KONINKLIJKE VLAAMSE ACADEMIE VAN BELGIE
VOOR WETENSCHAPPEN EN KUNSTEN

3RD ACTUARIAL AND FINANCIAL MATHEMATICS DAY

4 februari 2005

Michèle Vanmaele, Ann De Schepper, Jan Dhaene, Huguette Reynaerts, Wim Schoutens & Paul Van Goethem (Eds.)

CONTACTFORUM

KONINKLIJKE VLAAMSE ACADEMIE VAN BELGIE
VOOR WETENSCHAPPEN EN KUNSTEN

3RD ACTUARIAL AND FINANCIAL MATHEMATICS DAY

4 februari 2005

**Michèle Vanmaele, Ann De Schepper, Jan Dhaene,
Huguette Reynaerts, Wim Schoutens & Paul Van Goethem (Eds.)**

CONTACTFORUM

Handelingen van het contactforum "3rd Actuarial and Financial Mathematics Day" (4 februari 2005, hoofdaanvrager: Prof. M. Vanmaele, UGent) gesteund door de Koninklijke Vlaamse Academie van België voor Wetenschappen en Kunsten.

Afgezien van het afstemmen van het lettertype en de alinea's op de richtlijnen voor de publicatie van de handelingen heeft de Academie geen andere wijzigingen in de tekst aangebracht. De inhoud, de volgorde en de opbouw van de teksten zijn de verantwoordelijkheid van de hoofdaanvrager (of editors) van het contactforum.

KONINKLIJKE VLAAMSE ACADEMIE VAN BELGIE
VOOR WETENSCHAPPEN EN KUNSTEN

Paleis der Academiën
Hertogsstraat 1
1000 Brussel

D/2005/0455/05

Printed by *Universa Press, 9230 Wetteren, Belgium*

KONINKLIJKE VLAAMSE ACADEMIE VAN BELGIE
VOOR WETENSCHAPPEN EN KUNSTEN

3rd Actuarial and Financial Mathematics Day

PREFACE

The Contactforum "Actuarial and Financial Mathematics Day" has come to its third edition, and it definitely turns into an annual meeting between academics and practitioners. The large attendance at the symposium confirms the interest for strengthening the ties between the different research groups in actuarial and financial mathematics of the Flemish/Belgian universities on the one side, and professionals of the banking and insurance business on the other side. A contactforum like this seems to be a good formula for exchanging results and problems in this fascinating research field.

These transactions include two types of presentations. First, we have two invited papers of the guest speakers; this year, we could welcome Prof. dr. Rob Kaas from the University of Amsterdam, and Dr. Lutz Schloegl from Lehman Brothers International UK. Next, there are 8 contributions, presented by PhD students, postdocs, and practitioners.

We thank all our speakers, without whose effort the organization of the contactforum wouldn't be possible. We are also extremely grateful to our sponsors: the Royal Flemish Academy of Belgium for Science and Arts, and Scientific Research Network “Fundamental Methods and Techniques in Mathematics” of the Fund for Scientific Research - Flanders. They made it possible to spend the day in a very agreeable and inspiring environment.

The success of the meeting encourages us to continue with this yearly initiative. We are convinced that it provides a great opportunity to facilitate the exchange of ideas; it certainly stimulates the research in actuarial and financial mathematics in Flanders.

Ann De Schepper
Jan Dhaene
Huguette Reynaerts
Wim Schoutens
Paul Van Goethem
Michèle Vanmaele

KONINKLIJKE VLAAMSE ACADEMIE VAN BELGIE
VOOR WETENSCHAPPEN EN KUNSTEN

Contactforum, "3rd Actuarial and Financial Mathematics Day"

TABLE OF CONTENTS

Invited talks

Contributed talks

INVITED TALKS

COMPOUND POISSON DISTRIBUTIONS AND GLM'S — TWEEDIE'S DISTRIBUTION

Rob Kaas

Department of Quantitative Economics, Universiteit van Amsterdam, Roetersstraat 11, 1018 WB Amsterdam, The Netherlands
Email: `R.Kaas@UvA.NL`

Abstract

Generalized Linear Models are especially useful for actuarial applications, since they allow one to estimate multiplicative models, and also allow forms of heteroscedasticity such as they are found frequently in actuarial problems, of Poisson-type, of gamma-type with a fixed coefficient of variation, and in-between (Tweedie's class of Compound Poisson–gamma distributions).

1. INTRODUCTION

At the moment, many actuarial education programs do not contain any material on Generalized Linear Models (GLMs). And in fact, the textbook Modern Actuarial Risk Theory by Kaas et al. (2001) is only the first actuarial textbook devoting space to this subject. A new textbook on actuarial science and GLMs, however, is in preparation, in close cooperation between the KU Leuven and the University of Amsterdam.

One reason for advocating the use and study of GLMs for actuaries is that the generalizations that GLMs provide with respect to ordinary linear models are especially important for actuarial applications. Also, there are quite interesting connections with actuarial risk theory. Moreover, some renowned actuarial techniques are actually special cases of GLMs, like Bailey-Simon's rating method as well as the celebrated Chain Ladder method.

The importance of GLMs for actuarial practice is gaining recognition. For instance, a meeting of the Casualty Actuarial Society in 2004 was devoted entirely to GLMs. Also, GLMs are used as the standard method for all premium rating for personal lines in the UK, for graduation, and so on. A good review paper about this, aimed at an audience of statisticians, is Haberman and Renshaw (1996). In this paper it is demonstrated how GLMs can be used for a variety of actuarial statistical problems like survival modeling, graduation, multiple-state models, loss distributions, risk classification, premium rating and claims reserving in non-life-insurance.

In Section 2 we present the case for using GLMs and describe how they work. The quasi-likelihood described in Section 3 may be used to extend the class of distributions considered somewhat. Tweedie's class of distributions is the topic of Section 4, and we give an application in IBNR estimation in Section 5.

2. GENERALIZED LINEAR MODELS

An important problem a non-life actuary faces in his daily practice is the following: given data on a portfolio of risks classified by several characteristics, construct or analyze a rating system for this portfolio. Econometricians would resort to multiple linear regression to identify and calibrate the underlying data-generating mechanism. To apply linear regression properly, the effects of the covariates must be additive, the errors must be normally distributed, hence symmetric, and their variance must not depend on the mean (homoscedasticity). But in insurance applications (tariffs), the models used are generally multiplicative, hence linear only on the log-scale. Claim *numbers* are generally Poisson, with a variance equal to the mean, or Poisson-like with 'overdispersion' (variance/mean is a constant larger than 1). These distributions are not symmetric and heteroscedastic. Many softwares can perform Poisson regression. Claim *amounts* generally have a density shaped like the gamma density. So there is no left-hand tail and a significant right-hand tail, hence asymmetry. Often, rather than a constant variance, they exhibit a constant coefficient of variation σ/μ. Claim *totals* can often be thought of as generated by a compound Poisson process, with claim amounts as above. This means that they are neither continuous nor discrete. Then the variance, as a function of the mean, might be modeled as proportional to μ^p for some $p \in (1,2)$.

Transformations of observations are frequently used in econometrics to try to make the data better suited to the technique used. While this is OK in many situations, in insurance there is a natural and fixed dimension to the problem: one needs the actual amount of the premium to be asked or reserve to be held. And upon inverse transformation, desirable properties are often lost, for instance, one must be careful to remove the resulting bias. Sometimes even consistency is lost. Also, transformations do not solve everything. For instance if $Y \sim$ Poisson, then $Y^{1/2}$ has a more or less constant variance, $Y^{1/3}$ has more or less skewness zero, but $\log Y$ has additive effects instead of multiplicative. Actuaries tend to concentrate on predicting future values and on point estimates of parameters (for the tariff), and often do not care so much if a model is statistically valid.

A good way to overcome the problems mentioned is using the *Generalized Linear Models* introduced by Nelder and Wedderburn (1972). GLMs are flexible enough to encompass a large class of models applicable in actuarial statistics, yet their formulation is tight enough to allow the existence of one algorithm for the maximum likelihood estimation of all of them. GLMs are more helpful in actuarial statistics than ordinary multiple regression, since apart from normal distributions, GLMs explicitly allow Poisson, binomial, gamma and some other useful error distributions. Also, GLMs allow linearity on other scales than the identity scale (logarithmic, logit, probit, reciprocal and others). Note that in GLMs, not, e.g., $\mathrm{E}[\log Y_i]$ is taken linear on the covariate vector x_i, but $\log \mathrm{E}[Y_i]$.

A GLM has the following three components:

1. A *random* component:
   Independent observations $Y_1, \ldots, Y_n$ are available with a density from the exponential family, parameterized by μ_i, $i = 1, \ldots, n$ (denoting the mean) and ψ_i (called the dispersion parameter). Primary examples are:
   - Normal(μ_i, ψ_i), with variance $\psi_i(\mu_i)^0$
   - Poisson(μ_i, ψ_i), denoting multiples ψ_i times a Poisson(μ_i/ψ_i) random variable, with variance $\psi_i(\mu_i)^1$
   - compound Poisson r.v.'s with gamma claim severities, with for some $1 < p < 2$, variance $\psi_i(\mu_i)^p$
     (Tweedie's distributions; intensity and scale vary, the shape is fixed)
   - gamma$(\frac{1}{\psi_i}, \frac{1}{\psi_i \mu_i})$ distributions, having variance $\psi_i(\mu_i)^2$
   - inverse Gaussian$(\frac{1}{\psi_i \mu_i}, \frac{1}{\psi_i \mu_i^2})$ distributions, having variance $\psi_i(\mu_i)^3$
   - binomial, and negative binomial, with varying p-parameter
2. A *systematic* component:
   There is a linear predictor $\eta_i = \sum_j x_{ij}\beta_j$ for each observation $i = 1, \ldots, n$; here, the matrix $\mathbf{X} = ((x_{ij}))$ is the design matrix with covariates; $(\beta_1, \beta_2, \ldots)^T$ is the parameter vector.
3. Random and systematic component of a GLM are connected through a smooth and invertible *link function*: $\eta_i = g(\mu_i)$.

In matrix notation we simply have $g(\mu) = \eta = \mathbf{X}\beta$. In case of a logarithmic link function $\eta = \log \mu$, or equivalently an exponential mean function $\mu = \exp(\eta)$, we have linearity on the log-scale, so a multiplicative model. A standard Linear Model has normal errors and an identity link.

In fact, the observations in a GLM are assumed to be independent r.v.'s from an exponential dispersion family, consisting of densities of the form

$$f_Y(y; \theta, \psi) = \exp\left(\frac{y\theta - b(\theta)}{\psi} + c(y; \psi)\right), \quad y \in R_\psi. \tag{1}$$

The range R_ψ of the random variable may vary with ψ. The log-density having a term $y\theta/\psi$ makes θ the 'natural' parameter. It only affects the mean: using the cumulant generating function $\kappa(t) = (b(\theta + t\psi) - b(\theta))/\psi$ with these densities, see for instance Kaas et al. (2001), Corollary 8.6.3, it is easily shown that $\mu = b^{'}(\theta)\ \forall\psi$ and $\sigma^2 = \psi V(\mu)$, with $V(\mu) = b^{''}(\theta(\mu))$ the so-called variance function. It can also be shown that when the functions $b(.)$ and $c(.;.)$ as well as ψ are fixed, the subfamily arising by taking different θ consists of elements that are all Esscher-transforms of each other. A family with b, c and θ fixed and varying ψ can be generated by the operation of taking sample means. See Kaas et al. (2001), Section 8.6.

It is generally assumed that the precision of the ith observation varies in a known way, as when it is the mean of w_i i.i.d. observations. The number w_i is the natural weight, or exposure, or credibility for observation $i = 1, \ldots, n$. We take $\psi_i = \phi/w_i$, where the parameter ϕ may be known or unknown, but not a function of $\beta_1, \beta_2, \ldots$; it is a 'nuisance parameter', and ML-estimations of μ_i will never depend on the value chosen for ϕ.

The fit criterion used in GLM theory for statistical inference is the loglikelihood ratio. In case of normality, this is equal to the least squares distance to a 'full' model with a parameter for every observation i. In case of Poisson, gamma and other distributions, it is equal to other suitable distances between observations and fitted values. Analysis of residuals is performed by looking at the contribution of individual observations to this or some other distance. Analysis of deviance employs the scaled deviance $-2\log\Lambda$. Here Λ is the likelihood ratio: the likelihood of the current model, divided by the one of a saturated model. Every test (Student, F and so on) that can be used in ordinary linear models can be used asymptotically in GLMs.

Constructing a model involves determining covariates, link function and error type (variance function). For choosing covariates, goodness of fit must be weighed against manageability. One has the null model, in which all variation is random, the covariates have no influence; the only parameter is the overall mean μ, and the number of parameters equals 1. On the other end of the scale, there is the saturated or full model, ascribing all variation to the covariates; each observation has its own parameter. Both are in general unsatisfactory, and one will have to find a suitable model in-between.

The possible 'fitted values'/'predictions' $\mu_1, \ldots, \mu_n = g^{-1}(\eta_1), \ldots, g^{-1}(\eta_n)$ are the image under g^{-1} in $\mathbb{R}^n$ of a linear subspace with as dimension the number of parameters; we have $\eta = \mathbf{X}\beta$ and $\mu = g^{-1}(\eta)$.

3. QUASI-LIKELIHOOD

Consider the densities (1) for independent observations $y_1, y_2, \ldots, y_n$. Assume that there are parameters $\beta_1, \beta_2, \ldots$ leading to means μ_i and associated θ_i through the relations $\mu = g^{-1}(\mathbf{X}\beta)$ and $\mu(\theta) = b'(\theta)$. Let ψ be of the form $\psi_i = \phi/w_i$ for some fixed dispersion parameter ϕ and known weights w_i; the ith observation is an average of w_i r.v.'s with mean μ_i and dispersion parameter ϕ. Using the relations $\mu(\theta) = b'(\theta)$, therefore $\theta(\mu) = (b')^{-1}(\mu)$, as well as $V(\mu) = b''(\theta) = \partial\mu/\partial\theta$, we see that for the loglikelihood ℓ with a single observation of (1) we have:

$$\frac{\partial\ell}{\partial\mu} = \frac{\partial\ell}{\partial\theta}\frac{\partial\theta}{\partial\mu} = \frac{y - b'(\theta)}{\phi/w}\Big/ b''(\theta) = \frac{y-\mu}{\phi V(\mu)/w}, \quad y \in R_{\phi/w}.$$

From this we see directly that for this observation, the maximum of ℓ obtains when $\mu = y$. The likelihood ratio Λ is the ratio of the maximized likelihood under a model resulting in means $\mu_1, \ldots, \mu_n$ (depending on parameters $\beta_1, \beta_2, \ldots$), divided by the one maximized without imposing any restrictions on the means, i.e., under the 'full' model, and therefore

$$\log\Lambda = \ell(\mu_1, \ldots, \mu_n) - \ell(y_1, \ldots, y_n) = \frac{1}{\phi}\sum_{i=1}^{n} w_i \int_{y_i}^{\mu_i} \frac{y_i - \mu}{V(\mu)}\mathrm{d}\mu. \tag{2}$$

The *scaled deviance* is just $-2\log\Lambda$, while $D = -2\phi\log\Lambda$ is called the *deviance*.

Note that in formula (2) for the deviance, from the specific form of the density (1) only the mean-variance relationship $V(.)$ has remained. Tweedie (1984) has shown that there are exponential families having $V(\mu) = \mu^p$ for every $p \notin (0, 1)$. Performing the integration in (2) for a fixed variance function $V(\mu) = \mu^p$ and dispersion parameter ϕ, we get expressions for the corresponding deviance. For the case $p = 0$ we get the least-squares distance, corresponding to normal distributions with a fixed variance:

$$D_0 = \sum_1^n w_i (y_i - \mu_i)^2.$$

For $p = 1$, we have the Poisson distributions. For the Poisson distribution proper, mean and variance are equal (for $w = 1$), so $\phi = 1$ should hold. But if we extend the class of r.v.'s studied to multiples of ϕ times a Poisson(μ/ϕ) r.v., we still have a subclass of the exponential family, with arbitrary μ, ϕ combinations. The resulting deviance for this case is (with the expression in brackets equal to μ_i if $y_i = 0$):

$$D_1 = 2 \sum_1^n w_i \left(y_i \log\left(\frac{y_i}{\mu_i}\right) - (y_i - \mu_i) \right).$$

For $p = 2$, the gamma case with a fixed coefficient of variation, we get

$$D_2 = 2 \sum_1^n w_i \left(\frac{y_i}{\mu_i} - \log\left(\frac{y_i}{\mu_i}\right) - 1 \right).$$

For all $p \notin \{1, 2\}$, we get the following general expression for the (quasi-)deviance:

$$D_p = 2 \sum_1^n w_i \left(\frac{y_i^{2-p} - (2-p) y_i \mu_i^{1-p} + (1-p)\mu_i^{2-p}}{(1-p)(2-p)} \right). \tag{3}$$

Note that for $p = 0$ and in the limit for $p \to 1$ and $p \to 2$, D_p in (3) reduces to D_0, D_1 and D_2. For $p = 3$, the deviance D_3 is the one associated with inverse Gaussian distributions.

Maximizing the likelihood with respect to β for the distributions corresponding to $V(\mu) = \mu^p$ is tantamount to maximizing these expressions D_p. Note that in the actual likelihood, an indicator function appears, to reflect that $y \in R_{\phi/w}$. When estimation is done maximizing (2) instead of (1) (possibly also without the range constraints), one speaks of quasi-likelihood estimation. This leads to 'more' distributions than before if multiples of discrete random variables are allowed, such as in the case of so-called overdispersed Poisson random variables with a variance $\sigma^2 > \mu$ (underdispersion is a much less common phenomenon), but also if a mean-variance relationship $V(.)$ is used that does not occur in the exponential family. One example is $V(\mu) = \mu^2(1-\mu)^2$, $0 < \mu < 1$, for which the quasi-likelihood can be computed, but for which there is not an exponential family distribution having this log-likelihood ratio. The quasi-likelihood can be shown to have enough in common with ordinary log-likelihood ratios to allow many asymptotic results to still remain valid. See for instance McCullagh and Nelder (1989), Chapter 9.

The scaled deviance can be regarded as a distance in $\mathbb{R}^n$ between the vector of predictions and the vector of observed values. It is a sum of contributions for each observation taking into account its size, the contribution getting reduced if the observation is large (the larger the observation, the less precise, the less 'credible' it is).

Note that the deviance is measured in terms of the dispersion parameter ϕ. Also, the variance function determines the units of measurement for the deviance, so simply differencing these discrepancy measures across variance functions is not feasible. To compare different variance functions $V(\mu) = \mu^p$ it is necessary to widen the definition of quasi-likelihood. Following Nelder and Pregibon (1987), we look at the *extended* quasi-likelihood, for this case defined as

$$Q_p^+(\mu_1, \ldots, \mu_n, \phi; y_1, \ldots, y_n) = -\frac{1}{2}\sum_{i=1}^{n} \log\left(2\pi\phi y_i^p\right) - \frac{1}{2} D_p(y_1, \ldots, y_n; \mu_1, \ldots, \mu_n)/\phi, \qquad (4)$$

where D_p is as defined above, and ϕ is the dispersion parameter; $\exp(Q_p^+)$ is in fact the unnormalized saddlepoint approximation to the density for exponential families, see Barndorff-Nielsen and Cox (1979).

Note that the estimates for the parameters β obtained by maximizing Q_p^+ as in (4) coincide with the ML-estimates. The estimate of ϕ obtained by setting zero the partial derivative of Q_p^+ with respect to ϕ is the mean deviance.

It is easy to show that, apart from a linear transformation with coefficients not depending on μ, hence $\beta_1, \beta_2, \ldots$, the extended quasi-likelihood Q_p^+ in (4) is exactly equal to the likelihood in case $p = 0$ or $p = 3$ holds. For $p = 1$ and $p = 2$, this is approximately the case; the approximation is that in the likelihood, any factorial $k! = \Gamma(k+1)$ is replaced by Stirling's approximation

$$\Gamma(y+1) \approx (2\pi y)^{1/2}\, y^y\, \mathrm{e}^{-y}.$$

Sometimes, at the boundary with $y = 0$ it is preferable to replace Stirling's approximation by

$$\Gamma(y+1) \approx \left(2\pi\left(y + \frac{1}{6}\right)\right)^{1/2} y^y\, \mathrm{e}^{-y}.$$

See Nelder and Pregibon (1987), Section 4 for details.

4. TWEEDIE'S CLASS OF DISTRIBUTIONS

A subclass of the exponential family of distributions, named Tweedie's class in view of Tweedie (1984), with fixed $b(\cdot)$ and $c(\cdot;\cdot)$ functions and variable θ and ψ in (1) exists such that the variance function is of the form $V(\mu) = \mu^p$ for some exponent $p \in (1, 2)$. It consists of compound distributions. Assume specifically that $Y \sim$ compound Poisson(λ) with gamma(α, β) claim sizes. To get a family of distributions having mean $\lambda\alpha/\beta = \mu$ and variance $\lambda\alpha(\alpha+1)/\beta^2 = \psi\mu^p$ with $\mu > 0$ and $\psi > 0$, it suffices that the parameters satisfy:

$$\lambda = \frac{\mu^{2-p}}{\psi(2-p)}; \qquad \alpha = \frac{2-p}{p-1}; \qquad \frac{1}{\beta} = \psi(p-1)\mu^{p-1}. \qquad (5)$$

Note that all claim sizes have common α, hence the same shape, dictated by the value of p; the mean claim numbers and the scale vary to generate possible (μ, ψ) combinations. Clearly, it is possible to make other choices leading to the same mean-variance relation, but only this one leads to an exponential family subclass as desired.

We will demonstrate that for this particular choice of parameters (5), the mixed continuous/discrete density of Y can be written as in (1), both for $y = 0$ and for $y > 0$. Of course we have $\Pr[Y=0] = \mathrm{e}^{-\lambda}$, as well as

$$f_Y(y) = \mathrm{e}^{-\beta y}\,\mathrm{e}^{-\lambda} \sum_{n=1}^{\infty} \frac{\beta^{n\alpha}}{\Gamma(n\alpha)} y^{n\alpha-1}\,\frac{\lambda^n}{n!}, \qquad y > 0. \tag{6}$$

Now because of the choice in (5), $\lambda\beta^{\alpha}$ does not depend on μ, only on the parameter ψ and the constant p. Therefore the sum in (6) depends on ψ and y, but not on μ. To establish the result that (6) is of the form (1) with 'natural' parameterization, we define $c(y,\psi)$ as the logarithm of that sum, with $c(0,\psi) = 1$, and find θ by equating $-\beta = \theta/\psi$ as well as $\lambda = b(\theta)/\psi$. This gives

$$\theta = -\beta\psi = \frac{-1}{(p-1)\mu^{p-1}}, \quad \text{so} \quad \mu(\theta) = \left(-\theta(p-1)\right)^{-1/(p-1)}, \quad \text{and} \quad b(\theta) = \lambda\psi = \frac{\mu^{2-p}}{2-p}.$$

Note that the cases of a Poisson multiple ($p=1$) and a gamma variate ($p=2$) can be obtained as limits of this class. This fact may be verified by taking limits of the mgfs, or understood as follows. If $p \downarrow 1$, in the limit we get Poisson(μ/ψ) many claims that are degenerate on ψ. If $p \uparrow 2$, for the number of claims N we have $\mathrm{E}[N] = \lambda \to \infty$, as well as $\alpha \downarrow 0$ in such a way that $\lambda\alpha \to 1/\psi$. Replacing $N\alpha$ by $\lambda\alpha$, we see that the resulting limit distribution is the gamma$(1/\psi, 1/(\psi\mu))$ distribution.

Actual distributions in the exponential dispersion family with a variance function $V(\mu) = \mu^p$ exist for all values $p \notin (0,1)$. But for $p \in (0,1)$, still the quasi-likelihood can be maximized to obtain parameter estimates.

Many situations in actuarial statistics lead to observations that can be modeled well by a compound Poisson distribution, with a variance function 'between' $V(\mu) = \mu$ (Poisson) and $V(\mu) = \mu^2$ (gamma). The negative binomial(r,p) distribution (r fixed) also has a variance function with that property, since here we have $V(\mu) = \mu + \mu^2/r$. This is a linear combination on the identity scale; for Tweedie, it is on the log-scale.

5. EXAMPLE: AN IBNR PROBLEM

To illustrate the actuarial use of Tweedie's class of distributions, we generated an IBNR-triangle consisting of Tweedie distributed random outcomes. For this, we needed a way to generate gamma pseudo-random deviates. A fast and simple way of doing this is Algorithm GS, to be found in Ahrens and Dieter (1974). It works as follows. To draw from a gamma$(\varepsilon, 1)$ density

$$f(x) = \frac{1}{\Gamma(\varepsilon)}\, x^{\varepsilon-1}\mathrm{e}^{-x}, \quad x > 0,$$

with $0 < \varepsilon < 1$, it employs a rejection method, using as a majorant

$$q(x) = \frac{\mathrm{e}}{\mathrm{e}+\varepsilon}\varepsilon x^{\varepsilon-1}\mathrm{I}_{(0,1]}(x) + \frac{\varepsilon}{\mathrm{e}+\varepsilon}\mathrm{e}^{-(x-1)}\mathrm{I}_{(0,\infty)}(x-1).$$

It is easy to generate drawings from this mixture of densities. It can be achieved using the same random drawing to decide if the r.v. is < 1 or > 1, as well as its actual outcome. Outcome x from

	$j=1$	2	3	4	5	6	7	8	9	10
$i=1$	4290	3094	1146	1388	294	189	42	11	4	12
2	3053	2789	682	1476	253	101	79	15	8	
3	4389	2709	688	2050	353	266	109	48		
4	4144	2046	1642	1311	549	160	70			
5	2913	4079	1652	2501	395	221				
6	5757	5201	1178	2486	580					
7	4594	3928	1236	2730						
8	3695	3688	1301							
9	3967	4241								
10	4933									

Table 1: An IBNR triangle with Tweedie claim totals

$q(\cdot)$ is accepted as a drawing from $f(\cdot)$ with probability $f(x)/Mq(x) \in [0,1]$, where

$$M = \frac{\mathrm{e} + \varepsilon}{\mathrm{e}\varepsilon\Gamma(\varepsilon)}.$$

To generate a gamma$(k+\varepsilon, 1)$ drawing, we simply add k independent exponential(1) r.v.'s to the result of the above procedure, and a gamma$(k+\varepsilon,\beta)$ drawing of course results by dividing by β. This may not be the fastest way to generate gamma deviates, but it is readily available to everyone, and does not require too much programming effort.

To simulate an IBNR-problem, we used this method to generate drawings from Y_{ij}, $i,j = 1,\ldots,10$, $i+j \leq 11$, having Tweedie distributions with mean $\mu_{ij} = \mu r_i c_j \gamma^{i-1}\delta^{j-1}$ and variance $V(\mu) = \mu_{ij}^p$. The parameter values chosen were $p = 1.5$, $\psi = 2$, $\mu = 1$, $\gamma = 1.03$, $\delta = 0.9$. The r_i were known relative exposures for each row, the c_j given development factors (also in %) for each column in the IBNR-triangle. In fact,

$$r = (100, 110, 115, 120, 130, 135, 130, 140, 130, 120);$$
$$c = (30, 30, 10, 20, 5, 3, 1, 0.5, 0.3, 0.2).$$

The resulting IBNR-triangle is given in Table 1. To estimate the parameters γ, δ and μ we used Stata, with a log-link and with a user-written power variance function $V(\mu) = \mu^p$, generating a deviance D_p as given in (3). Note that this triangle looks convincingly like an incremental IBNR-triangle for total claims such as they occur in practice, except for the fact that it does not exhibit the bothersome negative numbers that most IBNR-models preclude but practice tends to generate anyway. In Table 2, we find the estimation results with some values of p. We list the estimated dispersion coefficient and the value of Q_p^+, computed through (4). Also, we computed the resulting estimate for the IBNR-reserve to be held, which is equal to the sum over $i,j = 1,\ldots,10$, $i+j > 11$ of all predicted values $\hat{\mu}_{ij} = \hat{\mu} r_i c_j \hat{\gamma}^{i-1}\hat{\delta}^{j-1}$, hence for the lower right triangle in Table 1. Note that the value of $\hat{\phi}$ varies very strongly with p, being about right only for p close to the actual value 1.5. The extended QL is maximal for $p \approx 1.8$, but the actual value $p = 1.5$ leads to an acceptable value as well. The reserve to be held is not overly sensitive to the value of p, just as, it turns out, are the parameter estimates $\hat{\gamma}$, $\hat{\delta}$ and $\hat{\mu}$. Observe that the required reserve increases with the exponent p.

p	$\hat{\phi}$	Q_p^+	Reserve
1.0	78.	−399	17287
1.1	36.	−393	17290
1.2	17.	−388	17295
1.3	8.0	−383	17307
1.4	3.8	−379	17329
1.5	1.9	−375	17369
1.6	.94	−372	17434
1.7	.49	−370	17535
1.8	.26	−369	17689
1.9	.15	−370	17912
2.0	.09	−373	18232

Table 2: Estimation results for the data in Table 1

The maximum extended quasi-likelihood estimate of the reserve equals 17689. The ML-estimate will be close to this value, but to compute it, values of (6) would have to be evaluated. While this is certainly doable, we point out that for our computations, we only needed to provide the standard software Stata with a subroutine to compute the quasi-deviance (3) for a power variance function $V(\mu) = \mu^p$, and do the calculations for a few selected values of p.

References

J.R. Ahrens and U. Dieter. Computer methods for sampling from gamma, beta, Poisson, and binomial distributions. *Computing*, 12:223–246, 1974.

O. Barndorff-Nielsen and D.R. Cox. Edgeworth and saddlepoint approximations with statistical applications (and discussion). *Journal of the Royal Statistical Society, B*, 41:279–312, 1979.

S. Haberman and A.E. Renshaw. Generalized linear models and actuarial science. *The Statistician*, 45:41–65, 1996.

R. Kaas, M.J. Goovaerts, J. Dhaene, and M. Denuit. *Modern Actuarial Risk Theory*. Kluwer, Dordrecht, 2001.

P. McCullagh and J.A. Nelder. *Generalized Linear Models*. Chapman and Hall, London, 1989.

J.A. Nelder and D. Pregibon. An extended quasi-likelihood function. *Biometrika*, 74:221–232, 1987.

J.A. Nelder and R.W.M. Wedderburn. Generalized linear models. *Journal of the Royal Statistical Society, A*, 135:370–384, 1972.

M.C.K. Tweedie. An index which distinguishes between some important exponential families. In *Statistics: Applications and New Directions. Proceedings of the Indian Statistical Institute Golden Jubilee International Conference*, pages 579–604, 1984.

STOCHASTIC METHODS FOR PORTFOLIO CREDIT DERIVATIVES

Lutz Schloegl

Fixed Income Quantitative Research, Lehman Brothers International (Europe), 25 Bank Street, London E14 5LE, United Kingdom
Email: `luschloe@lehman.com`

Abstract

Credit derivatives are an important meeting ground for actuarial and financial mathematics. This article is a brief introduction to the pricing of portfolio credit derivatives. We survey some of the stochastic methods currently used. These are illustrated with several of the main applications in portfolio credit derivatives such as the pricing of CDO and CDO^2 tranches.

1. CREDIT AT THE INTERSECTION BETWEEN DERIVATIVES AND INSURANCE

Credit derivatives occupy a unique position at the intersection between derivatives and insurance. The most liquid and basic credit derivative, the default swap, is an insurance contract between two counterparties on the credit risk of a reference entity. The *protection buyer* pays a regular premium until default or maturity of the trade, which is known as the *default swap spread* and is quoted on an annualized basis in basis points, i.e. hundredths of a percent of the trade notional. In return, the *protection seller* protects the buyer against the economic loss on the reference entity's bonds in the event of a default. At default, the contract is either subject to cash or physical settlement. In the case of physical settlement, the protection buyer delivers defaulted bonds to the seller and receives their par value in return. In the case of cash settlement, the protection seller pays the difference between par and the bonds' observed recovery rate, i.e. post-default price to the buyer. This is a typical insurance contract. In return for a (relatively) small premium, the protection buyer insures against a rare, but potentially large loss.

The digital nature of credit payouts highlights risks that are not so central to other derivative contracts, this has caused the market to evolve. Maturity mismatches are an important source of risk, and the market has evolved mitigation mechanisms against this. Protection on a new default swap contract commences at $T + 1$ *calendar* days after the trade date T. This is in contrast to other markets where settlement periods are usually expressed in business days to facilitate clearing and other back-office operations. This is particularly important, as credit relevant information is fairly often revealed when markets are closed. Similarly, default swaps are traded to fixed maturity dates. The so-called "IMM" dates are the 20th of March, June, September and December. This reduces maturity mismatches between different long and short positions and significantly facilitates the management of a default swap book.

Default swaps derive their importance not only from their role as insurance contracts, they are also the basic hedging tool for more complex credit derivatives. The exposures stemming from synthetic CDO tranches, CDO^2 trades, default swaptions, etc are all dynamically managed by credit derivative dealers using default swaps. In the context of a trading book, positions are continually marked to market. This means that sensitivities to spread movements are particularly important, more so as a derivative position can mutate from asset to liability and vice versa as the market moves. The nature of counterparty risk also becomes very different: contracts do not just cancel if the premium is no longer paid. Rather, the contract is marked to market and needs to be unwound. If the non-defaulting party is in the money, the market value of the contract becomes an unsecured claim on the defaulting counterparty. In particular, the seller of protection is also exposed to counterparty risk if the market tightens.

The standard approach to bootstrapping credit curves is to take a completely reduced-form view of the default event. The default time τ is a random variable with a distribution modelled via the hazard rate λ by specifying the conditional default probability as

$$P\left[\tau \leq t + \Delta t \middle| \, \tau > t\right] = \lambda(t)\Delta t. \tag{1}$$

The hazard rate λ is closely related to the credit spread. In the simple approach, it is treated as a deterministic function. More sophisticated models specify λ as a stochastic process. The unconditional survival probability $Q(0,t)$ to time T is obtained by integrating equation (1):

$$Q(0,T) = P[\tau > T] = \exp\left(-\int_0^T \lambda(s)ds\right).$$

On default, we assume that the loss is $1-R$, where R is a fixed recovery of par. The hazard rate is calibrated to default swap spreads, which are the actual market observables. Because spreads aggregate loss likelihood and severity, we need to disentangle these two aspects. The protection leg is priced by integrating

$$\text{Prot} = (1-R)\int_0^T B(0,t)Q(0,t)\lambda(t)dt. \tag{2}$$

Ignoring the issue of coupon accrual, the premium leg is the price of a risky annuity, also known as the risky PV01 (present value of a basis point). It is obtained by summing over the risky discount factors for the coupon dates $T_1,\ldots,T_n$. Denoting the accrual factor for each coupon period by α_i, we have

$$\text{Prem} = \sum_{i=1}^{n} \alpha_i B(0,T_i)Q(0,T_i). \tag{3}$$

The protection and premium leg values are not uniquely determined by the breakeven market spread because

$$s = \frac{\text{Prot}}{\text{Prem}} \approx \frac{(1-R)\sum_{m=1}^{M} B(0,t_m)\left(Q(0,t_{m-1}) - Q(0,t_m)\right)}{\sum_{i=1}^{n} \alpha_i B(0,T_i) Q(0,T_i)}.$$

In particular, different hazard rate curves can give the same default swap spread. Default probabilities are model-dependent quantities that depend quite strongly on the recovery rate assumptions that are used, and to a lesser extent on the interpolation methodology. Nevertheless, no-arbitrage links spreads, recovery rates, and default probabilities. A very useful approximation, the so-called credit triangle, can be derived by assuming a constant hazard rate and a continuously paid spread:

$$s \approx \lambda(1-R).$$

For example, a spread of 90bp and a recovery rate of 40% imply an annual default probability of approximately $\lambda = 1.5\%$. Using the credit triangle, one can compute the market value of protection bought at a spread of s_0:

$$\text{MTM} = (s - s_0) \sum_{i=1}^{n} \alpha_i B(0,T_i) e^{-\frac{sT_i}{1-R}}.$$

The recovery sensitivity of this MTM is quite low, particularly if the current market default swap spread s has not moved far away from s_0. This is good news if one is worried about marking a default swap book correctly. On the other hand, it implies that default swaps do not actually help us in disentangling default and recovery rate risk. The most certain thing about recovery rates is their uncertainty. The best data sources are the rating agencies or perhaps internal ratings, from a credit derivatives perspective one usually has to make fairly broad assumptions, for example a recovery rate of 40% for senior unsecured debt of investment grade companies.

