Rank	BIC Rank
Normal	1856.17	-1713.17	-1309.21	6	6	6
t	1918.32	-1774.32	-1367.54	5	5	2
JD	1920.62	-1774.64	-1362.22	2	4	5
VG	1922.51	-1777.51	-1367.91	1	1	1
NIG	1919.66	-1774.66	-1365.06	4	3	4
GHST	1920.00	-1775.00	-1365.40	3	2	3

Panel E: the Switzerland Mortality Data

Model	LLF	AIC	BIC	LLF Rank	AIC Rank	BIC Rank
Normal	1612.69	-1469.69	-1065.74	6	6	6
t	1806.21	-1662.21	-1255.43	5	4	4
JD	1815.79	-1669.79	-1257.36	3	3	3
VG	1826.48	-1681.48	-1271.88	2	2	2
NIG	1843.55	-1698.55	-1288.95	1	1	1
GHST	1806.74	-1661.74	-1252.14	4	5	5

Panel F: the U.S. Mortality Data

Model	LLF	AIC	BIC	LLF Rank	AIC Rank	BIC Rank
Normal	1199.95	-1074.95	-756.82	6	6	6
t	1273.77	-1147.77	-827.09	4	3	1
JD	1277.47	-1149.47	-823.71	1	1	4
VG	1271.16	-1144.16	-820.94	5	5	5
NIG	1274.83	-1147.83	-824.61	2	2	2
GHST	1274.19	-1147.19	-823.97	3	4	3

Table 3-3. Goodness-of-fit Tests for the Residuals of the Lee-Carter Model

Panel A: the Finland Mortality Data

Model	KS			AD			CvM		
	Statistic	Critical Value		Statistic	Critical Value		Statistic	Critical Value	
		5%	1%		5%	1%		5%	1%
Normal	0.061**	0.029	0.035	21.681**	2.443	3.914	3.365**	0.458	0.746
t	0.031*	0.029	0.035	2.624*	2.486	3.868	0.423	0.463	0.737
JD	0.024	0.029	0.035	1.769	2.491	3.799	0.252	0.452	0.716
VG	0.022	0.029	0.035	1.352	2.473	3.909	0.190	0.466	0.747
NIG	0.016	0.030	0.035	1.028	2.516	3.906	0.104	0.464	0.753
GHST	0.024	0.029	0.036	1.590	2.494	4.098	0.188	0.457	0.786

Note: * and ** denote significance at the 5% and 1% level, respectively.

Panel B: the France Mortality Data

Model	KS			AD			CvM		
	Statistic	Critical Value		Statistic	Critical Value		Statistic	Critical Value	
		5%	1%		5%	1%		5%	1%
Normal	0.094**	0.029	0.035	35.657**	2.443	3.914	5.671**	0.458	0.746
t	0.041**	0.029	0.035	7.765**	2.516	3.829	0.970**	0.463	0.724
JD	0.021	0.029	0.035	1.329	2.465	3.780	0.176	0.456	0.716
VG	0.040**	0.029	0.036	4.513**	2.467	3.852	0.462*	0.459	0.727
NIG	0.026	0.029	0.035	2.420	2.511	3.951	0.275	0.461	0.754
GHST	0.031*	0.029	0.035	4.020*	2.511	4.026	0.425	0.456	0.772

Note: * and ** denote significance at the 5% and 1% level, respectively.

Panel C: the Netherlands Mortality Data

Model	KS			AD			CvM		
	Statistic	Critical Value		Statistic	Critical Value		Statistic	Critical Value	
		5%	1%		5%	1%		5%	1%
Normal	0.062**	0.029	0.035	18.743**	2.443	3.914	3.118**	0.458	0.746
t	0.023	0.029	0.035	1.705	2.533	3.807	0.210	0.466	0.726
JD	0.018	0.029	0.035	0.858	2.473	3.762	0.130	0.458	0.710
VG	0.024	0.029	0.035	2.336	2.448	3.742	0.272	0.454	0.727
NIG	0.013	0.030	0.035	0.335	2.547	3.899	0.038	0.470	0.741
GHST	0.019	0.029	0.035	1.692	2.483	3.897	0.194	0.460	0.758

Note: * and ** denote significance at the 5% and 1% level, respectively.

Panel D: the Sweden Mortality Data

Model	KS			AD			CvM		
	Statistic	Critical Value		Statistic	Critical Value		Statistic	Critical Value	
		5%	1%		5%	1%		5%	1%
Normal	0.031*	0.029	0.035	3.721*	2.443	3.914	0.590*	0.458	0.746
t	0.015	0.029	0.035	0.571	2.488	3.881	0.084	0.466	0.751
JD	0.014	0.029	0.035	0.429	2.472	3.857	0.060	0.452	0.736
VG	0.024	0.029	0.035	1.805	2.477	3.909	0.282	0.466	0.747
NIG	0.014	0.029	0.035	0.379	2.495	3.893	0.047	0.461	0.740
GHST	0.013	0.029	0.035	0.370	2.488	3.879	0.043	0.468	0.732

Note: * and ** denote significance at the 5% and 1% level, respectively.

Panel E: the Switzerland Mortality Data

Model	KS			AD			CvM		
	Statistic	Critical Value		Statistic	Critical Value		Statistic	Critical Value	
		5%	1%		5%	1%		5%	1%
Normal	0.058**	0.029	0.035	17.921**	2.443	3.914	3.033**	0.458	0.746
t	0.025	0.029	0.035	2.720*	2.470	3.948	0.327	0.460	0.742
JD	0.027	0.029	0.035	2.621*	2.483	3.714	0.421	0.454	0.719
VG	0.016	0.029	0.035	0.865	2.456	3.805	0.108	0.454	0.730
NIG	0.017	0.030	0.036	0.699	2.579	3.978	0.081	0.473	0.756
GHST	0.025	0.029	0.035	2.677*	2.514	3.885	0.323	0.464	0.746

Note: * and ** denote significance at the 5% and 1% level, respectively.

Panel F: the U.S. Mortality Data

Model	KS			AD			CvM		
	Statistic	Critical Value		Statistic	Critical Value		Statistic	Critical Value	
		5%	1%		5%	1%		5%	1%
Normal	0.050**	0.039	0.046	7.501**	2.491	3.902	1.229**	0.464	0.748
t	0.047**	0.039	0.047	1.711	2.551	4.045	0.296	0.473	0.763
JD	0.047**	0.039	0.047	1.346	2.459	3.933	0.234	0.461	0.757
VG	0.038	0.039	0.046	2.291	2.488	3.880	0.361	0.460	0.745
NIG	0.038	0.039	0.046	1.430	2.479	4.001	0.246	0.455	0.756
GHST	0.039*	0.039	0.046	1.523	2.541	3.868	0.249	0.474	0.744

Note: * and ** denote significance at the 5% and 1% level, respectively.

Table 3-4 contains the results for the LLF, AIC, and BIC and their corresponding ranks in terms of the normal, *t*, JD, VG, NIG, and GHST distributions for the first difference of mortality indices. The Gaussian model is the worst, according to the LLF, AIC, and BIC. The LLF criterion also indicates that the best goodness-of-fit derives from the NIG model for the Netherlands but from the JD model for the five other countries. Because it introduces a penalty term for the effective number of parameters, the best in-sample goodness of fit changes for the *t* distribution, except for France and the Netherlands. According to the BIC, the NIG model again is the best fit for the Netherlands, the JD model is the best for France, and the *t* model is the best one for Finland, Sweden, Switzerland, and the United States. Table 3-5 lists the results for the KS, AD, and CvM tests, together with their critical values, pertaining to the error terms of the mortality indices. The results reject the notion that the error terms of the mortality indices for France, the Netherlands, and the United States come from a normal distribution. All three test results confirm that the error terms of the mortality indices come from non-Gaussian distributions. Therefore, the goodness-of-fit tests consistently indicate that non-Gaussian distributions provide better in-sample goodness of fit for the error terms of the mortality indices.

Table 3-4. Goodness-of-fit Tests for the First Difference in Mortality Indices

Panel A: the Finland Mortality Index

Model	LLF	AIC	BIC	LLF Rank	AIC Rank	BIC Rank
Normal	-208.78	210.78	213.38	6	6	6
t	-195.83	198.83	202.72	4	2	1
JD	-192.28	197.28	203.77	1	1	2
VG	-197.68	201.68	206.87	5	5	5
NIG	-195.34	199.34	204.53	2	3	3
GHST	-195.63	199.63	204.82	3	4	4

Panel B: the France Mortality Index

Model	LLF	AIC	BIC	LLF Rank	AIC Rank	BIC Rank
Normal	-222.77	224.77	227.36	6	6	6
t	-204.48	207.48	211.38	4	4	3
JD	-195.99	200.99	207.48	1	1	1
VG	-204.51	208.51	213.70	5	5	5
NIG	-198.66	202.66	207.85	2	2	2
GHST	-202.81	206.81	212.01	3	3	4

Panel C: the Netherlands Mortality Index

Model	LLF	AIC	BIC	LLF Rank	AIC Rank	BIC Rank
Normal	-209.29	211.29	213.88	6	6	6
t	-184.02	187.02	190.91	5	5	3
JD	-181.28	186.28	192.77	2	4	5
VG	-181.98	185.98	191.17	4	3	4
NIG	-179.93	183.93	189.12	1	1	1
GHST	-181.40	185.40	190.59	3	2	2

Panel D: the Sweden Mortality Index

Model	LLF	AIC	BIC	LLF Rank	AIC Rank	BIC Rank
Normal	-183.24	185.24	187.84	6	6	6
t	-175.64	178.64	182.53	3	1	1
JD	-174.79	179.79	186.28	1	4	5
VG	-176.59	180.59	185.78	5	5	4
NIG	-175.65	179.65	184.84	4	3	3
GHST	-175.49	179.49	184.68	2	2	2

Panel E: the Switzerland Mortality Index

Model	LLF	AIC	BIC	LLF Rank	AIC Rank	BIC Rank
Normal	-168.78	170.78	173.37	6	6	6
t	-150.54	153.54	157.43	3	2	1
JD	-147.09	152.09	158.58	1	1	2
VG	-153.58	157.58	162.77	5	5	5
NIG	-151.70	155.70	160.89	4	4	4
GHST	-150.37	154.37	159.56	2	3	3

Panel F: the U.S. Mortality Index

Model	LLF	AIC	BIC	LLF Rank	AIC Rank	BIC Rank
Normal	-92.74	94.74	97.34	6	6	6
t	-69.39	72.39	76.28	3	2	1
JD	-65.90	70.90	77.38	1	1	2
VG	-70.89	74.89	80.08	5	5	5
NIG	-70.73	74.73	79.92	4	4	4
GHST	-69.35	73.35	78.54	2	3	3

Table 3-5. Goodness-of-fit Tests for the First Difference in Mortality Indices

Panel A: the Finland Mortality Index

Model	KS			AD			CvM		
	Statistic	Critical Value		Statistic	Critical Value		Statistic	Critical Value	
		5%	1%		5%	1%		5%	1%
Normal	0.122	0.130	0.158	2.351	2.477	3.933	0.329	0.461	0.737
t	0.052	0.131	0.158	0.364	2.502	3.944	0.043	0.462	0.746
JD	0.053	0.130	0.159	0.218	2.529	3.941	0.037	0.457	0.753
VG	0.063	0.130	0.159	0.555	2.481	3.994	0.068	0.465	0.769
NIG	0.053	0.130	0.158	0.355	2.470	3.700	0.051	0.457	0.707
GHST	0.054	0.131	0.157	0.313	2.509	3.923	0.038	0.459	0.728

Note: * and ** denote significance at the 5% and 1% level, respectively.

Panel B: the France Mortality Index

Model	KS			AD			CvM		
	Statistic	Critical Value		Statistic	Critical Value		Statistic	Critical Value	
		5%	1%		5%	1%		5%	1%
Normal	0.156*	0.130	0.158	4.333**	2.477	3.933	0.720*	0.461	0.737
t	0.104	0.131	0.158	1.438	2.476	3.932	0.233	0.458	0.735
JD	0.065	0.131	0.158	0.411	2.513	3.980	0.077	0.459	0.762
VG	0.106	0.131	0.158	1.275	2.532	3.961	0.224	0.469	0.768
NIG	0.064	0.130	0.157	0.458	2.484	3.845	0.073	0.457	0.724
GHST	0.077	0.131	0.157	0.754	2.509	3.923	0.093	0.459	0.728

Note: * and ** denote significance at the 5% and 1% level, respectively.

Panel C: the Netherlands Mortality Index

Model	KS			AD			CvM		
	Statistic	Critical Value		Statistic	Critical Value		Statistic	Critical Value	
		5%	1%		5%	1%		5%	1%
Normal	0.144*	0.130	0.158	4.236**	2.477	3.933	0.720*	0.461	0.737
t	0.092	0.131	0.157	1.226	2.509	3.887	0.206	0.464	0.735
JD	0.069	0.131	0.159	0.361	2.551	3.899	0.070	0.469	0.760
VG	0.064	0.131	0.157	0.640	2.480	3.881	0.077	0.455	0.744
NIG	0.052	0.131	0.157	0.218	2.481	3.884	0.027	0.459	0.749
GHST	0.058	0.130	0.156	0.348	2.508	3.825	0.046	0.456	0.728

Note: * and ** denote significance at the 5% and 1% level, respectively.

Panel D: the Sweden Mortality Index

Model	KS			AD			CvM		
	Statistic	Critical Value		Statistic	Critical Value		Statistic	Critical Value	
		5%	1%		5%	1%		5%	1%
Normal	0.093	0.130	0.158	1.192	2.477	3.933	0.164	0.461	0.737
t	0.054	0.131	0.158	0.289	2.503	3.944	0.036	0.462	0.746
JD	0.062	0.131	0.158	0.215	2.493	3.935	0.034	0.469	0.739
VG	0.058	0.131	0.160	0.302	2.500	3.893	0.041	0.469	0.758
NIG	0.058	0.130	0.156	0.254	2.487	3.916	0.038	0.454	0.726
GHST	0.054	0.131	0.157	0.238	2.509	3.924	0.034	0.459	0.728

Note: * and ** denote significance at the 5% and 1% level, respectively.

Panel E: the Switzerland Mortality Index

Model	KS			AD			CvM		
	Statistic	Critical Value		Statistic	Critical Value		Statistic	Critical Value	
		5%	1%		5%	1%		5%	1%
Normal	0.104	0.130	0.158	2.244	2.477	3.933	0.310	0.461	0.737
t	0.060	0.131	0.157	0.345	2.511	3.887	0.052	0.467	0.735
JD	0.063	0.130	0.158	0.243	2.515	3.902	0.047	0.464	0.746
VG	0.066	0.131	0.157	0.552	2.521	3.940	0.073	0.465	0.761
NIG	0.064	0.131	0.157	0.411	2.443	3.830	0.061	0.457	0.733
GHST	0.061	0.130	0.156	0.323	2.507	3.825	0.055	0.456	0.728

Note: * and ** denote significance at the 5% and 1% level, respectively.

Panel F: the U.S. Mortality Index

Model	KS			AD			CvM		
	Statistic	Critical Value		Statistic	Critical Value		Statistic	Critical Value	
		5%	1%		5%	1%		5%	1%
Normal	0.140*	0.130	0.158	3.123*	2.477	3.933	0.456	0.461	0.737
t	0.077	0.131	0.157	0.508	2.508	3.887	0.083	0.466	0.733
JD	0.094	0.130	0.158	0.484	2.510	4.052	0.098	0.461	0.757
VG	0.109	0.132	0.159	0.926	2.552	4.004	0.136	0.474	0.767
NIG	0.076	0.130	0.157	0.546	2.477	3.867	0.076	0.460	0.743
GHST	0.076	0.130	0.156	0.493	2.507	3.824	0.081	0.456	0.728

Note: * and ** denote significance at the 5% and 1% level, respectively.

Mortality Projection

For out-of-sample performance, we apply the parameters estimated from 1900–1999 and obtain the mortality projection with 1,000,000 simulation paths. For each path, we can calculate the mean absolute percentage error (MAPE) as follows:

$$MAPE = 100\% \times \frac{1}{n} \sum_{i=1}^{n} \left| \frac{A_i - F_i}{A_i} \right|, \tag{3-13}$$

where A_i is the historical mortality rate and F_i is the forecast mortality rate. When we apply the calibrated parameters of the Lee-Carter model with the best BIC goodness-of-fit innovations, we find the impacts on different distributions of the mortality projection for the mean, 90th percentile, and 95th percentile of MAPE from 2000 to 2007, as we show in Table 3-6. Lower values indicate better predictive power for the fitted distribution. According to the average rank of the MAPE criterion, the normal distribution provides poor mortality projection performance; the *t* and its skew extension GHST provide the best mortality projection for all mortality data. Thus, the Lee-Carter model with non-Gaussian distributions provides a better mortality projection than that obtained from a normal distribution, in terms of the MAPE criterion.