After default swaps, synthetic CDO tranches are the most broadly traded credit derivatives. The insurance character is similar, the protection seller takes exposure to a band of losses to a given reference portfolio. The band is defined by a lower and an upper strike, K_1 and K_2, which are expressed as a percentage of the total portfolio notional. If L_T is the cumulative percentage loss to the portfolio, the percentage loss to the tranche is

$$L_T^{tr} = \frac{[L_T - K_1]^+ - [L_T - K_2]^+}{K_2 - K_1}. \tag{4}$$

The protection seller receives a spread s on the outstanding notional of the tranche. Once the portfolio losses exceed K_1, the seller makes a protection payment every time there is a loss, the notional of the tranche is reduced, and the tranche spread is only paid out on this reduced notional going forward. An important concept for CDO tranche pricing is the so-called tranche "survival probability" $Q^{tr}(0,T)$, which is the expected outstanding notional of the tranche

$$Q^{tr}(0,T) = 1 - \frac{E\left[[L_T - K_1]^+\right] - E\left[[L_T - K_2]^+\right]}{K_2 - K_1}. \tag{5}$$

With this, the protection and premium legs of the tranche swap become analogous to a single-name default swap, see equations (2) and (3):

$$\text{Prot} = -\int_0^T B(0,t)dQ(0,t)$$
$$\text{Prem} = s\sum_{i=1}^{n} \alpha_i B(0,T_i)Q^{tr}(0,T_i).$$

Equation (5) shows that tranche survival probabilities are effectively call spreads on the cumulative portfolio loss. This non-linearity explains why CDO tranches are correlation instruments, they depend not only on the overall risk in the portfolio (determined by $E[L_T]$), but also critically on the tendency of different credits to default (and survive) together. Hence, the main effort when developing portfolio credit models for tranche pricing is directed towards modelling the dependence structure between the different reference credits.

2. STOCHASTIC MODELLING TECHNIQUES

Equation (5) also implies that we need to concentrate on modelling cumulative portfolio loss distributions. We fix a time horizon T, each credit j defaults with probability p_j. One modelling framework is the so-called latent variable approach. With each credit j, we associate a random variable A_j, such that the credit defaults if A_j falls below a threshold K_j. The value of K_j is calibrated to the marginal default probability p_j:

$$P[A_j \leq K_j] = p_j. \tag{6}$$

The marginal distribution of A_j is only used to calibrate the threshold, default dependence is generated by the dependence structure of $A_1, \ldots, A_M$. Given this very general framework, there are still many ways to model the dependence between credits, of which we mention a few. A very popular approach is so-called times-to-default (TtD) modelling. For each credit, the basic modelling object is the random default time τ_j. It is generated by transforming the latent variable A_j via the marginal distribution function F_j. The famous Gaussian copula variant of this model, introduced by Li (2000), is obtained by choosing $A_1, \ldots, A_M$ as multivariate normal and setting

$$\tau_j = F_j^{-1}(\Phi(A_j)).$$

This specification immediately fits into the framework of equation (6) with $K_j = \Phi^{-1}(p_j)$. The TtD approach with a Gaussian copula is very tractable, but it also has fairly severe flaws. The dependence structure between credits is highly time-inhomogeneous, and the model does not produce realistic spread dynamics. It needs to be adjusted to match observed CDO tranche prices, i.e. fit to the "correlation smile". Nevertheless, it has become a de facto market standard. A more dynamic approach is achieved by defining a credit index process X_j for each credit j: default occurs at the first time X_j hits a time dependent barrier. This was originally proposed by Black and Cox (1976) and has more recently been revived by Hull and White (2001). A somewhat different approach uses the Cox process framework. Defaults are generated by jumps of point processes which are given as time changes of independent Poisson processes via dependent stochastic hazard rates. However, to produce realistic CDO tranche spreads, one needs to introduce jumps in the hazard rates or other feedback effects, because correlated diffusions do not generate sufficient levels of default dependence.

Despite the existence of a multitude of modelling approaches, several mathematical ideas have proven themselves very powerful in dealing with the high dimensionality inherent to credit portfolio analysis: conditional independence, asymptotic methods and semi-asymptotic sensitivity calculations.

2.1. Conditional Independence

Conditional independence is the idea that, conditional on some mixing variable η the credits in the portfolio are independent. The simplest example of this is the one-factor Gaussian copula, where each credit's $\mathcal{N}(0,1)$ distributed latent variable A_j is given by

$$A_j = \beta_j \, Z_{Mkt} + \sqrt{1-\beta_j^2}\, Z_j, \tag{7}$$

and the variables $Z_{Mkt}, Z_1, \dots, Z_M$ are i.i.d $\mathcal{N}(0,1)$. Conditional on the market factor Z_{Mkt}, all the credits are independent. In general, the conditional loss distribution is binomial and the unconditional loss distribution is obtained by integrating over η. Particularly in the one-factor framework, this is a straightforward numerical integration. A lot of research effort has been put into finding methods of computing the conditional loss distributions efficiently. Popular techniques include Fourier transforms, recursion techniques, as well as saddlepoint and other analytical approximations. The recursion method is particularly intuitive: for simplicity, we assume that each credit generates the same loss at default, so that the loss distribution can be expressed in integer multiples of an underlying loss unit. Denote the conditional default probability of credit j by p_j. For each $n \in \{0, 1, \dots, M\}$, $L^{(n)}$ is the conditional portfolio loss after n credits have been added to the portfolio, and $p_k^{(n)} = P\left[L^{(n)} = k\right]$. The start of the recursion is clear, because $p_0^{(0)} = 1$ and $p_k^{(0)} = 0$ for $k > 0$. The recursion step is

$$p_{k+1}^{(n)} = p_{k+1}^{(n-1)}(1-p_n) + p_k^{(n-1)} p_n, \tag{8}$$

with $p_0^{(n)} = p_0^{(n-1)}(1-p_n)$. Note that each credit makes a "default" and a "survival" contribution. When pricing a tranche, we can pick out the slice of the loss distribution we are interested in, depending on the strikes of the tranche. To compute sensitivities, we can use equation (8) to quickly unwind a step of the recursion. Because conditional default probabilities can be very small, numerical stability is key. Hence one needs to anchor the recursion either at a zero loss or the maximum loss.

2.2. The LHP Approximation and Semi-Asymptotic Extensions

Another important stochastic tool for the analysis of credit portfolios is the so-called Large Homogeneous Portfolio (LHP) approximation. This was introduced in a Gaussian context by Vasicek (1987), see also Vasicek (2002). In the conditional independence framework, one assumes the portfolio consists of equally-weighted homogeneous assets with default probability $p(\eta)$. As the number M of credits tends to infinity, the fraction of credits defaulting converges to $p(\eta)$ by the Law of Large Numbers. For simplicity, assume that recovery rates are zero, then the fractional loss of the portfolio also converges to $p(\eta)$. In the LHP approximation, we assume that the conditional loss is actually $p(\eta)$. We are replacing the conditional loss distribution with its conditional expectation, i.e. matching the first moment of the conditional loss distribution with a point mass. The great advantage of the LHP approximation is that it often gives analytical formulae and is therefore very useful in developing intuition for a given dependence structure. Since the original work Vasicek (1987), Schönbucher (2004) has applied the LHP approximation to some of the Archimedean copula family, Schloegl and O'Kane (2005) have analyzed the Student-t copula. In the Gaussian case, we start from equation (7). The default probability p and the correlation parameter β are common across all credits, the default threshold C is given by $C = \Phi^{-1}(p)$. Assuming $\beta > 0$, we can compute the unconditional loss distribution to be

$$P\left[L \leq \theta\right] = \Phi\left(\frac{\sqrt{1-\beta^2}\,\Phi^{-1}(\theta) - \Phi^{-1}(p)}{\beta}\right).$$

Because the LHP approximation assumes that the portfolio is homogeneous, it is not well suited to computing sensitivities to changes in individual issuer characteristics. A way to deal with this is via so-called semi-asymptotic methods. In a given portfolio, we model the credit we are interested in exactly, while treating the rest of the portfolio asymptotically. Emmer and Tasche (2003) use this approach to compute risk capital contributions. Lehman Brothers has utilized this method for computing sensitivities in a model we call LHP plus one asset, or LH+ for short, cf. Greenberg et al. (2004). To compute a stop-loss transform $E\left[[L-K]^+\right]$ for a given strike K, it is possible to identify thresholds $A < B$ for the market factor which determine whether the single credit is relevant for crossing the strike or not. Extending the LHP analysis then gives a formula in terms of bi- and trivariate normal distributions.

$$\begin{aligned} E\left[[L-K]^+\right] &= K\Phi_{2,\beta_0}(C_0, A) + (N_0 - K)\,\Phi_{2,\beta_0}(C_0, B) \\ &\quad + N\left[\Phi_{2,\beta}(C, A) + \Phi_{3,\Sigma}(C_0, C, B) - \Phi_{3,\Sigma}(C_0, C, A)\right]. \end{aligned} \tag{9}$$

Computing the spread delta effectively entails differentiating equation (9) with respect to the individual credit's default threshold C_0. This reduces the trivariate normal distribution to a bivariate one, giving a very tractable formula for the spread sensitivity. Further computational enhancements can be achieved by modelling the conditional loss distribution more exactly. One would immediately think of a Central Limit Theorem argument, i.e. matching the first two moments of the conditional loss distribution by fitting a normal distribution, as has been proposed by Finger (1999). However, the Central Limit Theorem can easily be a false friend in credit modelling, as we are dealing with rare events and are often concerned with the tail of the distribution. In fact, we have found that fitting a simple two-point distribution to the conditional loss by matching moments gives better results than a Gaussian approximation in the LH+ context.

2.3. CDO^2 Pricing

An interesting application of the modelling techniques we have discussed is the pricing of CDO^2 structures. The fundamental underlying to such a structure is a large pool of credits, the main constraint here is number of entities liquidly traded in the CDS market. The individual credits are assigned to different miniportfolios. The miniportfolios do overlap and the weighting of a particular credit is specific to each miniportfolio. A bespoke CDO tranche is chosen for each miniportfolio, and these minitranches form the reference set of a new structure, the synthetic CDO^2. Finally, a bespoke supertranche is selected from the squared structure. Of course, losses to the supertranche depend on losses affecting the minitranches, which in turn depend on the joint default behaviour of the individual credits. The seller of supertranche protection covers the losses affecting the supertranche, just as in a standard CDO tranche. Similarly, the contractual spread paid to the seller is based on the outstanding notional of the supertranche. The loss L_i^{tr} to the ith minitranche is a function of the loss to the ith miniportfolio as shown in equation (4). The percentage loss L^{sp} to the superportfolio is given as

$$L^{sp} = \frac{\sum_{i=1}^{N} N^{(i)} L_i^{tr}}{N_{tot}}.$$

Finally, the supertranche loss is given by

$$L^{st} = \frac{[L^{sp} - K_{st}]^+ - [L^{sp} - (K_{st} + w_{st})]^+}{w_{st}}. \tag{10}$$

Equation (10) shows that the supertranche loss is a compound option on the joint distribution of all the miniportfolio losses $L_1, \ldots, L_N$. This poses two challenges. Even if the individual credits are independent, $L_1, \ldots, L_N$ are not because of the overlap of the miniportfolios. Also, even if one has a very tractable model for the joint distribution of $L_1, \ldots, L_N$, one still needs to price a compound option. Naively treating the minitranches as effective CDS with some correlation is doomed to failure, as it is unclear which correlation to use. This is because tranches are highly non-linear payouts and some of the dependence stems from contagion between credits (this is captured by a factor model), whereas another part stems from the overlap between miniportfolios. Finally, the effective correlation is a function of the minitranche subordination. Senior tranches are more exposed to systemic risk, hence behave in a more correlated fashion.

One method to price CDO^2 tranches is to extend the recursion technique to higher dimensions, as shown in Baheti et al. (2005). Let us assume for simplicity that we are in the two-dimensional case, i.e. the superportfolio contains two minitranches. The conditional probability of joint losses after n credits is denoted by $p^n_{k_1,k_2}$. We add credits recursively with default probability π_j and loss weights of λ_i^j. Each credit makes a survival and a default contribution. We have both if $k_1 \geq \lambda_1^{n+1}$ and $k_2 \geq \lambda_2^{n+1}$. In this case the recursion is

$$p^{n+1}_{k_1,k_2} = (1 - \pi_{n+1})\, p^n_{k_1,k_2} + \pi_{n+1}\, p^n_{k_1-\lambda_1^{n+1},k_2-\lambda_2^{n+1}}.$$

Other points of the joint loss distribution only have a survival contribution

$$p^{n+1}_{k_1,k_2} = (1 - \pi_{n+1})\, p^n_{k_1,k_2}.$$

The main limitation of this approach is the fact that the probability space, and hence memory requirements, grow exponentially with the number of miniportfolios. The method is useful for relatively medium scale structures (around 7 minitranches), and also for many different similar products. For dealing with higher dimensional cases, other methods are needed. One approach is to condition on the market factor, simulate the joint loss distribution of $L_1, \ldots, L_N$, and evaluate the compound option via Monte Carlo. The moments of the conditional distribution are easy to calculate, because the individual credits are independent.

$$E\left[L_i \middle| Z\right] = \sum_{k=1}^{M} (1 - R_k) w_{i,k}\, p_k(Z) \tag{11}$$

$$\mathrm{cov}\left(L_i, L_j \middle| Z\right) = \sum_{k=1}^{M} (1 - R_k)^2 w_{i,k}\, w_{j,k}\, p_k(Z) \left(1 - p_k(Z)\right). \tag{12}$$

One way of accelerating the conditional Monte Carlo simulation is to use equation (11) and (12) to fit a multivariate normal distribution to the conditional losses. However, as mentioned before, one has to be wary of the Central Limit Theorem as a false friend in this type of application.

3. A BRIEF OUTLOOK

In the previous section we have detailed some of the stochastic techniques currently useful in credit derivatives pricing and have illustrated some of their applications. We have not discussed the very important topic of the correlation smile, i.e. the deviation between the market pricing of tranches and the simple Gaussian copula. Practitioners are very much searching for models which fit the observed market prices in the best possible manner. Also, in terms of modelling, one really needs to move beyond the simple times-to-default framework. It is too static, induces counterintuitive time-inhomogeneities and hence produces unrealistic spread dynamics. In fact, the challenge to all researchers in the credit derivatives field is to think more dynamically about modelling credit risk and portfolio credit derivatives. This will allow us to develop models which produce more realistic spread dynamics.

References

P. Baheti, R. Mashal, M. Naldi, and L. Schloegl. Squaring factor copula models. *Risk*, to appear, 2005.

F. Black and J.C. Cox. Valuing corporate securities: Some effects of bond indenture provisions. *Journal of Finance*, 31:351–367, 1976.

S. Emmer and D. Tasche. Calculating credit risk capital charges with the one-factor model. working paper, 2003.

C.C. Finger. Conditional approaches for CreditMetrics portfolio distributions. *CreditMetrics Monitor*, pages 14 – 33, April 1999.

A. Greenberg, D. O'Kane, and L. Schloegl. LH+: A fast analytical model for CDO hedging and risk management. *Lehman Brothers Quantitative Credit Research Quarterly*, 2004-Q2, 2004.

J.C. Hull and A. White. Valuing credit default swaps II: Modeling default correlations. *Journal of Derivatives*, 8(3):12–22, Spring 2001.

D.X. Li. On default correlation: a copula function approach. *Journal of Fixed Income*, 9:43–54, 2000.

L. Schloegl and D. O'Kane. A note on the large homogeneous portfolio approximation with the Student-t copula. *Finance and Stochastics*, to appear, 2005.

P.J. Schönbucher. Taken to the limit: Simple and not-so-simple loan loss distributions. *Wilmott magazine*, 2004.

O. Vasicek. Probability of loss on loan portfolio. working paper, KMV Corporation, 1987.

O. Vasicek. Loan portfolio value. *Risk*, 15(12):160–162, December 2002.

CONTRIBUTED TALKS

A JUMPS MODEL FOR CREDIT DERIVATIVES PRICING

Jessica Cariboni[†] **and Wim Schoutens**[§]

[†]*Department of Mathematics, K.U.Leuven, W. De Croylaan 54, B-3001 Leuven, Belgium*
European Commission, JRC, 21020, Ispra (VA), Italy
[§]*Department of Mathematics, K.U.Leuven, W. de Croylaan 54, B-3001 Leuven, Belgium*
Email: `jessica.cariboni@jrc.it, wim.schoutens@wis.kuleuven.ac.be`

Abstract

This work present a structural model for credit derivatives pricing. The firm asset value is assumed to follow the exponential of a Lévy process and default is triggered by the crossing of a predetermined barrier. This approach thus includes asymmetries, fat tails and instantaneous default.

In the case of the variance gamma Lévy process, we show how to price Credit Default Swaps (CDS) par spreads. The pricing is based on the numerical solution of a partial integral differential equation. The model is calibrated to different market CDS term structures.

1. INTRODUCTION

Equity pricing techniques have been used to assess credit risk since the development of structural models by Merton (1974) and by Black and Cox (1976). Merton defines an event of default to occur when the value of equity drops to zero. In contrast, Black and Cox's model default through an exogenous default barrier. With the Merton approach corporate bonds are treated as American style derivatives; in the Black and Cox approach corporate bonds are barrier style products. In the later, refer to as the Gaussian case, the asset value is modeled by a geometric Brownian Motion. Numerous extensions or modifications of both types of models have been developed. An example is given by CreditGrades$^{\mathrm{TM}}$ (RiskMetrics (2002)), which also assumes the asset price to follow a geometric Brownian motion. Default is defined here to occur if the asset value hits a low barrier, which is made stochastic to allow for higher default probabilities. The assumption of Brownian motion does not describe properly the distributions typically observed on the market, which are asymmetric and leptokurtic. Moreover the stochasticity of the barrier leads to the fact that one can have default even before one has started and leads to a unrealistic spread curve for the short-term.

This work presents a structural model where the asset price process is described by an exponential of a pure jumps Lévy process. Default is triggered by the crossing of a predetermined low barrier. Our model takes into account asymmetries and fat-tail behaviors and incorporates instantaneous default through jumps. Lévy based models have already proven their capabilities in equity models (Schoutens (2003)) and fixed income models (Eberlein and Raible (1999)). Other models that add jumps in the dynamics of the firm value are by Zhou (2001), Zhou (1997) and Hilberink and Rogers (2002). The rest of the paper is organized as follows. In the next section, we present Lévy processes, focussing on the VG process. In Section 3, we present our Lévy default model and relate CDS spreads to the prices of binary barrier options. Section 4 reports on numerical experiments, focussing on computational issues, on the sensitivity of the model to its parameters and on the model calibration. The last section concludes.

2. LÉVY SETTINGS

2.1. Lévy processes

Suppose $\phi(z)$ is the characteristic function of a distribution. If for every positive integer n, $\phi(z)$ is also the nth power of a characteristic function, we say that the distribution is infinitely divisible. One can define for every such an infinitely divisible distribution a stochastic process, $X = \{X_t, t \geq 0\}$, called Lévy process, which starts at zero, has independent and stationary increments and such that the distribution of an increment over $[s, s+t]$, $s, t \geq 0$, i.e. $X_{t+s} - X_s$, has $(\phi(z))^t$ as characteristic function.

The function $\psi(z) = \log \phi(z)$ is called the *characteristic exponent* and it satisfies the *Lévy-Khintchine formula* Bertoin (1990):

$$\psi(z) = \mathrm{i}\gamma z - \frac{\varsigma^2}{2} z^2 + \int_{-\infty}^{+\infty} (\exp(\mathrm{i}zx) - 1 - \mathrm{i}zx1_{\{|x|<1\}})\nu(\mathrm{d}x),$$

where $\gamma \in \mathbb{R}$, $\varsigma^2 \geq 0$ and ν is a measure on $\mathbb{R}\backslash\{0\}$ with $\int_{-\infty}^{+\infty}(1 \wedge x^2)\nu(\mathrm{d}x) < \infty$. We say that our infinitely divisible distribution has a triplet of Lévy characteristics $[\gamma, \varsigma^2, \nu(\mathrm{d}x)]$. The measure $\nu(\mathrm{d}x)$ is called the *Lévy measure* of X.

From the Lévy-Khintchine formula, one sees that a Lévy process consists of three independent parts: a linear deterministic part, a Brownian part, and a pure jump part. The Lévy measure $\nu(\mathrm{d}x)$ dictates how the jumps occur. Jumps of sizes in the set A occur according to a Poisson process with parameter $\int_A \nu(\mathrm{d}x)$.

2.2. The VG Process

The characteristic function of the VG law with parameters σ, ν, θ is given by

$$\phi_{VG}(u; \sigma, \nu, \theta) = (1 - \mathrm{i}u\theta\nu + \sigma^2\nu u^2/2)^{-1/\nu}$$

The corresponding Lévy measure ν_{VG} of the VG (σ, ν, θ) law is given by:

$$\nu_{VG}(\mathrm{d}x) \quad = \quad \begin{cases} C\exp(Gx)|x|^{-1}\mathrm{d}x & x < 0 \\ C\exp(-Mx)x^{-1}\mathrm{d}x & x > 0 \end{cases},$$

where

$$\begin{aligned} C &= 1/\nu > 0 \\ G &= \left(\sqrt{\frac{\theta^2\nu^2}{4}+\frac{\sigma^2\nu}{2}}-\frac{\theta\nu}{2}\right)^{-1} > 0 \\ M &= \left(\sqrt{\frac{\theta^2\nu^2}{4}+\frac{\sigma^2\nu}{2}}+\frac{\theta\nu}{2}\right)^{-1} > 0. \end{aligned}$$

The VG process $X^{(VG)} = \{X_t^{(VG)}, t \geq 0\}$ is a Lévy process where the increment $X_{s+t}^{(VG)} - X_s^{(VG)}$ over the time interval $[s, t+s]$ follows a VG$(\sigma, \nu/t, t\theta)$ law:

$$\begin{aligned} E[\exp(\mathrm{i}uX_t^{(VG)})] &= \phi_{VG}(u; \sigma\sqrt{t}, \nu/t, t\theta) \\ &= (\phi_{VG}(u; \sigma, \nu, \theta))^t \\ &= (1 - \mathrm{i}u\theta\nu + \sigma^2\nu u^2/2)^{-t/\nu}. \end{aligned}$$

A VG-process has no Brownian component and its Lévy triplet is given by $[\gamma, 0, \nu_{VG}(\mathrm{d}x)]$, where

$$\gamma = \frac{-C(G(\exp(-M)-1) - M(\exp(-G)-1))}{MG}.$$

When $\theta = 0$ then $G = M$ and the distribution is symmetric. Negative values of θ lead to the case where $G < M$, resulting in negatively skewness. Similarly, the parameter $\nu = 1/C$ primarily controls the kurtosis.

3. CDS PRICING UNDER VG SETTING

We assume we have to our disposal a risk-free bond $B = \{B_t, t \geq 0\}$ with price process

$$B = \{B_t = \exp(rt), t \geq 0\}.$$

Furthermore, we model the firm value, $S = \{S_t, t \geq 0\}$, by an exponential of a VG process. More precisely, we assume the following dynamics

$$S_t = S_0 \exp((r-q)t + X_t + \omega t),$$

where $S_0 > 0$ is the initial asset value, r is the constant continuously compounded interest rate, q is the asset continuous dividend yield and $X = \{X_t, t \geq 0\}$ is a VG process. The risk-neutral drift rate for the asset is $r - q$ and thus, to have $E[S_t] = S_0 \exp((r-q)t)$, we have to set

$$\omega = \nu^{-1} \log\left(1 - \frac{1}{2}\sigma^2\nu - \theta\nu\right).$$

For pricing of equity options under this model see e.g. Madan et al. (1998) or Schoutens (2003) for the European case; the pricing of equity options of an American nature is developped by Hirsa and Madan (2003).

We define a default event to occur the first time the asset value S_t crosses a deterministic barrier H which corresponds to the recovery value R of the firm's debt (cfr. Black and Cox (1976), Leland (1994), Longstaff and Schwartz (1995) and the CreditGrades$^{\text{TM}}$ model (2002)).

The risk-neutral probability of no-default between 0 and t, $P(t)$, is given by:

$$\begin{aligned}
P(t) &= P_Q\left(S_s > H, \text{for all } 0 \leq s \leq t\right); \\
&= P_Q\left(\min_{0\leq s\leq t} S_s > H\right); \\
&= E_Q\left[1\left(\min_{0\leq s\leq t} S_s > H\right)\right]
\end{aligned}$$

where $1(A)$ is equal to 1 if the event A is true and 0 otherwise; the subindex Q refers to the fact that we are working in a risk-neutral setting.

CDS provide an insurance against the defaulting of a company. The buyer of this protection pays a continuous spread, c, to the seller until the maturity T, unless default occurs. In this case, the buyer delivers a bond on the underlying defaulting asset in exchange for its face value. The price of a CDS is given by:

$$CDS = (1-R)\left(-\int_0^T \exp(-rs)\mathrm{d}P(s)\right) - c\int_0^T \exp(-rs)P(s)\mathrm{d}s,$$

The par spread c^* that makes the CDS price equal to zero is:

$$\begin{aligned}
c^* &= \frac{(1-R)\left(-\int_0^T \exp(-rs)\mathrm{d}P(s)\right)}{\int_0^T \exp(-rs)P(s)\mathrm{d}s} \\
&= \frac{(1-R)\left(1-\exp(-rT)P(T) - r\int_0^T \exp(-rs)P(s)\mathrm{d}s\right)}{\int_0^T \exp(-rs)P(s)\mathrm{d}s}
\end{aligned}$$

Let us denote by

$$BDOB(T,L) = \exp(-rT)E_Q\left[1\left(\min_{0\leq s\leq T} S_s > L\right)\right],$$

the price of a binary *down-and-out barrier* option with maturity T and barrier level L; its payout is 1 if S remains above the barrier during the lifetime, 0 otherwise. Since

$$BDOB(T,L) = \exp(-rT)P(t),$$

the par spread c^* can be rewritten in terms of the BDOB prices as:

$$c^* = \frac{(1-R)\left(1 - BDOB(T,L) - r\int_0^T BDOB(s,L)\mathrm{d}s\right)}{\int_0^T BDOB(s,L)\mathrm{d}s}.$$

In Cariboni and Schoutens (2004) it is shown that the price of the BDOB option can be estimated either via Monte Carlo (MC) simulation of the VG process (Schoutens (2003)) or via numerical solution of a partial integral differential equation (PDIE) based on the work of Hirsa and Madan (2003).

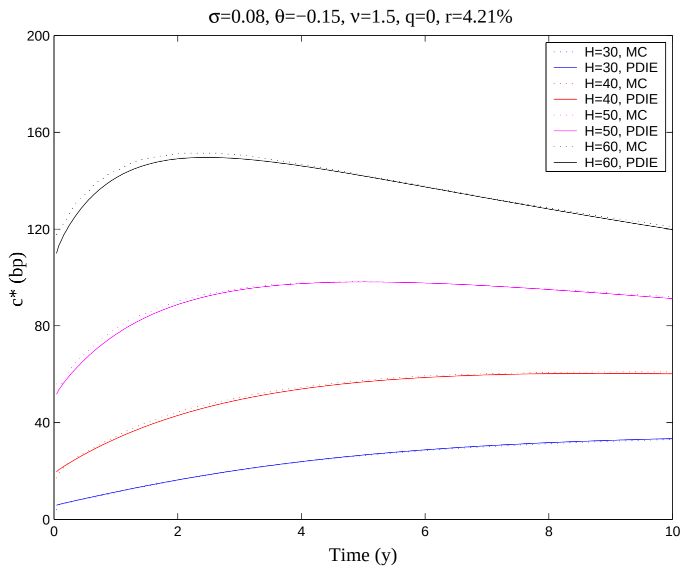


Figure 1: Comparison between MC and PDIE spreads for different values of the barrier.

4. NUMERICAL EXPERIMENTS

We price a BDOB and a CDS spread with time to maturity of $T = 1$ year. We set $S_0 = 100$, $H = 50$, $r = 0.0421$, and $q = 0$. The VG-parameters used are:

$$\sigma = 0.20722, \quad \nu = 0.50215, \quad \theta = -0.22898.$$

For the CDS we assume a recovery rate $R = 0.5$.

Model	M	N	c^* (in bp)	$BDIB$	cpu (in sec)
PDIE VG	100	100	129	0.0245	0.69
PDIE VG	150	150	131	0.0249	1.56
PDIE VG	200	200	132	0.0251	2.94
PDIE VG	250	250	132	0.0252	4.91
PDIE VG	500	250	132	0.0253	9.59
PDIE VG	250	500	132	0.0253	13.37
Model	iterations	N	c^* (in bp)	$BDIB$	cpu (in sec)
MC VG	10000	250	122	0.0233	268
MC VG	100000	250	132	0.0252	2198
MC VG	500000	250	132	0.0251	11040
MC VG	1000000	250	132	0.0253	22059

Table 1: c^* and BDIB prices.

In Table 1 we compare PDIE results with the ones obtained via MC simulation. In the PDIE approach N denotes the number of time steps taken (per year); M denotes the number of points in the log-strike dimension. Taking the MC price obtained by a million iterations as a very good proxy of the true price, meshes sizes greater than $N = 200$ and $M = 200$ for the PDIE algorithm give very accurate results in acceptable cpu times. Figure 1 shows the comparison between MC

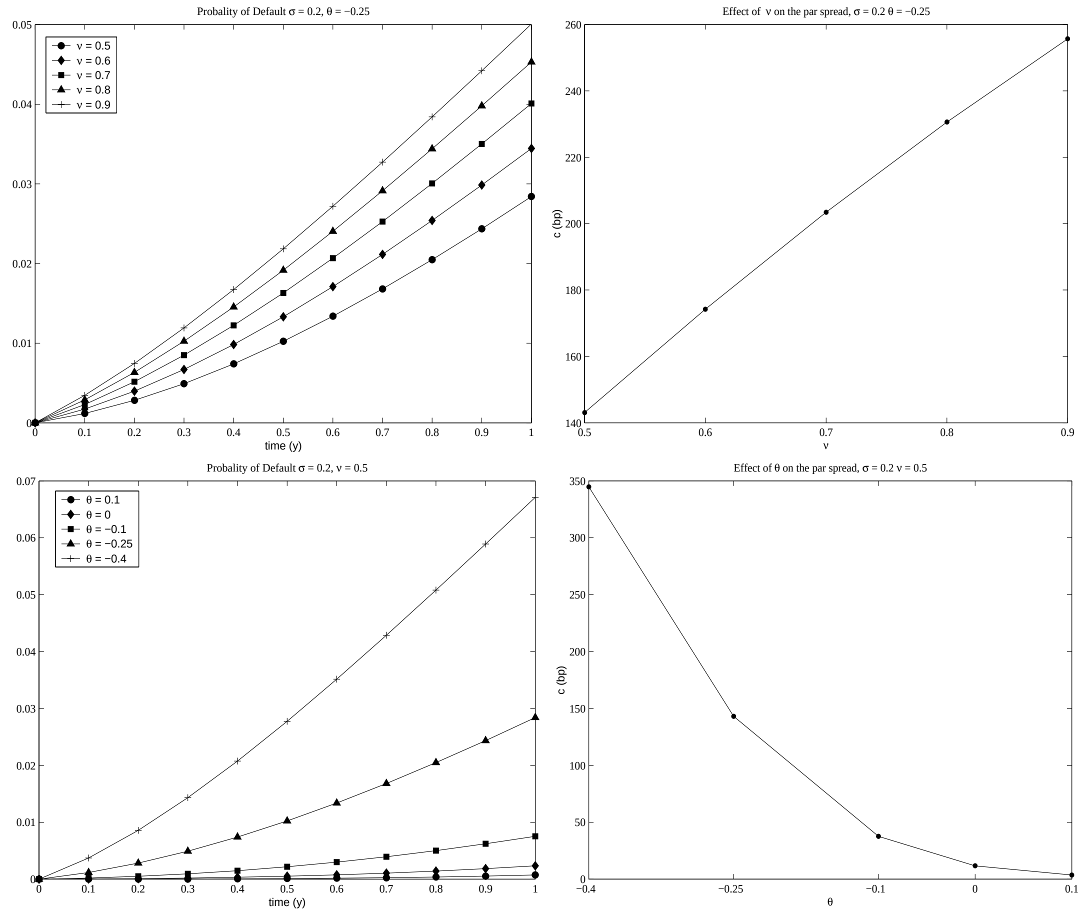


Figure 2: Kurtosis and skewness sensitivity: default probabilities and cds spreads.

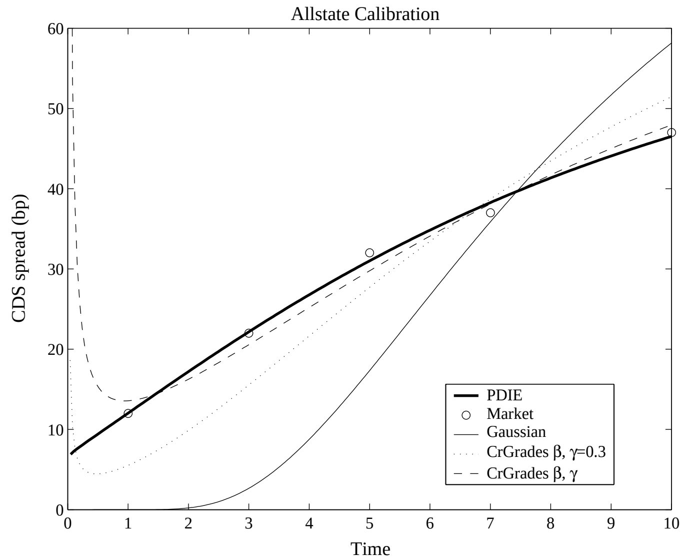


Figure 3: Calibrations for Allstate. The o-signs are the market quotes, the solid bold line is the best fit for the PDIE, the solid is the Gaussian case and the dotted and dashed ones refer to the CreditGrades$^{\text{TM}}$ calibration respectively on $(\beta, \gamma = 0.3)$ and on (β, γ).

and PDIE spreads for a different set of parameters ($\sigma = 0.08$, $\theta = -0.15$ and $\nu = 1.5$) and four different values for the barrier (namely $H = 30, 40, 50, 60$). Higher curves correspond to higher values of the barrier. Solid lines are obtained through the PDIE approach, dotted ones by using MC simulation. The maximum relative difference between MC and PDIE spreads is lower than 1% for all time horizons and all barrier values.

We investigate the sensitivity of the model to the VG parameters by varying the kurtosis parameter ν (Figure 2, top plots) and the skewness parameter θ (Figure 2, bottom plots). Higher kurtosis (i.e. higher ν's) and more negative skewness (i.e. smaller θ's) result in higher default probabilities and higher par spreads.

The model capabilities are finally tested by a calibration exercise to a series of CDS term structures, taken from the market as of the 26th of October 2004 (source Goldman Sachs). The Lévy model (VG), the Gaussian and the CreditGrades$^{\text{TM}}$ have been calibrated, as follows:

- Calibration of the VG model (using PDIE) on the σ, θ and ν VG parameters. Table 2 gives the optimal VG-parameters obtained through the calibration.

- Calibration of CreditGrades$^{\text{TM}}$ on the asset volatility β, which is the only parameter free to vary in the original model. The barrier volatility in set to $\gamma = 30\%$.

- Calibration of CreditGrades$^{\text{TM}}$ model on β and the barrier volatility γ.

- Calibration of the Gaussian model, this corresponds to the CreditGrades$^{\text{TM}}$ model with free β and $\gamma = 0$.

In all cases we use the Nelder-Mead simplex (direct search) method to minimize the difference between market CDS prices in the least-squares sense, i.e. we minimize the root mean square error

Company	σ	ν	θ	$rmse$	ape
Mbna Insurance	0.1141	2.2507	-0.0517	2.33	2.10 %
Wells Fargo	0.0182	2.2513	-0.0609	3.76	7.81 %
Wal-Mart	0.0465	0.4199	-0.1697	2.13	5.08 %
Merrill Lynch	0.1446	2.9404	-0.0008	2.15	2.44 %
Allstate	0.0645	2.0886	-0.0665	1.68	1.96 %
Amgen	0.1528	3.1698	0.0018	1.80	2.40 %
Ford Credit Co.	0.2041	0.9644	-0.0851	2.67	0.58 %
Wyeth	0.0111	0.7712	-0.1716	7.22	3.94 %
Autozone	0.2080	1.0109	0.0060	3.92	1.65 %
Bombardier	0.3553	2.8132	-0.0824	10.62	1.00 %

Table 2: Optimal VG parameters from the calibration on market CDS term structure.

($rmse$). Results are given in Table 3, which lists the market CDS spreads and the optimal values obtained from the calibrations. We also compute the ape, an overall measure of the fit quality:

$$ape = \frac{1}{\text{mean CDS spread}} \sum_{spreads} \frac{|\text{Market spread} - \text{Model spread}|}{\text{number of spreads}}$$

VG fits are almost always better than CreditGrades$^{\text{TM}}$ fits, especially for short time periods. It is also clear that the calibration of the Gaussian case is completely missing the market feature. Finally the CreditGrades$^{\text{TM}}$ calibration on two parameters (β and γ) typically results in better fits, but is problematic for very short maturities. Figure 3 plots the results of the calibrations for the Allstate insurance company. The market quotes are represented by the o-signs, the bold line is the VG best fit while the other lines show the CreditGrades$^{\text{TM}}$ calibrations. Specifically the solid line refers to the calibration in the gaussian case while the dotted and dashed lines refer respectively to the calibration on β and the calibration on β and γ.

5. CONCLUSIONS

In this work we have proposed a new structural model to price credit derivatives. We have assumed the asset price to follow the exponential of a pure-jump Lévy process and defined an event of credit default as the first crossing of the firm value of a preset barrier. The underlying distribution in Lévy models can be asymmetric and leptokurtic, matching typically observed empirical asset distributions. Moreover, the presence of jumps allows for unexpected default, which is instead introduced artificially, e.g. via a stochastic barrier, in continuous path models. Under this structural model, we have calculated the survival probability for the firm by relating it to the price of a binary down-and-out barrier option. For the variance gamma process we have tested the capabilities of our model by pricing credit default swaps. Compared to Monte Carlo, PDIE is also accurate and computationally much cheaper. The model has been calibrated to a series of market credit default structures, resulting in good matches of observed structure curves.

Company	**Moody**		$1y$	$3y$	$5y$	$7y$	$10y$
Mbna Insurance	Aaa	Market	21	36	46	51	61
		VG (PDIE)	21	35	46	53	60
		CG Model(β)	7	23	40	54	70
		CG Model(β, γ)	23	32	43	53	63
		Gaussian	0	6	29	52	77
Wells Fargo	Aa1	Market	3	10	20	23	32
		VG (PDIE)	5	10	17	24	33
		CG Model(β)	4	9	17	24	33
		CG Model(β, γ)	5	10	17	24	32
		Gaussian	0	1	8	21	39
Wal-Mart	Aa2	Market	1	9	17	22	32
		VG (PDIE)	1	8	17	24	31
		CG Model(β)	1	9	17	24	33
		CG Model(β, γ)	1	10	17	24	33
		Gaussian	0	1	7	27	37
Merrill Lynch	Aa3	Market	11	20	31	36	47
		VG (PDIE)	11	20	30	37	47
		CG Model(β)	5	15	27	38	50
		CG Model(β, γ)	12	19	29	37	47
		Gaussian	0	3	17	35	57
Allstate	A1	Market	12	22	32	37	47
		VG (PDIE)	12	22	31	38	47
		CG Model(β)	6	16	28	39	51
		CG Model(β, γ)	14	21	30	38	48
		Gaussian	0	3	17	36	58
Amgen	A2	Market	14	20	29	34	39
		VG (PDIE)	13	21	28	34	39
		CG Model(β)	5	13	24	34	46
		CG Model(β, γ)	15	19	26	33	41
		Gaussian	0	2	14	30	51
Ford Credit Co.	A3	Market	75	154	203	225	238
		VG (PDIE)	75	155	201	225	239
		CG Model(β)	49	151	204	223	245
		CG Model(β, γ)	73	158	202	224	237
		Gaussian	4	120	202	240	262
Wyeth	Baa1	Market	15	47	75	85	95
		VG (PDIE)	18	47	70	85	99
		CG Model(β)	12	41	67	86	104
		CG Model(β, γ)	19	45	68	85	100
		Gaussian	0	18	58	89	117
Autozone	Baa2	Market	25	65	102	117	127
		VG (PDIE)	24	67	99	117	127
		CG Model(β)	18	61	95	118	136
		CG Model(β, γ)	27	66	96	116	133
		Gaussian	0	34	88	123	151
Bombardier	Baa3	Market	320	405	425	425	425
		VG (PDIE)	322	398	426	432	422
		CG Model(β)	183	400	455	467	465
		CG Model(β, γ)	321	400	425	429	425
		Gaussian	74	384	466	485	485

Table 3: VG (PDIE), Gaussian and CreditGrades^{TM} (CG) best fits on market CDS term structures. Source: Goldman Sachs.

References

J. Bertoin. *Lévy Processes*. Number 121 in Cambridge Tracts in Mathematics. Cambridge University Press, Cambridge, 1990.

F. Black and J. Cox. The pricing of options and corporate liabilities. *Journal of Finance*, 31: 351–367, 1976.

J. Cariboni and W. Schoutens. Pricing credit default swaps under Lévy models. Technical Report 07, University Centre for Statistics, K.U.Leuven, 2004.

E. Eberlein and S. Raible. Term structure models driven by general Lévy processes. *Mathematical Finance*, 9(1):31–53, 1999.

B. Hilberink and L.C.G. Rogers. Optimal capital structure and endogenous default, finance and stochastics. *Finance and Stochastics*, 6:237–263, 2002.

A. Hirsa and D.B. Madan. Pricing American options under Variance Gamma. *Journal of Computational Finance*, 7(2), 2003.

H. Leland. Corporate debt value, bond convenants, and optimal capital structure. *Journal of Finance*, 49:1213–1252, 1994.

F. Longstaff and E. Schwartz. A simple approach to valuing risky fixed and floating rate debtebt value, bond convenants, and optimal capital structure. *Journal of Finance*, 50:789–819, 1995.

D.B. Madan, P. Carr, and E.C. Chang. The variance gamma process and option pricing. *European Finance Review*, 2:79–105, 1998.

R. Merton. On the pricing of corporate debt: the risk structure of interest rates. *Journal of Finance*, 29:449–470, 1974.

RiskMetrics. CreditGrades$^{\mathrm{tm}}$. Technical report, RiskMetrics Group, Inc, 2002.

W. Schoutens. *Lévy Processes in Finance: Pricing Financial Derivatives*. Wiley, 2003.

C. Zhou. A jump-diffusion approach to modelling credit risk and valuing defaultable securities. Technical report, Federal Reserve Board, Washington, 1997.

C. Zhou. The term structure of credit spreads with jump risk. *Journal of Banking and Finance*, 25: 2015–2040, 2001.

DISCRETE S-CONVEX EXTREMA, WITH APPLICATIONS IN ACTUARIAL SCIENCE

Cindy Courtois[†], Michel Denuit[†§] and Sébastien Van Bellegem[§]

[†]*Institut des Sciences Actuarielles, Univeristé catholique de Louvain, rue des Wallons 6, B-1348 Louvain-la-Neuve, Belgium*
[§]*Institut de Statistiques, Université catholique de Louvain, Voie du Roman Pays 20, B-1348 Louvain-la-Neuve, Belgium*
Email: `courtois@stat.ucl.ac.be`

Abstract

The purpose of this work is to derive expressions for the s-convex extrema in moment spaces when the support is discrete. As the s-convex maxima and minima are known up to $s = 4$ and $s = 3$ respectively, the aim is more precisely to derive the 4-convex minimum and the extrema for $s > 4$.
As an application, we use these extremal distributions to bound several quantities of interest in actuarial science, like the eventual probability of ruin in the compound binomial process.

1. INTRODUCTION

It is well established that the theory of stochastic orderings has a considerable interest in probability for theoretical and practical purposes (see, e.g., Oluyede (2004) and Shaked and Shanthikumar (1994)). For instance, it can be used to compare complex models with more tractable ones which are "riskier", leading thus to more conservative decisions.

In many situations, stochastic order relations are used to compare *real* random variables. Quite recently, various *discrete* stochastic orderings have been introduced to compare random variables that are *discrete* by nature as counts for instance (see, e.g., Fishburn and Lavalle (1995), Lefèvre and Picard (1993) and Lefèvre and Utev (1996)). A remarkable class investigated by Denuit and Lefèvre (1997) is the class of the *discrete s-convex orderings* among arithmetic random variables valued in some set $\mathcal{N}_n = \{0, 1, 2, \ldots, n\}$, $n \in \mathbb{N}$. Here s is any nonnegative integer smaller or equal to n.

Discrete s-convex orderings have been defined in Denuit and Lefèvre (1997) in the following way. Let Δ be the first order forward difference operator (with unitary increment) defined for each function $u : \mathcal{N}_n \to \mathbb{R}$ by $\Delta u(i) = u(i + 1) - u(i)$ for all $i \in \mathcal{N}_{n-1}$. Let Δ^k, $k \in \mathcal{N}_n$, be the

k-th order forward difference operator defined recursively by $\Delta^k u(i) = \Delta^{k-1}u(i+1) - \Delta^{k-1}u(i)$ for all $i \in \mathcal{N}_{n-k}$ (by convention, $\Delta^1 u \equiv \Delta u$ and $\Delta^0 u \equiv u$). If X and Y are two random variables valued in $\mathcal{N}_n$, X is said to be smaller than Y with respect to the discrete s-convex order if $\mathbb{E}\left[u(X)\right] \leq \mathbb{E}\left[u(Y)\right]$ for all $u \in \mathcal{U}_{s-cx}^{\mathcal{N}_n} = \{u : \mathcal{N}_n \to \mathbb{R} : \Delta^s u(i) \geq 0,\ \forall\ i \in \mathcal{N}_{n-s}\}$. In such a case, we write $X \preceq_{s-cx}^{\mathcal{N}_n} Y$.

Since the power functions $x \mapsto x^k$ and $x \mapsto -x^k$ both belong to $\mathcal{U}_{s-cx}^{\mathcal{N}_n}$ for $k = 1, 2, \ldots, s-1$, we immediately get the necessary condition

$$X \preceq_{s-cx}^{\mathcal{N}_n} Y \Rightarrow \mathbb{E}X^k = \mathbb{E}Y^k \text{ for } k = 1, 2, \ldots, s-1.$$

In other words, if $X \preceq_{s-cx}^{\mathcal{N}_n} Y$ then the $s-1$ first moments of X and Y necessarily match. Consequently, the ordering relation $\preceq_{s-cx}^{\mathcal{N}_n}$ can only be used to compare the random variables with the same first $s-1$ moments. This motivates to introduce the *moment space* $\mathcal{D}_s\left(\mathcal{N}_n; \mu_1, \mu_2, \ldots, \mu_{s-1}\right)$ which contains all random variables valued on $\mathcal{N}_n$ such that the first $s-1$ moments are fixed to $\mathbb{E}X^k = \mu_k$, $k = 1, \ldots, s-1$, where s is a prescribed nonnegative integer. One remarkable property of s-convex orderings is the following: Provided that the moment space satisfies some reasonable conditions (in particular this space is not void), the moment space contains a minimum random variable $X_{\min}^{(s)}$ and a maximum random variable $X_{\max}^{(s)}$ with respect to $\preceq_{s-cx}^{\mathcal{N}_n}$.

However, the proof of this existence result is implicit in the sense that a formula for $X_{\min}^{(s)}$ and $X_{\max}^{(s)}$ cannot be found easily, except in the simplest cases that we recall now.

If $s = 3$, the extrema $X_{\min}^{(3)}$ and $X_{\max}^{(3)}$ have been derived in Denuit and Lefèvre (1997). Let ξ_1 and ξ_2 be the integers in $\mathcal{N}_{n-1}$ such that $\xi_1 < \mu_2/\mu_1 \leq \xi_1 + 1$ and $\xi_2 < (n - \mu_1)^{-1}(n\mu_1 - \mu_2) \leq \xi_2 + 1$. Then the discrete 3-convex extremal distributions are given by

$$X_{\min}^{(3)} = \begin{cases} 0 & \text{with probability } p_1 = 1 - p_2 - p_3, \\ \xi_1 & \text{with probability } p_2 = \frac{(\xi_1+1)\mu_1 - \mu_2}{\xi_1}, \\ \xi_1 + 1 & \text{with probability } p_3 = \frac{\mu_2 - \xi_1\mu_1}{1+\xi_1}, \end{cases} \tag{1}$$

and

$$X_{\max}^{(3)} = \begin{cases} \xi_2 & \text{with probability } q_1 = \frac{(1+\xi_2)(n-\mu_1)+\mu_2-n\mu_1}{n-\xi_2}, \\ \xi_2 + 1 & \text{with probability } q_2 = \frac{(n+\xi_2)\mu_1 - \mu_2 - n\xi_2}{n-1-\xi_2}, \\ n & \text{with probability } q_3 = 1 - q_1 - q_2. \end{cases} \tag{2}$$

The proof of this result can be found in Denuit and Lefèvre (1997) and uses the theory of discrete Tchebycheff systems (see, e.g. Karlin and Studden (1966)).

If $s = 4$, the same argument is used in Denuit et al. (1999b) to derive the explicit formula for $X_{\max}^{(4)}$. Let ζ be the integers in $[0, n-2]$ such that $\zeta < (n\mu_1 - \mu_2)^{-1}(n\mu_2 - \mu_3) \leq \zeta + 1$. Then,

$$X_{\max}^{(4)} = \begin{cases} 0 & \text{with probability } v_1 = 1 - v_2 - v_3 - v_4, \\ \zeta & \text{with probability } v_2 = \frac{n\mu_1(\zeta+1) - \mu_2(\zeta+1+n)+\mu_3}{\zeta(n-\zeta)}, \\ \zeta + 1 & \text{with probability } v_3 = \frac{\mu_2(\zeta+n) - n\mu_1\zeta - \mu_3}{(\zeta+1)(n-\zeta-1)}, \\ n & \text{with probability } v_4 = \frac{\mu_3 - \mu_2(2\zeta+1)+\mu_1\zeta(\zeta+1)}{n(n-\zeta)(n-\zeta-1)}. \end{cases} \tag{3}$$

Surprisingly, no explicit formula for $X_{\min}^{(4)}$ is available in the literature. The point is that the argument based on the non-negativity of particular moment matrices is no longer valid for that case. The same phenomenon appears for the derivation of $X_{\min}^{(s)}$ or $X_{\max}^{(s)}$ with $s \geq 5$. In that sense the theory of discrete s-convex extremal distribution is limited to the case $s \leq 3$ and is partially solved for $s = 4$.

The present paper aims to go beyond this limitation and proposes new arguments, based on the so-called "majorant-minorant method" and the "cut-criterion", that allows to derive the explicit extremal distributions for all s. However these cases are far more complicated to deal with because a subtle discussion about the points of support of the extremal distribution is needed.

To illustrate that point, it is interesting to notice the close connection between the extrema (1)–(3) and the corresponding *continuous* extrema, for which a parallel theory is developed when the support of the random variable is the interval $[0, n]$. For instance, let us consider the case of $X_{\min}^{(3)}$. It can be shown (see Denuit et al. (1999a)) that the continuous 3-convex minimal distribution is given by

$$X_{\min}^{\text{cont.}(3)} = \begin{cases} 0 & \text{with probability } 1 - p, \\ \mu_2/\mu_1 & \text{with probability } p = \mu_1^2/\mu_2. \end{cases} \tag{4}$$

A comparison between (1) and (4) leads to the conclusion that the discrete extremal distribution can be easily obtained from the corresponding continuous extremal distributions since the probability mass $p = \mu_1^2/\mu_2$ of the continuous distribution is spread on $\xi, \xi + 1 \in \mathcal{N}_n$ such that $\xi < \mu_2/\mu_1 \leq \xi + 1$. This phenomenon also arises if we compare the discrete extremal distributions (2), (3) with their corresponding continuous extremal distribution. It is then tempting to conjecture that all discrete extrema can be obtained from their continuous extrema. This would be a right strategy to solve our problem since an explicit formula for continuous extremal distributions can be written for all s.

Surprisingly, this conjecture is wrong, as we can show with a simple example. Consider for instance the moment space fixed by the moments $(\mu_0, \mu_1, \mu_2, \mu_3) = (1, 6.625, 44.8525, 313.78825)$. One can see that the corresponding continuous 4-convex minimum is given by

$$X^{\text{cont.}} = \begin{cases} 6.4 & \text{with probability } 0.95, \\ 10.9 & \text{with probability } 0.05. \end{cases}$$

Using the theory that we develop in the present article, one can show that the discrete 4-convex minimum on $\mathcal{N}_n$ is given by

$$X^{\text{disc.}} = \begin{cases} 6 & \text{with probability } 0.490875, \\ 7 & \text{with probability } 0.487025, \\ 12 & \text{with probability } 0.016725, \\ 13 & \text{with probability } 0.005375. \end{cases}$$

In other words, the support of the discrete distribution does not appear as the neighbourhood in $\mathcal{N}_n$ of the supports of the continuous distribution. Moreover, if we discretize the continuous extremal distribution on the neighbouring support $\{6, 7, 10, 11\}$ one can see that the "probability mass" at 10 would be negative (-0.0794).

This example shows that it is challenging to find the form of the support of the discrete extremal distribution. This question is addressed in Section 2 of the article. In Subsection 2.1 we focus on the so-called "majorant/minorant method" to find the s-convex extrema. This section contains key results that characterize the discrete moment space. Then Subsection 2.2 recalls the cut-criterion Denuit and Lefèvre (1997). Subsection 2.3 derives the support of the 4-convex minimum.

Section 3 deals with an application of this theory. We compute lower and upper bounds for the probability of extinction in a Galton-Watson branching process and for the Lundberg's coefficient in the classical insurance risk model with discrete claim amounts.

Finally, Section 4 gives some conclusions as well as the generalization of the method developed in the paper to find the s-convex extrema for $s \geq 4$.

2. DERIVATION OF THE 4-CONVEX MINIMUM

2.1. S-convex extrema in moment spaces

As announced, random variables are assumed to take values on the state space $\mathcal{N}_n = \{0, 1, 2, \ldots, n\}$ for some non-negative integer n. We denote by $\mathcal{D}_s\left(\mathcal{N}_n; \mu_1, \mu_2, \ldots, \mu_{s-1}\right)$ the moment space of all the random variables valued in $\mathcal{N}_n$ and with prescribed first $s-1$ moments $\mu_k = \mathbb{E}X^k$, $k = 1, \ldots, s-1$. Henceforth, the moment sequence $(\mu_1, \mu_2, \ldots, \mu_{s-1})$ is supposed to be such that $\mathcal{D}_s\left(\mathcal{N}_n; \mu_1, \mu_2, \ldots, \mu_{s-1}\right)$ is non void (for conditions, see De Vylder (1996)).

We aim to derive random variables $X_{\min}^{(s)}$ and $X_{\max}^{(s)}$ belonging to $\mathcal{D}_s\left(\mathcal{N}_n; \mu_1, \mu_2, \ldots, \mu_{s-1}\right)$ and such that

$$X_{\min}^{(s)} \preceq_{s-cx}^{\mathcal{N}_n} X \preceq_{s-cx}^{\mathcal{N}_n} X_{\max}^{(s)} \text{ for all } X \in \mathcal{D}_s\left(\mathcal{N}_n; \mu_1, \mu_2, \ldots, \mu_{s-1}\right). \tag{5}$$

The determination of $X_{\min}^{(s)}$ and $X_{\max}^{(s)}$ involved in (5) has been discussed in Denuit and Lefèvre (1997)-Denuit et al. (1999b): using the cut-criterion on distribution functions (see Proposition 2.3 below), the extrema for $s = 1, 2, 3$ and the maximum for $s = 4$ were obtained explicitly. In this paper, using a method that we call the *Majorant/Minorant Method* (inspired from the so-called *method of admissible measures* in Kemperman (1987)), we find the form of the support of the 4-convex minimum.

Instead of solving (5) directly, we first look for the random variables that achieve the bounds

$$\max_{X \in \mathcal{D}_s(\mathcal{N}_n; \mu_1, \mu_2, \ldots, \mu_{s-1})} \mathbb{E}\left[X^s\right] \text{ and } \min_{X \in \mathcal{D}_s(\mathcal{N}_n; \mu_1, \mu_2, \ldots, \mu_{s-1})} \mathbb{E}\left[X^s\right]. \tag{6}$$

The extrema $X_{\min}^{(s)}$ and $X_{\max}^{(s)}$ necessarily achieve the bounds in (6).

Let us consider the problem of finding the random variables that realize the bounds in (6). We have the following result.

Property 2.1

(i) A random variable $X \in \mathcal{D}_s\left(\mathcal{N}_n; \mu_1, \mu_2, \ldots, \mu_{s-1}\right)$ achieves the maximum (6) if and only if X is sup-admissible, that is X is concentrated on the set

$$\left\{i \in \mathcal{N}_n : i^s = c_0 + c_1 \cdot i + c_2 \cdot i^2 + \cdots + c_{s-1} \cdot i^{s-1}\right\}$$

where the c_i's are real constants such that

$$i^s \leq c_0 + c_1 \cdot i + c_2 \cdot i^2 + \cdots + c_{s-1} \cdot i^{s-1}, \text{ for all } i \in \mathcal{N}_n.$$

(ii) *A random variable $X \in \mathcal{D}_s(\mathcal{N}_n; \mu_1, \mu_2, \ldots, \mu_{s-1})$ achieves the minimum (6) if and only if X is sub-admissible, that is X is concentrated on the set*

$$\left\{i \in \mathcal{N}_n : i^s = c_0 + c_1 \cdot i + c_2 \cdot i^2 + \cdots + c_{s-1} \cdot i^{s-1}\right\}$$

where the c_i's are real constants such that

$$i^s \geq c_0 + c_1 \cdot i + c_2 \cdot i^2 + \cdots + c_{s-1} \cdot i^{s-1}, \text{ for all } i \in \mathcal{N}_n.$$

Proof. We only prove (i); the proof for (ii) is similar.

Sufficient condition. Henceforth, we adopt the convention that $0^0 = 1$. Let X be a random variable in $\mathcal{D}_s(\mathcal{N}_n; \mu_1, \mu_2, \ldots, \mu_{s-1})$, i.e. $\sum_{i=0}^{n} \mathbb{P}[X = i]\, i^k = \mu_k \quad, k = 0, 1, \ldots, s-1$; which is concentrated on the set $\left\{i \in \mathcal{N}_n : i^s = \sum_{k=0}^{s-1} c_k i^k\right\}$, where the c_i's are real constants such that $i^s \leq \sum_{k=0}^{s-1} c_k i^k$ for all $i \in \mathcal{N}_n$. Let also Z be some random variable in $\mathcal{D}_s(\mathcal{N}_n; \mu_1, \mu_2, \ldots, \mu_{s-1})$, i.e. $\sum_{i=0}^{n} \mathbb{P}[Z = i]\, i^k = \mu_k \quad, k = 0, 1, \ldots, s-1$. We have

$$\begin{aligned}
\mathbb{E}[X^s] &= \sum_{i=0}^{n} \mathbb{P}[X = i]\, i^s = \sum_{i=0}^{n} \mathbb{P}[X = i] \sum_{k=0}^{s-1} c_k i^k = \sum_{k=0}^{s-1} c_k \sum_{i=0}^{n} \mathbb{P}[X = i]\, i^k \\
&= \sum_{k=0}^{s-1} c_k \mu_k = \sum_{k=0}^{s-1} c_k \sum_{i=0}^{n} \mathbb{P}[Z = i]\, i^k = \sum_{i=0}^{n} \mathbb{P}[Z = i] \sum_{k=0}^{s-1} c_k i^k \\
&\geq \sum_{i=0}^{n} \mathbb{P}[Z = i]\, i^s = \mathbb{E}[Z^s]
\end{aligned}$$

for all $Z \in \mathcal{D}_s(\mathcal{N}_n; \mu_1, \mu_2, \ldots, \mu_{s-1})$. So, $X = \arg\max_{Z \in \mathcal{D}_s(\mathcal{N}_n; \mu_1, \mu_2, \ldots, \mu_{s-1})} \mathbb{E}[Z^s]$.

Necessary condition. Let $X = \arg\max_{Z \in \mathcal{D}_s(\mathcal{N}_n; \mu_1, \mu_2, \ldots, \mu_{s-1})} \mathbb{E}[Z^s]$ and let us suppose that X is the s-convex maximum, i.e. $Z \preceq_{s-cx}^{\mathcal{N}_n} X$ for all $Z \in \mathcal{D}_s(\mathcal{N}_n; \mu_1, \mu_2, \ldots, \mu_{s-1})$. If X is not sup-admissible, by Kemperman (1987) there exists a sup-admissible random variable $Y \neq_d X$ such that $Y = \arg\max_{Z \in \mathcal{D}_s(\mathcal{N}_n; \mu_1, \mu_2, \ldots, \mu_{s-1})} \mathbb{E}[Z^s]$, which is impossible by Proposition 3.3 of Denuit et al. (1999b). Let us now prove by absurd that X is the s-convex maximum. If not, there exists some random variable $Y \in \mathcal{D}_s(\mathcal{N}_n; \mu_1, \mu_2, \ldots, \mu_{s-1})$, $Y \neq_d X$, such that $X \preceq_{s-cx}^{\mathcal{N}_n} Y$. By Proposition 3.1 of Denuit and Lefèvre (1997), it comes particularly that $\mathbb{E}[X^s] \leq \mathbb{E}[Y^s]$, which is impossible and ends the proof. $\square$

We even have the following result that enables us to identify the s-convex extrema with the random variables realizing the bounds (6). The discrete s-convex extrema are thus easily identified using Property 2.1.

Proposition 2.2 *Let X be some random variable in $\mathcal{D}_s(\mathcal{N}_n; \mu_1, \mu_2, \ldots, \mu_{s-1})$. Then X is the s-convex maximum (resp. minimum) if and only if $X = \arg\max_{Z \in \mathcal{D}_s(\mathcal{N}_n; \mu_1, \mu_2, \ldots, \mu_{s-1})} \mathbb{E}[Z^s]$ (resp. $X = \arg\min_{Z \in \mathcal{D}_s(\mathcal{N}_n; \mu_1, \mu_2, \ldots, \mu_{s-1})} \mathbb{E}[Z^s]$).*

Proof. The necessary condition has already been proved in the the proof of the necessary part of Property 2.1 and the sufficient condition is obvious using Proposition 3.1 of Denuit and Lefèvre (1997). $\square$

2.2. Cut-criterion

We now recall the cut-criterion on the distribution functions of Denuit and Lefèvre (1997) that allows us to compare two random variables in the s-convex sense.

Let u be any real-valued function defined on a subset $\mathcal{S}$ of $\mathbb{R}$. We introduce the operator S^- which, when applied to u, counts the number of sign changes of u over its domain S. More precisely, $S^-(u) = \sup S^-\left[u(x_1), u(x_2), \ldots, u(x_n)\right]$ where the supremum is extended over all $x_1 < x_2 < \ldots < x_n \in \mathcal{S}$, n is arbitrary but finite and $S^-\left[y_1, y_2, \ldots, y_n\right]$ denotes the number of sign changes of the indicated sequence $\{y_1, y_2, \ldots, y_n\}$, zero terms being discarded. The functions u_1 and u_2 are said to cross each other k times ($k = 0, 1, 2, \ldots$) if $S^-\left(u_1 - u_2\right) = k$. Moreover, if X and Y are random variables valued in $\mathcal{N}_n$ with respective distribution functions F_X and F_Y, we say that $F_X \geq F_Y$ near n if $F_X(k) \geq F_Y(k)$ for all $k \geq k_0$, with $k_0 \leq n-1$.

Proposition 2.3 (Denuit and Lefèvre (1997)) *Let X and Y be two random variables valued in $\mathcal{N}_n$, such that $\mathbb{E}\left[X^k\right] = \mathbb{E}\left[Y^k\right]$ for $k = 1, \ldots, s-1$. Then, $S^-(F_X - F_Y) \leq s-1$ together with $F_X \geq F_Y$ near $n \Rightarrow X \preceq_{s-cx}^{\mathcal{N}_n} Y$.*

2.3. Support of the 4-convex minimum

Using the cut-criterion, it can be verified that the possible structure of the supports of the 4-convex discrete extrema takes the form $\{\xi, \xi+1, \eta, \eta+1\}$ or $\{0, \zeta, \zeta+1, n\}$. It is interesting to note that those supports are identical to the ones that could be obtained calling upon the theory of the discrete Tchebycheff systems (see Karlin and Studden (1966)). The Majorant/Minorant Method is then used to derive the conditions on the support points ξ, η and ζ so that the random variable corresponding to such support has moments $\mu_1, \mu_2, \ldots, \mu_{s-1}$. This is done by computing the probabilities associated to the support points as solutions to some Vandermonde system and by checking that the resulting probabilities are positive.

Property 2.4 *Consider a moment space $\mathcal{D}_4\left(\mathcal{N}_n; \mu_1, \mu_2, \mu_3\right)$ with a given sequence of moments μ_1, μ_2, μ_3. If $\xi, \eta \in \mathcal{N}_n$ are such that $0 \leq \xi < \xi+1 < \eta < \eta+1 \leq n$ and define*

$$\begin{aligned}
\alpha_1 &:= -\mu_3 + \mu_2\left(2\eta + \xi + 2\right) - \mu_1\left[\left(\xi+1\right)\eta + \left(\xi+1\right)\left(\eta+1\right) + \eta\left(\eta+1\right)\right] + \left(\xi+1\right)\eta\left(\eta+1\right)\\
\alpha_2 &:= \mu_3 - \mu_2\left(\xi + 2\eta + 1\right) + \mu_1\left[\xi\eta + \xi\left(\eta+1\right) + \eta\left(\eta+1\right)\right] - \xi\eta\left(\eta+1\right)\\
\alpha_3 &:= -\mu_3 + \mu_2\left(2\xi + 2 + \eta\right) - \mu_1\left[\xi\left(\xi+1\right) + \xi\left(\eta+1\right) + \left(\xi+1\right)\left(\eta+1\right)\right] + \xi\left(\xi+1\right)\left(\eta+1\right)\\
\alpha_4 &:= \mu_3 - \mu_2\left(2\xi + 1 + \eta\right) + \mu_1\left[\xi\left(\xi+1\right) + \xi\eta + \eta\left(\xi+1\right)\right] - \xi\left(\xi+1\right)\eta
\end{aligned}$$

that are positive, then the discrete 4-convex minimal distribution of $\mathcal{D}_4(\mathcal{N}_n; \mu_1, \mu_2, \mu_3)$ is given by

$$X_{\min}^{(4)} = \begin{cases} \xi & \text{with probability } w_1 = \alpha_1 / \left(\eta - \xi\right)\left(\eta + 1 - \xi\right), \\ \xi + 1 & \text{with probability } w_2 = \alpha_2 / \left(\eta - \xi - 1\right)\left(\eta - \xi\right), \\ \eta & \text{with probability } w_3 = \alpha_3 / \left(\eta - \xi\right)\left(\eta - \xi - 1\right), \\ \eta + 1 & \text{with probability } w_4 = \alpha_4 / \left(\eta + 1 - \xi\right)\left(\eta - \xi\right). \end{cases}$$

Proof. The proof gives the minimal together with the maximal distribution (3). Using the majorant/minorant method, we find out the respective supports of the 4-convex extrema $X_{\max}^{(4)}$ and $X_{\min}^{(4)}$. To that end, we just compute the polynomials $p(i) = c_0 + c_1 i + c_2 i^2 + c_3 i^3$ of degree 3 (i.e. c_0, c_1, c_2 and $c_3 \in \mathbb{R}$) such that $X_{\max}^{(4)} \in \mathcal{D}_4\left(\mathcal{N}_n; \mu_1, \mu_2, \mu_3\right)$ (*resp.* $X_{\min}^{(4)}$) is concentrated on the set

$$\left\{i \in \mathcal{N}_n : i^4 = c_0 + c_1 i + c_2 i^2 + c_3 i^3\right\} = \{0, \zeta, \zeta+1, n\} \, (1 \leq \zeta \leq n-2)$$

$$(\textit{ resp. } \{\xi, \xi+1, \eta, \eta+1\} \, (0 \leq \xi < \xi+1 < \eta < \eta+1 \leq n))$$

and $i^3 \leq c_0 + c_1 i + c_2 i^2$ for all $i \in \mathcal{N}_n$ (*resp.* $\geq$).

The only polynomial of degree 3 that fulfills the conditions

$$\begin{aligned}
&0 = c_0 \\
&\zeta^4 = c_0 + c_1 \zeta + c_2 \zeta^2 + c_3 \zeta^3 \\
&(\zeta+1)^4 = c_0 + c_1 (\zeta+1) + c_2 (\zeta+1)^2 + c_3 (\zeta+1)^3 \\
&n^4 = c_0 + c_1 n + c_2 n^2 + c_3 n^3
\end{aligned}$$

is $p(i) = \zeta(\zeta+1) n i - [n(\zeta+1) + \zeta(\zeta+1) + n\zeta] i^2 + (\zeta + \zeta + 1 + n) i^3$. The zeros of the polynomial $x^4 - p(x)$ are of course 0, ζ, $\zeta+1$ and n and $x^4 - p(x)$ is always negative on $\mathcal{N}_n$. So, as we have checked that $i^4 \leq p(i)$ on $\mathcal{N}_n$, the random variable with support $\{0, \zeta, \zeta+1, n\}$ $(1 \leq \zeta \leq n-2)$ has to be $X_{\max}^{(4)}$.

The only polynomial of degree 3 that fulfills the conditions

$$\begin{aligned}
&\xi^4 = c_0 + c_1 \xi + c_2 \xi^2 + c_3 \xi^3 \\
&(\xi+1)^4 = c_0 + c_1 (\xi+1) + c_2 (\xi+1)^2 + c_3 (\xi+1)^3 \\
&\eta^4 = c_0 + c_1 \eta + c_2 \eta^2 + c_3 \eta^3 \\
&(\eta+1)^4 = c_0 + c_1 (\eta+1) + c_2 (\eta+1)^2 + c_3 (\eta+1)^3
\end{aligned}$$

is

$$\begin{aligned}
p(i) \;=\; & -\xi(\xi+1)\eta(\eta+1) \\
& + [(\xi+\xi+1)\eta(\eta+1) + \xi(\xi+1)(\eta+\eta+1)]\, i \\
& - [\eta(\eta+1) + (\xi+1)(\eta+1) + (\xi+1)\eta + \xi(\eta+1) + \xi\eta + \xi(\xi+1)]\, i^2 \\
& + (\xi+\xi+1+\eta+\eta+1)\, i^3
\end{aligned}$$

The zeros of the polynomial $x^4 - p(x)$ are of course ξ, $\xi+1$, η and $\eta+1$ and $x^4 - p(x)$ is always positive on $\mathcal{N}_n$. So, as we have checked that $i^4 \geq p(i)$ on $\mathcal{N}_n$, the random variable with support $\{\xi, \xi+1, \eta, \eta+1\}$ $(0 \leq \xi < \xi+1 < \eta < \eta+1 \leq n)$ has to be $X_{\min}^{(4)}$.