Table 3-6. Percentile of MAPE of Mortality Projection

Panel A: the Finland Mortality Data (Unit: %)

Model	Mean	90%	95%	Mean Rank	90% Rank	95% Rank	Average Rank
NIG-Normal	8.273	10.406	11.132	2	6	6	4.67
NIG-t	8.356	10.270	11.016	6	5	4	5.00
NIG-JD	8.301	10.195	11.047	5	2	5	4.00
NIG-VG	8.259	10.240	10.987	1	4	2	2.33
NIG-NIG	8.282	10.226	11.011	3	3	3	3.00
NIG-GHST	8.284	10.085	10.755	4	1	1	2.00

Note: X-Y model means that the error terms in Equations (3-1) and (3-2) are the X and Y models, respectively.

Panel B: the France Mortality Data (Unit: %)

Model	Mean	90%	95%	Mean Rank	90% Rank	95% Rank	Average Rank
JD-Normal	4.866	6.797	7.833	6	4	4	4.67
JD-t	4.772	6.562	7.692	2	2	3	2.33
JD-JD	4.862	6.893	8.359	5	6	5	5.33
JD-VG	4.789	6.579	7.644	3	3	2	2.67
JD-NIG	4.853	6.858	8.435	4	5	6	5.00
JD-GHST	4.475	5.808	6.655	1	1	1	1.00

Note: X-Y model means that the error terms in Equations (3-1) and (3-2) are the X and Y models, respectively.

Panel C: the Netherlands Mortality Data (Unit: %)

Model	Mean	90%	95%	Mean Rank	90% Rank	95% Rank	Average Rank
NIG-Normal	3.696	5.289	6.116	6	4	3	4.33
NIG-t	3.548	5.003	5.926	3	2	2	2.33
NIG-JD	3.588	5.334	6.573	4	5	6	5.00
NIG-VG	3.641	5.355	6.511	5	6	5	5.33
NIG-NIG	3.516	5.041	6.199	2	3	4	3.00
NIG-GHST	3.227	4.253	4.938	1	1	1	1.00

Note: X-Y model means that the error terms in Equations (3-1) and (3-2) are the X and Y models, respectively.

Panel D: the Sweden Mortality Data (Unit: %)

Model	Mean	90%	95%	Mean Rank	90% Rank	95% Rank	Average Rank
VG-Normal	8.020	9.853	10.446	4	6	4	4.67
VG-t	7.922	9.617	10.251	1	1	1	1.00
VG-JD	8.024	9.829	10.574	6	5	6	5.67
VG-VG	8.015	9.780	10.419	3	3	2	2.67
VG-NIG	8.020	9.789	10.473	5	4	5	4.67
VG-GHST	8.013	9.731	10.419	2	2	3	2.33

Note: X-Y model means that the error terms in Equations (3-1) and (3-2) are the X and Y models, respectively.

Panel E: the Switzerland Mortality Data (Unit: %)

Model	Mean	90%	95%	Mean Rank	90% Rank	95% Rank	Average Rank
NIG-Normal	3.816	4.798	5.321	5	6	5	5.33
NIG-t	3.783	4.652	5.187	4	4	4	4.00
NIG-JD	3.817	4.656	5.341	6	5	6	5.67
NIG-VG	3.763	4.601	5.046	2	2	2	2.00
NIG-NIG	3.765	4.616	5.107	3	3	3	3.00
NIG-GHST	3.741	4.542	5.012	1	1	1	1.00

Note: X-Y model means that the error terms in Equations (3-1) and (3-2) are the X and Y models, respectively.

Panel F: the U.S. Mortality Data (Unit: %)

Model	Mean	90%	95%	Mean Rank	90% Rank	95% Rank	Average Rank
t-Normal	3.266	4.247	4.676	6	6	6	6.00
t-t	3.191	3.971	4.357	2	2	2	2.00
t-JD	3.244	4.039	4.668	4	3	5	4.00
t-VG	3.265	4.118	4.548	5	5	4	4.67
t-NIG	3.228	4.047	4.457	3	4	3	3.33
t-GHST	3.180	3.943	4.314	1	1	1	1.00

Note: X-Y model means that the error terms in Equations (3-1) and (3-2) are the X and Y models, respectively.

3.3 Conclusions

Recently, many researchers have examined mortality rates and explored different models. Some studies demonstrate that mortality rate improvements also exhibit jump properties. We therefore attempt to incorporate five heavy-tailed distributions—*t*, JD, VG, NIG, and GHST—into the Lee-Carter model. Using mortality data from six countries, we apply the BIC and KS, AD, and CvM tests and find consistent support for the non-Gaussian residuals of the Lee-Carter model. Specifically, when we calibrate the parameters of the Lee-Carter model, the JD-JD model[6] is the best one for French mortality data, the NIG-NIG model is best for the Netherlands, the VG-*t* model offers the best goodness of fit for Swedish mortality data, the *t*-*t* model is best for the U.S. mortality data, and the NIG-*t* model is the best one for the mortality data from Finland and Switzerland. For forecasting mortality rates, we find that the normal distribution provides weak mortality projection performance, whereas *t* and its skew extension provide good mortality projections. Therefore, for applications of the Lee-Carter model, the heavy-tailed distributions appear to be the most appropriate choices for modeling long-term mortality data.

[6] The terminology "X-Y model" refers to the error terms in Equations (3-1) and (3-2), respectively.

國立政治大學
National Chengchi University

Chapter 4

Mortality Modeling with Non-Gaussian Innovations and Applications to the Valuation of Longevity Swaps

4.1 Stochastic Mortality Models with Cox Error Structures

In this section, we first review the RH model, in which the mortality index and cohort effect follow ARIMA models with normal innovations. However, according to the mortality data, the residuals exhibit non-Gaussian distribution. Consequently, we assume that the number of deaths follows a Cox process and that the death rates adhere to the RH model in which the residuals, the mortality indices, and the cohort effects follow three non-Gaussian distributions: JD, VG, and NIG. We also develop an iterative process for calibrating the corresponding parameters of the Cox process with leptokurtic intensity.

Renshaw and Haberman's (2006) Model

We analyze changes in mortality as a function of both age *x* and time *t*. For mortality forecasting, the cohort-based extension to the LC model proposed by Renshaw and Haberman (2006) is as follows:

$$ln\left(m_{x,t}\right)=\alpha_x+\beta_x k_t+\eta_x\gamma_{t-x}+e_{x,t}, \tag{4-1}$$

where $m_{x,t}$ is the death rate for age *x* in calendar year *t*, defined as running from time t to time $t+1$; α_x describes the average pattern of mortality over an age group; k_t explains the time trend of the general mortality level; β_x represents age-specific patterns of mortality change, indicating the sensitivity of the logarithm of the force of mortality at age *x* to variations in k_t; γ_{t-x} is a cohort effect; η_x controls age-specific cohort contributions to the mortality projection; and $e_{x,t}$ represents the error term, which is normally distributed with mean 0 and variance σ_e^2. This structure is designed to capture age–period–cohort effects.

To forecast future mortality dynamics, the mortality index k_t follows a one-dimensional random walk with drift (Lee and Carter, 1992), as follows:

$$k_t-k_{t-1}=\mu+\varepsilon_t, \tag{4-2}$$

where μ is a drift term and ε_t is a sequence of independent and identically zero-mean Gaussian random variables. Let the year of birth be equal to $c=t-x$. Following the model setup of Renshaw and Haberman (2006) and Cairns et al. (2010), we model the cohort factor γ_c as an ARIMA(1,1,0) process that is independent of k_t:

$$\Delta\gamma_c=\mu_\gamma+\alpha_\gamma\left(\Delta\gamma_{c-1}-\mu_\gamma\right)+\sigma_\gamma z_c, \tag{4-3}$$

where z_c is a sequence of independent and identically standard normal random variables.

Normality Test for the RH Model

According to table 1 in Dowd et al.'s (2010) article, the residuals of the Renshaw and Haberman (2006) model exhibit leptokurticity. In this subsection, we therefore apply the JB statistic (Jarque and Bera, 1980) to test empirically the normality of the three mortality data sets from England and Wales, France, and Italy for subjects aged 60–89 years during the period 1900–1983. The mortality data came from the Human Mortality Database (HMD) website.[7] Table 4-1 contains the results of the JB test for the residuals of the RH model, the first difference of the three countries' mortality indices, and the corresponding cohort effects from 1900 to 1983. The JB statistic rejects the assumption of normality for the residuals of the RH model and the cohort effects. Therefore, we use the heavy-tailed distributions—JD, VG and NIG—to model the non-Gaussian nature of the error terms of the RH model.

Table 4-1. The Jarque-Bera Test

	England and Wales	France	Italy
Residuals of the RH Model	373.952 [< 0.001]	1489.534 [< 0.001]	15531.672 [< 0.001]
First Difference in Mortality Indices	1.175 [0.477]	0.638 [0.500]	2.532 [0.182]
the Residuals of Cohort Effects	343.557 [< 0.001]	43.933 [< 0.001]	295.214 [< 0.001]

Note: The *p*-values of the Jarque-Bera test are in brackets.

[7] See http://www.mortality.org/.

Heavy-Tailed Distributions

We model the error terms of the RH model, $e_{x,t}$, ε_t, and z_{t-x}, using the three heavy-tailed distributions: JD, VG and NIG. In the subsequent subsection, we take $e_{x,t}$ as an example to describe the properties of these heavy-tailed distributions; analogous results are obtained for ε_t and z_{t-x}. We refer to these distributions in Chapter 2.

A Cox Process with Leptokurtic Intensity

We assume that $D_{x,t}$, or the number of deaths at age *x* during year *t*, adheres to a Cox process, also known as a doubly stochastic Poisson process. That is, $D_{x,t} \sim Cox\left(\lambda_{x,t}\right)$, where $\lambda_{x,t} = E_{x,t} m_{x,t}$ is a non-negative stochastic intensity process, and $E_{x,t}$ is the exposure to risk at age *x* during year *t*. When death rates adhere to the RH model, $\lambda_{x,t}$ can be modeled as

$$\lambda_{x,t} = E_{x,t} m_{x,t} = E_{x,t} \exp\left(\alpha_x + \beta_x k_t + \eta_x \gamma_{t-x} + e_{x,t}\right), \tag{4-4}$$

where $e_{x,t}$ is assumed to be an age- and period-homogeneous heavy-tailed distribution that captures leptokurticity. Let $d_{x,t}$ be the corresponding number of deaths actually observed. Conditional on $e_{x,t} = y$, the number of deaths $D_{x,t}$ becomes a Poisson distribution with intensity $E_{xt} \exp\left(\alpha_x + \beta_x k_t + \eta_x \gamma_{t-x} + y\right)$. As a result, the log-likelihood function based on the Cox regression model is defined as

$$LLF = \sum_{x,t} \int_{-\infty}^{\infty} \log f\left(D_{x,t} = d_{x,t} \mid e_{x,t} = y\right) f_{e_{x,t}}(y)\, dy, \tag{4-5}$$

where

$$\log f\left(D_{x,t}=d_{x,t}\,|\,e_{x,t}=y\right)$$

$$=d_{x,t}\log\left(E_{x,t}\exp\left(\alpha_x+\beta_x k_t+\eta_x\gamma_{t-x}+y\right)\right)-E_{x,t}\exp\left(\alpha_x+\beta_x k_t+\eta_x\gamma_{t-x}+y\right)-\log\left(d_{x,t}!\right). \quad (4\text{-}6)$$

Thus, to find the maximum likelihood estimates of the parameters of the Renshaw and Haberman (2006) model, we can maximize Equation (4-5) with respect to α_x, β_x, k_t, η_x, γ_{t-x}, and the error term distribution parameters. The closed-form solution of the log-likelihood function in Equation (4-5) is derived as follows:

$$LLF=\sum_{x,t}\left[d_{x,t}\left(\alpha_x+\beta_x k_t+\eta_x\gamma_{t\text{-}x}\right)\text{-}\left(E_{x,t}\exp\left(\alpha_x+\beta_x k_t+\eta_x\gamma_{t\text{-}x}\right)\right)M_{e_{x,t}}(1)\right]+C, \quad (4\text{-}7)$$

where $M_{e_{x,t}}(u)$ is the moment-generating function of $e_{x,t}$, and C represents a constant term equal to $\sum_{x,t}\left[d_{x,t}\log E_{x,t}-\log\left(d_{x,t}!\right)\right]$. The proof of Equation (4-7) is in Appendix C. Note that when $e_{x,t}$ is ignored while modeling the number of deaths (i.e., $M_{e_{x,t}}(1)=1$), the log-likelihood function defined in Equation (4-7) is precisely the same as that proposed by Wilmoth (1993), Brouhns et al. (2002), and Cairns et al. (2009).

Similar to the two-step procedures of Lee and Carter (1992), Brouhns et al. (2002), and Renshaw and Haberman (2006), we first calibrate the parameters α_x, β_x, k_t, η_x, and $\gamma_{t\text{-}x}$ with an updating scheme. Then, we estimate μ, μ_γ, α_γ, σ_γ, and the distribution parameters of the residuals of the mortality indices and cohort effects. In the first step, there are six sets of parameters, namely, the α_x, β_x, k_t, η_x, and $\gamma_{t\text{-}x}$ parameters, as well as the $e_{x,t}$ distribution parameters. Following Brouhns et al. (2002) and Renshaw and Haberman (2006), we use the following updating scheme: Let n_x be the total number of ages. Starting with $\eta_x=1/n_x$ and

$\gamma_{t\text{-}x}=0$ and then obtaining α_x, β_x, and k_t from the approximation method of the LC model, we calibrate the corresponding $e_{x,t}$ distribution parameters by maximizing the sample log-likelihood function:

$$LLF_{e_{x,t}}=\sum_{x,t}\log\left(f\left(e_{xt}\middle|\alpha_x,\eta_x,\gamma_{t\text{-}x},\beta_x,k_t\right)\right). \tag{4-8}$$

Then, with the $e_{x,t}$ distribution parameters estimated from Equation (4-8), we employ an iterating method to estimate the corresponding parameters of the RH model according to the elementary Newton method (Goodman, 1979; Brouhns et al., 2002; Renshaw and Haberman, 2006).

Following the estimating procedure of Renshaw and Haberman (2006), the parameters are estimated by iteration. In each iteration step, we update a single set of parameters; the other parameters are fixed at their current estimates using the following updating scheme:

$$update(\theta)=u(\theta)=\theta-\frac{\partial LLF/\partial\theta}{\partial^2 LLF/\partial\theta^2}. \tag{4-9}$$

Consequently, the updating scheme is as follows:

$$u(\alpha_x)=\alpha_x+\frac{\sum_t\left[d_{x,t}-E_{x,t}\exp\left(\alpha_x+\beta_x k_t+\eta_x\gamma_{t-x}\right)M_{e_{x,t}}(1)\right]}{\sum_t\left[E_{x,t}\exp\left(\alpha_x+\beta_x k_t+\eta_x\gamma_{t-x}\right)M_{e_{x,t}}(1)\right]}, \tag{4-10}$$

$$u(\gamma_z)=\gamma_z+\frac{\sum_{\substack{x,t\\z=t-x}}\left[d_{x,t}\eta_x-E_{x,t}\eta_x\exp\left(\alpha_x+\beta_x k_t+\eta_x\gamma_z\right)M_{e_{x,t}}(1)\right]}{\sum_{\substack{x,t\\z=t-x}}\left[E_{x,t}\eta_x^2\exp\left(\alpha_x+\beta_x k_t+\eta_x\gamma_z\right)M_{e_{x,t}}(1)\right]}, \tag{4-11}$$

$$u(\eta_x)=\eta_x+\frac{\sum_t\left[d_{x,t}\gamma_{t-x}-E_{x,t}\gamma_{t-x}\exp\left(\alpha_x+\beta_x k_t+\eta_x\gamma_{t-x}\right)M_{e_{x,t}}(1)\right]}{\sum_t\left[E_{x,t}\gamma_{t-x}^2\exp\left(\alpha_x+\beta_x k_t+\eta_x\gamma_{t-x}\right)M_{e_{x,t}}(1)\right]}, \tag{4-12}$$

$$u(k_t) = k_t + \frac{\sum_x \left[d_{x,t}\beta_x - E_{x,t}\beta_x \exp(\alpha_x + \beta_x k_t + \eta_x \gamma_{t-x}) M_{e_{x,t}}(1) \right]}{\sum_x \left[E_{x,t}\beta_x^2 \exp(\alpha_x + \beta_x k_t + \eta_x \gamma_{t-x}) M_{e_{x,t}}(1) \right]}, \text{ and} \quad (4\text{-}13)$$

$$u(\beta_x) = \beta_x + \frac{\sum_t \left[d_{x,t}k_t - E_{x,t}k_t \exp(\alpha_x + \beta_x k_t + \eta_x \gamma_{t-x}) M_{e_{x,t}}(1) \right]}{\sum_t \left[E_{x,t}k_t^2 \exp(\alpha_x + \beta_x k_t + \eta_x \gamma_{t-x}) M_{e_{x,t}}(1) \right]}. \quad (4\text{-}14)$$

We repeat the updating cycle (Equations (4-8)–(4-14)) and stop when the log-likelihood function in Equation (4-7) converges.[8] Model identification can be conveniently achieved with parameter constraints: $\sum_t k_t = 0$, $\sum_x \beta_x = 1$, $\sum_x \eta_x = 1$, and $\sum_t \gamma_{t-x} = 0$.[9]

After obtaining the mortality indices and cohort effects, we can calculate the parameters of Equations (4-2) and (4-3) by maximizing the log-likelihood function, as follows:

$$\sum_t \log\left(f(\varepsilon_t)\right) \text{ and } \sum_{s=t-x} \log\left(f(z_s)\right), \quad (4\text{-}15)$$

where $f(\varepsilon_t)$ and $f(z_s)$ are the probability density functions of ε_t and z_s, respectively.