Finally, we have to fix conditions on the support points to assure the non-negativity of their associated probabilities. The conditions on the support points of $X_{\max}^{(4)}$ are

$$\begin{aligned}
&0 < \zeta < \zeta+1 < n \\
&\mu_3 \leq -\zeta n \mu_1 + (\zeta+n)\mu_2 \\
&\mu_3 \leq \zeta(\zeta+1)n - [\zeta(\zeta+1) + n(\zeta+1) + n\zeta]\mu_1 + (\zeta+\zeta+1+n)\mu_2 \\
&\mu_3 \geq -\zeta(\zeta+1)\mu_1 + (\zeta+\zeta+1)\mu_2 \\
&\mu_3 \geq -(\zeta+1)n\mu_1 + (\zeta+1+n)\mu_2
\end{aligned}$$

and because we have $\zeta\left(\zeta+1\right)n-\left[\zeta\left(\zeta+1\right)+n\left(\zeta+1\right)+n\zeta\right]i+\left(\zeta+\zeta+1+n\right)i^2\geq i^3$ (cfr. 3-convex maximum) on $\mathcal{N}_n$ and $-\zeta\left(\zeta+1\right)i+\left(\zeta+\zeta+1\right)i^2\leq i^3$ on $\mathcal{N}_n$ (cfr. 3-convex minimum), the second and the third condition are respectively always verified and the system of conditions reduces to

$$0<\zeta<\zeta+1<n \text{ and } \zeta<\frac{n\mu_2-\mu_3}{n\mu_1-\mu_2}\leq\zeta+1.$$

Henceforth, we refind the 4-convex maximum (3). The conditions on the support points of $X_{\text{min}}^{(4)}$ are given by

$$\alpha_1\geq 0,\alpha_2\geq 0,\alpha_3\geq 0 \text{ and } \alpha_4\geq 0. \tag{7}$$

□

The solution (ξ,η) of (7) cannot be obtained explicitly. Nevertheless, it is easily obtained by testing each admissible pair (ξ,η) of $\mathcal{N}_n$.

3. APPLICATIONS

3.1. Theorical background

Given a random variable N valued in $\mathcal{N}_n$, n being a positive integer, a classical problem consists in solving the equation

$$\phi_N(z)=P_k(z), \tag{8}$$

in the unknown z, where $\phi_N(z)=\mathbb{E}\left[\mathrm{e}^{zN}\right]$ is the moment generating function of N, and where $P_k\left(\cdot\right)$ is a given non-decreasing polynomial function of degree k (usually, $k\leq 2$). When all that is known about N is that it belongs to $\mathcal{D}_s\left(\mathcal{N}_n;\mu_1,\mu_2,\ldots,\mu_{s-1}\right)$, then (8) cannot be solved explicitly. The aim of this subsection is to show that the s-convex extrema described previously allow accurate approximations for the solution of (8). The method using the continuous s-convex extrema could of course be applied here. Nevertheless, we get better bounds if we take into account the fact that N is now valued in the arithmetic grid $\mathcal{N}_n$ rather than in the interval $[0,n]$ (see Tables 1 and 2). The idea is to construct two functions $\phi_{\text{min}}^{(s)}\left(\cdot\right)$ and $\phi_{\text{max}}^{(s)}\left(\cdot\right)$ such that

$$\phi_{\text{min}}^{(s)}(z)\leq\phi_N(z)\leq\phi_{\text{max}}^{(s)}(z).$$

The sequence $\left[\mathrm{e}^{kz},\ k\in\mathbb{N}\right]$ being absolutely monotonic, we get from Denuit and Lefèvre (1997) that, $\phi_{\text{min}}^{(s)}(t)\leq\phi_N^{(s)}(t)\leq\phi_{\text{max}}^{(s)}(t)$ with $\phi_{\text{min}}^{(s)}(t)=\phi_{N_{\text{min}}^{(s)}}(t)$ and $\phi_{\text{max}}^{(s)}(t)=\phi_{N_{\text{max}}^{(s)}}(t)$, where the $N_{\text{min}}^{(s)}$ and $N_{\text{min}}^{(s)}$ are the stochastic extrema in $\mathcal{D}_s\left(\mathcal{N}_n;\mu_1,\mu_2,\ldots,\mu_{s-1}\right)$ with respect to the discrete versions of the s-convex stochastic orderings. These provide bounds on the root of the equation $\phi_N(z)=P_k(z)$, where P_k is a monotone polynomial function. Solving the equation $\phi_{\text{min}}^{(s)}(z)=P_k(z)$ yields the root $z_1^{(s)}$, say, and solving $\phi_{\text{max}}^{(s)}(z)=P_k(z)$ yields the root $z_2^{(s)}$, say. The solution $\tilde{z}$, say, of $\phi_N^{(s)}(z)=P_k(z)$ then satisfies $z_2^{(s)}\leq\tilde{z}\leq z_1^{(s)}$.

3.2. Lundberg's coefficient

In the classical discrete risk model, the discrete claim amounts $X_1, X_2, \ldots$ recorded by an insurance company are assumed to be independent and identically distributed with common distribution function F having finite $s-1$ moments, such that $F(0) = 0$. The number of claims in the time interval $[0, t]$ is assumed to be independent of the individual claim amounts and to form a Poisson process $\{N(t), t \geq 0\}$ with constant rate λ. Let the premium rate $c > 0$ be such that the inequality $c > \lambda\mathbb{E}[X_1]$ holds. Further, let $\psi(z)$ be the ultimate ruin probability with an initial capital z; that is, the probability that the process $Z(t) = \kappa + ct - \sum_{i=1}^{N(t)} X_i$, $t \geq 0$, describing the wealth of the insurance company, ever falls below zero. If the moment generating function of X exists, Lundberg's inequality provides an exponential upper bound on ψ, namely $\psi(\kappa) \leq \mathrm{e}^{-z\kappa}$, where z is the Lundberg's adjustment coefficient satisfying the integral equation $\phi_X(z) = 1 + \frac{cz}{\lambda}$.

As an illustration, let $n = 5$, $c = 12$, $\lambda = 10$ and $\mu_1 = 1$. First, consider $z^{(s)}_{\min}$ and $z^{(s)}_{\max}$ as functions of μ_2. When then get the numerical values depicted in Table 1. Second, let us fix $\mu_2 = 3$ and consider $z^{(s)}_{\min}$ and $z^{(s)}_{\max}$ as functions of μ_3 (see Table 2). It is seen that the bounds are quite accurate, and are particularly so when μ_2 is large.

μ_2	1.5	2	2.5	3
$z^{(3)}_{\min}$, continuous	0.2104138	0.1596485	0.1305692	0.1111426
$z^{(3)}_{\max}$, continuous	0.2361328	0.1771006	0.1416797	0.1180644
$z^{(3)}_{\min}$, discrete	0.2144848	0.1624468	0.1324108	0.1123238
$z^{(3)}_{\max}$, discrete	0.2330329	0.1771006	0.1409982	0.1180644
μ_2	3.5	4	4.5	
$z^{(3)}_{\min}$, continuous	0.09705668	0.08630042	0.07777624	
$z^{(3)}_{\max}$, continuous	0.1011998	0.08855031	0.07871084	
$z^{(3)}_{\min}$, discrete	0.09778207	0.08670383	0.07794723	
$z^{(3)}_{\max}$, discrete	0.1009502	0.08855031	0.07859318	

Table 1: Bounds on the Lundberg's coefficient z when $\mu_1 = 1$, $n = 5$, $c = 12$ and $\lambda = 10$.

The graph depicted in Figure 1 shows the exponential upper bounds on $\psi(\kappa)$ using the continuous and discrete 4-convex extrema and taking $n = 5$, $c = 12$, $\lambda = 10$, $\mu_1 = 1$, $\mu_2 = 3$ and $\mu_3 = 10$.

4. CONCLUDING REMARKS AND EXTENSION TO $S \geq 4$

Quite surprisingly, the discrete s-convex extrema cannot be obtained by discretizing the continuous ones (contrarily to the cases treated in Denuit and Lefèvre (1997)-Denuit et al. (1999b)). Using the Majorant/Minorant Method, we proved that the support of the discrete 4-convex minimum has to be of the form $\{\xi, \xi+1, \eta, \eta+1\}$ $(0 \leq \xi < \xi+1 < \eta < \eta+1 \leq n)$, when ξ and η are the solutions of (7).

μ_3	9.5	10	10.5	11
$z^{(4)}_{\min}$, continuous	0.1172383	0.1164471	0.1156892	0.1149623
$z^{(4)}_{\max}$, continuous	0.1173398	0.1166221	0.1159113	0.1152081
$z^{(4)}_{\min}$, discrete	0.1172558	0.1164697	0.1157054	0.1149623
$z^{(4)}_{\max}$, discrete	0.117302	0.1165591	0.1158351	0.1151295
μ_3	11.5	12	12.5	
$z^{(4)}_{\min}$, continuous	0.1142642	0.1135929	0.1129466	
$z^{(4)}_{\max}$, continuous	0.1145123	0.1138239	0.1131428	
$z^{(4)}_{\min}$, discrete	0.1142785	0.1136114	0.11296	
$z^{(4)}_{\max}$, discrete	0.1144	0.1136898	0.1129981	

Table 2: Bounds on the Lundberg's coefficient z when $\mu_1 = 1$, $\mu_2 = 3$, $n = 5$, $c = 12$ and $\lambda = 10$.

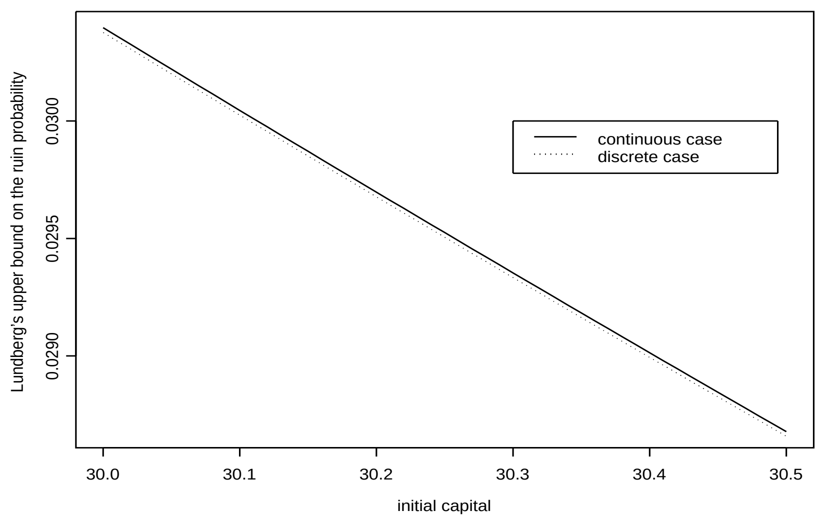


Figure 1: *Upper bound on* $\psi(\kappa)$ *taking* $n = 5$, $c = 12$, $\lambda = 10$, $\mu_1 = 1$, $\mu_2 = 3$ *and* $\mu_3 = 10$.

It is also interesting to note that the method proposed in this paper can be extended to any $s \geq 4$. It is done in the following way. Using the cut-criterion and Property 2.1, it can be seen that the most general form for the supports of the s-convex extrema, denoted by $\mathrm{Supp}_{X_{\min}^{(s)}}$ and $\mathrm{Supp}_{X_{\max}^{(s)}}$, are given as follows: for $s = 2m$, we have $\mathrm{Supp}_{X_{\min}^{(s)}} = \{\xi_1, \xi_1+1, \ldots, \xi_m, \xi_m+1\}$ ($0 \leq \xi_1 < \xi_1+1 < \ldots < \xi_m < \xi_m+1 \leq n$) and $\mathrm{Supp}_{X_{\max}^{(s)}} = \{0, \zeta_1, \zeta_1+1, \ldots, \zeta_{m-1}, \zeta_{m-1}+1, n\}$ ($0 < \zeta_1 < \zeta_1+1 < \ldots < \zeta_{m-1} < \zeta_{m-1}+1 < n$) while for $s = 2m+1$, we have $\mathrm{Supp}_{X_{\min}^{(s)}} = \{0, \xi_1, \xi_1+1, \ldots, \xi_m, \xi_m+1\}$ ($0 < \xi_1 < \xi_1+1 < \ldots < \xi_m < \xi_m+1 \leq n$) and $\mathrm{Supp}_{X_{\max}^{(s)}} = \{\zeta_1, \zeta_1+1, \ldots, \zeta_m, \zeta_m+1, n\}$ ($0 \leq \zeta_1 < \zeta_1+1 < \ldots < \zeta_{m-1} < \zeta_{m-1}+1 < n$).

Then, to express the conditions on the support points so that $X_{\min}^{(s)}$ and $X_{\max}^{(s)}$ have the required moments $\mu_1, \mu_2, \ldots, \mu_{s-1}$, we just have to compute the probabilities associated to the support points and to check that they are positive. We get the resulting probabilities using that

$$X \in \mathcal{D}_s(\mathcal{N}_n; \mu_1, \mu_2, \ldots, \mu_{s-1}) \text{ with } \mathrm{Supp}_X = \{a_0, a_1, \ldots, a_k\}$$

$$\Rightarrow \mathbb{P}[X = a_i] = \frac{\mathbb{E}\left[\prod_{j \neq i}(X - a_j)\right]}{\prod_{j \neq i}(a_i - a_j)} \; (i = 0, 1, \ldots, k).$$

The solution $(\xi_1, \ldots, \xi_{s/2}, \zeta_1, \ldots, \zeta_{(s/2)-1})$ (s even) (*resp.* $(\xi_1, \ldots, \xi_{(s-1)/2}, \zeta_1, \ldots, \zeta_{(s-1)/2})$ (s odd)) cannot be obtained explicitly. Nevertheless, it is easily obtained just by testing each admissible sequence $(\xi_1, \ldots, \xi_{s/2}, \zeta_1, \ldots, \zeta_{(s/2)-1})$ (*resp.* $(\xi_1, \ldots, \xi_{(s-1)/2}, \zeta_1, \ldots, \zeta_{(s-1)/2})$) of $\mathcal{N}_n$.

References

F.E. De Vylder. *Advanced Risk Theory. A Self-Contained Introduction*. Editions de l'Université Libre de Bruxelles - Swiss Association of Actuaries, Bruxelles, 1996.

M. Denuit, F.E. De Vylder, and Cl. Lefèvre. Extremal generators and extremal distributions for the continuous *s*-convex stochastic orderings. *Insurance: Mathematics and Economics*, 24: 201–217, 1999a.

M. Denuit and Cl. Lefèvre. Some new classes of stochastic order relations among arithmetic random variables, with applications in actuarial sciences. *Insurance: Mathematics and Economics*, 20:197–214, 1997.

M. Denuit, Cl. Lefèvre, and M. Mesfioui. On s-convex stochastic extrema for arithmetic risks. *Insurance: Mathematics and Economics*, 25:143–155, 1999b.

P.C. Fishburn and I.H. Lavalle. Stochastic dominance on unidimensional grid. *Mathematics of Operations Research*, 20:513–525, 1995.

S. Karlin and W.J. Studden. *Tchebycheff Systems: with Applications in Analysis and Statistics*. Wiley, New York, 1966.

J.H.B. Kemperman. Geometry of moment problem. *Proceedings of Symposia in Applied Mathematics*, 37:16–53, 1987.

Cl. Lefèvre and Ph. Picard. An unusual stochastic order relation with some applications in sampling and epidemic theory. *Advances in Applied Probability*, 25:63–81, 1993.

Cl. Lefèvre and S. Utev. Comparing sums of exchangeable Bernoulli random variables. *Journal of Applied Probability*, 33:285–310, 1996.

B.O. Oluyede. Inequalities and comparisons of the Cauchy, Gauss, and logistic distributions. *Applied Mathematics Letters*, 17:77–82, 2004.

M. Shaked and J.G. Shanthikumar. *Stochastic Orders and their Applications*. Academic Press, New York, 1994.

ACTUARIAL PRICING FOR A GUARANTEED MINIMUM DEATH BENEFIT IN UNIT-LINKED LIFE INSURANCE: A MULTI-PERIOD CAPITAL ALLOCATION PROBLEM

S. Desmedt[†], X. Chenut[†] and J.F. Walhin[†§]

[†]*Secura, Avenue des Nerviens 9-31, B-1040 Brussels, Belgium*
[§]*Institut de Sciences Actuarielles, Université Catholique de Louvain, 6 rue des Wallons, B-1348 Louvain-la-Neuve, Belgium*
Email: `stijn.desmedt@secura-re.com`

Abstract

We analyze an actuarial approach for the pricing and reserving of a guaranteed minimum death benefit (GMDB) in unit-linked life insurance. We explain two possible strategies to deal with such type of multi-period capital allocation problems. The first one uses no future information whereas the second one does. We explain how a cash-flow model can be used to perform the actuarial pricing and summarize a simulation strategy which can be used to derive approximate distribution functions for the future reserves, capitals and total solvency levels in the approach where future information is used. We test the simulation strategy and obtain useful information about the risk of a GMDB within the actuarial reserving strategy.

1. INTRODUCTION

When pricing and reserving for guarantees in unit-linked life insurance, one traditionally makes a distinction between the so-called financial (see e.g. Brennan and Schwartz (1976)) and actuarial approach. Under the actuarial approach, one does not apply a financial hedging strategy. Instead, capital is allocated to ensure for the necessary security. In the financial approach, pricing is done in the Black-Scholes-Merton framework. As mentioned in Hardy (2003), three types of potential costs are not incorporated in the risk-neutral price: transaction costs, hedging errors arising from discrete hedging intervals and additional hedging costs arising from the fact that log-returns are not normally distributed with fixed μ and σ. Based upon a cash-flow model, in the actuarial approach, the price is equal to the discounted average costs in the real world plus a loading for the capital.

Since a GMDB typically creates liabilities for multiple years, it is possible to review reserves and capitals when new information about the underlying asset and the mortality becomes available. We refer to IAA (2004), where it is advised to take future information into account on a regular basis of e.g. one year when assessing the solvency of long-term risks.

We describe two reserving and capital allocation strategies. The first takes no future information into account. Hence, at time 0, the reserves and the capitals which will be kept in the future are fixed. This strategy has some computational advantages but does not seem very rational. By taking the initial capital sufficiently large, a lot of security can be foreseen, but from the results for the second strategy, we will see it is very likely we will maintain an amount of capital which is too large in the future when applying the first strategy. Hence, in the cash-flow model, one will end up with a premium which is unnecessarily high. The second strategy takes future information into account on a yearly basis. Such a strategy seems more rational but is quite cumbersome to calculate. We briefly explain and verify an approximate simulation method to make the calculation of this strategy possible. Our approach allows us to estimate distribution functions for the future reserves and total solvency levels. This provides a lot of information about the risk which is borne by the (re)insurer reserving in an actuarial way.

Instead of working with a log-normal model for the underlying asset, as in Frantz et al. (2003), we use a more realistic regime-switching log-normal model with two regimes as described in Hardy (2001).

The remainder of this paper is organized as follows. In section 2, we show how one can deal with mono-period capital allocation problems using risk measures. We then suggest two multi-period capital allocation strategies in section 3. Section 4 explains how actuarial pricing can be performed using a cash-flow model. In section 5, we discuss how we modelled the unit-linked contract with GMDB and we briefly explain a simulation strategy for using future information. Finally, we compare the two multi-period strategies for the GMDB in section 6, where we also analyze the approximated distribution functions of the total solvency levels and the reserves in the approach using future information. We conclude in section 7.

2. MONO-PERIOD CAPITAL ALLOCATION

Assume we are exposed to a random loss X. For the moment we assume X is an insurance loss to which one is exposed for only one period and we do not take into account discounting. We define the pure premium *PP* as the average $E[X]$ of X and assume this pure premium is held as a reserve. For risky business, the pure premium itself will of course not guarantee that at the end of the period, there is enough security to withstand the losses which may occur. Therefore, companies exposed to risks need to hold capital as a safety margin against possible bad outcomes. We denote the capital as K. We define the total solvency level *TSL* as the sum of the reserve and the capital. This *TSL* should be sufficiently large such that future losses can be paid with a high probability.

Both in theory and in practice, risk measures are gaining more and more interest for assessing total solvency levels. We first define some well-known risk measures:

Definition 2.1 (Value-at-Risk) *For any $p \in (0, 1)$, the VaR at level p is defined and denoted by*

$$Q_p[X] = \inf\{x \in \mathbb{R} | F_X(x) \geq p\},$$

where $F_X(x)$ denotes the distribution function of X.

The *VaR* at level p of a risk X can be interpreted as the value at which there is only $1 - p$ % probability that the risk will have an outcome larger than that value.

Definition 2.2 (Tail Value-at-Risk) *For any $p \in (0,1)$, the TVaR at level p is defined and denoted by*

$$TVaR_p[X] = \frac{1}{1-p}\int_p^1 Q_q[X]dq.$$

Definition 2.3 (Conditional Tail Expectation) *For any $p \in (0,1)$, the CTE at level p is defined and denoted by*

$$CTE_p[X] = E[X|X > Q_p[X]]. \tag{1}$$

The *CTE* at level p of a loss X can be interpreted as the average of all possible outcomes of X which are above $Q_p[X]$. It is well-known that the *TVaR* and the *CTE* are the same for all p-levels if the distribution function of X is continuous.

Some interesting references on risk measures and their properties and uses are Dhaene et al. (2003), Dhaene et al. (2004a), Dhaene et al. (2004b) and Goovaerts et al. (2004).

3. MULTI-PERIOD CAPITAL ALLOCATION

We first introduce some notation. We denote 0 as the time at which the insurance is written and T as the time at which the last liabilities are possible. Suppose we are exposed to a risk

$$X = (X_1, \ldots, X_T),$$

where X_i is the outcome for the risk at the end of year t, for $t \in \{1, \ldots, T\}$. Denote the risk-free return for year t as Y_t. We define

$$Z = (Z_0 = X_1 e^{-Y_1}, \ldots, Z_{T-1} = X_T e^{-\sum_{t=1}^{T} Y_t}).$$

Z is the vector of all future yearly losses, discounted to the start of the first year. Furthermore, we introduce

$$D = (D_0 = \sum_{t=0}^{T-1} Z_t, D_1 = \sum_{t=1}^{T-1} Z_t e^{Y_1}, \ldots, D_{T-1} = Z_{T-1} e^{\sum_{t=1}^{T-1} Y_t}),$$

the vector of the discounted future costs at the start of the different years $t \in \{1, \ldots, T\}$.

As time passes, important information may become available. In the context of a GMDB, this information consists of the value of the underlying asset, the number of survivors and the return on the investments. Suppose the information is described by a filtration $\mathbb{F} = (\mathcal{F}_t)_{t\in[0,T]}$. We assume a time period of one year after which reserves and capitals may change.

For the ease of later computations (see section 4), we assume all payments are made at the end of the year in which the costs occur.

3.1. Approach Not Using Future Information

At time 0, we hold the pure premium

$$PP = E[D_0|\mathcal{F}_0] = R_0,$$

as a reserve. In addition, we could decide to hold a capital

$$K_0 = \rho[D_0 - R_0|\mathcal{F}_0],$$

where ρ is a risk measure. When not using future information, we define the future reserves as:

$$R_t = E[D_t|\mathcal{F}_0], \quad t \in \{1, \ldots, T-1\},$$

so the reserve at the start of year t is equal to the average discounted future costs at time $t-1$, given the information at time 0. Using the same risk measure as at time 0, and not incorporating future information, we can define

$$K_t = \rho[D_t - R_t|\mathcal{F}_0], \quad t \in \{1, \ldots, T-1\}.$$

The total solvency level at t is then defined and denoted as

$$TSL_t = R_t + K_t, \quad t \in \{1, \ldots, T-1\}.$$

It is very unlikely this approach will in reality be applied for reserving and capital allocation purposes. An insurer having future information available will normally take this into account. The use of this strategy should be seen as a possible fast computational approximation in the cash-flow model of the strategy described in section 3.2 (see section 4 for more details).

3.2. Approach Using Future Information

At time 0, we apply the same strategy as in section 3.1. At time $t \in \{1, \ldots, T-1\}$, we define the reserve as

$$R_t = E[D_t|\mathcal{F}_t],$$

where $\mathcal{F}_t$ denotes the information available at the start of year t. For the capital, we take

$$K_t = \rho[D_t - R_t|\mathcal{F}_t], \quad t \in \{1, \ldots, T-1\}.$$

As seen from time 0, R_t and K_t are random variables, since $\mathcal{F}_t$ is still unknown. At t however, R_t and K_t can be determined using the same methods as those which are used to determine pure premium and the capital at time 0. The total solvency level at t is again defined and denoted as

$$TSL_t = R_t + K_t, t \in \{1, \ldots, T-1\}.$$

The advantage of this approach is that the reserve and the capital are regularly adapted to new information. This also means that when the new information is bad, more reserves and capital than actually being available may need to be allocated. In other words, if the safety margin is not high enough given the new information, it is enlarged with new capital. Of course, this is subject to the assumption that one is able to allocate new capital if necessary. Three questions are very pertinent in this context:

1. Is it probable that, at a given moment in the future, new capital will need to be allocated in order to obtain a total solvency level which is sufficiently high?

2. Given that new capital needs to be allocated, how large can this amount be?

3. Can a company if necessary allocate additional capital to a risk?

In general, when new capital needs to be allocated, we may face problems of ruin. One can indeed question if one will find shareholders willing to invest in a product which is performing bad and leading to losses on the capital with a substantial probability. Therefore, a situation where it is probable that high amounts of capital need to be reinjected in the business is not preferable. For an answer to the first two questions in the context of a GMDB, we refer to section 6.2.

The answer to the third question depends upon a number of factors. If a company would only be exposed to risks of the same nature which are very correlated, then it may be very difficult to find additional capital if it is required. For a company with a more diversified portfolio of risks, this may however be possible.

4. ACTUARIAL PRICING USING A CASH-FLOW MODEL

In a cash-flow model, one models the average in-and outflows, taking the point of view of the shareholders. Assume the risk-free rate r is constant and equal to the rate of return on the bonds. Assume the tax rate is equal to γ and the return on the stocks, after taxation, is equal to δ.

We then have the following average outflows:

1. Net mean claim payments: $c_s(1-\gamma)$, where $s \in \{1/12, 2/12, \ldots, T\}$. Hence, costs can arise at the end of every month. For the ease of later computations, we assume claims which occur in a certain year are only paid at the end of that year. Hence, at the end of year t, the (re)insurer has to pay $c_t(1-\gamma)$, where

$$c_t = \sum_{s=1}^{12} e^{\frac{(12-s)r}{12}} c_{t-1+s/12}, \quad t \in \{1, \ldots, T\}.$$

2. Net mean change in the technical provisions: $\Delta p_t(1-\gamma)$, where $t \in \{0, \ldots, T\}$

3. Mean change in the allocated capital: Δk_t, where $t \in \{0, \ldots, T\}$

We have the following inflows:

1. Net mean return on the provisions: $R_t(p)(1-\gamma)$, where

$$R_t(p) = p_{t-1}(e^r - 1), \quad t \in \{1, \ldots, T\}.$$

2. Net mean return on the capital: $R_t(k)$, where

$$R_t(k) = k_{t-1}(e^\delta - 1) \quad \text{and} \quad t \in \{1, \ldots, T\}.$$

3. Net premium income: $TFP(1-\gamma)$, where *TFP* denotes the technico-financial premium which has to be determined.

It is at this point that we can understand the possible usefulness of the strategy described in section 3.1. To calculate the average in-and outflows, we need to determine the average reserves p_t and the average capitals k_t for all $t \in \{0, \ldots, T\}$. For the reserves, due to the iterativity property of the expectation, we have

$$p_t = E[R_t] = E[E[D_t|\mathcal{F}_t]] = E[D_t|\mathcal{F}_0], \quad \text{for all } t \in \{0, \ldots, T\}.$$

Hence, with respect to the reserves, there is no difference in the cash-flow model between the approach described in section 3.1 and section 3.2. Assume the capitals are calculated using the conditional tail expectation. For all $t \in \{1, \ldots, T\}$, we then have

$$k_t = E[K_t] = E[CTE_p[D_t - R_t|\mathcal{F}_t]] \neq CTE_p[D_t - R_t|\mathcal{F}_0].$$

We could use $CTE_p[D_t|\mathcal{F}_0]$ as an approximation of k_t but we certainly need to verify whether this approximation is good.

Once the average future cash-flows are known, we discount all of them with the cost of capital, which we assume to be known and constant. The technico-financial premium is defined as the value which makes the sum of the discounted inflows equal to the sum of the discounted outflows. Hence, the technico-financial premium is the value which solves the following equation:

$$\begin{aligned}&\sum_{t=0}^{T} e^{-tCOC}\left[\Delta p_t - R_t(p) + \sum_{s=1}^{12} e^{\frac{(12-s)r}{12}} c_{t-1+s/12}\right](1-\gamma)\\&= \sum_{t=0}^{T} e^{-tCOC}\left[R_t(k) - \Delta k_t\right] + TFP(1-\gamma),\end{aligned} \tag{2}$$

with the convention that $R_0(p) = R_0(k) = 0$ and $c_s = 0$ for all $s \leq 0$, and where *COC* stands for the cost of capital. Now suppose the average reserves are equal to the future mean claim payments, discounted at the risk free rate, i.e.

$$p_t = \sum_{s=1}^{12(T-t)} e^{-rs/12} c_{t+s/12}. \tag{3}$$

Using (3) it can be verified that the term $\Delta p_t - R_t(p) + \sum_{s=1}^{12} e^{\frac{(12-s)r}{12}} c_{t-1+s/12}$ in (2) is equal to

$$R_0 \quad \text{if} \quad t = 0, \tag{4}$$

$$0 \quad \text{if} \quad t \in \{1, \ldots, T\}. \tag{5}$$

The assumption that costs are only paid at the end of the year was taken in order to obtain (5). Using (4) and (5), we can write the technico-financial premium as

$$TFP = R_0 + \frac{\sum_{t=0}^{T} e^{-tCOC}[\Delta k_t - R_t(k)]}{1-\gamma}. \tag{6}$$

This means the technico-financial premium consists of two parts:

1. The reserve R_0 taken at time 0.

2. A loading for the average amounts of capital k_t which are allocated at the start of each year $t \in \{1, \ldots, T\}$.

5. SIMULATION STRATEGY

5.1. Modelling a GMDB

We suppose there is a group of $N_I = 1000$ insured aged $x = 50$ which all invest $C = 1$ into a risky asset $(S_t)_{t\geq 0}$. The (re)insurer provides a guarantee of $K = 1$ in case the insured dies before retirement. Hence, in case an insured dies at time t, the (re)insurer will pay $(K - S_t)_+$ for the guarantee. Note that $(K - S_t)_+$ is equal to the payoff function of a European put option with strike price K and maturity date t on the underlying asset $(S_i)_{i\in[0,t]}$. We suppose there is no surrender and there is an age of retirement $x_R = 65$ after which no payments are made. At retirement, no guarantee is given.

We use the regime-switching log-normal (RSLN) model with 2 regimes as described in Hardy (2001). This model provides us with monthly log-returns for the underlying asset. We denote the regime applying to the interval $[s, s+1)$ as κ_s. Hence $\kappa_s \in \{1, 2\}$. In a certain regime κ_s we assume the return Y_s satisfies:

$$Y_s = \log(\frac{S_{s+1}}{S_s}|\kappa_s) \sim N(\mu_{\kappa_s}, \sigma_{\kappa_s}).$$

Furthermore, the transitions between the regimes are assumed to follow a Markov Process characterized by the matrix P of transition probabilities

$$p_{ij} = Pr[\kappa_{s+1} = j|\kappa_s = i], \quad \text{for } i, j \in \{1, 2\}.$$

As parameters, we use the parameters estimated using the maximum log-likelihood techniques explained in Hardy (2001). We use the S&P 500 data (total returns) from January 1960 to December 2003. The estimated parameters are summarized in table 3.

To model mortalities, we use a Gompertz-Makeham approach. The survival probability of a person aged x is then described as

$${}_tp_x = \exp\left(-\alpha t - \frac{\beta e^{\gamma x}(e^{\gamma t} - 1)}{\gamma}\right),$$

for some $\alpha > 0$, $\beta > 0$ and γ. We use the first set of Gompertz-Makeham parameters in table 3 up to age 65 and the second set for ages higher than 65. As for the underlying asset, we model mortalities on a monthly basis, generating from a binomial distributions.

5.2. Using Future Information

First we describe a naive simulation strategy which could be used to incorporate future information (see figure 1). At time 0 we make N_S simulations going from 0 to T and we use these to determine the distribution function of D_0. At time 1, we then obtain N_S values for the underlying asset and N_S values for the number of people who are still alive. This leads to $N = N_S^2$ combinations. From each of these combinations, we should then make new simulations from 1 to T to determine N distribution functions from which reserves and capitals can be calculated. p_1 and k_1 are determined as the averages of these N reserves and capitals respectively. For $t \in \{2, \ldots, T\}$, p_t and k_t could be determined using a similar strategy.

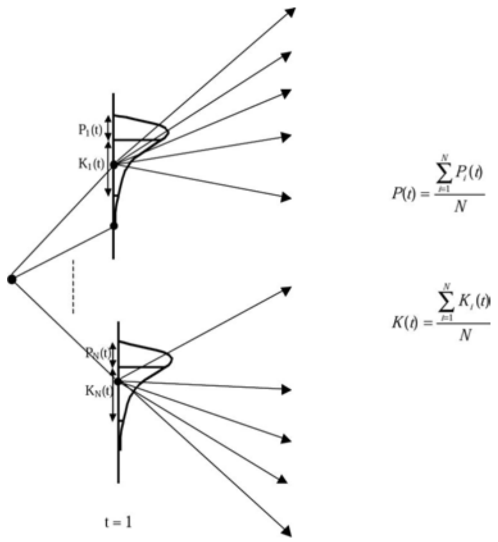


Figure 1: Tree simulations to obtain p_t and k_t.

It can easily be understood that this approach will require a huge amount of computation time. Therefore, we turn to an approximation strategy. This strategy is described in detail in Desmedt et al. (2004). We here only describe intuitively the simplifications which are made to the strategy described above.

- Instead of using all future values for the underlying asset and for the mortality to estimate the required reserve and capital, we limit ourselves to a only N_A values for the underlying asset and N_O classes for the mortality. Since the influence of the position of the underlying asset on the estimated costs is a lot more important than that of the mortality, we take $N_O < N_A$.
- Both for the underlying asset and for the mortality, we avoid resimulating at future points in time.

5.3. Verification

5.3.1. Put Option

In Hardy (2001), theoretical formulas for the *VaR* and *CTE* of a European put option with payoff function $X = (G - S_n)_+$ on a risky asset $(S_s)_{s\in[0,n]}$ are derived under the log-normal model. As pointed out by Hardy, when $p < \Pr[S_n > G]$, the definition of the *CTE* given by (1) does not give suitable results. Therefore, she redefines the *CTE* as

$$CTE_p[X] = \frac{(1-\beta')E[X|X > Q_p[X]] + (\beta' - p)Q_p[X]}{1-p}, \tag{7}$$

where $\beta' = \max\{\beta \mid Q_p[X] = Q_\beta[X]\}$. In what follows, we will use definition (7) for the *CTE*. When calculating the *CTE* from simulations, we take $CTE_p[X]$ equal to the average of the $(1-p)\%$ worst outcomes. If $S_n \sim \text{LN}(n\mu, \sqrt{n}\sigma)$, then for $p \geq \Pr[S_n > G]$

$$CTE_p[X] = G - S_0 \frac{e^{n\mu + n\sigma^2/2}}{1-p}\Phi(-\Phi^{-1}(p) - \sqrt{n}\sigma).$$

On the other hand, if $p < Pr[S_n > G]$, one can verify the *CTE* is given by

$$CTE_p[X] = \frac{1 - Pr[S_n > G]}{1-p} CTE_{Pr[S_n > G]}[X].$$

To test the simulation strategy for the underlying asset, we wish to calculate

$$E[CTE_p[(G - S_n)_+ | S_y]], \text{ where } y \in \{1, \ldots, T\}, \tag{8}$$

the expectation of the *CTE* of the European put option, given the value of the underlying asset at $y \in \{1, \ldots, T\}$, both theoretically and using the simulation strategy.

First we calculate the theoretical value of (8), as if $p \geq \Pr[S_n > G|S_y]$ were valid for all possible values S_y as a (sometimes rough) approximation of $E[CTE_p[(G - S_n)_+|S_y]]$. Hence we can write

$$\begin{aligned} E[CTE_p[(G - S_n)_+|S_y]] &\approx \int_0^\infty [G - x\frac{e^{(\mu+\sigma^2/2)(n-y)}}{1-p}\Phi(-\Phi^{-1}(p) - \sqrt{n-y})] f_{S_y}(x)dx \qquad (9) \\ &= G - \frac{e^{(\mu+\sigma^2/2)(n-y)}}{1-p}\Phi(-\Phi^{-1}(p) - \sqrt{n-y}) \int_0^\infty x f_{S_y}(x)dx \\ &= G - \frac{e^{(\mu+\sigma^2/2)n}}{1-p}\Phi(-\Phi^{-1}(p) - \sqrt{n-y}). \end{aligned}$$

The approximation made in (9) will improve for high values of p but even for very high p-values it will still need to be corrected to get a good approximation, if we go into the future. In Desmedt et al. (2004), it is described in detail how this correction can be made.

In table 1, we summarize the results of this test for the *CTE* at level $p = 0.99$ for a maturity guarantee of $G = 1 = S_0$ after 10 years. Given the information we have at 0, we wish to compare the average simulated (A_S) and theoretical (A_T) *CTE* at level 0.99 of the maturity guarantee at the start of every year, taking into account the information about the underlying asset we have at that moment. We made 15000 simulations for the underlying asset for 10 years where the returns are log-normally distributed with parameters $\mu = 0.085$ and $\sigma = 0.20$. In column 1, we specify the year at the start of which the average *CTE*'s at level 0.99 are compared. In the fourth column, we calculate the difference between A_S and A_T, relative to A_T. In column 5, we calculate the difference between A_S and A_T, relative to the level of the guarantee.

Year	A_T	A_S	$\frac{\lvert A_S - A_T \rvert}{A_T}$	$\lvert A_S - A_T \rvert$
1	0.557	0.568	2.0%	1.1%
2	0.508	0.512	0.8%	0.4%
3	0.453	0.450	0.7%	0.3%
4	0.397	0.399	0.5%	0.2%
5	0.339	0.345	1.8%	0.6%
6	0.286	0.297	3.8%	1.1%
7	0.234	0.246	5.1%	1.2%
8	0.183	0.183	0.0%	0.0%
9	0.132	0.138	4.5%	0.6%
10	0.082	0.088	7.3%	0.6%

Table 1: Test average *CTE* at level 0.99 of a put option.

We see that for all compared years, the theoretical and average *CTE* are fairly close.

5.3.2. Iterativity Property of the Expectation

In figure 2, we compare the average reserves for the parameters as specified in table 3. For all $t \in \{0, \ldots, T\}$, the dashed line represents the approximation of $E[E[D_t|S_t, N_t]]$. The full line represents $E[D_t|S_0, N_0]$.We see that the two lines fit very well onto each other. This means that under these conditions, our approximation strategy passes the test of the iterativity property of the expectation. By taking N_A smaller under the same conditions, we would see the iterativity property is less well satisfied. The investigation of some other conditions learned that when costs are more likely (e.g. when $K > S_0$), less values can be taken for the underlying asset to satisfy the iterativity property in the same way. On the other hand, when costs are less likely (e.g. when $K < S_0$), N_A needs to be larger to obtain results of the same quality as of those in figure 2.

6. RESULTS

6.1. Average Capitals and Technico-Financial Premium

In figure 3, we see that the capitals start at the same value in the two multi-period strategies. At the start of the second year, the strategy using future information requires on average less capital than the one not using future information. The differences increase if we move further into the future.

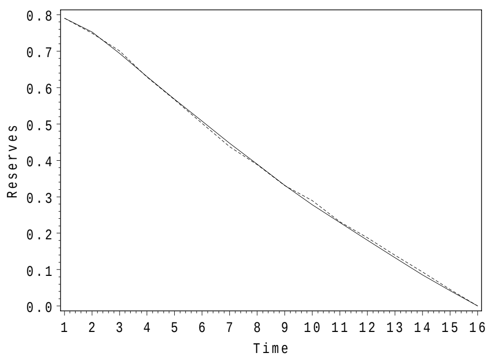


Figure 2: Comparison of the average reserves.

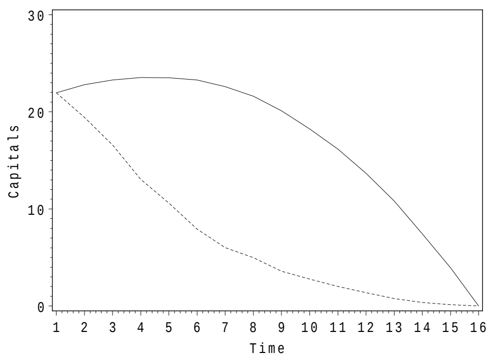


Figure 3: Comparison of the average capitals.

As we see in table 2, the differences between the average capitals in the two strategies have an important influence on the technico-financial premium.

	Approach section 3.1	Approach section 3.2
PP	0.79	0.79
K_0	21.97	21.97
TFP	14.05	6.50

Table 2: Comparison technico-financial premium.

The technico-financial premium in the approach not using future information is more than twice as large as in the approach in which future information is taken into account.

6.2. Distribution Functions of Future Total Solvency Levels

In figure 4, we show the approximated distribution functions of the future total solvency levels, when they are adapted to the information at the start of year $t \in \{1, \ldots, T\}$, which are obtained using the simulation strategy. The vertical line is the total solvency level at time 0. To characterize the other lines, we can use the following rule: the higher the line is at the left, the further in time the situation which is represented.

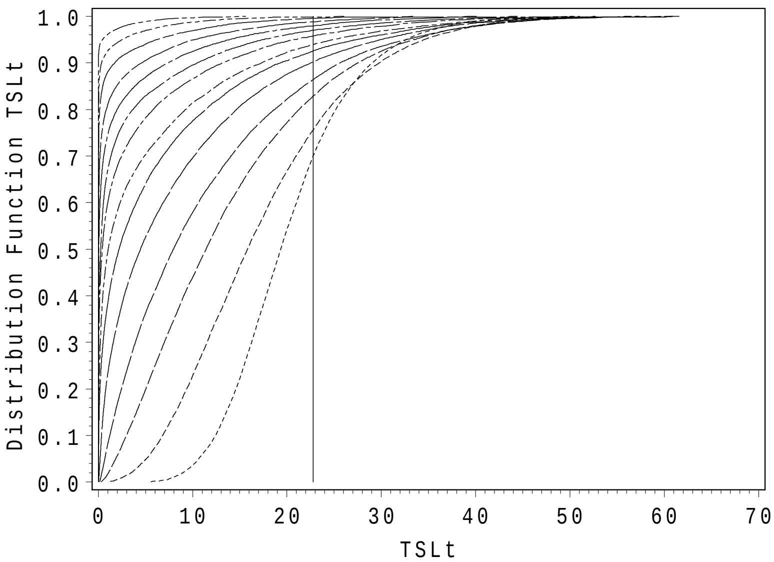


Figure 4: Approximate distribution functions for the total solvency levels at the start of each year when using future information.

We see that in about 30% of the cases, the total solvency level at the start of the second year should be larger than the initial total solvency level. The largest estimations for the required total solvency level at the start of the second year are about twice as large as the initial total solvency level. On the other hand, the lowest estimations are about one third of the initial total solvency level. Hence, already after one year, the differences in mainly the underlying asset can be such that we obtain a very wide range in possible required total solvency levels. At the start of the third year, the largest estimations become even larger but the probability that the total solvency level should be larger than the initial total solvency level decreases to about 25%. Then for a few years, the largest estimations remain at a comparable level. We also see that the probabilities that the total solvency levels should be larger than the initial total solvency level continue decreasing. At about half the lifetime of the risk, the largest estimations start to decrease too. From the start of year five on, there are estimations for the required total solvency level of 0.

7. CONCLUSION

In this paper, we suggested a multi-period capital allocation strategy which incorporates future information. We explained how this strategy can be used within a cash-flow model to price a multi-period risk. An interesting example of such a risk is a GMDB in unit-linked life insurance, especially when pricing and reserving is done in an actuarial way. After explaining and verifying a simulation strategy for the approach using future information, we compare it with an approach not using future information. For the average reserves, we see both strategies lead to the same results, due to the iterativity property of the expectation. For the average capitals however, we observe important differences between the two methods. As a consequence, the technico-financial premium for the method not using future information is more than twice as large as in the approach using future information. Our simulation strategy also provides approximations for the distribution functions for the future reserves and capitals. From these, interesting information about possible variations of reserves and capitals can be withdrawn.

8. USED PARAMETERS AND NOTATIONS: SUMMARY

Simulations		
Number of simulations	N_S	15000
Number of values for underlying asset	N_A	500
Number of classes for mortality	N_O	5
Contractual Parameters		
Portfolio composition	$\{N_I, x, C\}$	$\{1000, 50, 1\}$
Initial value underlying asset	S_0	1
Guarantee at death	K	1
Age of retirement	x_R	65
Mortality Parameters		
Gompertz-Makeham parameters	α	0.000591068646661458
	β_{0-65}	0.0000073759357103 7331
	γ_{0-65}	0.11807173977857
	β_{65-99}	0.000619125291109306
	γ_{65-99}	0.0532009916754107
Parameters Cash Flow model		
Risk free rate	r	0.0425
Tax rate	γ	0.4
Average return on investments in shares	δ	0.0505
Cost of capital	*COC*	0.085
Parameters RSLN model		
Average log-return in regime 1	μ_1	0.0135
Average log-return in regime 2	μ_2	-0.0109
Volatility in regime 1	σ_1	0.0344
Volatility in regime 2	σ_2	0.0645
Probability to move from regime 1 to 1	p_{11}	0.0483
Probability to move from regime 2 to 1	p_{21}	0.1985

Table 3: Used parameters and notations: summary.

References

P Artzner, F Delbaen, J.M. Eber, and D. Heath. Coherent measures of risk. *Mathematical Finance*, 9:203–228, 1999.

N.L. Bowers, H.U. Gerber, and J.C. Hickman. *Actuarial Mathematics*. Society of Actuaries, 1986.

M.J. Brennan and E.S. Schwartz. The pricing of equity-linked life insurance policies with an asset value guarantee. *Journal of Financial Economics*, 3:195–213, 1976.

S. Desmedt, X. Chenut, and J.F. Walhin. Actuarial pricing for minimum death guarantees in unit-linked life insurance: A multi-period capital allocation problem. *Proceedings of the AFIR Colloquium, Boston*, 2004.

J. Dhaene, M.J. Goovaerts, and R. Kaas. Economic capital allocation derived from risk measures. *North American Actuarial Journal*, 7(2):44–59, 2003.

J. Dhaene, R. Laeven, S. Vanduffel, G Darkiewicz, and M.J. Goovaerts. Can a coherent risk measure be too subadditive? *DTEW Research Report 0431,K.U.Leuven*, 2004a.

J. Dhaene, S. Vanduffel, Q.H. Tang, M.J. Goovaerts, R. Kaas, and D. Vyncke. Capital requirements, risk measures and comonotonicity. *Belgian Actuarial Bulletin*, 4:53–61, 2004b.

C. Frantz, X. Chenut, and J.F. Walhin. Pricing and capital allocation for unit-linked life insurance contracts with minimum death guarantee. *Proceedings of the AFIR Colloquium, Maastricht*, 2003.

M.J. Goovaerts, E. Van den Borre, and R. Laeven. Managing economic capital and virtual economic capital within financial conglomerates. *Proceedings of the ASTIN Colloquium, Bergen*, 2004.

M. Hardy. A regime-switching model of long-term stock returns. *North American Actuarial Journal*, 5(2):41–53, 2001.

M. Hardy. Bayesian risk management for equity-linked insurance. *Scandinavian Actuarial Journal*, 3:185–211, 2002.

M. Hardy. *Investment Guarantees: Modeling and Risk Management for Equity-Linked Life Insurance*. Wiley Finance, 2003.

J.C. Hull. *Options, Futures & other Derivatives*. Prentice-Hall International, Inc., 2000.

IAA. A global framework for insurer solvency assessment. *Report of the Insurer Solvency Assessment Working Party*, 2004.

R. Kaas, M.J. Goovaerts, J. Dhaene, and M. Denuit. *Modern Actuarial Risk Theory*. Kluwer Academic Publishers, 2001.

RUIN THEORY WITH K LINES OF BUSINESS

Stéphane Loisel

Université Claude Bernard Lyon I, Ecole ISFA, 50 avenue Tony Garnier, 69007 Lyon, France
Email: `stephane.loisel@univ-lyon1.fr`

Abstract

This paper deals with the evolution of the reserves of an insurance company with $K \geq 1$ lines of business facing dependent risks. We consider risk measures based on the behavior of the multivariate risk process from an academic point of view. To deal with multivariate risk processes, we propose a multi-risks model. We then explain how to determine the optimal reserve allocation of the global reserve to the lines of business in order to minimize those risk measures. The impact of dependence on the risk perception and on the optimal allocation is studied and used to test the consistency of the risk measures. This paper is mainly based on the two following papers Loisel (2004 2005a).

1. INTRODUCTION

This paper describes the work I presented at the 3rd AFM Day conference. It is based on the two following papers Loisel (2004 2005a), which the interested reader is encouraged to consult for further details.

We consider here the case of an insurance company with $K \geq 1$ lines of business. Some authors, like Cossette and Marceau (2000), and many others, considered multi-risk models. However, in most cases, they focus on the unidimensional risk process representing the total wealth of the company.

From now on, we consider a fixed accounting time horizon T, which may be infinite, and study the evolution of the wealths of the lines of business of the company between times 0 and T. With three lines of business (three kinds of activities), for example liability, disablement and driving insurance, it is not the same situation to have $(1M, 2M, -2.8M)$ (*id est* 1 million euros for the first branch, 2 million euros for the second one and to be short of 2.8 million euros for the last branch), or to have $(0, 0.1M, 0.1M)$. Considering only the total wealth (0.2 million euros) does not reflect the situation of the company very well. A few years ago, a holding company mainly had two large airlines companies. The first one was doing well, say its wealth was 10M dollars. The second one was undergoing a bad period, with a debt of, say, 2M dollars. Even if the subcompanies were collateralized, the holding company was estimated 4M dollars, instead of 8M dollars, by the market.

The reason was that analysts expected the healthy line of business to be penalized by the other one, which was in the red. To be able to detect such penalties, and to compute probabilities of such unfavorable events, the multi-dimensional process has to be studied as an additional indicator.

The solvency II project is an additional motivation for these considerations. For an international insurance group, the required capital, formerly determined at the group level, which took into account possible mutualization of risks supported by the different subcompanies, will possibly be determined additionally subcompany by subcompany. This would breed for all subcompanies a need for a much higher capital.

We consider in section 2 risk measures based on the behavior of the multivariate risk process from an academic point of view. To deal with multivariate risk processes, we propose in section 3 a multi-risks model. We then explain in section 4 how to determine the optimal reserve allocation of the global reserve to the lines of business in order to minimize those risk measures. The impact of dependence on the risk perception and on the optimal allocation is studied and used to test the consistency of the risk measures in section 5.

2. RISK MEASURES WITH K LINES OF BUSINESS

For a unidimensional risk process, one classical goal is to determine the minimal initial reserve u_ϵ needed for the probability of ruin to be less than ϵ.
In a multidimensional framework, modelling the evolution of the different lines of business of an insurance company by a multirisk process $(u_1 + X_t^1, \ldots, u_K + X_t^K)$ ($R_k = u_k + X_t^k$ corresponds to the wealth of the k^{th} line of business at time t), one could look for the global initial reserve u which ensures that the probability of ruin ψ satisfies

$$\psi(u_1, \ldots, u_K) \leq \epsilon$$

for the optimal allocation $(u_1, \ldots, u_K)$ such that

$$\psi(u_1, \ldots, u_K) = \inf_{v_1 + \cdots + v_K = u} \psi(v_1, \ldots, v_K).$$

There exist different ruin concepts for multivariate processes. Most of them may be represented by the multivariate, time-aggregated claim process to enter some domain of $\mathbb{R}^K$, called insolvency region (see Picard et al. (2003), Loisel (2004)). Considering insolvency regions of the kind

$$\{x \in \mathbb{R}^K, \quad x_1 + \cdots + x_K > u + ct\},$$

corresponds to the classical unidimensional ruin problem for the global company. However, one could consider that ruin occurs when at least one line of business gets ruined :

$$\psi(u_1, \ldots, u_K) = P(\exists k \in [1, K], \exists t > 0, u_k + X_t^k < 0).$$

To measure the severity of ruin, one may consider penalty functions which quantify the penalty undergone by the company due to insolvency of some of its lines of business.

Instead of the probability of crossing some barriers, it may thus be more interesting to minimize the sum of the expected cost of the ruin for each line of business until time T, which may be represented by the expectation of the sum of integrals over time of the negative part of the process. In both cases, finding the global reserve needed requires determination of the optimal allocation.

The multidimensional risk measure A, which does not depend on the structure of dependence between lines of business, is one example of what can be considered :

$$A(u_1, \ldots, u_K) = \sum_{i=k}^{K} EI_T^k$$

where

$$EI_T^k = E\left[\int_0^T |R_t^k| 1_{\{R_t^k < 0\}} dt\right]$$

with $R_t^k = u_k + X_t^k$ under the constraint $u_1 + \cdots + u_K = u$.

Another possibility would be to minimize the sum

$$B = \sum_{k=1}^{K} E\tau_k'(u_1, \ldots, u_K)$$

where

$$E\tau_k'(u_1, \ldots, u_K) = E\left(\int_0^T 1_{\{R_t^k < 0\}} 1_{\{\sum_{j=1}^K R_t^j > 0\}} dt\right).$$

To determine the total initial reserve u needed to have an acceptable risk level requires to find the optimal allocation. Before tackling this problem in section 4, we propose a dependence structure in section 3 to be able to work on risk measures like B or ψ which take dependence into account.

3. A MULTIDIMENSIONAL RISK MODEL

We consider the process modelling the wealth of the K lines of business of an insurance company. Typical lines of business are driving insurance, house insurance, health, incapacity, death, liability,... Two main kinds of phenomena may generate dependence between the aggregated claim amounts of these lines.

- Firstly, in some cases, claims for different lines of business may come from a common event: for example, a car accident may cause a claim for driving insurance, liability and disablement insurance. Hurricanes might cause losses in different countries. This should correspond to simultaneous jumps for the multivariate process. The most common tool to take this into account is the Poisson common shock model.

- Secondly, there exist other sources of dependence, for example the influence of the weather on health insurance and on agriculture insurance. In this case, claims seem to outcome

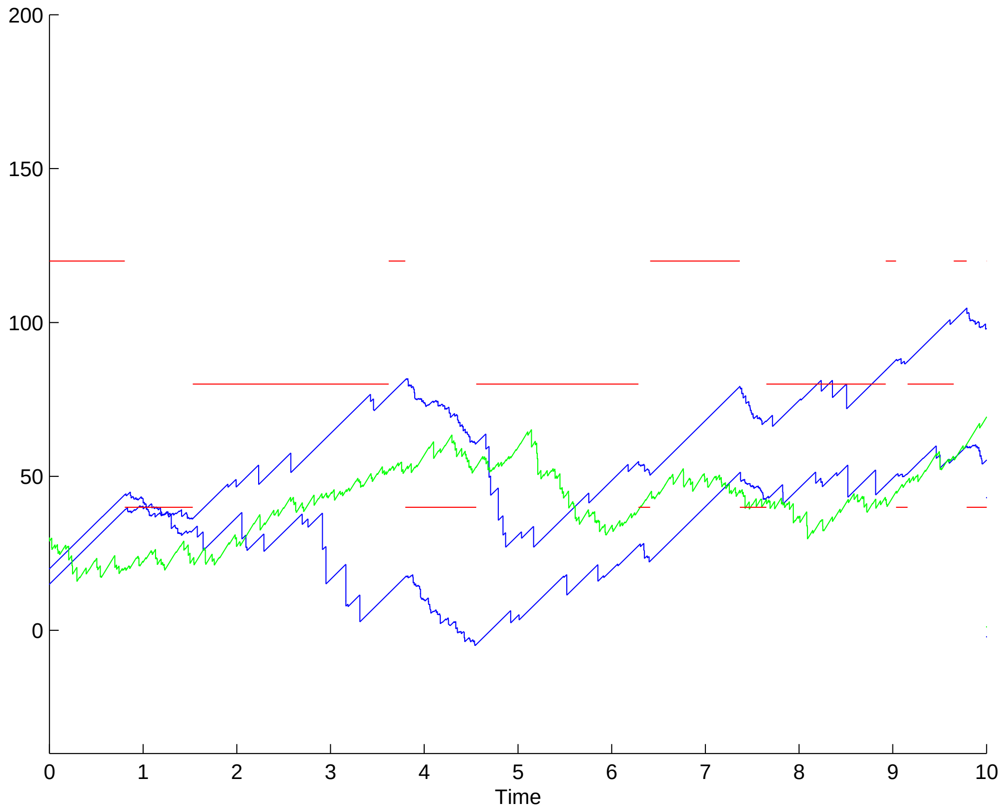


Figure 1: Sample path for three lines of business: The grey one does not depend on the state of the environment. The two dark lines of business have identical parameters, and are independent conditionally on the environmental state. Occupation periods for environment states are given by horizontal lines.

> independently for each branch, depending on the weather. This seems to correspond rather to models with modulation by a Markov process which describes the evolution of the state of the environment.

Common Poisson shock models are quite easy to understand. To illustrate the other notion, figure 1 shows a sample path of the surpluses of the 3 lines of business of the insurance company, under a Markovian environment (horizontal curves correspond to occupation times), but without common shock. The set of states of the environment has cardinality three. State 3 (highest horizontal level on the graph) is the most favorable for the company, almost no claim occurs for lines 1 and 2 (dark curves) in this state. State 1 is the least favorable state for the company, claim frequencies and severities are higher for lines 1 and 2. Events for the third line of business (grey curves) are independent from the state of the environment. One can see the strong positive dependence between lines 1 and 2 (dark curves), but also their independence conditionally to the environment state. At some moment, the two dark curves separate each other because of this conditional independence.

Let us define more precisely the model we propose, which takes into account these two different sources of dependence: the Markov-modulated, Common Shock, Multivariate Compound Poisson Process model ((MM,CS)-MCPP), Conditionally on the state of the environment, the multivariate claim process is modelled by a Compound Poisson Process with Common Shocks. The intensity, the claim size distribution and even the common shock parameters may vary in function of the state of the environment, which is modelled by a Markov process.

The environment state process, denoted by $J(t)$, is a Markov process with state space $S = \{1, \ldots, N\}$, initial distribution μ and intensity matrix A.
For example, state 1 might correspond to periods of frequent heavy rains, or very hot weather, to hurricane seasons. It might also indicate frequent speed controls, law modifications,...

There are $m \geq 1$ different types of shock. If $J(t) = i$, then shocks of type e $(1 \leq e \leq m)$ occur according to a Poisson process with intensity $\lambda_{e,i}$. These shock counting processes are independent conditionally on $J(.)$.
For example, a shock may be a big car accident, a particular hurricane, an explosion, a medical mistake with consequences in a hospital,...

If $J(t) = i$, at the r^{th} occurrence of type e $(1 \leq e \leq m)$, the Bernoulli vector $I_i^{e,r} = (I_{1,i}^{e,r}, \ldots, I_{Ki}^{e,r})$ indicates whether a loss occurs for branch $k \in [1, K]$, and the potential losses are represented by $X_i^{e,r} = (X_{1,i}^{e,r}, \ldots, X_{K,i}^{e,r})$.
For example, a car accident may cause claims in driving insurance, liability, incapacity, death.

For a fixed state i and fixed shock type e, the successive $I_i^{e,r}$ are i.i.d., the successive $X_i^{e,r}$ are i.i.d.

Besides, the $I_i^{e,r}$ are independent from the $X_i^{e,r}$. However, for a fixed event i, e, r, the loss triggers $(I_{1,i}^{e,r}, \ldots, I_{K,i}^{e,r})$ and the potential losses generated by this event $X_i^{e,r} = (X_{1,i}^{e,r}, \ldots, X_{K,,i}^{e,r})$ may be dependent. In most real-world cases, the by-claims amounts seem to be positively correlated. This is the reason why we allow this kind of dependence. Between time 0 and time t, denote by $N_i^e(t)$ the number of shocks of type e that occurred while J was in state i. Then the aggregate claim amount vector up to time t is $S(t) = (S_1(t), \ldots, S_K(t))$ where for a branch $k \in [1, K]$,

$$S_k(t) = \sum_{i=1}^{N} \sum_{e=1}^{m} \sum_{r=1}^{N_i^e(t)} I_{k,i}^{e,r} . X_{k,i}^{e,r}$$

In case of no common shock, $S_k(t)$ (the aggregate claim amount vector up to time t for branch k) is a compound Cox process with intensity $\lambda_{k,J(t)}$ and claim size distribution $F_{k,J(t)}$.

$(S(t), J(t))$ is a Markov process (and of course $S(t)$ is not!). This has an impact on computation times, because we have to keep track of the environment state during the computations.

In this model, the use of Monte Carlo methods is necessary if the number of states of the environment or the number of lines of business is too large.

For small values of these parameters, we explain in Loisel (2004) how to compute finite-time ruin probabilities using an algorithm generalizing the algorithm of Picard et al. (2003). In case of financial interactions between some lines of business, we can also use martingale methods based on results of Asmussen and Kella (2000) and Frostig (2004) to provide a theoretical way to compute the expected time to ruin of the main line of business, and the impact of the other lines on the time to ruin and on the dividends paid to the shareholders until ruin of the main line (see Loisel (2005b)).

We will only recall here results from Loisel (2004) on the impact of dependence on multidimensional, finite-time ruin probabilities in section 5.

4. OPTIMAL RESERVE ALLOCATION

Using differentiation theorems from Loisel (2005a), it is possible to determine a very intuitive optimal reserve allocation strategy to minimize the functional A defined in section 2. The problem is to minimze

$$A(u_1, \ldots, u_K) = \sum_{i=k}^{K} EI_T^k$$

where

$$EI_T^k = E\left[\int_0^T |R_t^k| 1_{\{R_t^k<0\}} dt\right]$$

with $R_t^k = u_k + X_t^k$ under the constraint $u_1 + \cdots + u_K = u$. This does not depend on the dependence structure between the lines of business because of the linearity of the expectation. Under hypothesis that make it possible to consider finite expectations and to differentiate (see Loisel (2005a)), denote $v_k(u_k)$ the differentiate of EI_T^k with respect to u_k. Using Lagrange multipliers implies that if $(u_1, \ldots, u_K)$ minimizes A, then $v_k(u_k) = v_1(u_1)$ for all $1 \leq k \leq K$. Compute $v_k(u_k)$:

$$v_k(u_k) = \left(E\left[\int_0^T |R_t^k| 1_{\{R_t^k<0\}} dt\right]\right)' = -E\tau^k = -\int_0^T P\left[\{R_t^k < 0\}\right] dt$$

where τ^k represents the time spent in the red between 0 and T for line of business k.
The sum of the average times spent under 0 is a decreasing function of the u_k. So A is strictly convex. On the compact space

$$\mathcal{S} = \{(u_1, \ldots, u_K) \in (\mathbb{R}^+)^K, \quad u_1 + \cdots + u_K = u\},$$

A admits a unique minimum.

Theorem 4.1 *The optimal allocation is thus the following: there is a subset $J \subset [1, K]$ such that for $k \notin J$, $u_k = 0$, and for $k, j \in J$, $E\tau_k = E\tau_j$.*

The interpretation is quite intuitive: the safest lines of business do not require any reserve, and the other ones share the global reserve in order to get equal average times in the red for those lines of business.

Relaxing nonnegativity, on $\{u_1 + \cdots + u_K = u\}$, if $(u_1, \ldots, u_K)$ is an extremum point for A, then for the K lines of business, the average times spent under 0 are equal to one another. If it is a minimum for the sum of the times spent below 0 for each line of business, then the average number of visits is proportional to the marginal premium income rates c_k, and in infinite time the ruin probabilities are in fixed proportions. However the existence of a minimum is not guaranteed, because $(u_1, \ldots, u_K)$ is no longer compact. The problem would be more tractable with the average time in the red or with minimization on the c_k, because some factors penalize very negative u_k in these cases.

The multidimensional risk measure A, which does not depend on the structure of dependence between lines of business, is one example of what can be considered. Another possibility was to minimize the sum

$$B = \sum_{k=1}^{K} E\tau_k'(u_1, \ldots, u_K)$$

where

$$E\tau_k'(u_1, \ldots, u_K) = E\left(\int_0^T 1_{\{R_t^k<0\}} 1_{\{\sum_{j=1}^K R_t^j>0\}} dt\right).$$

Here B takes dependence into account, and the following proposition shows what can be done:

Proposition 4.2 *Assume that for each line k, $X_t^k = c_k t - S_t^k$, with all $c_k > 0$ and where the S_t^k satisfy some technical hypotheses (see Loisel (2005a)). Define B by $B(u_1, \ldots, u_K) = \sum_{k=1}^K E(\tau_k'(u_1, \ldots, u_K))$ for $u \in \mathbb{R}^K$. B is differentiable on $(\mathbb{R}_*^+)^K$, and for $u_1, \ldots, u_K > 0$,*

$$\frac{\partial B}{\partial u_k} = -\frac{1}{c_k} E N_k^0(u, T),$$

where $N_k^0(u, T) =$ Card $\left(\{t \in [0, T], \quad \left(R_t^k = 0\right) \cap \left(\sum_{j=1}^K R_t^j > 0\right)\}\right)$.

All these results are drawn from Loisel (2005a). The proofs and some examples may be found there.

5. IMPACT OF DEPENDENCE

The fact that the optimal reserve allocation strategy for the functional A does not depend on the dependence structure between risks makes it a good benchmark to compare with other risk measures, for example the probability that at least one line of business gets ruined, which we call from now on multidimensional ruin probability (m.r.p.).

Suppose that your lines of business face dependent risks, with fixed marginals. An incontestable fact is that any actuary would prefer negative dependence between risks, in order to profit from their mutualization. This is in agreement with the univariate risk model, in which positive dependence between risks increases the probability of ruin for the global company. However, this is in total contradiction with the results obtained in Loisel (2004) for the m.r.p.. It is shown that positive dependence between risks decrease the probability that at least one line of business gets ruined. This is quite intuitive since for m.r.p. you do not care if only one or all your lines are ruined.

In particular we have the following result. Assume that the transition rate matrix of the environment process is stochastically monotone, and that one can order the states from the least to the most favorable state for the company for all lines of business at the same time. Then, the m.r.p. in the multidimensional model in which all lines of business are impacted by the same environment is less than the m.r.p. in the similar model in which the K lines of business are impacted by independent copies of the environment process. Picard et al. (2003) had also proved for processes with independent increments, if the risks were $PUOD$, then the times to ruin were $PLOD$.

This shows that the m.p.r. is interesting as a complementary information about the solvency of the lines of business, and may be used with other criteria to determine the reserve allocation once the risks have been selected, but that to select risks and to determine the capital requirements, one should use as before risk measures on the aggregated process first. Using only m.r.p. would be a poor idea.

6. CONCLUSION

We proposed and studied risk measures and models for multidimensional risk models. The impact of dependence on the risk perception and on the optimal reserve allocation strategy gives an idea on how to use them. Interesting problems to consider are the problems of parameter estimation and the link with credit risk theory. It would also be interesting to consider more general risk processes to take investment into account.

References

S. Asmussen and O. Kella. A multi-dimensional martingale for Markov additive processes and its applications. *Adv. in Appl. Probab.*, 32(2):376–393, 2000. ISSN 0001-8678.

H. Cossette and E. Marceau. The discrete-time risk model with correlated classes of business. *Insurance Math. Econom.*, 26(2-3):133–149, 2000. ISSN 0167-6687. Liber amicorum for Etienne De Vylder on the occasion of his 60th birthday.

E. Frostig. The expected time to ruin in a risk process with constant barrier via martingales. *Working paper*, 2004.

S. Loisel. Finite-time ruin probabilities in the markov-modulated multivariate compound poisson model with common shocks, and impact of dependence. *Working paper, Cahiers de recherche de l'ISFA*, (WP2026), 2004.

S. Loisel. Differentiation of some functionals of risk processes and optimal reserve allocation. *to appear in Journal of Applied Probability*, 42(2), 2005a.

S. Loisel. Time to ruin, insolvency penalties and dividends in a markov-modulated multirisk model with common shocks. *Working paper, Cahiers de recherche de l'ISFA*, (WP2027), 2005b.

P. Picard, C. Lefèvre, and I. Coulibaly. Multirisks model and finite-time ruin probabilities. *Methodol. Comput. Appl. Probab.*, 5(3):337–353, 2003. ISSN 1387-5841.

A MOTIVATION FOR CONDITIONAL MOMENT MATCHING

Roger Lord[†] [§]

[†]*Tinbergen Institute, Erasmus University Rotterdam, P.O. Box 1738, 3000 DR Rotterdam, The Netherlands.*
[§]*Modelling and Research (UC-R-355), Rabobank International, P.O. Box 17100, 3500 HG Utrecht, The Netherlands.*
Email: `lord@few.eur.nl`, `roger.lord@rabobank.com`

Abstract

One can find approaches galore in the literature for the valuation of Asian basket options. When the number of underlyings is large one has to resort to bounds or approximations to value these options. In this respect, Curran (1994) and Rogers and Shi (1995) very successfully applied a conditioning approach. Recently, Lord (2005) combined their approach with the traditional ad-hoc moment matching approaches, to obtain an approximation which is extremely accurate and has an analytical bound on its error. Here we review this approach and extend the results to multiple conditioning variables, along the lines of Vanmaele et al. (2004).

1. INTRODUCTION

This paper deals with the pricing of European options on arithmetic averages. If the average is a time average of a single underlying asset, these options are referred to as Asian options. Another possibility is that the average is taken over several assets at the same time instant; these options are referred to as basket options. Of course, mixtures of Asian and basket options exist in the market. For instance, the average could be taken over time and over several assets, creating what we will refer to as an Asian basket option.

The reason for the existence of such options is clear. As far as basket options are concerned, large companies may want to buy some downside protection on their investments. One possibility to achieve this would be to buy an option on a basket that is representative for the investments of the firm. What about Asian options, i.e. options on a time average of a single underlying? A pure European option on this asset would exhibit a large dependence on the final value of the underlying asset, and as such the option is quite sensitive to large shocks or price manipulation. To avoid such issues, many financial contracts often contain a so-called 'Asian tail', which means that the final payoff is based on the average price of the underlying over a time interval before the expiry date.

Recently Schrager and Pelsser (2004) have shown that unit-linked guarantees contain rate of return guarantees, which closely resemble Asian options. Given the fact that fair value calculations are currently the talk of the town, it is highly important to be able to value these types of contracts.

In the Black-Scholes model it is already not straightforward to price these types of contracts, the main reason for this being that no closed-form probability law exists for the sum of correlated lognormal random variables. As the valuation of these options is already mathematically interesting in the Black-Scholes model, many papers, including ours, are based in this lognormal setting. Within this model many an approach has been used to value or approximate these options. Broadly speaking we can divide these methods into five classes: approaches based on Monte Carlo simulation, the numerical solution of partial differential equations (PDEs), integral transforms, analytical approximations or analytical bounds. We will not attempt to give an overview of all these approaches here, we refer the interested reader to Lord (2005) and references therein.

In principle, the most flexible approach when considering multiple underlyings is probably a Monte Carlo simulation. Aside from the fact that it is very easy to implement, a large advantage is that we can easily allow for more realistic dynamics in the model. However, even though excellent control variates exist within the Black-Scholes model, the method can still be somewhat computationally intensive, not to mention the additional problem of computing sensitivities with respect to model and market parameters. Nowadays however, many structured products with a basket as their underlying, use caps and floors on the performance of individual underlyings, so that Monte Carlo simulation or the PDE approach is the only method that can be used. Here we ignore these additional features, and consider a simple European arithmetic Asian basket option.