4.2 Empirical Analysis

In this section, we investigate the goodness-of-fit distributions for the number of deaths, the first differences of the mortality indices, and the cohort effects. Using the

[8] The criterion used to stop the iterative fitting procedure is a very small relative change in the log-likelihood function. We adopt 10^{-7} as the default value.

[9] Similar to Brouhns et al. (2002), after updating the k_t parameters, we impose a centering constraint $\sum_t k_t = 0$ by removing $\sum_t k_t$ from k_t. After updating the β_x parameters, a scaling constraint $\sum_x \beta_x = 1$ must be imposed by dividing the estimates for β_x by $\sum_x \beta_x$ and multiplying the estimates for k_t by the same number. Following the analogical procedure, the constraints of η_x and γ_{t-x} are also achieved.

mortality data from 1900–1983, we fit the residuals of the RH model to four distributions: normal, JD, VG, and NIG. We then fit the mortality indices and cohort effects from the best-fitting model according to the Bayesian information criterion (BIC) to the same four distributions. Finally, we project the subsequent 25-year mortality rates (1984–2008).

In-Sample Goodness of Fit

Using mortality data from 1900 to 1983, we first investigate the goodness-of-fit distributions of the number of deaths for England and Wales, France, and Italy. Table 4-2 presents the LLF, Akaike information criterion (AIC), and BIC statistics[10] for the number of deaths at which the residuals of the RH model adhere to the normal, JD, VG, and NIG models. All three criteria indicate that the normal distribution is the worst fitting model for the number of deaths. They also indicate that the VG model is consistently the best model for the number of deaths in the three mortality data sets. Therefore, we use the mortality indices and cohort effects obtained from the VG model to investigate the pattern of the error terms of the time and cohort effects.

Table 4-2. Goodness-of-fit Measures for the Number of Deaths

Model	England and Wales			France			Italy		
	LLF	AIC	BIC	LLF	AIC	BIC	LLF	AIC	BIC
Normal	-32727	33011	33839	-30356	30640	31468	-45130	45414	46242
JD	-32557	32844	33681	-30237	30524	31361	-43719	44006	44843
VG	**-32554**	**32840**	**33674**	**-30084**	**30370**	**31204**	**-42900**	**43186**	**44020**
NIG	-32556	32842	33676	-30236	30522	31356	-44314	44600	45434

The test results for the first difference in mortality indices are in Table 4-3, which contains the LLF, AIC, and BIC statistics for the normal, JD, VG, and NIG distributions. The Gaussian model is the worst according to the LLF criterion, which

[10] $AIC = -LLF + NP$ and $BIC = -LLF + 0.5 \times NP \times \log(NS)$, where NP is the effective number of parameters being estimated and NS is the number of observations.

also indicates that the best fit for the three mortality data sets derives from the JD model. The best in-sample goodness of fit for mortality indices changes for the normal distribution of all mortality data, because the BIC introduces a penalty term for the effective number of parameters.

Table 4-3. Goodness-of-fit Tests for the First Difference in Mortality Indices

Model	England and Wales			France			Italy		
	LLF	AIC	BIC	LLF	AIC	BIC	LLF	AIC	BIC
Normal	-164.87	166.87	**169.29**	-174.07	**176.07**	**178.49**	-179.71	**181.71**	**184.13**
JD	**-161.89**	166.89	172.93	**-173.71**	178.71	184.75	**-178.59**	183.59	189.64
VG	-162.67	**166.67**	171.50	-173.85	177.85	182.69	-178.67	182.67	187.51
NIG	-163.49	167.49	172.33	-173.84	177.84	182.68	-178.68	182.68	187.52

In Table 4-4 we present the LLF, AIC, and BIC statistics for the normal, JD, VG, and NIG distributions for cohort effects. The LLF, AIC, and BIC statistics consistently indicate that the best fit for Italy derives from the JD model, but for England and Wales, it derives from the NIG model. For France, the best model for cohort effects is the JD model according to LLF, but in terms of the AIC and BIC, the best is the VG model. All three criteria indicate that the normal distribution is the worst fitting model for the cohort effects. Consequently, with mortality data from three countries over the period 1900–1983, in-sample model selection criteria indicate a preference for modeling the RH model with non-Gaussian innovations.

Table 4-4. Goodness-of-fit Tests for the Residuals of Cohort Effects

Model	England and Wales			France			Italy		
	LLF	AIC	BIC	LLF	AIC	BIC	LLF	AIC	BIC
Normal	-96.11	99.11	103.18	-107.63	110.63	114.69	-125.75	128.75	132.81
JD	-84.27	90.27	98.4	**-97.57**	103.57	111.69	**-105.31**	**111.31**	**119.44**
VG	-83.83	88.83	95.6	-97.79	**102.79**	**109.57**	-115.01	120.01	126.79
NIG	**-83.48**	**88.48**	**95.25**	-97.86	102.86	109.64	-111.45	116.45	123.22

Mortality Projection

To assess out-of-sample performance, we apply the parameters estimated from the time period 1900–1983 to obtain 25-year mortality projections, calculating the mean absolute percentage error (MAPE) as follows:

$$MAPE = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{A_i - F_i}{A_i}\right|, \tag{4-16}$$

where A_i is the logarithm of the historical mortality rate; F_i is the natural logarithm of the forecast mortality rate; and n is the number of observations.

By applying the calibrated parameters of the RH model to the VG innovations (the best model according to BIC), we reveal the impact of the different distributions on the mortality projection for MAPE from 1984 to 2008 (Table 4-5). A lower value indicates better predictive power for the distribution. For comparison, we also provide the mortality projection of the original RH model with four forecasting distributions—normal, JD, VG and NIG (the original RH-Normal model corresponds to the M2 model of Cairns et al., 2009). The VG-NIG model[11] is the best mortality projection for the mortality data of England and Wales. The VG-VG model provides the best one for the mortality data from France and Italy. As a result, in terms of the MAPE criterion, the RH model with non-Gaussian innovations provides better mortality projection than that obtained from the original RH model with normal innovations.

[11] A VG-NIG model corresponds to a VG error term in the RH model and to NIG distributions for the time and cohort effects.

Table 4-5. MAPE of Logarithm of Mortality Projection in 1984-2008 (Unit: %)

Model	England and Wales	France	Italy
Original RH-Normal	4.9344	5.1999	7.2268
Original RH-JD	4.9343	5.1915	7.2822
Original RH-VG	4.9393	5.1000	7.9326
Original RH-NIG	4.9319	5.1891	7.2983
VG-Normal	4.8072	5.1719	7.0612
VG-JD	4.8064	5.1619	7.1135
VG-VG	4.807	**5.0595**	**6.9687**
VG-NIG	**4.8063**	5.1598	7.1293

Note: Original RH-Normal is the same as M2 of Cairns et al. (2009). The X-Y model corresponds to an X error term in the RH model and to Y distributions for the time and cohort effects.

4.3 Application: The Valuation of Longevity Swaps

In this section, we first price a longevity swap. Using the mortality data of England and Wales from 1900 to 2008, we then re-fit the RH model to attain the fair swap premium of the longevity swap for both the original RH model (M2) and the best projection model. Finally, we provide the VaR and CTE of the longevity swaps.

Pricing Longevity Swaps

The traditional method of transferring longevity risk in a pension plan or an annuity book is to sell the liability through an insurance or reinsurance contract, known as pension buyouts. These tactics have attracted increasing attention since 2006, especially in the United Kingdom. However, such transactions involve the transfer of all risks, including longevity and investment risk. To transfer longevity risk only to capital markets, Blake and Burrows (2001) first advocate the use of longevity bonds, whose coupon payments depend on the proportion of the population surviving to particular ages. Bauer (2006) and Barbarin (2008) also apply the Heath-Jarrow-Morton methodology (see Heath et al., 1992) to price longevity bonds. The EIB/BNP longevity bond was the first securitization instrument designed to

transfer longevity risk but ultimately was withdrawn. The lack of success in issuing longevity bonds led to new securitization instruments, such as longevity swaps,[12] which were pioneered in capital markets by J.P. Morgan and Canada Life in July 2008. As Blake et al. (2012) show, 16 publicly announced longevity swaps were executed between 2007 and 2012 in the United Kingdom. In this context, the valuation of longevity swaps represents an important research topic for developing capital market solutions for longevity risk.

Longevity swaps have been widely explored in prior literature (Dawson, 2002; Lin and Cox, 2005; Dowd et al., 2006; Dawson et al., 2010; Biffis et al., 2011; Wang and Yang, 2012). Dowd et al. (2006) introduce the mechanism for transferring longevity risk; this instrument involves exchanging actual pension payments for a series of pre-agreed fixed payments. On each payment date, the fixed-rate payer (e.g., pension plan) receives from the hedge supplier a random mortality-dependent payment and, in return, makes a fixed payment to the hedge supplier. Dowd et al. (2006) demonstrate that the hedge is almost perfect when the reference index is based on the survivor experience of the insurer's annuity book. If the expected reference indices and insurers' own survivor experiences are highly correlated, the longevity swap can still hedge the insurer against a considerable amount of the aggregate longevity risk it faces. In this article, following the vanilla longevity swap structure analyzed by Dowd et al. (2006) and Dawson et al. (2010), we discuss a *T*-year bespoke longevity swap linked to a benchmark cohort of a given initial age for the England and Wales mortality data.[13]

For a given time horizon *T*, we consider a filtered probability space

[12] For the recent development of longevity-linked securities, see Blake et al. (2012) and reference therein.

[13] To bear no basis risk, the variable payments in bespoke longevity swaps are designed to match precisely the mortality experience of each individual hedger.

$\left(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t=0}^{T}, P\right)$ on which the death time is modeled as a stopping time τ with respect to $\mathcal{F}_t$. As mentioned by Biffis et al. (2010) and Hainaut (2012), $\mathcal{F}$ is the enlarged filtration $\mathcal{H} \vee \mathcal{G}$ where $\mathcal{H}$ is the filtration related to risk factors and $\mathcal{G}$ is the complement, such that τ is a stopping time on $\mathcal{F}$. Conditional on the path followed by the mortality rates, the *t*-year survival probability that a 65-year-old person in calendar year 2008+*t* reaches age 65+*t* is of the form:

$$S(t) = P\left(\tau > t \middle| \mathcal{H}_T\right) = \exp\left(-\int_0^t m_{65+s,2008+s} ds\right). \tag{4-17}$$

We assume that the mortality rates are constant within certain age and time windows but may vary from one window to the next. Specifically, given any integer age x and calendar year t, we presume that

$$m_{x+\xi,t+\tau} = m_{x,t} \quad \text{for} \quad 0 \le \xi, \tau < 1. \tag{4-18}$$

Thus,

$$S(t) = \exp(-\sum_{h=0}^{t-1} m_{65+h,2008+h}). \tag{4-19}$$

To transfer longevity risk, on each of the payment dates *t*, the fixed-rate payer pays the notional principal multiplied by a prespecified fixed proportion $(1+\pi)H(t)$ to the floating-rate payer and receives the notional principal multiplied by $S(t)$, where $H(t)$ is anticipated by using the best estimate of the underlying mortality model, and π is the swap premium that would be set so that the initial value of the swap is zero for each party.

The distribution function of $S(t)$ under the real-world (physical) probability measure *P* is

$$F_t(y) = Prob_P(S(t) \le y). \tag{4-20}$$

Wang (2000) proposes a distortion operator to change the probability measure from the real-world probability measure *P* to an equivalent martingale measure *Q*, with the following transformation:[14]

$$\tilde{F}_t(y) = \Phi\left(\Phi^{-1}\left(F_t(y)\right) + \lambda\right), \tag{4-21}$$

where λ is a parameter called the market price of risk, and Φ is the standard normal distribution function. Therefore, as shown by Denuit et al. (2007), the expectation value of $S(t)$ under the equivalent martingale measure *Q* is defined as

$$E_Q\left[S(t)\right] = \int_0^1 \left(1 - \tilde{F}_t(y)\right) dy = \int_0^1 \left(1 - \Phi\left(\Phi^{-1}\left(F_t(y)\right) + \lambda\right)\right) dy . \tag{4-22}$$

Let M be the total annuities issued to an initial population that consists of persons aged 65 years who also are alive in 2008. Under the equivalent martingale measure *Q*, the fair value of a pay-fixed longevity swap at issue year 2009, denoted by LS_0, can be calculated as

$$LS_0 = E_Q\left[\sum_{t=1}^{T} \exp\left(-\int_0^t r(u)\,du\right) M\left(S(t) - (1+\pi)H(t)\right)\right], \tag{4-23}$$

where $r(t)$ is the risk-free rate. We also consider the term structure of the interest rate in our valuation framework. Let $B(t,T)$ denote the price of a zero-coupon bond issued at time *t* that pays \$1 at time *T*, $t \le T$. With the assumption that mortality rates and financial risk are independent, the fair value of a pay-fixed longevity swap takes the form:

$$LS_0 = M\sum_{t=1}^{T} B(0,t) E_Q\left[S(t)\right] - M(1+\pi)\sum_{t=1}^{T} B(0,t) H(t) . \tag{4-24}$$

[14] The Wang transform represents only one possible choice among several incomplete market pricing methods. For example, Biffis et al. (2010) provide the equivalent changes of measures that preserve the structure of the LC model and the tractability of the doubly stochastic setup. The specification of both a real-world and an equivalent martingale measure raises the issue of whether the doubly stochastic setting applies under the two measures. For more details, please refer to the Proposition 3.2 in Biffis et al. (2010).

The fair swap premium π, which is set when the initial value of the swap equals zero, is given by

$$\pi = \frac{\sum_{t=1}^{T} B(0,t) E_Q\left[S(t)\right]}{\sum_{t=1}^{T} B(0,t) H(t)} - 1. \tag{4-25}$$

The analytical computation of $E_Q\left[S(t)\right]$ is difficult to implement. We explain briefly the Monte Carlo algorithm to compute the expected value of the *t*-year survival probability under the equivalent martingale measure *Q* in Appendix D.

Numerical Analysis

To simulate the mortality rates, we first re-fit the RH model with four distributions—normal, JD, VG, and NIG—to the mortality data of England and Wales from 1900 to 2008 in Table 4-6. Similar to the results based on the 1900–1983 period, the best model for England and Wales is still the VG model. Consequently, we use the best prediction models presented in Table 4-5 to simulate mortality rates, which is the VG-NIG model for England and Wales.

Table 4-6. Goodness-of-fit Measures for the Number of Deaths in 1900-2008

Model	England and Wales		
	LLF	AIC	BIC
Normal	-41760	42094	43111
JD	-41616	41953	42980
VG	**-41524**	**41860**	**42883**
NIG	-41526	41862	42886

In this section, we provide a numerical example of the longevity swaps based on a cohort of 65-year-old persons in calendar year 2008. The initial term structure is obtained from the U.S. Department of the Treasury.[15] We also assume that $M = 1$.

[15] See http://www.treasury.gov/resource-center/data-chart-center/interest-rates/pages/TextView.aspx?

Figure 4-1 depicts the swap premium curve by varying the level of the risk-adjusted parameter λ. The fair swap premium is higher for a longer duration swap, because long-duration contracts are usually more expensive for covering longevity risk. The lower λ implies higher survival probabilities, so the fair swap premiums should be bigger for lower λ. In addition, the fair swap premiums of the RH model (the M2 model of Cairns et al., 2009) are higher than those of the best prediction model, which means that the fixed-rate payer (longevity risk hedger) can pay lower swap premium, according to the best prediction model.

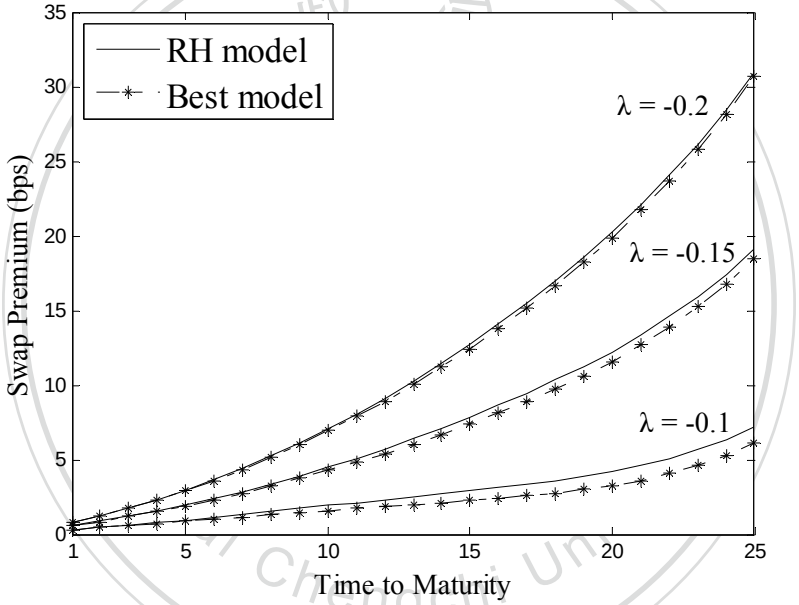


Figure 4-1. Swap Premium Curves for Distinct Level of Risk-Adjusted Parameter λ

Table 4-7 reveals the fair swap premiums with time to maturity equal to 25 years when λ is -0.1, -0.15, and -0.2, with parallel shifts upward of 0%, 2%, and 4% in the yield curve. From Table 4-7, we see that the lower the λ and the interest rate are, the higher is the fair swap premium. Similarly, the fair swap premiums of the RH

data=yieldYear& year=2008. The 1-year, 2-year, 3-year, 5-year, 7-year, 10-year, 20-year, and 30-year yield rates are 0.37%, 0.76%, 1%, 1.55%, 1.87%, 2.25%, 3.05%, and 2.69% on December 31, 2008, respectively. We use the linear interpolation to obtain other yield rates.

model are higher than those of the best prediction model, even when the yield curve moves up in parallel. Longevity risk hedgers can use the best prediction model to price longevity swaps and pay lower swap premiums.

Table 4-7. Swap Premiums for Different Interest Rates (Units: bps)

Yield Rates	Model	λ = -0.1	λ = -0.15	λ = -0.2
Original yield curve	RH	7.15	19.07	30.91
	Best	6.08	18.42	30.67
Parallel shift up of 2%	RH	6.11	16.32	26.48
	Best	5.15	15.73	26.24
Parallel shift up of 4%	RH	5.21	13.94	22.63
	Best	4.36	13.40	22.38

Note: Time to maturity is 25 years.

As market conditions change (e.g., mortality patterns, a parallel shift in yield curve), the marking-to-market (MTM) procedure could mean that the longevity swap switching status in the hedger's balance sheet falls between that of an asset and that of a liability. Assume that λ is -0.1 and the maturation time is 25 years, as in our baseline case. The initial swap premiums are 7.15 bps and 6.08 bps for the RH and best prediction models in the baseline case, respectively. In Table 4-8, applying Equation (4-24), we report the impacts of market condition changes (a parallel shift in yield curve and different risk-adjustment parameters λ) on the MTM profits or losses of the longevity swaps. When the yield curve moves up in parallel, *ceteris paribus*, the fair value of the longevity swap decreases, which means that a parallel shift up in the yield curve leads to a loss for the fixed-rate payer (hedger). In addition, a lower level of the risk-adjustment parameter results in a higher expected value of survival probability (higher mortality improvement), which in turn leads to a higher value of the longevity swap. Because the U.S. Fed reiterated its plan to keep its key short-term interest rate near zero until at least late 2014, it may be not favorable for the hedgers

(e.g., pension funds, annuity providers) to hedge their exposure to longevity risk through longevity swaps in this low interest rate environment. However, as shown in Table 4-8, the risk-adjustment parameter has a larger impact than the parallel shift up in the yield curve on the fair value of the longevity swap. Consequently, as life expectancy increases dramatically in developed countries, it is reasonable to find the recent surge in transactions in longevity swaps.

Table 4-8. The MTM Values of Longevity Swaps

Model	Yield Rates	λ		
		-0.1	-0.15	-0.2
RH	Original yield curve	0	0.0184	0.0366
	Parallel shift up of 2%	-0.0013	0.0116	0.0245
	Parallel shift up of 4%	-0.0021	0.0072	0.0164
Best	Original yield curve	0	0.0190	0.0379
	Parallel shift up of 2%	-0.0012	0.0122	0.0255
	Parallel shift up of 4%	-0.0018	0.0077	0.0172

Note: Assume that λ is -0.1 and maturation time is 25 years in the baseline case.

From the standpoint of the pay-fixed payer of a longevity swap, the unexpected loss at time t is of the form:

$$L(t) = M\left((1+\pi)H(t) - S(t)\right), \quad t = 1,...,T. \tag{4-26}$$

The present value of the total unexpected loss, denoted as PVL, is given by

$$PVL = \sum_{t=1}^{T} B(0,t)L(t). \tag{4-27}$$

Figure 4-2 depicts the pdf of PVL for the RH model and the best prediction model of England and Wales mortality data; it also marks the areas for the other three subplots in the upper left-hand panel. We find that the pdf of PVL for the best prediction model possesses leptokurticity and a high tip. In addition, Table 4-9 presents the VaR and CTE of the PVL with maturation times of up to 25 years. It is

clear that, compared with the RH model, the best prediction model has higher VaR and CTE. Because shorter-duration contracts cover less longevity risk, the VaR and CTE values are smaller for shorter duration longevity swaps. The differences of RH and the best prediction model are larger for longer durations. Therefore, the loss distribution of longevity swaps is centralized and heavy-tailed, especially for longer duration contracts. It is critical to have a good mortality model to calculate accurate loss distributions.

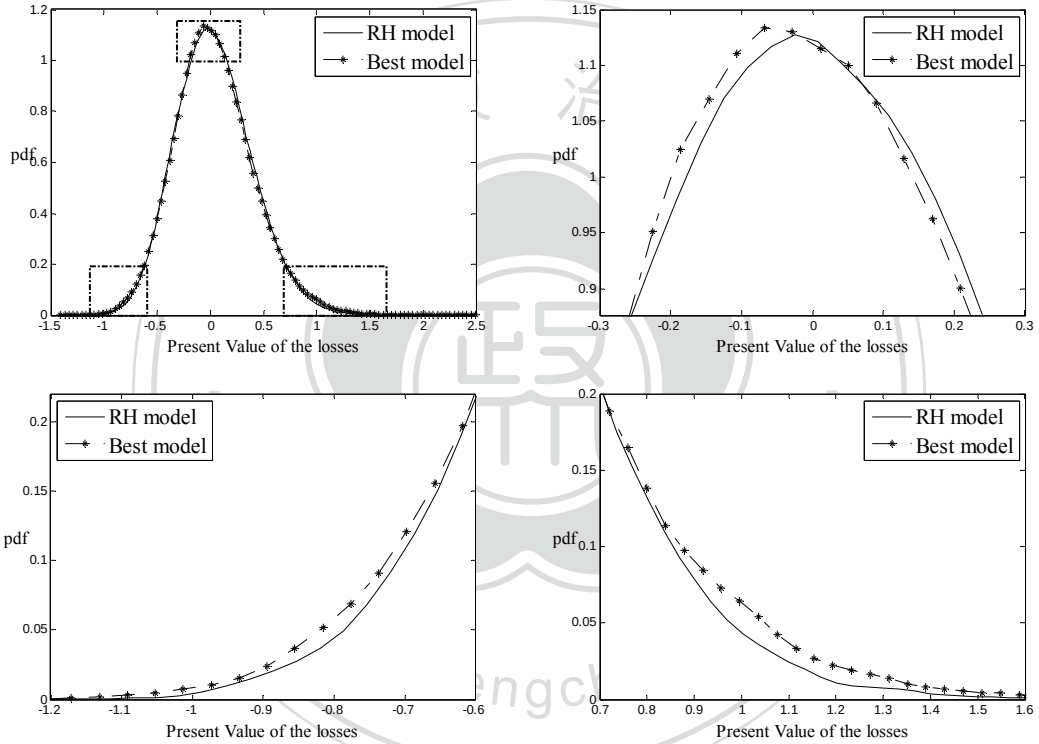


Figure 4-2. Probability Density Functions of Present Value of the Losses
($\lambda = -0.1$, $T = 25$)

Table 4-9. The VaR and CTE of the Losses for Different Maturation Times ($\lambda = -0.1$)

Time to Maturity	Model	VaR95	VaR99	CTE95	CTE99
10	RH	0.0800	0.1188	0.1047	0.1380
	Best	0.0847	0.1355	0.1171	0.1656
15	RH	0.2136	0.3061	0.2713	0.3596
	Best	0.2254	0.3452	0.3002	0.4176
20	RH	0.4137	0.5940	0.5256	0.6951
	Best	0.4355	0.6590	0.5757	0.7926
25	RH	0.6720	0.9563	0.8497	1.1165
	Best	0.7057	1.0526	0.9233	1.2528

4.4 Conclusions and Suggestions

Many researchers have examined mortality rates and explored various models. Some studies have demonstrated that improvements in the LC model occur when the model is adjusted by fitting the Poisson regression model to the number of deaths and considering an age–period–cohort extension of the LC model. Under the Poisson error structure though, intensity consists of the death rate, which is commonly modeled by stochastic mortality models. In addition, empirical results demonstrate that mortality rate improvements exhibit jump properties. We therefore attempt to provide an iterative fitting algorithm for estimating the Cox regression model, under which death rates adhere to the RH model with three heavy-tailed distributions—JD, VG, and NIG.

Using three mortality data sets from England and Wales, France, and Italy, we find consistent support for the non-Gaussian residuals of the RH model. Specifically, when we calibrate the parameters of the RH model, the VG model provides the best fit for the three countries according to the BIC criterion. For mortality projection from

the three mortality data sets, we find that the normal distribution provides weak mortality projection performance, whereas the non-Gaussian distributions provide good mortality projections. In the longevity swap application, we demonstrate that the swap curves of the original RH model are higher than those of the RH model with non-Gaussian innovations, which means that a longevity risk hedger, a fixed-rate payer of a longevity swap, can pay lower swap premium by using the RH model with non-Gaussian innovations. In addition, the VaR and CTE of the original RH model are lower than those of the RH model with non-Gaussian innovations. Choosing an appropriate leptokurtic model is critical to avoiding an underestimation of the loss reserve. As a result, when applying the RH model to long-term mortality data, the non-Gaussian distributions appear to be the most appropriate choices.

國立政治大學
National Chengchi University

Chapter 5

Pricing High-Dimensional Bermudan Options with Lévy Processes Using Low Discrepancy Mesh Methods

5.1 Multivariate Affine Generalized Hyperbolic Distributions

Let $Z=(Z_1,...,Z_n)'$ be a random vector which consists of n mutually independent random variables with univariate standard GH distributions[16] and we denote by $Z_j \sim stdGH(z;\omega_j)$, where $\omega_j := (\alpha_j, \beta_j, \lambda_j)$ Then, we can define MAGH distributions as follows.

Definition 5-1 (MAGH distribution). Let $X=(X_1,...,X_n)'$ be an n-dimensional

[16] See Chapter 2 for the definition of standard GH distributions.

MAGH distributed with mean vector $M \in R^n$ and covariance matrix $\Sigma \in \mathbb{R}^{n\times n}$ if X can be expressed as an affine transformation of a vector Z: $X = M + AZ$, where $A \in \mathbb{R}^{n\times n}$ is a lower triangular matrix such that $AA' = \Sigma$.[17] We denote by $X \sim MAGH(\omega, \Sigma, M)$, where the parameter vectors $\omega := (\omega_1, ..., \omega_n)'$.

As Schmidt et al. (2006) and Fajardo and Farias (2010) pointed out, this definition is responsible for simplifying the estimation procedure and allows us to model more leptokurtic data. Following the similar algorithm used by Schmidt et al. (2006) and Fajardo and Farias (2009, 2010), we use the following steps to estimate the parameter of MAGH distributions.

Step 1: Get $Z = B(X - M)$, where B is the inverse Cholesky factorization of the covariance matrix Σ. Then Z is a set of independent $stdGH(\omega_j)$.

Step 2: Estimate the univariate stdGH by using maximum likelihood estimation.

The procedure leads to a simplification on the parameter estimation and allows us to estimate n one-dimensional distributions, instead of the simultaneous estimation of $5n + n(n-1)/2$ parameters.

According to Definitions 5-1, using the Jacobian determinant, the density function of X adhering to an MAGH density can be represented as

$$f_X(x) = |A|^{-1} \prod_{j=1}^{n} f_{Z_j}(z_j), \tag{5-1}$$

where $x = (x_1, ..., x_n)'$ and $z = (z_1, ..., z_n)' = B(x - M)$. In addition, the characteristic function (CF) of X is given by

[17] Distinct form Schmidt et al. (2006) and Fajardo and Farias (2011) who use the univariate GH distributions with zero location and unit scaling to construct MAGH distributions, we use standard GH margins to construct MAGH distributions. As a result, this setup is easier to understand because M and Σ are directly the mean vector and covariance matrix, respectively.

$$\phi_X(\omega) = E\left(\exp\left(i\omega' X\right)\right) = \exp\left(i\omega' M\right)\prod_{k=1}^{n}\phi_{Z_k}(\psi_k), \tag{5-2}$$

where $\omega = (\omega_1,...,\omega_n)'$ and $\psi' = (\psi_1,...,\psi_n) = \omega' A$. We prove Equation (5-2) in Appendix E.

Remark 5-1. Because the variance gamma (VG) distribution is the limiting case of GH distribution, when $Z_j \sim stdVG(z;\alpha_j,\beta_j)$, we obtain $X = (X_1,...,X_n)$ follows an *n*-dimensional multivariate affine variance gamma random vector, denoted by $X \sim MAVG(\omega,\Sigma,M)$, where $\omega := (\omega_1,...,\omega_n)'$ and $\omega_j := (\alpha_j,\beta_j)$. Similarly, normal inverse Gaussian (NIG) distribution is a special case of the GH distribution with $\lambda = -0.5$. When $Z_j \sim stdNIG(z;\alpha_j,\beta_j)$, we obtain $X = (X_1,...,X_n)$ adheres to an *n*-dimensional multivariate affine normal inverse Gaussian (MANIG) random vector, denoted by $X \sim MANIG(\omega,\Sigma,M)$, where $\omega := (\omega_1,...,\omega_n)'$ and $\omega_j := (\alpha_j,\beta_j)$.

MAGH Processes for Asset Returns

In this section, we would introduce the MAGH model for the assets returns. The risky asset returns over a small time interval are defined as follows:

$$R_j(t) = \log\left(S_j(t)\right) - \log\left(S_j(t-1)\right), \quad j = 1,...,n-1, \quad t = 1,...,T \tag{5-3}$$

where $S_j(t)$ is the j^{th} asset price at time t. The return on the risk-free asset over the same time interval equals r_f. Following the work of Fajardo and Farias (2010), we construct the assets returns using the MAGH distributions; that is

$$R(t) = \begin{bmatrix} R_1(t) \\ \vdots \\ R_n(t) \end{bmatrix} = \begin{bmatrix} m_1 \\ \vdots \\ m_n \end{bmatrix} + \begin{bmatrix} a_{11} & 0 & \cdots & 0 \\ a_{21} & a_{22} & \cdots & \vdots \\ \vdots & \vdots & \ddots & 0 \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{bmatrix} \begin{bmatrix} Z_1 \\ \vdots \\ Z_n \end{bmatrix} = M + AZ, \tag{5-4}$$

where m_j is mean of the asset return $R_j(t)$ and A is a lower triangular matrix such that the covariance matrix Σ is equal to AA'. According to Equation (5-2), the CF of the logarithm of assets returns is of the form:

$$\phi_R(\omega) = E\left(\exp\left(i\omega' R(t)\right)\right) = \exp\left(i\omega' M\right)\prod_{k=1}^{n}\phi_{Z_k}(\psi_k), \tag{5-5}$$

For Derivative pricing, we have to construct a risk-neutral measure to ensure that there is no arbitrage opportunities in the market described by the model (see Harrison and Kreps, 1979; Harrison and Pliska, 1981, 1983). Gerber and Shiu (1994) first employ the use of the Esscher transform for option valuation in an incomplete market. In this line, Fajardo and Mordecki (2006) and Fajardo and Farias (2010) extend one-dimensional Esscher transform to multidimensional derivative pricing[18].