Since financial institutions demand quick and accurate answers for the value of a derivative and its Greeks, large parts of the literature have focused on analytical approximations and bounds. Probably one of the most widely known and used approximations is that of Levy (1992), who approximates the arithmetic average with a lognormal random variable, such that the first two moments coincide with that of the true distribution. A problem that this approximation shares with other ad-hoc moment matching approaches, is that the size of their error is not known analytically. Furthermore, most of these approximations only tend to work well for low to moderate volatility environments. The first shortcoming has certainly motivated researchers to come up with sharp lower and upper bounds on the value of these options. A seminal paper in this area is that of Rogers and Shi (1995). In the context of Asian options, they derived a sharp lower bound and an upper bound on the value of an Asian option. The technique used to derive the lower bound is remarkably simple, but very effective — they condition on a variable that is highly correlated with the basket, and then apply Jensens inequality to find a very sharp lower bound. Curran (1994) arrived at the same lower bound, and was the first to observe that the payoff of Asian basket options can be split into two parts: one that can be calculated exactly, and one that has to be approximated. This approach yielded Currans so-called 'sophisticated' approximation. Recently, Lord (2005) showed that this approximation of Curran actually diverges when the strike price tends to infinity. Combining the ideas of Rogers and Shi and Curran, he introduces the class of partially exact and bounded (PEB) approximations, which are guaranteed to lie between Rogers and Shis lower bound, and a sharpened version of Rogers and Shis upper bound (due to Nielsen and Sandmann (2003) and Vanmaele et al. (2005)).

In this paper we will, in the next section, first review the conditioning approaches of Rogers and Shi and Curran. In Vanmaele et al. (2004) it is shown how to extend the lower bound of Rogers and Shi so that we can condition on two random variables. Here we trivially extend this lower bound, as well as the sharpened upper bound of Rogers and Shi, to allow for an arbitrary number of conditioning variables. In the third section we review the results of Lord (2005), which provide a clear motivation for conditional moment matching. Building on the results of the second section, we can extend the PEB approximations to allow for multiple conditioning variables. Finally, we show that a recent bounded approximation of Vanmaele et al. (2004), which also matches the first two conditional moments, satisfies all requirements of a PEB approximation. We end the paper with a brief numerical illustration and some conclusions and recommendations. The focus throughout the paper will not be on exact expressions required to calculate the various bounds and approximations, but on the rationale behind the approaches.

Before starting the next section, we will first introduce some notation. As mentioned, we will base ourselves in the Black-Scholes framework. For notational convenience we will work with a constant parameter Black-Scholes model, although all results still hold when these parameters are deterministic functions of time, and the growth and spot rate are Gaussian. We assume the underlying assets S_i, $i = 1, \ldots, N$ and the money market account M evolve according to the following stochastic differential equation (SDE):

$$\begin{aligned} \frac{dS_i(t)}{S_i(t)} &= \mu_i dt + \sigma_i dW_i(t) \\ \frac{dM(t)}{M(t)} &= rdt \end{aligned}$$

where all Brownian motions are correlated with instantaneous correlation matrix R. Throughout the document we will assume, without loss of generality, that the current date is 0. The underlyings of all options we consider in this paper will be an Asian basket, which at the maturity date T will be defined as $B(T)$ in:

$$\begin{aligned} B(T) &= \sum_{i=1}^{N} w_i A_i(T) \\ A_i(T) &= \int_0^T S_i(t)\rho_i(t)dt. \end{aligned}$$

Here, the weights w_i are positive and sum to 1, and similarly all ρ_i are non-negative functions, integrating to 1 over $(0, T)$. We will only consider newly issued, non-forward-starting call options on this Asian basket. This is no loss of generality. Put options can be priced via the Asian put-call parity, whereas running average options can be treated as newly issued ones, with a correction to the strike price. Finally, forward-starting options pose no problems when interest rates are deterministic or Gaussian. For ease of exposure we will mostly deal with forward prices in our analysis. The forward price of the Asian basket call option is equal to its expected value under the risk-neutral probability measure $\mathbb{E}_0^{\mathbb{Q}}$, conditional upon all information known at time 0:

$$c_B(T, K) = \mathbb{E}_0^{\mathbb{Q}}[(B(T) - K)^+]$$

In the remainder we will leave out the superscript indicating the measure and the subscript indicating at which time the expectation is evaluated, unless any confusion can arise. Having introduced the notations we will use, we are now ready to turn to the next section.

2. THE CONDITIONING APPROACHES

The most successful approximations and bounds all rely heavily on results first derived by Rogers and Shi (1995) and Curran (1994). We briefly review their approaches here, whereafter we extend them to allow for multiple conditioning variables, something which was done for two conditioning variables by Vanmaele et al. (2004). We will here use Currans idea of decomposing the Asian option into two parts: one that can be calculated exactly, and one that has to be approximated. Suppose that we have a normally distributed random variable Λ with the convenient property that $\Lambda \geq \lambda(K)$ implies that $B(T) \geq K$. Examples of such random variables will be given shortly. Following Curran, we can then write:

$$\begin{aligned} c_B(T,K) &= \mathbb{E}[(B(T)-K)^+ 1_{[\Lambda<\lambda(K)]} + (B(T)-K)^+ 1_{[\Lambda\geq\lambda(K)]}] \\ &= \mathbb{E}[(B(T)-K)^+ 1_{[\Lambda<\lambda(K)]} + (B(T)-K)\, 1_{[\Lambda\geq\lambda(K)]}] \\ &\equiv c_1(T,K,\Lambda) + c_2(T,K,\Lambda) \end{aligned}$$

As is shown in Curran (1994), it is quite straightforward to calculate the c_2-part, using the convenient property that normally distributed random variables are still normally distributed upon conditioning on a correlated normal random variable. We will therefore refrain from reproducing the exact formulae here. This leaves us with the calculation of the c_1-part, which we can bound or approximate. Let us first however consider several possible random variables Λ, which have the above property. A very natural candidate for such a Λ is the logarithm of the geometric average, which for the Asian basket will be defined as:

$$\begin{aligned} G(T) &= \prod_{i=1}^{N} G_i(T)^{w_i} \\ G_i(T) &= \exp\left(\int_0^T \ln S_i(t)\rho_i(t)dt\right) \end{aligned}$$

An application of the weighted Jensens inequality shows that $B(T) \geq G(T)$, with equality attained if and only if all components of the average are equal. Defining $\Lambda_{GA} = lnG(T)$, it is then obvious that when $\Lambda_{GA} \geq lnK$, we indeed have $B(T) \geq K$. Other possible conditioning variables, see e.g. Vanmaele et al. (2004), arise from a first order approximation of the Asian basket $B(T)$ in its driving Brownian motions. In the setting of an Asian basket option, we then obtain the following conditioning variables and their corresponding thresholds:

$$\begin{aligned} \Lambda_{FA1} &= \sum_{i=1}^{N} w_i \int_0^T S_i(0)\exp\left((\mu_i - \frac{1}{2}\sigma_i^2)t\right)\cdot(1+\sigma_i W_i(t))\,\rho_i(t)dt \\ \Lambda_{FA2} &= \sum_{i=1}^{N} w_i \int_0^T S_i(0)\cdot\left(1+(\mu_i-\frac{1}{2}\sigma_i^2)t+\sigma_i W_i(t)\right)\rho_i(t)dt \end{aligned} \tag{1}$$

The corresponding thresholds are in both cases equal to $\lambda_{FA1}(K) = \lambda_{FA2}(K) = K$. We note that the higher the correlation of Λ with $B(T)$ is, the larger the relative contribution of c_2 to the option price will be. In practice, any of the above conditioning variables is quite highly correlated with $B(T)$, provided that the volatilities of the underlying assets are not too high. This is one of the key points as to why these conditioning approaches work so well - c_2 constitutes a large part of the option price, so that any approximation we make in c_1 will not have a large impact. The larger the volatilities and maturities are, the more important it becomes to approximate c_1 accurately.

We now turn to the approximating part c_1. Both Rogers and Shi and Curran used Jensens inequality to find a lower bound on the value of these options. A lower bound on c_1 simply follows from:

$$\begin{aligned} c_1(T,K,\Lambda) &= \mathbb{E}\left[(B(T)-K)^+ 1_{[\Lambda<\lambda(K)]}\right] \\ &= \mathbb{E}[\mathbb{E}[(B(T)-K)^+ 1_{[\Lambda<\lambda(K)]} \mid \Lambda]] \\ &\geq \mathbb{E}[\mathbb{E}[(B(T)-K)\, 1_{[\Lambda<\lambda(K)]} \mid \Lambda]^+] \end{aligned}$$

so that then the lower bound becomes the sum of this lower bound and c_2:

$$\begin{aligned} LB(T,K,\Lambda) &= \mathbb{E}[(B(T)-K)^+ 1_{[\Lambda<\lambda(K)]}] + c_2(T,K,\Lambda) \\ &= \mathbb{E}\left[(\mathbb{E}[B(T) \mid \Lambda] - K)^+\right] \end{aligned} \tag{2}$$

This lower bound can in principle be applied using an arbitrary conditioning variable, not only conditioning variables for which we have the aforementioned property. In Lord (2005) it is shown how to calculate (2) in closed-form for an arbitrary conditioning variable, and an arbitrary correlation structure between the various underlyings. This greatly facilitates the computations required for the lower bound, as otherwise we would have to resort to a numerical integration over a discontinuous integrand.

Another approach to approximate c_1 will be pursued in the following section. We will now extend the lower bound so that we can condition on multiple random variables. For two conditioning variables this idea was first pursued in Vanmaele et al. (2004), so this is merely a trivial extension of their results. Suppose that we have a conditioning variable Λ and a set of conditioning variables $\mathcal{Z}$, such that for any realisation of the random variables in $\mathcal{Z}$, $\Lambda \geq \lambda(K)$ implies that $B(T) \geq K$. The lower bound on c_1 then becomes:

$$\begin{aligned} c_1(T,K,\Lambda) &= \mathbb{E}[(B(T)-K)^+ 1_{[\Lambda<\lambda(K)]}] \\ &= \mathbb{E}\left[\mathbb{E}[(B(T)-K)^+ 1_{[\Lambda<\lambda(K)]} \mid \Lambda,\mathcal{Z}]\right] \\ &\geq \mathbb{E}\left[\mathbb{E}[(B(T)-K) \mid \Lambda,\mathcal{Z}]^+ \cdot 1_{[\Lambda<\lambda(K)]}\right] \end{aligned}$$

so that the resulting lower bound is:

$$LB(T,K,\Lambda,\mathcal{Z}) = \mathbb{E}\left[\mathbb{E}[(B(T)-K) \mid \Lambda,\mathcal{Z}]^+ \cdot 1_{[\Lambda<\lambda(K)]}\right] + c_2(T,K,\Lambda) \tag{3}$$

Note that the first part in (3) will typically have to be calculated via a multivariate numerical integration, whereas the second part is the same as before, and can hence be done in closed-form. Let us now turn to an analysis of the error made by approximating the value of the Asian basket option by the lower bound in (3). This upper bound, based on the lower bound, was first derived by Rogers and Shi (1995). It is based on the following inequality:

$$\begin{aligned} 0 &\leq \mathbb{E}[X^+] - \mathbb{E}[X]^+ = \frac{1}{2}\left(\mathbb{E}\left[|X|\right] - |\mathbb{E}[X]|\right) \\ &\leq \frac{1}{2}\mathbb{E}\left[|X-\mathbb{E}[X]|\right] \leq \frac{1}{2}\sqrt{Var(X)} \end{aligned}$$

More recently, Nielsen and Sandmann (2003) and Vanmaele et al. (2005) sharpened this upper bound considerably. We here extend their sharpened version to allow for multiple conditioning variables. Proceeding as above, we find:

$$\begin{aligned} 0 &\leq c_B(T,K) - LB(T,K,\Lambda,\mathcal{Z}) \\ &= \mathbb{E}\left[\mathbb{E}[(B(T)-K)^+ 1_{[\Lambda<\lambda(K)]} \mid \Lambda,\mathcal{Z}] - \mathbb{E}[(B(T)-K)\, 1_{[\Lambda<\lambda(K)]} \mid \Lambda,\mathcal{Z}]^+\right] \\ &\leq \frac{1}{2}\mathbb{E}\left[Var\left(B(T)1_{[\Lambda<\lambda(K)]} \mid \Lambda,\mathcal{Z}\right)^{1/2}\right] \\ &\equiv \epsilon_1(T,K,\Lambda,\mathcal{Z}) \end{aligned} \tag{4}$$

This yields an upper bound which again has to be calculated via a multivariate numerical integration. Nielsen and Sandmann and Vanmaele et al., using only one conditioning variable, go one step further to derive a slightly larger upper bound, that can be calculated in closed-form. Here it is equal to:

$$\begin{aligned} \epsilon_1(T,K,\Lambda,\mathcal{Z}) &= \frac{1}{2}\mathbb{E}[Var\left(B(T)1_{[\Lambda<\lambda(K)]} \mid \Lambda,\mathcal{Z}\right)^{1/2}] \\ &\leq \frac{1}{2}\sqrt{\mathbb{E}[Var\left(B(T) \mid \Lambda,\mathcal{Z}\right)\cdot 1_{[\Lambda<\lambda(K)]}]\cdot\mathbb{E}[1_{[\Lambda<\lambda(K)]}]} \\ &\equiv \epsilon_2(T,K,\Lambda,\mathcal{Z}) \end{aligned} \tag{5}$$

Both error estimates yield an upper bound which is equal to $UBi(T,K,\Lambda,\mathcal{Z}) = LB(T,K,\Lambda,\mathcal{Z}) + \varepsilon_i(T,K,\Lambda,\mathcal{Z})$, for $i = 1,2$. It can be calculated in closed-form because we can write:

$$\mathbb{E}[Var\left(B(T) \mid \Lambda,\mathcal{Z}\right)\cdot 1_{[\Lambda<\lambda(K)]}] = \mathbb{E}\left[\mathbb{E}\left[Var\left(B(T) \mid \Lambda,\mathcal{Z}\right) \mid \Lambda\right]\cdot 1_{[\Lambda<\lambda(K)]}\right]$$

As shown in Nielsen and Sandmann and Vanmaele et al., this expression can be calculated in closed-form. We do not reproduce the formulae here, as it only distracts from the rest of the text and the calculations are exactly the same as in the aforementioned articles. Note that the variance of $B(T)$ given Λ and $\mathcal{Z}$ is zero if the set $\{\Lambda,\mathcal{Z}\}$ contains all random variables within $B(T)$. Then the results above imply that the lower bound exactly coincides with the true value of the Asian basket option. Finally, we mention that Rogers and Shis upper bound corresponds to the limit of UB_1 for K tending to infinity.

3. THE BENEFITS OF CONDITIONAL MOMENT MATCHING

As mentioned in the introduction, many original approximations merely substitute the arithmetic average by a tractable random variable, which has the same first couple of unconditional moments. An example of this is Levy (1992) approximation, which fits a lognormal random variable to the arithmetic average. These types of approximations typically only work well when volatilities and maturities are low. Furthermore, the size of the error made can not easily be estimated. Here we show that conditional moment matching does yield an analytical error estimate. The proposed approximation follows from:

$$\begin{aligned} \tilde{c}_B(T,K,\Lambda,\mathcal{Z}) &= \mathbb{E}[\left(\tilde{B}(T)-K\right)^+ 1_{[\Lambda<\lambda(K)]} + (B(T)-K)\, 1_{[\Lambda\geq\lambda(K)]}] \\ &\equiv \tilde{c}_1(T,K,\Lambda) + c_2(T,K,\Lambda) \end{aligned} \tag{6}$$

i.e. it again exists of an approximating part and an exact part. For $\Lambda \geq \lambda(K)$ we can take our approximating random variable $\tilde{B}(T)$ to be equal to $B(T)$, yielding the exact c_2-part. For λ smaller than $\lambda(K)$, we have to make an approximation. Given certain criteria that $\tilde{B}(T)$ must fulfill, which follow in the next theorem, we can find an analytical error estimate as derived in Lord (2005). Here we extend this result to allow for multiple conditioning variables.

Theorem 3.1 *If we impose the following two conditions on the approximating random variable* $\tilde{B}(T)$*:*

$$\begin{aligned} \mathbb{E}[\tilde{B}(T) \mid \Lambda = \lambda, \mathcal{Z} = z] &= \mathbb{E}[B(T) \mid \Lambda = \lambda, \mathcal{Z} = z] \\ Var[\tilde{B}(T) \mid \Lambda = \lambda, \mathcal{Z} = z] &\leq Var[B(T) \mid \Lambda = \lambda, \mathcal{Z} = z] \end{aligned} \tag{7}$$

for $\lambda \in (-\inf, \lambda(K))$*, the resulting approximation in (6) lies between* $LB(T, K, \Lambda, \mathcal{Z})$ *and* $UB_i(T, K, \Lambda, \mathcal{Z})$*.*

Proof:
The proof follows along the same lines as (4)-(5):

$$\begin{aligned} 0 &\leq \tilde{c}_B(T, K, \Lambda) - LB(T, K, \Lambda, \mathcal{Z}) \\ &= \mathbb{E}\left[\mathbb{E}[\left(\tilde{B}(T) - K\right)^+ 1_{[\Lambda<\lambda(K)]} \mid \Lambda, \mathcal{Z}] - \mathbb{E}[\left(\tilde{B}(T) - K\right) 1_{[\Lambda<\lambda(K)]} \mid \Lambda, \mathcal{Z}]^+\right] \\ &\leq \frac{1}{2}\mathbb{E}\left[Var\left(\tilde{B}(T)1_{[\Lambda<\lambda(K)]} \mid \Lambda, \mathcal{Z}\right)^{1/2}\right] \leq \epsilon_1(T, K, \Lambda, \mathcal{Z}) \end{aligned}$$

It is clear that the first equality holds, due to the construction in (6) and the fact that the conditional moments are equal for $\Lambda \leq \lambda(K)$. The rest of the derivation is similar to (4)-(5). It immediately follows that:

$$LB(T, K, \Lambda, \mathcal{Z}) \leq \tilde{c}_B(T, K, \Lambda, \mathcal{Z}) \leq UB_1(T, K, \Lambda, \mathcal{Z})$$

which concludes the proof of the theorem. $\square$

This theorem directly motivates why it is good to match conditional moments. Intuitively we can indeed expect to obtain better results than by just matching unconditional moments. The above theorem gives a rigorous (and typically sharp) error bound for this. Note that the moments do not have to be exactly matched - the conditional variance may actually be smaller. Approximations satisfying (6) and (7) are dubbed partially exact and bounded (PEB) approximations. The lower bound $LB(T, K, \Lambda, \mathcal{Z})$ is a special case hereof. In Vanmaele et al. (2004) another route is attempted. Without delving into details, they construct an approximation via a (conditionally) convex combination of the lower bound and the partially exact and comonotonic upper bound (PECUB), the so-called LBPECUB approximation. The conditional weights are chosen by ensuring that the first two conditional moments are matched exactly. As such, it satisfies the criteria for it to be a PEB approximation, and hence it is bounded above by the UB_1 as well as of course by the PECUB upper bound.

We note that in practice the approximating part in (6) will have to be calculated via a numerical integration. From a computational point of view one would therefore not like to use too many conditioning variables. Typically one conditioning variable may already be more than enough, as has been shown in Lord (2005) for a pure Asian option, and as we will demonstrate for a pure basket option in the next and final section.

4. NUMERICAL ILLUSTRATION AND CONCLUSIONS

To illustrate the effectivity of conditional moment matching we will here provide a numerical example for a pure basket option. The example has been taken from Milevsky and Posner (1998), and also features in Vanmaele et al. (2004). The basket underlying the option is the weighted average of the normalized G-7 stock indices. Weights, volatilities, dividend yields and correlations can be found in the tables below.

Country	**Index**	**Weight**	**Volatility**	**Dividend yield**
Canada	TSE 100	10%	11.55%	1.69%
Germany	DAX	15%	14.53%	1.36%
France	CAC 40	15%	20.68%	2.39%
U.K.	FTSE 100	10%	14.62%	3.62%
Italy	MIB 300	5%	17.99%	1.92%
Japan	Nikkei 225	20%	15.59%	0.81%
U.S.	S&P 500	25%	15.68%	1.66%

Table 1: Weights, volatilities and dividend yields of the basket

	Canada	*Germany*	*France*	*U.K.*	*Italy*	*Japan*	*U.S.*
Canada	1.00	0.35	0.10	0.27	0.04	0.17	0.71
Germany		1.00	0.39	0.27	0.50	−0.08	0.15
France			1.00	0.53	0.70	−0.23	0.09
U.K.				1.00	0.45	−0.22	0.32
Italy					1.00	−0.29	0.13
Japan						1.00	−0.03
U.S.							1.00

Table 2: Upper triangular part of the instantaneous correlation matrix

As we use the normalized values of the indices, this effectively means we assume the initial spot value equals 1 for each index. In the following table we compare the lower and upper bounds using one or two conditioning variables to the 'true' value obtained from a Monte Carlo simulation with 5 000 000 paths, using antithetic variables and using the geometric basket as the control variate. Results are only shown for the most extreme example in Vanmaele et al., namely for a maturity of 10 years. Vanmaele et al. only considered three strike prices, 0.95, 1 and 1.05. However, the forward price of the basket (the mean of its unconditional distribution) can be calculated as 1.578288, so that we found it important to include higher strike prices in the table as well. The choices for conditioning variables are the same as in this article, the first conditioning variable is FA_1 (cf. (1)), the second is similar to FA_2, apart from the fact that the sign of the one but last Brownian motion is reversed.

Strike	**MC**	**(StdErr)**	$\mathbf{LB_{FA1}}$	$\mathbf{UB_{FA1}}$	$\mathbf{LB_{FA1,FA2*}}$	$\mathbf{UB_{FA1,FA2*}}$
0.95	33.6590	(0.0036)	33.6221	33.8765	33.6461	33.7814
1	31.1333	(0.0037)	31.0758	31.3974	31.1103	31.2849
1.05	28.6603	(0.0038)	28.5870	28.9879	28.6335	28.8553
1.25	19.6748	(0.0042)	19.5002	20.3444	19.6046	20.1044
1.5	11.1785	(0.0045)	10.8999	12.5203	11.0481	12.0654
1.578288	9.1918	(0.0044)	8.8975	10.7762	9.0927	10.2872

Table 3: Upper and lower bounds based on one or two conditioning variables

The upper bounds are the UB_2 upper bounds, see equation (5). Indeed, conditioning on more random variables sharpens the lower and upper bounds considerably, as was already demonstrated in Vanmaele et al., but is now also apparent from the new upper bound.

To show that conditional moment matching actually works remarkably well, we compare various conditional moment matching approximations to the true value of the option. As mentioned earlier, the LBPECUB approximations considered in Vanmaele et al. (2004) are convex combinations of the lower bound and the PECUB upper bound. Results in their paper were shown for using the geometric average as the conditioning variable. Two distinctions can be made on the choice of the weights for each bound: $z(\lambda)$ indicates that the first two conditional moments are matched exactly (yielding a PEB approximation as noted in section 3), whereas z^u indicates that the lower bound and the PECUB upper bound are weighted using a global weight stemming from another approximation in Vyncke et al. (2004). The latter is not a conditional moment matching approximation, but it works rather well.

The Curran2M+ and Curran3M+ approximations are PEB approximations considered in Lord (2005) that fit a shifted lognormal random variable to the basket. The 2M+ approximation considered here uses a shift equal to the conditioning variable Λ_{FA1}; the remaining two parameters are chosen such that the first two conditional moments are matched exactly. The 3M+ approximation is a slight take on this: the shift is now also considered as a parameter, so that the first three conditional moments can be fit exactly. In both approximations we condition on Λ_{FA1}.

Strike	**MC**	**(StdErr)**	$\mathbf{LBPECUB_{GA}}$		**2M+**	**3M+**
			$\mathbf{z}(\lambda)$	$\mathbf{z}^u$		
0.95	33.6590	(0.0036)	33.6379	33.6543	33.6682	33.6612
1.00	31.1333	(0.0037)	31.0971	31.1239	31.1389	31.1318
1.05	28.6603	(0.0038)	28.6149	28.6554	28.6693	28.6631
1.25	19.6748	(0.0042)	—	—	19.6704	19.6769
1.50	11.1785	(0.0045)	—	—	11.1606	11.1768
1.578288	9.1918	(0.0044)	—	—	9.1791	9.1919

Table 4: Several approximations for the value of a basket option

We did not get round to implementing the LBPECUBGA approximation ourselves, so that we here only reproduce the values given in Vanmaele et al. They only considered strike values up to 1.05; as the forward price of the basket is 1.578288 for a 10-year contract, we also found it important to consider slightly higher strike prices. As can be seen from the table, the 3M+ approximation seems to give results that are very close to the true values and this has only been achieved by using one conditioning variable. The conditional moment matching approximation of Vanmaele et al., using $z(\lambda)$, seems to yield too low values. However, their approximation which uses z^u is a clear contender, yielding results which are comparable to those of the 2M+ approximation. Considering the computational effort, which has been investigated in Lord (2005), we have a slight overall preference for the 2M+ approximation, although this is of course subject to discussion.

Concluding, in this paper we reviewed the conditioning approaches of Rogers and Shi (1995) and Curran (1994). Rogers and Shis lower and (sharpened) upper bounds, as well as the PEB approximations of Lord (2005), have been extended along the lines of Vanmaele et al. (2004) to allow for multiple conditioning variables. Finally, we have shown that the LBPECUB convex combination of the lower bound and the PECUB upper bound, considered in Vanmaele et al., is indeed a PEB approximation, and as such is bounded from above by the (sharpened) Rogers and Shi upper bound. In a numerical example the effectivity of conditional moment matching has been demonstrated.

References

M. Curran. Valuing Asian and portfolio options by conditioning on the geometric mean price. *Management Science*, 40(12):1705–1711, 1994.

E. Levy. Pricing European average rate currency options. *Journal of International Money and Finance*, 11:474–491, 1992.

R. Lord. Partially exact and bounded approximations for arithmetic Asian options. *submitted, URL http://www.few.eur.nl/few/people/lord/PEBAsian.pdf*, 2005.

M.A. Milevsky and S.E. Posner. A closed-form approximation for valuing basket options. *Journal of Derivatives*, 4:54–61, 1998.

J.A. Nielsen and K. Sandmann. Pricing bounds on Asian options. *Journal of Financial and Quantitative Analysis*, 38(2):449–473, 2003.

L.C.G. Rogers and Z. Shi. The value of an Asian option. *Journal of Applied Probability*, 32: 1077–1088, 1995.

D.F. Schrager and A.A.J. Pelsser. Pricing rate of return guarantees in regular premium unit linked insurance. *Insurance: Mathematics and Economics*, 35(2):369–398, 2004.

M. Vanmaele, G. Deelstra, and J. Liinev. Approximation of stop-loss premiums involving sums of lognormals by conditioning on two variables. *Insurance: Mathematics and Economics*, 35(22): 343–367, 2004.

M. Vanmaele, G. Deelstra, J. Liinev, J. Dhaene, and M.J. Goovaerts. Bounds for the price of discretely sampled arithmetic Asian options. *forthcoming in: Journal of Computational and Applied Mathematics*, 2005.

D. Vyncke, M. Goovaerts, and J. Dhaene. An accurate analytical approximation for the price of a European-style arithmetic Asian option. *Finance*, 25:121–139, 2004.

ON MEASURING THE EFFICIENCY OF THE SOCIAL SECURITY SYSTEM REFORMS. THE CASE OF POLAND

Piotr Mularczyk[§,†] **and Joanna Tyrowicz**[§]

[§]*Faculty of Economics, Warsaw University, Poland*
[†]*Faculty of Applied Economics, Katholieke Universiteit Leuven, Belgium*
Email: `PMularczyk@wne.uw.edu.pl, JTyrowicz@wne.uw.edu.pl`

Abstract

As other European countries, transition economies face the reform of the social security system. As one of the first, Poland has introduced a pension reform in 1999, which changed a standard pay-as-you-go system into one constructed of three pillars and based on addressed contributions. The five years since the reform allow us to take a first look at the reform, both in terms of assessing the legal implementation as well as the realization of main assumptions and aims. In this paper we consider the effectiveness of the reform. We find that in many aspects this reform should not be considered successful.

1. INTRODUCTION

To a probably larger extent than in Western European countries, transition economies face the necessity to reform their social security systems. Movement towards market-based economic setting accompanied by demographic changes, imposes an intense time constraint on their governments to design and implement a new system. However, the particular characteristics of the new solutions as well as the very foundations of the new system are not as evident as one might think.

There were two fundamental prerequisites for the reform in Poland: on one hand the state pay-as-you-go (PAYG) system seized being capable of serving future generations for demographic reasons. On the other hand, the national saving rate was believed to be too low. In addition, reform was expected to ameliorate the problem of capital market shallowness and increase the efficiency as well as the transparency of national savings management. In this paper we attempt to verify whether these major aims were satisfied. We analyse the institutional set-up of the new system and point to the inefficiency of the current pension system, as well as its ineffectiveness since not only is it relatively costly but also provides no risk hedge. The major hypothesis we try to support is that from the users' point of view **the current system is in its idea equivalent to the obligatory bank savings, although significantly more expensive**.

The paper is organised as follows. We first approach the officially acknowledged reasons for the reform, thereby suggesting the initial requirements for the new system. In the second section we discuss the macroeconomic and demographic characteristics of Poland which serves the purpose of presenting the social security landscape on the eve of reform. We further move to the micro level foundations, analysing the new system in section three. Eventually, we approach the issue of efficiency, presenting the observed performance of the system and modelling its cost to the beneficiaries. The paper is rounded of by conclusions.

2. REASONS FOR THE REFORM

Most European countries maintain a pay-as-you-go retirement system, sometimes allowing for tax redemptions in case of additional participation in non-obligatory and private pension plans. The pay-as-you-go system has emerged as a result of the welfare economics doctrine. With the post-war demographic booms and trente gloires of economic expansion, outlooks for this solution seemed not only sustainable but also strictly dominating other alternatives in terms of political consensus and social support. The problems became evident as late as in 1980s when in most European countries demographic structure started to change in the adverse direction hazarding the liquidity of social security funds.

Poland is by no means exception from this pattern. Due to changes in population age scheme, social security system required subsidisation during the entire 1990s, with numerous examples of insolvency and delayed benefit payments. High unemployment (both observed unemployment and the 'voluntary' one resulting from foreign transmittances) have further deteriorated the situation of society's changing demographic structure. In addition, there is also an issue of savings habits which constituted a serious problem for capital formation process.

2.1. Demographic issues

As in the rest of Europe, the demographic structure of Polish society was gradually changing. The post-war population boom started approaching the retirement age, whereas new generations are less numerous. The shift is clearly visible both for men and women. The decrease in number of births is accompanied by a generally longer life time. In addition, for both genders, the population of between the age of 20 and 35 years is decreasing, whereas the opposite may be observed in the range of people aged 35 to 60 years. These elements considered together with the growing unemployment, combine to a serious downfall in contributions to the social security funds.

This process is accompanied by various structural changes in the Polish society. Traditionally, farmers' contributions to the social security fund are rather symbolic, whereas welfare of this social group vitally depends on the state transfers. Secondly, an increasingly important share of unemployment originates from the structural mismatch. On the top, Polish economy was gradually moving from 7% real GDP growth rate in 1997 to much lower levels owing to the general global economic slowdown.

2.2. Savings

For the forty years of the post-war period the economic system created little incentives for postponing consumption. First, in the presence of the socialist state, citizens rarely take up the responsibility for their financial condition. Secondly, the trust in the value of Polish money was rather diminishable — the main currencies for wealth accumulation were US dollar and German mark, none of which could be legally deposited above certain predefined and extremely low level. On the other hand, the financial system itself was rather a tease, since there were no internal financial markets and international capital flows were strictly controlled.

In addition, trust in national institutions in this respect was significantly decreased by 1956 'reform', when the government overnight devaluated the currency by 60% without any compensation. The observed growing disparities between internal and external prices and skyrocketing black market premium further diminished the propensity to institutionalise savings. In addition, there are also psychological aspects in play — in case of a centrally planned economy many goods and services were simply underprovided, which excessively raised consumption in case of accidental availability. In addition, the 1990s brought a period of hyperinflation, while at the beginning deposits were not indexed.

2.3. Financial markets

The Polish economy is troubled by a capital gap. Continuous inflow of foreign investment partly alleviated the problem but ushered trade balance deficit and increased vulnerability of the Polish economy to foreign shocks. In addition, the assets available on the markets were so far insufficient to induce deepening of the financial markets by foreign investors, which is often raised in public debates[1]. Hence, home investors are believed to be crucial to facilitate the process of developing the Polish financial markets.

However, with virtually absent private pension funds (private plans are rare even among top tier managers) there were only banks and investment funds. The first focus on the credit activity. As to the latter, the participation in the stock market is still low among the citizens - lower than the European average, which is in general much lower than the US levels. Shifting retirement savings from the Social Security Fund to private and independent pension funds was believed to trigger the process of deepening the Polish financial markets (Giunrken 1999).

[1] An extensive coverage of these debates may be found on Polish Insurance Portal (www.ubezpieczyciel.pl, in Polish).

3. THE NEW SYSTEM VERSUS ITS ASSUMPTIONS

In 1999 government introduced four general reforms: health care, education, administration and social security system, all of which required creating a new institutional architecture as well as introducing new rules concerning contributions and benefits. Five principles can be identified as having guided these changes (Gomulka and Styczen 1999). They have laid the foundation for the new system.

- **The diversification principle:** the necessity of enhancing the security and efficiency of the pension system should be obtained by diversifying the system between state and private funds as well as between obligatory and voluntary contributions. The system of retirement pensions for workers consists of three pillars: (I) state PAYG, (II) private open pension funds and (III) other pension funds. Only pillars II and III are private, whereas only pillars I and II receive obligatory contributions (see figure 1).

- **The distribution principle:** maintain the PAYG rule in the public part of the new system, while making it less redistributive and more transparent in order to immune the system to temporary political pressures. Obligatory contributions remain proportional to earnings (a form of a payroll tax, 19.52% of gross wage), but are also a subject to a cut-off point equal to 2.5 times the average wage. Half of the contribution is paid by employers and the other half by employees and the payment is made upfront each month.

- **The capital-funding principle:** make it capital funded as well as induce appropriate regulation of the private part. The private pension funds which constitute pillar II receive a specified part of all the obligatory contributions (7.3% of gross wage, the remainder is left in the PAYG state pillar I). The efficiency of the pension funds is recognised in two ways: contributors are freely allowed to change the fund and a floor was imposed on the fund triennial rates of return.

- **The savings principle:** the savings measures introduced must cover the cost of the reform. The reform clearly implies the deterioration of the public finance in the initial phase, as part of the obligatory contributions is diverted to pillar II. To contain this deterioration, supplementary reforms were introduced, namely: tighter criteria for handicapped benefits, raising the effective retirement age, reducing the scope for early retirement and reducing the growth rate of benefits.

- **The gradual phasing-in principle:** the phasing-in of the new system must be spread over a prolonged period of time, and should not involve people near retirement. A gradual phasing-in of the reform is intended, above all, to protect the rights of older workers (actual founders of Polish economy, as it was destroyed in nearly 80% after the second world war), as they remain in the old system. On the other hand this helps to limit the cost of reform implementation.

As the authors of the reform admit, the necessary process of shifting away from the inefficient and unsustainable PAYG system was prolonged over at least 20 years. In addition, even after this period, PAYG remains one of the pension sources, providing 30% of the benefit[2].

3.1. The architecture of the new system

According to the rules outlined above, each citizen born after 1969 is obliged to participate in both I and II pillar (the choice of II pillar is left to the worker). Citizens born between 1949 and 1969 are free to choose, whether they join the II pillar at all or remain in the PAYG system solely (regardless of their choice, their pension is calculated according to new principles). Finally, citizens born before 1949 do not have the choice, as they are obligatory members of the PAYG system[3]. The government is the lender of last resort for the second pillar, which effectively neglects the risk of loosing savings in result of a bankruptcy (Nowak 1999).