Let $(\eta_t)_{t=0}^{T}$ is a stochastic process defined as

$$\eta_T = \prod_{t=1}^{T}\xi_t, \tag{5-6}$$

where

$$\xi_t = \frac{\exp\left(\theta' R(t)\right)}{E_P\left(\exp\left(\theta' R(t)\right)\right)}, \tag{5-7}$$

and $\theta = (\theta_1, ..., \theta_n)'$. It is then straightforward to verify that $E_P(\eta_T) = 1$ and $E_P\left(\eta_T \middle| F_t\right) = \eta_t$, where P is the physical probability measure (i.e., real-world probability measure). Or equivalently, $(\eta_t)_{t=0}^{T}$ is a martingale under P. Define a new martingale measure Q_θ by

[18] Distinct from Fajardo and Farias (2010), we provide explicitly a procedure to obtain the Esscher transform parameters as well as the corresponding MAGH parameter setup under the risk-neutral measure.

$$\left.\frac{dQ_\theta}{dP}\right|_{F_T} = \eta_T = \prod_{t=1}^{T} \xi_t \,. \tag{5-8}$$

We provide the multidimensional Esscher transform for MAGH in Proposition 5-1.

Proposition 5-1. *Under the martingale measure Q_θ, the distribution of Z_j adheres to a $GH(\alpha_j, \beta_j + \varphi_j^\theta, \lambda_j, \delta_j, \mu_j)$ law, where $\varphi_\theta{}' = \left(\varphi_1^\theta, ..., \varphi_n^\theta\right) = \theta' A$. In addition, to ensure that the discounted stock price process is a Q_θ-martingale, the Esscher transform parameter θ_j is chosen to satisfy the following equation:*

$$r_f = m_j + \sum_{k=1}^{j} \left(\log\left(\phi_{Z_k}\left(-i(\varphi_k^\theta + a_{jk}) \right) \right) - \log\left(\phi_{Z_k}\left(-i\varphi_k^\theta \right) \right) \right). \tag{5-9}$$

The detailed proof is shown in Appendix F.

5.2 Low Discrepancy Mesh (LDM) Method

The LDM method of Boyle et al. (2003) is similar to the MCM method of Broadie and Glasserman (2004) except it exploits explicitly the greater uniformity of the low discrepancy sequence for generating the mesh points, but the MCM method generates the mesh points by crude Monte Carlo points. Boyle et al. (2003) demonstrated the power of the LDM method for pricing high-dimensional American style options when the underlying asset prices follow multivariate lognormal processes. In this paper, we show that the LDM method can also be extended to the MAVG and MANIG distributions.

Let $S(t) = \left(S_1(t), ..., S_n(t)\right)$ be a vector-valued process which denotes the prices of the n underlying asset at time t with fixed initial price $S(0)$. The payoff of derivative security depends on these underlying asset prices. We assume that

derivative security can be exercised at one of the d possible exercise time points $0 = t_0 < t_1 < ... < t_d = T$, excluding initial time. From the standard option pricing theory, the value of the Bermudan option V can be formulated as an optimization problem of the form (see Duffie, 1996)

$$V \equiv \max_{\tau} E_{Q_\theta}\left[h\left(\tau, S(\tau)\right)\right], \tag{5-10}$$

where τ is a stopping time taking values in the set $\{t_1,...,T\}$, and $h(t,x) \geq 0$ is the payoff at time t in state x if the option is exercised. It is well known that the above optimization problem can be solved using the principle of dynamic programming. By defining $V(t,x)$ as the value of the Bermudan option at time t in state x, the dynamic programming can be described as follows:

Step 1: Starting from the maturity T of the option and set $V(T,x) = h(T,x)$.

Step 2: For $i = d-1,...,1,$ recursively calculate

$$V(t_i,x) = \max\left(h(t_i,x), B(t_i,t_{i+1})E_{Q_\theta}\left[V\left(t_{i+1}, S(t_{i+1})\right)\middle| R(t_{i+1}) = x\right]\right), \quad i = d-1,...,1, \tag{5-11}$$

where $B(t_i,t_{i+1})$ is the discounting factor from t_i to t_j, $i,j = 0,...,d$ and $i < j$. Here $B(t_i,t_{i+1})E_{Q_\theta}\left[V\left(t_{i+1}, S(t_{i+1})\right)\middle| R(t_{i+1}) = x\right]$ can be interpreted as the continuation value of the option if it is kept alive. Hence, $V(t,x)$ corresponds to the maximum of the exercise value and the continuation value.

Step 3: For $i = 0$, calculate

$$V\left(0, S(0)\right) = B\left(0,t_1\right)E_{Q_\theta}\left[V\left(t_1, S(t_1)\right)\middle| R(0) = \ln S(0)\right], \tag{5-12}$$

the time-0 value of the Bermudan option is given by $V = V\left(0, S(0)\right)$.

The difficulty in the above backward recursive algorithm is to provide an

effective algorithm to estimate the continuation value; i.e. the conditional expectations

$$E_{Q_\theta}\left[V\left(t_{i+1}, S(t_{i+1})\right)\middle|R(t_i)=x\right] \tag{5-13}$$

The essence of the MCM method or the LDM method is provide an effective approach of approximating (5-13) using only the N mesh points $R_{t_i}(j)$, $j=1,...,N$, at each time point.

To explain this, it is convenient to first define the transition probability of underlying asset as

$$\text{Prob}_{Q_\theta}\left(S(t_{i+1})\in A\middle|R(t_i)=R_{t_i}(j)\right)=\int_A f\left(t_i, R_{t_i}(j), t_{i+1}; u\right)du\,, \tag{5-14}$$

where $R_{t_i}(j)$ is the j^{th} node at time t_i; $f\left(t_i, R_{t_i}(j), t_{i+1}; u\right)$ represents the transition probability from node $R_{t_i}(j)$ at time t_i to node u at time t_{i+1}. By construction, $R_{t_0}(\cdot)$ corresponds to the logarithm of initial underlying asset price $\ln S(0)$. Then, as shown by Broadie and Glasserman (2004), the conditional expectations for all j and $i=d-1,...,1$ is

$$\begin{aligned}
E_{Q_\theta}\left[V\left(t_{i+1}, S(t_{i+1})\right)\middle|R(t_i)=R_{t_i}(j)\right] &= \int_{\mathbb{R}^n}\hat{V}\left(t_{i+1},u\right)f\left(t_i, R_{t_i}(j), t_{i+1}; u\right)du \\
&= \int_{\mathbb{R}^n}\hat{V}\left(t_{i+1},u\right)\frac{f\left(t_i, R_{t_i}(j), t_{i+1}; u\right)}{g\left(t_{i+1}; u\right)}g\left(t_{i+1}; u\right)du \\
&= \frac{1}{N}\sum_{k=1}^{N}\hat{V}\left(t_{i+1}, X_{t_i}(k)\right)\frac{f\left(t_i, R_{t_i}(j), t_{i+1}; R_{t_{i+1}}(k)\right)}{g\left(t_{i+1}; R_{t_{i+1}}(k)\right)}\,,
\end{aligned} \tag{5-15}$$

where $f\left(t_i, R_{t_i}(j), t_{i+1}; R_{t_{i+1}}(k)\right)\Big/g\left(t_{i+1}; R_{t_{i+1}}(k)\right)$ is the Radon-Nikodym derivatives. By the simple change of measure approach, the above result indicates that if the mesh points were simulated from an arbitrary $g\left(t_{i+1};\cdot\right)$ instead of $f\left(t_i, R_{t_i}(j), t_{i+1};\cdot\right)$, then the conditional expectation can be approximated by (5-15).

What remains is the specification of the mesh density $g(t_{i+1};\cdot)$. As argued in Boyle et al. (2003), a reasonable choice of the mesh density is the marginal density. In our case, we similarly use the marginal density. We simulate mesh points at each time step using a simpler method; i.e. choose marginal distribution as $g(t_{i+1};\cdot)$, the first and second moments of which match those of $f(0, R(0), t_{i+1};\cdot)$ but the distribution of Z_j, j=1,...,n, remain the same. In addition, in order to implement the framework easily, we simulate the underlying prices by the view of logarithm underlying returns. Because we assume the underlying returns follow MAVG or MANIG distribution, instead of underlying price.

To summarize, the LDM method can be implemented using the following two steps:

Step 1 (Generate mesh points):

At time t_1, we use low discrepancy sequence and inverse method of cumulative density function to simulate the state variables $R_{t_1}(1),...,R_{t_1}(N)$ from MAVG or MANIG distributions with the density, as follows,

$$R(t_1) = M + AZ\,. \tag{5-16}$$

At time t_i, $i>1$, we use low discrepancy sequence and inverse method of cumulative density function to simulate these state variables $R_{t_i}(1),...,R_{t_i}(N)$ from MAVG or MANIG distributions with the density, as follows,

$$R(t_{i+1}) = M\,t_{i+1} + A\sqrt{t_{i+1}}Z\,. \tag{5-17}$$

Step 2 (Backward recursion):

We calculate the Equation (5-15) according to transition probabilities and the

underlying payoffs. Then, we can obtain the value of Bermudan option following the backward recursion from Equation (5-11).

5.3 Empirical and Numerical Analyses

The Data Description

In this paper, we first calibrate the MAVG and MANIG parameters of the daily stock returns of Apple (AAPL), Yahoo (YHOO) and Google (GOOG) for pricing Bermudan options. Our sample period runs from August 1st, 2007 to July 31th, 2012, with data from the Yahoo Finance. Table 5-1 presents the descriptive statistics for the raw data.

Table 5-1. Descriptive Statistics

	Apple	Yahoo	Google
Mean	0.1198%	-0.0305%	0.0167%
Median	0.1529%	0.0000%	0.0414%
Maximum	13.0194%	39.1817%	18.2251%
Minimum	-19.7470%	-23.4025%	-12.3402%
Std. Dev.	2.4052%	2.9359%	2.2229%
Skewness	-0.4244	1.3125	0.3782
Excess Kurtosis	6.0008	30.0603	8.3703
JB test	1928.32	47801.91	3708.28
Observation	1260	1260	1260

The descriptive statistics present that all risky assets we used exhibit non-zero skewness and positive excess kurtosis. Based on the Jarque–Bera (JB) test statistics, the null hypothesis is significantly rejected, which means that the empirical distribution of the return series do not follow the normality assumption.

Parameter Calibration

We first obtain the sample mean vector $\hat{M}$ and sample covariance matrix $\hat{\Sigma}$.

Based on the sample covariance matrix, we can obtain its inverse Cholesky factorization, together with three residual series. We then estimate the corresponding univariate stdVG and stdNIG distributions by using maximum likelihood estimation. The results are presented in Table 5-2.

Table 5-2. Estimated Parameters for MAVG and MANIG

Model	Name	α	β	$\hat{M}$		$\hat{\Sigma}$		Loglikehood
	Apple	24.8844	-0.8460	0.3019	0.1458	0.0731	0.0862	-1693.17
MAVG	Yahoo	20.8623	0.6451	-0.0768	0.0731	0.2172	0.0652	-1553.69
	Google	20.8135	-0.8580	0.0421	0.0862	0.0652	0.1245	-1582.33
	Apple	13.5471	-0.9611	0.3019	0.1458	0.0731	0.0862	-1688.76
MANIG	Yahoo	9.6213	0.2875	-0.0768	0.0731	0.2172	0.0652	-1528.79
	Google	9.6980	-0.3460	0.0421	0.0862	0.0652	0.1245	-1559.01

For derivative pricing, we calculate φ_θ and the Esscher parameter θ. As we mention in proposition 5-1, we solve numerically the Equation (5-9) to obtain φ_θ as well as the Esscher parameter θ. The estimation results, when the risk-free interest rate is equal to 2%, are in Table 5-3.

Table 5-3. φ_θ and Esscher Parameters θ for MAVG and MANIG

Model	Name	φ_θ	θ
	Apple	-0.9262	-3.6788
MAVG	Yahoo	0.3892	0.7343
	Google	0.3980	1.4952
	Apple	-0.9236	-3.6713
MANIG	Yahoo	0.3887	0.7331
	Google	0.3979	1.4950

Note: The risk-free interest rate is 2%.

Numerical Results

In this section we provide some numerical evidence on the effectiveness of the LDM method. We consider the following Bermudan options:

- Put option on single asset: $\max\left(K - S_t, 0\right)$
- Put option on the maximum of two assets: $\max\left(K - \max\left(S_t^1, S_t^2\right), 0\right)$
- Put option on maximum of three assets: $\max\left(K - \max\left(S_t^1, S_t^2, S_t^3\right), 0\right)$

We further assume that $K = \$100$ and $r = 2\%$, these options have 4 and 8 exercisable opportunities and that the underlying asset prices follow either MAVG or MANIG. For each option, LDM method and MCM method are implemented with mesh points $N = \{512, 1024, 2048, 4096\}$ and using the mesh density as described in Equation (5-16) and (5-17). For the LDM method, the scrambled Sobol' sequence is used to generate the necessary mesh points and this procedure is replicated 30 times in order to provide an estimate of the standard error of the LDM estimate. The same procedure is applied to the MCM method except that random sequences are used. The results are reported in Tables 5-4 ~ 5-9.

When the number of asset is equal to one, the MAVG or MANIG processes reduce to the VG or NIG processes. Let us first focus on Tables 5-4 and 5-5. Because the contingent claim in this case is a standard put option which depends only on a single asset, their option prices can be approximated to a high degree of precision using the multinomial tree (MT), proposed by Këllezi and Webber (2004), even though the underlying asset follows a complicated process such as VG or NIG. In what follow, we assume that the Bermudan put option prices obtained from the MT with 4096 nodes are the "correct" prices so that these prices are used as benchmark against the simulated mesh estimates. Based on these results, we draw the following remarks:

1. In agreement with Boyle et al. (2003), both the mesh estimates of LDM and MCM are high biased. In addition, the mesh estimates decline as we increase the

mesh points.

2. The bias of the LDM method is remarkably small, assuming the prices from the MT is the correct value. Consequently, the LDM method is extremely effective. For example, when $N = 4096$, the maximum relative error is no more than 0.1% for both 4 and 8 exercises points and under both VG and NIG processes.
3. The MCM method, on the other hand, is clearly inferior, as signified by the much larger standard errors and relative errors. More specifically, consider the case with $S0 = 90$ where the MT yields an option price of 12.9839 for the NIG process. The LDM method with 4096 mesh points gives the closer estimate with standard error of 0.0002. The error relative to the MT is therefore 0.0054%. In contrast, the corresponding MCM estimate is 13.1974 with standard error 0.0552. This leads to a relative error of 1.6443% and comparing to the LDM method, the standard error of the MCM method is about 305 times larger.

Let us consider the results in the remaining table. The options in these cases depend on more than one asset and this brings out the weakness of the MT. The MT becomes computational inefficient when there is more than one asset. The LDM method, on the other hand, is more flexible in that it can easily be accommodated even when there is more than one underlying asset. This is the advantage of the LDM method.

By considering option that depends on more than one asset also poses additional challenge on assessing the relative efficiency of the mesh estimates due to the lack of appropriate benchmark. Here we can only compare to LSM estimates of Longstaff and Schwartz (2001) based on 1,000,000 simulated paths and 30 trials. It should be emphasized that the mesh estimate and the least square estimate may not be directly comparable for at least the following two reasons. One is that the mesh estimate is

high bias while the least square estimate is low bias. The other is that for a given number of sample points, the magnitude of bias is not known and hence their accuracies can be difficult to gauge. On the other hand, the numerical results in the case of single asset examples are encouraging. These results indicate that the bias of the LDM method is extremely insignificant. This provides some confidence on the reliability of the LDM method and thus its accuracy in high-dimensional applications. The extensive numerical examples conducted by Boyle et al. (2003) also supported this even though the underlying state variables in their examples are multivariate normally distributed.

From the results in Tables 5-6 ~ 5-9, we draw the following remarks:

1. The mesh estimates are higher than the corresponding least square estimates. This is consistent with the property that the mesh estimate is high bias while the least square estimate is low bias. This property also ensures that the true value will lie between these two bounds.
2. Even with more complicated options with payoffs depend on the maximum of multiple assets, the LDM method consistently outperforms MCM method, as signified by the much smaller standard errors under both MAVG and MANIG and $d = 4$ and 8. This suggests that LDM method can be effective for pricing high-dimensional Bermudan option even though the underlying asset follows a more complicated multivariate processes.

In conclusion, the above numerical examples have demonstrated the relative efficiency of the LDM method. This method not only has competitive advantage for pricing high-dimensional Bermudan option, the method is flexible to accommodate any complicated stochastic process. This is a clear advantage of the LDM method as many of the existing numerical methods will breakdown when we move from

univariate to multivariate case. An example is the MT which is proven to be effective as long as the option depends on a single asset. When the option depends on more than one asset, the MT is no longer feasible.