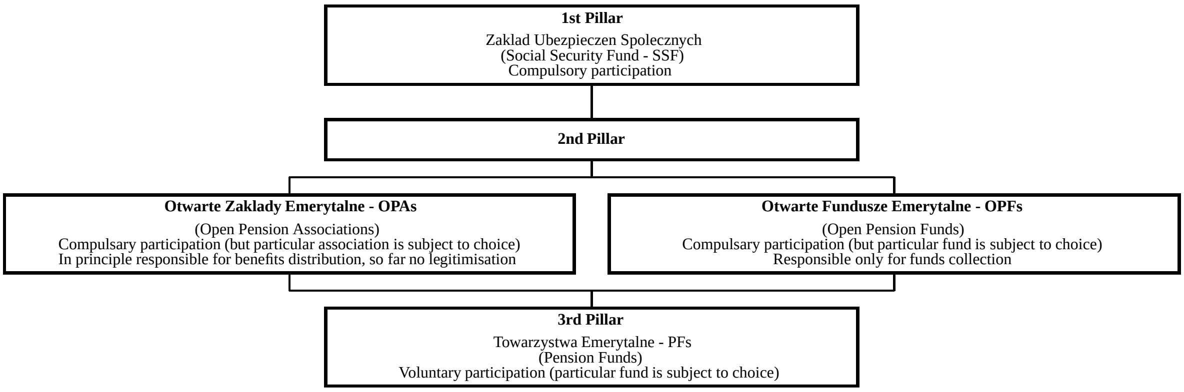


Figure 1: The architecture of the New Social Security System

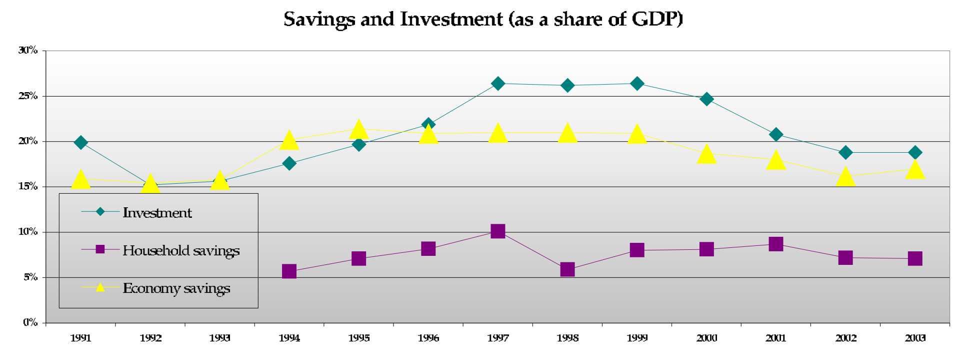
Everyone, regardless of age, is entitled to join any pension fund within the scheme of the III pillar. However, contribution is limited to 7% of wage and is not tax deductible. In addition, the solution of Individual Pension Accounts was introduced in 2004, allowing to deposit voluntary contributions in banks and financial institutions (the equivalent of approximately 800-850 euro on the annual basis) with the redemption of the tax on interest.

3.2. The impact on savings

The analysis of the household behaviour shows a considerable growth in the savings rate in the period prior to the reform. However, although official statistics exhibit the growth in total economy savings from 16% of GDP in 1991 to 21% in 1997, with the average real GDP growth rate for this period exceeding 7%, reality is much less optimistic (Liberda 1999). Firstly, 1996 brought the change in the system of national account and in this year savings to GDP ratio amounted to ap-

[2]The problem of employed in the agriculture remains essentially unaddressed, as the scope of reform excluded this social group from the reform thus strengthening the persistence of the old system, where the incompatibility between contributions and benefits is especially visible — in the agriculture pension system mainly serves the purpose of providing the safety net to this social group.

[3]The payment scheme, however, has been changed in order to facilitate maintaining liquidity in the future.



Source: Annual reports of Institute of Business Cycles, Prices and Trade (Instytut Koniunktur, Cen i Handlu Zagranicznego) and Liberda (2000).

Figure 2: Savings and investment (as a share of GDP).

proximately 18%. Secondly, this is mainly attributable to the institutionalisation of the previously unofficial savings. More realistic estimates support the thesis of 2-3 percentage points increase. Furthermore, this tendency was reversed in the later years yielding the persistent average saving rates of 17-18% (Liberda 2003).

Authors of the reform expected to induce a change in household savings behaviour by introducing compulsory participation in pension funds and thus strengthening the expansion potential of the private pension plans. Obviously, since propensity to save may crucially depend upon the GDP growth rate, the economic slowdown might have overshadowed the impact of the reform. However, one seems to be unable to prove this hypothesis based on available data. In other words, although one cannot state that reform failed to achieve the aim of stimulating household savings, the opposite seems largely disputable.

3.3. The impact on the depth of the capital market

With the wind of 1989 and gradual development of the banking system, some changes occurred. Despite initially strong impact of inflation, people eagerly opened accounts in new, commercial banks. The popularity of vista accounts and short term deposits is believed to be the key factor of success in the stage of forming financial markets as well as in the initial investment process (Jurkowski 2001). However, Polish economy suffers from a capital gap, which necessitates implementing a strategy aiming at introducing capital accumulating institutions and creating investment friendly environments.

Unfortunately, OPFs are rather inefficient in transforming the participants' savings into investment capital. They avoid direct capital involvement (via for instance capital funds) and are reluctant as stock exchange players, the latter being partly justifiable by the shallowness of the stock market. More importantly, government and communal bonds continuously contribute to almost 70% of their assets. Furthermore, these bonds are rarely purchased on the primary market realising the buy-and-hold strategy.

Based on these results it is difficult to support for the thesis of financial market deepening argument. Although all OPFs investment occurs exclusively via the market mechanism, this group of investors is not particularly active on the markets — neither risk-less nor involving significant uncertainty[4].

4. THE VALUE-FOR-MONEY APPROACH AS AN EFFICIENCY CRITERION

As presented in the previous section, the reform has not achieved its initial aims. One lacks the ground to state that it significantly affected the savings patterns, while it only marginally contributed to deepening the financial markets. Above all, with the current status, pension funds reform did not introduce the risk sharing mechanisms leaving the problem of benefit payouts essentially unaddressed. OPFs are not pension funds in the canonical understanding of the word, as they only play a role of contribution collectors and managers.

Despite all these shortcomings the new system involving Open Pension Funds might provide benefits in terms of value-for-money analysis. To be able to verify this hypothesis we need to specify a benchmark case. For simplicity we assume a slightly peculiar and perfidious example of virtually no system at all — instead of introducing the pension funds the system would impose an obligation to deposit monthly exactly the same contributions on a bank account without the right to retrieve these holdings. We further assume that there is no cost associated to this saving strategy, while in return it brings the average long term commercial bank deposit rate.

The choice of the benchmark was dictated by the observation emphasised in the previous sections that OPFs are only responsible for collecting and managing the contributions but they bear no actuarial risk. The main advantage the new system enjoys over other solutions in this respect concern the capital adequacy rules and investment prudence. Thus, although a bank account would not be as secure without additional regulations, in predictable conditions the new system is functionally not very distant from the current Polish solution.

In this section we compare the Open Pension Funds to this virtual benchmark case. We briefly present the costs associated with this institutional solution, analysing at the same time whether the current regulation introduces incentives for both efficiency of savings management and cost rationalisation. Lastly, we try to specify minimum conditions for future OPFs performance to insure strict domination of this solution[5].

[4]The results suggest also that this is a consistent pattern across the funds implying low incentives for investment strategies diversification (Mularczyk and Tyrowicz 2004). One needs to admit that in the previous system, current contributions were neither accumulated nor invested. If the PAYG system liabilities were below contemporaneous contributions, the surplus was taken over by the budget. In the case of Social Security Fund liquidity problems, the government covered the deficit. Currently, contributions constitute savings and although they still finance the budget deficit, it occurs via the market mechanism.

[5]Mularczyk (2002) tested for levels of substitution between men and women in the first pillar, pointing also to the weaknesses of the floor mechanism - the results suggest that the system discriminates between the participants for the moment of joining which cannot become a decision variable. Furthermore, as an experiment demonstrated, protection against speculation on the event of fund bankruptcy is also weak with time. Mularczyk and Tyrowicz (2004), performed a panel data analysis basing on the Markowitz model, indicating that funds benchmark poorly against the market.

4.1. The costs of the OFPs

Open Pension Funds are privately funded institutional investors as well as legal entities obliged by the act of 1997 to perform certain reporting activities. Consequently, they face two types of costs - those associated with organising and operating a financial institution (benefits collection, funds management, etc.) and t hose following from regulator decision (essentially reporting to the Pension Funds Superintendence as well as participants). They may seek two sources to cover these costs: own capital and fees charged to the beneficiaries (Hadyniak and Monkiewicz 1999, Nowak 1999). Whereas the first option is rarely applied for the obvious reasons, the latter is strongly regulated.

First, OPFs are entitled to a certain part of each first contribution. This rate is decreasing with the participation in the fund. The regulator assumes that being a member of a fund for a longer period of time entitles the participant to a special treatment, which is reflected in the fee charged. The details are given in table 1.

Year of participation	1	2	3	4	5	6-10	11-15	16-20	21-25	26	27	28-...
Average fee (%)	8.94	8.88	6.43	6.27	6.23	6.22	6.18	5.26	5.12	5.03	5.01	4.97

Source: Own calculations based on Superintendence of Pension Funds Quarterly Bulletins

Table 1: The management fee on each first contribution

Secondly, OPFs are also allowed to charge a management fee of approximately 0.02% to 0.05% of the accumulated capital, monthly at the end of each month. The size of the fee is inversely related to the size of fund, which follows from an amendment of 2004. Previously it used to be a constant percentage of 0.05% monthly regardless of the number of participants. The details of the current regulation are outlined in table 2.

Size (in bln PLN)	0-8000	8000-20000	20000-35000	35000-65000	65000-...
Max. monthly fee (% of capital)	0.0450	0.040	0.032	0.023	0.015

Source: Superintendence of Pension Funds

Table 2: Structure of monthly management fees depending on OPFs size

While this new regulation allows the participants to internalise obvious returns to scale previously enjoyed exclusively by the fund, one can find little support to justifying the very presence of this fee. The efficiency argument is often raised here, namely that funds will thus have an incentive to actively manage the participants' savings invest as they will have their share in the high returns earned for the members. However, funds are allowed to charge this fee irrespectively of their performance, whilst regardless of their performance with the accumulation of members and their contributions this fee is strictly increasing in time. Furthermore, the criteria according to which the respective levels were specified are not evident.

4.2. OPFs versus a bank account

In this section we intend to asses the efficiency of OPFs as capital accumulators vis-à-vis a standard financial institution that is a bank. We compare the observed risk-less interest rates with OPFs results and — in a very simplified way — estimate the differential necessary ensure OPFs outperform the bank account strategy.

Consider a very basic financing model to compare two different types of medium and long run investment strategies (OPF versus bank strategy). Assume:

- K_i : monthly accumulated amount transfer to OFE or an alternative system;
- α_i : part of K_i OPFs management fee, which they are allowed to deduce from the contributions;
- $r_{\text{OFE}} = \sum_i \text{weight}_i \cdot \dfrac{K_{1,i} - K_{0,i}}{K_{0,i}}$: average weighted OPFs rates of return (weights based on the market shares);

  K_0 initial capital at time 0, K_1 capital after a specified period (usually monthly);
- r_{free} : risk free interest rate identified with the return from annual T-bills ($r_{\text{free,T}}$) or commercial bank deposits ($r_{\text{free,D}}$), depending on the scenario analysed.

We consider future values (FV) of a 480 monthly investment strategy: from 25 till 65 years of age. Since any assumptions on the interest rates patterns over such a long time span is a disputable issue, we avoid this problem by comparing the cash flows directly at time t.

We further assume that such a long run investment strategy is essentially costless under the bank strategy. This assumption has two grounds. First, most citizens already have an independent bank account. Secondly, there are many bank alternatives where opening and maintaining an account actually is costless (e.g. internet banking). We assume that the return in case of such a strategy is determined by the interest on treasury bills (the proxy for risk-less interest rate).

Based on the above specification, the cash-flows per a person are given by:

$$FV(r_{\text{free}}) = \sum_{i=1}^{480} K_i \cdot (1 + r_{\text{free}})^{480-i} - \quad \text{future value of benefit}$$

$$FV(r_{\text{OFE}}; \alpha) = \sum_{i=1}^{480} K_i \cdot (1 - \alpha_i)^{480-i} \cdot (1 + r_{\text{OFE}})^{480-i} - \quad \text{future value of benefit}$$

Comparing these two amounts allows to formulate the following hypothesis:

$$FV(r_{\text{OFE}}; \alpha) - FV(r_{\text{free}}) \geq 0.$$

It is thus sufficient to compare $(1-\alpha_i)^{480-i} \cdot (1+r_{\text{OFE}})^{480-i}$ and $(1+r_{\text{free}})^{480-i}$ under the assumption of strict stochastic domination. Taking appropriate roots we obtain a system of equations for each i:

$$r_{\text{OFE}} \geq \frac{(1 + r_{\text{free}})}{(1 - \alpha_i)} - 1, \quad \text{where} \quad 1 - \alpha \in (0, 1).$$

Of course above inequality is a stochastic one and we do not know the processes followed by r_i^{OFE},r_i^{free}, neither do we know K_i. What we have at our disposal is the complete information about the realisation of these processes up to know. We know r_i^{OFE} and r_i^{free} in the period 1999 to 2004 as well as the pattern of α_i evolution (constructed from tables 1 and 2). Based on this data set and an additional assumption of scaling K_i to 1 we performed two simulations.

- **Simulation I**

  We compute the gap between the accumulated returns to two types of investments: OPFs and our artificial one controlling for the costs associated to OPFs. The artificial benchmark is constructed basing on the average long term deposit rate in the commercial banks as reported by the Central Bank (monthly data on three year or more deposits, annual rates). From these time series we conclude that the artificial investment strategy was working better than OPFs. This outcome can be attributed to both the structure of OPFs rates of return and significant costs (α_i factor in our model).

  We then ask the data by how much on average must OPFs outperform simple bank deposits over the next thirty five years to ensure that the new system is efficient from the value-for-money point of view. For the reasons specified above it seems important to analyse the accumulated increase in value and not only contemporaneous rates of return. We compare the observed risk-less interest rates with OFEs results and — in a very simplified way — estimate the differential necessary ensure OFEs outperform the bank account strategy. The following formula is applied:

$$\sum_{i=1}^{63}\left[(1-\alpha)\prod_{m=1}^{i}[1+r_m^{\mathrm{OFE}}]-\prod_{m=1}^{i}[1+r_m^{\mathrm{free}}]\right]-\sum_{j=64}^{416}\left[\frac{(1-\alpha)}{[1+r^{\mathrm{OFE}}]^j}-\frac{1}{[1+r^{\mathrm{free}}]^j}\right]>0,$$

  where r_m^i denotes observed rates of return, while estimates of the expression in second brackets on the RHS of the above equation is only considered in a form of a differential. As demonstrated above, the first term at the left-hand side is clearly negative. These simple calculations demonstrate that OPFs would need to maintain on average returns 0.4 - 0.5 percentage points higher than the commercial banks deposit rates. The results are presented below.

  The results demonstrate, that values for matter significantly in comparing the completely risk-free investment strategy and participating in the pension fund. It is only natural to require OPFs to realise the effective returns above the bank deposit. With the interest rates decreasing in Poland on its accession to European Monetary Union, OPFs are rather likely to achieve the rates of return required in this simulation. However, one should be only moderately optimistic about this result.

- **Simulation II**

  Open Pension Funds are believed to play the role of crucial institutional investors on the Polish capital market. If so, the secure investment rather than secure deposit should be treated as a benchmark. Thus we repeated the above procedure for the most profitable low risk investment opportunity over the relevant time span. Public offerings of the government bonds in the period 1999-2004 were believed to be the most attractive from the risk-return trade-off point of view. We used the annual yields of these bonds, as indicated at the day of offering as a reference level (regardless of maturity, dates of public offerings and OPFs unit values accorded). This is equivalent to a simple buy & sell strategy.

Amount invested monthly	1 PLN	
OFE cost structure	As in table 1 plus average of table 2	
Bank deposit cost structure	None	
No. of monthly observations of observed returns	64	
No. of monthly contributions till retirement	480 (40 years)	
Earned in OFE	76.9 PLN	
Earned on	Commercial bank deposit	T-bills trading
	77.08 PLN	87,9 PLN
Differential for $r_t^{\text{free}} = 0.00$	0.48%	0.515%
Differential for $r_t^{\text{free}} = 0.04$	0.41%	0.469%
Differential for $r_t^{\text{free}} = 0.08$	0.37%	0.457%

Source: Own calculations. OPFs data from Superintendence of Pension Funds, commercial bank data from Polish Central Bank.

For the derivations we assumed a constant differential between the two considered rates of return

Table 3: Simulation – OFE versus a bank deposit *a risk-free investment strategy*

Clearly, the gap between the effective OPFs returns and the government bond yields is much higher than in the previous case. Furthermore, although the budged needs will eventually decrease over time, the effect of the current public offerings will last for even thirty years in some cases. Thus, although government and communal bonds constitute the vast majority of OPFs assets, funds should not be perceived as active investors on this market due to the high returns differential.

Summarising, treating bank as a benchmark for analysing the efficiency of the OPFs is a rather perfidious example. We did that only in order to show that effectively OPFs rates of return are beaten even by the essentially risk free rates. Relatively low OPFs rates in the beginning make it difficult to obtain relatively high benefits in the end due to accumulation process, although the observed rates of return at the end might suggest that OPFs strategy outperforms the risk-free one. There is also an additional effect of the timing of joining a fund. Most of the potential beneficiaries belong to a fund starting 1999 due to a relatively short decision time (only one year). Currently, only new generations may add to the status, which is very little when compared to the initial situation[6].

[6]However, there are sound arguments against leaving retirement savings in the banks domain. Primarily, banks bankruptcy may have many sources and thus providing full guarantees might be much less attractive from the government and the society point of view. Also supervision would be significantly more difficult, as standards for pension

5. CONCLUSIONS

There can be no doubt that the PAYG system could not be maintained — it was inefficient and failed to respond to demographic and economic changes in the Polish society. As a reform a three pillar system was designed in which only the first pillar remains of pay-as-you-go nature, whereas the other two have the contribution account character. The obligatory participation in the second pillar had two major aims: increasing national savings and stimulating the formation of saving habits.

In the paper we attempted to demonstrate that the current solution is no different in its nature from the regular bank account, as second pillar OPFs are merely an obligatory money-box. We further verified that this solution is inefficient in a sense that it is more expensive than a bank account solution used as a benchmark case. The data show that OPFs invest most of their funds in government bonds which certainly does not stimulate the growth of the Polish economy. In addition, there is no mechanism that could alleviate the shortcomings of the OPFs at the moment of retirement, as the OPAs are likely to be as inefficient.

In the case of each welfare state — socialist or Western European — citizens are used to rest the burden of future on the state's shoulders. Investors refuse to consider long-term opportunities, whereas consumers are characterised by myopia and excessive claims. The process of shaping proper saving behaviour patterns is a long one and requires a lot of time to be completed. However, the current Polish pension fund system is not necessarily moving in the right direction.

References

Act of 28.08.1997. About the organisation and functioning of the pension funds. (Dz. U. nr97.139.934), 1997.

P. Bernstein. *Against the gods. The remarkable story of risk.* John Wiley & Sons, 1996.

Central Statistical Office. Annual bulletins. (1998-2002), 1998.

W. van Giunrken. Social security for the excluded majority. Case studies of developing countries, Geneva, 1999.

S. Gomulka and M. Styczen. Estimating the impact of the 1999 Pension Reform in Poland, 2000-2050, Case CEU Working papers, 27, 1999.

D.B. Hadyniak and J. Monkiewicz. Fundusze Emerytalne II Filar, Warsawa, 1999.

K. Jajuga and T. Jajuga. Inwestycje - Instrumenty finansowe, ryzyko finansowe, inzynieria finansowa, PWN, Warsawa, 1999.

A. Jurkowski. Bankowosc detaliczna, NBP, Departament Analiz i Badan, Warsawa, 2001.

fund management are much easier to set and to execute (system is more transparent). The main argument on our side follows from the observation that from the user point of view the difference between the OPF and the bank account is not that evident, especially considering that the banking sector is also fully guaranteed by the state.

R. Kaas, M. Goovaerts, J. Dhaene, and M. Denuit. *Modern Actuarial Risk Theory*. Kluwer Academic Publishers, 2001.

B. Liberda. Household Saving in Poland, CASE Studies and Analyses, 187, Warsawa, 1999.

B. Liberda. Oszczednosci w gospodarce, PTE-Bellona, Warsawa, 2003.

B. Liberda and T. Tokarski. Determinants of Saving and Economic Growth in Poland in comparison to the OECD countries, CASE CEU Working papers, 24, Warsawa, 1999.

Money.pl. website. http://www.money.pl.

P. Mularczyk. Actuarial Modelling of the new three-pillar pension system in Poland, Warsaw university, Faculty of Economics, mimeo, 2002.

P. Mularczyk and J. Tyrowicz. A pension plan or a money box? the reform of the social security system in poland. In *3rd Meeting on Social Security and Complementary Pension Systems, CIEF, IDEFE, ISEG*. Technical University of Lisbon, 2004.

K. Nowak. Nowe emerytury, Poznan, 1999.

OECD. Electronic finance: Economics and institutional factors, 2001.

H.H. Panjer. *Financial Economics – with Applications to Investments, Insurance and Pensions*. The Actuarial Foundation, 1998.

O. Schneider. Implicit public debt of the Czech social security system, CASE Studies and Analyses, 167, Warsawa, 1999.

Superintendence of the Banking sector. Annual bulletins and quarterly bulletins (1999-2003).

Superintendence of the Pension Funds. Annual bulletins and quarterly bulletins (1999-2003). http://www.knuife.gov.pl.

PARAMETER ESTIMATION FOR AN IG-OU STOCHASTIC VOLATILITY MODEL

Luis Valdivieso

Universitair Centrum voor Statistiek, K.U.Leuven, W. de Croylaan 54, 3001 Heverlee, Belgium
Email: `lvaldiv@pucp.edu.pe`

Abstract

Given a self-decomposable distribution D, a Non-Gaussian stationary process of Ornstein-Uhlenbeck type can be constructed in such a way that their marginals coincide with D. This is called a D-OU process. Bandorff-Nielsen and Shepard have recently extended the classical Black-Scholes model for stock prices by allowing the volatility parameter to be either a D-OU process or a superposition of these. In this article we propose a maximum likelihood methodology to estimate the parameters for a general D-OU process. In particular we apply these results when D is an inverse Gaussian distribution. Within this framework we present a simulation study, which includes the case of superposition.

1. INTRODUCTION

Financial time series of different assets are widely documented as sharing a set of well established stylized features. For example, log-returns are known to display heavy tailed distributions, aggregational gaussianity, quasi long-range dependence and correlate negatively with their stochastic volatilities. These empirical factors are successfully captured by a new class of models in which the stochastic variance of log-returns is constructed via a one or more mean reverting, stationary processes of Ornstein-Uhlenbeck type driven by subordinators. The latter stochastic volatility models have been introduced and studied by Bandorff-Nielsen and Shepard (2001) and are further termed BN-S models.

In their seminal work, Bandorff-Nielsen and Shepard document the impossibility to obtain direct maximum likelihood estimators for the parameters indexing the stochastic volatility model and they review a series of different alternatives to deal with this problem.

The main contribution of this paper lies in the presentation of a new methodology to obtain the maximum likelihood estimates for a general D-OU process. In particular, we could apply these results to estimate the volatility parameters of the BN-S model under an IG-OU volatility model. The IG-OU specification is of particular interest since in the simplest BN-S model, log returns are approximately distributed according to a normal inverse Gaussian law. This distribution has been shown to provide very good fit to log-returns of stock prices (see for instance Prause (1999) and Raible (2000)).

In order to asses the goodness of our approach, we performed a simulation study. In this, we generated some non-superposed and superposed IG-OU processes and then we compared our estimates with the underlying parameters of these processes.

2. THE BN-S MODEL

In the Black-Scholes world the logarithm of a stock-price process , $\mathbf{X} = \{X(t) = Ln(S(t))\}$, satisfies the stochastic differential equation

$$dX(t) = (\mu - \frac{1}{2}\sigma^2)dt + \sigma dW(t) ,$$

where μ and $\sigma > 0$ are fixed parameters and $\mathbf{W} = \{W(t)\}$ is a standard Brownian Motion. As is widely documented, this model presents a series of empirical limitations. First, log-returns do not seem to behave according to a normal distribution and second, the estimated volatilities change stochastically over time. More precisely, these volatilities tend to be clustered and their up jumps are usually associated to down jumps in the stock price. This last feature is known as the leverage effect. To circumvent these inadequacies Bandorff-Nielsen and Shepard (2001) proposed to extend the Black-Scholes model by making $\sigma^2 = \{\sigma^2(t)\}$ stochastic and by including a leverage term in the log-price process. Their model states that

$$dX(t) = (\mu - \frac{1}{2}\sigma^2(t))dt + \sigma(t)dW(t) + \rho dZ(\lambda t)$$

and

$$d\sigma^2(t) = -\lambda\sigma^2(t)dt + dZ(\lambda t) , \tag{1}$$

where $\lambda > 0$, $\rho < 0$ is a parameter which accounts for the leverage effect, and $\mathbf{Z} = \{Z(t)\}$ is a zero-drift Lévy process with non-negative increments. In this setting, the Brownian motion process $\mathbf{W}$ and the Lévy process $\mathbf{Z}$ are independent and the filtration $\mathbf{F}$ is taken to be the usual augmentation of the filtration generated by the pair $(\mathbf{W}, \mathbf{Z})$. In both processes $X(0)$ and $\sigma^2(0)$ are assumed to be independent random variables.

A further extension in the BN-S model is given by considering in (1) not only a single volatility process, but also a convex linear combination (superposition) of these processes.

As with other complex models, the BN-S model is arbitrage free but incomplete, which means that there is more than one equivalent martingale measure (EMM). This problem is studied by Nicolato and Venardos (2003), who pay particular attention to the class $\mathcal{M}'$ of structure-preserving EMM and in special to the subclass $\mathcal{M}^{IG} \subset \mathcal{M}'$ of EMM preserving (1) when $\sigma^2 = \{\sigma^2(t)\}$ is a stationary process with inverse Gaussian law, or briefly, an IG-OU volatility process.

Using standard vanilla call option prices on the S&P 500 index and the characteristic-based pricing formula developed by Carr and Madan (1998), Schoutens (2003) calibrates the parameters of the BN-S model under an IG-OU volatility process. He shows that this model largely outperforms the one obtained under the Black-Scholes model and other plausible Lévy models with constant volatility.

3. THE IG-OU PROCESS

Let $\mathbf{Z} = \{Z(t)\}$ be an univariate Lévy process with generating triplet $(\sigma_0, \gamma_0, \upsilon_0)$ and let $\lambda > 0$. $\mathbf{X} = \{X(t)\}$ is called a process of Ornstein-Uhlenbeck type generated by $(\sigma_0, \gamma_0, \upsilon_0, \lambda)$ if it is càdlàg and satisfies the stochastic differential equation

$$\begin{cases} dX(t) = -\lambda X(t)dt + dZ(\lambda t) \\ X(0) = X_0 \ , \end{cases}$$

being X_0 a random variable independent of $\mathbf{Z}$. Under this setting $\mathbf{Z}$ is termed the background driving Lévy process (BDLP).

Using the theory of integration of left-continuous predictable process with respect to semimartingales and in particular to Lévy processes (see Protter (1990)), we obtain

$$X(t) = X_0 + Z(\lambda t) - \lambda \int_0^t X(s)ds,$$

or recursively

$$X(t+\Delta) = e^{-\lambda\Delta}\left(X(t) + e^{-\lambda t}\int_t^{t+\Delta} e^{\lambda s} dZ(\lambda s)\right). \tag{2}$$

As a result $\mathbf{X}$ is a Markov process. Furthermore, the following property is obtained.

Proposition 3.1 *For any* $t, \Delta > 0$:

$$e^{-\lambda t}\int_t^{t+\Delta} e^{\lambda s} dZ(\lambda s) \overset{d}{=} \int_0^{\Delta} e^{\lambda s} dZ(\lambda s).$$

Proof. It follows from the integration by parts rule for semimartingales (see Protter (1990)). $\square$

The random variable $Z^*(\Delta) = \int_0^{\Delta} e^{\lambda s} dZ(\lambda s) \overset{d}{=} \int_0^{\lambda\Delta} Z(s)ds$ will play a central role in our analysis.

Definition 3.1 *A random variable X has a self-decomposable distribution if for any $0 < a < 1$, it is possible to find a random variable Y_a, independent of X, such that*

$$X \overset{d}{=} aX + Y_a.$$

The following result, due to Sato (1999), explores the stationarity of a process of Ornstein-Uhlenbeck type.

Proposition 3.2 *If $\mathbf{X}$ is a process of Ornstein-Uhlenbeck type generated by $(\sigma_0, \gamma_0, \upsilon_0, \lambda)$ such that*

$$\int_{|x|>2} Ln(|x|) d\upsilon_0(x) < \infty, \tag{3}$$

then $\mathbf{X}$ has a unique self-decomposable stationary distribution μ.

Conversely, for any $\lambda > 0$ and any self-decomposable distribution D, there exists a unique triplet $(\sigma_0, \gamma_0, \upsilon_0)$ satisfying (3) *and a process of Ornstein-Uhlenbeck type $\mathbf{X}$ generated by $(\sigma_0, \gamma_0, \upsilon_0, \lambda)$ such that D is the stationary distribution of $\mathbf{X}$.*

The stationary process $\mathbf{X}$ in the converse result of proposition 3.2 is called a D-OU process. As mentioned in section 2, this was the process describing the stochastic volatility behaviour in the BN-S model.

A convenient way to analyze the relation between a D-OU process and its corresponding BDLP, **Z**, is through their cumulant functions:

$$C_D(\vartheta) = LnE[e^{i\vartheta D}] \quad \text{and} \quad C_{Z(1)}(\vartheta) = LnE[e^{i\vartheta Z(1)}].$$

As proved in Bandorff-Nielsen (1988), these satisfy the following two properties.

Proposition 3.3 *For any $\vartheta \in \mathbb{R}$:*

$$C_{Z(1)}(\vartheta) = \vartheta \frac{dC_D(\vartheta)}{d\vartheta}.$$

Proposition 3.4 *For any $\Delta > 0$ and $\vartheta \in \mathbb{R}$:*

$$E[e^{i\vartheta Z^*(\Delta)}] = e^{\lambda \int_0^\Delta C_{Z(1)}(\vartheta e^{\lambda s})ds}.$$

The dependence structure in a D-OU process is characterized as follows.

Proposition 3.5 *For any $t \geq 0$ and $\Delta \in \mathbb{R}$:*

$$\rho(\Delta) = \frac{Cov(X(t), X(t+\Delta))}{\sqrt{V(X(t))V(X(t+\Delta))}} = e^{-\lambda|\Delta|},$$

Proof. It follows easily from the stationary property and (2). □

We close this section by introducing the IG-OU process.

Definition 3.2 *A random variable X has an inverse Gaussian distribution with parameters $a > 0$ and $b > 0$, or briefly $X \sim IG(a, b)$, if it has a density*

$$f(x) = \frac{ae^{ab}}{\sqrt{2\pi x^3}} e^{-\frac{1}{2}(\frac{a^2}{x} + b^2 x)}, \; \forall x > 0.$$

The characteristic function of $X \sim \mathrm{IG}(a, b)$ is proved to be

$$\hat{P}_X(\vartheta) = e^{a(b - \sqrt{b^2 - 2i\vartheta})},$$

which means that this is an infinitely divisible distribution and we can define an inverse Gaussian Lévy process. Furthermore, Halgreen (1979) shows that $X \sim \mathrm{IG}(a, b)$ is self-decomposable. Then proposition 3.2 can be invoked to justify the existence of an $\mathrm{IG}(a, b)$-OU process.

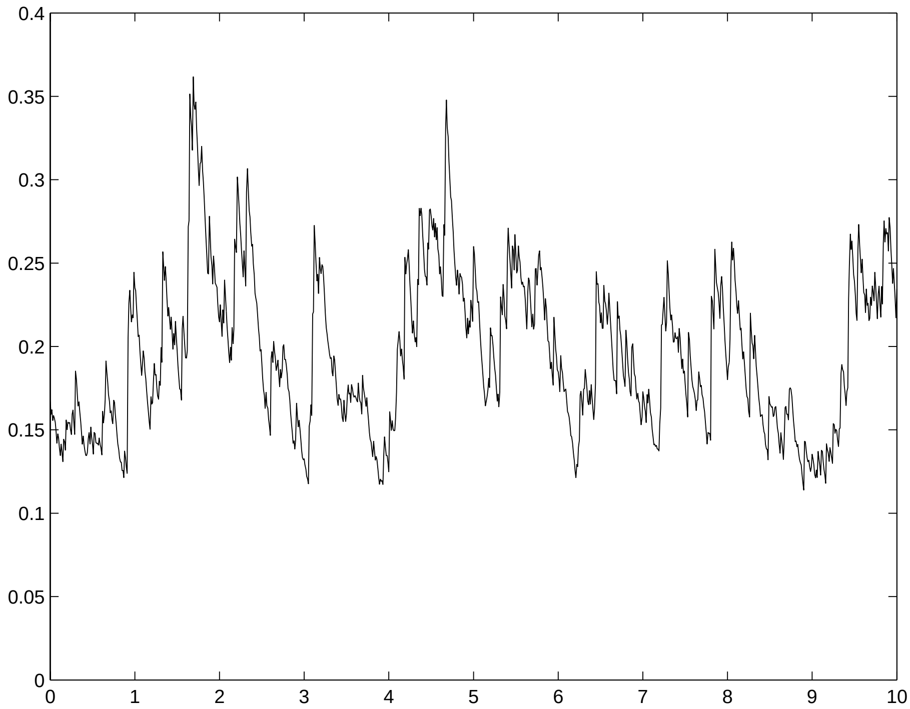


Figure 1: Simulation of an IG$(2, 10)$-OU sample path with $\lambda = 5$ and 1001 points.

4. LIKELIHOOD INFERENCE FOR A D-OU PROCESS

Suppose that $\mathbf{X} = \{X(t)\}$ is a D-OU process, where D depends on an unknown parameter $\theta \in \mathbb{R}^m$, and we are interested in estimating θ based on a set of $n+1$ observations $x_0, x_1, \ldots, x_n$ from the sample $X(0), X(\Delta), X(2\Delta), \ldots, X(n\Delta)$ of $\mathbf{X}$. By the Markov property, the likelihood function of this sample is

$$L(\theta) = f_{X(0)}(x_0) \prod_{k=1}^{n} f_{X(k\Delta)|X((k-1)\Delta)=x_{k-1}}(x_k). \tag{4}$$

According to (2) and proposition 3.1

$$X(k\Delta) \stackrel{d}{=} e^{-\lambda\Delta}(X((k-1)\Delta) + Z^*(\Delta)), \quad \forall k = 1, 2, \ldots, n. \tag{5}$$

This equality leads to the following simplification

$$L(\theta) = f_{X(0)}(x_0) e^{n\lambda\Delta} \prod_{k=1}^{n} f_{Z^*(\Delta)}(e^{\lambda\Delta} x_k - x_{k-1}). \tag{6}$$

Hence, the likelihood function depends only on D, via $f_{X(0)}$, and on the $Z^*(\Delta)$ density.

Briefly, our estimation methodology will be based on the evaluation of (6) via a fast Fourier transform. To be more precise, let us assume that D is an inverse Gaussian distribution with parameters $a > 0$ and $b > 0$. Then, by proposition 3.3, the cumulant function of the corresponding BDLP becomes

$$C_{Z(1)}(\vartheta) = \frac{ai\vartheta}{\sqrt{b^2 - 2i\vartheta}}.$$

Consequently, by proposition 3.4

$$\begin{aligned}\hat{P}_{Z^*(\Delta)}(\vartheta) &= e^{\lambda \int_0^\Delta C_{Z(1)}(\vartheta e^{\lambda s})ds} \\ &= e^{\lambda \int_0^\Delta \frac{ai\vartheta e^{\lambda s}}{\sqrt{b^2-2i\vartheta e^{\lambda s}}}ds} \\ &= e^{a \int_\vartheta^{\vartheta e^{\lambda\Delta}} \frac{i}{\sqrt{b^2-2iy}}dy}\end{aligned}$$

and

$$\hat{P}_{Z^*(\Delta)}(\vartheta) = e^{a(\sqrt{b^2-2i\vartheta}-\sqrt{b^2-2i\vartheta e^{\lambda\Delta}})}. \tag{7}$$

This explicit representation motivates the following scheme to estimate $\theta = (a, b)$, when $\lambda > 0$ is given.