Table 5-4. Put Option on Single Asset (4 exercise points)

Underlying asset is assumed to follow VG process

S0	Multinomial Tree	Nodes	Low Discrepancy Mesh			Monte Carlo Mesh		
			Estimate	Std. err.	Rel. err.	Estimate	Std. err.	Rel. err.
90	12.9793	512	13.0027	0.0036	0.1809%	13.9341	0.1598	7.3567%
		1024	12.9892	0.0026	0.0763%	13.4425	0.1065	3.5687%
		2048	12.9824	0.0006	0.0242%	13.2185	0.0646	1.8432%
		4096	12.9800	0.0002	0.0059%	13.1895	0.0550	1.6196%
100	7.3786	512	7.4044	0.0035	0.3496%	8.0490	0.1405	9.0859%
		1024	7.3901	0.0024	0.1548%	7.6847	0.0888	4.1480%
		2048	7.3825	0.0006	0.0520%	7.5256	0.0570	1.9925%
		4096	7.3799	0.0002	0.0168%	7.5293	0.0446	2.0421%
110	3.8334	512	3.8562	0.0033	0.5957%	4.2943	0.1207	12.0226%
		1024	3.8443	0.0023	0.2833%	4.0474	0.0750	5.5831%
		2048	3.8371	0.0005	0.0978%	3.9202	0.0471	2.2644%
		4096	3.8349	0.0002	0.0390%	3.9323	0.0330	2.5794%

Underlying asset is assumed to follow NIG process

S0	Multinomial Tree	Nodes	Low Discrepancy Mesh			Monte Carlo Mesh		
			Estimate	Std. err.	Rel. err.	Estimate	Std. err.	Rel. err.
90	12.9839	512	13.0060	0.0039	0.1702%	13.9336	0.1559	7.3149%
		1024	12.9918	0.0019	0.0612%	13.4433	0.1056	3.5387%
		2048	12.9862	0.0004	0.0179%	13.2217	0.0646	1.8320%
		4096	12.9846	0.0002	0.0054%	13.1974	0.0552	1.6443%
100	7.3991	512	7.4238	0.0038	0.3328%	8.0677	0.1369	9.0349%
		1024	7.4087	0.0018	0.1291%	7.7008	0.0886	4.0772%
		2048	7.4022	0.0005	0.0408%	7.5458	0.0566	1.9822%
		4096	7.4003	0.0002	0.0152%	7.5536	0.0449	2.0874%
110	3.8684	512	3.8907	0.0037	0.5777%	4.3283	0.1174	11.8910%
		1024	3.8779	0.0018	0.2460%	4.0798	0.0744	5.4674%
		2048	3.8714	0.0005	0.0799%	3.9549	0.0463	2.2367%
		4096	3.8697	0.0003	0.0347%	3.9709	0.0332	2.6506%

Table 5-5. Put Option on Single Asset (8 exercise points)

Underlying asset is assumed to follow VG process

S0	Multinomial Tree	Nodes	Low Discrepancy Mesh			Monte Carlo Mesh		
			Estimate	Std. err.	Rel. err.	Estimate	Std. err.	Rel. err.
90	12.9901	512	13.0591	0.0094	0.5306%	14.6584	0.2761	12.8425%
		1024	13.0132	0.0025	0.1776%	13.9463	0.1507	7.3606%
		2048	12.9982	0.0011	0.0621%	13.6739	0.1132	5.2640%
		4096	12.9927	0.0007	0.0201%	13.3532	0.0820	2.7948%
100	7.3850	512	7.4521	0.0089	0.9080%	8.5045	0.2280	15.1593%
		1024	7.4104	0.0025	0.3433%	7.9991	0.1297	8.3154%
		2048	7.3949	0.0011	0.1334%	7.8607	0.0880	6.4417%
		4096	7.3884	0.0007	0.0458%	7.6017	0.0707	2.9336%
110	3.8366	512	3.8962	0.0082	1.5543%	4.5906	0.1856	19.6533%
		1024	3.8608	0.0023	0.6325%	4.2262	0.1072	10.1563%
		2048	3.8465	0.0010	0.2586%	4.1448	0.0635	8.0331%
		4096	3.8402	0.0007	0.0945%	3.9660	0.0549	3.3728%

Underlying asset is assumed to follow NIG process

S0	Multinomial Tree	Nodes	Low Discrepancy Mesh			Monte Carlo Mesh		
			Estimate	Std. err.	Rel. err.	Estimate	Std. err.	Rel. err.
90	12.9946	512	13.0536	0.0079	0.4544%	14.6545	0.2652	12.7741%
		1024	13.0125	0.0024	0.1381%	13.9448	0.1483	7.3126%
		2048	13.0003	0.0009	0.0438%	13.6914	0.1141	5.3626%
		4096	12.9959	0.0003	0.0105%	13.3628	0.0813	2.8336%
100	7.4054	512	7.4637	0.0075	0.7878%	8.5156	0.2191	14.9916%
		1024	7.4256	0.0024	0.2727%	8.0120	0.1264	8.1917%
		2048	7.4127	0.0009	0.0980%	7.8919	0.0891	6.5695%
		4096	7.4073	0.0004	0.0254%	7.6259	0.0704	2.9780%
110	3.8715	512	3.9242	0.0071	1.3613%	4.6131	0.1771	19.1551%
		1024	3.8913	0.0022	0.5118%	4.2532	0.1025	9.8605%
		2048	3.8789	0.0009	0.1914%	4.1878	0.0644	8.1712%
		4096	3.8736	0.0004	0.0536%	4.0023	0.0545	3.3796%

Table 5-6. Put Option on the Maximum of Two Assets (4 exercise points)

Underlying assets follow MAVG process

S0	LSM	Std. err	Nodes	Low Discrepancy Mesh		Monte Carlo Mesh	
				Estimate	Std. err.	Estimate	Std. err.
90	9.3129	0.0013	512	9.4167	0.0194	9.7923	0.0977
			1024	9.3639	0.0065	9.5016	0.0459
			2048	9.3449	0.0047	9.4264	0.0345
			4096	9.3366	0.0035	9.3724	0.0240
100	4.3920	0.0010	512	4.4534	0.0149	4.7246	0.0836
			1024	4.4276	0.0062	4.5164	0.0378
			2048	4.4087	0.0027	4.4558	0.0299
			4096	4.4031	0.0021	4.4219	0.0217
110	1.8642	0.0007	512	1.8942	0.0135	2.0841	0.0729
			1024	1.8865	0.0068	1.9539	0.0277
			2048	1.8750	0.0029	1.9007	0.0218
			4096	1.8691	0.0023	1.8832	0.0157

Underlying assets follow MANIG process

S0	LSM	Std. err	Nodes	Low Discrepancy Mesh		Monte Carlo Mesh	
				Estimate	Std. err.	Estimate	Std. err.
90	9.3238	0.0013	512	9.4351	0.0204	9.8055	0.0939
			1024	9.3781	0.0070	9.5169	0.0458
			2048	9.3578	0.0050	9.4427	0.0343
			4096	9.3488	0.0037	9.3878	0.0240
100	4.4069	0.0010	512	4.4745	0.0158	4.7432	0.0809
			1024	4.4460	0.0066	4.5334	0.0384
			2048	4.4250	0.0028	4.4734	0.0299
			4096	4.4190	0.0022	4.4400	0.0218
110	1.8831	0.0007	512	1.9169	0.0141	2.1055	0.0708
			1024	1.9084	0.0066	1.9718	0.0281
			2048	1.8951	0.0028	1.9194	0.0215
			4096	1.8884	0.0024	1.9034	0.0158

Table 5-7. Put Option on the Maximum of Two Assets (8 exercise points)

Underlying assets follow MAVG process

S0	LSM	Std. err	Nodes	Low Discrepancy Mesh		Monte Carlo Mesh	
				Estimate	Std. err.	Estimate	Std. err.
90	9.6176	0.0014	512	10.4165	0.1556	10.6105	0.1295
			1024	9.9317	0.0732	10.2371	0.0699
			2048	9.7773	0.0235	9.9601	0.0486
			4096	9.7176	0.0148	9.8262	0.0335
100	4.5026	0.0012	512	5.1152	0.1290	5.1846	0.0988
			1024	4.7630	0.0535	4.9176	0.0648
			2048	4.6309	0.0200	4.7540	0.0499
			4096	4.5838	0.0138	4.6485	0.0349
110	1.9141	0.0007	512	2.1548	0.0507	2.2997	0.0631
			1024	2.1103	0.0349	2.1608	0.0502
			2048	2.0139	0.0187	2.0740	0.0360
			4096	1.9784	0.0137	2.0100	0.0267

Underlying assets follow MANIG process

S0	LSM	Std. err	Nodes	Low Discrepancy Mesh		Monte Carlo Mesh	
				Estimate	Std. err.	Estimate	Std. err.
90	9.6218	0.0014	512	10.4384	0.1324	10.6749	0.1329
			1024	9.9446	0.0773	10.2744	0.0693
			2048	9.7965	0.0256	9.9886	0.0509
			4096	9.7300	0.0159	9.8367	0.0339
100	4.5139	0.0012	512	5.1409	0.1059	5.2386	0.1021
			1024	4.7821	0.0554	4.9467	0.0630
			2048	4.6565	0.0221	4.7872	0.0520
			4096	4.6028	0.0147	4.6651	0.0351
110	1.9314	0.0008	512	2.1818	0.0502	2.3424	0.0672
			1024	2.1316	0.0360	2.1831	0.0477
			2048	2.0403	0.0208	2.1059	0.0376
			4096	2.0000	0.0143	2.0298	0.0264

Table 5-8. Put Option on the Maximum of Three Assets (4 exercise points)

Underlying assets follow MAVG process

S0	LSM	Std. err	Nodes	Low Discrepancy Mesh		Monte Carlo Mesh	
				Estimate	Std. err.	Estimate	Std. err.
90	7.6986	0.0010	512	7.9565	0.0400	8.1163	0.0738
			1024	7.8522	0.0181	8.0086	0.0642
			2048	7.7897	0.0119	7.7956	0.0277
			4096	7.7503	0.0044	7.7705	0.0183
100	3.0933	0.0008	512	3.2602	0.0312	3.3545	0.0594
			1024	3.1867	0.0140	3.3000	0.0527
			2048	3.1645	0.0129	3.2074	0.0248
			4096	3.1232	0.0044	3.1426	0.0143
110	1.0887	0.0007	512	1.1854	0.0233	1.2236	0.0444
			1024	1.1495	0.0106	1.2025	0.0422
			2048	1.1245	0.0086	1.1521	0.0167
			4096	1.1065	0.0037	1.1099	0.0101

Underlying assets follow MANIG process

S0	LSM	Std. err	Nodes	Low Discrepancy Mesh		Monte Carlo Mesh	
				Estimate	Std. err.	Estimate	Std. err.
90	7.7247	0.0010	512	7.9900	0.0396	8.1534	0.0724
			1024	7.8802	0.0186	8.0289	0.0552
			2048	7.8186	0.0123	7.8286	0.0279
			4096	7.7776	0.0044	7.8009	0.0182
100	3.1145	0.0009	512	3.2852	0.0309	3.3844	0.0588
			1024	3.2083	0.0139	3.3141	0.0433
			2048	3.1872	0.0123	3.2338	0.0256
			4096	3.1446	0.0044	3.1674	0.0144
110	1.1063	0.0007	512	1.2056	0.0232	1.2465	0.0442
			1024	1.1652	0.0094	1.2110	0.0317
			2048	1.1431	0.0082	1.1725	0.0175
			4096	1.1234	0.0036	1.1300	0.0103

Table 5-9. Put Option on the Maximum of Three Assets (8 exercise points)

Underlying assets follow MAVG process

S0	LSM	Std. err	Nodes	Low Discrepancy Mesh		Monte Carlo Mesh	
				Estimate	Std. err.	Estimate	Std. err.
90	8.1648	0.0012	512	9.2977	0.0946	9.5514	0.1752
			1024	8.8570	0.0781	8.8936	0.0669
			2048	8.5516	0.0356	8.6437	0.0462
			4096	8.3509	0.0130	8.4415	0.0336
100	3.2085	0.0011	512	3.8896	0.0614	4.0688	0.1175
			1024	3.7233	0.0446	4.0192	0.1156
			2048	3.5160	0.0281	3.5742	0.0411
			4096	3.3758	0.0125	3.4239	0.0351
110	1.1295	0.0007	512	1.4140	0.0435	1.4878	0.0614
			1024	1.3549	0.0368	1.3544	0.0397
			2048	1.2934	0.0250	1.3177	0.0278
			4096	1.2265	0.0100	1.2564	0.0248

Underlying assets follow MANIG process

S0	LSM	Std. err	Nodes	Low Discrepancy Mesh		Monte Carlo Mesh	
				Estimate	Std. err.	Estimate	Std. err.
90	8.1899	0.0012	512	9.4202	0.0998	9.6708	0.1776
			1024	8.9340	0.0894	8.9673	0.0679
			2048	8.6019	0.0383	8.7218	0.0527
			4096	8.3893	0.0133	8.4782	0.0328
100	3.2271	0.0011	512	3.9609	0.0622	4.1446	0.1199
			1024	3.7211	0.0519	3.7251	0.0552
			2048	3.5530	0.0288	3.6342	0.0459
			4096	3.4068	0.0137	3.4517	0.0345
110	1.1461	0.0007	512	1.4514	0.0433	1.5384	0.0668
			1024	1.3770	0.0354	1.3844	0.0379
			2048	1.3149	0.0236	1.3584	0.0318
			4096	1.2500	0.0108	1.2782	0.0241

5.4 Conclusions

The current challenge to price high dimensional Bermudan derivatives is to obtain a procedure that is computationally efficient. In addition, due to the styled facts such as nonnegative skewness and excess kurtosis, large negative returns have been demonstrated to occur more frequently than predicted under the assumption of normality. Consequently, an inappropriate asset model may result in a serious mispricing of derivatives. In this paper we attempt to handle these two problems simultaneously by incorporating the MAGH models of Schmidt et al. (2006) and the LDM method of Boyle et al. (2003) to price the multidimensional Bermudan derivatives. For derivative pricing, we also derive the MAGH parameter setup under the risk-neutral measure. From the numerical results, we demonstrate that our method is computationally efficient and easy to implement. Furthermore, for pricing multidimensional Bermudan derivatives, the LDM estimates are high bias which the LSM estimates are low bias, this property also ensures that the true value will lie between these two bounds.

It is a clear advantage of the LDM method as many of the existing numerical methods will breakdown when we move from univariate to multivariate case for pricing high-dimensional Bermudan options. Consequently, because the LDM method is flexible to accommodate any complicated stochastic process with a stochastic interest rate, it will be a topic for future research. In addition, in this paper we focuses directly on the method itself, we did not include any types of variance reduction techniques. However our approach could also be enhanced by including variance reduction techniques.

國立政治大學
National Chengchi University

Chapter 6

Conclusions

This thesis offers three applications with heavy-tailed distributions. In Chapter 3, we attempt to incorporate five heavy-tailed distributions—*t*, JD, VG, NIG, and GHST—into the Lee-Carter model. Using mortality data from six countries, the JD-JD model[19] is the best one for French mortality data, the NIG-NIG model is best for the Netherlands, the VG-*t* model offers the best goodness of fit for Swedish mortality data, the *t*-*t* model is best for the U.S. mortality data, and the NIG-*t* model is the best one for the mortality data from Finland and Switzerland. For forecasting mortality rates, *t* and its skew extension provide good mortality projections.

In Chapter 4, we refine the model, proposed by Renshaw and Haberman (2006), with heavy-tailed distributions. Under the Poisson error structure, however, the intensity is composed of the death rate, which is commonly modeled by stochastic mortality models. We attempt to provide an iterative fitting algorithm for estimating

[19] The terminology "X-Y model" refers to the error terms in Equations (3-1) and (3-2), respectively.

the Cox regression model under which death rates adhere to the RH model with three heavy-tailed distributions—JD, VG and NIG. Using three mortality datasets from England and Wales, France and Italy, we find consistent support for the non-Gaussian residuals of the RH model. Specifically, when we calibrate the parameters of the RH model, the VG model provides the best fit for the three countries according to the BIC criterion. For mortality projection from the three mortality datasets, we find that the non-Gaussian distributions provide good mortality projections. In the longevity swap application, we demonstrate that the swap curves of the original RH model are higher than those of the RH model with non-Gaussian innovations, which means that a longevity risk hedger, a fixed-rate payer of a longevity swap, can pay lower swap premium by using the RH model with non-Gaussian innovations. In addition, the VaR and CTE of the original RH model are lower than those of the RH model with non-Gaussian innovations. Choosing an appropriate leptokurtic model is critical to avoiding an underestimation of the loss reserve. Therefore, for applications of the Lee-Carter model and RH model, the heavy-tailed distributions appear to be the most appropriate choices for modeling long-term mortality data.

In Chapter 5, we attempt to handle these two problems simultaneously by incorporating the MAGH models of Schmidt et al. (2006) and the LDM method of Boyle et al. (2003) to price the multidimensional Bermudan derivatives. For derivative pricing, we also derive the MAGH parameter setup under the risk-neutral measure. From the numerical results, we demonstrate that our method is computationally efficient and easy to implement. Furthermore, for pricing multidimensional Bermudan derivatives, the LDM estimates are high bias which the LSM estimates are low bias, this property also ensures that the true value will lie between these two bounds. It is a clear advantage of the LDM method as many of the

existing numerical methods will breakdown when we move from univariate to multivariate case for pricing high-dimensional Bermudan options. Consequently, because the LDM method is flexible to accommodate any complicated stochastic process with a stochastic interest rate, it will be a topic for future research.

國立政治大學
National Chengchi University

Appendix A

We introduce another popular representation of the VG distribution and its standardization in Appendix A. According to the following relationships of parameters:

$$\alpha = \frac{\sqrt{\frac{2\sigma^2}{\kappa} + \theta^2}}{\sigma^2}, \tag{A-1}$$

$$\beta = \frac{\theta}{\sigma^2}, \tag{A-2}$$

$$\lambda = \frac{1}{\kappa}, \tag{A-3}$$

we can rewrite the Equation (2-11) as follows:

$$f_{VG}\left(x\middle|\kappa,\theta,\sigma,\mu\right) = \frac{\sqrt{2}\left(\frac{\left(x-\mu\right)^2}{2\sigma^2/\kappa+\theta^2}\right)^{\frac{1}{2\kappa}-0.25}}{\kappa^{1/\kappa}\Gamma\left(1/\kappa\right)\sigma\sqrt{\pi}} e^{\frac{\theta}{\sigma^2}(x-\mu)} K_{1/\kappa-0.5}\left(\frac{\sqrt{\left(\frac{2\sigma^2}{\kappa}+\theta^2\right)\left(x-\mu\right)^2}}{\sigma^2}\right). \tag{A-4}$$

However, the characteristic function of the VG distribution is of the form:

$$\phi_{VG}\left(\omega\middle|\kappa,\theta,\sigma,\mu\right) = e^{i\mu\omega}\left(1-i\theta\kappa\omega+0.5\kappa\sigma^2\omega^2\right)^{\frac{-1}{\kappa}}. \tag{A-5}$$

The first two moments of the VG distribution are

$$E\left(X\right) = \mu+\theta, \tag{A-6}$$

$$Var\left(X\right) = \sigma^2+\kappa\theta^2. \tag{A-7}$$

Let $\mu=-\theta$ and $\sigma=\sqrt{1-\kappa\theta^2}$ such that $E\left(X\right)=0$ and $Var\left(X\right)=1$. The standardized VG distribution can be represented by two parameters, we write

$$f_{VG}\left(x\middle|\kappa,\theta\right)=\frac{\sqrt{2}\left(\frac{\left(x+\theta\right)^2}{2\left(1-\kappa\theta^2\right)/\kappa+\theta^2}\right)^{\frac{1}{2\kappa}-0.25}}{\kappa^{1/\kappa}\Gamma\left(1/\kappa\right)\sqrt{\pi\left(1-\kappa\theta^2\right)}}e^{\frac{\theta}{\left(1-\kappa\theta^2\right)}\left(x+\theta\right)}K_{1/\kappa-0.5}\left(\frac{\sqrt{\left(\frac{2\left(1-\kappa\theta^2\right)}{\kappa}+\theta^2\right)\left(x+\theta\right)^2}}{\left(1-\kappa\theta^2\right)}\right). \tag{A-8}$$

However, the characteristic function of the VG distribution is of the form:

$$\phi_{VG}\left(\omega\middle|\kappa,\theta\right)=e^{-i\theta\omega}\left(1-i\kappa\theta\omega+0.5\kappa\left(1-\kappa\theta^2\right)\omega^2\right)^{\frac{-1}{\kappa}}. \tag{A-9}$$

This is another form of the VG distribution and its standardization.

Appendix B

We introduce another popular representation of the NIG distribution and its standardization in Appendix B. According to the following relationships of parameters:

$$\alpha = \frac{\sqrt{\frac{\sigma^2}{\kappa} + \theta^2}}{\sigma^2}, \tag{B-1}$$

$$\beta = \frac{\theta}{\sigma^2}, \tag{B-2}$$

$$\delta = \frac{\sigma}{\sqrt{\kappa}}, \tag{B-3}$$

we can rewrite Equation (2-15) as follows:

$$f_{NIG}\left(x|\kappa,\theta,\sigma,\mu\right) = \frac{\sqrt{\frac{1}{\kappa}+\frac{\theta^2}{\kappa\sigma^2}}}{\pi}\exp\left(\frac{1}{\kappa}+\frac{\theta^2}{\sigma^2}\left(x-\mu\right)\right) \times \frac{K_1\left(\sqrt{\frac{\sigma^2+\kappa\theta^2}{\kappa\sigma^2}\left(\frac{1}{\kappa}+\frac{1}{\sigma^2}\left(x-\mu\right)^2\right)}\right)}{\sqrt{\frac{\sigma^2}{\kappa}+\left(x-\mu\right)^2}}. \tag{B-4}$$

However, the characteristic function of the NIG distribution is of the form:

$$\phi_{NIG}\left(\omega|\kappa,\theta,\sigma,\mu\right) = \exp\left(i\mu\omega + \frac{1}{\kappa}\left(1-\sqrt{1-2i\kappa\theta\omega+\kappa\sigma^2\omega^2}\right)\right). \tag{B-5}$$

The first two moments of the NIG distribution are

$$E\left(X\right) = \mu + \theta, \tag{B-6}$$

$$Var\left(X\right) = \sigma^2 + \kappa\theta^2. \tag{B-7}$$

Let $\mu = -\theta$ and $\sigma = \sqrt{1-\kappa\theta^2}$ such that $E\left(X\right)=0$ and $Var\left(X\right)=1$. The

standardized NIG distribution can be represented by two parameters, we write

$$f_{NIG}\left(x\middle|\kappa,\theta\right)=\frac{\sqrt{\frac{1}{\kappa}+\frac{\theta^2}{\kappa\left(1-\kappa\theta^2\right)}}}{\pi}\exp\left(\frac{1}{\kappa}+\frac{\theta^2}{\left(1-\kappa\theta^2\right)}\left(x+\theta\right)\right)$$
$$\times\frac{K_1\left(\sqrt{\frac{1}{\kappa\left(1-\kappa\theta^2\right)}\left(\frac{1}{\kappa}+\frac{1}{\left(1-\kappa\theta^2\right)}\left(x+\theta\right)^2\right)}\right)}{\sqrt{\frac{1-\kappa\theta^2}{\kappa}+\left(x+\theta\right)^2}}. \tag{B-8}$$

However, the characteristic function of the VG distribution is of the form:

$$\phi_{NIG}\left(\omega\middle|\kappa,\theta\right)=\exp\left(-i\theta\omega+\frac{1}{\kappa}\left(1-\sqrt{1-2i\kappa\theta\omega+\kappa\left(1-\kappa\theta^2\right)\omega^2}\right)\right) \tag{B-9}$$

This is another form of the NIG distribution and its standardization.

Appendix C

When the death rates follow the RH model, the explicit solution of the log-likelihood function in Equation (4-5) can be rewritten as follows:

$$\begin{aligned} LLF &= \sum_{x,t}\int_{-\infty}^{\infty} \log f\left(D_{x,t}=d_{x,t}\,|\,e_{x,t}=y\right) f_{e_{x,t}}(y)\,dy \\ &= \sum_{x.t}\left(\int_{-\infty}^{\infty} d_{x,t}\left(\log E_{x,t}+\alpha_x+\beta_x k_t+\eta_x\gamma_{t-x}+y\right) f_{e_{x,t}}(y)\,dy\right. \\ &\qquad \left.-\int_{-\infty}^{\infty} E_{x,t}\exp\left(\alpha_x+\beta_x k_t+\eta_x\gamma_{t-x}\right)\exp(y) f_{e_{x,t}}(y)\,dy-\log\left(d_{x,t}!\right)\right). \end{aligned} \tag{C-1}$$

Because $E\left(e_{x,t}\right)=0$, we have

$$\begin{aligned} LLF &= \sum_{x,t}\left[d_{x,t}\left(\log E_{x,t}+\alpha_x+\beta_x k_t+\eta_x\gamma_{t-x}\right)\right. \\ &\qquad \left.-E_{x,t}\exp\left(\alpha_x+\beta_x k_t+\eta_x\gamma_{t-x}\right)\int_{-\infty}^{\infty}\exp(y) f_{e_{x,t}}(y)\,dy-\log\left(d_{x,t}!\right)\right] \\ &= \sum_{x.t}\left[d_{x,t}\left(\alpha_x+\beta_x k_t+\eta_x\gamma_{t-x}\right)-\left(E_{x,t}\exp\left(\alpha_x+\beta_x k_t+\eta_x\gamma_{t-x}\right)\right) M_{e_{x,t}}(1)\right] \\ &+\sum_{x.t}\left[d_{x,t}\log E_{x,t}-\log\left(d_{x,t}!\right)\right]. \end{aligned} \tag{C-2}$$

This completes the proof of Equation (4-7).

國立政治大學
National Chengchi University

Appendix D

The procedure of computing the expected value of the *t*-year survival probability under the equivalent martingale measure *Q* is as follows.

Step 1: After calibrating the parameters of the RH model, we use a Monte Carlo simulation with *N* iterations to generate the futures mortality rates and the survival probabilities under the real-world probability measure *P*. According to the *N* simulated values of *t*-year survival probabilities, we can construct the corresponding empirical cumulative distribution function (cdf) $F_t(\cdot)$ and its inverse cdf $F_t^{-1}(\cdot)$ under *P*.

Step 2: We know that the probability-integral transform of a random variable is distributed as standard uniform. Consequently, we have, according to Equation (4-21),

$$\tilde{F}_t(S(t)) = U = \Phi\left(\Phi^{-1}\left(F_t(S(t))\right) + \lambda\right), \tag{D-1}$$

where U is a standard uniform random variable. Rearranging Equation (D-1) and drawing *N* random numbers from a standard uniform distribution, we can generate *N* possible values of the *t*-year survival probabilities under the equivalent martingale measure *Q*, as follows:

$$S(t) = F_t^{-1}\left(\Phi\left(\Phi^{-1}\left(U\right) - \lambda\right)\right). \tag{D-2}$$

Averaging the *N* values of the *t*-year survival probabilities produces the expected value of *t*-year survival probability under *Q*. A higher value of *N* leads to a more precise setup for $F_t(\cdot)$, $F_t^{-1}(\cdot)$, and $E_Q\left[S(t)\right]$. We use *N* = 100,000.

國立政治大學
National Chengchi University

Appendix E

Let $\omega = (\omega_1, ..., \omega_n)'$, the characteristic function of X is of the form:

$$\phi_X(\omega) = E\left(\exp(i\omega' X)\right)$$

$$= E\left(\exp\left(i\omega'(AZ + M)\right)\right) = \exp(i\omega' M) E\left(\exp(i\omega' AZ)\right). \quad \text{(E-1)}$$

Let $\psi' = \omega' A = (\psi_1, ..., \psi_n)$, because $Z = (Z_1, ..., Z_n)'$ is a random vector which consists of n mutually independent random variables, we have

$$\phi_X(\omega) = \exp(i\omega' M) E\left(\exp(i\psi' Z)\right)$$

$$= \exp(i\omega' M) \prod_{k=1}^{n} E\left(\exp(i\psi_k Z_k)\right) = \exp(i\omega' M) \prod_{k=1}^{n} \phi_{Z_k}(\psi_k). \quad \text{(E-2)}$$

This completes the proof of Appendix E.

國立政治大學
National Chengchi University

Appendix F

According to Lemma 5.2.2 of Shreve (2004), the characteristic function of logarithm of stock return under the martingale probability measure Q_θ is given by

$$E_{Q_\theta}\left(\exp\left(i\omega' R(t)\right)\middle|F_{t-1}\right)=E_P\left(\frac{\eta_t}{\eta_{t-1}}\exp\left(i\omega' R(t)\right)\middle|F_{t-1}\right)=E_P\left(\xi_t\exp\left(i\omega' R(t)\right)\middle|F_{t-1}\right)$$

$$=E_P\left(\frac{\exp\left(\theta' R(t)\right)}{E_P\left(\exp\left(\theta' R(t)\right)\right)}\exp\left(i\omega' R(t)\right)\middle|F_{t-1}\right)=\frac{E_P\left(\exp\left(i(\omega-i\theta)' R(t)\right)\right)}{E_P\left(\exp\left(\theta' R(t)\right)\right)}$$

$$=\frac{\exp\left(i(\omega-i\theta)'M\right)\prod_{k=1}^{n}\phi_{Z_k}\left(\psi_k-i\varphi_k^\theta\right)}{\exp\left(\theta' M\right)\prod_{k=1}^{n}\phi_{Z_k}\left(-i\varphi_k^\theta\right)}=\exp\left(i\omega' M\right)\prod_{k=1}^{n}\left(\frac{\phi_{Z_k}\left(\psi_k-i\varphi_k^\theta\right)}{\phi_{Z_k}\left(-i\varphi_k^\theta\right)}\right). \qquad \text{(F-1)}$$

Because Z_k adheres to a $stdGH(\alpha_k,\beta_k,\lambda_k)$ law, by using Equation (2-8), and μ and δ fulfill Equations (2-35) and (2-36), we have

$$\frac{\phi_{Z_k}\left(\psi_k-i\varphi_k^\theta\right)}{\phi_{Z_k}\left(-i\varphi_k^\theta\right)}=\frac{\mathrm{e}^{i(\psi_k-i\varphi_k^\theta)\mu_k}\left(-\frac{\alpha_k^2-\beta_k^2}{\alpha_k^2-(\beta_k+i(\psi_k-i\varphi_k^\theta))^2}\right)^{\lambda_k/2}\frac{K_{\lambda_k}\left(\delta_k\sqrt{\alpha_k^2-(\beta_k+i(\psi_k-i\varphi_k^\theta))^2}\right)}{K_{\lambda_k}\left(\delta_k\sqrt{\alpha_k^2-\beta_k^2}\right)}}{\mathrm{e}^{\varphi_k^\theta\mu_k}\left(-\frac{\alpha_k^2-\beta_k^2}{\alpha_k^2-(\beta_k+i(-i\varphi_k^\theta))^2}\right)^{\lambda_k/2}\frac{K_{\lambda_k}\left(\delta_k\sqrt{\alpha_k^2-(\beta_k+i(-i\varphi_k^\theta))^2}\right)}{K_{\lambda_k}\left(\delta_k\sqrt{\alpha_k^2-\beta_k^2}\right)}}$$

$$=\mathrm{e}^{i\psi_k\mu_k}\left(-\frac{\alpha_k^2-(\beta_k+\varphi_k^\theta)^2}{\alpha_k^2-(\beta_k+\varphi_k^\theta+i\psi_k)^2}\right)^{\lambda_k/2}\frac{K_{\lambda_k}\left(\delta_k\sqrt{\alpha_k^2-(\beta_k+\varphi_k^\theta+i\psi_k)^2}\right)}{K_{\lambda_k}\left(\delta_k\sqrt{\alpha_k^2-(\beta_k+\varphi_k^\theta)^2}\right)}. \qquad \text{(F-2)}$$

In view of Equation (F-2), Z_k adheres to a $GH(\alpha_k,\beta_k+\varphi_k^\theta,\lambda_k,\delta_k,\mu_k)$ law under the risk-neutral measure Q_θ.

Based on the risk-neutral pricing theory, pricing the contingent claims is done under the martingale probability measure Q_θ which makes the discounted stock

price into a Q_θ-martingale, which can be represented as $E_Q\left(S_t \middle| \Im_{t-1}\right) = S_{t-1}\exp\left(r_f\right)$, that is

$$E_Q\left(S_j(t)\middle|\Im_{t-1}\right) = S_j(t-1)E_Q\left(\exp\left(\ln\left(\frac{S_j(t)}{S_j(t-1)}\right)\right)\middle|\Im_{t-1}\right)$$

$$= S_j(t-1)E_{Q_\theta}\left(\exp\left(R_j(t)\right)\middle|F_{t-1}\right) = S_j(t-1)e^{r_f}. \quad \text{(F-3)}$$

Or equivalently, we have

$$e^{r_f} = E_{Q_\theta}\left(\exp\left(R_j(t)\right)\middle|F_{t-1}\right) = E_P\left(\xi_t\exp\left(R_j(t)\right)\middle|F_{t-1}\right) = \frac{E_P\left(\exp\left((\theta+I_j)'R(t)\right)\right)}{E_P\left(\exp\left(\theta'R(t)\right)\right)}$$

$$= \frac{\exp\left((\theta+I_j)'M\right)E_P\left(\exp\left((\varphi_\theta{}' + I_j{}'A)Z\right)\right)}{\exp\left(\theta'M\right)E_P\left(\exp\left(\varphi_\theta{}'Z\right)\right)} = e^{m_j}\prod_{k=1}^{n}\left(\frac{E_P\left(\exp\left((\varphi_k^\theta + a_{jk}1_{(j\geq k)})Z_k\right)\right)}{E_P\left(\exp\left(\varphi_k^\theta Z_k\right)\right)}\right)$$

$$= e^{m_j}\prod_{k=1}^{j}\left(\frac{E_P\left(\exp\left((\varphi_k^\theta + a_{jk})Z_k\right)\right)}{E_P\left(\exp\left(\varphi_k^\theta Z_k\right)\right)}\right), \quad \text{(F-4)}$$

where I_j is a column vector which the *j*th coordinate is one and the other is zero and $1_{(\bullet)}$ is indicator function. Consequently, we obtain *the parameter* φ_j^θ *by using the following equation:*

$$r_f = m_j + \sum_{k=1}^{j}\left(\log\left(\phi_{Z_k}\left(-i(\varphi_k^\theta + a_{jk})\right)\right) - \log\left(\phi_{Z_k}\left(-i\varphi_k^\theta\right)\right)\right). \quad \text{(F-5)}$$

Because $\varphi_\theta{}' = \theta'A$, we have $\varphi_j^\theta = \sum_{m=j}^{n} a_{mj}\theta_m$. As a result, after obtaining φ_j^θ, *j*=1...,*n*, we can obtain $\theta_n = \varphi_n^\theta / a_{nn}$ and $\theta_j = \left(1/a_{jj}\right)\left(\varphi_j^\theta - \sum_{m=j+1}^{n} a_{mj}\theta_m\right)$ for *j*=1...,*n*-1. This completes the proof of Appendix F.