- Find initial estimates of a and b:

  Using (5), we propose to consider the n independent and identically distributed random variables

  $$Y_k = \int_{\lambda(k-1)\Delta}^{\lambda k\Delta} e^s dZ(s) = e^{\lambda\Delta} X(k\Delta) - X((k-1)\Delta), \text{ with } k = 1, 2, \ldots, n,$$

  and match the expected value and variance of these with the statistics

  $$\bar{Y} = \frac{1}{n}\sum_{k=1}^{n} Y_k \text{ and } S_Y^2 = \frac{1}{n}\sum_{k=1}^{n}(Y_k - \bar{Y})^2.$$

  This procedure generates the initial estimators

  $$\hat{a}_0 = \frac{\hat{b}_0\bar{Y}}{(e^{\lambda\Delta}-1)} \quad \text{and} \quad \hat{b}_0 = \frac{1}{S_Y}\sqrt{\frac{\bar{Y}(e^{2\lambda\Delta}-1)}{e^{\lambda\Delta}-1}}.$$

- Use (7) and the fast Fourier transform to evaluate $f_{Z^*(\Delta)}$.

- Use numerical methods to optimize (6).

If λ is not given, proposition 3.5 suggests it be estimated by solving:

$$\hat{\lambda}_0 = \arg\min_{\lambda} \sum_{k=1}^{n-2} (\hat{acf}(k) - e^{-\lambda k\Delta})^2,$$

where $\hat{acf}(k)$ denotes the empirical autocorrelation function of lag k based on the data $x_0, x_1, \ldots, x_n$. Once $\hat{\lambda}_0$ is obtained, we can opt for two methods. These will be called the $\pm\lambda$ and $-\lambda$ method . In the first we substitute λ with $\hat{\lambda}_0$ in the scheme above, while in the second we take $(\hat{a}_0, \hat{b}_0, \hat{\lambda}_0)$ as an initial estimator in the search for $\theta = (a, b, \lambda)$ that maximizes (6).

5. THE SUPERPOSITION CASE

Let $\mathbf{X_1} = \{X_1(t)\}$ be a D$_1$-OU process and $\mathbf{X_2} = \{X_2(t)\}$ be an independent D$_2$-OU process, satisfying the stochastic differential equations

$$dX_1(t) \ = -\lambda_1 X_1(t)dt + dZ_1(\lambda_1 t)$$

and

$$dX_2(t) = \ -\lambda_2 X_2(t)dt + dZ_2(\lambda_2 t).$$

In this section, we will be interested in analyzing the estimation problem for a superposed model $\mathbf{X} = \{X(t)\}$ defined by

$$X(t) = \omega X_1(t) + (1-\omega)X_2(t), \ \ \text{where} \ \ \omega \in [0,1].$$

In comparison to a D-OU process, this model presents a more interesting correlation pattern.

Proposition 5.1 *The autocorrelation function of* $\mathbf{X}$ *has, for some* $c \geq 0$*, the form:*

$$\rho(\Delta) = ce^{-\lambda_1|\Delta|} + (1-c)e^{-\lambda_2|\Delta|}, \ \ \forall \Delta \in \mathbb{R}.$$

Proof. It is a straightforward consequence of proposition 3.5. $\square$

In terms of applications, quasi-long-range dependence processes are better described by these superposed models.

In order to study the inference problem, let us assume that we have a set of $n+1$ observations $x_0, x_1, \ldots, x_n$ from the sample $X(0), X(\Delta), X(2\Delta), \ldots, X(n\Delta)$ of $\mathbf{X}$. Although in this case (4) can not be reduced to (6), we can still make use of the general estimation scheme proposed in section 4. To start, (5) now takes the form

$$X(k\Delta) = \omega e^{-\lambda_1 \Delta} X_1((k-1)\Delta) + (1-\omega)e^{-\lambda_2\Delta} X_2((k-1)\Delta) + Z^*(\Delta),$$

where

$$Z^*(\Delta) = \omega e^{-\lambda_1\Delta} \int_0^\Delta e^{\lambda_1 s} dZ_1(\lambda_1 s) + (1-\omega)e^{-\lambda_2\Delta} \int_0^\Delta e^{\lambda_2 s} dZ_2(\lambda_2 s).$$

Since we do not have observations from $\mathbf{X_1}$ and $\mathbf{X_2}$, we could condition on $\mathbf{X_1}$ to obtain

$$P(X(k\Delta) \leq x | X((k-1)\Delta) = x_{k-1})$$

$$= \int_0^\infty P(Z^*(\Delta) \leq x - e^{-\lambda_2\Delta} x_{k-1} - \xi\omega(e^{-\lambda_1\Delta} - e^{-\lambda_2\Delta})) \frac{f(x_{k-1}, \xi)}{f_{X((k-1)\Delta)}(x_{k-1})} d\xi,$$

where $f(.,.)$ denotes the conjoint density function of $\mathbf{X}$ and $\mathbf{X_1}$ at time $(k-1)\Delta$. Moreover, this density can be explicitly obtained. For example, if $\mathbf{X_1}$ is an IG(a_1, b)-OU process and $\mathbf{X_2}$ an independent IG(a_2, b)-OU process, then

$$f(y_1, y_2) = \begin{cases} \frac{a_1 a_2 \sqrt{1-\omega}}{2\pi(y_2(y_1-\omega y_2))^{\frac{3}{2}}} \ e^{-\frac{1}{2}\left(\frac{a_1^2}{y_2} + \frac{a_2^2(1-\omega)}{y_1-\omega y_2} + b^2(y_2 + \frac{y_1-\omega y_2}{1-\omega}) - 2b(a_1+a_2)\right)}, & \forall \ 0 < y_2 < \frac{y_1}{\omega} \\ 0, & \text{otherwise}. \end{cases}$$

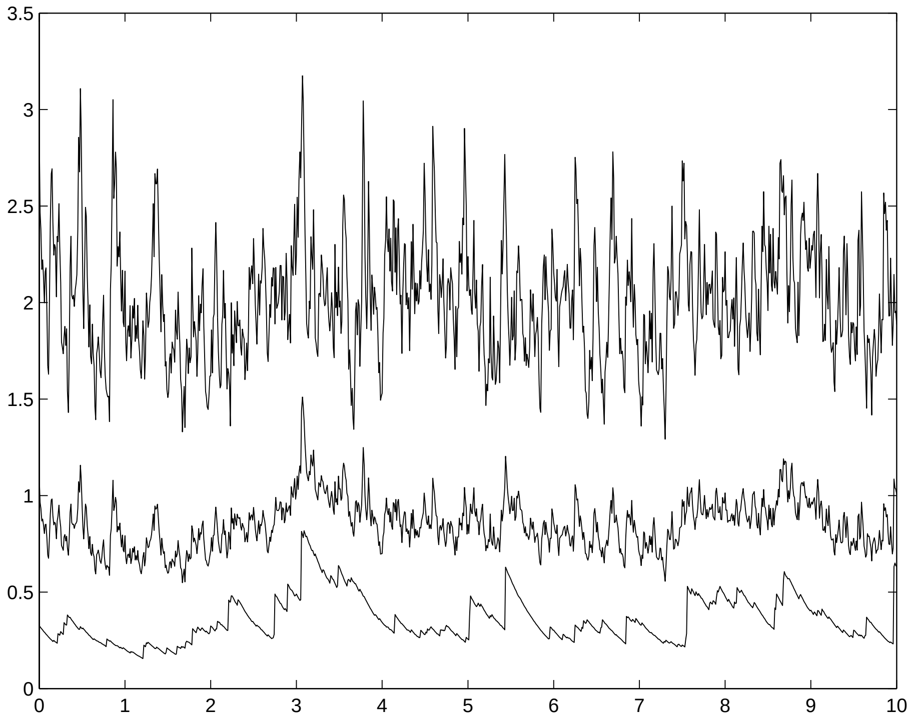


Figure 2: From bottom to top, simulated sample paths of an IG$(2,5)$-OU process $\mathbf{X_1}$ with $\lambda_1 = 2$, a superposed process $\mathbf{X} = 0.7\mathbf{X_1} + 0.3\mathbf{X_2}$, and an IG$(10,5)$-OU process $\mathbf{X_2}$ with $\lambda_2 = 30$.

Summarizing, the conditional density $f_{X(k\Delta)|X((k-1)\Delta)=x_{k-1}}(x_k)$ in (4) takes the form:

$$\frac{1}{f_{X((k-1)\Delta)}(x_{k-1})}\int_0^{\frac{x_{k-1}}{\omega}} f_{Z^*(\Delta)}(x_k - e^{-\lambda_2\Delta}x_{k-1} - \omega\xi(e^{-\lambda_1\Delta} - e^{-\lambda_2\Delta}))f(x_{k-1},\xi)d\xi. \quad (8)$$

Note that this integral can be approximated, for instance with a Simpson's rule, and so (8) will depend basically on the $Z^*(\Delta)$ and $X((k-1)\Delta)$ densities. Consequently, (4) could be evaluated, if we knew the characteristic functions of $Z^*(\Delta)$ and $X((k-1)\Delta)$.

Let us consider for example, that $\mathbf{X_1}$ is an IG(a_1,b)-OU process and $\mathbf{X_2}$ is an independent IG(a_2,b)-OU process. Then, by definition 3.2 and (7)

$$\hat{P}_{Z^*(\Delta)}(\vartheta) = e^{a_1(\sqrt{b^2-2i\vartheta\omega e^{-\lambda_1\Delta}}-\sqrt{b^2-2i\vartheta\omega})+a_2(\sqrt{b^2-2i\vartheta(1-\omega)e^{-\lambda_2\Delta}}-\sqrt{b^2-2i\vartheta(1-\omega)})}$$

and

$$\hat{P}_{X((k-1)\Delta)}(\vartheta) = e^{a_1(b-\sqrt{b^2-2i\vartheta\omega})+a_2(b-\sqrt{b^2-2i\vartheta(1-\omega)})}, \ \ \forall k = 1,2,\ldots,n.$$

As a result, approximated likelihood estimates of $\theta = (a_1, a_2, b)$ can be obtained by adapting the estimation scheme in section 4. Table 3 presents some estimation results under this procedure.

6. SOME SIMULATION RESULTS

In order to asses our approach, we performed two simulation studies. In the first, we considered two sample paths of an IG$(2, 10)$-OU process $\mathbf{X} = \{X(t)\}$ with $\lambda = 5$. The paths were simulated by means of the Euler scheme

$$X(0.01k) \approx e^{-5\lambda}(X(0.01(k-1)) + \sum_{k=1}^{100} e^{k\hbar} Z(\hbar)), \quad \forall\, k = 1, 2, \dots, 1000.$$

where $\mathbf{Z} = \{Z(t)\}$ is the corresponding BDLP of $\mathbf{X}$ and $\hbar = 0.0005$. The simulation of $\mathbf{Z}$ benefited from the fact that $\mathbf{Z}$ can be decomposed as a sum of an inverse Gaussian Lévy process and an independent compound Poisson process (see Bandorff-Nielsen (1988) for details). The first of the simulated sample paths is shown in Figure 1.

Once the simulated data was obtained, we proceeded to follow the scheme in section 4. The resulting estimates are shown in Tables 1 and 2. The first rows in each table present the estimates of $a = 2$ and $b = 10$ when λ is given, while the second and third rows present the results under the $\pm\lambda$ and $-\lambda$ methods. All the procedures were implemented in Matlab.

Sample path 1

Method	Initial estimates			Likelihood estimates			-Loglikelihood
$+\lambda$	a	b	λ	a	b	λ	
(2.265)	2.033	10.309	5	1.9915	10.315	5	-3746.3
$\pm\lambda$	a	b	λ	a	b	λ	
(3.75)	1.9082	9.6663	4.3614	1.3618	8.5796	4.3614	-3007.8
$-\lambda$	a	b	λ	a	b	λ	
(32.922)	1.9082	9.6663	4.3614	2.052	10.471	5.0298	-3748.1

Table 1: Estimation for an IG$(2, 10)$-OU process with $\lambda = 5$. The numbers in brackets indicate the computational time in seconds.

Sample path 2

Method	Initial estimates			Likelihood estimates			-Loglikelihood
$+\lambda$	a	b	λ	a	b	λ	
(2.75)	2.0165	10.418	5	1.9064	10.068	5	-3801
$\pm\lambda$	a	b	λ	a	b	λ	
(3.125)	2.2783	11.797	6.5284	2.8614	14.15	6.5284	-3537.2
$-\lambda$	a	b	λ	a	b	λ	
(12.266)	2.2783	11.797	6.5284	1.9274	10.121	5.0101	-3801.3

Table 2: Estimation for an IG$(2, 10)$-OU process with $\lambda = 5$. The numbers in brackets indicate the computational time in seconds.

In the second simulation study, we considered two sample paths of the superposed process $\mathbf{X} = 0.7\mathbf{X_1} + 0.3\mathbf{X_2}$, where $\mathbf{X_1}$ is an IG$(2,5)$-OU process with $\lambda_1 = 2$ and $\mathbf{X_2}$ is an independent IG$(10,5)$-OU process with $\lambda_2 = 30$. In this opportunity, λ_1, λ_2 and ω were given and we aimed to estimate the parameter $\theta = (a_1, a_2, b) = (2, 10, 5)$. The results are presented in Table 3 and the first of the simulated sample paths is shown in Figure 2.

Sample path	a_1	a_2	b
1	2.0074	9.2591	4.8854
2	1.3069	12.334	5.3315

Table 3: Approximated likelihood estimates for the process $\mathbf{X} = 0.7\mathbf{X_1} + 0.3\mathbf{X_2}$, composed by an IG$(2,5)$-OU process $\mathbf{X_1}$ with $\lambda_1 = 2$ and an independent IG$(10,5)$-OU process $\mathbf{X_2}$ with $\lambda_2 = 30$.

References

O.E. Bandorff-Nielsen. Processes of normal inverse Gaussian type. *Finance and Stochastic*, 2: 41–68, 1988.

O.E. Bandorff-Nielsen and N. Shepard. Non-Gaussian Ornstein-Uhlenbeck-based models and some of their uses in financial economics. *Journal of the Royal Statistical Society*, Series B: 167–241, 2001.

P. Carr and D. Madan. Option valuation using the fast Fourier transform. *Journal of Computational Finance*, 2:61–73, 1998.

C. Halgreen. Self-decomposability of the generalized inverse Gaussian and hyperbolic distributions. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete*, 47:13–18, 1979.

E. Nicolato and E. Venardos. Option pricing in stochastic volatility models of Ornstein-Uhlenbeck type. *Mathematical Finance*, 13:445–466, 2003.

K. Prause. The Generalized Hyperbolic model: estimation, financial derivatives, and risk measures. PhD thesis, 1999.

Ph. Protter. *Stochastic Integration and Differential Equations*. Springer, 1990.

S. Raible. Lévy processes in finance: theory, numerics, and empirical facts. PhD thesis, 2000.

K. Sato. *Lévy Processes and Infinitely Divisible Distributions*. Cambridge University Press, Cambridge, 1999.

W. Schoutens. *Lévy Processes in Finance: Pricing Financial Derivatives*. Wiley, Chichester, 2003.

PRICING POWER CONTRACTS BY UTILITY OPTIMIZATION: WORKED EXAMPLES

Michel Verschuere[†] **and Josef Teichmann**[§]

[†]*Quantitative Analyst VERBUND-Austrian Power Trading AG, Am Hof 6a, A-1010 Vienna*
[§]*Technical University of Vienna, E105, Wiedner Hauptstrasse 8-10, A-1040 Vienna, Austria.*
Email: michel@fam.tuwien.ac.at, jteichma@fam.tuwien.ac.at

Abstract

We consider a power market featured of a balancing mechanism, encouraging market participants to accurately match supply with demand for their customers. We illustrate the pricing of typical power supply contracts in a series of simple cases involving only one time step. Our pricing approach relies upon the application of utility optimization principles on a suitable functional involving the hedging strategy in forward contracts.

1. INTRODUCTION

Several power markets around the globe have undergone widespread structural change since the beginning of the nineties, in a process of deregulation. The intention behind deregulation attempts is to open national power markets for competition, where they used to be mainly state controlled. As a consequence of this deregulation, electrical energy is nowadays increasingly traded at exchanges such as NordPool (Oslo, Norway), Powernext (Paris, France), APX (Amsterdam, The Netherlands) and EEX (Leipzig, Germany) to mention just a few examples within the European context.

Together with the creation of transparent markets for electrical energy comes the question of the market based pricing of customer deals. After all, price data recorded at the various exchanges becomes the benchmark — and opportunity cost level — for all market players, as soon as traded volumes express sufficient liquidity. Offering consumer deals below market price represents loss of potential revenues as well as an increased risk exposure. Conversely, market players would price themselves out of the market when being too generous in computing risk premia. Power market participants face prices that have proven to be severely volatile in the past, due to the absence of any physical buffer for the product, as electrical energy is not economically storable once it is generated. As a result, all participants — producers and suppliers — must increasingly seek the careful balance between over- and underestimating cost levels, where their sole guidance is being provided by market prices themselves.

In the present note, we consider a discrete time power market that is featured of forward- and spot markets as well as a balancing mechanism ensuring network stability. Notations have been inherited from Teichmann and Verschuere (2005), where a continuous time version was also formulated. Existing spot markets for electrical power are usually of the day-ahead type, where prices are determined by a supply and demand mechanism for any hourly delivery interval of the next day. Forward markets trade contracts that entail either physical- or financial delivery against future spot prices. Such standardized, exchange traded forward instruments possess primitive load dispatching characteristics, with delivery periods ranging from days, over weeks, months up to entire calender years. As an illustration, German EEX power futures are quoted in Base and Peakload version, where Peak products go online at 1 MW from 08.00h until 20.00h Monday to Friday while the baseload contracts entail the delivery of 1 MW around the clock during the delivery period specified.

A last — but essential — component to any power market is a balancing mechanism, allowing for real time price formation for any electrical energy required for maintaining the balance between infeed and offtake of physical power from the high voltage grid. This service is usually offered by the Transport System Operator (TSO) himself and implies that sufficient reserve capacity is held available at short notice. As demand forecasting in power markets is at least as challenging as weather prediction, such differences between forecasted infeed and offtake and their values at delivery and must be compensated in real-time in order to guarantee reliability of supply. This service represents an intrinsic cost for the market as an entity, but actual charges are somewhat loaded in addition to these core expenses, with the objective to encourage the increased accuracy of load prediction by market participants.

The remainder of this paper is structured as follows: We define a discrete time market model in the next section and illustrate the structure of the UK and Belgian balancing mechanisms. We next formulate two propositions on the existence of optimal forward investment strategies in this environment. Section 3 describes a pricing approach to typical contract types present in power markets based on Utility Maximization. The valuation of primitive versions to these contracts is then described in section 4 and we finish by formulating some conclusions.

2. OPTIMAL FORWARD INVESTMENT IN A DISCRETE SETTING

Take a discrete probability space $(\Omega, \mathcal{F}, P)$ with a filtration $(\mathcal{F}_n)_{n=0,\dots,N}$. We consider time of delivery at time N. We denote the bank account process by $(B_n)_{n=0,\dots,N}$ and prices of forwards on electricity delivered at time $T = N$ by $(F_n)_{0 \le n \le N}$. The strategy of investment is denoted by $(\pi_n)_{n=0,\dots,N-1}$ and we assume B, F and π to be adapted processes. Prices are assumed to be positive adapted processes. The costs which appear up to time $N-1$ are consequently the costs of acquisition $\sum_{i=0}^{N-1} \pi_i F_i$. We assume a utility function, i.e. a strictly increasing, concave C^2-function $u : \mathrm{dom}(u) \to \mathbb{R}$ with usual conditions. Furthermore, we assume a forecasting random variable D_N, which forecasts the amount of MWh at time $T = N$ used by the customers. Note that certainly forecasting gets better as the instant of delivery closes in, but in order to calculate a reasonable price per MWh in advance, one has to work with a long-term forecast. This updating can be included in a re-calculation of the optimization problem for shorter periods or via parameters.

The customers are assumed to pay the price x per MWh. The balancing at time of delivery N constitutes an additional cost C_N: the precise structure of these costs is market specific. We specify in more detail the structure of balancing charges in the Belgian and U.K.-market below. However, the joint feature of any system is that the demand at delivery has to be precisely fulfilled, hence the costs are determined by the difference of the actual demand D_N of the customers and the already bought delivery contracts $\sum_{i=0}^{N-1} \pi_i$. Consequently, we assume a cost function $c : \mathbb{R} \times \mathbb{R}^2 \to \mathbb{R}$ depending on this difference $D_N - \sum_{i=0}^{N-1} \pi_i$ and on additional random variables reflecting the price behaviour between $N-1$ and N.

Example 2.1 *In the Belgian market the function c is convex, increasing in the first variable and depends in the second variable linearly on the day ahead price of the Dutch market one day before, so $C_N := c((D_N - \sum_{i=0}^{N-1} \pi_i), F_{N-1}) = p(D_N - \sum_{i=0}^{N-1} \pi_i)F_{N-1}$.*

Example 2.2 *In the U.K. market we can model the costs C_N at time N via a function c depending on $D_N - \sum_{i=0}^{N-1} \pi_i$ and several prices which appear before delivery, when the actual demand gets clear. The simplest choice would be $p(D_N - \sum_{i=0}^{N-1} \pi_i)F_N$, with p as in the previous example. Note that we are given and applying a spot price F_N here. The choice of p reflects your belief in the liquidity of the market.*

In line with the previous statements, we intend to solve the following optimization problem for F, D_N and x fixed.

$$E_P[u(-\sum_{i=0}^{N-1} \pi_i B_i^{-1} F_i - B_N^{-1} c(D_N - \sum_{i=0}^{N-1} \pi_i, F_{N-1}, F_N) + B_N^{-1} x D_N)] \to \max, \tag{1}$$

where we maximize over all adapted strategies $(\pi_n)_{n=0,\dots,N-1}$. The structure of the cost functional appearing as an argument to the utility $u(\cdot)$ in (1) is clear: It represents the aggregated procurement costs and sales revenues associated to the physical delivery of D_N MWh of power at time $T = N$ and all cash flows have been discounted to time zero using the bond process $(B_n)_{n=0,\dots,N-1}$.

We now formulate two propositions and the reader is referred to Teichmann and Verschuere (2005) for their proof.

Proposition 2.1 *Assume that the previous optimization problem has a solution $(\widehat{\pi_n})_{n=0,\dots,N-1}$, then there exists an equivalent measure Q such that the process*

$$(F_0, B_1^{-1} F_1, \dots, B_{N-1}^{-1} F_{N-1}, B_N^{-1} c'(D_N - \sum_{i=0}^{N-1} \widehat{\pi}_i, F_{N-1}, F_N))$$

is a martingale. In particular the discounted forward price process $(B_n^{-1} F_n)_{n=0,\dots,N-1}$ is a martingale.

Let us now assume the existence of an equivalent martingale measure Q for the process $(B_n^{-1} F_n)_{n=0,\dots,N-1}$ and the existence of an $\mathcal{F}_{N-1}$-measurable random variable Π_{N-1}, such that

$$B_{N-1}^{-1} F_{N-1} = E_Q[B_N^{-1} c'(D_N - \Pi_{N-1}, F_{N-1}, F_N) | \mathcal{F}_{N-1}].$$

Furthermore we assume that $\frac{c(x,\cdot)}{x} \to \infty$ as $x \to \infty$ and $\frac{c(x,\cdot)}{x} \to 0$ as $x \to -\infty$. Under these conditions we can actually prove the existence of a solution of our optimization problem.

Proposition 2.2 *If the previous regularity conditions are satisfied, then there exists an optimizer* $(\widehat{\pi_n})_{n=0,\ldots,N-1}$.

Proposition 2.1 tells us that existence of an optimal strategy in the sense of (1) is equivalent to the existence of an equivalent martingale measure to the discounted extended futures price process. The notion extended here means that the futures process is completed at time T by marginal procurement costs in the balancing regime, i.e. by $c'(D_N - \sum_{i=0}^{N-1} \widehat{\pi}_i, F_{N-1}, F_N)$, where the prime denotes differentiation with respect to the first argument of c.

From proposition 2.2, we then learn that the existence of an optimal trading strategy is guaranteed under quite natural convexity conditions on the mapping $x \to c(x,\cdot)$, provided one can identify an $\mathcal{F}_{N-1}$-measurable random variable Π_{N-1}, which is relatively easy when considering specific example cases.

Remark 2.1 The applied method is the classical method from financial mathematics of optimizing expected utility (see for instance Schachermayer (2004)). Here we obtain in fact an incomplete problem, since we can only trade up to time $N-1$ and we have to face the risk which appears in the last tick. Hence we can hope for pricing mechanisms which are a mixture of actuarial and financial methods of pricing.

3. PRICING SELECTED POWER CONTRACTS

In the present section, we introduce three different customer contracts tailored to the specific situation of a power market. The first contract offers a predetermined load at a future date in exchange for a fixed EUR/MWh price per delivered commodity unit. Although it constitutes a very primitive example, it provides some intuition on how price risk is to be mitigated if considered in the UK environment. The second example concerns the delivery of a flexible load, again at a fixed unit price to be determined. Such contracts are often referred to as balancing deals and we consider their pricing in the Belgian market, where load uncertainty proves the only source of risk to be mitigated. Case three is a mixture of the previous two cases, where we allow for correlations between the stochastic variables price and load.

Our pricing approach relies upon a utility maximization principle similar to the one introduced before. It may be employed for pricing more complicated customer contracts, including power derivatives. We next show how one should proceed in general, but restrict all examples in section 4 to the selection of three basic customer contracts in order to keep calculations explicit and transparent. Our goal here is to get some intuitive understanding on price determinants in power contracts.

We thus retain the previous notation and introduce a function $h : \mathbb{R}^3 \to \mathbb{R}$ such that the image $h(D_N, \sum_{i=0}^{N-1} \pi_i, F_N)$ is the $\mathcal{F}_N$-measurable random variable representing the time $T = N$ payoff of a contingent claim defined in terms of the specified variables. We wish to determine the utility-optimal price for this contract, i.e. we address the optimization problem

$$E_P[u(-\sum_{i=0}^{N-1}\pi_i B_i^{-1}F_i - B_N^{-1}c(D_N - \sum_{i=0}^{N-1}\pi_i, F_{N-1}, F_N) \\ + B_N^{-1}h(D_N, \sum_{i=0}^{N-1}\pi_i, F_N) + \xi)] \to \max, \tag{2}$$

while maximizing over all adapted strategies $(\pi_n)_{n=0,\ldots,N-1}$. The constant ξ represents the optional up-front premium, to be paid at time zero in exchange for the rights on the contingent claim h. Note that the form of the functional in (1) constitutes a special case for a specific supply contract where $h(u, v, w) = x\,u$, where x is the constant customer price in EUR per delivered MWh. Once the optimal trading strategy $(\widehat{\pi_n})_{n=0,\ldots,N-1}$ has been identified the utility indifference price for the contingent claim h is implied from the utility indifference condition

$$E_P[u(-\sum_{i=0}^{N-1}\widehat{\pi}_i B_i^{-1}F_i - B_N^{-1}c(D_N - \sum_{i=0}^{N-1}\widehat{\pi}_i, F_{N-1}, F_N) \\ + B_N^{-1}h(D_N, \sum_{i=0}^{N-1}\widehat{\pi}_i, F_N) + \xi)] = u(0), \tag{3}$$

where ξ is the optional up-front premium, to be paid at time zero. This relation reflects the idea that selling and hedging the claim in the market at the indifference price is equivalent to not doing anything at all in terms of the utility function u.

Most types of power contracts can be moulded in the form of the payoff function h as we illustrate in the next four examples.

1. **Fixed load contracts:** In this case, there is no up-front premium ξ and the payoff h takes the form

$$h(D_N, \sum_{i=0}^{N-1}\pi_i, F_N) = xD_N, \tag{4}$$

and the challenge is to determine the indifference price x, for $D_N = D$ a constant load at time $T = N$.

2. **Variable load contracts:** This contracts offers flexible coverage of the time $T = N$ load D_N without need for any up-front premium ξ and the payoff h takes the form

$$h(D_N, \sum_{i=0}^{N-1}\pi_i, F_N) = xD_N, \tag{5}$$

The issue is to determine the indifference price x, whilst D_N is an $\mathcal{F}_{N-1}$-measurable random variable of finite variance.

3. **CAP option contract:** In this case, a premium ξ must be paid in advance in exchange for the time-$T = N$ payoff h

$$h(D_N, \sum_{i=0}^{N-1} \pi_i, F_N) = \min\{F_N, A\}. \tag{6}$$

The contract thus entails delivery of $D_N = 1$ MWh at time-$T = N$ at a price limited to the upper bound $A \geq 0$. The main task is to determine the premium ξ.

4. **CALL option contract:** In this case, a premium ξ must be paid in advance in exchange for the time-$T = N$ payoff h

$$h(D_N, \sum_{i=0}^{N-1} \pi_i, F_N) = -\max\{F_N - K, 0\} := -(F_N - K)^+. \tag{7}$$

One recognizes the payoff for a shorted CALL option at strike K in h. The contract thus entails delivery of 1 MWh at time-$T = N$ for K EUR/MWh in case the time $T = N$ price exceeds the strike K. In case the option is not exercised, any power procured up to time $N - 1$ will be settled in the balancing regime. The goal is now to determine the premium ξ.

In the next section we address the pricing question for the first two contract types as well as in a mixed case in simple one-step examples. We skip the valuation of CAP and CALL option contracts in order to obtain explicit results in all cases. We thereto note that the utility indifference pricing condition (3) reduces to pure optimal forward investment in case $u(x) = x$, as the optimal strategy is then determined only by the futures price dynamics and the balancing regime; The derivatives' payoff does not intervene.

4. WORKED EXAMPLES

In the single step examples below, we select the "utility function" $u(x) = x$ and balancing penalty function $p(x) = \exp(x) - 1$. Balancing fees in the Belgian and UK environment only differ by the benchmark futures price, respectively F_0 (day-ahead market price in EUR /MWh) and F_1 (the forward price observed in the balancing market).

1. **Fixed load contracts:** Focusing on the one step case in the Belgian market, we have $D_1 = D \geq 0$ and the optimal strategy $\widehat{\pi}_0$ solves

$$E_P[-\pi_0 F_0 - p(D - \pi_0)F_0 + xD] \to \max .$$

We trivially find $\widehat{\pi}_0 = D$.

The UK environment is a bit richer, because the futures price F_1 is essentially random. Take $F_1 = F_0 \exp(\mu + \sigma Z)$, with Z a standard normal random variable, μ a drift and $\sigma > 0$ the forward price volatility. We optimize

$$E_P[-\pi_0 F_0 - p(D - \pi_0)F_1 + xD] \to \max,$$

to find that $\widehat{\pi}_0 = D+\mu+\sigma^2/2$. This means that one procures extra because of the uncertainty in futures prices at delivery in order to stay away from the unattractive balancing regime. The customer price x becomes

$$x = \frac{F_0}{D}(\widehat{\pi}_0 + 1 - \exp(\mu + \sigma^2/2)),$$

which is strictly less than F_0 if $\mu = 0$, indicating that over-procurement of the future load D yields a slight advantage in costs in the presence of the balancing mechanism.

2. **Variable load contracts:** We restrict ourselves to the Belgian market case. The optimization problem reads

$$E_P[-\pi_0 F_0 - p(D - \pi_0)F_0 + xD] \to \max,$$

and the random variable load is chosen as $D_1 = (m + bZ)^2$, with Z again standard normal. After a little calculating, we arrive at the optimal strategy $\widehat{\pi}_0$, namely

$$\widehat{\pi}_0 = \frac{m^2}{1-2b^2} - \frac{1}{2}\log(1-2b^2),$$

which is strictly larger than the average load $E_P[D] = m^2 + b^2$ provided $0 < b < \frac{1}{\sqrt{2}}$. Once more, the presence of a balancing regime implies that over procuring expected load levels is the optimal strategy in avoiding high balancing charges. The customer price x then proves a little higher than the time zero futures price F_0, as

$$x = F_0 \frac{\widehat{\pi}_0}{m^2 + b^2} \geq F_0,$$

and this must be considered a canonical risk premium.

3. **Combined case:** We consider the UK market where futures prices and F_1 and load D_1 are correlated stochastic variables. We choose $F_1 = F_0 \exp(\mu + \sigma Z_1)$ and $D_1 = (m + bZ_2)^2$, whereby $Z_1,\ Z_2$ are bivariate normally distributed with correlation $\rho \geq 0$. We thus have

$$p(z_1, z_2) = \frac{1}{2\pi\sqrt{1-\rho^2}} \exp(-\frac{z_1^2}{2} - \frac{z_2^2}{2} + \rho z_1 z_2),$$

as a bivariate density. The optimal investment strategy $\widehat{\pi}_0$ is then found from

$$\widehat{\pi}_0 = \log(E_P[\exp((m + bZ_2)^2 + \mu + \sigma Z_1)]),$$

which gives

$$\begin{aligned}\widehat{\pi}_0 =& \frac{(2mb + \rho\sigma)^2}{2(1-\rho^2-2b^2)} + \mu + m^2 + \frac{\sigma^2}{2} \\ & - \frac{1}{2}\log(1-\rho^2) - \frac{1}{2}\log(1-\rho^2-2b^2).\end{aligned}$$

Focusing to the case $\mu = 0$ and $b,\ \sigma,\ \rho$ small, we arrive at the second order result for $\widehat{\pi}_0 \sim m^2 + b^2 + \rho^2 + \sigma^2/2$. We recognize our earlier results in this approximate solution if $\rho = 0$ but generally, the optimal strategy also includes loading that is entirely due to the correlation ρ, which is intuitively clear. When price and load correlate, balancing penalties may become even more vulnerable and one procures additional load up front to mitigate this risk.

5. CONCLUSIONS

We have illustrated the utility optimized valuation of power contracts in a set of three one step examples. We derived closed formulas for prices and premia for fixed- and variable load customer deals, as well as the optimal investment in a combined case. The intention was to get intuitive understanding in the price formation for such deals in power markets featured of a balancing mechanism. The reason why we tackled the pricing question here by means of utility considerations instead of risk neutral techniques can be motivated by observing that electrical power is not economically storable. As a result, replicating portfolios can not be constructed using spot power quantities and essential risk remains while hedging contingent claims. Our results are an application of earlier work Teichmann and Verschuere (2005), where one can also find continuous time equivalents to the basic propositions formulated in section 2.

References

W. Schachermayer. Utility maximisation in incomplete markets. In M. Frittelli and W. Runggaldier, editors, *Stochastic Methods in Finance, Lectures given at the CIME-EMS Summer School in Bressanone/Brixen*, volume 1856 of *Springer Lecture Notes in Mathematics*, pages 225–288, 2004.

J. Teichmann and M. Verschuere. Optimal forward investment in power markets. Working paper, 2005.

De Koninklijke Vlaamse Academie van België voor Wetenschappen en Kunsten coördineert jaarlijks tot 25 wetenschappelijke bijeenkomsten, ook contactfora genoemd, in de domeinen van de natuurwetenschappen (inclusief de biomedische wetenschappen), menswetenschappen en kunsten. De contactfora hebben tot doel Vlaamse wetenschappers of kunstenaars te verenigen rond specifieke thema's.

De handelingen van deze contactfora vormen een aparte publicatiereeks van de Academie.

Contactforum "3rd Actuarial and Financial Mathematics Day" (4 februari 2005, Prof. M. Vanmaele)

De "3rd Actuarial and Financial Mathematics Day" kende net als de vorige edities een groot succes. Dankzij dit jaarlijks evenement worden de contacten tussen de verschillende onderzoekers en onderzoeksgroepen van de Vlaamse universiteiten (KULeuven, UA, UGent en VUB) in deze domeinen verder aangehaald. Daarenboven biedt het contactforum een uitstekende gelegenheid om resultaten van het gevoerde onderzoek voor te stellen aan collega's uit de bank- en verzekeringswereld, en de terugkoppeling te maken aan de hand van problemen vanuit de praktijk. Naast twee gastsprekers (een academicus en een practitioner) kwamen doctoraatsstudenten, postdocs en mensen uit de bedrijfswereld aan bod. In deze publicatie vindt u een neerslag van de voorgestelde onderwerpen. Alle onderwerpen kunnen gesitueerd worden in het ruime gebied van financiële en actuariële toepassingen van wiskunde, maar met een grote variatie: de bijdragen gaan van verdelingen tot schattingen, van benaderingsmethoden naar exacte prijszetting, en van allocatieproblemen naar vraagstukken in verband met efficiëntie.