References

Aas, K., Haff, I. H., 2006. The Generalized Hyperbolic Skew Student's t-distribution. Journal of Financial Econometrics 4, 275-309.

Akaike, H, 1974. A New Look at the Statistical Model Identification. IEEE Transactions on Automatic Control AC-19, 716-723.

Amin, K., 1993. Jump Diffusion Option Valuation in Discrete Time. Journal of Finance 48, 1833-1863.

Anderson, T. W., 1962. On the Distribution of the Two-Sample Cramér-Von Mises Criterion. The Annals of Mathematical Statistics 33, 1148-1159.

Barbarin J., 2008. Heath-Jarrow-Morton Modelling of Longevity Bonds and the Risk Minimization of Life Insurance Portfolios. Insurance Mathematics and Economics 43, 41-55.

Barndorff-Nielsen, O. E., 1977. Exponentially decreasing distributions for the logarithm of particle size. Proceedings of the Royal Society of London 353, 409-419.

Barndorff-Nielsen, O. E., 1978. Hyperbolic distributions and distributions on hyperbolae. Scandinavian Journal of Statistics 5, 151-157.

Barndorff-Nielsen, O. E., 1995. Normal Inverse Gaussian Processes and the Modeling of Stock Returns. Technical Report 300, Department of Theoretical Statistics, Institute of Mathematics.

Barndorff-Nielsen , O. E, Pedersen, J., Sato, K. I., 2001. Multivariate Subordination Self-Decomposability and Stability. Advance Application Probability 33, 160-187.

Barndorff-Nielsen, O. E., Shephard, N., 2001. Non-Gaussian Ornstein-Uhlenbeck-Based Models and Some of Their Uses in Financial Economics. Journal of the Royal Statistical Society B 63, 167-241

Bauer, D., 2006. An Arbitrage-Free Family of Longevity Bonds, Discussion Paper, Ulm University.

Biffis, E., 2005. Affine Processes for Dynamic Mortality and Actuarial Valuations. Insurance: Mathematics and economics 37, 443-468.

Biffis, E., Blake, D., Pitotti, L., Sun, A., 2011. The Cost of Counterparty Risk and Collateralization in Longevity Swaps, Pensions Institute Discussion Paper PI-1107, June.

Biffis, E., Denuit, M., Devolder, P., 2010. Stochastic Mortality under Measure Changes. Scandinavian Actuarial Journal 4, 284-311.

Bishop, C. M., 2006. Pattern Recognition and Machine Learning. Springer.

Blake, D., Burrows, W., 2001. Survivor Bonds: Helping to Hedge Mortality Risk. Journal of Risk and Insurance 68, 339-348.

Blake, D., Cairns, A. J. G., Coughlan, G., Dowd, K., MacMinn, R., 2012. The New Life Market, Discussion Paper.

Blasild, P., Jensen, J. L., 1981. Multivariate Distributions of Hyperbolic Type. In Statistical Distributions in Scientific Work-Proceedings of theNATO Advanced Study Institute held at the Università degli studi di Trieste 4, 45-66.

Bølviken, E., Benth, F. E., 2000. Quantification of Risk in Norwegian Stocks via the Normal Inverse Gaussian Distribution. Proceedings of the AFIA 2000 Colloquium, Tromsø, Norway, 87-98.

Boyle, P. P., Kolkiewicz, A. W., Tan, K. S., 2003. Pricing American Style Options Using Low Discrepancy Mesh Method. Submitted for Publication.

Broadie, M., Glasserman, P., 2004. A Stochastic Mesh Method for Pricing High-Dimensional American Options. Journal of Computational Finance 7, 35-72.

Brouhns, N., Denuit, M., Vermunt, J. K., 2002. A Poisson Log-Bilinear Regression Approach to the Construction of Projected Life Tables. Insurance: Mathematics and Economics 31, 373-393.

Cairns, A. J. G., Blake, D., Dowd, K., 2006. A Two-Factor Model for Stochastic Mortality with Parameter Uncertainty: Theory and Calibration. Journal of Risk and Insurance 73, 687-718.

Cairns, A. J. G., Blake, D., Dowd, K., Coughlan, G. D., Epstein, D., Khalaf-Allah, M., 2010. A Framework for Forecasting Mortality Rates with an Application to Six Stochastic Mortality Models. Pensions Institute Discussion Paper PI-0801, March.

Cairns, A. J. G., Blake, D., Dowd, K., Coughlan, G. D., Epstein, D., Ong, A., Balevich, I., 2009. A Quantitative Comparison of Stochastic Mortality Models Using Data From England and Wales and the United States. North American Actuarial Journal 13, 1-35.

Carr, P., Geman, H., Madan, D. P., Yor, M., 2002. The Fine Structure of Asset Returns: An Empirical Investigation. Journal of Business 75, 305-332.

Carr, P., Madan, D. P., 1999. Option Valuation Using the Fast Fourier transform. Journal of Computational Finance 2, 61-73.

Chen, H., Cox, S. H., 2009. Modeling Mortality with Jumps: Applications to Mortality Securitization. Journal of Risk and Insurance 76, 727-751.

Chernobai, A. S., Rachev, S. T., Fabozzi, F. J., 2007. Operational Risk: A Guide to Basel II Capital Requirements, Models, and Analysis. John Wiley & Sons, Inc..

Clark, P. K., 1973. A Subordinated Stochastic Process Model with Finite Variance for Speculative Prices. Journal of the Econometric Society 41, 135-156.

Cont, R., Tankov, P., 2004. Financial Modelling with Jump Processes. Chapman and Hall/CRC Financial Mathematics Series.

Cox, D. R., 1955. Some Statistical Methods Connected with Series of Events (with Discussion). Journal of the Royal Statistical Society, Series B 17, 129-164.

Cox, S. H., Lin, Y., Wang, S. S., 2006. Multivariate Exponential Tilting and Pricing Implications for Mortality Securitization. Journal of Risk and Insurance 73, 719-736.

Dawson, P., 2002. Mortality Swaps, Mimeo, Cass Business School.

Dawson, P., Blake, D., Cairns, A. J. G., Dowd, K., 2010. Survivor Derivatives: A Consistent Pricing Framework. Journal of Risk and Insurance 77, 579-596.

Demarta, S., McNeil, A. J., 2005. The t Copula and Related Copulas. International Statistical Review 73, 111-129.

Denuit, M., Devolder, P., Goderniaux, A. C., 2007. Securitization of Longevity Risk: Pricing Survivor Bonds With Wang Transform in The Lee-Carter Framework. Journal of Risk and Insurance 74, 87-113.

Dowd, K., Blake, D., Cairns, A. J. G., Dawson, P., 2006. Survivor Swaps. Journal of Risk and Insurance 73, 1-17.

Dowd, K., Cairns, A. J. G., Black, D., Coughlan, G. D., Epstein, D., Khalaf-Allah, M., 2010. Evaluating the Goodness of Fit of Stochastic Mortality Models. Insurance: Mathematics and Economics 47, 255-265.

Eberlein, E., Keller, U., 1995. Hyperbolic Distributions in Finance. Bernoulli 1, 281-299.

Eberlein, E., Madan, D. B., 2009. On Correlating Lévy Processes. Working paper.

Fajardo, J., Farias, A., 2009. Multivariate Affine Generalized Hyperbolic Distributions: An Empirical Investigation. International Review of Financial Analysis 18, 174-184.

Fajardo, J., Farias, A., 2010. Derivative Pricing Using Multivariate Affine Generalized Hyperbolic Distributions. Journal of Banking and Finance 34, 1607-1617.

Fajardo, J., Mordecki, E., 2006. Pricing Derivatives on Two Dimensional Lévy Processes. International Journal of Theoretical and Applied Finance 9, 185-197.

Fu, M. C., Laprise, S. B., Madan, D. B., Su, Y., Wu, R., 2001. Pricing American Options: A Comparison of Monte Carlo Simulation Approaches. Journal of Computational Finance 2, 62-73.

Gerber, H., Shiu, E., 1994. Option Pricing by Esscher Transforms. Transactions of the Society of Actuaries 46, 99-191.

Giacometti, R., Ortobelli, S., Bertocchi, M. I., 2009. Impact of Different Distributional Assumptions in Forecasting Italian Mortality Rates. Investment Management and Financial Innovations 6(3), 186-193.

Glasserman, P., 2003. Monte Carlo Methods in Financial Engineering. New York, Springer.

Goodman. L. A., 1979. Simple Models for the Analysis of Association in Cross-Classifications Having Ordered Categories. Journal of the American Statistical Association 74, 537-552.

Haberman, S., Renshaw, A. E., 2009. On Age-Period-Cohort parametric motality rate projections. Insurance: Mathematics and Economics 45, 255-270.

Hainaut, D., 2012. Multidimensional Lee-Carter Model with Switching Mortality Processes. Insurance: Mathematics and Economics 50, 236-246.

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

Harrison, J. M., Kreps, D. M., 1979. Martingales and Arbitrage in Multiperiod Securities Markets. Journal of Economic Theory 20, 381-408.

Harrison, J. M., Pliska, S. R., 1981. Martingales and Stochastic Integrals in the Theory of Continuous Trading. Stochastic Processes and Their Applications 11, 215-280.

Harrison, J.M., Pliska, S. R., 1983. A Stochastic Calculus Model of Continuous Trading: Complete Markets. Stochastic Processes and their Applications 15, 313-316.

Heath, D., Jarrow, R., Morton, R., 1992. Bond Pricing and the Term Structure of Interest Rates: A New Methodology for Contingent Claim Valuation. Econometrica 60, 77-105.

Hirsa, A., Madan, D., 2004. Pricing American Options under Variance-Gamma. Journal of Computational Finance 7, 63-80.

Jarque, C. M., Bera, A. K., 1980. Efficient Tests for Normality, Homoscedasticity and Serial Independence of Regression Residuals. Economic Letters 6, 255-259.

Jones, M. C., Faddy, M. J., 2003. A Skew Extension of the t Distribution, with Applications. Journal of the Royal Statistical Society B 65, 159-174.

Këllezi, E., Webber, N., 2004. Valuing Bermudan Options When Asset Returns Are Lévy processes. Quantitative Finance 4, 87-100.

Kolmogorov A. N., 1933. Grundbegriffe der Wahrscheinlichkeitsrechnung. Springer, Berlin. English translation (1950): Foundations of the theory of probability. Chelsea, New York.

Lee, R., 2000. The Lee-Carter Method for Forecasting Mortality, with Various Extensions and Applications. North American Actuarial Journal 4, 80-93.

Lee, R. D., Carter, L. R., 1992. Modeling and Forecasting U.S. Mortality. Journal of the American Statistical Association 87, 659-675.

Li. S. H., Chan, W. S., 2007. The Lee-Carter Model for Forecasting Mortality, Revisited. North American Actuarial Journal 11, 68-89.

Lillestøl, J., 2000. Risk Analysis and the NIG Distribution. Journal of Risk 2, 41-56.

Lin, Y., Cox, S. H., 2005. Securitization of Mortality Risks in Life Annuities. Journal of Risk and Insurance 72: 227-252.

Lin, Y., Cox, S. H., 2008. Securitization of Catastrophe Mortality Risks. Insurance Mathematics and Economics 42, 628-637.

Loeys, J., Panigirtzoglou, N., Ribeiro, R., 2007. Longevity: A Market in the Making, J. P. Morgan Research Publication.

Longstaff, F. A., Schwartz, E. S., 2001. Valuing American Options by Simulation: A Simple Least-Squares Approach. The Review of Financial Studies 14, 113-147.

Luciano, E., Schoutens, W., 2006. A Multivariate Jump-Driven Financial Asset Model. Quantitative Finance 385-402.

Luciano, E. Semeraro, P., 2007. Extending Time-Changed Lévy Asset Models through Multivariate Subordinators. Working paper.

Luciano, E., Semeraro, P., 2010. A Generalized Normal Mean Variance Mixture for Return Processes in Finance. International Journal of Theoretical and Applied Finance 13, 415-440.

Luciano, E., Vigna, E., 2005. Non Mean Reverting Affine Processes for Stochastic Mortality. ICER Applied Mathematics Working Paper No.4 Available at SSRN: http://ssrn.com/abstract=724706.

Madan D. B., Carr, P. P., Chang, E. C., 1998. The Variance Gamma Process and Option Pricing. European Finance Review 2, 79-105.

Madan, D. B., Seneta, E., 1987. Chebyshev Polynomial Approximations and Characteristic Function Estimation. Journal of the Royal Statistical Society Series B 49, 163-169.

Madan D. B., Seneta, E., 1990. The Variance Gamma (VG) Model for Share Market Returns. Journal of Business 63, 511-524.

Maller, R. A., Solomon, D. H., Szimayer, A., 2006. A Multinomial Approximation for American Option Prices in Lévy process Models. Mathematical Finance 16, 613-633.

Mandelbrot, B., Taylor, H., 1967. On the Distribution of Stock Prices Differences. Operations Research 15, 1057-1062.

Matache, A. M., Nitsche, P. A., Schwab, C., 2005. Wavelet Galerkin Pricing of American Options on Lévy Driven Assets. Quantitative Finance 5, 403-424.

Mencia, F. J., Sentana, E., 2004. Estimation and Testing of Dynamic Models with Generalised Hyperbolic Innovations. CMFI Working Paper 0411, Madrid, Spain.

Milidonis, A., Lin, Y., Cox, S. H., 2011. Mortality Regimes and Pricing. North American Actuarial Journal 15, 266-289.

Mulinacci, S., 1996. An Approximation of American Option Prices in a Jump-Diffusion Model. Stochastic Processes and their Applications 62, 1-17.

Pitacco, E., 2004. Survival Models in Dynamic Context: A Survey. Insurance: Mathematics and Economics 35, 279-298.

Prause, K., 1997. Modelling Financial Data Using Generalized Hyperbolic Distributions. FDM Preprint 48, University of Freiburg.

Prause, K., 1999. The Generalized Hyperbolic Models: Estimation, Financial Derivatives and Risk Measurement. PhD Thesis, Mathematics Faculty, University of Freiburg.

Renshaw, A. E., Haberman, S., 2003. Lee-Carter Mortality Forecasting with Age-Specific Enhancement. Insurance: Mathematics and Economics 33, 255-272.

Renshaw, A. E., Haberman, S., 2006. A Cohort-Based Extension to the Lee-Carter

Model for Mortality Reduction Factors. Insurance: Mathematics and Economics 38, 556-570.

Rydberg, T. H., 1997. The Normal Inverse Gaussian Lévy process: Simulation and Approximation. Communications in Statistics: Stochastic models 13, 887-910.

Schwarz, G., 1978. Estimating the Dimension of a Model. Annals of Statistics 6, 461-464.

Schmidt, R., Hrycej, T., Stutzle, E., 2006. Multivariate Distribution Models with Generalized Hyperbolic Margins. Computational Statistics and Data Analysis 50, 2065-2096.

Semeraro, P., 2008. A Multivariate Variance Gamma Model for Financial Application. Journal of Theoretical and Applied Finance 11,1-18.

Stephens, M. A., 1974. EDF Statistics for Goodness of Fit and Some Comparisons. Journal of the American Statistical Association 69, 730-737.

Wang, C. W., Huang, H. C., 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—Issues and Practice 36, 675-696.

Wang, C. W., Yang, S. S., 2012. Pricing Survivor Derivatives with Cohort Mortality Dependence under the Lee-Carter Framework. Forthcoming in Journal of Risk and Insurance.

Wilmoth, J. R., 1993. Computational Methods for Fitting and Extrapolating the Lee-Carter Model of Mortality Change. Technical Report, Department of Demography, University of California, Berkeley.